Feature Extraction for Range Image Interpretation using Local Topology Statistics

Tsz-Wai Rachel Lo
Department of Computing Science
Faculty of Information and Mathematical Sciences
University of Glasgow

A thesis submitted for the degree of
Doctor of Philosophy

January 2009

©Tsz-Wai Rachel Lo, 2009

Abstract

This thesis presents an approach for interpreting *range images* of known subject matter, such as the human face, based on the extraction and matching of local features from the images. In recent years, approaches to interpret two-dimensional (2D) images based on local feature extraction have advanced greatly, for example, systems such as *Scale Invariant Feature Transform* (SIFT) can detect and describe the local features in the 2D images effectively. With the aid of rapidly advancing three-dimensional (3D) imaging technology, in particular, the advent of commercially available surface scanning systems based on photogrammetry, image representation has been able to extend into the third dimension. Moreover, range images confer a number of advantages over conventional 2D images, for instance, the properties of being invariant to lighting, pose and viewpoint changes. As a result, an attempt has been made in this work to establish how best to represent the local range surface with a *feature descriptor*, thereby developing a matching system that takes advantages of the third dimension present in the range images and casting this in the framework of an existing scale and rotational invariance recognition technology: SIFT.

By exploring the statistical representations of the local variation, it is possible to represent and match range images of human faces. This can be achieved by extracting unique mathematical keys known as feature descriptors, from the various automatically generated stable keypoint locations of the range images, thereby capturing the local information of the distributions of the mixes of surface types and their orientations simultaneously. Keypoints are generated through scale-space approach, where the (x, y) location and the appropriate scale σ are detected. In order to achieve invariance to in-plane viewpoint rotational changes, a consistent canonical

orientation is assigned to each keypoint and the sampling patch is rotated to this canonical orientation. The mixes of surface types, derived using the shape index, and the image gradient orientations are extracted from each sampling patch by placing nine overlapping Gaussian sub-regions over the measurement aperture. Each of the nine regions is overlapped by one standard deviation in order to minimise the occurrence of spatial aliasing during the sampling stages and to provide a better continuity within the descriptor.

Moreover, surface normals can be computed from each of the keypoint location, allowing the local 3D pose to be estimated and corrected within the feature descriptors since the orientations in which the images were captured are unknown a priori. As a result, the formulated feature descriptors have strong discriminative power and are stable to rotational changes.

Acknowledgements

I would like to thank my PhD supervisor, Dr. J. Paul Siebert, for his guidance, support and patience throughout the course of this work. Paul introduced me to the world of computer vision four years ago and had been there constantly to guide me through this journey. There had been tears along the way, but his enthusiasm in this work helped me through.

To the rest of my supervisory team: firstly my second supervisor, Dr. Iadh Ounis, for his support, guidance and ideas from a different perspective, away from computer vision. To my third supervisor, Prof. Ashraf Ayoub, for his feedback from a clinical perspective.

To my external and internal examiners, Prof. John Illingworth and Prof. Mark Girolami, for their interest in this work as well as taking the time to study this thesis extensively and their valuable input to improve this thesis. To my convenor, Prof. Muffy Calder, for all the support and encouragement before and during the viva.

I would like to acknowledge (in chronological order) the bodies that funded this work: BBSRC, Tissue Science Laboratories plc, the Department of Computing Science, EPSRC and the Chief Science Office.

To past and present members of the Glasgow Computer Vision & Graphics Group for providing a warm research environment. It was nice to know that I was never alone, especially during those long hours at the office. A big thank you to all my friends, both within and outwith the DCS. It had not been an easy four years, but they had not abandoned me and were always there for me. Special thanks to Ross McIlroy and Stephen Strowes for doing extra bedtime reading for this thesis.

To Craig Macdonald, for all his love, kindness and support over the years. Thank you for travelling down this PhD journey with me and always believing in me, even when I had little faith in myself.

Last, but by no means least, to my parents and "big" sister, for their unconditional love, support and encouragement over the years. I can never thank them enough for everything they have done for me. To quote my sister, without them, life would not have any meanings.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

The main aim of this research project is to develop a methodology for the interpretation of 2.5D range images by extracting stable feature descriptors, which serve to provide point-based correspondences between compared range surfaces. This can be achieved by capturing the underlying surface information from the range images at a set of key correspondences known as keypoints. The methodology adopted here has been inspired by the two-dimensional (2D) Scale Invariant Feature Transform (SIFT) (Lowe, 2004) in which descriptors, comprising the local distribution function of the image gradient orientations, are extracted at each sampling keypoint location over a local measurement aperture. Following an investigation into 2.5D local feature representations, the 2D SIFT concept has been adapted to the 2.5D domain in this work by concatenating the histogram of the range surface topology types and the histogram of the range gradient orientations to form a feature descriptor. These histograms are sampled within a measurement window centred over each mathematically derived keypoint location. Furthermore, the local slant and tilt at each keypoint location is estimated by extracting range surface normals, allowing the three-dimensional (3D) pose of each keypoint to be estimated and used to adapt the descriptor sampling window to potentially provide a more reliable match under out-of-plane viewpoint rotation.

1.1 Aims and Objectives

A range image comprises a 2D matrix in which each element encodes not the intensity of the light focused on an optical imaging sensor, but the distance (or range) of the nearest world surface to each element in the imaging plane (Besl, 1998). Due to this

availability of measurements in the third dimension, this imaging modality provides direct and explicit geometric information, thereby allowing users to interpret the shape of the imaged surfaces directly. Moreover, range images are partially invariant to lighting, pose and viewpoint changes (Gordon, 1992), which confers a number of added advantages over an analysis based on 2D images alone (Bowyer et al., 2006). Accordingly, range images can capture surface shape variation, irrespective of illumination variations (Hesher et al., 2003). These predominant properties have therefore motivated the goals of 2.5D machine understanding of human faces in this research project.

1.1.1 Scientific Questions

At the beginning of this research project, the following questions were posed:

- How to represent local features on a range map?
  - The goal is to formulate a stable feature descriptor that is suitable for representing the range images.
  - A number of representation could be used here, for example the surface types derived from the signs of the mean (H) and Gaussian (K) curvatures, the principal ($k1$ and $k2$) curvatures or the shape index.
- What architecture would best serve local feature based matching in the 2.5D domain?
  - The Elastic Bunch Graph (Wiskott et al., 1997) approach or SIFT algorithm are both potential choices.
- Can local feature matching be achieved in the 2.5D domain?
- Validation of the engine:
  - What methodology to adopt in order to validate matching 2.5D descriptors within a system?
  - To what degree is the system invariant to viewpoint rotational and scale changes?
  - What benefits does this 2.5D technique give over the traditional 2D interpretations?

1.1.2 Motivation

There is currently increasing interest in the use of 3D models for clinical photography, for example, clinical applications now include surface anatomy visualisation and assessment (pre and post surgery intervention) for the face, breast, foot, hand and spine. Conventional methods of surface anatomy assessment are based on the subjective interpretation of the individual surgeons on either the 2D photographs or directly on the patients. With the advent of commercially available surface scanning systems[1] (Figure 1.1) based on digital photogrammetry, clinical photography has been extended into the third dimension. An example of a practical ongoing 3D medical imaging project includes the extraction of geodesic curves on 3D models of children who have had facial deformity, compared to curves extracted from 3D models of children of the same age group who have not had any facial surgery. This process determines the symmetry between the groups, thereby allowing the surgeons to refine their surgical techniques and improve the quality of medical services provided.

(a) 3D Capture System (b) Foot Scanner

Figure 1.1: Examples of 3D scanners: (a) 3D capture system developed by Dimensional Imaging Ltd. (b) Foot scanner developed by Precision 3D Ltd.

Conventional approaches to interpret 2D or 2.5/3D images of surface anatomy are traditionally based on manual landmark placement. Landmarks can provide identification of different anatomical features, for instance on a human face, where examples of

[1]For example: Dimensional Imaging Ltd. (http://www.di3d.com), Precision 3D Ltd. (http://www.precision3d.co.uk), 3dMD Inc. (http://www.3dmd.com)

useful anatomical landmarks include the corner of the eyes and the tip of the nose etc. Given a set of landmarks it becomes possible to compute the Euclidean or Geodesic distances between pairs of landmarks and then to characterise this information in terms of a distance matrix. This distance matrix is capable of capturing the essential configuration statistics of the landmarks. Hence the normal and pathological distributions of the variation of a biological structure can be characterised using multivariate statistical approaches, for example, Principal Components Analysis (PCA). Unfortunately manual designation of landmarks is error prone and subjective, requiring considerable clinical skill. Since manual landmark placement is labour intensive and time consuming, it represents a bottleneck in terms of throughput and compromises the objectivity, if not accuracy, of the placed landmarks. An automated means of locating these key locations would therefore enable large numbers of images to be processed to support large scale clinical protocols such as screening, assessment and diagnosis.

The aim of this research project is to progress machine point-based interpretation of range images of human faces. Initially, these images were captured under highly constrained conditions, where images have been captured with a specific pose and at a specific orientation. These constrains were subsequently relaxed to allow interpretation of a wider range of poses and orientations as the project matured.

1.2 Background

Traditional approaches in machine interpretation of 3D surface manifolds have been based on the classification of the different types of surface topology using differential geometry (do Carmo, 1976). This approach has been widely used since the 1980s, where the signs of the H and K curvatures are used to segment any smooth and differentiable surface into eight surface types (Besl and Jain, 1985; Ittner and Jain, 1985). This approach forms the basis of 3D shape analysis using differential geometry that is now used widely. In the 1990s, Gordon (1992) and Lee and Milios (1990) reported 3D shape analysis based methods to achieve face recognition by means of range images. In this period, the use of local feature descriptors formed by extracting the surface curvatures from 3D images for facial recognition, was also widely reported and examples include work by Chen and Bhanu (2004); Dorai and Jain (1997*b*); Hetzel et al. (2001); Moreno et al. (2003); Wang et al. (2002); Xu et al. (2004*a*). However, most

of the methods mentioned above are limited to the categorisation of single surface types only and typically require a user-defined threshold in order to segment the object with respect to the H, K, the $k1$ and $k2$ curvatures. Since different thresholds are usually required for different types of image, this process falls short of providing fully automated interpretation. Furthermore, little work, until recently (Akagündüz and Ulusoy, 2007; Li and Guskov, 2007; Lo et al., 2006, 2007; Norman et al., 2006; Pansang et al., 2005), has been reported on the invariance limits of viewpoint rotational changes achieved using the above techniques. A more in-depth discussion can be found in Chapter 2.

1.2.1 Local Measurement Technique

In the range image analysis approach presented in this work, a *feature descriptor*, comprises a histogram of the local shape information, based on differential geometry and a histogram of the local image gradients orientation, is extracted over an appropriate *measurement aperture* (sampling window). This feature descriptor is a distinct mathematical key that must be capable of encapsulating the predominant "shape signature" of the underlying surface and be capable of providing sufficient descriptive richness to discriminate between different local surface shapes, while retaining invariance to changes in viewpoint rotation. The size of the measurement aperture is the vital key to the discrimination between feature descriptors, for example if the support region is too small, it may not be capable of capturing enough information to represent the underlying surface, whereas if the support region is too large, the localisation of the feature descriptors could be compromised. Moreover, the sampling measurement aperture is expected to capture a *mixture* of surface types in this work. For instance, the descriptor extracted from the *pronasale* (tip of the nose) keypoint will be dominated by a single surface type whereas the descriptor extracted from the *exocanthion* (outer corner of the eye) keypoint location will be expected to contain a wider mix of surface types. This concept is illustrated in Figure 1.2, showing the unique mixtures of surface types and their relative frequencies (normalised to probability densities for matching purposes) captured at three different locations on a 2.5D face image. Therefore, instead of attempting to segment surfaces into a piecewise patchwork of single surface types, the distributions of the underlying surface topology types are extracted.

By extracting distinct feature descriptors from keypoints locations on a 2.5D range image, it is possible to conduct comparison between and within the population of human

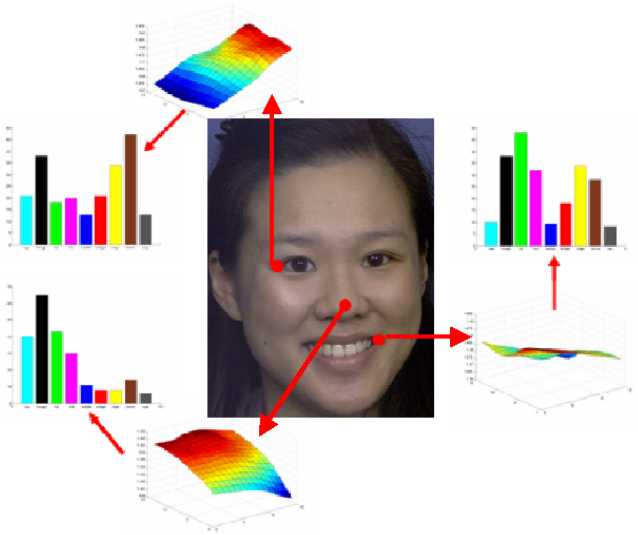


Figure 1.2: Surface types and their histograms (illustrated as bar graphs) extracted from three different keypoint locations on a face range data.

faces depict viewpoints rotational and scale changes, providing an initial matching between the range images based on the descriptors. More information on the rationale and methodology can be found in Chapters 5 and 7 of this thesis. The next section outlines the general approach taken for this work.

1.3 Overview of the General Approach

Figure 1.3 illustrates the overview of the 2.5D SIFT system developed here, from the image capture process to the keypoint matching process.

Stereo-pair images are captured using a single pod system and processed using C3D2 (Ju et al., 2003; Siebert and Marshall, 2000; Siebert and Urquhart, 1994) to produce range images. These range images are then pre-processed in order to suppress random noise. Mathematical keypoints, along with their appropriate scale, σ, are detected on the range images based on Lowe's SIFT methodology (Lowe, 2004) using *scale-space* representation (Lindeberg, 1994*a*,*b*). In the basic formulation, a consistent *canonical orientation*, θ, is assigned to each keypoint location, based on the local image gradient orientation properties, following Lowe's methodology. Subsequently, a more advanced version of the 2.5D SIFT was developed that allocates a *canonical slant* ϕ and *canonical tilt* τ to each keypoint location, thereby allowing the local pose to

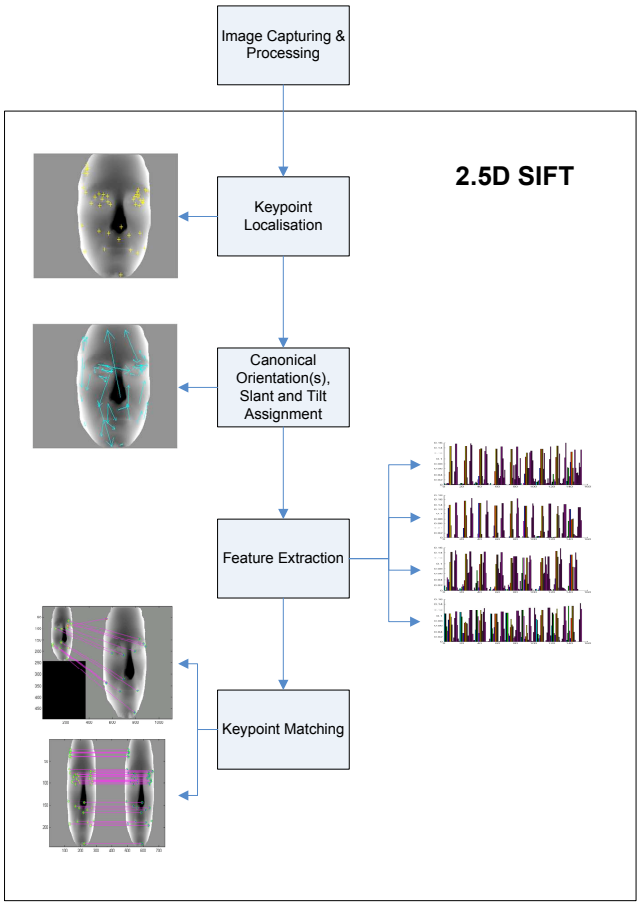


Figure 1.3: Overall approach for this research project.

be estimated and the sampling to be adjusted accordingly. Feature descriptors are then extracted on each (x, y) location, over a measurement aperture of σ defining the scale of the keypoint, using the appointed canonical orientation and the canonical slant and tilt to facilitate invariance to 3D rotational changes. Keypoint matching can then be performed on the extracted feature descriptors from different images, where a candidate match is found using the nearest-neighbour algorithm. False matches are initially rejected using the log likelihood ratio test. In order to verify the matches between two different images, a similarity transform is computed between the two sets of descriptors by means of a Hough Transform. Clusters of matching features with a consistent interpretation are identified. If three or more entries are located in each cluster, an affine transform fitting procedure can be applied to the cluster in order to

recover the affine pose between the matched features and also identify outliers. The details for each stage will unfold in later chapters of this thesis.

1.4 Contributions

This thesis makes the following key contributions to the existing literature:

- The formulation of a stable feature descriptor suitable for representing local features on range images.
- Integration of this feature descriptor within a standard matching framework (SIFT), that affords rotation and scale invariance.
- Feature descriptors are rotationally invariant to viewpoint changes to all three Euler's angles.
- Improvement of the SIFT matching algorithm.

The work presented in this thesis has appeared in the following publications:

- **Tsz-Wai Rachel Lo, J. Paul Siebert, Ashraf F. Ayoub**: An Implementation of the Scale Invariant Feature Transform in the 2.5D Domain. In *Proceedings of MICCAI 2007 Workshop on Content-based Image Retrieval for Biomedical Image Archives: Achievements, Problems, and Prospects*, pages 73-82, 29 October 2007, Brisbane, Australia.
- **Tsz-Wai Rachel Lo, J. Paul Siebert**: SIFT Keypoint Descriptors for Range Image Analysis. Presented in *British Machine Vision Association and Society for Pattern Recognition One Day Symposium: The Inaugural Student Papers Meeting*, 28 March 2007, London, United Kingdom.
- **Tsz-Wai Rachel Lo, J. Paul Siebert, Ashraf F. Ayoub**: Robust Feature Extraction for Range Images Interpretation using Local Topology Statistics. In *Proceedings of MICCAI 2006 Workshop on Craniofacial Image Analysis for Biology, Clinical Genetics, Diagnostics and Treatment*, pages 75-82, 5 October 2006, Copenhagen, Denmark.

The following papers have been submitted for consideration of being published:

- **Tsz-Wai Rachel Lo, J. Paul Siebert**: SIFT Keypoint Descriptors for Range Image Analysis. Submitted to *Annals of the BMVA* for consideration.
- **Tsz-Wai Rachel Lo, J. Paul Siebert**: Local Feature Extraction and Matching on Range Images: 2.5D SIFT. Submitted to *Computer Vision and Image Understanding, Special Issue on 3D Representation for Recognition* for consideration.

1.5 Hypothesis

This thesis argues that by exploiting statistical representations of local surface properties, range images can be represented and matched based on direct recovery of the local surface topology sampled at discrete locations on the range manifold. The use of range data allows surface normals to be recovered directly, such that the pose of the locally sampled descriptors can be estimated and corrected. The above advantages, along with the lighting and pose invariant properties of range images, imply that 2.5D image interpretation techniques can potentially offer improvements over standard 2D techniques.

1.6 Overview of Thesis

This thesis is organised as follows:

- Chapter 2 gives the background and the literature review for this research project.
- Range surfaces analysis is discussed in Chapter 3.
- Chapter 4 introduces the different types of feature descriptors investigated in this thesis.
- The full pipeline of 2.5D SIFT will be introduced and presented in Chapter 5.
- Chapter 6 presents the validation of the 2.5D SIFT.
- The investigation conducted for the 3D pose estimation and correction using surface normals is discussed in Chapter 7.

- Chapter 8 gives the outline and findings for the full validation for this research project, in particular the validation of the pose corrected 2.5D SIFT.
- Finally, Chapter 9 details the contribution of this research project and draws this thesis to a conclusion. Potential future work will be suggested and discussed.

Chapter 2

Background and Literature Review

This chapter discusses the various existing methodologies in the image representation of 2D and 3D in the current literature. Existing feature extraction methods based on shape representation and statistical topology are discussed. Shape analysis is widely used in different disciplines, from science disciplines, for instance computer vision, image analysis, medical, biological, geography and genetics to name a few, to art disciplines in archaeology, theology and museum work etc. In this work, the applications in the computer vision, image analysis and medical field are of interest. This chapter first differentiates the differences between the different imaging modalities, then it focuses on the review of the existing literature on feature extraction of the 3D images, in particular the 2.5D range images. An abundance of literature relating to feature extraction has been produced over the last two decades and therefore a selection of literature has been chosen to be presented in this chapter relating to the feature extraction methodology based on shape analysis. The advantages and shortfalls of the approaches are discussed and the chapter concludes by suggesting ways to improve the existing literature, in order to advance range image feature extraction.

2.1 Image Representation and Interpretations

Traditional image representation is based on 2D intensity images, where the intensity values of light focused on an optical sensor are stored in a 2D matrix (or array), in which each value in the matrix is known as a *pixel*. This image modality is relatively cheap and easy to come by, offering reasonable results for facial recognition for decades.

Zhao et al. gives an extensive review on facial recognition using 2D images (Zhao et al., 2003). However, 2D images are not invariant to lighting, viewpoint, pose and shape variations. As a result, researchers have been looking into a different imaging modality which exploits the third dimension in order to accommodate the shortfalls that 2D intensity images suffer.

Machine interpretation that exploits the third dimension, based on 2.5D range image analysis began in early 1980s. A *range image* (or *depth maps*) comprises a 2D matrix in which each element encodes not the brightness or colour information, but records the depth at which the ray associated with each pixel first intersects the scene observed by a camera (Forsyth and Ponce, 2003). It is a projection via the perspective centre and is therefore in register with the 2D image. In other words, a range image is a large matrix of distance-measurements from the sensor coordinate system to surface points onto the objects in a scene (Besl and Jain, 1986). The measurement unit is normally in metres. For the convenience of the discrete representation based on pixels values, a *heightfield* can sometimes be used to represent range data in which it describes a surface in the Cartesian coordinates (i.e. forms a constant grid of consistent incremental values in the x and y direction, in other words an orthographic projection). Heightfields are assumed to be a piecewise constant function with every pixel corresponding to a height values. A number of measurement techniques can be used to generate range images of human faces, for instance, lasers, depth from shading, texture, motion and stereo. In this work, stereo-pair images are employed to create the range images, based on the distances between the cameras to the real-world surface.

2.1.1 Characteristics of Range Map and Assumptions Made

The range map grid, unlike heightfield (Figure 2.1) which contains uniform (x, y), is of uniform angular sampling (Figure 2.3) and therefore contains perspective in the same manner as standard 2D image (Figures 2.2).

In the standard camera projection model, the distance Z to a point on the surface relates non-linearly to the position of the point projected in the imaging place as follows:

$$\begin{aligned} x &= -f\frac{X}{Z} \\ y &= -f\frac{Y}{Z} \end{aligned} \tag{2.1.1}$$

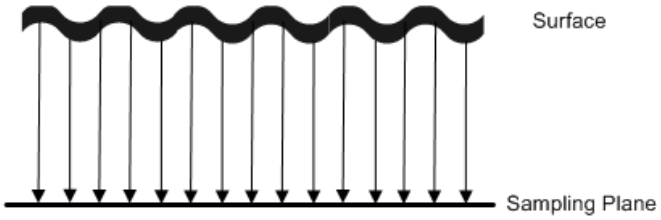


Figure 2.1: Heightfields.

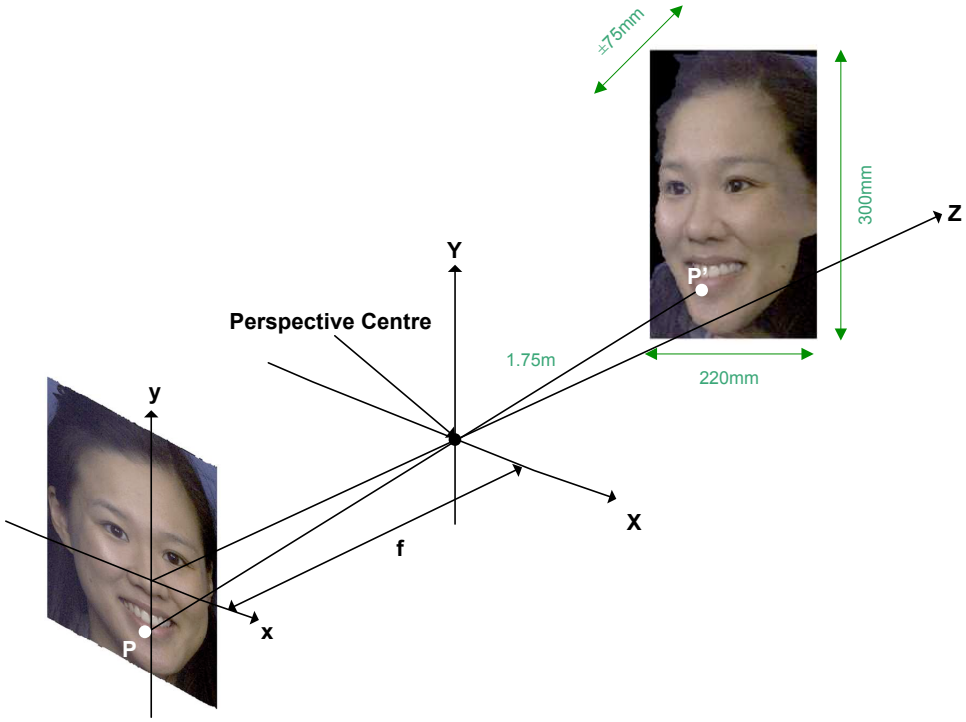


Figure 2.2: Formation of a standard image (for illustration purposes only).

In the configuration used in this work, the lenses which are employed to capture the images have comparatively long focal lengths (180mm, corresponds to a coefficient of variation of $\approx \pm 10°$) and the distance between the camera and the imaged world surface is large (approximately 1.75m), compared to the depth range imaged ($\approx \pm 75mm$). This configuration corresponds to that of a *weak perspective* where the average variation of

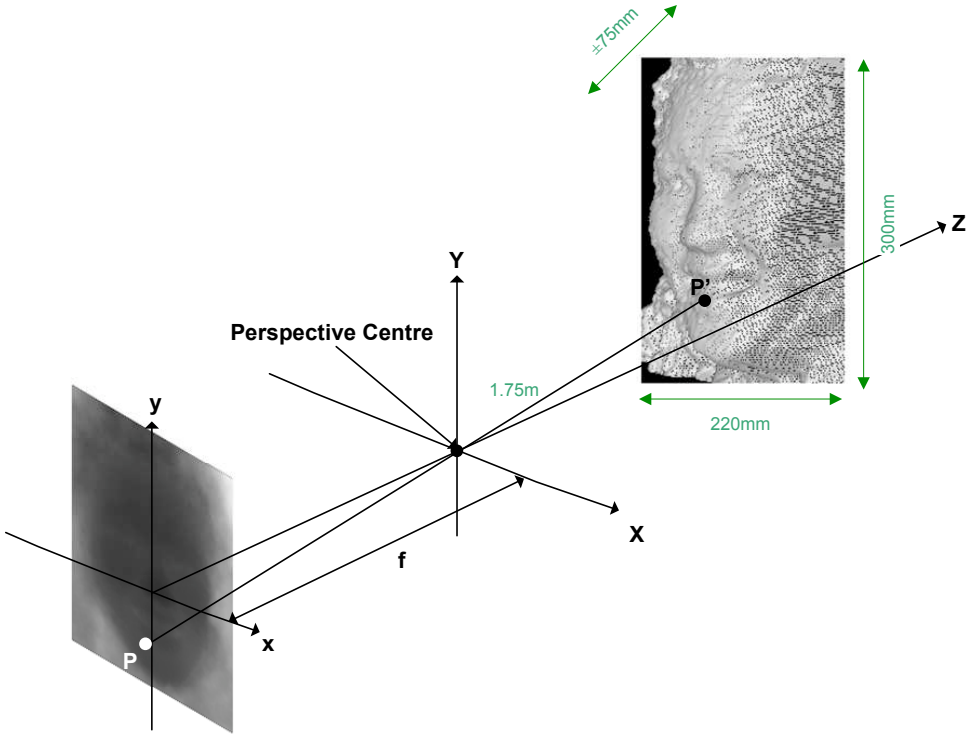


Figure 2.3: Formation of a range image (for illustration purposes only).

the depth of the object (ΔZ) along the line of sight is small compared to the surface mean depth $\bar{Z}$ and the field of view is small. Hence the projection model can be linearised as follows.

Since $\bar{Z} \gg (\Delta Z)$, the weak perspective camera model (linear) applies and therefore the following equations (Trucco and Verri, 1998) hold:

$$\begin{aligned} x &= -f\frac{X}{\bar{Z}} \\ y &= -f\frac{Y}{\bar{Z}} \end{aligned} \tag{2.1.2}$$

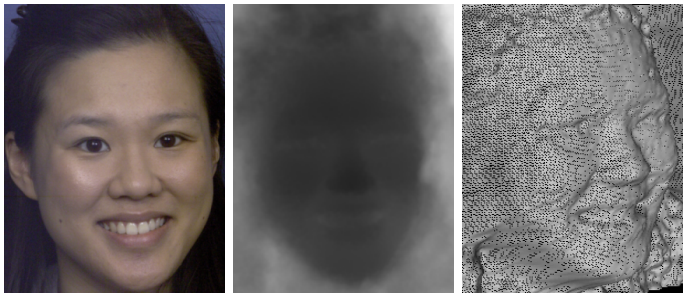
Accordingly, the perspective shifts due to surface variation are ignored here. It should also be noted that the range maps contain exactly the same degree of perspective distortion as that of the standard 2D images formed from the same imaging geometry. Therefore the use of range maps is expected to be no more onerous in terms of matching than standard 2D images.

Since range images can provide an extra dimension (third dimension) compared to 2D intensity image, (i.e. it provides the Z, the depth or the height information, which unlike 2D intensity information, is dependent only on the surface shape (Coombes, 1993)), they are invariant to lighting as each value does not encode the intensity values, and can overcome viewpoint limitations (Hesher et al., 2003; Medioni and Waupotitsch, 2003). Furthermore, Gordon (1992) stated that by using curvature descriptors extracted from range images, it is possible to describe the 3D surface, thereby achieving higher accuracy than using 2D intensity images alone. Given the increasing availability of this imaging modality as well as range data having an inherently greater degree of invariance to illumination and greater resilience to variation in facial expressions (i.e. pose), makeup, disguise and aging (Chellapa et al., 1995), a range image based analysis confers a number of advantages over any analysis based on 2D images alone (Bowyer et al., 2006; Moses et al., 1997; Xu et al., 2004*b*).

Further representation of 3D images includes a point cloud, which is a set of Cartesian coordinates in a 3D space, derived either from one range image or several range images once they are transformed into a common coordinate system. A point cloud is capable of providing a whole view of the 3D object, compared to range images which show the distance metric. However, the connectivity between the 3D points or the neighbourhood relationships and the normals are not explicit in a point cloud. Therefore it is difficult to study the shape of 3D surfaces or extract features, such as curves, directly from a point cloud. Polygon meshes are another popular method to describe any 3D object. They can be obtained from range images or point clouds by triangulation. Polygon meshes can describe the shape of any free-form object accurately. However, polygons are not in a matrix form and do not contain uniform topology, therefore interpretations using polygons are cumbersome.

Figure 2.4 shows an example of an intensity 2D image, a 2.5D range image and a 3D polygon mesh.

Measurements can be performed directly on the range images using differential geometry. In order to perform range image analysis, landmarks (or *keypoints*, see Section 2.2.2) are usually used to identify areas of interest and subsequently extracting the relevant information from these locations for further measurements. The next section details the types of landmarks available.



(a) 2D (b) 2.5D (c) 3D

Figure 2.4: Examples of (a) a 2D intensity, (b) a range image and (c) a 3D polygon mesh (without texture).

2.2 Landmarks

According to Dryden, a landmark is "a point of correspondence on each object that matches between and within populations" (Dryden and Mardia, 1998). Therefore landmarks define set locations which provide the means to compare sparse samples on range surfaces. There exists three types of landmarks: anatomical, mathematical and pseudo landmarks (Bookstein, 1997).

2.2.1 Anatomical Landmarks

An anatomical landmark is assigned by an expert, usually clinicians, and these correspond between organisms in some biologically meaningful way. Medical terms are usually used to identify different types of anatomical landmarks, for example *pronasale* (tip of the nose) and *exocanthion* (the outer corner of the eyes). Figure 2.5 shows an example of a face annotated with anatomical landmarks (Ferrario et al., 1998). Anatomical landmarks are useful and meaningful, however the placement of anatomical landmarks is subjective, error-prone and extremely time consuming.

2.2.2 Mathematical Landmarks

Mathematical landmarks are points located on an object according to some mathematical or geometrical property of the figure, for example, at a point of high curvature, usually calculated using differential geometry. The use of mathematical landmarks is extremely useful in automatic recognition and/or analysis. Mathematical landmarks are also known as *keypoints*.

Figure 2.5: Example of anatomical landmarks.
(Ferrario et al., 1998)

2.2.3 Pseudo Landmarks

Pseudo landmarks are constructed points on the object, either around the outline or in between anatomical or mathematical landmarks. These are useful in matching surfaces as it provides statistical analysis of the surface.

Keypoints are useful to provide an initial location of meaningful locations on the surface, while shape representation can be used to differentiate between different keypoint locations. There are many methodologies for describing a surface using differential geometry and these are discussed in the next section.

2.3 3D Shape Representation

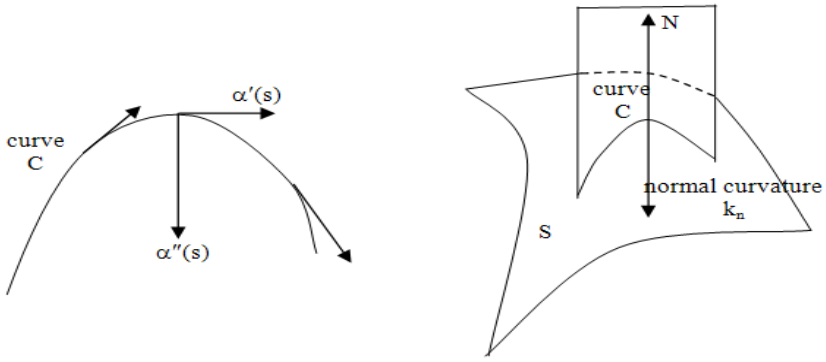
3D shape-based object and face recognition can be classified into different categories:

1. Curvature based methods
2. Principal Component Analysis (PCA) based methods
3. Point set (or point signature) based methods
4. Feature descriptors based methods
5. Multi-modal using 2D and 3D techniques

Combination of different shape-based analysis is often found to be beneficial in 3D object and face recognition systems. For instance, range images can be segmented into different sections using differential geometry, allowing feature descriptors to be extracted from each section, as shown in work of Gordon (1992).

2.3.1 Basic Concepts

From the basic concept of 2D curvature, this can be extended to the 3D domain in which shape can be represented using curvatures. Figure 2.6 illustrates the definition of curve curvatures.



(a) Curve Curvature (b) Normal Curvature

Figure 2.6: The definition of curvature: (a) curve curvature; (b) normal curvature.

Definition 2.3.1 Curve Curvature: Let α: $I \rightarrow R^3$ be a curve parameterised by arc length $s \in I$. Then $|a''(s)| = k(s)$ is called the curvature of the curve a at s.

Definition 2.3.2 Normal Curvature: The normal curvature is the curvature of any curve lying in an intersection plane which contains the surface normal direction at that given point.

Definition 2.3.3 Principal Curvatures: The principal curvatures ($k1$ and $k2$) are the two extremes of normal curvature at the given point, namely the maximum and minimum curvatures. Their corresponding directions are known as principal directions and are orthogonal.

Definition 2.3.4 Mean Curvature: The mean curvature (H) is the average of the principal curvatures:

$$H = \frac{k1 + k2}{2} \tag{2.3.1}$$

Definition 2.3.5 Gaussian Curvature: The Gaussian curvature (K) is the product of the principal curvatures:

$$K = k1 \times k2 \tag{2.3.2}$$

In addition, $k1$ and $k2$ can be obtained from H and K using Equation 2.3.3:

$$k1, k2 = H \pm \sqrt{H^2 - K} \tag{2.3.3}$$

where $k1 \geq k2$.

Definition 2.3.6 Umbilical Point: An umbilical point is a point on a surface at which the curvature is the same in two or more directions. In other words,

$$k1 = k2 \tag{2.3.4}$$

Definition 2.3.7 Shape Index: The bounded $[-1, 1]$ shape index, S (Koenderink and van Doorn, 1992), is used to measure quantitatively the shape of the local surface, derived by $k1$ and $k2$ curvatures and is invariant to curvatures. This representation is fundamental in this work since it is capable of characterising the underlying shape and is a pure manifestation of the object, mapping directly to human perspective. Figure 2.7 shows the nine shapes represented by the shape index scale.

$$S = \frac{2}{\pi} \tan^{-1} \left[\frac{k2 + k1}{k2 - k1} \right] \tag{2.3.5}$$

Definition 2.3.8 Degree of Curvedness: The degree of curvedness (Koenderink, 1990), R, is defined as:

$$\begin{aligned} R &= \sqrt{2H^2 - K}, \quad \text{or} \\ R &= \sqrt{\frac{(k1^2 + k2^2)}{2}} \end{aligned} \tag{2.3.6}$$

More information and details on differential geometry can be found in do Carmo (1976). The next section discusses the current literature of face recognition of range images using the above defined curvatures methodologies.

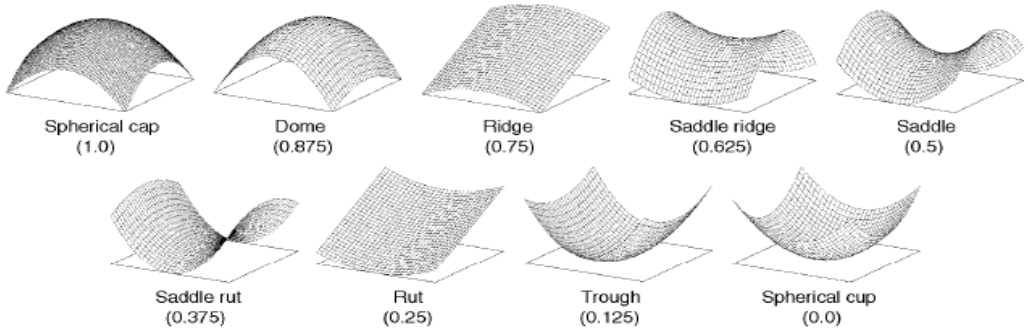


Figure 2.7: The nine shapes represented by different values of the shape index . (Dorai and Jain, 1997*a*)

2.4 Curvature Based Methods in Object Recognition

This section reviews the literature based on curvature measurements of 3D images, in particular range images of human faces. The description of 3D facial surfaces using curve curvature dates back to 1926, where the renowned German mathematician Klein marked out parabolic curves in an attempt to put facial aesthetics on a mathematical foundation (Hilbert and Cohn-Vossen, 1952). However, these curves are deemed to be unsuitable for producing a robust description of the face since the extracted curves are complex (Brady et al., 1985).

2.4.1 Classical Techniques - Prior to this Project

Machine interpretation of 3D surfaces, based on the classification of surface types, has been widely used since the 1980s. Ittner and Jain (1985) investigated the effectiveness of six different curvature measurements in view of the identification of four surface primitives, namely sphere, plane, cylinder and cone. These six measurements involved were the average curvature, the principal curvatures, the mean curvature, the Gaussian curvature and the curvature ratio. Fan et al. (1986) calculated the surface curvature in four arbitrary different directions and obtained the curvature extrema and zero-crossings for each of these one-dimensional curves, which were used to detect surface and depth discontinuities. Cartoux et al. (1989) segmented a range image based on the principal curvature and found a plane of bilateral symmetry through the face in order

to normalise the pose. Yokoya and Levine (1989) segmented range images using the H and K curvatures.

In late 1980s, Besl and Jain (1985, 1988) segmented the surfaces into eight types, namely peak, ridge, saddle ridge, flat (plane), minimal surface, pit, valley and saddle valley, based on the signs of the mean and Gaussian curvatures, as shown in Table 2.1 (Jain et al., 1995). These surface types can be used to describe any smooth and differentiable surface. This method forms the basis of 3D shape analysis using differential geometry that are now used extensively.

	$K(i,j) > 0$	$K(i,j) = 0$	$K(i,j) < 0$
$H(i,j) < 0$	Peak	Ridge	Saddle Ridge
$H(i,j) = 0$	**UNDEFINED**	Flat	Minimal Surface
$H(i,j) > 0$	Pit	Valley	Saddle Valley

Table 2.1: Eight surface types, based on the signs of H and K curvatures.

Since then, investigation has been conducted on the representation of the human face using features based on shape and curvature of the face surface. Based on the surface types derived from the signs of H and K curvatures, Lee and Milios (1990) segmented the range images of the human face and matched convex regions of different individuals, instead of using the entire face, for recognition. In early 1990s, Gordon (1991, 1992) proposed the use of the principal curvatures in order to segment a facial range image using ridge and valley lines. She defined ridge lines as the local maxima in the principal curvature along the line of maximum curvature while valley lines as the local minima in principal curvature along the line of minimum curvature, as shown in Figure 2.8. Thresholds have been applied to the $k1$ and $k2$ curvatures, with $k1$ greater than a certain threshold while using $k2$ less than a certain threshold, producing a much cleaner and more stable classification, compared to Haralick's work (Haralick, 1983) on extracting ridges and valleys on 2D intensity images. Tanaka et al. (1998) used curvature-based segmentation and represented the face using an extended Gaussian image (EGI). Wang et al. (2000) presented a new shape-based approach for 3D brain surface correspondence using geodesic paths and geometrical feature.

More recently, Kim et al. (2001) proposed the use of range image with real-time normalisation and feature extraction using one of the curvature characteristics, the principal curvatures. Segmentation on the faces was performed based on thresholded

(a) Ridge Lines (b) Valley Lines

Figure 2.8: (a) ridge lines, local maxima of k1 and (b) valley lines, local minima of k2. (Gordon and Vincent, 1992)

principal curvature values as shown in Figure 2.9. Features are then extracted by matching the templates of mouth, nose and eye filter to the segmented facial range data. Campbell and Flynn (2002) also segmented 3D objects into different regions in which the shape index and the degree of curvedness were employed as region attributes in order to compare a specific object with different prototypes for recognition. In 2003, Moreno et al. (2003) used the signs of the H and K curvatures as point classification for isolating regions of pronounced curvatures. They employed three point types, namely hyperbolic points ($K > 0$), convex points ($H < 0$ and $K > 0$) and elliptic concave points ($H > 0$ and $K > 0$). In order to isolate regions of high curvature avoiding points in which low curvature were obtained, they experimented with different curvature thresholds and chose to use $H_t = \pm 0.05$ and $K_t = \pm 0.005$ in which the following segmented regions of a face was achieved as shown in Figure 2.10.

2.4.2 Contemporary Work Subsequent to the Start of this Project

Lee and Shim (2004) proposed an implementation of 3D recognition based on the "depth-weighted Hausdorff distance" using the surface curvatures extracted from 3D images of the human faces. Bhanu and Zhou (2004) used a curvature-based approach for fiducial extraction while Lu and Jain (2005) used the local shape index to find similar points between images. Kim et al. (2005) used curvatures to localise the nose.

Colombo et al. (2006) detected salient face features such as the eyes and nose,

Figure 2.9: Segmentation results of facial range images based on principal curvatures $(k1, k2)$ (black: $k1 > 0.5$, grey: $k1 < 0$ and $-0.3 < k2 < 0$, white: $k1 < 0$ and $k2 < -0.3$).
(Kim et al., 2001)

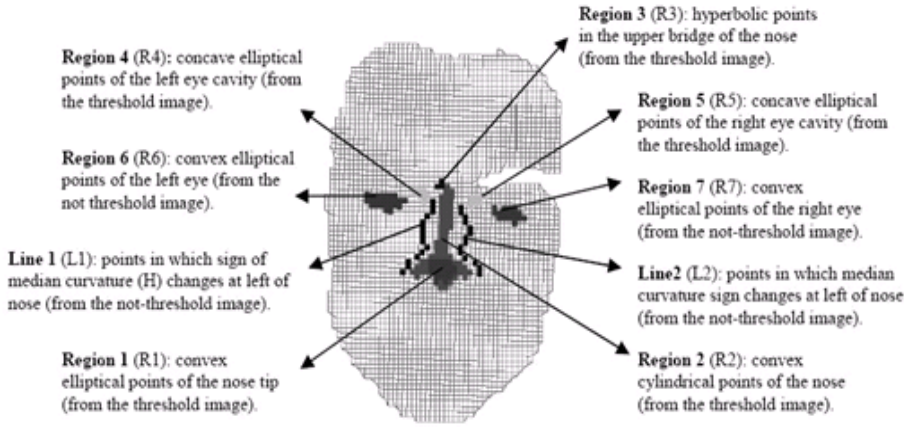


Figure 2.10: Segmented regions of a face.
(Moreno et al., 2003)

through an analysis of the curvature of the surface as shown in Figure 2.11. Lu et al. (2004, 2006) used the shape index to segment the range images of human faces as shown in Figure 2.12. Lee et al. (2006) employed the signs of the principal curvatures to detail the characterisation of the face. Chang et al. (2006) detected the nose tip, eye cavities and the nose bridge using the signs of H and K curvatures and user-defined thresholds as shown in Figure 2.13. Sun and Yin (2006) selected features from range images of faces using principal curvatures. Gökberk et al. (2006) reviewed and

compared a few 3D face registration and recognition algorithms. They compared the use of 3D point coordinates, surface normals, curvature-based descriptors, range images and facial profile curves, all extracted from 3D shape information.

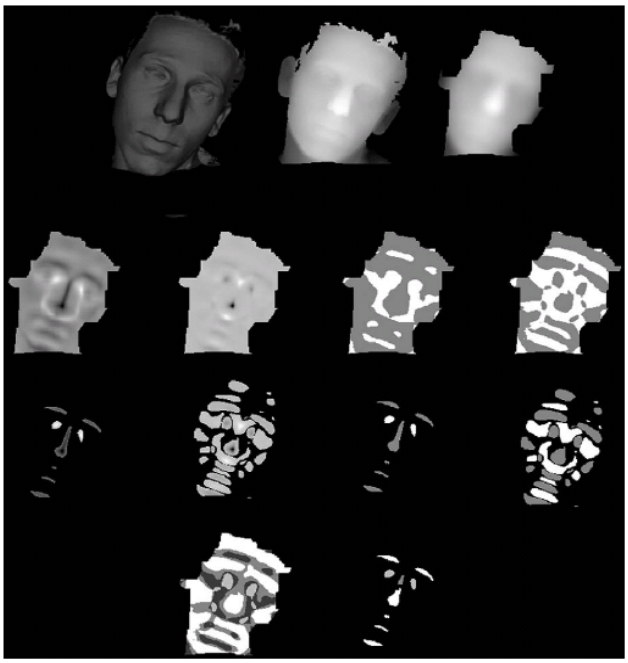


Figure 2.11: The analysis of the curvature of the face surfaces for the detection of salient face features. This first row of this figure illustrates the polygon model, projected range image and its smoothed version. The second row shows the H and K curvature maps where the darker the zone, the higher the curvature regions. Third row presents the thresholded H and K maps and the final row shows the HK-Classification maps where the face is thresholded into different regions, according to the signs of the H and K. (Colombo et al., 2006)

Further example of uses of curve curvatures include McFarlane et al. (2005) in which they used the signs of H and K curvatures in order to segment the range images of pigs into different regions.

The use of curve curvature based methods to interpret 3D images has many advantages, for instance measurements can be conducted on the range images directly and the measurements have the intrinsic properties of describing the local shape. Furthermore, curvature based methods provide a good pictorial descriptor of the surface segmentation when combined with an appropriate colour coding scheme. However, this methodology suffers from sensitivity to the noise that exists in the range data and

Figure 2.12: A face that has been segmented using the shape index. The darker the regions the lower the shape index value whereas the lighter the regions the higher the shape index values.
(Lu and Jain, 2006)

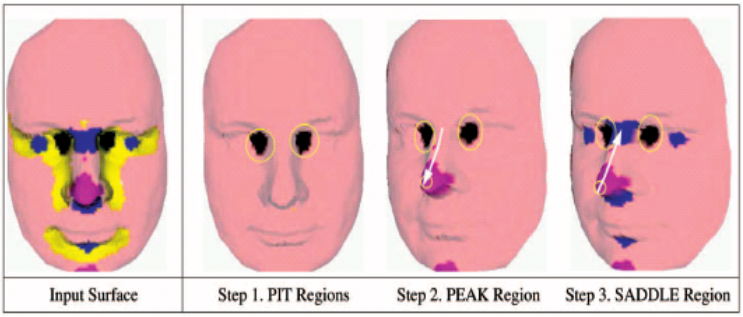


Figure 2.13: Using the signs of H and K curvatures to detect area of interests.
(Chang et al., 2006)

therefore pre-processing techniques such as smoothing are carried out on the range images prior to the calculations of the curvatures.

Apart from using curve curvatures for 3D image interpretations, other methods are discussed in the following sections.

2.5 Principal Component Analysis (PCA) Based Methods

Other approaches for 3D face recognition include the extension of conventional dimensionality reduction techniques, for instance PCA, to the 3D images or a combination of 2D intensity and 3D images (see Section 2.8). Examples of work that employed PCA

include Achermann et al. (1997), Mavridis et al. (2001), Chang et al. (2003), Hesher et al. (2003), Heseltine et al. (2004), Pan et al. (2005) and Russ et al. (2006).

2.5.0.1 Principal Component Analysis

Principal component analysis (PCA) is a vector space transform often used to reduce multidimensional data sets to lower dimensions for analysis purposes. The new features are linear functions of the old features designed so that a lower dimensional linear subspace contains the majority of variation of the data points from their mean (Forsyth and Ponce, 2003). It is important for the lower dimensions data set to have an accurate representation of the original data set so that no important data is lost. This method, known as the Karhuenen-Loéve transform, is a classical technique from statistical pattern recognition (Duda and Hart, 1973; Duda et al., 2000; Fukunaga, 1990; Oja, 1983). The method is described as follows:

Let x_i be a set of N column feature vectors in $\mathbb{R}^d$. The mean of the dataset, μ, is obtained:

$$\mu = \frac{1}{N}\sum_{i=1}^{N} x_i \tag{2.5.1}$$

The covariance matrix, C_x, of the population is estimated by:

$$C_x = \frac{1}{N-1}\sum_{i=1}^{N}(x_i - \mu)(x_i - \mu)^T \tag{2.5.2}$$

The mean can then be used as an origin and the offsets from the mean $(x_i - \mu)$ can be studied. Since the PCA features are linear combinations of the original features, it is possible to consider the projection of these offsets onto various different directions. A unit vector v represents a direction in the original feature space and it can be interpreted as a new feature $v(x)$. The value of u on the ith data point is given by $v(x_i) = v^T(x_i - \mu)$. A good feature captures as much of the variance of the original dataset as possible. Note

that v has zero mean and the variance of v, $var(v)$ is as follows:

$$\begin{aligned} var(v) &= \frac{1}{N-1}\sum_{i=1}^{N} v(x_i)v(x_i)^T \\ &= \frac{1}{N}\sum_{i=1}^{N-1} v^T(x_i-\mu)(v^T(x_i-\mu))^T \\ &= v^T\left\{\sum_{i=1}^{N-1}(x_i-\mu)(x_i-\mu)^T\right\}v \\ &= v^T C_x v \end{aligned} \tag{2.5.3}$$

It is necessary to maximise $v^T C_x v$ subject to the constraint that $v^T v = 1$. This can be solved by finding the eigenvector of C_x corresponding to the largest eigenvalue. The data can be projected onto a space perpendicular to the eigenvector and as a result a collection of $d-1$ dimensional vectors are obtained. Therefore the eigenvectors of C_x, $[v_1, v_2, ..., v_d]$, where the order is given by the size of the eigenvalue in descending order (i.e. v_1 has the largest eigenvalues), gives a set of features with the following two properties:

1. They are independent as the eigenvectors are orthogonal and
2. Projection onto the basis $v_1, v_2, ..., v_k$ gives the k-dimensional set of linear features that preserves the most variance.

In MATLAB, the function `princomp` gives the principal-component vectors and related quantities (Gonzalez et al., 2004).

2.6 Point Set Based Approaches

In order to standardise face pose, researchers often locate some common features on the 3D images, this methodology is known as the point-set (or point signatures) based method. For example, Nagamine et al. (1992) found five feature points around the face and used these feature points to match various curves or profiles through the face data. Chua et al. (2000); Chua and Jarvis (1997) used "point signatures" for 3D face recognition in which they locate reference points on the faces in order to standardise different pose and facial expressions. Achermann and Jiang (2000) applied an extension of Hausdorff distance matching for range images matching. Pan et al. (2003) presented

an automatic 3D face verification approach by first registering and then comparing the range data using a Hausdorff distance approach and a PCA based approach. Xu et al. (2004*a*) converted the original 3D point cloud data into a regular mesh and located the nose region and used it as an anchor to find other local regions.

More recent and notable 3D face recognition methodologies using point-set based approach include work by Bronstein et al. (2005) in which they addressed the problem of facial expressions explicitly. Here, they converted the 3D face data to an *eigenform* that is invariant to the type of shape deformation that is modelled. Chang et al. (2005) described a "multi-region" approach to 3D face recognition in which overlapping subregions around the nose are independently matched using iterative closest point (ICP).

The point-set approach is similar to the well known 2D approaches such as SIFT (Lowe, 2004) and Elastic Bunch Graph (EBGM) (Wiskott et al., 1997). However, it is not possible to apply these approaches directly onto 3D images without any modifications since the nature of 2D images are very different from 3D images. Al-Osaimi et al. (2007) presented a methodology for extracting interest-points for range images in which a range image is represented by two images with blob-like patterns that have easily detectable peaks and can be efficiently extracted using a convolution kernel.

2.6.1 Scale Invariant Feature Transform

Lowe derived a 2D method for image feature generation called Scale Invariant Feature Transform (SIFT) (Lowe, 1999, 2001) in which it transforms an image into a large collection of local feature vectors, each of which is invariant to image translation, scaling and rotation. This approach was described in details in his later paper (Lowe, 2004).

A feature should be invariant to image scaling and rotation and should also be partially invariant to changes in illumination and 3D camera viewpoint. In addition, features should be highly distinctive. There are four major stages in SIFT in order to generate a set of image features. First of all, a scale-space detection was employed to identify potential interest points that are invariant to scale and orientation. This can be achieved using a Laplacian-of-Gaussian (LOG) (Lindeberg, 1994*b*; Mikolajczyk, 2002; Mikolajczyk and Schmid, 2004) difference-of-Gaussian (DOG) function. A pyramid was built for each image in order to obtain a multi-scale representation. The second

key stage was keypoint localisation in which a detailed fit is performed of the identified potential candidate to its nearby data for location, scale and the ratio of principal intensity curvatures. Keypoints were then selected based on measures of stability. The third step was orientation assignment in which orientation histograms (36-bins) are computed within a region of the keypoints. The final step was to extract a keypoint descriptor in which a Gaussian weighing function is used to assign weight to the magnitude of each sample point in order to avoid sudden changes in the descriptor with small changes in the position of the window. A subregion of a window of size 4×4 and eight directions was applied to each orientation histogram. As a result, a 128-element (4×4×128) feature vector was created for each keypoint and this feature vector is invariant to image scale and rotation. Figure 2.14 shows a toy example of a keypoint descriptor extracted from a 2×2 descriptor array computed from an 8×8 set of samples.

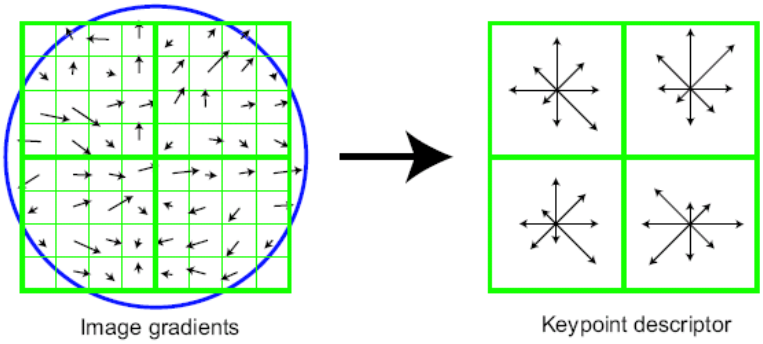


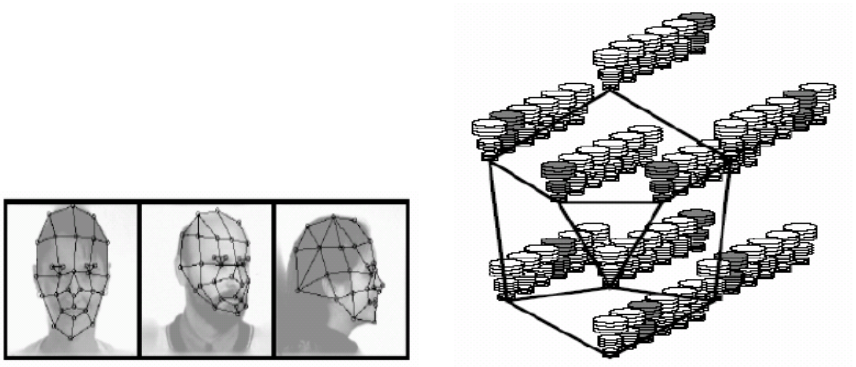
Figure 2.14: SIFT Keypoint descriptor created by accumulating image gradient magnitudes and orientations weighted by a Gaussian window. Gradients are binned into orientation histograms over sub-regions in the descriptor (right).
(Lowe, 2004)

This can be applied to object recognition in which each potential keypoint is matched independently to a database of keypoints extracted from the training images.

2.6.2 Elastic Bunch Graph

Wiskott et al. (1997) proposed a 2D face representation and recognition system in which they described manually selected fiducial points on the face as a labelled graph consisting of N nodes. These nodes are labelled as sets of *Gabor jets*. A Gabor jet comprises of responses of local visual content to scale and rotated versions of a *mother*

wavelet. The jets are then collected in a stacked data structure called a *face bunch graph*. The face bunch graph has a node for every landmark on the face. Every node is a collection of model jets for the corresponding landmark (a *bunch*). The face bunch graph serves as a database of landmark descriptions. Once the face graph is created, the image is discarded, and the face graph becomes the internal representation of that image. As a result, the similarity between two face graphs (representing two face images) can be computed. There are two ways to measure the similarity of the face graphs, first by comparing the geometry of the graph (landmark locations) and secondly by comparing the similarity of the Gabor jets (landmark jets). Figure 2.15 shows the location of the fiducial points registered onto a face image and the Elastic Bunch Graph representation of a face with Gabor jets. This algorithm was tested using the FERET (Phillips et al., 2000) database of faces with different expression and poses.



(a) Location of fiducial points (b) Elastic Bunch Graph

Figure 2.15: (a) Locations of fiducial points registered on to a face image with different pose orientations. (b) Elastic Bunch Graph representation of a face with Gabor jet responses at different orientations and scales centred at fiducial points.
(Wiskott et al., 1997)

2.7 Feature Descriptors Based Methods

In order to characterise and match 3D objects, it is possible to encode information of the underlying shape surface in the form of a feature descriptor, otherwise known as a feature vector. A feature descriptor is a numerical representation of the objects, usually represented in the form of 1D or 2D histograms. For example, curvature information

can be encoded in the form of feature descriptors in order to characterise and match 3D objects. This section details some of the works in the literature that employ feature descriptors extracted from 3D images to represent and match 3D objects, in particular of human faces. The idea of using a feature vector to describe a feature became more popular in recent years.

2.7.1 Typical Classical Techniques used Prior to the Start of this Work

Histograms, which are the count of the quantised values of the H and K curvatures, were calculated to construct the curvature signature for 3D objects, as presented by Mustafa et al. (1999). These were combined with surface colour signature to match the objects in the scene to the training models for identification. Hetzel et al. (2001) created a set of local features of 3D objects based on the pixel depth, the surface normals and the shape index, represented in a multi-dimensional histogram. Moreno et al. (2003) first segmented the face using Gaussian curvatures, as described in Section 2.4 above, then created a feature vector based on these segmented regions.

Further examples of local 3D keypoint descriptors include point signatures (Chua et al., 2000; Chua and Jarvis, 1997) in which "signatures" are extracted from arbitrary points and these signatures are used to vote for models with similar signatures. For a given point, a contour on the surface is defined around the point of interest. Each point on the contour may be characterised by the signed distance from the point of interest to a point on the contour and a clockwise rotation from the reference vector about the normal. However, since local representations of 3D surfaces can be sensitive to noise, which can affect the features derived from differential quantities such as curvatures and surface normals. As a result, many new recognition systems have adopted geometric representations which combine local and global representations together, examples include spin images (Johnson, 1997; Johnson and Hebert, 1999) and COSMOS (Dorai and Jain, 1997*a*).

2.7.2 Contemporary Work Subsequent to the Start of this Project

Chen and Bhanu (2004, 2007) computed 2D histogram consisting of the shape index and the angles between the normal of a reference point and of its neighbours of 3D objects. Lee et al. (2005) proposed an approach to 3D face recognition based on the

curvatures values at eight feature points on the face. Huang et al. (2006) performed face recognition by combining the global statistics of geometrical features with the local statistics of the correlative features of facial surfaces (using shape index).

Li and Guskov (2007) tackles the difficult recognition problem caused by viewpoint rotations by extracting local shape feature descriptors from surface patches detected by salient features. These feature descriptors were matched with a pyramid kernel function. Akagündüz and Ulusoy (2007) extracted feature descriptors using the H and K curvatures from various data points at different scales, allowing face recognition to be performed at a more accurate level.

The above feature descriptors are based on the use of curve curvature as part of the feature descriptors. The following describes a few of the works in the literature which employ other information as part of the feature descriptor. Lee et al. (2003) formed a feature vector based on the contours along the face at a sequence of depth values, having first identified the nose tip. As mentioned in Section 2.6, Xu et al. (2004*a*) first located the nose region and used it as an anchor to find other local regions of mouth, nose, left and right eye, and formed a feature descriptor around the these local regions. Shan et al. (2006) used histograms of shape signature or prototypical shapes called *shapemes* to recognise partially observed query objects.

More recently, Mian et al. (2007) first identified stable keypoints on the range images of faces, then extracted local features by fitting a surface to the neighbourhood of a keypoint and sampling it on a uniform grid. Huang et al. (2007) performed face recognition using the statistics of the differences between the pixels values of different range images.

Feature descriptors play an important role in machine interpretation of 3D images as it characterises the 3D information into a simple descriptor form which could be used to represent the surfaces. Since 3D shape does not depend on the illumination and pose, it allows the interpretation to be conducted directly on the surfaces. However, 2D cues do have their advantages as they are generally cheaper to compute and easier to come by. The next section describes some of the existing literature in the attempt to combine 2D and 3D modalities together to achieve machine interpretation of images.

2.8 2D and 3D Multi-modal/Fusion Methods

Research for machine interpretation of 3D images of human faces dates back to 1980s, however the combination of 2D and 3D methodologies did not appear until around 2000. By using the geometry information available in 3D data along with 2D information, it is possible to cope with the pose and illumination variations that inhibit the performance of 2D face recognition (Malassiotis and Strintzis, 2005). This section discusses some of the methodologies in combining 2D and 3D algorithms together.

Following from the work conducted on real-time profile recognition by Beumier and Acheroy (1997), they used a central profile and a lateral profile in both 2D and 3D and approached multi-modal recognition by using a weighted sum of 3D and 2D similarity measures (Beumier and Acheroy, 2001).

Moving away from profile measurements, Wang et al. (2002) combined features extracted from both range data and intensity images together, where feature points were described by Gabor filter responses in the intensity images and point signatures in the range data. Bronstein et al. (2003) proposed the use of multi-modal 3D and 2D recognition using eigen decomposition of flattened textures and canonical images. This approach used an isometric transformation to 3D face analysis in order to cope with the variation due to facial expression. Tsalakanidou et al. (2003) used range images combined with colour images (instead of using intensity alone) to attempt multi-modal face recognition. Chang et al. (2003) reported on a PCA-based recognition by combing 2D and 3D using a weighted sum of the distances from the individual 3D and 2D face spaces. Similarly, Godil et al. (2004) employed PCA for the matching of 2D and range image together while Papatheodorou and Reuckert (2004) conducted multi-modal 2D and 3D face recognition using a generalisation of ICP based on point distances in a 4D space of $(x, y, z$,intensity). This approach integrated shape and texture information at an early stage as opposed to other work presented hitherto, where the decision was made independently for each modality and these were then combined together.

Heseltine (2005) introduced a 3D face database providing 3D texture mapped face models, as well as 2D images captured at the same instant, allowing a direct comparison of 3D and 2D techniques. Different methodologies in system combination were tested, including combination by dimensional accumulation, elimination and genetic selection.

As a result, 2D and 3D data can be combined effectively and producing good error rates, with a clear advantage over single subspace systems.

Hüsken et al. (2005) presented strategies for the fusion of 2D and 3D face recognition algorithms. Here the matching of each modality is conducted independently and the fusion of the results from the two modalities is done at the score level. They claimed the benefit of this approach is determined by two aspects: the accuracy of the single algorithms to be combined and their statistical independence as the error in one of the algorithms could be counterbalanced by the other modality. Maurer et al. (2005) combined 2D and 3D face recognition by means of a weighted sum rule to fuse the results from each modality, with one exception that if the shape score is very high, the texture score is subsequently disregarded.

Most recently, Mian et al. (2008) presented a feature-based algorithm for the recognition of textured 3D faces where they have combined features extracted from 2D and 3D domain. 2D features were extracted using standard SIFT while 3D features were extracted by fitting a surface to the neighbourhood of a keypoint and sampling it on a uniform grid. These features are combined at the feature and score-level.

Building from the previous work with ICP style matching of 3D shape by Lu and Jain (2005), they created a 2D and 3D multi-modal system using linear discriminant analysis (LDA) for the 2D matching component. BenAbdelkadera and Griffin (2005) demonstrated that by using both 2D texture and 3D depth modality improved recognition rate. However, these results were based on high-resolution data captured within a controlled environment and on manual marked facial feature points.

J. Cook et al. (2006) combined 2D and 3D face recognition by using Log-Gabor Templates for decomposing the image into different sub-regions. Mian et al. (2006*a*) presented an algorithm which used 2D and 3D multi-modal local features. 3D local features were extracted based on their previous work on tensor representation (Mian et al., 2006*b*) which makes the features invariant to pose, while the 2D local features were extracted using Lowe's SIFT (Lowe, 2004). The results of the 2D and 3D local features are fused at the rank level using a confidence weighted sum rule.

More recently, Riccio and Dugelay (2007) claimed that working in 3D is more costly, compared to using 2D alone, therefore they proposed the enrolment in 3D but identification performed from 2D images. Ansari et al. (2008) presented a multi-modal approach for 3D face modelling and recognition using 2D and 3D information. They

used the 2D data to compute the disparity map and extracted facial features points and used these information to build a 3D model which incorporated the depth information with the facial features labels, as well as the profile information.

In order to combine 2D and 3D analysis together, most of the work presented here separated each of the modality and the fusion occurred at the matching stage with the exception of work by Godil et al. (2004). Moreover, since the different modality can offer different information, a fusion rule is often applied at the matching stage. For instance, in the work by Maurer et al. (2005), if the score from the shape (3D) modality is particularly high, the score from the texture (2D) modality is discarded since shape analysis is more reliable than texture analysis.

2.9 Summary and Discussions

In this chapter, the different image representation and interpretation methods have been discussed. While 2D can provide a cheaper means of representing an object, it does not overcome problems caused by viewpoint and light illumination. With the advance of technology, it is now possible to obtain 3D images relatively cheaply. In particular, the use of 2.5D range images have been popular over the last two decades. This imaging modality can provide direct and explicit representation of the objects, and are invariant to lighting, pose and even viewpoint rotations.

The different methodologies in representing 3D shapes were also discussed in this chapter. These were separated into different categories, namely curvature-based methodologies, PCA based methods, point set (or point signature) based methods, feature descriptors based methods and finally multi-modal by combining 2D and 3D techniques together. While sometimes these techniques are used on their own to represent and/or match any 3D shapes, often a mixture of techniques are employed. For instance, curve curvature based methodologies are useful in segmenting the surface into different regions, thereby allowing features descriptors to be extracted from each region for matching purposes. Similarly, keypoints can be located on the 3D images followed by the extraction of features descriptors. This concept is demonstrated in Lowe's state-of-the-art 2D SIFT in which keypoints are located and features are extracted from the 2D images and similarly in Wiskott's Elastic Bunch Graph. However, these systems only

apply to 2D images and no attempts have been made to extend these systems onto the range surface.

In order to provide a full and in depth representation of 3D surfaces, it is necessary to understand and categorise the underlying surface information. The most popular 3D shape representation methodology lies within the extraction of curve curvatures based on differential geometry, which is a popular choice since the 1980s. This is because measurements can be made directly onto the surfaces, especially with the availability of range images. However, this method is prone to the sensitivity of noise that exists in the data and as a result, the range images are normally pre-processed in order to suppress this noise. Moreover, a manually-defined threshold is typically assigned to the H, K, $k1$ and $k2$ curvatures, in order to segment the images into different regions, meaning different types of images may require different manually-defined thresholds. Furthermore, from the image representation point of view, any surface (in particular of a human face), is expected to contain *mixes* of surface types, but not the typical crude methodology of attempting to segment a surface into a piecewise patchwork of single surface types, thereby preventing a true and accurate representation of a surface to be made.

Current state-of-the-art methodologies on feature extraction of 3D images are based on differential geometry and local statistics of the surface, using either the H and K or the $k1$ and $k2$ curvatures. The feature vector extracted usually captures only the surface types or in the case of Lowe's 2D SIFT, the distribution of the orientation of the 2D images at keypoint locations, but not taking both information into account simultaneously. To this end, at the beginning of this research project, the following gaps in the current literature of local feature representation and matching of 3D images were identified: firstly, the important factor of distinguishing a wide range of pattern types present at any surface. Secondly, the scale and rotation invariant properties in which any feature descriptor is expected to have. For example, Rao (1994) investigated feature vectors at multiple scales and orientations in 2D images and has shown that they have useful properties such as rotation (in-plane) and scale invariance.

In order to advance the current state-of-the-art research and address the shortfall of the current feature extraction of 3D images, it is proposed in this project to firstly investigate the representation of a surface, by encapsulating the underlying surface topology and their orientations simultaneously to form a feature descriptor which

captures both the topology and the orientation of the surface. These are discussed in Chapters 3 and 4. Secondly, since the shape of an image surface should remain constant under viewpoint changes, the local distribution of curvatures observed on its surface should therefore also remain constant. Therefore it is proposed in this project to investigate the invariance properties of the extracted feature descriptors to all three Euler's viewpoints rotational changes. Finally, it is important to be able to match the feature descriptors extracted from range images of different sizes and depicting the same target objects at different scales. As a result, scale issues are addressed in this project.

In the next chapter, a pilot investigation conducted on the extraction of feature descriptors from known keypoint locations is detailed and discussed. Here, the underlying surface information of the range images are determined and the surface types and their orientations are extracted from each manually defined keypoint locations, allowing the features descriptors to be matched and compared.

Chapter 3

Range Surface Analysis

Following the literature review, this chapter is devoted to a pilot investigation into the extraction of local feature descriptors from range images of human faces. This chapter will first provide a detail description of the data collection process used in this research project. The rationale and methodology of the feature descriptor extraction process will be presented later in this chapter, where feature descriptors were extracted from known locations of the range images based on the anatomical landmarks placed on the 3D models.

3.1 Motivation

The objective of this pilot investigation is to determine a useful way to describe a surface sampled at a landmark location on a range image, thereby allowing effective interpretations of 2.5D range images. As outlined in Chapter 2, it is possible to extract different forms of information based on differential geometry, that could be used to characterise the local range image surface.

The data used in this pilot study were computed from stereo-pair images captured using 13.5 Mpixel digital cameras and processed using a stereo-photogrammetry package, C3D2 (Ju et al., 2003). There are two main outputs from C3D2: the range image and the 3D polygon model in VRML format. The VRMLs can be loaded into the Facial Analysis Tool (Mao, 2005) where anatomical landmarks (used in this pilot investigation) were placed on the models by a professional clinician. While it would be sensible to use the 2D images, which are in line with the range maps, for the placement of these landmarks, the clinicians do require a full 3D interpretation and interaction on

the 3D models in order to place landmarks accurately. These placed landmarks can be related onto the range images using *Barycentric Coordinates* and *Texture Coordinates* (see Appendix B for details). Investigation and interpretation of the properties of the local surfaces defined by landmarks can be performed directly on the range images once the landmarks have been related onto them from the 3D polygon meshes. By fitting a *finite support region* (window) over each landmark location, predominant signature of the finite support region can be identified by means of a *feature descriptor*. The support region is expected to contain mixtures of surface types, and their statistics (surface type distributions) are predicted to capture the semantics of the underlying surfaces. The size of this finite support region depends entirely on the characteristics of the landmarks, if the support region is small, it may not be capable of capturing enough information to describe uniquely the surface associated with a particular landmark, whereas if the support region is too large, the localisation of the landmark will be compromised.

Once the range images have been pre-processed, different types of information can be obtained from the range images at set landmark locations: First of all, shape information can be obtained based on the signs of the mean (H) and Gaussian (K) curvatures. Secondly, it is possible to determine the orientation of the underlying surface (i.e. direction) and use this information to categorise the underlying surface. By extracting features based on the histograms that characterise the *local* surface curvature, it is possible to create a composite 16-element feature descriptor (eight for the surface types and eight for the orientation) that contains a histogram of the relative frequencies of the mixes of different surface types (peak, ridge, saddle ridge, flat, minimal surface, pit, valley or saddle valley) and the orientation obtained from the direction of the principal maximum curvature ($k1$), based on differential geometry. This proposed feature descriptor is potentially capable of discriminating the landmarks by characterising the local surface shape on which they have been located based on these features.

Figure 3.1 shows the steps involved to extract a feature descriptor from the regularised range image. Each step involved in the process will be described in detail in the remainder of this chapter as follows: Section 3.2 details the data collection process involved in this work. Section 3.3 describes the data regularisation process taken in this work in order to pre-process the range data. The pilot investigation conducted in

devising a suitable feature descriptor for range image representation is detailed in Section 3.4. Section 3.5 outlines the validation for this pilot study. Section 3.6 presents the initial results obtained from the matching of the feature descriptors. Finally, Section 3.7 concludes this chapter.

3.2 Data Collection

The data set used in this research project comprises 2.5D range images and their corresponding 2D stereo-pair images. The data set employed for the pilot study contains 60 range images of female faces, captured at a fixed pose similar to a standard passport photograph, except subjects have their eyes closed and facial muscles relaxed. This data set was captured at different time points at the Southern General Hospital as part of a longitudinal study (i.e. a study of changes over time) investigating the properties of collagen injections (Downie et al., 2008).

The process in deriving the range images is described in this section. Figure 3.2 shows the steps involved in the capturing and the building of 3D models.

3.2.1 Image Capturing

The configuration of the stereo-pair cameras are illustrated in Figure 3.3, showing the single pod system mounted on a camera rig, attached with two portable flash units and connected to a PC computer. The cameras are of the following specifications: DCS 14N Pro Kodak Digital Cameras and each picture is of 4500×3000 pixels resolution. Image capture is manually initiated via a user interface, that synchronises simultaneous capture of the stereo-pairs. The stereo-pair images are then transferred from the cameras to the computer where models can be built. The program, *Capture Control*, developed by Ewan Borland at the University of Glasgow, is used to capture all the images (as well as the images used for the calibration of the system).

A dental chair with adjustable height and headrest is placed in front of the digital cameras. Subjects are asked to sit on the dental chair and the height of the seat can be adjusted to ensure the subject's face is in line with the cameras. The adjustable headrest enables the distance of the subject to the cameras to be standardised. A sequence of facial exercises is then undertaken by each subject in order to help them relax their facial expression to ensure the fixed pose criteria is met. The resulting images

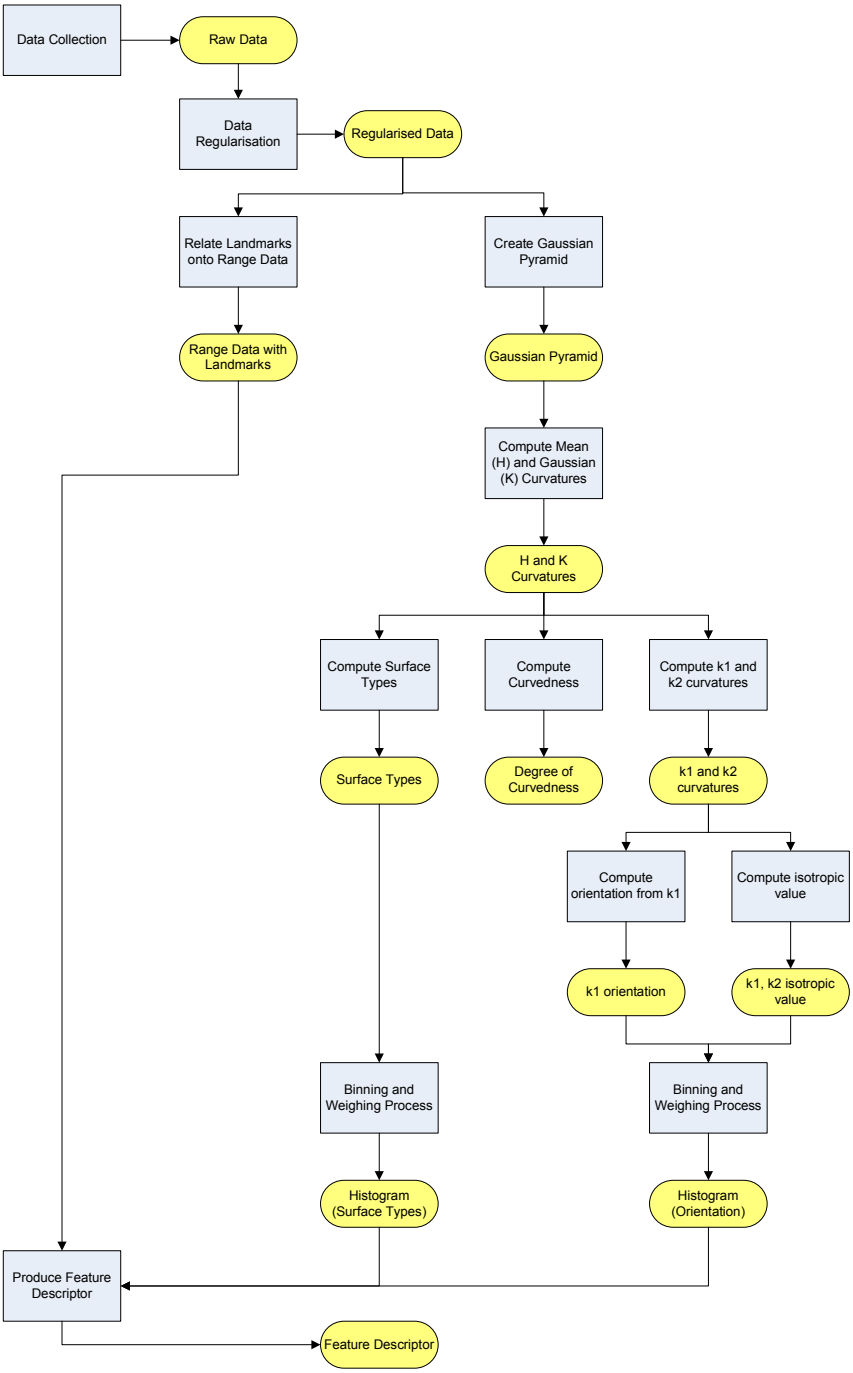


Figure 3.1: Feature descriptor extraction pipeline.

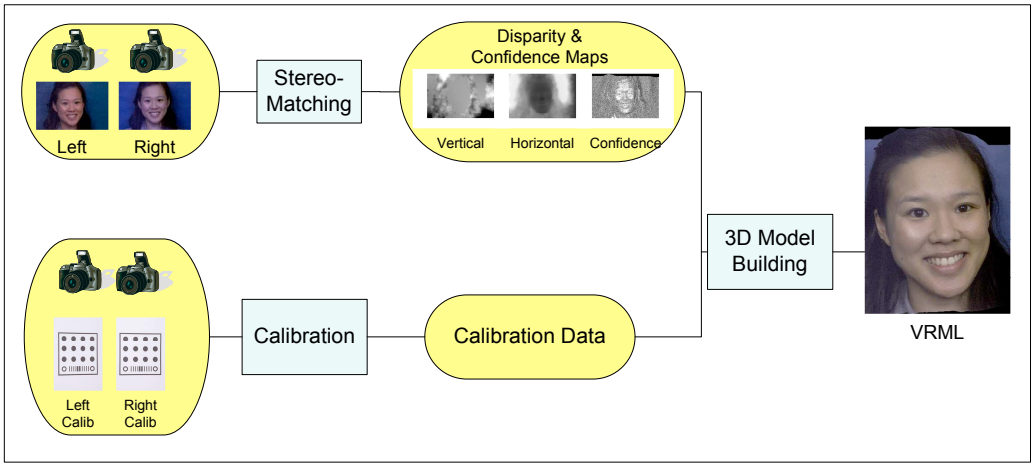


Figure 3.2: Pipeline for building a 3D model.

have the appearance of a standard passport photograph, except the subjects have their eyes closed and all facial muscles relaxed. These images are captured using Capture Control and then checked visually on the camera display before being transferred onto the computer. Multiple images are taken in order to accommodate pose and acquisition errors and the most suitable set of images with clear focus and subject's pose relaxed are selected. Once the images have been captured and checked, they are exported as HIPS format images (Landy and Sperling, 1984), ready for the corresponding 3D models to be built using C3D2.

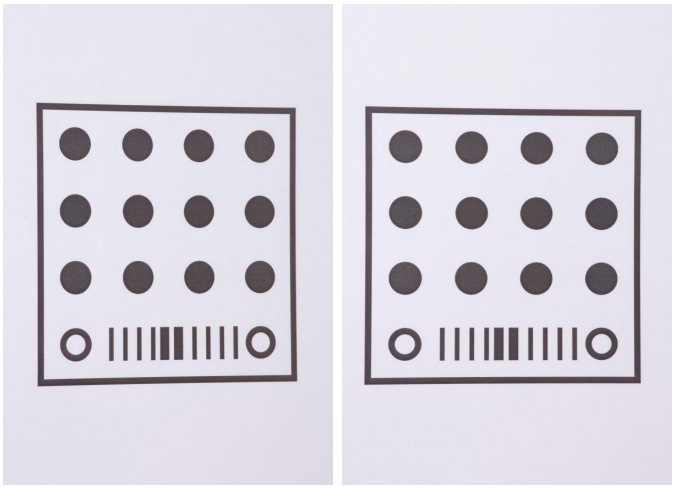
3.2.2 Models Building: C3D2

C3D2 was initially developed by Turing Institute, Glasgow (Siebert and Marshall, 2000; Siebert and Urquhart, 1994), and was later improved by Ju et al. (2003). This software is used for the calculation of the calibration error of the system as well as the construction of range images and 3D polygon models from the stereo-pair of 2D images. There are three steps in the building of 3D models: 1) stereo matching, 2) surface re-section (photogrammetry) and 3) polygonisation. The models are built using the direct range mesh method. The settings used for the data set used in this work can be found in Appendix A.

Figure 3.3: Single pod stereo-pair system with two cameras mounted on a rig, along with two portable flash units.

3.2.2.1 Calibration

In order to recover the metric range values, a calibration procedure is required. The calibration target is captured in 13 different orientations within the field of view of the cameras and these images are used to calculate the geometry of the cameras and their relative orientations by means of bundle adjustment (Urquhart, 1997). This information is used to recover the range values from the disparity maps produced from the stereo-pairs, thus enabling a 3D model to be built. The calibration target used in this work, type *Plane Domino 48* (illustrated in Figure 3.4), is placed on the dental chair in front of the camera. Figure 3.4 shows the images captured from the calibration target by the left (a) and the right (b) cameras. The captured calibration image is projected onto a virtual plane and is called the virtual back-projected image. The calibration error is computed by calculating the minimised sum of the root-mean-square (RMS) between the 12 dark circles on the virtual back-projected image and the circles on the real back-projected image. Values less than 1.5 pixels of RMS error are acceptable.



(a) Left (b) Right

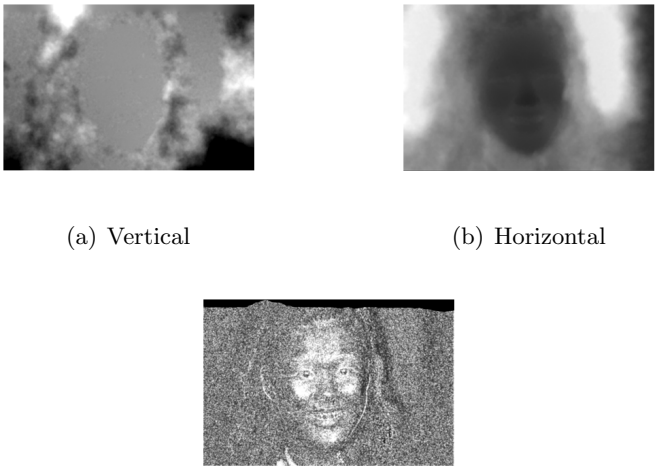
Figure 3.4: Calibration images from (a) left camera and (b) right camera

3.2.2.2 Range Surface

The starting point for building a 3D model is the stereo-pair of photographs captured from the high-resolution digital cameras. A scale-space based matching algorithm in C3D2 computes the dense disparity map from the stereo-pair. This map can be displayed as vertical (Figure 3.5(a)) and horizontal (Figure 3.5(b)) disparity maps. In addition, a confidence map (Figure 3.5(c)) is produced which indicates the reliability of each match value, the lighter the shading, the greater the confidence. The calibration data is used to transform the match data to produce depth values for each pixel. This is summarized in a range map as shown in Figure 3.6. From the range map, the 3D coordinates for the model are constructed and the model is built. The model is a triangulated mesh with 2D photographic texture superimposed onto the mesh. This can then be exported as a VRML file (as shown in Figure 3.7) which is a commonly used format for 3D files and can be viewed with 3D viewer, for example, GLView.

3.2.3 Manual Landmark Placements

Anatomical landmarks have been placed on the 3D polygon mesh models by a professional clinical expert using *Facial Analysis Tool*, developed by Zhifang Mao and Zhili Mao (Mao, 2005), to provide set locations on the face from which subsequent analysis can provide standardised measurements. A total of 28 landmarks were placed on



(a) Vertical (b) Horizontal

(c) Confidence

Figure 3.5: (a) Vertical disparity; (b) horizontal disparity; (c) confidence map.

each model, as shown below in Figure 3.8. Table 3.1 details the information for each landmark.

Once the landmarks have been placed on the 3D polygon model, they are saved to a text file which contains the x, y and z coordinates of the landmarks and their corresponding normal vectors. This data can be read into the Facial Analysis Tool at a later time point. By relating the landmarks from the 3D model onto the corresponding 2.5D range image, subsequent analysis on the discrimination of the landmarks can be calculated. This process is described in Section 3.3.3 below.

3.3 Data Regularisation and Representation

By manipulating on the range image directly, it potentially allows interpretation to be performed at full resolution, as opposed to approximately 5% of the available information as is the case for the (highly decimated) polygon meshes models used by the clinicians. Noise removal on the range images is vital due to the inherent range noise present of the order of $\pm$0.2-0.3mm RMS, determined by C3D2. This has been achieved using one of the available built-in functions in the *Heritable Image Processing System* (HIPS) (Landy and Sperling, 1984). HIPS, first developed in the 1980s, is a image

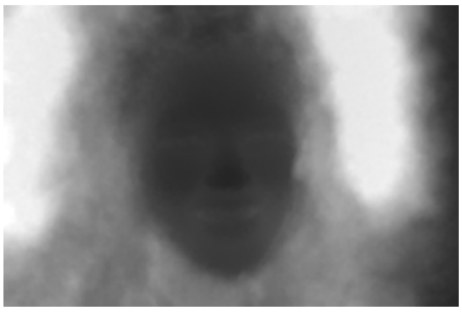


Figure 3.6: Range image.

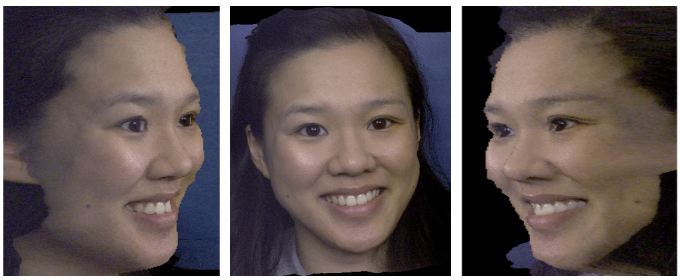


Figure 3.7: Examples of VRMLs.

processing package written in C that provides several hundred useful image processing filter processes.

3.3.1 Noise Removal

Each range image contains a header which provides information regarding the data, for example, the number of pixels contained in the range image and the pixel format. However, these must be eliminated prior to reading into MATLAB for analysis since MATLAB can only deal with matrices but not matrices with headers. This can be achieved by using two of the built-in HIPS functions, `ptoa` and `stripheader`, which would convert the pixel values to `ASCII` and remove the header from the range image respectively.

Moreover, HIPS also provides an effective noise removal function, using the Adaptive-Surface-Labelling (ASL) technique (Mowforth and Jin, 1986) in which a smoothing filter is applied to each image of the input sequence to suppress random noise.

No.	Label	Anatomical Name
1	exR	right exocanthion
2	enR	right endocanthion
3	enL	left endocanthion
4	exL	left exocanthion
5	sbalR	right subalare
6	sbalL	left subalare
7	chR	right cheilion
8	chL	left cheilion
9	n	nasion
10	prn	pronasale
11	acR	right alar
12	acL	left alar
13	cphR	right crista philtri
14	cph	crista philtri
15	cphL	left crista philtri
16	li	labiale inferius
17	sto	stomion
18	sto	stomion (open mouth)
19	liUR	mid-point right upper lip
20	liUL	mid-point left upper lip
21	liLLR	lateral one third right lower lip
22	liMLR	medial one third right lower lip
23	liMLL	medial one third left lower lip
24	liLLL	lateral one third left lower lip
25	exR-chL	mid-point on the surface curve joining landmarks 1 and 7
26	exR-chR	mid-point on the surface curve joining landmarks 4 and 8
27	men	menton
28	colm	subnasal point

Table 3.1: Anatomical landmarks placed by a professional clinician.

Further noise removal can be achieved using a Gaussian Pyramid (Burt and Adelson, 1983). The Gaussian Pyramid also offers a multi-scale representation of the image. The original size of the range image is 1498 pixels by 2249 pixels and at this scale, representative features are extremely difficult to detect as the support region required to sample local surface structure will vary significantly for different types of anatomic structure. Using the Gaussian Pyramid, analysis can be performed on different scales simultaneously by applying the *same* extraction window to each level of the pyramid.

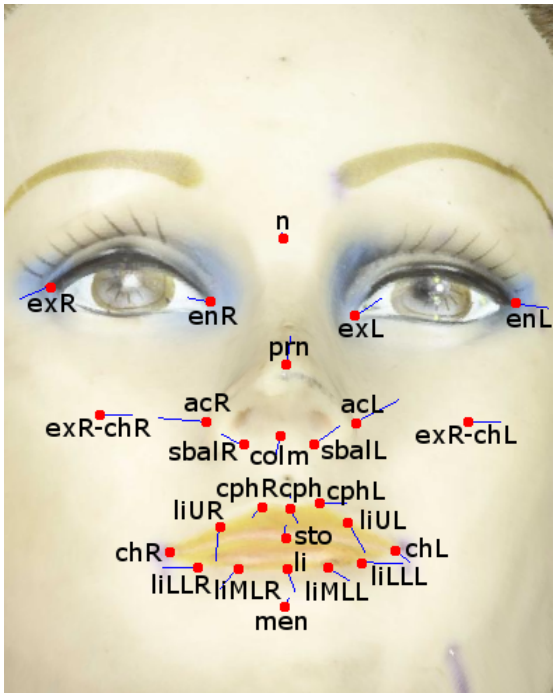


Figure 3.8: Anatomical landmarks (red dots) placed on a 3D model.

3.3.2 Gaussian Pyramid

The Gaussian pyramid contains a hierarchy of low-pass filtered versions of the original image, such that successive levels comprise smaller images containing correspondingly lower frequency information. Each layer is smoothed by a symmetric Gaussian kernel and re-sampled to compute the next layer. As a result, noise can be further suppressed using a Gaussian pyramid as well generating a multi-scale representation. Figure 3.9 illustrates an image pyramid, where Level 1 corresponds to the original image and each subsequent level contains a low-pass filtered versions of the original image.

For the purpose of this work, a half-octave Gaussian pyramid is used to reduce the size of the range images. Table 3.2 shows the image size at each of the half-octave Gaussian pyramid, where Level 1 represents the original range image and Levels 2 to 9 represent the down-scaled images and their respective sizes.

3.3.3 Relating Landmarks on Range Images

As mentioned in Section 3.2.3, anatomical landmarks have been assigned by a clinical professional on the 3D polygon meshes using the Facial Analysis Tool. In order to

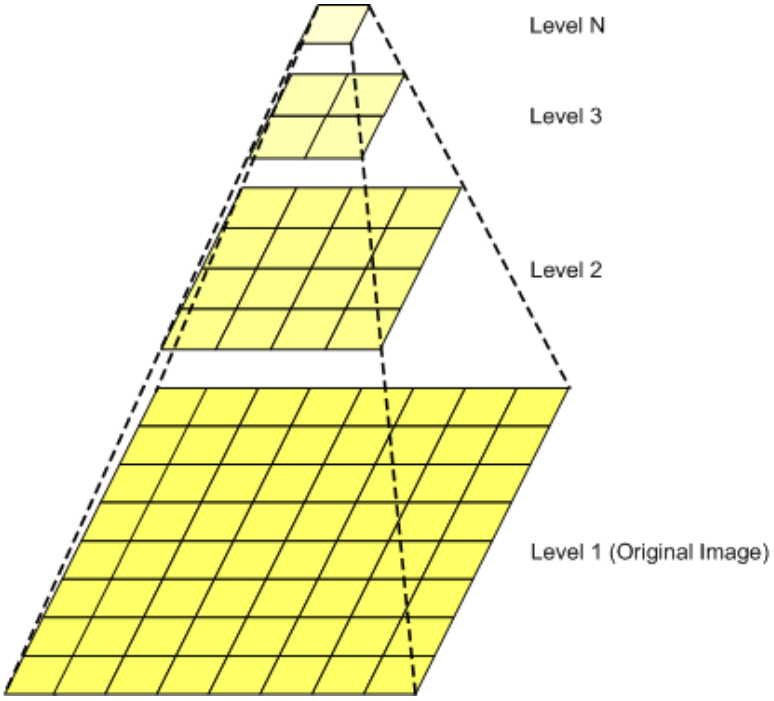


Figure 3.9: Example illustrating a Gaussian pyramid.

perform investigation and surface analysis of the characteristics of the landmarks and their surrounding pixels, the location of the landmarks must be identified on the range image. This can be achieved using *Barycentric coordinates* which allows the *Texture Coordinates* of the landmarks to be located. Texture coordinates are the means by which texture image positions are assigned to vertices. The details for the methodology in relating landmarks on the range images can be found in Appendix B.

The feature extraction process around each landmark can begin once the data has been pre-regularised and the Gaussian pyramid for the range image has been built. The methodology taken to extract the feature descriptor is described in Section 3.4 below.

3.4 Feature Extraction

This section describes the methodology and the steps taken in building a feature descriptor for each of the landmarks and their surrounding pixels, over a finite support region.

Level	size (pixels)
1	1498×2249
2	995×1495
3	746×1121
4	494×744
5	370×557
6	244×369
7	182×275
8	119×181
9	88×134

Table 3.2: Size of the image at each pyramid level.

3.4.1 Methodology

In order to compute a feature descriptor which contains a predominant signature of each landmark, it is important to identify the underlying surface types and orientation of the landmarks and their surrounding pixels. Thereby the underlying surface can be expressed in terms of statistical descriptions of the surface shape. These can be calculated based on differential geometry. The steps that lead to the formation of the histograms, which contains the relative frequencies of the underlying surface types and the principal directions of the range image, are as follows:

1. Calculate the first partial derivatives with respect to x and y axis of the range image (Jain et al., 1995):

$$f_x = \frac{\partial F}{\partial x} = F(i+1, j) - F(i, j) \tag{3.4.1a}$$

$$f_y = \frac{\partial F}{\partial y} = F(i, j+1) - F(i, j) \tag{3.4.1b}$$

   where (i, j) is the i^{th} and j^{th} pixel value for the range data F.

2. Calculate the second partial derivatives with respect to x and y axis of the range image (Jain et al., 1995), based on the first derivatives:

$$f_{xx} = \frac{\partial^2 F}{\partial x^2} = F(i+1, j) + F(i-1, j) - 2F(i, j) \tag{3.4.2a}$$

$$f_{yy} = \frac{\partial^2 F}{\partial y^2} = F(i, j+1) + F(i, j-1) - 2F(i, j) \tag{3.4.2b}$$

$$f_{xy} = \frac{\partial^2 F}{\partial x \partial y} = F(i+1, j+1) + F(i-1, j-1) - F(i+1, j-1) - F(i-1, j+1) \quad (3.4.2c)$$

where (i, j) is the i^{th} and j^{th} pixel value for the range data F.

3. The mean curvature (H) can be computed based on the first and second derivatives:

$$H(i,j) = \frac{(1 + f_y^2(i,j))f_{xx}(i,j) + (1 + f_x^2(i,j))f_{yy}(i,j) - 2f_x(i,j)f_y(i,j)f_{xy}(i,j)}{2(\sqrt{1 + f_x^2(i,j) + f_y^2(i,j)})^3} \quad (3.4.3)$$

4. Similarly, the Gaussian curvature (K) can be computed based on the first and second derivatives:

$$K(i,j) = \frac{f_{xx}(i,j)f_{yy}(i,j) - f_{xy}^2(i,j)}{(1 + f_x^2(i,j) + f_y^2(i,j))^2} \quad (3.4.4)$$

5. Generate the surface types for the 3D surface, based on the signs of the H and K curvatures, using Table 2.1 as shown in Section 2.4. There are eight types available from this representation, namely one of peak, ridge, saddle ridge, flat (plane), minimal surface, pit, valley and saddle valley. Eight different colours can be assigned to each of the surface types, as shown in Figure 3.10 below. Figure 3.11 illustrates a face that has been categorised into the eight surface types.

6. Compute the degree of curvedness of the surface, which is used as a weight to indicate how strong the curvature is at that particular point:

$$\text{curvedness} = \sqrt{2H^2 - K} \quad (3.4.5)$$

7. Calculate the principal curvatures (maximum and minimum curvatures $k1$ and $k2$ respectively), using Equation 2.3.3 as shown in Section 2.3.1.

8. Calculate the maximum curvature orientation ($k1$ direction), as shown in Equation 3.4.6. This can then be used to categorise the surface into eight orientations, covering the full 360° of orientation, as shown in Figure 3.12.

$$\theta_{k1} = tan^{-1}\frac{\partial y}{\partial x} \quad (3.4.6)$$

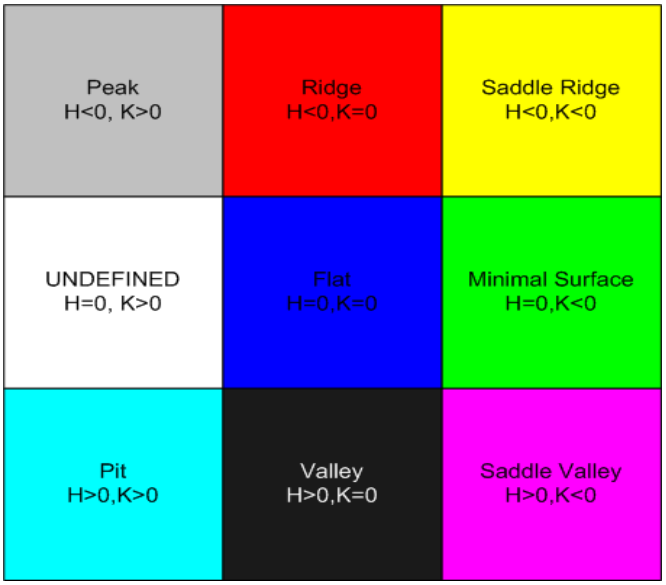


Figure 3.10: Colour assigned to the eight different surface types derived from the signs of H and K curvatures.

Figure 3.11: Example of a face that has been categorised into the eight surface types. The colouring of this figure corresponds to the colour scheme showed in Figure 3.10.

where ∂y and ∂x are the first derivatives of the $k1$ curvatures.

9. Compute a normalised isotropic value, which can be used to favour the surface contributions where the rate of orientation change (of the curvature) is changing dynamically. This value can then be employed as a weight in the histogram and

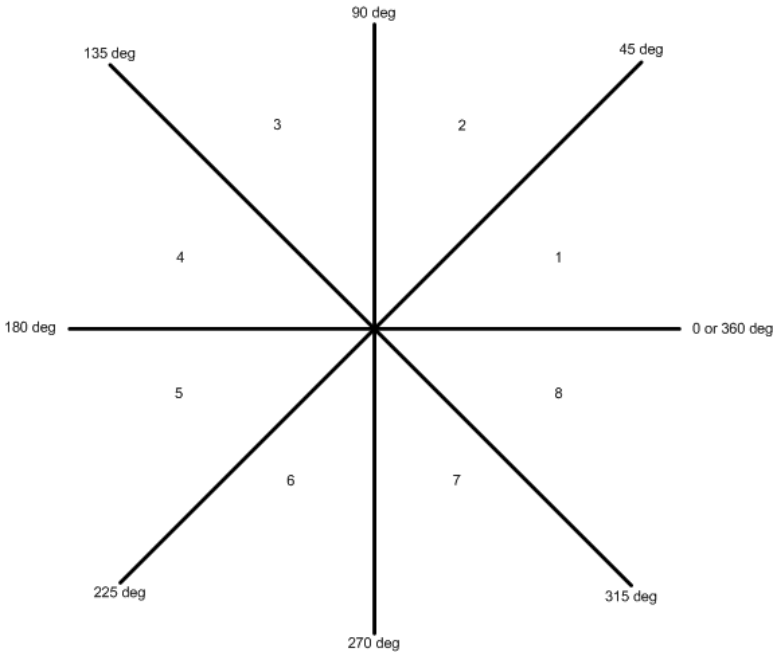


Figure 3.12: Orientations can be separated into eight sections, covering the full 360° of orientation.

is potentially more diagnostic/descriptive of the sample patch:

$$\frac{\left|\frac{\partial^2\theta}{\partial x^2}+\frac{\partial^2\theta}{\partial y^2}\right|}{\sqrt{\left(\frac{\partial^2\theta}{\partial x^2}\right)^2+\left(\frac{\partial^2\theta}{\partial y^2}\right)^2}} \tag{3.4.7}$$

where θ is the $k1$ orientation.

10. Fit a finite support region (window) over the landmark and compute the histograms of the surface types and orientations within that window respectively, using the curvedness and the isotropic values for the $k1$ direction as the weights of the histograms respectively. The histogram for the surface types over the support region is computed by accumulating the frequencies of each of the surface types based on the signs of H and K. There are a total of eight bins in the histogram, one for each surface type. Each sample added to the histogram is weighted by the curvedness which provides an indicator of how strongly a particular surface type is occurring at that point. Similarly, compute the histogram for the orientation using eight bins, each sample added to the histogram weighted by the

isotropic value which indicates the orientation strength. The resulting histograms are presented as follows, for surface type and orientation, respectively:

$$S = (s_1, s_2, s_3, s_4, s_5, s_6, s_7, s_8)$$

$$D = (d_1, d_2, d_3, d_4, d_5, d_7, d_7, d_8)$$

11. Normalise the histograms locally by the sum of counts in its corresponding histogram.

12. A feature descriptor can be computed by combining the histograms of the surface types and the orientations together to form a single vector of length 16, the first eight elements represents the histogram from the surface types and the latter eight represents the elements from the histogram of the orientation. The structure of the feature descriptor can be found in the next section.

3.4.2 Structure of the Feature Descriptor

By combining the histogram for the surface type and the histogram for the orientation over a landmark together, a single 16-element feature descriptor which provides information about a landmark and its surrounding pixels over a finite support region can be computed and represented as follows:

$$\text{Feature Descriptor} = (s_1, s_2, s_3, s_4, s_5, s_6, s_7, s_8, d_1, d_2, d_3, d_4, d_5, d_7, d_7, d_8)$$

where $s_1, ..., s_8$ is the histogram of the surface types and $d_1, ..., d_8$ is the histogram of the orientation. This feature descriptor has been locally normalised individually.

Such a feature descriptor is computed for each of the landmarks and their surround pixels over a finite support region of 5×5, 7×7, 11×11, 13×13 and 17×17 for each of the nine levels in the Gaussian pyramid. This offers a multi-scale representation and the aim is to locate the level and the size of the finite support region which exhibits the best discrimination value between the 28 landmarks.

Definition 3.4.1 Z-normalisation:

$$sdata = \frac{cdata - \text{mean}(cdata)}{\text{std}(cdata)} \tag{3.4.8}$$

where mean(*cdata)* and std(*cdata)* is the mean and the standard deviation across the entire population. This offers a global normalisation of the data set.

A pilot investigation was conducted on the feature descriptors using two sets of data, firstly a synthetic set of range data of an orthogonal spatial sine waves, generated using MATLAB and secondly 60 test data comprising 60 range images of human faces. The 60 test data are divided into two groups: the training (known) data and the testing (unknown) data. Each group comprises 30 faces, each of which has 28 landmark classes, representing the 28 different landmarks. By knowing in advance which feature descriptor is used to represent a particular landmark, it becomes possible to compute the matching properties of the formulated descriptor, using the K-nearest-neighbour (KNN) algorithm.

3.5 Validation of the Feature Descriptor on Synthetic Data

A 2.5D synthetic range image, created from orthogonal spatial sine waves, has been designed and generated in order to validate the feature extraction process (See Figure 3.13). By knowing in advance which surface types are present in the synthetic test data, it becomes possible to validate the surface curvature and surface type labelling process and also the construction and utility of the feature descriptor. The surface types present in the synthetic test data would be 'peak' where the sine waves are at their peaks, 'pit' where the trough of the waves appear and saddle ridge and saddle valley in between, as shown in Figure 3.14.

Moreover, the stability of the calculations of the curvatures, including their ability to cope with variation in surface pose with respect to the camera, can be determined by rotating the 2.5D synthetic range data (Figure 3.15). Experiments show that the calculation of the surface types can manage with rotation up to $\frac{\pi}{4}$ degrees.

3.6 Validation of the Feature Descriptor

K-nearest-neighbour method can be used to estimate the matching properties of the feature descriptors. This can be achieved by measuring the Euclidean distance between the test data and the training data. The test data consists of feature descriptors of unknown landmarks whereas the training data consists of the feature descriptors of known landmarks. The smaller the distance between the test data and the training

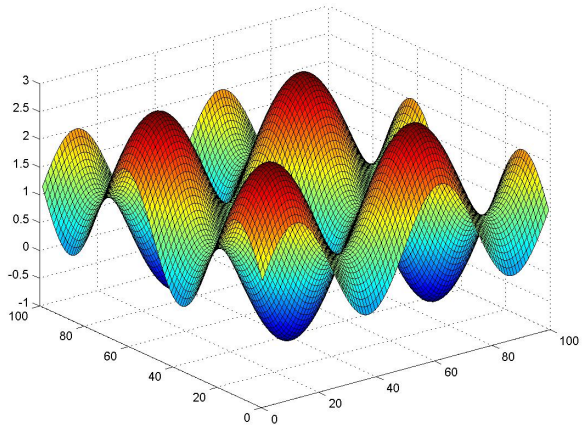


Figure 3.13: Orthogonal spatial sine waves.

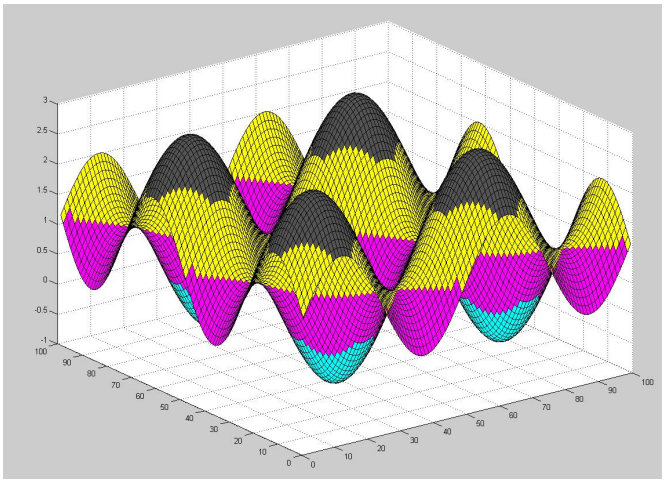


Figure 3.14: Surface types of synthetic waves. The colouring of this figure corresponds to the colour scheme showed in Figure 3.10.

data, the more likely it is that the test data is correctly labelled as the landmark represented by the feature descriptor in the training data. By knowing in advance the correct landmark associated with a feature descriptor in the test data, it is possible to identify the number of correctly labelled landmarks in the test data, based on the training data. Each of the testing data and training data contains 30 faces each with 28 landmarks. The methodology is:

1. Read the training data into MATLAB. The training data consists of the locally

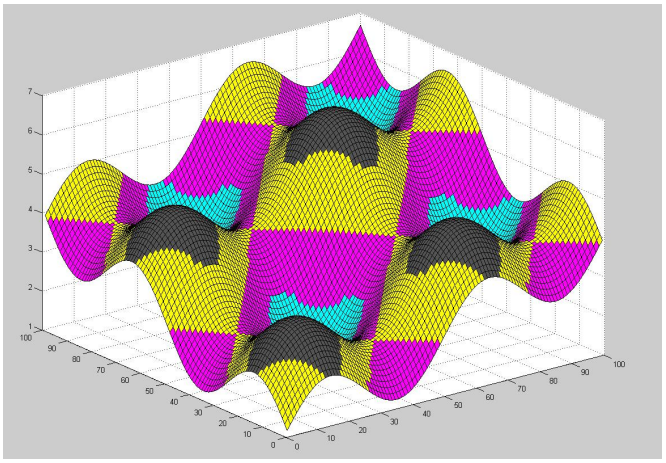


Figure 3.15: Surface types of synthetic waves with out-of-plane orientation of $\frac{\pi}{8}$. The colouring of this figure corresponds to the colour scheme showed in Figure 3.10.

normalised feature descriptors of 30 faces, each with 28 landmarks (i.e. a total of 840 feature descriptors).

2. Find the mean and standard deviation for each landmark across the population.

3. Normalise the training data globally using z-normalisation (as described in Equation 3.4.8).

4. Read the testing data into MATLAB. The testing data comprises of feature descriptors of 30 different faces to the training data, each with 28 landmarks in which it is not known a-priori the association between the landmark class and the feature descriptors.

5. Normalise the testing data individually using the mean and standard deviation of the training data. This ensures both set of data are projected onto the same space.

6. Compute the Euclidean distance between the testing data and the training data, using the following equation:

$$\text{Distance} = \sqrt{(Data_{test} - Data_{training})^2} \tag{3.6.1}$$

7. Rank the results in ascending order (i.e. the smallest distance ranks first).

8. Extract the top k values, providing a label to the unknown landmark in the testing data feature descriptor. k is selected to be 1, 3, 5, 7 and 9 for this experiments.

9. Label the unknown feature descriptor with a landmark class, using the majority voting scheme. If the vote is tied, then repeat using $k = k - 1$, until one single label has been identified.

10. Compute a *confusion matrix* (of size 28×28) based on the assigned labels. Each entry in the matrix represents the number of feature descriptors identified as that particular landmark for each of the testing class against the training class. For instance, entry (23,23) will represent the number of feature descriptors identified as landmark 23 from the population.

11. Compute the percentage of correctly labelled landmarks for each k using the following:

$$\% \text{ of correctly labelled landmarks} = \frac{\#\text{correctly labelled}}{\#\text{sample}} \times 100 \tag{3.6.2}$$

The above algorithm is performed on all nine levels and on different finite support regions (5×5, 7×7, 11×11, 13×13 and 17×17) over the landmarks. The results are then analysed by finding the percentage of correctly labelled landmarks for the test data, generated using the above algorithm which can be viewed visually using the built-in MATLAB function `bar(KNN)` as shown in Figure 3.16 below. Ideally, all k labels should be correctly labelled. However, this is extremely difficult to achieve because of the nature of the surface types and orientations could be quite similar between two landmarks, for instance the corner of the eyes. Each landmark result can be examined individually visually using the built-in MATLAB function `bar(KNN(i,:))` as shown in Figure 3.17 below. The peak should ideally represent the correctly labelled landmarks. In this particular example, ($k = 7$), which was extracted from landmark number nine from level nine with a finite support region of 5×5, 15 out of 30 feature descriptors were labelled correctly. The results are presented in the next section.

3.6.1 Results

Following from the algorithm described in the previous section, the matching ability of the feature descriptors can be investigated using different values of k, where $k =$

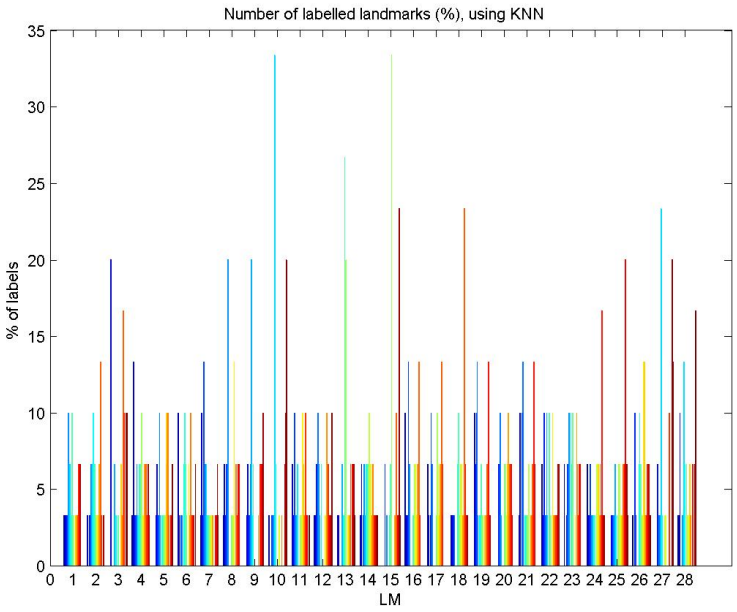


Figure 3.16: Percentage of correctly labelled landmarks using the k-nearest-neighbour algorithm.

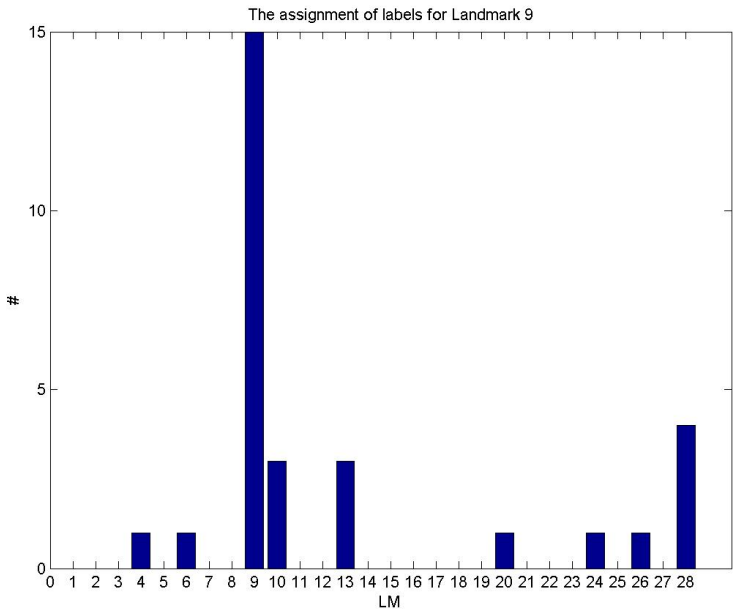


Figure 3.17: Labels assigned to feature descriptor whose true class is landmark 9.

$1, 3, 5, 7$ or 9. The data set was separated into two subsets: the training data, containing 840 feature descriptors of the landmarks in which the position of the landmarks are known to the system and 840 feature descriptors of unknown location of landmarks as the testing data. This experiment was carried out on all nine levels of the half-octave Gaussian pyramid and over different sizes of finite support region. The aim of this investigation is to determine the level and the size of the finite support region which provides the most correctly labelled landmarks.

The average percentage of the correctly labelled landmark for each of the k value can be computed and ranked in descending over all nine levels and different sizes of the finite support region, for each k.

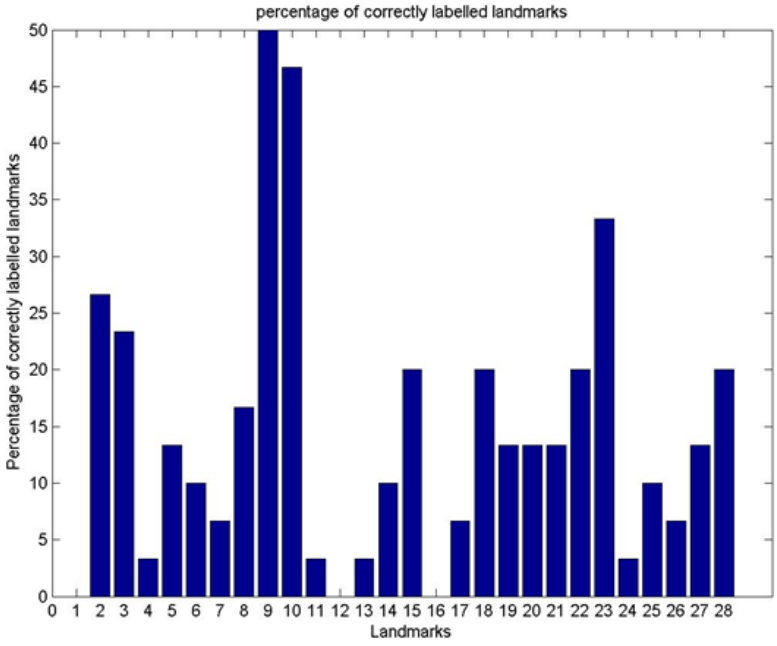


Figure 3.18: Percentage of the correctly labelled keypoints in the data set: Level 9 with window 5×5, $k = 1$

By ranking the percentages in descending order over all nine levels and different finite support region sizes, at level 9 with a finite support of 5×5 region with k equals 1 yields the best result, with an average of 15% of the testing data being correctly labelled, as shown in Figure 3.18 above. Level 9 with a finite support of 5×5 region yield the best average results, ranging from 14.04% to 15%. Overall, more labels

are correctly identified when the image size is smaller. This is because features are extremely difficult to detect when the image is at its full size.

This experiment provides a preliminary evaluation to the classification of landmarks only and the resulting classifier is regarded as low to weak in terms of recognition power.

In addition, different landmarks may require to be detected at different levels, depending on the nature of the landmark. For example, the number of correctly labelled landmarks for landmark 9, representing the tip of the nose, was achieved optimally at level 9 of the Gaussian pyramid, with a finite support of 5×5 region and $k = 7$, with 16 being correctly labelled out of 30 (over 53% of accuracy) as shown in Figure 3.17.

Furthermore, the properties of mirroring landmarks could have provided the ambiguities in locating the landmarks successfully. For instance, the surface types surrounding the landmark which represents the left outer corner of the eye may be similar to the surface types surround the right outer corner of the eye.

3.7 Summary and Discussion

This chapter explores the pilot investigation conducted in this research project on the derivation of the feature descriptor in which a basic structure was formed for fixed landmarks. The statistics of the local topology was considered, a methodology that is away from the classical single surface type classification present in the typical literature. Here, the underlying surface types of the range images are categorised and histogrammed into eight types, allowing the mixtures of surface types to be presented for each landmark location. Furthermore, the orientation of the sampling window was histogrammed into eight bins, covering the full 360° of rotations. As a result, by taking the relative frequencies of the surface types and their orientations into account simultaneously, it is possible to encapsulate the underlying signature of a local surface within a measurement aperture that samples an anatomic landmark location.

Whilst it is potentially useful to take the surface types and their orientations into account simultaneously, capturing both the quantisation and the nature of the underlying surface, there are several deficiencies in this derived feature descriptor: firstly, it does not differentiate well between mirrored landmarks. For instance, the classifier potentially gets confused between the left exocanthion (outer corner-of-the-eye) landmark and the right exocanthion landmark. Furthermore, using the eight surface types

derived by the signs of H and K curvatures are not adequate as a representation of the range images of human faces. This is because no surface on the face is flat unless a user-defined threshold is introduced and therefore not all the surface types represented by the signs of H and K curvatures are explored fully. Ideally, in order to categorise a range image of a human face to full effect, a classification that is more continuous is required and as a result, the shape index should be considered instead.

In order to address these deficiencies, the next chapter examines alternative means of deriving a more suitable feature descriptor for range image representation of a human face where the signs of H and K curvatures will be replaced by the shape index, a continuous representation of a surface along with other means of representing the orientations. Different formulations and combination of the feature descriptor will be investigated. Moreover, in order to couple the information obtained from the surface types and their orientations together, a 2D feature descriptor is investigated and the results are addressed in the next chapter.

Chapter 4

Formulation of the Feature Descriptors

The concept of building a feature descriptor for range image analysis based on its local surface information was discussed in the previous chapter, where the frequencies of the mixes of surface types and their orientations were taken into account simultaneously. The surface types was presented by means of the signs of the mean (H) and Gaussian (K) curvatures while the orientations were represented by the gradients of the maximum principal ($k1$) curvatures. Despite the concept of combining the surface types and their orientations having the potential to express the richness of the representation, the obtained results were poor. One major deficiency is the lack of continuity in the quantisation stage of the surface types, since no surface is flat in a range image of a human face, unless an user-defined threshold is introduced. Therefore, it is necessary to address this issue by exploring alternative methodologies in extracting the surface information. This chapter will explore the different possible structures for a stable feature descriptor for range image analysis, demonstrating the invariance and discriminability of the basic feature descriptor to viewpoint rotational changes.

4.1 Objectives

Following the previous chapter where a feature descriptor was formulated in order to capture the underlying surface, the use of histograms appears to be able to capture the surface information. However, there was a flaw in the quantisation part of the feature descriptors, resulting in the matching ability of the feature descriptors being comparatively weak. As a result, an alternative representation of the surface types that

avoids explicit quantisation was formed by using the shape index as detailed in this chapter.

The two core elements in the feature descriptors are concatenated together and therefore, the resulting feature descriptors are decoupled (i.e. not correlated). As a result, a 2D feature descriptor has been proposed in order to improve and capture the association between the components and is detailed in Section 4.3.1.

In order to study the invariance and discriminability properties of the feature descriptors, rotational changes have been applied to the range images and feature descriptors have been extracted from the rotated images at standard locations. The vector dot product between the rotated feature descriptors and the original (un-rotated) feature descriptors can be measured and hence the invariance measure can be determined. The discriminability of the feature descriptors extracted from the rotated range data can be measured using the KNN algorithm, both outlined in Sections 4.4.2 and 4.4.3.

4.2 Surface Features Representations

The characteristic of the feature descriptor is its ability of capturing not only the mixes of surface types but also the orientation extracted from the landmarks and their surrounding pixels. This section details an investigation conducted on the surface types and orientation components of the feature descriptor in order to observe their contribution towards the feature descriptor.

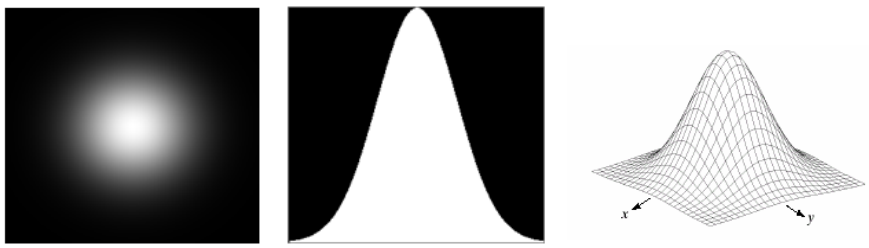
4.2.1 Gaussian Derivatives

In order to suppress further the random noise produced in the calculations of the H, K, $k1$ and $k2$ curvatures, *Gaussian* derivatives can be used, presenting with more robust and less noisy results (Marr, 1982). A 2D Gaussian kernel (Equation 4.2.1) with the appropriate sigma (σ) in both directions is applied to the range image in order to smooth the first (Equation 4.2.2) and second (Equation 4.2.3) derivatives.

$$G(x,y) = \frac{1}{2\pi\sigma^2} e^{-\left(\frac{x^2+y^2}{2\sigma^2}\right)} = \left[\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\left(\frac{x^2}{2\sigma^2}\right)}\right] \times \left[\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\left(\frac{y^2}{2\sigma^2}\right)}\right] \quad (4.2.1)$$

$$G'(x,y) = \frac{-\sqrt{x^2+y^2}}{2\pi\sigma^4} e^{\frac{-(x^2+y^2)}{2\sigma^2}} \tag{4.2.2}$$

$$G''(x,y) = -\frac{1}{2\pi\sigma^4} e^{\frac{-(x^2+y^2)}{2\sigma^2}} \left[1 - \frac{x^2+y^2}{\sigma^2}\right] \tag{4.2.3}$$



(a) 2D image (b) 1D slice through (c) 3D view

Figure 4.1: A Gaussian kernel displayed as (a) an image, (b) a 1D slice through and (c) a 3D view.

4.2.2 Quantisation of the Shape Analysis

In Chapter 3, the relative frequencies of the surface types and their orientations were taken into account simultaneous in order to form a feature descriptor. The surface types were defined by the signs of H and K curvatures, where eight surface types were represented as peak, ridge, saddle ridge, flat, minimal surface, pit, valley or saddle valley. The orientation component was obtained from the direction of the principal maximum curvature ($k1$), based on differential geometry. Whilst the idea of combining the relative frequencies of the surface types and their orientations had the potential to improve the richness of the representation, the lack of continuity in the histogram of the surface types was noticed. Therefore, in the following sessions, alternative methodologies to extracting the surface types and orientations are presented.

4.2.2.1 Surface Types

It is observed that a number of the defined surface types do not contribute towards the feature descriptor after close examination. For instance, the definition of "flat" is when both H and K curvatures have zero values which is impossible to meet under the

defined modality here. H and K will only have zero values if and only if a user-defined threshold is introduced to the modality, taken into account the biological form (human face) and the constraints of the subjects under this investigation, i.e. "flat" is not a suitable surface type to use with this type of biological form. In other words, the user will have to define a particular threshold in which all values below it will, by default, have zero value. However, a user-defined threshold is impractical as it varies according to the subject matter. As a result, alternative methods in deriving the different surface types have been investigated, in particular the [-1,1] bounded shape index introduced by Koenderink and van Doorn (1992), in the following form as shown in Equation 4.2.4. This can be used to classify into nine surface types, namely cup, through, rut, saddle rut, saddle, saddle ridge, ridge, dome and cap, as shown in Figure 4.2. The surface type "flat" is considered undefined in this representation. The nine surface types defined by shape index are of similar nature to the eight surface types described by the surface types using signs of H and K curvatures with the exception of the "flat" type and with a more refined ranges of saddles, which is highly suitable under the defined constraints and purposes of this investigation. Hence, shape index is a more suitable and a better candidate for consideration in terms of the construction of the feature descriptor under the current circumstances, constraints and purposes of the investigation.

$$s = \frac{2}{\pi} tan^{-1} \left[\frac{(k2 + k1)}{(k2 - k1)} \right] \tag{4.2.4}$$

where $k1 > k2$.

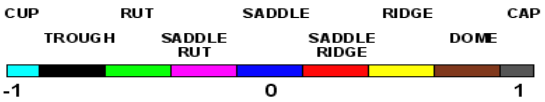


Figure 4.2: Surface types defined using shape index.

4.2.2.2 Orientation

The orientation component presented in the feature descriptor, described in Chapter 3, consists of the maximum $k1$ curvature orientation, covering the 360° of orientation, along with the isotropic value which indicates the relative signal strength of the orientation as the weight. The principal curvatures are calculated using the H and K curvatures as shown in Equation 2.3.6 in Section 2.3.1, which consists of the first and

Figure 4.3: Example of a face that has been categorised into the nine surface types derived using the shape index. The colouring of this figure corresponds to the colour scheme showed in Figure 4.2.

second order derivatives of the range image itself. The isotropic value used as the weight is in form of the second derivatives of the maximum curvature, as shown in Equation 3.4.7 in Section 3.4.1 and is therefore a fourth order derivatives of the range image. Concerns have been raised at the high degree of derivatives taken into account. The aim of the feature descriptor is to be able to capture the orientation information of the range image and therefore, it is possible to compute the orientation based on the range image using the first derivatives of the range images as an alternate methodology in extracting the orientation components instead. Eight bins can be used, covering the 360° or 2π orientation.

Furthermore, in order to confer viewpoint rotation invariance, the highest peak of the orientation histogram is detected and the histogram is then normalised to its canonical (orientation) form by rotating the orientation histogram until the peak is situated at the first histogram position.

4.3 Representation Properties

The feature descriptor derived for this work comprises two core elements: the distribution of the relative frequencies of the surface types and the distributions of the frequencies of the orientation. Based on these two core elements, it is possible to exploit different combinations for the feature descriptor, containing surface types and orientation. For instance, surface types can be derived using signs of H and K curvatures (into eight types) or using the shape index (into nine types). Orientation can be de-

rived using the high order $k1$ curvature or based on the first Gaussian derivatives of the image itself. Furthermore, the weighting system applied to the feature descriptor can be separated into different categories: no weighting (i.e. add one to each appropriate bin), the curvedness as a weight for the surface type histogram or the isotropic value derived in Section 3.4.1 for the orientation histogram.

4.3.1 Coupled Histogram

The feature descriptor described hitherto has been a single one-dimensional 16-element feature descriptors, constructed by concatenating the histogram of the mixes of surface types and the histogram of the distribution of the orientation components. However, the combination of these two local histograms results in a decoupled feature descriptor. As a result, a two-dimensional feature descriptor, using mixes of surface types and orientation as bins, has been formulated.

4.3.1.1 Methodology of the Construction of the 2D Feature Descriptors

The calculations for the derivation of the surface types and orientation are identical to the process described in Section 3.4.1. When constructing the histograms for the 2D feature descriptors, an eight by eight matrix has been formed (or in the case of using the surface types derived from the shape index, a nine by eight matrix is used), with the individual rows representing the eight (or nine) surface types and the columns representing the eight-binned orientation histogram. Each pixel within the sampling aperture is examined and its representing surface types bin (x) and orientation bin (y) have been identified and the product of the weights of these two components is added to the (x, y) position of the matrix. The orientation component is rotated, as before, by locating the highest peak of the orientation (for each column) and is normalised to canonical (orientation) form by rotating the entire column until the peak is located at the first position of the column. Finally, the feature descriptor is normalised locally by the sum of all its bins in order to produce a standard unit vector.

Shown in Figure 4.4 is the 2D feature descriptor extracted from the face range image sample 1, with sample aperture size 17×17 pixels, on landmark 10 (tip of the nose).

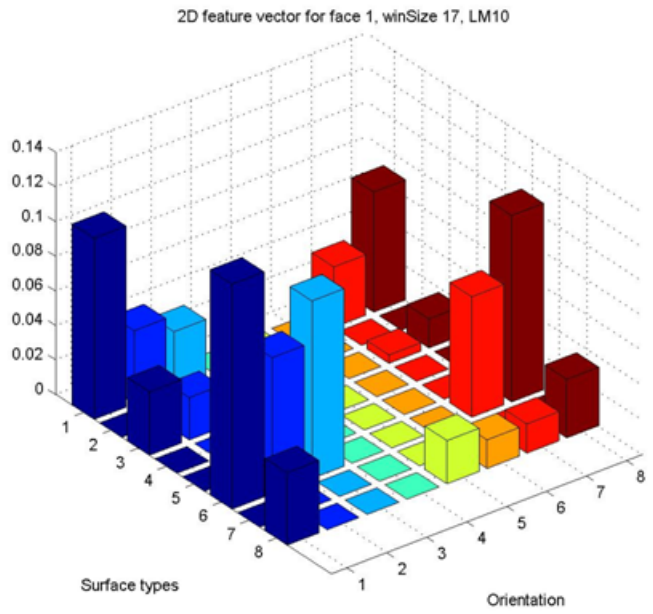


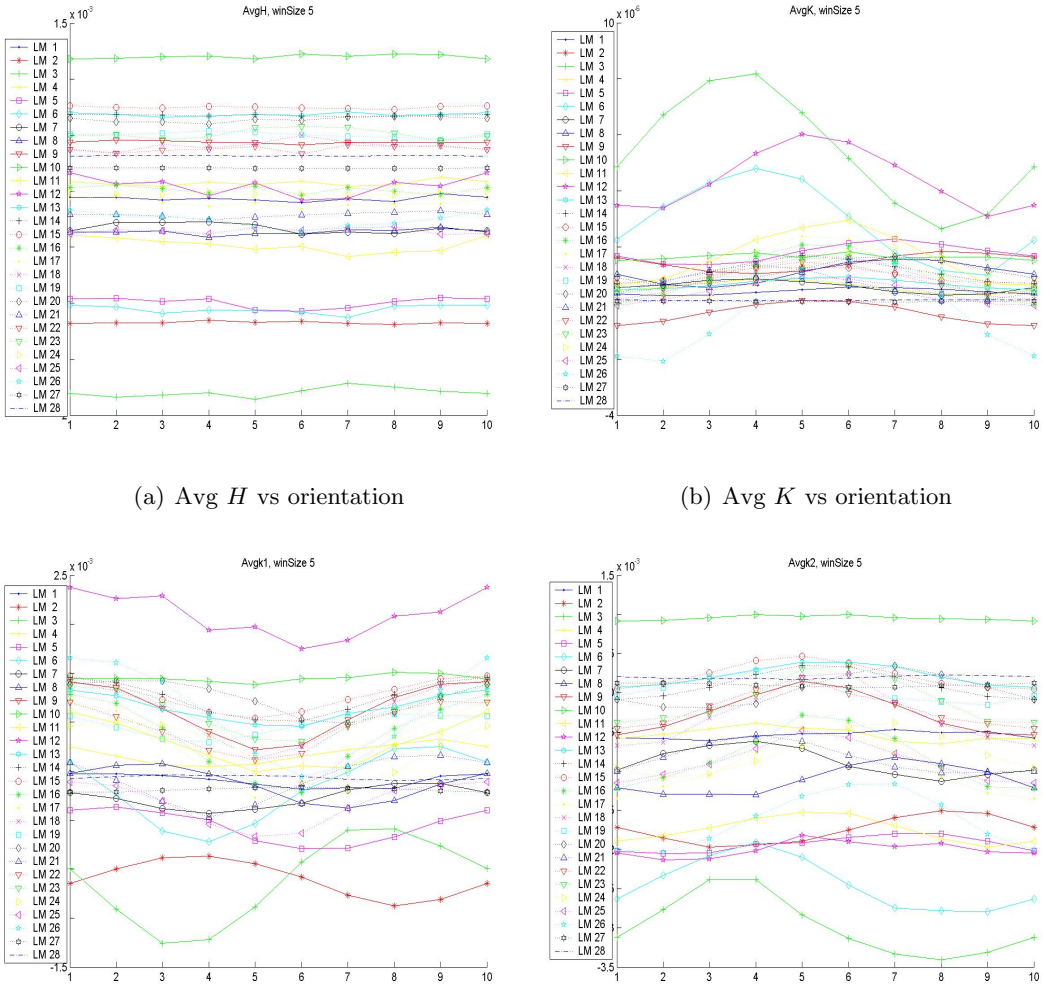
Figure 4.4: 2D feature descriptor for face sample 1 on LM10 (tip of the nose), with sampling aperture 17x17.

4.3.2 Combinations of the Feature Descriptors

Surface properties can be explored using surface types, derived from 3D shape analysis. For example, it is possible to use the values of the H or K curvatures, derived using the first and second derivatives of the image, and produce a histogram of the distribution of the relative frequencies of the values. Similarly a histogram of the distribution of the frequencies of the principal curvatures can be obtained and included as the feature descriptor. Many combinations can be employed to create a feature descriptor, however, the properties and aim of the feature descriptor must be examined: it should be invariant to rotational changes. As a result, the search for a suitable candidate for the feature descriptor can be reduced and examined by investigating the individual candidates and their invariance properties with respect to viewpoint rotational changes on the landmark locations over a measurement aperture (support region).

The candidates to be considered are degree of curvedness, H, K, $k1$ and $k2$ curvatures. By extracting these properties from each landmark location, using different

support regions, over rotational changes, it is possible to identify the candidates which are suitable to be considered as a component in the feature descriptor by taking the average of the values over the population (50 samples). The suitable candidates have to be invariant to viewpoint rotational changes. This investigation uses in-plane rotation (about the z-axis) up to and including 90° and the obtained graphs are shown in Figure 4.5.



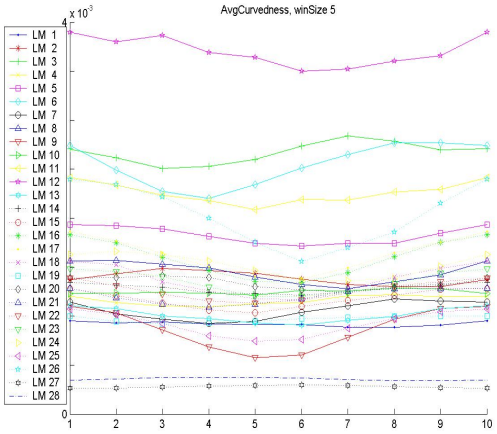
(a) Avg H vs orientation

(b) Avg K vs orientation

(c) Avg $k1$ vs orientation

(d) Avg $k2$ vs orientation

Figure 4.5: (a) Average of H; (b) average of K; (c) Average of $k1$; (d) average of $k2$ taken over 50 samples, extracted from different landmarks with support region 5×5. Each different coloured line represents the results obtained from different landmarks, over in-plane rotational changes.



(e) Avg curvedness vs orientation

Figure 4.5: (e) Average of curvedness taken over 50 samples, extracted from different landmarks with support region 5×5. Each different coloured line represents results obtained from different landmarks, over in-plane rotational changes.

As shown from the graphs in Figure 4.5, the average of the H curvature over the population against rotational changes is stable (Figure 4.5a), therefore, H can be taken into consideration as part of the feature descriptor. Similarly, degree of curvedness (Figure 4.5e) is stable against degree of rotation as well and as a result, it can be considered as a component for the feature descriptor, as well as acting as a weight for the surface types classification. Shown in the following sections are the combinations of the feature descriptors that have been investigated in this work.

4.3.2.1 1D Feature Descriptor Combinations

Shown in the Table 4.1 is the combinations investigated for 1D feature descriptor (i.e concatenated statistics).

	Components	Weighting used for Surface types	Weighting used for Orientation	Length
1	$(HK)^T$	+1	N/A	8
2	(HK)	Curvedness	N/A	8
3	(k1 Orientation)^T	N/A	+1	8
4	(k1 Orientation)^T	N/A	Isotropic Value	8
5	(Shape Index)^T	+1	N/A	9
6	(Shape Index)^T	Curvedness	N/A	9
7	(Image Orientation)^T	N/A	+1	8
8	(HK+k1 Orientation)^T	+1	+1	16
9	(HK+k1 Orientation)^T	Curvedness	Isotropic Value	16
10	(HK+k1 Orientation)^T	Curvedness	+1	16
11	(HK+Image Orientation)^T	+1	+1	16
12	(HK+Image Orientation>)^T	Curvedness	+1	16
13	(Shape Index+k1 Orientation)^T	+1	+1	17
14	(Shape Index+k1 Orientation)^T	Curvedness	+1	17
15	(Shape Index+k1 Orientation)^T	Curvedness	Isotropic Value	17
16	(Shape Index+Image Orientation)^T	+1	+1	17
17	(Shape Index+Image Orientation)^T	Curvedness	+1	17

Table 4.1: Combinations for the construction of 1D feature descriptors.

Note: In Table 4.1, HK denotes the surface types extracted using signs of H and K; $k1$ orientation denotes the orientation extracted from the $k1$ curvature; *Shape Index* denotes the surface types extracted using shape index and *Image Orientation* denotes the orientation extracted from the first Gaussian derivatives of the range image.

4.3.2.2 2D Feature Descriptor Combinations

Table 4.2 shows the combinations of the 2D feature descriptor (i.e. joint probability density functions pdfs) investigated in this work:

	Components	Weighting used for Surface types	Weighting used for Orientation	Length
1	(HK,k1 Orientation)^T	+1	+1	8×8
2	(HK,k1 Orientation)^T	Curvedness	Isotropic Value	8×8
3	(HK,k1 Orientation)^T	Curvedness	+1	8×8
4	(HK,Image Orientation)^T	+1	+1	8×8
5	(HK,Image Orientation)^T	Curvedness	+1	8×8
6	(Shape Index,k1 Orientation)^T	+1	+1	9×8
7	(Shape Index,k1 Orientation)^T	Curvedness	Isotropic Value	9×8
8	(Shape Index,k1 Orientation)^T	Curvedness	+1	9×8
9	(Shape Index,Image Orientation)^T	+1	+1	9×8
10	(Shape Index,Image Orientation)^T	Curvedness	+1	9×8

Table 4.2: Combinations used for 2D feature descriptors.

All of the feature descriptors shown in the Table 4.2 are tested using the bounded [0,1] vector dot product for their invariance against in-plane rotational changes and also their discriminability using the KNN algorithm.

In-plane rotational data are simulated using MATLAB's built-in function, `imrotate`, in which the range images are rotated about the z-axis with bicubic interpolation.

4.4 Validation

The methodology in extracting the 1D and 2D feature descriptors has been described in the previous sections. Alternative methods, and more appropriate to the task of landmark automation using surface analysis, namely the shape index, has also been introduced the previous section. This section details the validation process involved in this investigation.

4.4.1 Data

Similar to the data set used in the pilot investigation detailed in Chapter 3, a data set comprising 50 female frontal facial range images has been employed in this investigation. For the purposes of this investigation, experiments were conducted on a single-scale image representation, where the range image is of size 88×134 pixels, corresponding to Level 9 of the Gaussian half octave pyramid.

4.4.2 Invariance - Vector Dot Product

The invariance properties of the feature descriptors can be validated using the vector dot product, bounded to [0,1], in which the cosine angle between two descriptors are computed (Equation 4.4.1), 1 being the two descriptors are identical. The vector dot product between the feature descriptors extracted from the rotated data and those extracted from the un-rotated data is computed and indicates the similarity between these vectors and as a result the invariance to viewpoint rotational changes to the surface image.

$$cos\theta = \frac{a.b}{|a||b|} \tag{4.4.1}$$

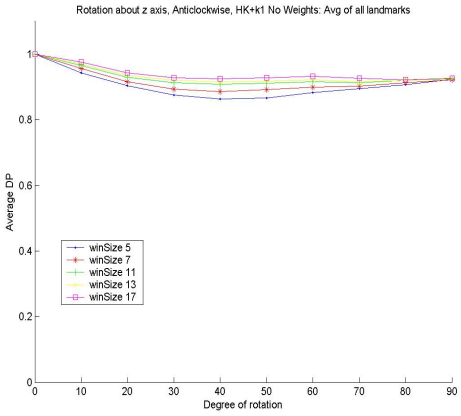
where θ is the angle between the two vectors a and b, $|a|$ and $|b|$ denote the length of a and b. If both a and b have length one (i.e. they are unit vectors), their dot product simply gives the cosine of the angle (bounded [0,1]) between them.

The invariance of the feature descriptors with respect to in-plane rotation has been investigated using different combination of feature descriptors based on the relative frequencies of surface types and their relative frequencies of the orientation, using the bounded [0,1] vector dot product. The results are presented as follows:

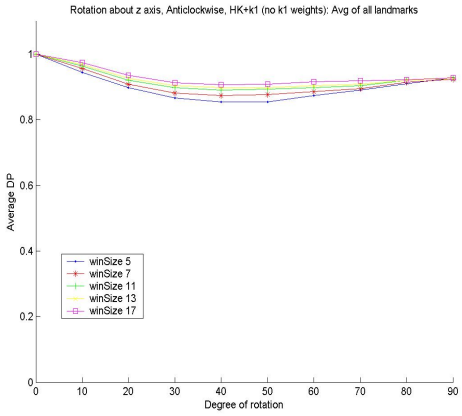
4.4.2.1 Results - In-plane Rotation, 1D Feature Descriptors

The average invariance of the entire sample set (50 faces) and all 28 anatomical landmarks, over in-plane anticlockwise rotational changes up to and including 90° of rotation, are presented here.

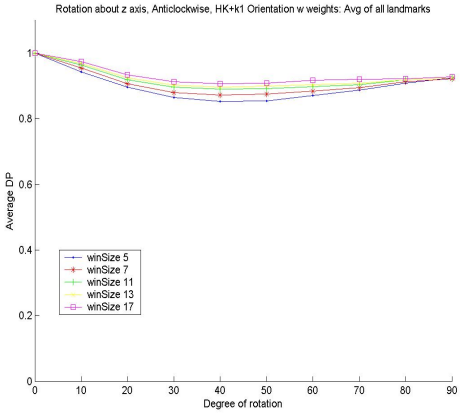
The results for the invariance measure of different feature descriptors combinations are shown in Figure 4.6. The invariance of each of the combinations is good, with all maintaining over 0.75 in invariance (1 being identical). In all cases, measurement aperture size 17 by 17 pixels remains the most invariant over the rotation.



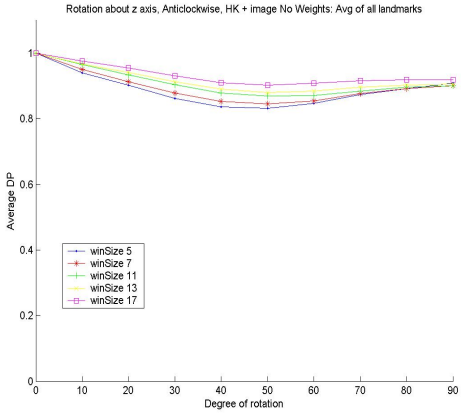
(a) $(\text{HK}(+1)+\text{k1}(+1))^T$



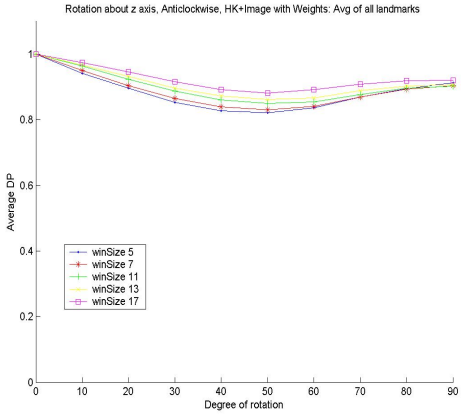
(b) $(\text{HK}(+\text{curvedness})+\text{k1}(+1))^T$



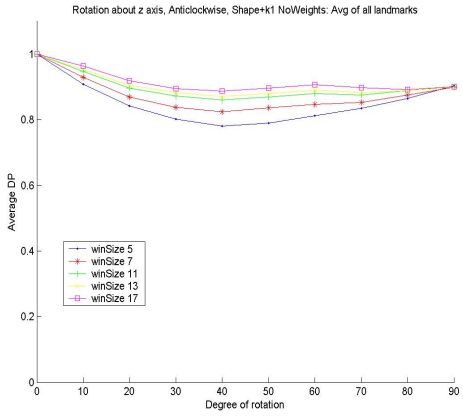
(c) $(\text{HK}(+\text{curvedness})+\text{k1}(+\text{isotropic val}))^T$



(d) $(\text{HK}(+1)+\text{Image direction}(+1))^T$

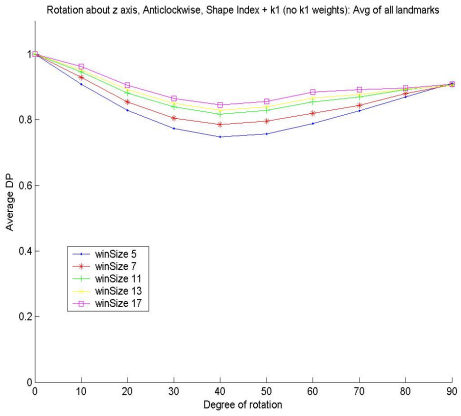


(e) $(\text{HK}(+\text{curvedness})+\text{Image}(+1))^T$

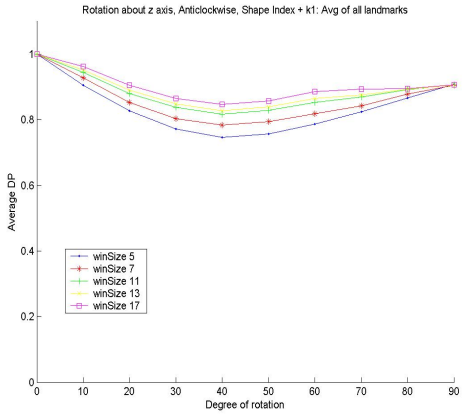


(f) $(\text{Shape Index}(+1)+\text{k1}(+1))^T$

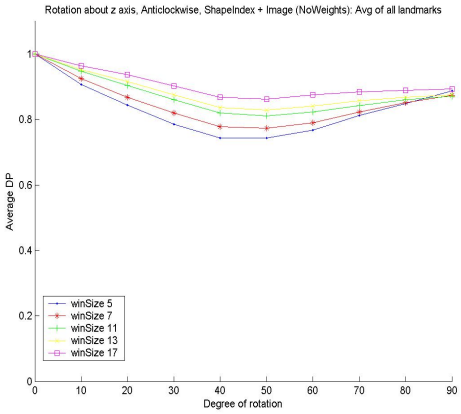
Figure 4.6: Invariance of 1D feature descriptors w.r.t. in-plane rotations.



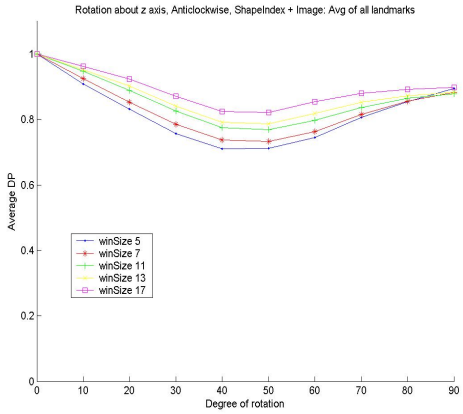
(g) (Shape Index(+curvedness)+k1(+1))^T



(h) (Shape Index (+curvedness)+ k1 (+isotropic value))^T



(i) (Shape Index(+1)+Image direction(+1))^T

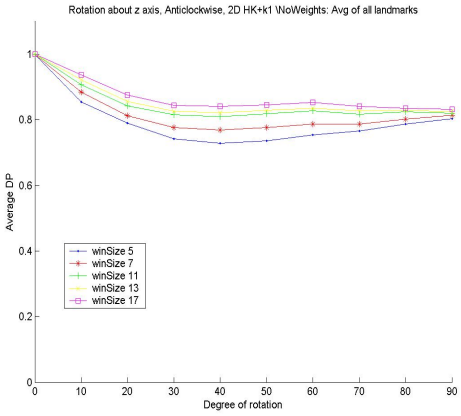


(j) (Shape Index (+curvedness)+ Image direction(+1))^T

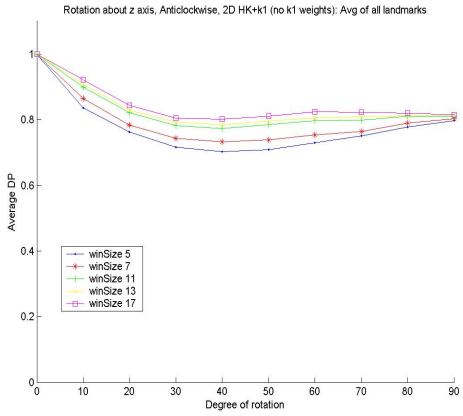
Figure 4.6: Invariance of 1D feature descriptors w.r.t. in-plane rotations (cont.).

4.4.2.2 Results - In-plane Rotation, 2D Feature Descriptors

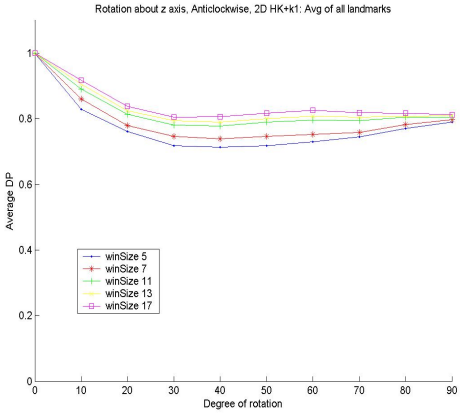
The average results for the invariance of different feature descriptors have been obtained using 50 face samples and 28 anatomical landmarks. In-plane rotational changes up to and including 90° of rotation has been performed.



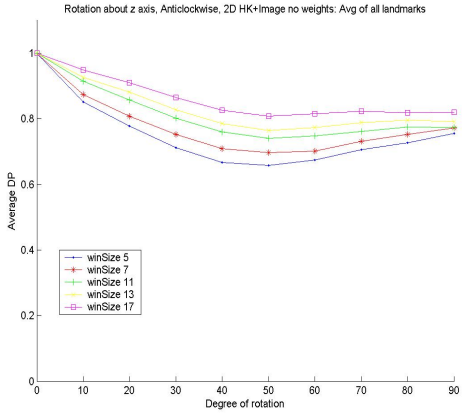
(a) $(\text{HK}(+1),\text{k1}(+1))^T$



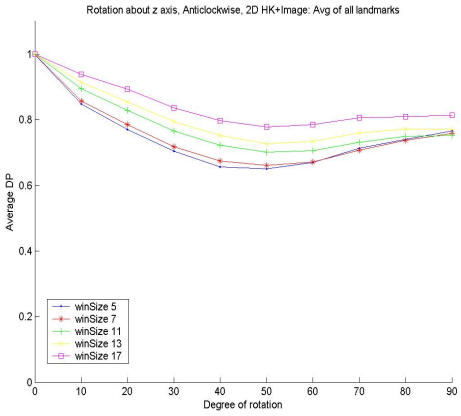
(b) $(\text{HK}(+\text{curvedness}),\text{k1}(+1))^T$



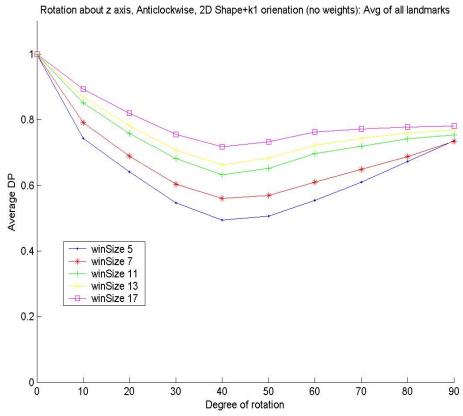
(c) $(\text{HK}(+\text{curvedness}),\text{k1}(+\text{isotropic value}))^T$



(d) $(\text{HK}(+1),\text{Image direction}(+1))^T$

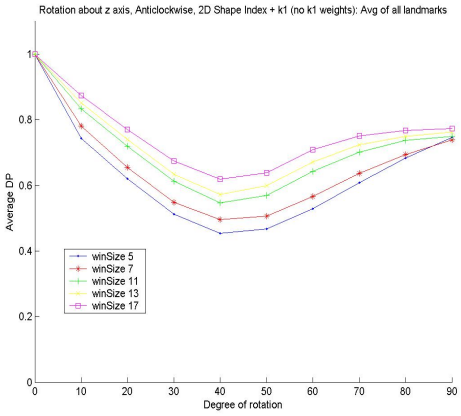


(e) $(\text{HK}(+\text{curvedness}),\text{Image}(+1))^T$

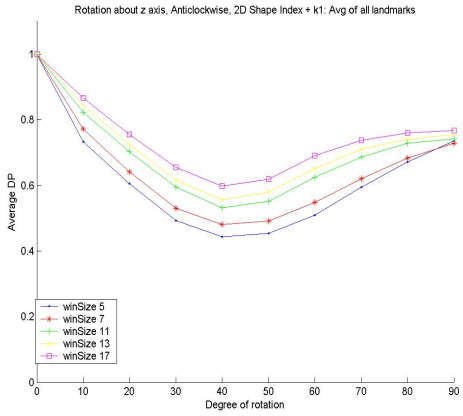


(f) $(\text{Shape Index}(+1),\text{k1}(+1))^T$

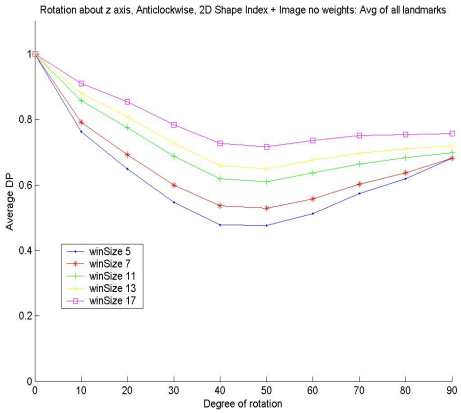
Figure 4.7: Invariance for 2D feature descriptors w.r.t. in-plane rotations.



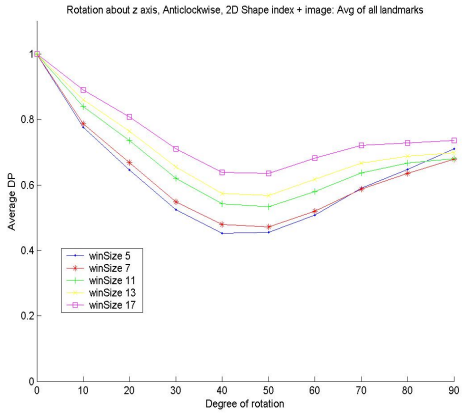
(g) (Shape Index(+curvedness),k1(+1))^T



(h) (Shape Index (+curvedness), k1 (+isotropic value))^T



(i) (Shape Index(+1),Image direction(+1))^T



(j) (Shape Index (+curvedness), Image direction(+1))^T

Figure 4.7: Invariance for 2D feature descriptors w.r.t. in-plane rotations (cont.).

The average invariance for 2D feature descriptors extracted from un-rotated and rotated images shows that it remains over 0.7, using support region 17×17, in all cases.

4.4.3 Matching of the Feature Descriptors using K-Nearest-Neighbour

The matching ability of the feature descriptors can be obtained using the KNN algorithm in which the process was described in Section 3.6. The data set is separated into training and testing data subsets where the Euclidean distance between the training and testing data are computed and a landmark label is provided for each of the testing

data using the majority voting system with k selected to be 1, 3, 5, 7 and 9 respectively.

Mirrored landmarks are investigated by separating the landmarks into left and right sections, depending on the location on the face, using the mid-line drawn from the image, i.e. landmarks representing the *nasion*, *pronasale* and *stomion*. All landmarks which are located to the left of these landmarks are categorised into the left section and similarly, all the landmarks situated on the right of the midpoints are included in the right section. All the landmarks located on the mid-line are included in both left and right sections.

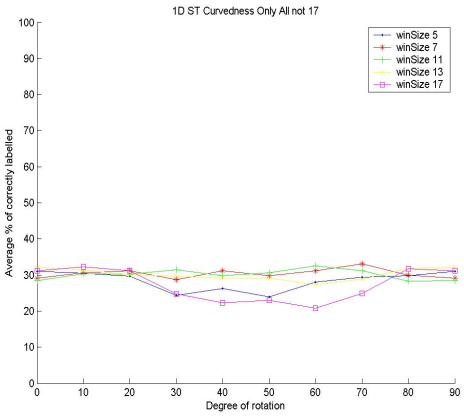
4.4.3.1 Results - In-plane Rotation, 1D Feature Descriptors

Based on the initial findings from the invariance of the feature descriptors in the previous section, shown in the following are a selection of the discriminability graphs obtained using different combination of the 1D feature descriptors.

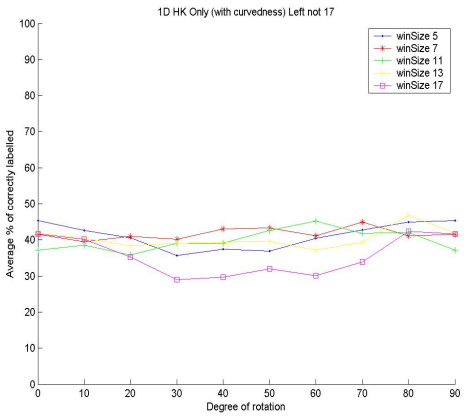
First of all, Figure 4.8(a) shows the average recognition accuracy of a feature descriptor comprised of the surface types, derived from the signs of the H and K curvatures, with the degree of curvedness as weighting system. On average, just over 30% of the landmarks were correctly labelled. Figure 4.8(b) and (c) show the average percentage of the correctly labelled landmarks obtained from the landmarks extracted from the left and right side of the face respectively.

Secondly, Figure 4.9(a) shows the average matching accuracy of a feature descriptor comprised of the surface types, derived from the signs of H and K curvatures, with the degree of curvedness as a weighting system, concatenated with the $k1$ orientation histogram, with the isotropic value as a weighting system. The average percentage of the correctly labelled landmarks is approximately 50%, at window size 17 by 17 pixels when 28 landmarks are taken into consideration. This is an improvement compared to using only one component in the feature descriptors.

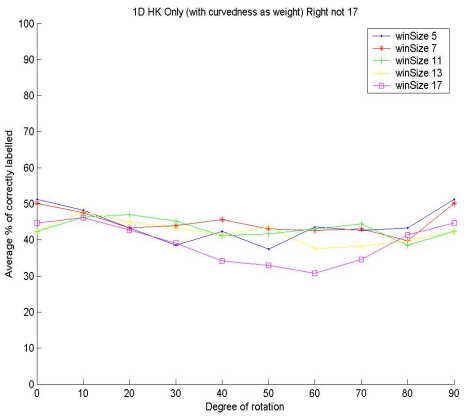
Figure 4.10(a) shows the average percentage of correctly labelled landmarks of a feature descriptor comprised of the relative frequencies of the surface types, derived from the shape index, weighted by the degree of curvedness, concatenated with the distribution of the relative frequencies of the orientation, derived using the first Gaussian derivatives of the image. This combination of the 1D feature descriptors demonstrates an improvement in terms of percentage of correctly labelled landmarks. On average, taken over 28 landmarks, approximately 60% of the landmarks were labelled correctly.



(a) Average of all landmarks

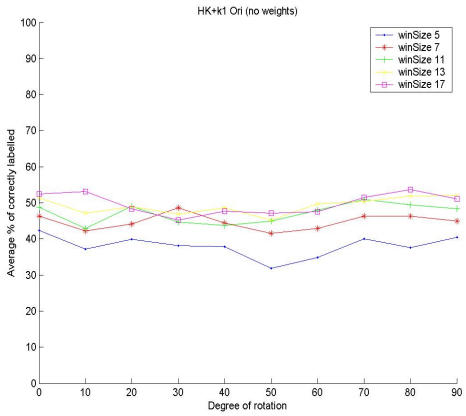


(b) Average of all left landmarks

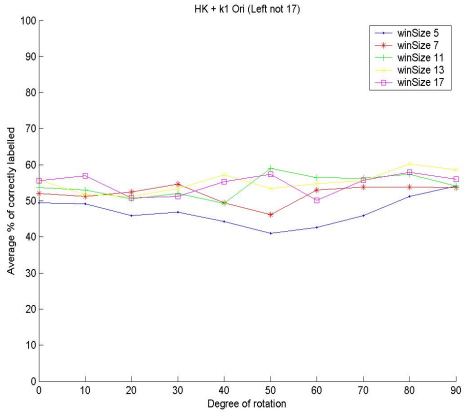


(c) Average of all right landmarks

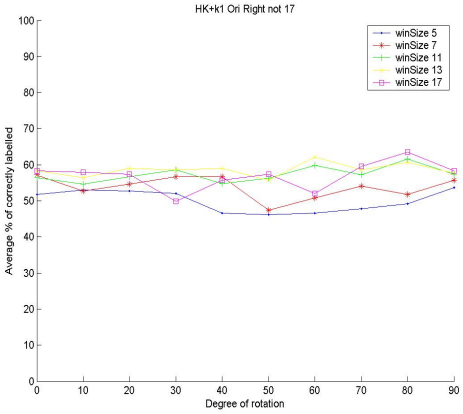
Figure 4.8: Matching accuracy of feature descriptor (HK with curvedness only) against in-plane rotational changes.



(a) Average of all landmarks

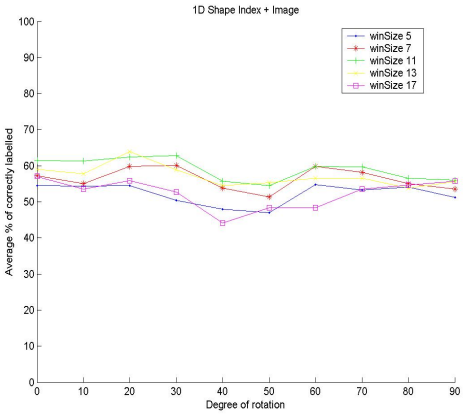


(b) Average of all left landmarks

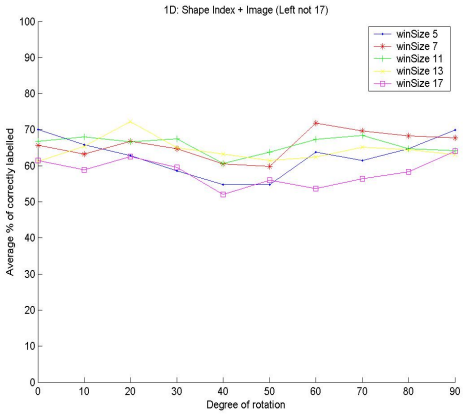


(c) Average of all right landmarks

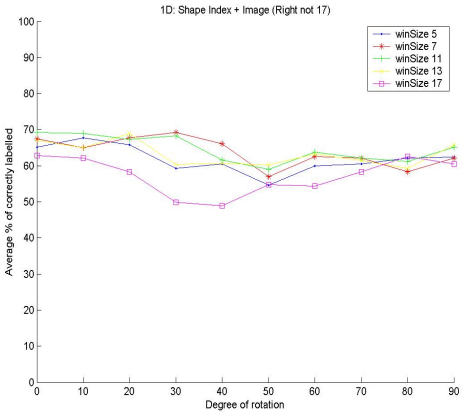
Figure 4.9: Percentage of correctly labelled landmarks for feature descriptor comprised surface types obtained using signs of H and K, concatenate with the $k1$ orientation (with curvedness and isotropic value as weights respectively).



(a) Average of all landmarks



(b) Average of all left landmarks

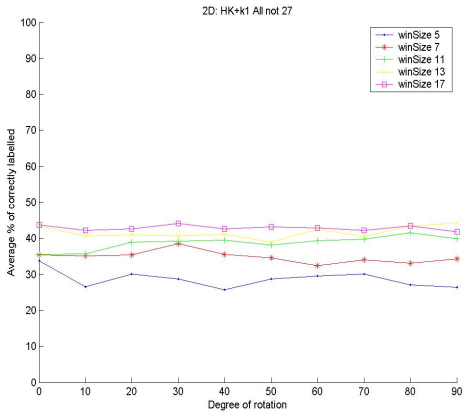


(c) Average of all right landmarks

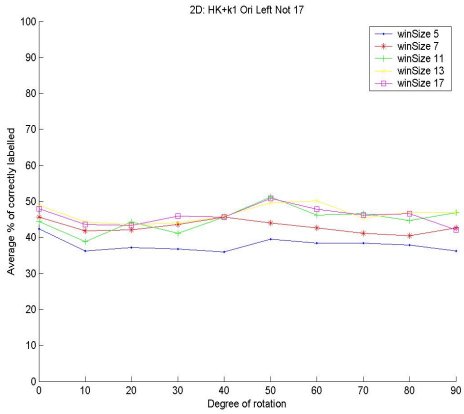
Figure 4.10: Percentage of correctly labelled landmarks for feature descriptor comprised surface types derived from the shape index, weighted by the degree of curvedness, concatenated with the distribution of the relative frequencies of the orientation, derived from the first Gaussian derivatives of the image.

4.4.3.2 Results - In-plane Rotation, 2D Feature Descriptors

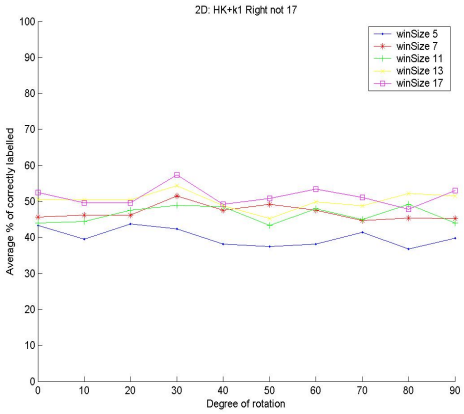
Based on the initial findings from the 1D feature descriptors, the results of two of different combinations of the 2D feature descriptors are presented here. Figure 4.11(a) shows the average percentage of correctly labelled landmarks of a 2D feature descriptor comprised of the relative frequencies of the surface types derived from the signs of H and K curvatures, combined with the $k1$ orientation, taken over 28 landmarks is approximately 45% using a measurement aperture of 17×17.



(a) Average of all landmarks



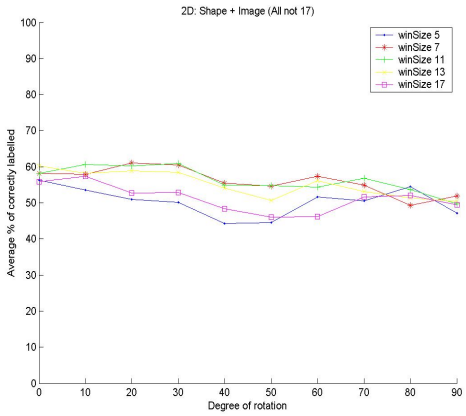
(b) Average of all left landmarks



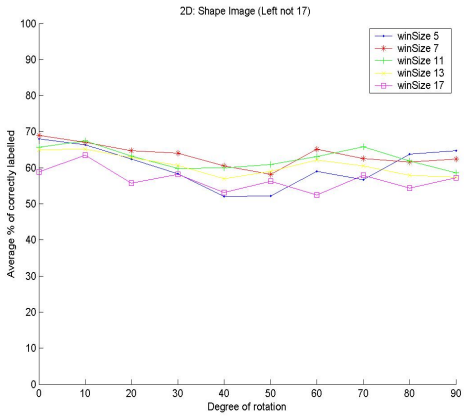
(c) Average of all right landmarks

Figure 4.11: Average percentage of correctly labelled landmarks of a 2D feature descriptor: Surface types (based on the signs of H and K) combined with the $k1$ orientation (with curvedness and isotropic value as weight respectively).

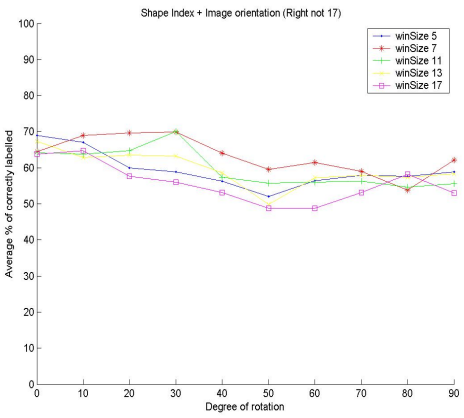
Another combination of the feature descriptor, comprised the nine surface types derived using shape index and the orientation extracted from the first derivatives of the image, provides, on average when taken over 28 landmarks, over 60% of correctly labelled landmarks. When taking 17 landmarks into account (from the right side of the image), approximately 70% of the landmarks are correctly labelled. These are shown in Figure 4.12.



(a) Average of all landmarks



(b) Average of all left landmarks



(c) Average of all right landmarks

Figure 4.12: Average percentage of the correctly labelled landmarks for a 2D feature descriptor comprising the surface types (derived from shape index) and image orientation.

4.5 Summary and Discussions

An in depth investigation for a suitable feature descriptor for range image representation was discussed in this chapter, based on the basic structure of the feature descriptor described in Chapter 3, where the underlying information of the mixes of surface types, along with their mixes of directional/orientations are captured simultaneously. The orientation component of the feature descriptor was normalised to its canonical form in order to achieve the viewpoint invariant properties. This was accomplished by rotating the orientation histogram until the peak is located at the first histogram position. This normalisation process is not required in the surface types histogram because the underlying surface does not change over rotations. By normalising the orientation histogram to its canonical form, the pattern of the orientation distribution histogram will remain constant over rotational change and therefore potentially improves the viewpoint rotational invariant properties that a feature descriptor should have.

In this chapter, alternative structures of the feature descriptors were explored in which shape index was used as the representation of the different underlying surface types of the image as opposed to the signs of the H and K curvatures. Results show that the use of shape index was more appropriate for range image representation of a human face, since every surface types defined by the shape index can be captured and represented within the feature descriptor. Furthermore, the idea of coupled feature descriptors was explored in which a 2D feature descriptors were devised based on the initial findings. While the results obtained for the 2D feature descriptors were more stable, the performance was only marginally better. This could be explained by the "curse of dimensionality" (Bellman, 1957; Duda and Hart, 1973) - the higher the dimension of the feature descriptor, the less discriminable it becomes. Moreover, it was noted that there were many zero entries within the 2D feature descriptor, meaning not every bin was quantised. The results are summarised in Table 4.3 below.

Different combinations of the feature descriptors were explored in this chapter as well. It is noticed that the surface types derived using shape index and orientation derived using the first Gaussian derivatives provide, on average, 60% of discriminability over rotational changes. This is because the surface types defined using the shape index are more suitable for classifying face range data as the shape index provides a wider mixes of surface types, compared to using the signs of H and K curvatures alone.

Dimension	Avg Invariance	Discriminability		
		<HK>	<HK,k1>	<Shape Index, Image Orientation>
1D	>0.75	30%	50%	60%
2D	>0.7	N/A	45%	60%

Table 4.3: Summary of results obtained from this chapter, addressing the invariance and discriminability of different combinations of feature descriptors, against in-plane rotational changes.

Furthermore, the orientation estimates obtained from the range image are based on the first Gaussian derivatives, whereas the direction estimates obtained from $k1$ curvature are based on the second order Gaussian derivatives, resulting in a lower error rate. Therefore, the combination of the histogram of surface types derived from the shape index and the orientation histogram by means of the first image gradient component provides a stronger feature descriptor.

From the findings derived from this chapter, it was noted that a feature descriptor containing either only the surface types or the orientations alone was not enough for range image analysis. Moreover, it was noted that the 1D feature descriptor (containing the information of the relative frequencies of the shape index and the relative frequencies of the image gradient orientations) was the most suitable candidate for a range image representation.

At this stage, a suitable architecture was required to incorporate this feature descriptor so that a complete range image analysis system can be developed. Two predominant systems were considered: SIFT (Lowe, 2004) or the Elastic Bunch Graph (Wiskott et al., 1997). While the Elastic Bunch Graph algorithm was potentially less complex and it has been well validated on the face data, the applied algorithm was very specific to the face only. Since the aim of this work was to provide a potential means to interpret other anatomical surfaces, a more general purpose system was therefore required. Even though the SIFT platform did not have the in-built ability to cope with different biological forms in its existing form, it was believed that it could potentially be adapted to apply to different parts of the body (or even objects). Moreover, as the code to implement standard 2D SIFT was available (El-Maraghi, 2004; Lowe, 2005) and given the time scope of this research project, it was felt that SIFT was a better candidate. As a

result, the decision was made to extend SIFT into the 2.5D domain, providing a more direct route to completing and validating a full 2.5D matching system.

The next chapter will investigate the incorporation of the feature descriptor, derived from the shape index and the image gradients of the range image, into a 2.5D SIFT platform where keypoints and their appropriate scales will be defined on the range images automatically.

Chapter 5

Building a Scale and Rotation Invariant Framework – 2.5D SIFT

A feature descriptor suitable for range image analysis has been devised from the previous two chapters where the underlying surface information is captured by means of the surface types and their orientations simultaneously. The extraction of the feature descriptor described hitherto has been reliant on manual placement of landmarks on the range images and the use of a fixed scale measurement aperture size. However, the process of manual placement of landmarks are subjective, time consuming, error-prone and requires a considerable amount of clinical skills. Therefore, it is necessary to derive an alternative method of localisation of the landmarks by means of mathematical keypoints. This is then followed by the feature extraction process, thereby allowing a fully automatic range image analysis to be conducted. Since the form of the devised feature descriptors is analogous to the structure of Lowe's 2D SIFT (Lowe, 2004), it is possible to incorporate the 2.5D feature descriptor extraction process into a framework similar to Lowe's SIFT. To this end, this chapter will address the extension of standard 2D SIFT into the 2.5D domain.

5.1 Objectives

Feature descriptors are distinctive mathematical keys, capable of providing sufficient descriptive richness to discriminate between different descriptors. Therefore, a feature

descriptor should encapsulate the key characteristics and provide a predominant signature of the underlying surface information and is also expected to be invariant to viewpoint rotational changes. A feature descriptor has been derived from Chapters 3 and 4, which is suitable for range image analysis as it can capture the underlying shape signature together with its orientation. In order to extract feature descriptors from range images, it is necessary to define the locations at which the feature descriptors will be extracted. These locations could be anatomical landmarks, where each landmark will correspond to a certain meaningful biological representation of the feature, for instance the corners of the eye, tip of the nose etc. These landmarks are traditionally placed by professional clinicians since the placement of landmarks require an extensive knowledge of the anatomy of the human body. However, the manual designation of landmarks is labour intensive, time-consuming, error-prone and subjective and therefore an alternative means of deriving key locations for feature extraction is required.

Moreover, following from the previous chapter, in order to accomplish the viewpoint rotational invariant properties for the feature descriptor, the orientation component of the feature descriptor is required to be normalised to its canonical form, i.e. rotated to its peak orientation. Currently, the canonical orientation is defined at the extraction of the feature descriptors stage where eight histogram bins are used to cover the full 360° of rotation. Therefore, the canonical orientation is only accurate to approximately 45°. While eight bins is sufficient for distributing the values in a feature descriptor, it does not provide an accurate measurement of the canonical orientation. As a result, an extra stage is required prior to the feature extraction process, where the canonical orientation of the patch can be determined more precisely.

The aim of this chapter is present the methodology employed to extend the existing 2D SIFT platform into the 2.5D domain, incorporating the 2.5D feature descriptor derived from Chapter 4. There are four main stages in 2.5D SIFT, as follows:

1. **Keypoint localisation**, where the position (x, y) of any stable keypoints, along with their appropriate scale σ, are detected on the range images using *scale-space* representation.

2. **Canonical orientation(s) assignment**, where a consistent **canonical orientation** θ, is assigned to each keypoint landmark location, based on the local image

gradient orientation properties. Multiple canonical orientations can be assigned to a keypoint location, resulting in multiple descriptors for the keypoint.

3. **Extraction of feature descriptor**, based on the keypoints locations, along with their appropriate scale and their canonical orientation(s) information.

4. **Keypoint matching**, where different range images can be matched.

Figure 5.1 illustrates the steps taken for the a full 2.5D SIFT to be accomplished.

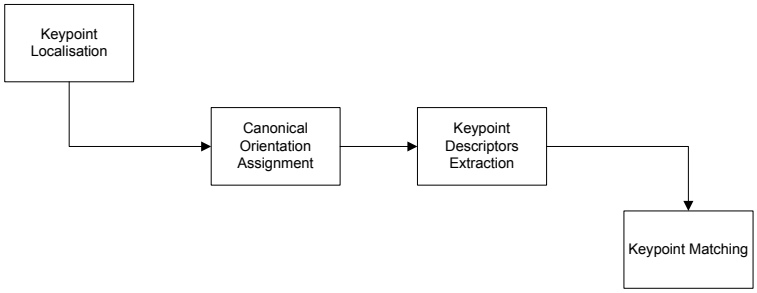


Figure 5.1: Flowchart illustrating the stages involved in 2.5D SIFT.

The remainder of this chapter is organised in the chronological order in which each step in the 2.5D SIFT was investigated. Firstly, following the previous chapter in which the orientation component of the feature descriptor was normalised to its canonical form, Section 5.2 details the steps involved in deriving the canonical orientation(s) from known keypoint locations, prior to the feature extraction stage. Section 5.3 discuss the methodology taken in order to strengthen the discriminative power of the feature descriptor and the incorporation of this feature descriptor to the 2.5D SIFT. Section 5.4 outlines the steps taken to derive the stable keypoints locations and their appropriate scale, which defines the magnitude of the finite support region. Finally, Section 5.5 presents a matching algorithm used to compare and match the individual feature descriptors between images.

5.2 Canonical Orientations

It was deduced from Chapter 4 that the invariance of the feature descriptor improved significantly when the orientation component of the feature descriptor was normalised to its canonical form, i.e. the orientation component histogram was rotated until the peak

is situated at the first histogram position. Since each sampling window will contain a mix of orientations, it is important to identify the most commonly occurring orientation, providing a standardised description of the sampling window over rotational changes.

In Chapter 4, this process is conducted during the feature descriptor extraction stage where eight histogram bins are used to define the orientation component of the feature descriptor. Therefore, the canonical orientation is accurate to 45°, or 1-bin, accuracy. In order to improve this accuracy, a canonical orientation, which is accurate to $\approx \pm 1^{\circ}$ degree using 360 sampling histogram bins, can be assigned prior to the feature extraction stage. Upon the assignment of the canonical orientation to each keypoint, by rotating the measurement aperture sampling patch to its canonical orientation at the feature extraction stage, the same effect can be achieved.

Multiple canonical orientations can also be assigned to a keypoint, resulting in several descriptors for the keypoint, each rotated to align to different orientated structures, and therefore potentially improving the recognition rate. A modified version of Lowe's orientation assignment algorithm has been used for this work. The steps involved in the orientation assignment for each keypoint location over a measurement aperture ($round(4\sigma)$, where σ is the scale at which the keypoints are detected) are as follows:

1. A circular Gaussian mask, set to the detected measurement aperture scale σ, is used to sample the image and the Gaussian is centred on the keypoint location with sub-pixel accuracy (See Section 5.2.1 for more details on sub-pixel accuracy).

2. The local image gradient orientations (Equation 5.2.1) and magnitude (Equation 5.2.2) within the sampling mask are computed using the Gaussian first derivatives (Equation 4.2.2) of the image. A histogram is formulated that comprises 360 bins, each bin containing a relative frequency entry for each of the 360° potentially detectable orientations. Each detected orientation entry is weighted by its corresponding Gaussian derivative magnitude value prior to being accumulated in the appropriate histogram bin. The magnitude is used to provide a measure of the signal strength of the orientated features.

$$\theta_{local} = tan^{-1}\frac{\partial y}{\partial x} \tag{5.2.1}$$

$$\mid \theta_{local} \mid = \sqrt{\partial^2 x + \partial^2 y} \tag{5.2.2}$$

3. The values of the orientation histogram are stabilised, in terms of orientation continuity, by distributing each entry over a number of bins by means of a 1D symmetric Gaussian convolution kernel of $\sigma \approx$7-bins. This step anti-aliases the orientation histogram and stabilises the keypoint canonical orientation allocation process by providing estimates of orientation that change smoothly as the input visual stimulus changes in orientation.

4. The orientation peaks in the filtered histogram are located and each peak within 80% of the magnitude of the largest peak is deemed to represent a keypoint. Sub-bin orientation precision is obtained by interpolation. A quadratic polynomial is fitted to the three histogram values closest to the peak, as shown in Equation 5.2.3 below:

$$\begin{pmatrix} (\theta_{peak} - \Delta\theta)^2 & \theta_{peak} - \Delta\theta & 1 \\ \theta_{peak}^2 & \theta_{peak} & 1 \\ (\theta_{peak} + \Delta\theta)^2 & \theta_{peak} + \Delta\theta & 1 \end{pmatrix} \begin{pmatrix} a \\ b \\ c \end{pmatrix} = \begin{pmatrix} H(\theta_{peak} - \Delta\theta) \\ H(\theta_{peak}) \\ H(\theta_{peak} + \Delta\theta) \end{pmatrix} \tag{5.2.3}$$

   where $H(\theta_{peak})$ is the histogram value at θ_{peak}.

5. The canonical orientation of the descriptor is given by:

$$\theta_{canonical} = -\frac{b}{2a} \tag{5.2.4}$$

5.2.1 Sub-Pixel Accuracy

The keypoint locations can be generated at floating points coordinates (x_i, y_i) in order to achieve sub-pixel accuracy. Sampling image pixels (x, y) with a kernel based on the keypoints centred at (x_i, y_i) requires the calculation of the horizontal and vertical sub-pixel offset $(\text{offset}_x, \text{offset}_y)$ of the centre of the floating point position from the actual kernel integer location. The equation for a symmetric two-dimensional Gaussian square kernel with scale σ_i used to place Gaussian support regions on an image with sub-pixel accuracy is as follows:

$$G = \frac{1}{2\pi\sigma^2} e^{-\frac{(x_i - round(x_i) - \text{offset}_x)^2 + (y_i - round(y_i) - \text{offset}_y)^2}{2\sigma^2}} \tag{5.2.5}$$

As a result, the centre of the measurement aperture window is required to be adjusted to the sub-pixel offsets accordingly. The computation of the sub-pixel offset

depends on whether the rounded integer size of the support region ($round(4\sigma)$) is odd or even. The change (error) in the position of the centre of the keypoint by the rounding operation can be significant. Figure 5.2 illustrates the calculation of the centre of a support region for even and odd sized kernels.

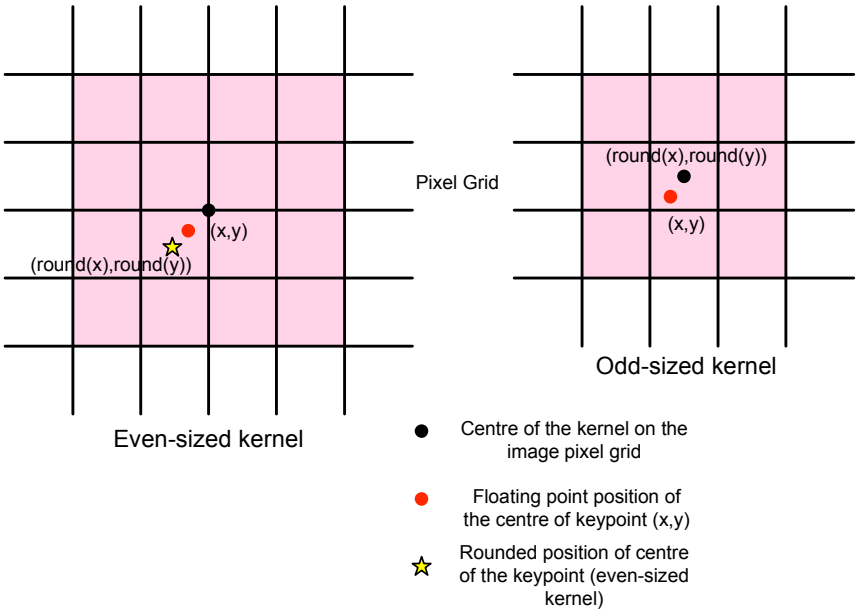


Figure 5.2: Calculating the centre of the support region for even and odd kernel. In the example illustrated here, the kernels are square, not circular.

In order to generate a Gaussian circular support region, apply the equation of an unit circle to the Gaussian square support region to filter any region outwith the circle to zero. In other words, if the location of the pixel within the square kernel satisfies the following equation:

$$(x_i - \text{offset}_x)^2 + (y_i - \text{offset}_y)^2 > \left(\frac{\text{kernel size}}{2}\right)^2 \tag{5.2.6}$$

Finally, the filter coefficients of the Gaussian support region G are normalised to sum to unity to satisfy the following equation:

$$\sum_{\forall x, \forall y} G = 1 \tag{5.2.7}$$

Number of bins	Precision to nearest °
8	45
10	36
18	20
36	10
360	1

Table 5.1: Table showing the different numbers of bins used to derive the optimal number of bins for the categorisation histogram.

Number of bins
0
3
5
7
9

Table 5.2: Table showing the different number of bins tested to derive the optimal number of bins used for the smoothing of the categorisation histogram with 360-bins.

5.2.2 Validation

In order to validate the canonical orientation assignment process, a set of synthetic 2.5D data simulated with known in-plane orientations are used for the validation. The synthetic range data of an orthogonal spatial sine waves is once again employed by firstly locating the fixed keypoints locations on the image and then rotating it with known amount of in-plane orientation. The set of un-rotated and rotated images are then feed into the canonical orientation algorithm where the canonical orientation is estimated and recovered.

The purpose of this experiment is twofold: it first tests the sub-pixel accuracy algorithm to shift the Gaussian centre kernel accordingly and secondly it tests the robustness of the recovery of the canonical orientation from known amount of rotations. The experiment was conducted repeatedly using different numbers of histogram bins for the categorisation (Table 5.1) as well as the smoothing of the histogram (Table 5.2) stages. Experiment shows that by using 360 bins covering the full 360° of rotation and using seven bins to smooth the histogram in this case yields the best results where the recovered canonical orientation is corrected to $\pm 1.0°$.

Figure 5.3 gives the results from a pilot investigation conducted in which a patch is synthetically rotated to a known orientation (denoted as black $\circ$ in the graph) and

the orientation is subsequently recovered using the orientation recovery algorithm as described above, shown as red + on the graph, illustrating the accuracy of the canonical orientation recovery algorithm.

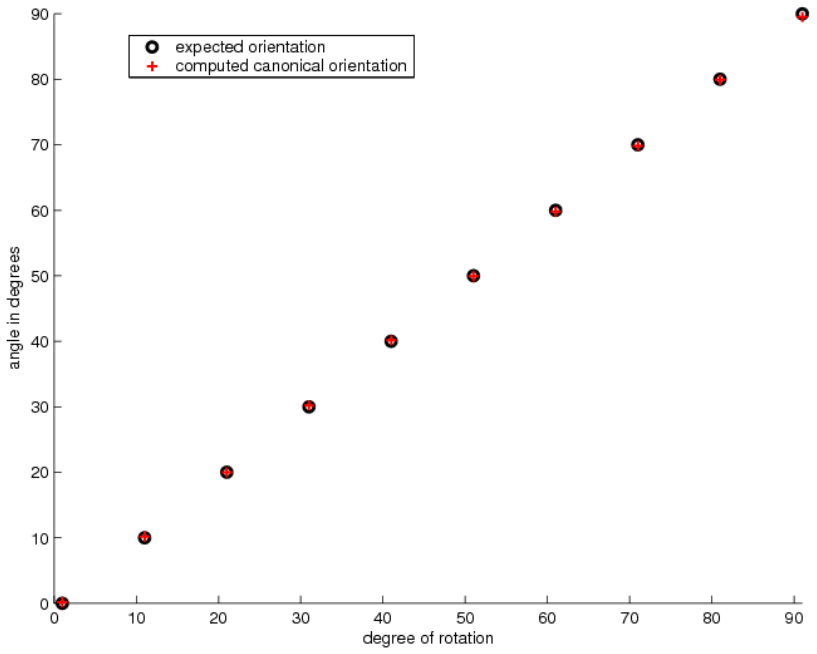


Figure 5.3: This graph shows the recovered orientation of a synthetically rotated patch using the orientation recovery algorithm (red +), against known orientation (black ∘).

5.3 Keypoint Descriptors

For each keypoint (x, y) location, a feature descriptor is extracted over a measurement aperture of σ defining the scale of the keypoint, using the appointed canonical orientation(s) θ. The keypoint location can be a floating point and therefore appropriate adjustment is required to be made in order to place the measurement aperture over the keypoint to sub-pixel accuracy.

In order to incorporate the feature descriptor described hitherto into a suitable 2.5D SIFT framework, the following steps are proposed:

1. The image patch, comprises the Gaussian circular measurement aperture placed over the keypoint, is rotated to its canonical orientation in order to achieve view-

point rotation invariance. Differential geometry is used to compute the signs of the H, K, $k1$ and $k2$ curvatures from the first and second Gaussian derivatives (see Chapter 4), which provide more stable and accurate estimates. Thereby it is possible to categorise the underlying distribution of the surface types, using the bounded $[-1, 1]$ local shape index (Equation 4.2.4). The degree of local curvedness (Equation 2.3.6), along with the local image gradient orientation (Equation 5.2.1) and the corresponding local magnitude estimate (Equation 5.2.2) can also be computed from the first and second Gaussian derivatives.

2. Nine Gaussian weighted sub-regions (Balasuriya and Siebert, 2006; Balasuriya, 2005) are placed over the sampling patch, as shown in Figure 5.4. Each of the nine regions is overlapped by one standard deviation in order to minimise the spatial aliasing occurred during the sampling stages and to provide greater continuity within the descriptor. Since overlapping the Gaussian sub-regions results in the feature descriptors extracted from adjacent subregions being correlated, this reduces spatial aliasing and also enforces spatial continuity that occurs during sampling. For example, small shifts in the location of the keypoint will now result in small (continuous) changes in the magnitude of the composite keypoint descriptor (and its component vectors). The choice of how many sub-regions to employ is a trade-off between excessive dimensionality and feature discriminability, particularly to symmetric patterns. A sampling configuration comprising 3×3 overlapped matrix had been found by Balasuriya to achieve a good working compromise.

3. For each of the nine sub-regions placed over the sampling patch, a local distribution histogram of the relative frequencies of the nine surface types is computed, weighted by the degree of curvedness. Similarly, an eight-element histogram, covering the 360° range of orientations, can be formulated, weighted by the magnitude. Each histogram is normalised to unity magnitude (i.e. to a unit vector) and the influence of large histogram values in each normalised histogram is reduced by clipping the value at a threshold of $\frac{1}{\sqrt{a}}$ where a is the number of bins in the histogram. This means that the distribution of orientations has greater emphasis. The histograms are then concatenated to form H_i, which is then normalised to

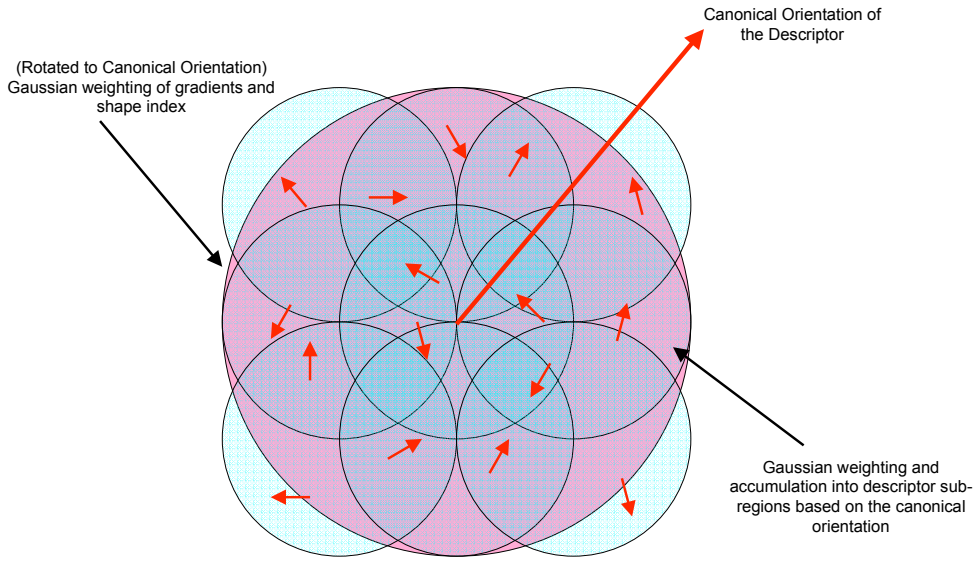


Figure 5.4: Placement of the nine sub-regions, with the spatial support at one standard deviation, over the keypoint location.

unity magnitude.

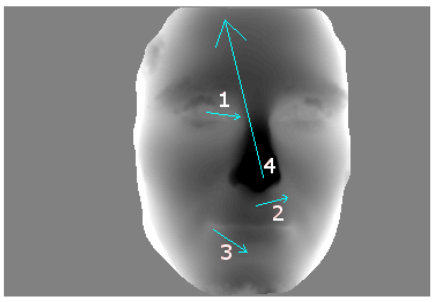
$$\widehat{LocalHist_i} = \left(\widehat{H_{surface}}\right)\left(\widehat{H_{orientation}}\right) \tag{5.3.1}$$

This normalisation step provides invariance to curvature scale (used for weights). The clipping (as per Lowe (2004)) prevents range surface outliers, i.e. spikes, from affecting the local signatures.

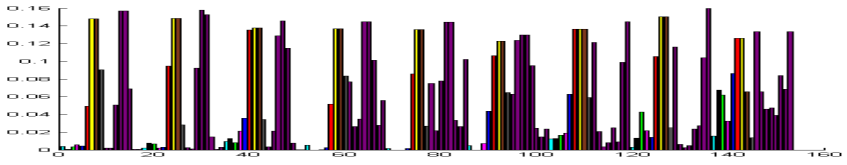
4. The nine normalised histograms $LocalHist_i$ are concatenated to form the final feature descriptor:

$$Descriptor_{\theta_{canonical}} = \left(\widehat{LocalHist_1} + \widehat{LocalHist_2} + ... + \widehat{LocalHist_9}\right) \tag{5.3.2}$$

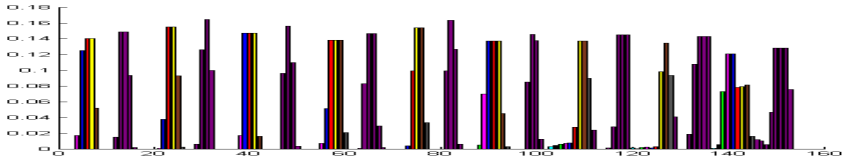
Figure 5.5 illustrates a selection of feature descriptors extracted from four different keypoints on the face range image. The keypoints selected are shown in Figure 5.5(a), where the corresponding descriptor number have been labelled accordingly. This figure illustrates the uniqueness of the feature descriptors extracted from different locations on the face.



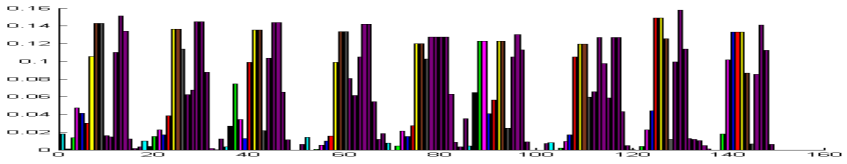
(a) Selected keypoints



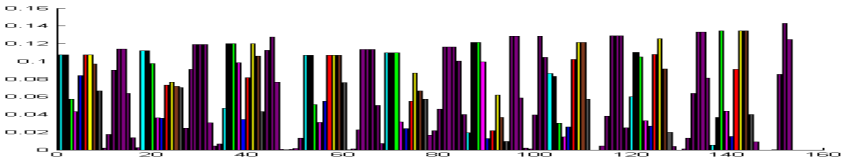
(b) Descriptor extracted from keypoint 1



(c) Descriptor extracted from keypoint 2



(d) Descriptor extracted from keypoint 3



(e) Descriptor extracted from keypoint 4

Figure 5.5: A selection of descriptors extracted from four different keypoints on the face range image, over nine overlapping sub-regions. The colours correspond to the different surface types the underlying information have been assigned to (see Figure 4.2 for the colour chart), along with the orientation (shown as purple bars).

5.4 Keypoint Localisation

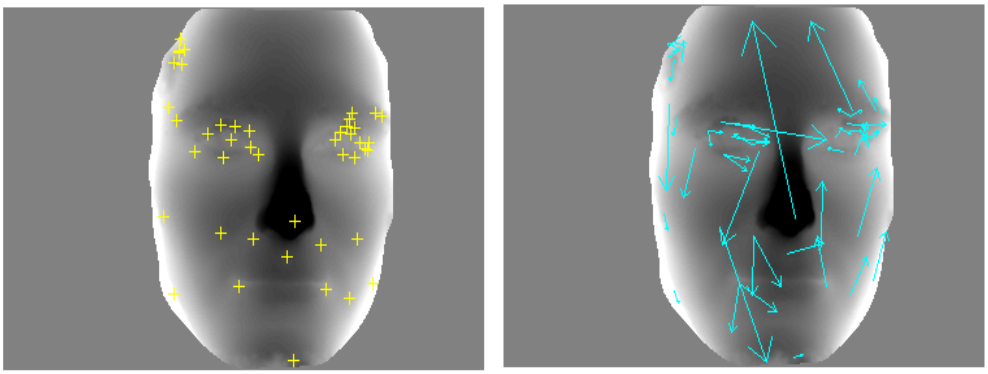
The (x, y) location of stable keypoints are detected on the range images using a modification of Lowe's keypoint localisation algorithm, in order to accommodate the 2.5D modality accordingly. A Gaussian-tapered segmentation mask is applied to the range image in order to isolate the area of interest while avoiding sharp boundaries which would result in providing false keypoints, and the image is then z-normalised to mean $\mu = 0$ and standard deviation $\sigma = 1$ to standardise the (potentially large) dynamic ranges of values present in range image. The z-normalised image is then blurred with a factor of 0.5 to suppress aliasing and is then up-sampled by a factor of two using linear interpolation. Linear interpolation was selected for this work because the computational effort taken for bicubic interpolation is too expensive for the size of images used in this work.

This is followed by the creation of a discrete *scale-space* representation (Lindeberg, 1994*a*,*b*; Mikolajczyk, 2002; Mikolajczyk and Schmid, 2004) of the range image using the Gaussian and the Difference-of-Gaussian (DOG) pyramids (Lowe, 2004) with sub-interval layers. The signal maxima and the minima are detected within the DOG scale-space and potential keypoints with a low contrast (compared to a user-defined contrast threshold of 0.003) are rejected. The H, K, $k1$ and $k2$ curvatures are then computed for each sub-level using the first and second Gaussian derivatives parameterised with σ corresponding to that of the scale-space. This process provides more stable range surface gradient estimates by employing Gaussian smoothing in the calculation of the derivatives. By comparing the ratio of the principle curvatures $\frac{k1}{k2}$ to a curvature threshold $r = 5$, spatially compact feature locations are successfully located. As a result, a set of (x, y, σ) values to each keypoint location are detected.

Figure 5.6(a) shows a set of keypoints located on a range image using the 2.5D keypoint localisation algorithm and Figure 5.6(b) illustrates the position, scale (shown as the magnitude of the arrows) and the canonical orientation(s) (shown as the directions of the arrows) for each keypoint.

5.5 Keypoint Matching

The discriminability of the extracted feature descriptors against viewpoint rotational changes can be determined by matching the feature descriptors extracted from differ-



(a) Keypoints located (b) Keypoints, scale and orientations

Figure 5.6: (a) Keypoint locations (shown as +) extracted using the modified SIFT keypoint localisation algorithm. (b) The scale (demonstrated by the magnitude of the arrows) and the canonical orientation(s) (the directions of the arrows) for each keypoint locations.

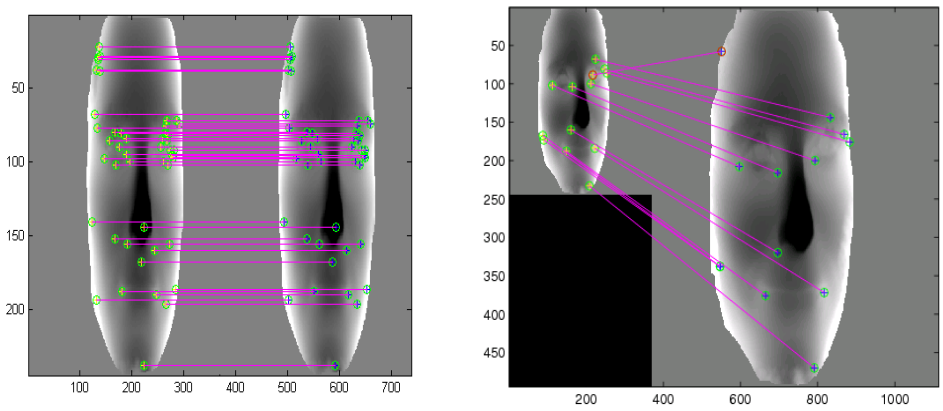
ent set of images, captured at different angles. Following the methodology proposed by Lowe (2004), by using the nearest neighbour algorithm, a candidate is located by computing and ranking (in ascending order) the angle between the descriptors using vector dot product. False matches can be initially rejected using the likelihood ratio test if the ratio between the potentially best matched descriptor to its next best matched descriptor is above a `distRatio` threshold of 0.8 (Equation 5.5.1).

$$\text{Match} = \begin{cases} 1, & \text{if } \dfrac{\text{val}_1}{\text{val}_2} < \text{distRatio} \\ 0, & \text{otherwise} \end{cases} \tag{5.5.1}$$

In order to verify matches between two different range images (captured at different angles), a similarity transform is computed between the two sets of descriptors by means of the Hough Transform. Clusters of matching features with a consistent interpretation (i.e. matches between features exhibiting the same relative shift in orientation, translation and scale) are identified, in other words, a similarity transform between a test set of descriptors and an image in the database is computed using the Hough transform. If three or more entries are located in each cluster, it is possible to apply a robust affine transform fitting procedure to the cluster in order to recover the affine pose between the matched features and also identify outliers. This process matches reliably a set of

extracted feature descriptors to sets of feature descriptors contained in a database and extracted from range images captured at different angles.

Figure 5.7(a) shows a self-matching range image at the same scale while Figure 5.7(b) shows the same range image being matched to an enlarged version of the image.



(a) Self-matching, fixed scale

(b) Self-matching, varied scale

Figure 5.7: Examples of point-to-point matching of the range images where (a) shows a self-matching range image and (b) shows the same image being matched to an enlarged version of the image.

5.5.1 Hough Transform

The Hough transform (Ballard, 1981; Duda and Hart, 1972) is used to identify clusters of features that have a consistent interpretation of an object hypothesis by a voting procedure, where the object hypothesis contains not only the object label but also its position, scaling and rotation (in this work). The Hough transform is especially useful when there are a high proportion of outliers in the matched feature descriptors.

The Hough transform maps descriptor matches from spatial coordinates in the visual scene to a hypothesis voting accumulator space to eliminate outlying object, position or pose hypotheses which accumulate fewer votes. Feature descriptor matches vote into the Hough accumulator space which is parameterised by the underlying degrees of freedom considered within the problem domain: translation (in plane), rotation (in plane) and scale in size.

5.5.2 Affine Transformation

Lowe's methodology (Lowe, 2004) is applied directly for the estimation refinement here. Once the Hough transform has identified three or more entries in each cluster, the affine pose between the matched features can be recovered, thereby allowing outliers to be located, using affine transformation.

If $f(x, y)$ and $f'(x', y')$ are the feature descriptors from training and test respectively, the transformation of the object from the training image to the test image may be accurately given as follows:

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} m_1 & m_2 \\ m_3 & m_4 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} t_x \\ t_y \end{bmatrix} \tag{5.5.2}$$

where m_1, m_2, m_3, m_4 and t_x, t_y are the parameters of the affine transformation of the object from the training appearance view to the test scene. These may be determined by solving the following the least squares system where a single match $f(x, y)$ and $f'(x', y')$ is indicated. Since there are six unknowns, at least three match pairs (six equations) will be needed to determine transformation parameters.

$$\begin{bmatrix} x' \\ y' \\ \vdots \end{bmatrix} = \begin{bmatrix} x & y & 0 & 0 & 1 & 0 \\ 0 & 0 & x & y & 0 & 1 \\ & & \cdots & & & \\ & & \cdots & & & \end{bmatrix} \begin{bmatrix} m_1 \\ m_2 \\ m_3 \\ m_4 \\ t_x \\ t_y \end{bmatrix} \tag{5.5.3}$$

5.6 Summary and Discussions

This chapter presents a complete implementation of the SIFT algorithm in the 2.5D domain. The aim was to incorporate the feature descriptor devised from the previous two chapters into a suitable framework, thereby allowing range images to be represented and matched. The nature of the devised feature descriptor inspired the idea of the 2.5D SIFT implementation since the structure of the feature descriptor is analogues to Lowe's 2D SIFT feature descriptors. Similar to standard SIFT, there are four main stages in the 2.5D SIFT: keypoint localisation, canonical orientation assignment, feature descriptor extraction and finally the keypoint matching.

By adopting statistical normalisation of the input range images it becomes possible to set a consistent set of parameters appropriate to detecting stable keypoint locations and their appropriate scales (independently of the dynamic range of the input range maps or their content). In the standard SIFT formulation, viewpoint rotation invariance is achieved by rotating the range image patch sampled at a keypoint location to its canonical orientation. The stability and accuracy of this keypoint orientation estimator was improved in this work by increasing the orientation histogram resolution, smoothing this histogram in line with the increased angular resolution and then, by applying polynomial interpolation, to recover the canonical orientation with sub-bin ($\pm 1^{\circ}$) precision.

In order to capture a perceptually significant description of the local surface patches sampled at keypoint locations, the 2.5D keypoint descriptors had been formulated to sample the underlying relative frequencies of surface types present. The feature descriptor is based on histogramming the shape index computed over the sampled patch and is weighted by the degree of local surface curvedness. A second component of the keypoint descriptor comprises the histogrammed local orientations within the patch weighted by the local gradient magnitude values. Thereby the keypoint signatures based on surface shape, degree of curvature and the relative orientations have been captured. Furthermore, potential sampling effects caused by spatial aliasing within the standard SIFT keypoint descriptor have been minimised by placing nine Gaussian weighted sub-regions, with spatial support of one standard deviation, over each sampled keypoint location at the detected scale. The aim of this feature descriptor is to increase the invariance properties of the feature descriptor to both Euler's in-plane and out-of-plane rotations.

The next chapter will present the validation on this 2.5D SIFT system using range images of a human face and a mannequin head captured at different angles, in order to determine the invariance properties of the feature descriptors against rotational changes. The validation will therefore be broken down into two main categories: a) in-plane and b) out-of-plane rotational changes. The in-plane rotational changes can be simulated synthetically using MATLAB's built-in `imrotate` function, along with the `bicubic` interpolation in order to avoid aliasing of the resulting rotated images. The out-of-plane rotational data used in the validation consists of real data captured by the single-pod stereo-pair system described in Chapter 3.

Keypoints are then located within each of the data set, followed by the canonical orientation assignment, thereby allowing feature descriptors to be extracted for each keypoint from each of the test images. These feature descriptors, along with the (x, y, σ, θ) information, are recorded in a database where each of the descriptors extracted from the baseline range image can be compared to each of the feature descriptors extracted from the testing images, resulting in an estimation of the stability of the feature descriptors against rotational changes.

Chapter 6

Keypoint Matching Validation

The theory and methodology of the 2.5D SIFT system was presented in Chapter 5 where stable keypoints are located at various locations in the range images, followed by the recovery of the canonical orientation(s) of surface patches, thereby allowing feature descriptors to be extracted at each keypoint location using the (x, y, σ, θ) information. This chapter will validate the invariance properties for the extracted feature descriptors from the 2.5D SIFT system.

6.1 Objectives

The aim of this chapter is to investigate and address the invariance properties of the feature descriptors extracted from the 2.5D SIFT system against rotational changes. For the purposes of this validation, the investigation was to be conducted on fixed scale images (i.e. images of fixed size). The scale of the image chosen for this investigation is approximately of size 244×369 pixels, which is similar to the standard scale Lowe used in his experiments for 2D SIFT (Lowe, 2004).

A set of stereo-pair images (of a human face and a mannequin face respectively) have been captured and processed where the resulting range images are employed as the test data for this investigation. The process involved in capturing and processing stereo-pair images was described in Section 3.2. The data set comprises images captured face-on and then at approximately 10° intervals in both clockwise and anticlockwise directions up to approximately 90° of rotation (where the profile of the face will be captured by the cameras) in order to generate Euler's out-of-plane rotations about the yaw axis.

In order to simulate Euler's in-plane rotational changes, the range images have been rotated synthetically using MATLAB's standard `imrotate` function at 10° intervals.

Once the range images have been pre-processed and standardised, including the size reduction of the range images using an half-octave Gaussian pyramid and the derivation of the resulting Gaussian mask image, they can be passed into the 2.5D SIFT system described in the previous chapter where feature descriptors are extracted from each keypoint location. In order to determine the invariance properties of the 2.5D SIFT system, the feature descriptors extracted from each range image are compared against the feature descriptors extracted from the range images captured at a different angles. The Hough Transform (HT) is then used to detect potential matches and a match is confirmed by fitting an affine transform on each cluster of three or more potential matches where outliers can be identified and rejected.

The results obtained can then be represented in a match-matrix showing the percentages of matched and filtered keypoints obtained for matches between range images depicting all combinations of viewpoints. An initial indication of the performance of the system can therefore be established by noting the degree of invariance to rotation attained for each acquired viewpoint.

The remainder of this chapter is organised as follows: Section 6.2 details the plan for the validation of the 2.5D SIFT system, in particular the invariance of the feature descriptors against rotational changes. Section 6.3 presents the validation results obtained for this investigation and finally Section 6.4 concludes this investigation.

6.2 Validation Plan and Methodology

This section presents the proposed plan to validation the 2.5D SIFT system with respect to rotational changes. It is important for a feature descriptor to represent the distinct surface shape signature in order to differentiate between surface locations, it is also vital for the feature descriptor to exhibit strong invariance under rotational changes about all three Euler's exes. However, in this pilot investigation, only the Euler's in-plane rotation and out-of-plane rotation about the yaw axis were taken into consideration. The invariance of the feature descriptors against rotational changes can be measured by attempting to match the feature descriptors extracted from a certain range image against a set of feature descriptors extracted from another (rotated) range image. The

percentage of matched keypoints can be determined and as a result, the invariance can be measured.

6.2.1 Data Set

It was possible to capture only a modest data set within the scope of this research project in order to validate the 2.5D SIFT system. This data comprises two sets of range images captured at different viewpoint angles (one set of range images of a human face and one set of range images of a mannequin head). The aim of the validation is to test the invariance of the feature descriptors against viewpoint rotational changes, therefore two main types of data set have been used for this investigation, namely the range images simulating the Euler's in-plane and out-of-plane rotational changes. A total of 38 captured out-of-plane images (2×19) and 72 synthetically generated in-plane (2×36) rotational images are used for this investigation.

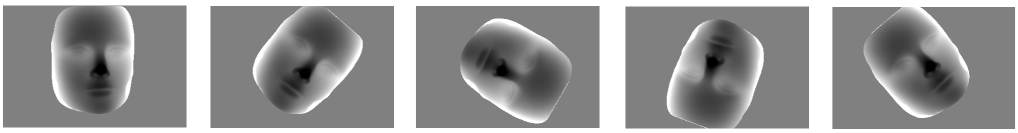
The range images of a human face and a mannequin head, captured face-on (i.e. at 0°), are referred to as the baseline images here. These have been generated from the stereo-pair images captured using a single pod stereo-pair capturing system and processed by a stereo-pair package, C3D2, where the details can be found in Section 3.2.

The range images are pre-processed using HIPS in order to eliminate the header and suppress any random noise using ASL (see Section 3.3.1). Furthermore, for the purposes of this investigation, the images are down-scaled from 1498×2249 pixels to 244×369 pixels (equivalent of Level 6 of the half-octave pyramid) using an half-octave Gaussian pyramid.

The baseline images are then used to generate a set of in-plane rotated images, detailed in the next section.

6.2.1.1 In-Plane Rotational Changes

Based on the baseline images, in-plane rotational changes can be simulated using MATLAB's built-in `imrotate`, along with the `bicubic` interpolation option in order to reduce aliasing caused by the rotation of images. Since MATLAB's built-in rotation function is efficient, it can provide a faster and an accurate means to generate in-plane rotated range images. As a pilot investigation, each baseline (face-on) range image is rotated at 10° intervals (clockwise) up to 350° of rotation. Figure 6.1 illustrates a selection of the baseline range image of a mannequin head rotated to different angles.



(a) 0° (b) 40° (c) 120° (d) 200° (e) 320°

Figure 6.1: A selection of in-plane rotated range images of a mannequin head, captured at (a) 0° (baseline), and then rotated synthetically using MATLAB at (b) 40°, (c) 120° (d) 200° and (e) 320° in the clockwise direction.

6.2.1.2 Out-of-Plane Rotational Changes

The out-of-plane test data set comprise range images captured at baseline (i.e. face-on at 0°) and then at approximately 10° intervals about the yaw axis in both clockwise and anticlockwise directions (i.e. the face is rotated left/right horizontally) up to and including 90°.

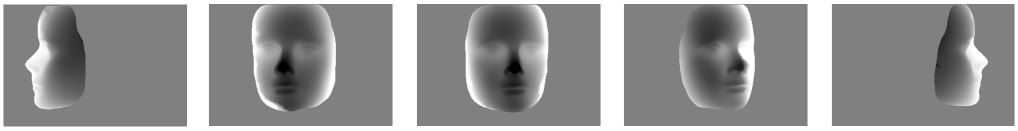
Here, in order to capture the different rotational images, each subject (i.e. a human and a mannequin head respectively) was located on a seat placed over a reticule marked in polar degrees of 10° intervals, allowing the seat to be rotated to ±90°. The seat is rotated accordingly and the images are taken at different angles, producing a set of stereo-pair images of Euler's out-of-plane rotations about the yaw axis. While this is a comparatively crude method for estimating the rotation about the yaw axis, it is adequate as a proof-of-concept at this stage of the work.

Figure 6.2 shows a selection of out-of-plane rotated range images of a mannequin head, captured using the methodology named above. Figure 6.3 illustrates the same images with a Gaussian mask, normalised to [0,1], placed over the range images in order to segment the area of interest.

(a) -90° (b) -40° (c) 0° (d) 40° (e) 90°

Figure 6.2: A selection of out-of-plane rotated range images of a mannequin head, captured at (a) 90° clockwise, (b) 40° clockwise, (c) 0° (baseline), (d) 40° anticlockwise and (e) 90° anticlockwise.

The next section outlines the overall approach taken for the validation.



(a) -90° (b) -40° (c) 0° (d) 40° (e) 90°

Figure 6.3: A selection of out-of-plane rotated range images of the mannequin head with masking, captured at (a) 90° clockwise (b) 40° clockwise, (c) 0° (baseline), (d) 40° anticlockwise and (e) 90° anticlockwise.

6.2.2 Overall Approach

This section outlines the overall approach taken to validate the 2.5D SIFT system against rotational changes. Given a set of (pre-processed) range images captured at different angles, the (x, y, σ, θ) information can be extracted using a scale-space keypoint localisation approach, followed by the extraction of the feature descriptor for each keypoint. Thus, the $(x, y, \sigma, \theta, \text{Descriptor})$ can be recorded in a local database for each range image. The steps for feature extraction and matching are as follows:

1. For each range image, detect the stable keypoints (i.e. their real-value (x, y) location) and their appropriate scale σ using the methodology outlined in Section 5.4.

2. Canonical orientation(s) can be determined for each keypoint using its appropriate σ based on the methodology shown in Section 5.2 with sub-pixel accuracy. The σ determines the size of the overall circular Gaussian measurement aperture (size = 4σ).

3. Extract the feature descriptors from each of the keypoint location using the appropriate σ, and its canonical orientation(s) θ, using the method shown in Section 5.3.

4. Create a database with n entries, where n is the number of range images employed for the test data set, one for each range image for each rotational change. The database can be created within MATLAB using the `cell array` data structure, allowing multiple information to be stored simultaneously, where each level created contains one range image corresponds to each viewpoint captured.

5. Store the relevant information, each of the corresponding $(x, y, \sigma, \theta, \text{Descriptor})$, into the database for each range image, along with the corresponding normalised image.

6. The discriminability of the feature descriptors can be calculated by the matching algorithm (using log-likelihood), given in Section 5.5, between the feature descriptors extracted from the baseline and the test range images in the Hough Transform space. Initial **putative** matches between the images are detected using the nearest-neighbour algorithm, and these potential matches are confirmed as either **positive matches** or **outliers** within the HT. While the HT itself is not infallible in recovering outliers, its use allows a more accurate calculation of the discriminability. The positive matches cannot be guaranteed to be all correct. However, from visual inspection of the outliers, the HT appears to be effective.

7. The final results contain the number of keypoints in the tested range image, the number of potential matches between the tested range image and the baseline image, the number of positive matches and the number of outliers. This can then be used to determine the percentage of matched keypoints (Equation 6.2.1), which are then presented in a match-matrix used to represent all combinations of the matching of the images. See Section 6.2.3.1 for details.

$$\%\text{ of positive (filtered) matches} = \frac{\#\text{ of positive matches}}{\#\text{ of potential matches}} \times 100 \qquad (6.2.1)$$

8. The results can also be presented in a Receiver Operating Characteristics (ROC) space, providing a visual interpretation of the rate of classifications. Details of ROC can be found in Section 6.2.3.2

6.2.3 Presentation of Results

Each percentage in the match-matrix illustrates the percentage of the positive and filtered matches between two range images, indicating the invariance of the feature descriptors between these two range images under rotation. The matching results obtained for every possible combination between the rotated range images can be presented mathematically as well as visually. This section details two of the methodologies used in this research work.

6.2.3.1 Match-Matrix

A *match-matrix* is a summary of results typically used in an unsupervised learning environment and is similar to the *confusion matrix* used in Section 3.6, where it was used in a supervised learning environment. The match-matrix used in this work comprises the percentages of positive matches obtained for matches between range images depicting all combinations of viewpoint (i.e. it covers all possible rotational viewpoints used in the data set). The main (or leading) diagonal entries correspond to the percentage of matched keypoints between self-matching images (i.e. an range image matching with itself).

Table 6.1 illustrates a toy-example of a match-matrix obtained for all the possible combinations of the out-of-plane rotational results from -30° to 30°. Each row corresponds to an image stored in the database and each column corresponds to the tested image. For example, entry $M_{2,3}$ of this match-matrix corresponds to the percentage of positive matches between range images captured at -20° and -10°, which is 95% in this particular toy-example.

		Test Image						
		-30°	**-20°**	**-10°**	**0°**	**10°**	**20°**	**30°**
Reference Image	**-30°**	100	90	75	0	0	0	0
	-20°	85	100	95	80	0	0	0
	-10°	54	75	100	85	62	0	0
	0°	0	32	65	100	82	68	0
	10°	0	0	27	67	100	74	52
	20°	0	0	0	43	76	100	87
	30°	0	0	0	0	23	46	100

Table 6.1: A toy-example of a match-matrix for all the possible combinations of the out-of-plane rotational results.

The match-matrix can also be represented visually using bar charts. Each height of the columns represent the percentages of matched keypoints, as a result, the higher the column, the better the results. Figure 6.4 shows the graph obtained from the toy-example match-matrix in Table 6.1.

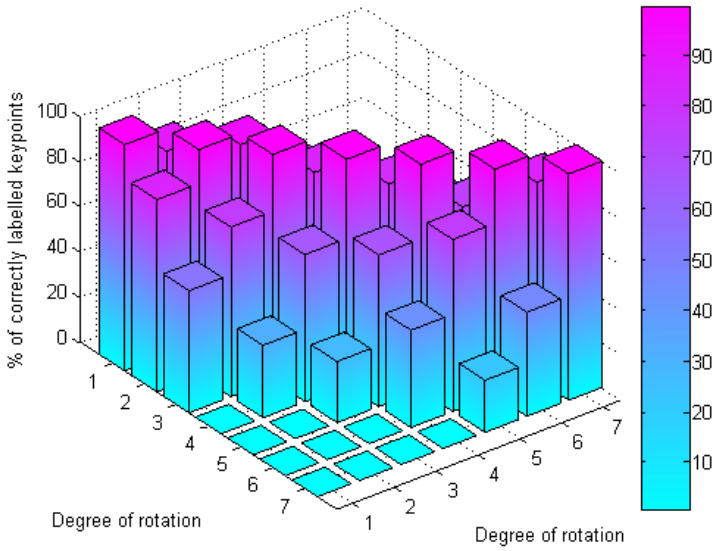


Figure 6.4: Match-matrix results represented in a graph obtained from the toy-example in Table 6.1.

6.2.3.2 Receiver Operating Characteristics (ROC)

Receiver Operating Characteristics (ROC) is a graphical plot of the *true positive rate* (TPR) versus the *false positive rate* (FPR). The TPR is also known as the *sensitivity* whereas the FPR is also known as *1-specificity*. TPR determines a classifier or a diagnostic test performance on classifying positive instances correctly among all positive samples available during the test. FPR, on the other hand, defines how many incorrect positive results occur among all negative samples available during the test. As a result, ROC graphs are useful for comparing classifiers as well as providing a meaningful visualisation tool. ROC are commonly used in medical decision making (Altman and Bland, 1994; Bewick et al., 2004; Egan, 1975; Park et al., 2004; Zou, 2002), but in recent years has been used increasingly in machine learning and data mining research (Fawcett, 2006; Spackman, 1989). ROC is also similar to the *Precision-Recall* (PR) graph in Information Retrieval (IR) (Belew, 2000; van Rijsbergen, 1979). The subtle difference between ROC and PR is that ROC curves are commonly used to present results for binary decision problems in machine learning whereas PR curves are normally employed to deal with highly skewed datasets in order to obtain a more informative picture of the

algorithm's performance (Davis and Goadrich, 2006). As a result, it is more suitable to use ROC as opposed to the PR curve in this work, since the results will be a binary decision (i.e. either it is a match or it is not a match) . This section describes how to derive an approximate ROC from the 2.5D SIFT matching algorithm.

The TPR and FPR, along with the precision and recall values, can be determined from a confusion matrix containing information about the true positive (TP), false positive (FP), false negative (FP) and true negative (TN) rates, illustrated in Table 6.2. The TPR and FPR can then be determined as shown in Equation 6.2.2.

	Correct results/classification	
Obtained results/ classification	True Positive (TP)	False Positive (FP)
	False Negative (FN)	True Negative (TN)

Table 6.2: Confusion matrix for the four possible outcomes: true positive (TP), false positive (FP), false negative (FN) and false positive (FP). This can then be used to calculate the TPR and FPR, as well as the precision/recall value.

$$\begin{aligned} TPR &= \frac{TP}{TP+FN} = \frac{TP}{\text{Total Positives}} \\ FPR &= \frac{FP}{FP+TN} = \frac{FP}{\text{Total Negatives}} \end{aligned} \tag{6.2.2}$$

Similarly, the precision and recall values can be determined from the confusion matrix table, as shown in Equation 6.2.3. Note that recall is the same as sensitivity, which is equivalent to the TPR.

$$\begin{aligned} Recall &= \frac{TP}{TP+FN} \\ Precision &= \frac{TP}{TP+FP} \end{aligned} \tag{6.2.3}$$

A *ROC space* is defined by the FPR and TPR as the x and y axes respectively (i.e. plotting TPR against FPR), which depicts relative trade-offs between the true positive and the false positive. Since TPR is equivalent to sensitivity and FPR is equal to 1-specificity, the ROC graph is sometimes called the sensitivity vs 1-specificity plot. Each prediction result or one instance of a confusion matrix represents one point in the ROC space.

Figure 6.5 illustrates the ROC space with two toy example prediction points (blue points). The diagonal line (red line), known as the *line of no discrimination*, divides the ROC space in areas of good or bad classification, where points above the diagonal line indicate good classification results (i.e. better than random guessing), while points below the line indicate wrong results. The best possible prediction method would yield a point in the upper left corner or coordinate $(0, 1)$ of the ROC space, representing a 100% sensitivity (all true positives are found) and 100% specificity (no false positives are found), known as a *perfect classification*. A completely random guess would give a point along the line of no-discrimination from the left bottom to the top right corners. In other words, the closer the points are to the y-axis (i.e. smaller FPR) and the closer the points are to the top left hand corner (where TPR=1), the better and more reliable the classifications are.

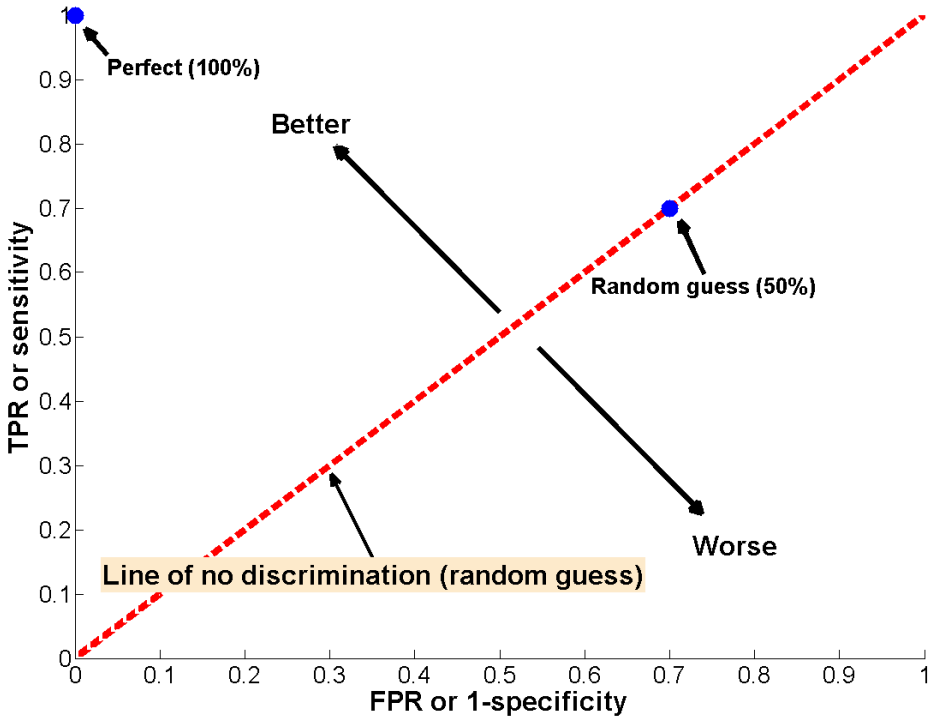


Figure 6.5: A graph demonstrating the ROC space. The diagonal (red) line shows the line of no discrimination (i.e. random guess) and any points located above the line is considered to be a good classification whereas any points below the line is consider a bad classification.

In this work, for each pairing of the test range images, the probability of the TP, FP, TN and FN can be obtained as follows:

$$\begin{aligned} TP &= \frac{\text{\# of positive matches}}{\text{total \# of keypoints detected}} \\ FP &= \frac{\text{\# of false matches}}{\text{total \# of keypoints detected}} \\ TN &= \frac{\text{total \# of keypoints detected} - \text{\# of potential matches}}{\text{total \# of keypoints detected}} \\ FN &= 1 - TP - FP - TN \end{aligned} \tag{6.2.4}$$

It is not necessary to convert the TP, FP, TN and FN into probabilities in ROC. However, the number of keypoints detected for each image is unique in this work, it is, therefore, better to convert the values of TP, FP, TN and FN into probabilities for a consistent representation, as shown above.

The TPR and FPR for each possible image pairing can be calculated using Equation 6.2.2 and then plotted within a ROC space, thereby providing a visual means of interpreting the classification results.

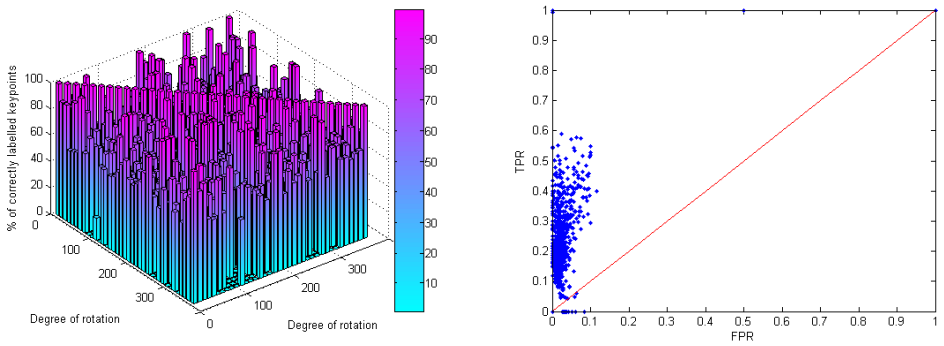
6.3 Validation Results

This section presents the matching results obtained for both in-plane and out-of-plane images based on the methodology outline in the previous section. The purpose of this validation of the 2.5D SIFT system is twofold: firstly to test the invariance properties on Euler's in-plane rotational changes and secondly to test its invariance to the Euler's out-of-plane rotations. Results are presented in both the match-matrix forms and in the ROC space.

6.3.1 In-Plane Rotations

There are two sets of in-plane rotation data: range images of a human face and range images of a mannequin head. The baseline (no rotation) image was rotated synthetically using MATLAB in clockwise direction increments of 10° up to 350° (i.e. a full circle). Keypoints and feature descriptors were then extracted from each range image and were then stored in a database for matching. Results are shown below.

Table 6.3 presents the match-matrix obtained by exploring all the combinations of the feature descriptors extracted from the synthetically in-plane rotated range images from 0° at increments of 10° in a clockwise direction up to 350°. This match-matrix is presented visually in Figure 6.6(a). There are some instances in the matrix where 0% of matched keypoints are returned. This is caused by the insufficient putative matches between the images. Based on the obtained results, the feature descriptors showed promising self-matching properties. Moreover, close examination of the results reveals over 68.7% of the test images are matched at over a 50% keypoint matching rate, while over 46.8% of the test images are matched at over 60% keypoint matching rate. The results are plotted in a ROC space as shown in Figure 6.6(b), showing the majority of the matches are above the line of no discrimination.



(a) Match-matrix (b) ROC

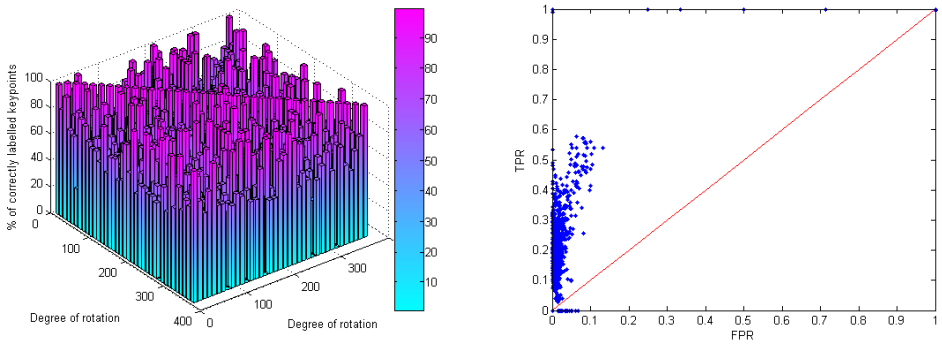
Figure 6.6: (a) Graph illustrating the match-matrix results of the percentage of matched keypoints, produced from the in-plane rotational data of a **mannequin head** (from 0° at 10° clockwise increments up to 350°). (b) Results plotted in ROC space.

	0°	10°	20°	30°	40°	50°	60°	70°	80°	90°	100°	110°	120°	130°	140°	150°	160°	170°
0°	100	83.78	88.24	60	58.33	0	0	0	0	0	50	0	0	0	62.5	0	0	100
10°	87.5	100	85.71	93.75	66.67	100	0	0	57.14	75	0	0	0	0	0	75	62.5	63.64
20°	87.5	84.38	100	72.73	55.56	57.14	0	0	60	66.67	25	0	0	75	0	75	0	0
30°	54.55	89.47	73.08	100	73.33	66.67	66.67	75	55.56	57.14	0	57.14	60	66.67	71.43	75	66.67	63.64
40°	55.56	40	66.67	68.42	100	63.64	70	87.5	0	0	66.67	0	0	0	66.67	66.67	60	77.78
50°	57.14	57.14	84.62	57.14	67.65	100	80.77	66.67	54.55	57.14	57.14	60	75	55.56	83.33	62.5	60	66.67
60°	100	60	53.33	55.56	57.14	72.73	100	72.22	75	60	60	25	0	66.67	57.14	71.43	0	71.43
70°	60	70	62.5	60	60	84.21	68.97	100	71.43	62.5	62.5	66.67	0	57.14	100	100	0	61.54
80°	0	0	0	0	66.67	57.14	66.67	76.67	100	78.26	57.14	66.67	0	0	75	57.14	0	66.67
90°	0	0	0	0	60	70	66.67	53.33	65.52	100	78.95	62.5	57.14	75	62.5	0	0	66.67
100°	16.67	53.85	62.5	60	80	60	60	63.64	73.68	82.61	100	78.57	60	0	70	70	75	56.25
110°	0	0	0	0	0	75	55.56	87.5	57.14	63.64	81.82	100	92.86	72.73	66.67	66.67	0	66.67
120°	57.14	0	75	60	55.56	66.67	60	71.43	83.33	0	76.92	79.17	100	69.57	71.43	60	66.67	60
130°	0	70	75	66.67	66.67	0	66.67	60	62.5	57.14	0	60	65	100	88.46	69.23	53.85	75
140°	55.56	55.56	33.33	60	58.33	72.73	72.73	62.5	77.78	0	0	66.67	57.14	65.38	100	74.36	71.43	66.67
150°	57.14	85.71	100	100	62.5	60	62.5	0	0	60	60	100	60	58.33	74.19	100	81.25	73.33
160°	62.5	62.5	60	75	55.56	55.56	57.14	66.67	75	75	75	0	0	70	75	83.33	100	78.05
170°	70	66.67	60	66.67	57.14	0	57.14	62.5	57.14	57.14	0	0	0	53.85	62.5	87.5	73.68	100
180°	75	20	57.14	62.5	75	0	57.14	66.67	75	0	0	60	0	62.5	0	80	85.71	63.16
190°	62.5	60	55.56	57.14	66.67	71.43	66.67	75	84.62	63.64	0	0	57.14	70	70	91.67	70	82.14
200°	60	57.14	55.56	57.14	0	55.56	100	0	60	60	100	0	0	70	83.33	57.14	55.56	100
210°	62.5	0	55.56	66.67	66.67	0	0	0	0	0	57.14	0	66.67	66.67	66.67	60	54.55	63.64
220°	66.67	66.67	66.67	60	62.5	100	0	60	0	0	0	75	60	0	0	75	62.5	63.64
230°	75	60	57.14	62.5	63.64	66.67	55.56	66.67	50	0	0	0	0	100	60	50	0	55.56
240°	88.89	57.14	60	60	70	66.67	75	0	0	66.67	0	100	0	100	62.5	66.67	0	66.67
250°	0	0	100	100	100	71.43	57.14	0	60	0	60	100	100	100	55.56	66.67	57.14	0
260°	0	0	60	70	83.33	60	50	87.5	60	0	0	0	0	66.67	0	0	0	0
270°	55.56	66.67	0	0	0	62.5	100	75	0	75	0	75	0	0	0	0	60	90
280°	71.43	60	0	60	75	60	100	62.5	60	0	0	75	0	0	0	0	0	20
290°	0	80	75	0	0	0	62.5	0	0	0	0	66.67	66.67	66.67	0	0	0	100
300°	0	62.5	0	0	0	0	0	0	0	0	0	57.14	0	57.14	0	0	60	62.5
310°	60	0	66.67	0	0	0	0	75	0	0	0	0	60	0	57.14	54.55	100	55.56
320°	0	75	20	71.43	60	66.67	0	75	0	0	0	0	0	62.5	57.14	60	55.56	66.67
330°	55.56	0	75	66.67	55.56	57.14	0	0	0	0	83.33	0	0	66.67	0	40	71.43	90.91
340°	73.91	80	66.67	66.67	0	55.56	66.67	20	0	0	0	0	0	66.67	60	62.5	88.24	70
350°	78.79	72	100	66.67	75	57.14	83.33	57.14	0	0	0	0	0	0	64.71	66.67	54.55	66.67

	180°	190°	200°	210°	220°	230°	240°	250°	260°	270°	280°	290°	300°	310°	320°	330°	340°	350°
0°	90.91	57.14	72.73	62.5	66.67	75	0	0	0	100	57.14	0	0	66.67	66.67	63.64	66.67	70.97
10°	80	75	57.14	57.14	57.14	60	0	66.67	57.14	0	0	0	0	100	0	0	75	63.16
20°	100	62.5	93.33	60	60	0	0	0	62.5	57.14	0	83.33	0	66.67	0	66.67	0	88.89
30°	66.67	75	57.14	62.5	0	40	0	75	0	0	66.67	0	0	0	75	66.67	62.5	53.85
40°	100	75	40	61.54	66.67	70	0	0	66.67	0	60	0	0	60	66.67	70	75	85.71
50°	60	0	40	55.56	91.67	75	62.5	62.5	75	0	0	80	20	66.67	75	66.67	60	62.5
60°	84.62	66.67	0	66.67	100	75	66.67	75	75	0	0	100	60	55.56	60	0	0	66.67
70°	77.78	72.73	71.43	83.33	62.5	100	57.14	75	62.5	0	0	75	0	66.67	75	62.5	57.14	63.64
80°	66.67	72.73	66.67	0	66.67	0	66.67	0	0	71.43	83.33	0	40	85.71	66.67	75	0	25
90°	60	62.5	60	100	57.14	60	0	0	0	0	100	50	0	66.67	0	0	75	60
100°	0	75	66.67	0	62.5	71.43	60	60	0	0	0	0	80	60	71.43	100	55.56	54.55
110°	0	0	0	0	75	0	0	57.14	0	60	0	0	0	83.33	60	0	0	66.67
120°	50	66.67	66.67	55.56	0	60	75	60	66.67	0	100	0	57.14	58.33	75	57.14	0	57.14
130°	66.67	100	71.43	60	66.67	0	0	66.67	57.14	57.14	85.71	0	66.67	57.14	66.67	57.14	57.14	0
140°	60	57.14	75	75	0	100	75	75	60	66.67	60	57.14	0	70	61.54	63.64	62.5	64.29
150°	70	0	0	0	66.67	60	0	0	0	0	0	66.67	50	66.67	60	75	66.67	75
160°	85	66.67	100	0	60	60	66.67	0	62.5	0	0	0	0	62.5	72.73	54.55	60	84.62
170°	72.22	68.42	55.56	57.14	0	60	62.5	0	57.14	60	57.14	0	0	66.67	100	77.78	60	57.14
180°	100	76.92	80.77	72.73	66.67	60	0	0	0	60	60	0	0	83.33	62.5	0	60	100
190°	80	100	83.33	100	57.14	66.67	0	75	0	0	0	0	0	66.67	60	57.14	66.67	100
200°	80.95	80.56	100	81.08	63.64	75	0	0	57.14	20	0	75	0	60	77.78	62.5	66.67	58.33
210°	57.14	88.24	88.57	100	68.97	100	62.5	57.14	57.14	0	57.14	53.33	0	57.14	0	75	0	57.14
220°	60	60	88.89	67.74	100	72	100	60	60	0	0	0	0	0	0	0	0	55.56
230°	71.43	60	92.86	52.94	61.54	100	73.08	71.43	55.56	100	0	0	0	0	66.67	62.5	55.56	75
240°	90	60	62.5	25	80	70.83	100	72.73	75	71.43	0	0	20	0	0	75	0	60
250°	0	75	60	100	57.14	60	67.86	100	78.26	63.64	71.43	0	71.43	88.89	60	0	14.29	0
260°	100	0	75	62.5	57.14	58.33	76.92	76	100	62.96	80	60	75	75	0	0	66.67	40
270°	83.33	0	0	0	28.57	66.67	55.56	69.23	75	100	83.33	83.33	66.67	60	0	0	60	66.67
280°	63.64	57.14	40	66.67	60	66.67	0	70	70	82.61	99.38	88	63.64	60	0	60	62.5	57.14
290°	80	57.14	100	60	62.5	90	62.5	100	60	87.5	69.7	100	77.27	22.22	60	60	62.5	66.67
300°	83.33	80	0	57.14	0	57.14	71.43	0	75	57.14	76.92	77.78	100	76	70	57.14	33.33	83.33
310°	55.56	66.67	55.56	70	66.67	66.67	66.67	54.55	62.5	66.67	60	70	84.62	100	57.14	80	66.67	54.55
320°	60	0	71.43	60	20	60	66.67	0	0	0	0	57.14	66.67	73.08	100	94.74	69.23	60
330°	66.67	57.14	57.14	55.56	10	54.55	0	0	0	0	0	66.67	57.14	61.54	78.79	100	80	75
340°	66.67	100	60	66.67	0	66.67	60	62.5	0	100	0	71.43	75	75	78.57	73.08	100	87.88
350°	66.67	66.67	14.29	75	80	57.14	57.14	57.14	0	60	0	0	75	75	57.14	72.73	84.85	100

Table 6.3: Match-matrix results of the percentage of the matched and filtered keypoints by HT, produced from the in-plane rotational data of a **mannequin head** (from 0° at 10° clockwise increments up to 350°).

Table 6.4 presents the match-matrix obtained by exploring all the combinations of the feature descriptors extracted from the in-plane rotated range images of a human face from 0° at 10° increments in the clockwise directions up to 350°. The examination of this matrix illustrates over 79.9% of the images are matched at over 50% keypoint matched rate while over 55% of the test images are matched at 60% keypoint matching rate. Once again, the feature descriptors show strong self-matching properties. Note that there are fewer zero entries in this match-matrix, compared to the match-matrix obtained from the mannequin head data (Table 6.3). This is because there are more depth features within a human face whereas the depth features in a mannequin head is much smoother. Figure 6.7(a) shows this match matrix in a graph and Figure 6.7(b) illustrates the results in ROC space, demonstrating the majority of the matches are above the line of no discrimination.



(a) Match-matrix (b) ROC

Figure 6.7: (a) Graph showing the match-matrix results of the percentage of matched keypoints, produced from a set of in-place rotational data of a **human face** (from 0° at 10° increments in the clockwise direction up to 350°). (b) Results plotted in ROC space.

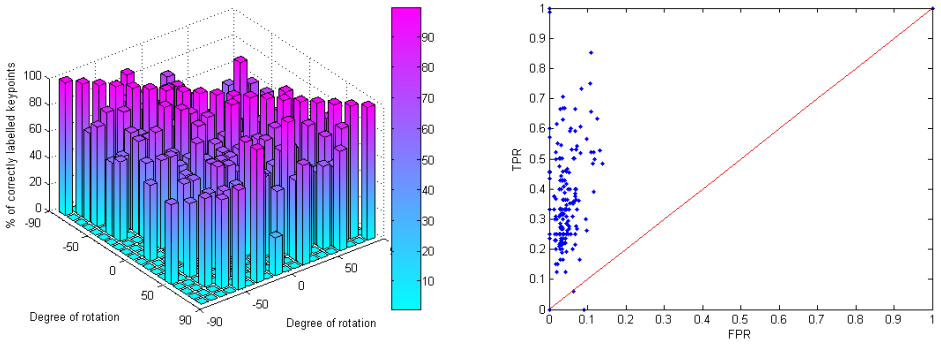
	0°	10°	20°	30°	40°	50°	60°	70°	80°	90°	100°	110°	120°	130°	140°	150°	160°	170°
0°	98.15	75.44	65	90	100	83.33	66.67	0	62.5	0	66.67	0	60	60	81.82	0	55.56	55.56
10°	81.03	100	71.7	71.43	75	57.14	0	50	80	37.5	62.5	0	0	66.67	54.55	85.71	66.67	63.64
20°	92.59	66.67	100	63.64	92.86	75	54.55	0	0	0	60	60	66.67	57.14	100	60	63.64	58.33
30°	69.23	78.95	79.59	100	69.64	76.19	86.67	0	0	0	0	0	62.5	0	66.67	75	75	70.59
40°	66.67	83.33	77.78	77.55	99.63	66.67	65	55.56	57.14	60	60	60	100	75	66.67	0	0	0
50°	28.57	58.33	100	69.23	66.67	100	71.43	80	66.67	0	25	0	0	66.67	57.14	83.33	63.64	87.5
60°	83.33	0	57.14	86.67	70.83	68.18	100	80	58.33	66.67	0	0	100	0	60	16.67	0	75
70°	80	55.56	66.67	14.29	90.91	80	77.78	100	83.78	53.33	57.14	57.14	0	0	20	57.14	75	100
80°	70	66.67	62.5	60	83.33	66.67	81.82	78	100	65.85	61.11	54.55	0	66.67	80	63.64	60	0
90°	66.67	72.73	71.43	61.54	70	42.86	100	70	78.79	100	88.24	94.12	71.43	71.43	60	54.55	64.71	53.85
100°	66.67	66.67	53.85	0	62.5	75	55.56	75	95	86.96	100	79.63	63.64	66.67	66.67	85.71	60	60
110°	60	57.14	55.56	69.23	55.56	66.67	0	57.14	77.78	94.44	72.34	99.62	80	69.23	60	66.67	75	60
120°	88.89	57.14	57.14	25	60	66.67	0	0	0	75	75	87.1	100	78.85	75	61.54	63.64	57.14
130°	55.56	0	66.67	60	69.23	100	62.5	0	0	60	0	78.57	85.37	100	78.72	61.9	66.67	81.82
140°	69.23	62.5	66.67	75	75	60	100	57.14	0	0	0	60	73.68	73.47	100	70.91	66.67	55.56
150°	70	61.9	78.57	60	78.57	73.33	66.67	77.78	70	88.24	0	0	75	59.09	76.36	100	70.97	60.71
160°	55.56	75	54.55	80	53.33	73.33	55.56	0	0	87.5	0	0	0	57.14	70.59	72.73	100	79.66
170°	57.14	75	55.56	54.55	20	55.56	0	66.67	0	0	75	0	0	57.14	71.43	53.33	82.69	100
180°	86.36	75	66.67	57.14	62.5	77.78	80	0	66.67	0	0	0	0	57.14	0	66.67	78.95	75.76
190°	66.67	58.82	55.56	0	68.75	70	62.5	66.67	0	0	0	66.67	66.67	0	100	90.91	75	78.38
200°	100	58.33	75	52.63	58.33	70	90	57.14	80	0	75	0	62.5	0	0	66.67	60	57.14
210°	54.55	68.75	64.29	87.5	66.67	57.14	90	55.56	57.14	90	0	0	60	75	60	0	62.5	75
220°	62.5	100	87.5	62.5	100	66.67	66.67	55.56	62.5	62.5	0	0	0	60	0	85.71	100	75
230°	62.5	53.85	63.64	90	52	73.08	75	60	66.67	55.56	71.43	0	0	57.14	66.67	66.67	58.33	70
240°	60	0	81.82	53.85	54.55	93.33	60	0	70	0	100	0	80	60	0	0	0	75
250°	60	66.67	100	60	56	57.89	60	100	55.56	75	77.78	0	100	90	60	57.14	20	75
260°	57.14	60	60	0	60	0	25	75	60	0	0	0	0	66.67	55.56	0	55.56	72.73
270°	0	60	75	75	55.56	75	55.56	0	57.14	71.43	0	60	100	80	100	83.33	66.67	60
280°	72.73	57.14	62.5	62.5	40	75	60	63.64	66.67	88.89	0	80	100	60	70	60	60	55.56
290°	55.56	55.56	80	60	57.14	60	0	57.14	71.43	66.67	57.14	75	100	0	0	62.5	75	55.56
300°	100	70	60	66.67	66.67	60	80	0	0	57.14	77.78	88.89	100	75	0	57.14	55.56	66.67
310°	62.5	100	0	62.5	0	66.67	0	0	75	66.67	66.67	0	0	72.73	100	100	62.5	64.29
320°	80	58.33	58.33	60	57.14	0	0	60	66.67	0	0	0	83.33	91.67	80	50	55.56	81.82
330°	64.29	66.67	80	85.71	66.67	55.56	0	60	60	57.14	75	75	100	57.14	80	81.82	62.5	86.67
340°	73.68	52.94	56.25	57.14	66.67	75	60	60	66.67	0	50	57.14	62.5	100	88.89	54.55	71.43	100
350°	70.97	69.7	66.67	70	54.55	54.55	0	57.14	62.5	100	0	0	66.67	0	66.67	57.14	100	61.54

	180°	190°	200°	210°	220°	230°	240°	250°	260°	270°	280°	290°	300°	310°	320°	330°	340°	350°
0°	82.35	85.71	100	93.75	57.14	66.67	60	66.67	57.14	66.67	75	60	0	75	60	54.55	68.18	77.78
10°	92.31	53.85	88.89	53.85	57.14	57.14	85.71	16.67	85.71	100	100	0	57.14	60	75	66.67	66.67	77.27
20°	57.14	73.33	66.67	80.95	78.95	80	90	0	0	0	66.67	33.33	0	54.55	64.71	60	75	71.43
30°	55.56	61.11	60	72	100	75	0	62.5	0	0	0	0	66.67	75	55.56	62.5	80	60
40°	100	83.33	85.71	100	58.33	100	60	66.67	57.14	62.5	0	0	62.5	62.5	0	66.67	0	75
50°	66.67	55.56	78.57	61.54	63.16	78.95	93.75	72.73	60	66.67	0	0	0	62.5	62.5	75	55.56	100
60°	90	54.55	53.85	100	83.33	55.56	100	62.5	53.85	60	66.67	75	71.43	0	0	0	0	53.85
70°	55.56	57.14	80	100	62.5	64.29	72.73	60	75	0	0	60	0	57.14	70	0	55.56	57.14
80°	44.44	66.67	61.54	100	80	0	54.55	100	57.14	66.67	83.33	60	75	75	66.67	66.67	60	0
90°	58.33	75	60	62.5	28.57	0	75	75	75	73.33	57.14	66.67	66.67	66.67	55.56	57.14	83.33	54.55
100°	0	0	0	28.57	60	66.67	66.67	75	0	0	50	88.89	87.5	100	75	62.5	50	0
110°	60	60	62.5	66.67	55.56	66.67	60	55.56	0	100	0	75	87.5	60	62.5	60	70	100
120°	66.67	0	0	75	60	0	66.67	60	0	0	57.14	60	62.5	66.67	0	60	88.89	60
130°	52.94	58.33	66.67	66.67	0	60	55.56	75	66.67	40	0	55.56	57.14	100	69.23	83.33	100	66.67
140°	61.54	0	57.14	63.64	54.55	60	57.14	0	0	0	0	57.14	57.14	57.14	100	57.14	81.82	50
150°	57.89	63.64	53.33	75	56.25	28.57	60	16.67	60	83.33	60	60	63.64	60	90.91	87.5	54.55	66.67
160°	70.97	81.25	77.78	26.67	0	75	57.14	0	0	0	0	0	57.14	83.33	0	70	63.64	87.5
170°	77.78	66.67	57.14	100	69.23	75	75	66.67	0	0	0	83.33	0	60	75	50	0	57.14
180°	100	77.03	80.95	100	66.67	0	57.14	62.5	0	0	62.5	62.5	0	0	66.67	80	62.5	89.47
190°	78.87	100	72.92	95.45	69.23	66.67	66.67	0	0	0	60	60	62.5	57.14	70	55.56	81.82	71.43
200°	71.43	75.47	100	71.43	63.89	60	55.56	0	55.56	0	66.67	50	100	57.14	57.14	0	62.5	53.85
210°	71.43	60.87	82.22	100	72	70	53.85	60	80	71.43	66.67	0	0	54.55	75	0	66.67	60
220°	100	92.31	82.35	74.6	99.65	79.03	65.22	57.14	0	0	66.67	62.5	0	66.67	75	55.56	60	57.14
230°	66.67	75	81.82	64.29	75.38	100	73.17	70	0	57.14	60	66.67	100	0	57.14	60	54.55	57.14
240°	100	81.82	100	66.67	75	71.74	99.63	78.72	54.55	66.67	0	0	100	0	60	100	0	75
250°	60	60	66.67	88.89	75	72	72.22	100	85	78.57	55.56	25	0	62.5	60	57.14	55.56	62.5
260°	0	66.67	0	86.67	83.33	71.43	72.73	84.21	100	73.33	80.95	77.78	0	88.89	62.5	57.14	100	66.67
270°	75	57.14	0	0	0	57.14	57.14	76.92	80.65	100	81.25	81.25	55.56	75	66.67	0	0	70
280°	62.5	60	20	40	75	57.14	60	80	72.73	92.59	99.6	74.42	64.71	66.67	75	75	75	54.55
290°	66.67	75	66.67	60	0	60	0	0	62.5	92.31	81.08	100	78.43	60	60	60	0	66.67
300°	55.56	0	77.78	60	55.56	66.67	0	75	0	75	72.73	84.85	100	78.95	75	55.56	70	66.67
310°	58.33	66.67	60	70	0	57.14	75	0	75	50	75	75	84.44	100	77.78	83.33	55.56	66.67
320°	87.5	85.71	75	0	0	77.78	0	0	75	0	60	75	70	79.66	100	73.21	72.73	70
330°	53.85	83.33	60	80	83.33	100	57.14	66.67	66.67	60	75	57.14	63.64	66.67	75	100	70.91	70.59
340°	55.56	100	60	66.67	53.85	60	83.33	0	60	85.71	100	57.14	75	66.67	78.95	74.14	100	68.85
350°	55.56	66.67	80	62.5	55.56	55.56	60	0	40	71.43	40	57.14	0	75	90	85	80	100

Table 6.4: Match-matrix results of the percentage of matched and filtered keypoints, produced from the in-plane rotational data of a **human head** (from 0° at 10° clockwise increments up to 350°).

6.3.2 Out-of-Plane Rotations

Table 6.5 illustrates the match-matrix obtained by exploring all the combinations of the feature descriptors extracted from the out-of-plane rotated range images of a mannequin head from -90° to 90°. This match-matrix is presented in Figure 6.8(a). Closer examination of the matrix shows that not only do the feature descriptors have perfect self-matching properties (100% recognition rate), over 46.8% of the test images are matched at over 50% of the keypoints matching rate while over 35.2% of the test images are matched at over 60% of the keypoints matching rate. Figure 6.8(b) shows the results in ROC space, illustrating that the majority of the results are above the line of no discrimination.



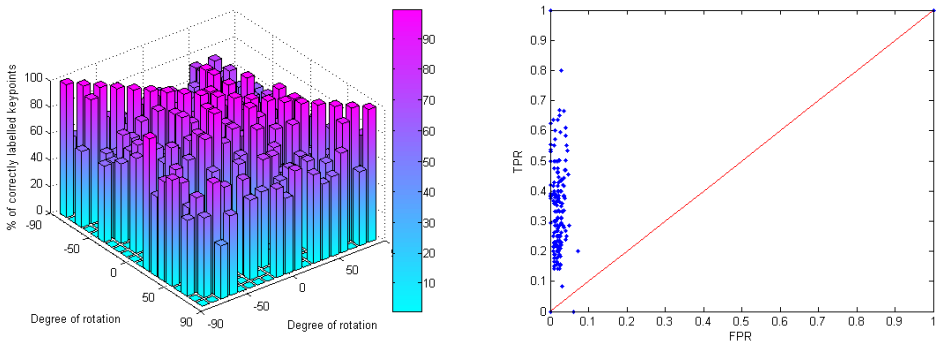
(a) Match-matrix (b) ROC

Figure 6.8: (a) Graph illustrating the match-matrix results of the percentage of the matched keypoints, produced from a set of out-of-plane rotational data of a **mannequin head** (from -90° to 90°). (b) Matching results plotted in a ROC space.

	-90°	-80°	-70°	-60°	-50°	-40°	30°	-20°	-10°	0°	10°	20°	30°	40°	50°	60°	70°	80°	90°
-90°	100	0	0	0	0	0	0	0	0	0	0	75	0	0	0	0	0	0	0
-80°	0	100	55.56	57.14	40	55.56	60	0	60	0	0	66.67	0	0	0	0	0	0	0
-70°	0	66.67	100	69.23	66.67	100	0	80	40	66.67	57.14	0	0	0	0	50	62.5	0	0
-60°	0	75	83.33	100	86.67	62.5	57.14	75	66.67	60	0	0	0	0	0	0	0	0	0
-50°	0	0	0	84.62	100	81.25	91.67	72.73	75	66.67	0	75	0	0	0	0	0	0	0
-40°	0	55.56	57.14	72.73	68.42	100	81.48	94.74	94.12	60	40	0	0	0	66.67	75	75	0	0
-30°	0	60	0	70	69.23	66.67	100	70	70	62.5	66.67	66.67	0	0	100	0	0	0	0
-20°	0	0	0	57.14	77.78	94.12	77.27	100	80.77	80	0	75	75	0	0	0	75	0	0
-10°	0	0	0	0	0	62.5	92.86	87.5	100	57.14	66.67	66.67	60	0	0	0	0	0	0
0°	0	0	0	66.67	66.67	57.14	60	81.82	85.71	100	76.47	75	60	0	0	0	0	0	0
10°	0	57.14	0	0	0	0	75	60	69.23	72.22	100	66.67	66.67	0	0	75	75	0	0
20°	0	0	0	0	60	0	0	0	66.67	100	66.67	100	79.17	66.67	57.14	0	0	0	0
30°	0	0	0	60	60	0	66.67	66.67	66.67	70	60	72.22	100	62.5	100	0	0	0	0
40°	0	0	0	0	0	0	66.67	66.67	60	60	57.14	75	81.82	100	66.67	0	66.67	0	0
50°	60	0	55.56	0	57.14	75	0	55.56	85.71	62.5	55.56	60	60	66.67	100	80	73.33	0	0
60°	0	0	0	0	0	0	0	0	0	60	0	75	75	0	85.71	100	85.71	75	0
70°	0	0	66.67	0	66.67	0	0	0	0	0	0	100	55.56	60	54.55	78.26	100	66.67	0
80°	0	0	0	66.67	0	0	0	75	0	57.14	0	60	66.67	0	62.5	60	85.71	100	0
90°	0	0	0	0	75	0	100	0	28.57	0	0	75	0	0	0	75	0	0	100

Table 6.5: Match-matrix results of the percentage of matched and filtered keypoints, produced from the out-of-plane rotational data of a **mannequin head**, from -90° to 90° of rotations.

Table 6.6 presents the match-matrix obtained by exploring all the combinations of the feature descriptors extracted from the out-of-plane rotated images of a human face from -90° to 90°. This match-matrix is also shown in Figure 6.9(a). This shows that over 51.8% of the test images are matched at over 50% of recognition rate, with over 36.3% of the images are matched at over 60%. Figure 6.9(b) illustrates the results in ROC space, showing the majority of the results are above the line of discrimination.



(a) Match-matrix (b) ROC

Figure 6.9: (a) Graph showing the match-matrix of the percentage of matched keypoints, produced from a set of out-of-plane range images of a **human face** from -90° to 90°. (b) Matching results plotted in ROC space.

	-90°	**-80°**	**-70°**	**-60°**	**-50°**	**-40°**	**30°**	**-20°**	**-10°**	**0°**	**10°**	**20°**	**30°**	**40°**	**50°**	**60°**	**70°**	**80°**	**90°**
-90°	100	58.33	60	0	0	0	54.55	0	57.14	0	66.67	0	0	57.14	75	0	75	0	0
-80°	0	100	80	57.14	0	66.67	57.14	66.67	66.67	0	0	0	0	66.67	75	0	66.67	57.14	0
-70°	57.14	93.33	100	75	70	0	0	0	0	0	0	0	0	60	0	0	75	60	0
-60°	0	0	0	100	66.67	60	0	0	0	60	62.5	71.43	0	75	0	0	0	0	60
-50°	0	60	57.14	55.56	100	86.67	80	0	87.5	0	60	0	0	0	60	0	0	0	0
-40°	57.14	57.14	75	70	80	100	94.12	66.67	60	0	0	100	0	60	0	0	66.67	0	0
-30°	0	60	66.67	0	62.5	88.89	100	85.71	66.67	71.43	75	100	0	0	0	0	0	0	0
-20°	0	0	75	0	66.67	60	87.5	100	80	91.67	0	60	60	57.14	0	0	75	60	0
-10°	0	71.43	55.56	0	60	66.67	88.89	81.82	100	100	83.33	100	60	100	0	0	66.67	0	0
0°	0	0	0	0	75	0	75	66.67	90	100	90.91	72.73	87.5	55.56	57.14	0	0	0	0
10°	0	0	0	0	0	66.67	83.33	100	60	83.33	100	88.89	90	60	66.67	75	40	50	0
20°	100	0	0	0	0	0	83.33	100	75	83.33	75	100	84.21	85.71	100	75	0	0	66.67
30°	60	0	75	0	0	0	62.5	0	20	66.67	100	90.48	100	80	78.57	0	0	0	62.5
40°	75	0	0	0	0	0	0	0	66.67	0	87.5	0	82.35	100	100	0	55.56	0	66.67
50°	80	75	0	0	0	0	0	0	0	0	60	75	55.56	86.67	100	60	0	0	0
60°	0	60	0	0	0	0	0	0	0	75	75	60	60	66.67	0	100	60	0	62.5
70°	57.14	71.43	0	0	71.43	0	71.43	0	57.14	11.11	60	66.67	0	66.67	0	60	100	66.67	0
80°	0	60	83.33	0	0	0	60	0	0	0	0	0	0	0	0	0	90.91	100	76.92
90°	0	0	40	60	0	0	0	0	66.67	57.14	75	0	66.67	57.14	60	75	0	55.56	100

Table 6.6: Match-matrix results of the percentage of matched and filtered keypoints, produced from the out-of-plane rotational data of a **human head**, from -90° to 90° of rotations.

6.4 Summary and Discussions

This chapter presents the results and findings from the validation of the 2.5D SIFT system, using both in-plane and out-of-plane range images of a human face and a mannequin head. The images were first captured at the baseline (face-on, no rotation) position and then at 10° intervals up to and including 90° of rotation in both clockwise and anticlockwise directions (i.e. fully lateral) to create a set of data to simulate out-of-plane images. In-plane rotational images were generated synthetically using MATLAB where each of the images is rotated clockwise at 10° interval up to 350° (i.e. a full circle). In order to suppress noise introduced in the synthetic rotation algorithm, the bicubic interpolation was used on each rotational image.

The positions of the keypoints, along with their appropriate scale, were located on each of the range images. The canonical orientation(s) were then subsequently assigned to each keypoints based on the image gradients. Based on the (x, y, σ, θ) information, feature descriptors were extracted from each keypoint location. Matching can then be conducted on the images by comparing the feature descriptors using nearest-neighbour-algorithm where potential matches were identified and then verified using the Hough Transform.

Each pairwise combination of images is compared, thereby formulating a match-matrix which can be used to interpret the obtained results. The match-matrix contains the percentage of positive and filtered matches for each pairing. Furthermore, the

results can be plotted within a ROC space and if they appear above the line of no discrimination, the results are acceptable.

Table 6.7 summaries the results obtained in this chapter. This shows that the feature descriptors demonstrates a good self-matching properties and have a good recognition rate at over 50% of keypoint matching rate, in particular with respect to in-plane rotational changes. These results were also reflected within the ROC space, concluding there are few FP results, illustrating the feature descriptors are stable to rotational changes.

Subject	**Rotations**	**# pairings**	**Images Achieving Matching Rate >**		**Self-matching %**
			50%	**60%**	
Human Face	**In-plane**	1369	>79.9%	>50.1%	>99.9%
	Out-of-plane	361	>51.8%	>36.3%	100%
Mannequin Head	**In-plane**	1369	>68.7%	>46.8%	>99.9%
	Out-of-plane	361	>46.8%	>35.2%	100%

Table 6.7: Summary of the results obtained using 2.5D SIFT on in-plane and out-of-plane images.

The results obtained in this chapter shows that the feature descriptors shows an useful degree of invariance to in-plane rotations, however the robustness of the feature descriptors against out-of-plane rotational would benefit by improvement. Since 2.5D data are reliable, there is the potential to exploit the 3D surface information afforded by the range imaging modality by finding local pose of keypoint sampling patches and expressing this information within the feature extraction framework to improve (out-of-plane) classification. To this end, the slant and tilt information can be deduced by calculating the *surface normals* on the range images and thereby allow the extraction of the *canonical normals* in the form of the *canonical slant* and *canonical tilt*. Having computed the slant and the tilt at each keypoint location over a measurement aperture, the current circular Gaussian window used for feature extraction can be warped to the shape of an elliptical window. By this means it is possible to correct the 3D pose, allowing matching of the 2.5D keypoints to be more stable to changes in pose angle. The methodology is proposed and outlined in the next chapter.

Chapter 7

Local Pose Estimation

An implementation of a 2.5D SIFT framework has been described in Chapter 5, in which the invariance properties of the feature descriptors were investigated in Chapter 6. In order to fully exploit the 3D information afforded by the range imaging modality, it is possible to deduce the local slant and tilt information from the range images, enabling the interpretation of the three Euler angle viewpoint rotational changes. To this end, this chapter is devoted to the theory and methods for deriving the slant and tilt information from the surface normals of the range images, and the integration of this information into the existing feature extraction framework by means of the concepts of canonical slant and canonical tilt. Thereafter it becomes possible to adapt the circular Gaussian measurement aperture described in Section 5.3 into an elliptical window, in order to facilitate invariance to 3D rotational changes.

7.1 Objectives

The implementation of the 2.5D SIFT framework described hitherto has focused on local feature matching that relies on estimating and recovering the in-plane rotation at each sampling keypoint. However, it is necessary to take into account the degree of out-of-plane rotation exhibited at each keypoint as well, since the relative viewpoint orientation at the keypoint sampling location (compared to that of the matching keypoint extracted from a different view) is not known a priori with respect to the range image capture device gaze direction. For instance, as an approximately planar surface patch is rotated to a different viewpoint orientation, the area representing this rotated patch will be reduced (or enlarged, depending on the degree of orientation), compared

to the original patch. Hence, if no estimation of the local slant and tilt is taken into account, the keypoint matching will result in comparing different regions of the surface. This concept is illustrated in Figures 7.1 and 7.2 where Figure 7.1 shows a 20×20 patch extracted from a synthetic sinusoid range image. The sinusoid range image is then rotated to a different viewpoint, where the same fixed 20×20 measurement aperture is applied to the rotated image, resulting in a different patch (Figure 7.2) from the baseline image being extracted. However, if the local slant and tilt are taken into account, and the circular measurement aperture is adapted to an elliptical shape, the extracted patches sample the same underlying surface information. As a result, it is necessary to estimate and recover the 3D pose of the keypoints by estimating their local slant and tilt. Thereafter it becomes possible to take the canonical surface orientation direction of each keypoint into account and correct this in order to allow 2.5D SIFT keypoints representing the same range surface patch, but captured from different viewpoints, to be matched.

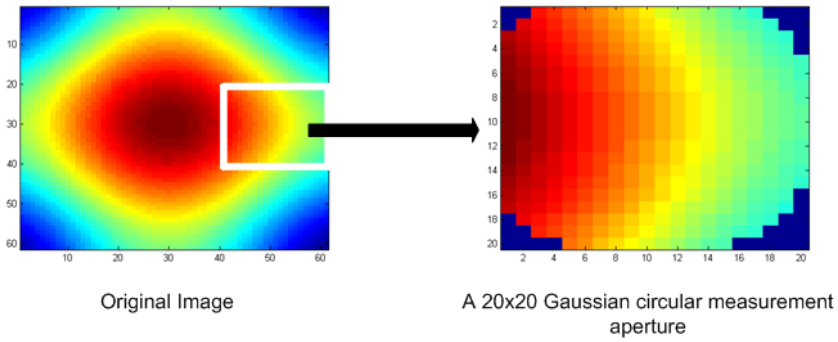


Figure 7.1: A 20×20 patch is extracted from the original image, using the Gaussian circular measurement aperture.

Two basic assumptions are made here: firstly the local (x, y) spacing is sufficiently uniform to allow normals to be calculated directly on the range map without transferring to X, Y, Z space coordinates; and secondly, the assumption of approximate local planarity in this work implies there is only one *canonical slant* and one *canonical tilt* assigned for each keypoint location at this stage of the work. Accordingly, there could be more than one dominant direction of the population of the *surface normals* (i.e. the *canonical normals*), and therefore represents a limitation of the pose corrected version of 2.5D system. Further investigation is required in order to address this issue fully.

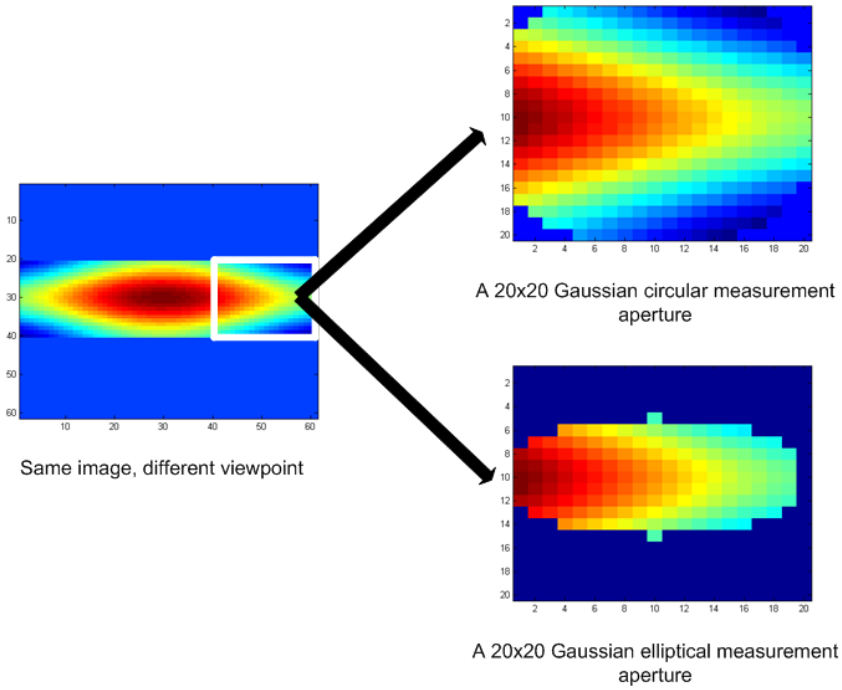


Figure 7.2: A rotated version of the original image, where a 20×20 patch is extracted using a Gaussian circular measurement aperture, resulting in a different patch to be extracted. This can be corrected by applying a Gaussian elliptical measurement aperture, according to the local slant and tilt, as shown in the bottom right diagram.

The slant and tilt can be derived from the surface normals of the range images and the canonical slant and canonical tilt can subsequently be determined for each keypoint using a similar methodology to that used for the calculation of the canonical orientations described in Section 5.2. Once the slant and tilt have been computed for each keypoint location, the circular Gaussian support region used for the feature extraction can be warped into the shape of an elliptical measurement aperture, according to the magnitude of the slant and tilt. If zero slant and tilt is present for a particular keypoint, the measurement aperture will remain circular (i.e. no adjustment is required). The slant determines the aspect ratio of the projected ellipse, whereas the tilt determines the orientation of the ellipse. By this means it is possible to correct the 3D pose (to incorporate orientation, slant and tilt), allowing matching of the 2.5D keypoints to be more stable to changes in pose angle.

This chapter investigates the derivation of the slant and tilt, and the methodology

used to incorporate this extra information into the existing feature extraction framework. This enables the local 3D pose to be estimated and corrected, hence providing a better framework for range image analysis. This chapter contains only the theory and methodology for the estimation along with the correction of the 3D pose, where validation of the process will be addressed and investigated in the next chapter. The remainder of this chapter is organised as follows: Section 7.2 presents the theory and algorithms used to derive the slant and tilt from range images; Section 7.3 details how the canonical slant and canonical tilt components are incorporated into the feature extraction process by adapting the circular Gaussian window into an elliptical window.

7.2 Local Surface Pose Estimation

This section describes the algorithm employed to calculate the slant and tilt from range images based on the surface normals. Having computed the slant and tilt, it is possible to derive the canonical slant and tilt for each keypoint location. The methodology used to calculate the slant and tilt is presented in this section.

7.2.1 Basic Concepts

Figure 7.3 illustrates the three Euler angle viewpoint rotational changes. The in-plane (roll) rotation can be estimated by means of the local image gradient orientations, as discussed in Section 5.2. The out-of-plane (yaw and pitch) rotations can in turn be estimated by the slant and tilt angles. Table 7.1 details the different types of angles that can be extracted from the three Euler viewpoint rotational changes.

	Rotation	**Angle**
In-plane	roll	orientation θ
Out-of-plane	yaw	slant ϕ
	pitch	tilt τ

Table 7.1: The different angles that can be extracted from the three Euler angle viewpoint rotations.

7.2.1.1 Calculations of Slant and Tilt

The surface normals $[Nx, Ny, Nz]$ can be computed from the first Gaussian derivatives of the range images, with the appropriate σ_G within the scale-space (Section 5.4), as

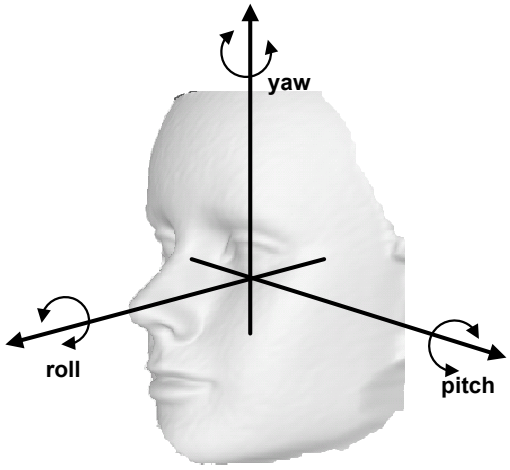


Figure 7.3: Diagram illustrating the three Euler angle viewpoint rotational changes. These are the in-plane *roll* rotation, out-of-plane *yaw* (left/right) rotation and out-of-plane *pitch* (up/down) rotation. Different angles can be estimated accordingly, as shown in Table 7.1.

shown in Equation 7.2.1 (Sze et al., 1998) where f_x and f_y are the first Gaussian derivatives of the range images in the x and y direction respectively.

$$[Nx, Ny, Nz] = \frac{[-f_x, -f_y, 1]}{(1 + f_x^2 + f_y^2)^{\frac{1}{2}}} \tag{7.2.1}$$

The slant ϕ and tilt τ can then be computed using Equations 7.2.2 and 7.2.3 (Norman et al., 2006) respectively:

$$\phi = tan^{-1}\left(\frac{\sqrt{N_x^2 + N_y^2}}{Nz}\right) \tag{7.2.2}$$

$$\tau = tan^{-1}\left(\frac{Nx}{Ny}\right) \tag{7.2.3}$$

The next section outlines the derivation of the canonical slant and tilt, by incorporating these basic concepts into the canonical orientation assignment stage, thereby providing a vector describing the state of a sampled keypoint patch in terms of the $(x, y, \sigma, \theta, \phi, \tau)$ information.

7.2.2 Derivation of the Canonical Slant and Canonical Tilt

Based on the position (x, y) and scale σ of each keypoint location, a consistent canonical slant and tilt is assigned using the surface normals $[Nx, Ny, Nz]$, as shown in Equation 7.2.2 and Equation 7.2.3.

Similar to the canonical orientation assignment, described in Section 5.2, two histograms of 360-bins each are employed to record the values of the slant and tilt from each keypoint with respect to its appropriate scale σ. Each item within the slant and tilt histogram is weighted by a standard Gaussian envelope (of sigma σ) in order to preserve continuity at the centre weight. In other words, the centre of the measurement aperture (i.e. the sub-pixel location corresponding the keypoint itself) has the strongest weight. The further away from the centre of the measurement aperture, the smaller the weight.

Once the slant and tilt histograms have been derived, the values of the histograms are smoothed and stabilised by distributing each entry over a number of bins, by means of a 1D symmetric Gaussian convolution kernel of $\sigma \approx 7$-bins. The peaks are then located and filtered to sub-bin precision using interpolation, where a quadratic polynomial is fitted to the three histograms values on the peak and either side of the peak, thereby formulating the canonical slant and tilt respectively, as shown in Equation 7.2.4 and 7.2.5 below.

$$\begin{pmatrix} (\angle_{peak} - \Delta\angle)^2 & \angle_{peak} - \Delta\angle & 1 \\ \angle_{peak}^2 & \angle_{peak} & 1 \\ (\angle_{peak} + \Delta\angle)^2 & \angle_{peak} + \Delta\angle & 1 \end{pmatrix} \begin{pmatrix} a \\ b \\ c \end{pmatrix} = \begin{pmatrix} H(\angle_{peak} - \Delta\angle) \\ H(\angle_{peak}) \\ H(\angle_{peak} + \Delta\angle) \end{pmatrix} \tag{7.2.4}$$

where $\angle$ represents either ϕ or τ. $H(\angle_{peak})$ is the histogram value at $\angle_{peak}$.

The canonical slant and tilt of the sampling patch is given by:

$$\angle_{canonical} = -\frac{b}{2a} \tag{7.2.5}$$

where $\angle_{canonical}$ is either the $\phi_{canonical}$ or $\tau_{canonical}$.

From the above, the derived $(x, y, \sigma, \theta, \phi, \tau)$ information is used to extract feature descriptor(s) at each keypoint location using the canonical slant and tilt to correct the 3D pose. This is detailed in the next section.

7.3 Pose Correction

In order to conduct a full 3D pose estimation and correction for range image analysis, it is necessary to take into account all three of the Euler's rotations, namely the in-plane (roll) and out-of-plane (yaw and pitch) rotations. The in-plane rotational changes have already been taken into account, as shown in Chapter 5, where each range image patch has been rotated to its canonical orientation at the feature extraction stage. Here, it is essential to incorporate the canonical slant and tilt into the feature extraction stage, thereby completing the full 3D pose correction. This section details the methodology developed to extend the current feature extraction stage to include the slant and tilt information.

7.3.1 Incorporation of the Slant and Tilt into the Feature Descriptors

The slant and tilt information can be expressed within the feature extraction stage by means of warping the circular Gaussian measurement aperture into an elliptical window, depending on the degree of the local slant and tilt presented within each individual keypoint sampling patch, using the following approximation. The elliptical regions are formed based on the canonical slant and canonical tilt, where the slant gives the aspect ratio of the projected ellipse and the tilt determines the orientation of the ellipse (Norman et al., 2006).

In order to generate a Gaussian elliptical measurement aperture at a keypoint location (x_i, y_i), first create a symmetric two-dimensional Gaussian square kernel with scale σ_i corresponds to the keypoint scale used to place the Gaussian support region on an image with sub-pixel accuracy as shown in Equation 5.2.5, where the centre of the measurement aperture is adjusted to the sub-pixel x and y offsets accordingly. Then apply the equation of an ellipse to the Gaussian square kernel to adjust the aspect ratio of the measurement aperture, in order to satisfy the following equation:

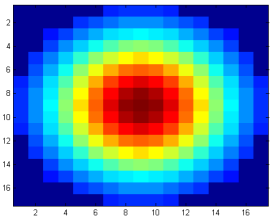
$$\frac{(x_i - \text{offset}_x)^2}{a^2} + \frac{(y_i - \text{offset}_y)^2}{b^2} - 1 \geq 0 \tag{7.3.1}$$

where $\frac{b}{a}$ = aspect ratio. Therefore, a = radius = width of kernel and $b = acos(\phi)$.

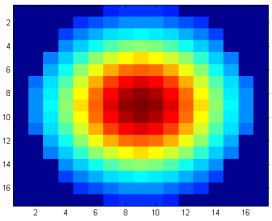
If $a = b = 1$, the Gaussian measurement support region will remain circular. Similar to the computation of the Gaussian circular measurement aperture, the filter coefficients

of the Gaussian elliptical support region G are normalised to sum to unity to satisfy Equation 5.2.7.

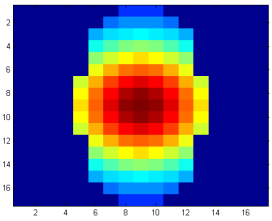
Figure 7.4(a)-(g) illustrate seven examples of the Gaussian elliptical measurement apertures (with no offsets), comprising of different values of slant and tilt, where Figure 7.4(a) illustrates an unit circular Gaussian support region, obtained when the values of the slant and tilt are both zero.



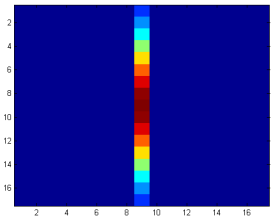
(a) Tilt=0°, slant=0°



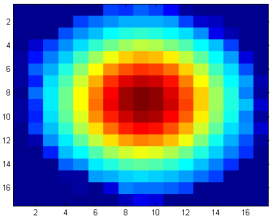
(b) Tilt=0°, slant=30°



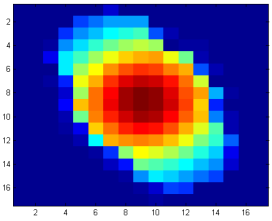
(c) Tilt=0°, slant=60°



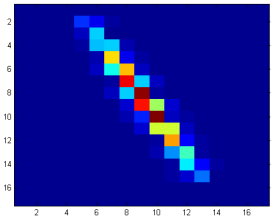
(d) Tilt=0°, slant=90°



(e) Tilt=30°, slant=30°



(f) Tilt=30°, slant=60°



(g) Tilt=30°, slant=90°

Figure 7.4: A few examples showing the Gaussian elliptical windows of different values of slant and tilt. Each measurement aperture is of size 17×17, with $\sigma = 4.25$.

The new feature descriptor extraction process, which incorporates the slant and tilt information to express a feature descriptor with the ability to correct the 3D pose fully, is slightly altered from the standard 2.5D feature descriptor extraction process presented in Section 5.3. The process is presented as follows:

1. Similar to the feature descriptor extraction methodology outlined in Section 5.3, the sampled range image patch is rotated to its canonical orientation in order to facilitate viewpoint rotational invariance. The values of H, K, $k1$ and $k2$ curvatures can be computed from the first and second Gaussian derivatives, with the appropriate σ. It is then possible to categorise the underlying distribution of the surface types, using the bounded $[-1, 1]$ shape index (Equation 4.2.4). The degree of the local curvedness (Equation 2.3.6), can be computed from the first and second Gaussian derivatives and is used as a weight for the shape index. Moreover, the local image gradient orientation (Equation 5.2.1) and the corresponding local magnitude estimate (Equation 5.2.2) can also be computed.

2. Nine *elliptical* Gaussian weighted sub-regions, overlapped by one standard deviation, are placed over each sampling patch, according to the appropriate scale, as shown in Figure 7.5. The elliptical regions are formed based on the canonical slant ϕ and tilt τ, where ϕ gives the aspect ratio of the projected ellipse and τ determines the orientation of the ellipse. In other words, if the value of the calculated slant and tilt are both zero, then the Gaussian measurement aperture will remain circular. This elliptical measurement aperture, along with the rotation of the sampling patch, accommodates the three Euler's viewpoint rotational changes. Overlapping of the Gaussian sub-regions reduces the spatial aliasing that occurs during the sampling stages and also enforces the spatial continuity of the sampling patch. For example, a small shift in the location of the keypoint should result in small (continuous) changes in the magnitude of the extracted keypoint descriptor.

3. Following from Section 5.3, for each of the nine elliptical sub-regions placed over the sampling patch, the local distribution of the relative frequencies of the nine surface types is represented within a histogram, weighted by the degree of curvedness. Similarly, an eight-element histogram, covering the 360° range of orientations, can be formulated, and is weighted by its magnitude. Each histogram is normalised to unity magnitude (i.e. to a unit vector) and the influence of large histogram values in each normalised histogram is reduced by clipping the value at a threshold of $\frac{1}{\sqrt{a}}$, where a is the number of bins in the histogram. The histograms are then concatenated to form Hi, and the combined histogram is then

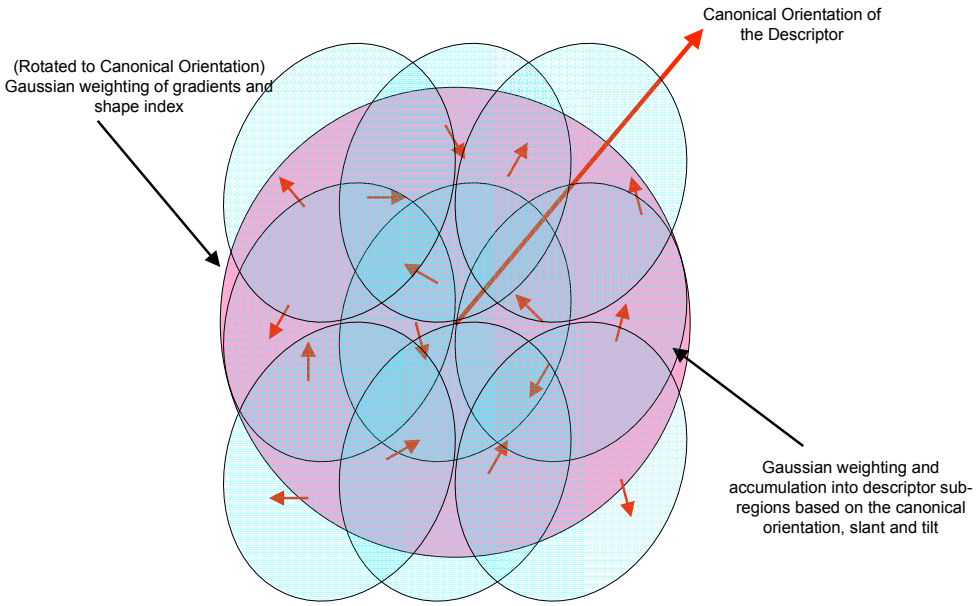


Figure 7.5: Placement of the nine elliptical sub-regions, with the spatial support at one standard deviation, over the keypoint location is illustrated.

normalised to unity magnitude.

$$\widehat{LocalHist_i} = \left(\widehat{Hi_{surface}}\right)\left(\widehat{Hi_{orientation}}\right) \tag{7.3.2}$$

4. The nine normalised histograms $LocalHist_i$ are juxtaposed to form the final feature descriptor for each keypoint location:

$$Descriptor_{\theta_{canonical}} = \left(\widehat{LocalHist_1}, \widehat{LocalHist_2}, ..., \widehat{LocalHist_9}\right) \tag{7.3.3}$$

7.4 Summary and Discussions

This chapter presents the methodology used to adapt the 2.5D feature descriptor described in Section 5.3 to accommodate pose estimation involving out-of-plane rotations. Local pose estimation plays an important role in feature extraction since the orientation in which the image was captured at is not known a priori. By computing the canonical slant and the canonical tilt, along with the canonical orientation at each keypoint location, the $(x, y, \sigma, \theta, \phi, \tau)$ sextet can be extracted, thereby allowing feature descriptors to be extracted from each keypoint location. Feature descriptors are extracted at each

keypoint location by fitting nine elliptical sub-regions, overlapped by one standard deviation, over the keypoint. The image patch used to sample each keypoint for the range image is rotated to its canonical orientation prior to this feature extraction process. The canonical slant and tilt provides the aspect ratio and the orientation of the projected sampling Gaussian ellipsoid respectively. This process potentially allows the 3D pose to be estimated and corrected for each keypoint on the range image, since it ensures that the underlying surface patches are being compared consistently, which cannot be accomplished by using a fixed window sampling technique. Moreover, this local pose estimation and correction technique can potentially be applied to 2D sampling windows in order to improve 2D matches, provided the corresponding range images are available to allow the estimation of the local slant and tilt.

However, there are limitations to the pose corrected version of the 2.5D SIFT: firstly it only allows local pose estimation to be conducted at this stage of the work. Furthermore, due to the local planarity assumption adopted here, only one canonical slant and one canonical tilt are assigned to each keypoint location and further investigation is required to determine if this assignment is sufficient. Otherwise, it may be necessary to extend the system to employ multiple canonical slants and tilts, which was beyond the scope of the current investigation.

The next chapter will investigate the invariance properties of the pose corrected version of 2.5D SIFT system by firstly validating the incorporation of the canonical slant and tilt into the feature descriptor. Furthermore, it is important to be able to match feature descriptors extracted from range images of different sizes and depicting the same target object at different scales. As a result, the main validation is divided into three parts: firstly, the viewpoint rotational invariant properties of the pose corrected version of the 2.5D SIFT system is investigated. Secondly, the scale invariance properties of the feature descriptors by matching images at different scales is investigated. Finally, performance comparisons between standard 2D SIFT on 2D images and 2.5D SIFT on 2.5D images, against rotational, changes will be made.

Chapter 8

Validation

In the previous chapter, an adapted version of the 2.5D SIFT system was devised where the local slant and tilt information was taken into account by means of the canonical slant and canonical tilt, thereby allowing the local pose of each keypoint location to be estimated and corrected. This is achieved by firstly calculating the canonical slant and tilt for each keypoint, followed by the incorporation of this information into the feature descriptor extraction process, where the circular Gaussian support region is warped into an elliptical measurement aperture. The canonical slant and tilt determines the aspect ratio and the orientation of the ellipse respectively. The purpose of this chapter is twofold: it firstly addresses the performance rate of this pose estimated and corrected version of the 2.5D SIFT system against all three Euler angle changes, as well as investigating the performance of the system when applied to range images of different sizes and depicting the same target object at different scales. Secondly, this chapter will compare the performance rate between standard 2D SIFT on 2D images and the pose corrected version of 2.5D SIFT with range images.

8.1 Objectives

The aim of this chapter is to validate the pose corrected 2.5D SIFT system (denoted 2.5D_{pc}) against all three Euler's rotational changes. Table 8.1 shows the list of experiments conducted in this chapter. The adopted 2.5D SIFT system has been extended to estimate and adapt to the local 3D pose at each keypoint location, using the slant and tilt information, and therefore can potentially improve the matching performance of images expressing out-of-plane rotations. Apart from the invariance properties of

the feature descriptors to rotational changes, it is important to address scale issues, since the range images to be compared can be of different sizes. Accordingly, the standard scale extraction mechanisms of the 2.5D SIFT algorithm will be applied to range images to estimate the scale of local structures represented as keypoints. Moreover, a comparison between the performance of the standard 2D SIFT and the 2.5D SIFT is conducted in this validation. The results for each validation will be presented in two ways: the match-matrix illustrating the percentage of matched keypoints and the matches plotted in a ROC space, showing visually the TPR versus the FPR.

System	Images	Scale	Rotations
$2.5D_{pc}$ SIFT	Range Images	Fixed	In-plane (Synthetic) Out-of-plane (Real) Out-of-plane (Synthetic)
		Variable	In-plane (Synthetic) Out-of-plane (Real) Out-of-plane (Synthetic)
2D SIFT	Intensity Images	Fixed	In-plane (Synthetic) Out-of-plane (Real) Out-of-plane (Synthetic)
		Variable	In-plane (Synthetic) Out-of-plane (Real) Out-of-plane (Synthetic)

Table 8.1: Table illustrating the experiments involved in this chapter.

The validation of the $2.5D_{pc}$ system conducted in this chapter will follow closely to the protocol outlined in Section 6.2.2, where feature descriptors are extracted from the scale-space defined keypoint locations. The canonical orientation, canonical slant and canonical tilt are all taken into account during the feature descriptor extraction stage, where the sampling patch is rotated to its canonical orientation and the circular Gaussian sub-sampling regions are adapted into elliptical apertures defined by the canonical slant and canonical tilt accordingly. In order to be able to compare the performance of the originally proposed 2.5D SIFT with this $2.5D_{pc}$ SIFT, the data set employed for the first part of the investigation is identical to the data set described in Section 6.2.1 (i.e. synthetic in-plane rotated range images are simulated using MATLAB and out-of-plane images are captured using the stereo-pair imaging system described in Section 3.2.). A further data set simulating out-of-plane viewpoint changes (about

both the yaw and pitch axes) is generated synthetically using MATLAB based on the methodology detailed in Appendix C, for rotations up to $\pm 40^{\circ}$.

A MATLAB version of the standard 2D SIFT, released by Lowe (2005), is used in this work to investigate the performance rate between the standard 2D SIFT and 2.5D SIFT. The standard 2D SIFT provides the $(x, y, \sigma, \theta, 128\text{-element descriptor})$ for each 2D image processed. The feature descriptors extracted from 2D SIFT encapsulate the image gradient orientation information only. The $(x, y, \sigma, \theta, 128\text{-element descriptor})$, along with the image itself, can be recorded in a database, where the images can be matched using the Hough Transform (HT) as in the 2.5D SIFT system. As a result, the performance rate can be obtained for each of the Euler rotations.

It is worth noting that the majority of the range image matching methodologies in the existing literature, in particular with respect to the out-of-plane rotational changes, cite the acceptable range (degrees) where matching can be sustained to be $\approx \pm 20^{\circ}$ about the yaw axis and $\approx \pm 15^{\circ}$ about the pitch axis (Akagündüz and Ulusoy, 2007; Pansang et al., 2005). Moreover, synthetically rotated range images will potentially lose information at over 30° of out-of-plane rotations (Li and Su, 2006).

The remainder of this chapter is organised as follows: Section 8.2 describes the data set used for the different experiments conducted in this chapter. Section 8.3 investigates and addresses the performance rate of the 2.5D_{pc} SIFT system against rotational changes on a) single-scale 2.5D range images (Section 8.3.1) and b) range images of different scales (Section 8.3.2). Section 8.4 determines the performance rate of standard 2D SIFT on 2D images against rotational changes and finally, Section 8.5 summaries the findings of the validation.

8.2 Data Set

In order to validate the 2.5D SIFT system, there are three different sets of data used for this investigation, namely the images that simulate the Euler's in-plane rotations (i.e. about the roll axis), the real out-of-plane (about the yaw axis) images captured using the stereo-pair system described in Section 3.2 and finally the images simulating the out-of-plane (about both the yaw and pitch axis) rotations, derived using the algorithm outlined in Appendix C. Table 8.2 details the data sets used in this validation.

Type	Method	Modality	Degrees of rotation	Subjects
In-plane (about roll axis)	MATLAB simulated using `imrotate` from baseline (0°) images	2D & $2.5D_{pc}$	$0° - 350°$	1×Human Face & 1×Mannequin Head
Out-of-plane (about yaw axis)	Real data captured using stereo-pair system	2D & $2.5D_{pc}$	$-90° - 90°$	1×Human Face & 1×Mannequin Head
Out-of-plane (about yaw axis)	MATLAB simulated using algorithm described in Appendix C from baseline (0°) images	2D & $2.5D_{pc}$	$-40 - 40°$	1×Human Face & 1×Mannequin Head
Out-of-plane (about pitch axis)	MATLAB simulated using algorithm described in Appendix C from baseline (0°) images	2D & $2.5D_{pc}$	$-40 - 40°$	1×Human Face & 1×Mannequin Head

Table 8.2: Data sets employed for the validation conducted in this chapter.

Each of the 2.5D images used in this validation is pre-processed to suppress the random noise that exists in the data and furthermore the range images are reduced to approximately 244×369 pixels from 1498×2249 pixels using an half-octave Gaussian pyramid. Similarly, the corresponding 2D images are pre-processed using an half-octave pyramid to reduce the image sizes.

Apart from the rotation invariance properties, a feature descriptor should also be scale invariant. Thus, the investigation of the scale invariance properties of the $2.5D_{pc}$ SIFT is conducted by matching the feature descriptors extracted from images of different sizes. This is achieved by enlarging and reducing the size of the range images using a half-octave Gaussian pyramid. Table 8.3 shows the size of the range images generated in this work and the images that are used in this experiment are marked with a $\star$. For the purposes of this experiment, the baseline image will be at a scale that corresponds to Level 6 of the half-octave Gaussian pyramid (marked as bold in the table) and the images between Levels 4 and 8 will be used in order to keep the computational costs

to a minimum.

Levels	Size of images
1	1498×2249
2	995×1495
3	746×1121
4*	494×744
5*	370×557
6*	**244×369 (Baseline)**
7*	182×275
8*	119×181
9	88×134

Table 8.3: Size of the range images produced from each of the half-octave Gaussian pyramid level. Level 1 corresponds to the original range image size. The baseline image is selected to be of size 244×369 pixels and is used to compare to range images of other sizes (the entries marked with a $\star$ are used in this experiment).

Using the feature descriptors extracted from the 244×369 pixels range images as the baseline here, the feature descriptors extracted from the remainder set of data (i.e. same viewpoint angles but of different sizes) can be compared to the baseline images, thereby forming a match-matrix that shows the percentage of the matched and filtered keypoints at each scale. The protocol follows the methodology outlined in Section 6.2.2.

The next section outlines the results obtained using the pose corrected version of the 2.5D SIFT system on both in-plane and out-of-plane range images.

8.3 Performance Rate of 2.5D$_{\mathbf{pc}}$ SIFT

This section presents the results obtained from the validation of 2.5D$_{\mathrm{pc}}$ SIFT system against Euler angle changes on a) images of fixed scale and b) images of different scales. In order to compare the performances between the originally proposed 2.5D SIFT (without the 3D pose estimation and correction) and the updated version of the 2.5D SIFT (with the 3D pose estimation and correction built-in the system), the validation process will follow closely to the protocol detailed in Section 6.2.2, with the appropriate adaptation to allow the incorporation of the local 3D pose estimation and correction. The results are presented in two ways: firstly the percentages of the matched and filtered keypoints are presented in a match-matrix, showing all the combinations of

the range image pairings (between rotated images). This will be presented in a match-matrix table as well as visually by means of a 3D bar chart. Secondly, the results will be presented in a ROC space, thereby providing a visual measure of the reliability of the matches between the feature descriptors extracted from range images with different rotational changes.

The results are also summarised for each experiment and the statistical significance of differences, using the Wilcoxon Matched-Pairs Signed-Rank test, between the performance of the 2.5D$_{\text{pc}}$ SIFT and the baseline 2.5D SIFT (where appropriate) is presented by means of one of the following symbols:

- $\ll$: This result is significantly worse ($p < 0.01$) than the baseline.
- $<$: This result is significantly worse ($p < 0.05$) than the baseline.
- $=$: This result has no statistically significant difference ($p > 0.05$) from the baseline.
- $>$: This result is significantly better ($p < 0.05$) than the baseline.
- $\gg$ This result is significantly better ($p < 0.01$) than the baseline.

Stable keypoints are located on each of the range images, along with their appropriate scale σ, thereby allowing the canonical orientation(s) θ and the canonical slant ϕ and canonical tilt τ to be identified. Therefore, feature descriptors can be extracted for each $(x, y, \sigma, \theta, \phi, \tau)$ accordingly by firstly rotating the sampling patch to its θ followed by the extraction of the surface types and orientation histograms from each of the nine overlapping Gaussian elliptical sub-regions.

The resulting $(x, y, \theta, \phi, \tau, \text{descriptor})$, along with the normalised range image, are stored within a database. A similarity transform is computed between the sets of feature descriptors extracted from one range image and the feature descriptors extracted from a different range image (captured at a different viewpoint angle or at different image scale) by means of the Hough Transform. As a result, clusters of matching features with a consistent interpretation (i.e. matches between features exhibiting the same relative shift in orientation, translation and scale) are identified. If three or more entries are located in each cluster, it is possible to apply a robust affine transform fitting procedure to the cluster in order to recover the affine pose between the matched features and also

identify outliers. Note that the use of HT as a filtering process is itself not 100% reliable and is a known limitation of this work. However, it is difficult to establish ground truths, even if the rotations between matched range images are known in advance, therefore as an approximation of the matched keypoints, the use of the HT as a filtering process is deemed adequate at this stage of the work. The percentage of the matched and filtered keypoints (Equation 6.2.1) can be determined, along with the approximation of the true positive rate and the false positive rate (Equation 6.2.2). This step is performed repeatably for each pairing of the range images (of different viewpoint angles or different sizes) and as a result, a match-matrix and a graph plotted within a ROC space can be formed.

The remainder of this section is organised into two major parts: firstly the investigation conducted on fixed scale images (of size 244×369 pixels), against rotational changes (both in-plane and out-of-plane); Secondly, the investigation conducted range images of different scales, against rotational changes, will be presented.

8.3.1 Fixed Scale 2.5D$_{\mathbf{pc}}$ SIFT

The performance of the 2.5D$_{\mathrm{pc}}$ SIFT system against rotational changes on fixed scale range images (of size 244×369 pixels) is analysed here. Feature descriptors are extracted from each of the test range images, thereby allowing feature descriptors extracted from range images captured at a different angle to be matched. Every pairing of the feature descriptors extracted from range images captured at different angles will be compared, resulting in a match-matrix of results. For each bar graph representing the match-matrix, the lower axes of the graph show the viewpoint angles of each pair of the compared range images, while the height of the columns shows the percentage of matched keypoints. A total of 72 range images simulating in-plane rotations (2×36), 38 captured out-of-plane images (2×19) and 36 range images simulating out-of-plane rotations (2×9 rotated images about the pitch axis and 2×9 rotated images about the yaw axis) are used in this experiment.

8.3.1.1 Synthetic In-plane Rotations

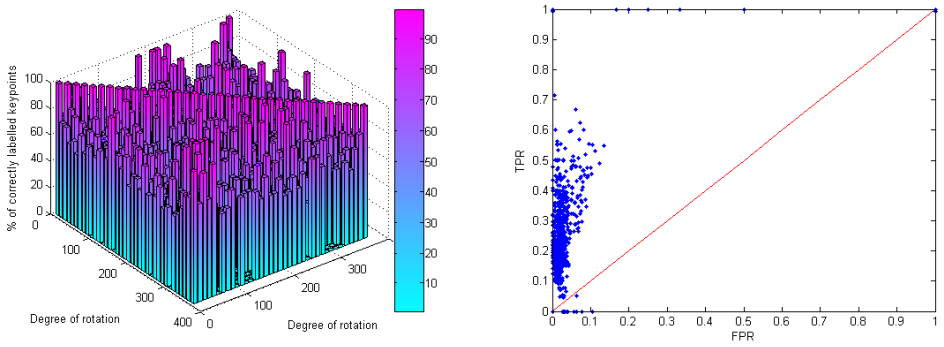
There are two sets of 2.5D in-plane rotation data: range images of a human face and range images of a mannequin head. The baseline (no rotation) image was rotated synthetically using the MATLAB function `imrotate`, with the bicubic interpolation

methodology. It was rotated in clockwise direction in increments of 10° up to 350° (i.e. a full circle, 37 images per set of data). Keypoints and feature descriptors were then extracted from the range images and stored in a database for matching. Results are shown here.

Table 8.4 summarises the results obtained by exploring all combinations of the feature descriptors extracted from the synthetically in-plane rotated range images of a mannequin head and a human face respectively. Statistically significant improvements from the baseline (2.5D SIFT) are shown using one of these five symbols to denote significance: $\ll$, $<$, $=$, $>$, $\gg$.

System	**Subject**	**Pairings**	**Average Matching Rate**	**Images Achieving Matching Rate > 50%**	**60%**	**Match-Matrix**	**ROC**
2.5D$_{pc}$	Mannequin head	1369	45.3%$\gg$	64.1%	43%	Table 8.5 Figure 8.1	Figure 8.1
	Human face	1369	64.9%$\gg$	88%	63.6%	Table 8.6 Figure 8.2	Figure 8.2

Table 8.4: Summary of results obtained by exploring all pairwise combinations of the feature descriptors extracted from the rotated in-plane range images of a mannequin head and a human face, using the 2.5D$_{pc}$ SIFT system.



(a) Match-Matrix (b) ROC

Figure 8.1: (a) The match-matrix results of the percentage of matched keypoints, produced from a set of in-plane rotational data of a **mannequin head** (from 0° at 10° increments in the clockwise direction up to 350°), using the 2.5D$_{pc}$ SIFT system. (b) The matching results presented in a ROC space. This figure can be compared with Figure 6.6, where the results were obtained from the same data using the 2.5D SIFT.

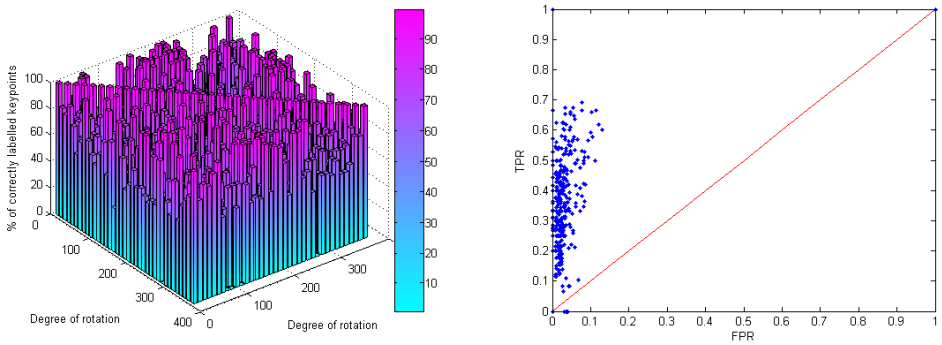
	0°	10°	20°	30°	40°	50°	60°	70°	80°	90°	100°	110°	120°	130°	140°	150°	160°	170°
0°	100	74.36	96	75	90.91	0	60	62.5	75	0	66.67	0	0	66.67	60	75	60	60
10°	71.79	100	88.1	86.67	66.67	0	0	69.23	0	0	0	0	0	57.14	0	55.56	100	66.67
20°	70.59	82.35	100	67.44	73.33	77.78	75	71.43	0	0	0	0	0	75	60	66.67	0	72.73
30°	60	78.57	65.79	100	70	70.59	57.14	0	57.14	0	0	85.71	66.67	0	66.67	66.67	0	75
40°	63.64	54.55	61.54	70	99.38	66.67	70	55.56	20	60	0	0	75	0	66.67	57.14	60	66.67
50°	55.56	77.78	54.55	76.47	63.16	100	72.41	65	77.78	0	0	60	57.14	0	0	57.14	60	54.55
60°	71.43	75	70	80	58.82	66.67	100	78.38	63.64	60	62.5	75	57.14	66.67	66.67	0	62.5	53.85
70°	60	75	55.56	60	66.67	72.22	71.43	99.43	87.5	71.43	57.14	60	0	0	57.14	66.67	0	57.14
80°	66.67	66.67	66.67	66.67	62.5	66.67	66.67	75	100	85.71	63.64	60	0	75	60	60	0	75
90°	0	75	0	66.67	0	87.5	71.43	93.33	84	99.38	69.23	63.64	57.14	75	0	75	0	66.67
100°	62.5	66.67	66.67	0	75	62.5	75	62.5	60	70.83	100	90.91	61.54	0	60	62.5	0	55.56
110°	69.23	60	60	63.64	57.14	57.14	55.56	70	60	62.5	69.57	100	80	60	57.14	53.85	37.5	66.67
120°	0	0	0	0	0	75	0	66.67	66.67	100	55.56	80.95	100	73.91	82.35	83.33	60	0
130°	66.67	70	0	50	71.43	61.54	83.33	60	57.14	57.14	0	63.64	66.67	100	79.31	54.55	69.23	53.85
140°	100	55.56	55.56	62.5	63.64	57.14	80	70	66.67	66.67	0	100	66.67	74.19	100	74.19	60	90
150°	0	60	75	0	75	71.43	62.5	60	0	0	0	0	0	63.64	78.05	100	75	80
160°	75	55.56	87.5	66.67	0	57.14	55.56	58.33	60	60	0	0	0	66.67	68.75	81.08	100	82.22
170°	66.67	55.56	60	71.43	60	40	57.14	62.5	66.67	66.67	0	0	42.86	66.67	0	72.73	76.74	99.43
180°	77.78	60	57.14	55.56	40	60	0	60	0	0	0	0	0	0	83.33	66.67	81.82	70.59
190°	57.14	75	57.14	60	62.5	71.43	75	60	58.33	0	75	0	57.14	0	60	57.14	78.57	84
200°	66.67	85.71	80	63.64	66.67	57.14	57.14	0	66.67	77.78	66.67	0	0	28.57	83.33	63.64	57.14	60
210°	66.67	57.14	25	0	0	60	66.67	0	60	0	66.67	75	66.67	100	75	0	0	66.67
220°	60	62.5	83.33	66.67	81.82	0	0	0	0	0	0	57.14	0	57.14	0	0	75	52.94
230°	80	66.67	66.67	66.67	66.67	60	0	0	0	0	0	0	0	0	0	0	60	62.5
240°	60	0	0	60	62.5	55.56	71.43	0	0	0	0	0	66.67	57.14	55.56	66.67	71.43	60
250°	0	75	0	57.14	62.5	60	57.14	0	66.67	0	75	0	0	83.33	0	66.67	66.67	71.43
260°	62.5	57.14	60	58.33	83.33	60	88.89	66.67	0	0	0	0	0	40	20	0	75	66.67
270°	71.43	0	60	0	0	0	0	0	0	0	0	0	100	0	0	0	20	60
280°	57.14	55.56	75	57.14	0	0	100	62.5	0	0	40	0	60	75	0	77.78	0	71.43
290°	60	80	0	60	100	16.67	100	100	100	0	0	0	60	0	0	0	0	100
300°	0	55.56	0	0	0	0	0	0	0	0	0	0	57.14	66.67	0	60	60	66.67
310°	57.14	0	62.5	62.5	80	0	0	0	0	0	0	0	0	66.67	66.67	75	66.67	62.5
320°	60	75	33.33	0	100	60	0	0	0	0	0	0	0	53.85	66.67	57.14	63.64	75
330°	66.67	60	75	66.67	0	66.67	66.67	0	0	0	0	0	0	60	60	0	57.14	60
340°	76.92	57.14	57.14	55.56	57.14	66.67	60	75	0	0	0	0	0	0	0	80	55.56	84.62
350°	78.79	93.33	80	0	66.67	66.67	0	60	0	0	0	0	0	0	62.5	60	60	55.56

	180°	190°	200°	210°	220°	230°	240°	250°	260°	270°	280°	290°	300°	310°	320°	330°	340°	350°
0°	69.23	55.56	85.71	66.67	57.14	66.67	0	75	75	66.67	0	0	0	0	62.5	55.56	70	83.33
10°	55.56	100	100	0	100	0	0	0	0	0	0	75	0	10	66.67	57.14	100	78.57
20°	77.78	66.67	60	81.82	0	0	0	0	0	0	0	66.67	75	0	0	55.56	57.14	57.14
30°	60	25	55.56	66.67	0	0	0	0	0	0	0	0	57.14	60	0	75	60	60
40°	83.33	100	0	91.67	55.56	53.85	0	0	0	0	0	0	0	66.67	57.14	66.67	66.67	54.55
50°	83.33	83.33	0	57.14	76.92	66.67	57.14	0	60	0	75	100	75	0	62.5	0	0	62.5
60°	81.82	57.14	66.67	100	55.56	76.92	57.14	0	60	60	0	0	60	0	77.78	60	75	66.67
70°	66.67	0	0	0	0	0	0	0	0	60	0	0	60	66.67	66.67	57.14	0	0
80°	0	55.56	66.67	62.5	75	83.33	0	54.55	62.5	0	0	75	60	83.33	60	75	57.14	66.67
90°	75	60	0	0	0	0	0	0	0	0	60	0	57.14	66.67	0	0	75	62.5
100°	62.5	0	0	83.33	62.5	54.55	60	60	0	0	0	0	75	0	66.67	75	0	0
110°	70	62.5	62.5	85.71	0	60	0	100	0	0	83.33	0	0	66.67	62.5	75	0	55.56
120°	75	0	0	0	0	0	0	57.14	75	0	0	0	0	0	40	0	0	0
130°	66.67	100	66.67	60	66.67	0	0	100	75	60	0	0	60	0	20	62.5	75	75
140°	66.67	57.14	60	0	57.14	0	60	75	66.67	75	100	0	0	66.67	66.67	80	58.33	66.67
150°	70	0	66.67	60	0	66.67	0	0	0	0	0	0	0	100	58.33	75	90	57.14
160°	82.35	70	75	57.14	0	0	0	60	0	0	0	0	0	66.67	62.5	62.5	55.56	83.33
170°	70.59	64.29	60	0	57.14	0	0	0	60	0	0	0	0	0	60	0	57.14	60
180°	100	85	96.3	57.14	57.14	0	0	0	75	0	0	60	0	75	0	0	71.43	70
190°	83.33	100	77.14	87.5	66.67	0	20	75	0	0	0	0	0	0	66.67	0	57.14	55.56
200°	75	82.61	100	72.22	81.82	66.67	0	0	0	0	0	0	66.67	66.67	55.56	60	66.67	66.67
210°	62.5	57.89	76.32	100	65.22	55.56	57.14	60	75	60	75	0	66.67	0	0	0	55.56	0
220°	76.92	55.56	78.57	76.92	100	63.33	80	60	85.71	0	0	0	0	57.14	62.5	83.33	0	100
230°	83.33	0	63.64	66.67	78.57	99.38	80	80	92.86	0	0	0	0	0	0	0	0	0
240°	83.33	60	57.14	66.67	66.67	66.67	99.32	60.87	60	75	0	0	60	60	0	55.56	0	60
250°	0	66.67	66.67	62.5	60	62.5	82.61	100	80.95	60	62.5	0	0	60	0	0	0	0
260°	100	0	0	0	60	100	75	76.19	100	70.83	62.5	75	0	0	66.67	0	0	75
270°	100	0	0	0	66.67	62.5	75	76.92	74.07	99.4	77.27	54.55	75	0	0	66.67	83.33	0
280°	62.5	66.67	0	60	62.5	75	0	69.23	60	86.96	100	70.83	72.22	60	75	83.33	75	62.5
290°	77.78	57.14	0	57.14	60	0	100	100	62.5	66.67	70.83	100	66.67	66.67	0	55.56	0	0
300°	0	100	0	0	0	57.14	60	0	0	57.14	87.5	79.17	100	69.57	85	52.63	83.33	0
310°	0	60	62.5	0	75	75	0	57.14	0	60	75	69.23	70.83	100	75	63.64	78.57	75
320°	54.55	66.67	55.56	60	71.43	66.67	0	60	0	0	0	0	62.5	60.87	100	81.48	55.56	66.67
330°	58.33	0	60	33.33	100	62.5	60	0	0	0	0	0	50	77.78	75.86	100	82.14	70
340°	0	60	58.33	0	60	57.14	71.43	57.14	0	0	0	80	66.67	53.85	57.14	80	100	78.57
350°	80	57.14	57.14	0	85.71	40	0	40	0	0	0	0	0	57.14	66.67	69.23	78.13	100

Table 8.5: Match-matrix results of the percentage of matching keypoints, produced from the in-plane rotational data of a **mannequin head** (from 0° at 10° clockwise increments up to 350°), using the 2.5D$_{pc}$ SIFT system.

	0°	10°	20°	30°	40°	50°	60°	70°	80°	90°	100°	110°	120°	130°	140°	150°	160°	170°
0°	100	76.74	87.5	84.62	92.86	100	57.14	0	0	60	0	75	60	62.5	60	100	62.5	95.45
10°	83.33	100	79.59	80.95	100	100	0	0	0	0	0	75	100	57.14	66.67	66.67	90	95
20°	80	73.91	100	83.72	76.67	75	62.5	75	0	0	0	0	66.67	55.56	55.56	63.64	75	57.14
30°	88.24	78.26	80.49	100	85.11	58.82	60	71.43	80	0	62.5	0	100	57.14	62.5	62.5	100	100
40°	75	77.27	81.82	79.07	100	73.47	66.67	53.33	0	0	0	60	66.67	57.14	100	75	62.5	75
50°	55.56	66.67	80	71.43	73.58	100	75.76	66.67	66.67	83.33	0	0	0	60	0	0	100	75
60°	66.67	62.5	81.82	73.91	80	72.22	100	82.14	87.5	62.5	57.14	66.67	80	57.14	83.33	57.14	75	55.56
70°	100	62.5	66.67	58.33	59.09	77.27	79.41	100	83.33	80	90	60	75	83.33	80	75	66.67	58.33
80°	71.43	75	60	62.5	66.67	63.64	60	77.14	100	80.77	78.57	92.86	100	60	62.5	57.14	71.43	77.78
90°	62.5	75	53.85	100	50	100	75	75	74.19	100	77.78	82.14	80	85.71	58.33	71.43	0	58.33
100°	57.14	60	55.56	66.67	75	57.14	62.5	66.67	77.78	75	100	77.78	84.21	55.56	62.5	55.56	57.14	70
110°	66.67	66.67	66.67	60	77.78	70	0	54.55	54.55	87.5	78.13	100	90.32	88.24	63.64	62.5	0	66.67
120°	77.78	100	66.67	66.67	0	60	0	66.67	60	93.33	86.36	82.76	100	83.72	76.67	70.59	100	70
130°	54.55	61.54	54.55	57.14	66.67	60	0	75	0	57.14	71.43	70.83	86.84	100	73.58	72	62.5	80.95
140°	81.82	62.5	58.33	66.67	75	100	75	0	0	57.14	60	78.57	84	81.25	100	78.38	59.09	68.18
150°	33.33	87.5	77.78	75	0	66.67	57.14	0	57.14	0	62.5	66.67	60	73.33	86.05	100	78.05	69.7
160°	100	66.67	70	100	57.14	83.33	0	66.67	0	0	0	60	57.14	72.73	75	72.5	100	74.51
170°	75	62.5	66.67	70	72.73	90	57.14	75	0	0	0	40	75	57.14	62.5	80	84.62	100
180°	52.94	88.89	62.5	0	83.33	53.85	0	55.56	0	75	0	75	100	71.43	75	66.67	88.89	79.17
190°	58.33	78.57	60	77.78	78.57	76.92	66.67	0	60	0	66.67	60	57.14	60	70	100	71.43	80.56
200°	71.43	86.67	75	75	57.14	72.73	62.5	66.67	0	66.67	71.43	60	88.89	75	66.67	57.14	66.67	76.19
210°	60	100	100	71.43	60	66.67	0	100	0	75	25	0	66.67	75	60	0	62.5	78.57
220°	61.54	100	78.57	60	100	92.31	55.56	75	0	0	100	60	0	66.67	57.14	55.56	0	90
230°	57.14	54.55	63.64	93.33	100	95	75	57.14	66.67	88.89	0	75	100	60	100	0	71.43	55.56
240°	53.85	87.5	81.25	55.56	60	100	60	100	66.67	60	0	100	60	57.14	90	0	57.14	100
250°	57.14	54.55	75	66.67	56.25	62.5	0	60	66.67	75	0	75	60	83.33	81.82	0	62.5	62.5
260°	56.25	60	62.5	54.55	54.55	75	60	0	62.5	0	60	80	60	66.67	77.78	66.67	57.14	0
270°	91.67	57.14	55.56	100	57.14	66.67	55.56	0	75	66.67	71.43	50	100	85.71	100	20	60	55.56
280°	0	76.47	60	64.29	0	62.5	60	0	0	0	40	60	20	83.33	0	66.67	70	57.14
290°	63.64	63.64	57.14	58.33	60	80	66.67	57.14	71.43	0	62.5	60	60	70	55.56	57.14	54.55	58.33
300°	55.56	60	71.43	55.56	11.11	66.67	62.5	75	60	60	0	60	66.67	55.56	60	71.43	85.71	73.33
310°	83.33	100	12.5	62.5	66.67	60	54.55	66.67	66.67	66.67	55.56	62.5	87.5	66.67	100	87.5	88.89	66.67
320°	55.56	60	87.5	100	75	60	0	0	66.67	62.5	77.78	55.56	100	81.82	66.67	92.86	100	76.92
330°	68.42	80	55.56	55.56	62.5	66.67	75	0	75	60	0	100	63.64	100	61.11	94.74	100	100
340°	74.36	85.71	55.56	66.67	0	66.67	0	0	62.5	60	0	66.67	100	75	80	100	100	70
350°	79.25	87.88	73.91	66.67	92.86	70	66.67	60	75	75	75	60	55.56	57.14	60	66.67	55.56	92.86
	180°	**190°**	**200°**	**210°**	**220°**	**230°**	**240°**	**250°**	**260°**	**270°**	**280°**	**290°**	**300°**	**310°**	**320°**	**330°**	**340°**	**350°**
0°	93.33	80	54.55	100	100	90.91	66.67	50	0	40	60	0	66.67	55.56	55.56	60	80	78.72
10°	62.5	100	82.35	85.71	100	61.9	100	83.33	75	0	0	25	57.14	60	100	78.57	86.67	81.08
20°	77.78	56.25	94.74	100	73.33	76.47	80	66.67	0	60	0	83.33	66.67	85.71	75	55.56	63.64	72.22
30°	100	62.5	100	100	100	100	63.64	66.67	66.67	57.14	66.67	60	100	57.14	75	66.67	77.78	56.25
40°	80	87.5	57.14	71.43	94.74	100	71.43	62.5	57.14	77.78	20	85.71	62.5	62.5	60	0	75	57.14
50°	78.57	62.5	73.33	56.25	100	96.3	66.67	54.55	87.5	0	62.5	60	66.67	60	63.64	55.56	62.5	87.5
60°	80	100	100	100	61.54	81.25	57.14	66.67	60	66.67	0	57.14	83.33	60	63.64	60	57.14	71.43
70°	84.62	88.89	60	61.54	91.67	35.29	88.89	60	60	100	0	83.33	100	66.67	53.85	60	66.67	55.56
80°	50	66.67	62.5	75	71.43	77.78	54.55	66.67	57.14	0	75	57.14	53.33	69.23	21.43	71.43	71.43	91.67
90°	70	62.5	55.56	53.85	58.33	70	60	75	57.14	75	0	66.67	60	77.78	100	57.14	75	63.64
100°	78.57	63.64	55.56	55.56	80	77.78	0	0	66.67	57.14	57.14	75	71.43	72.73	70	62.5	57.14	66.67
110°	0	57.14	62.5	54.55	57.14	62.5	75	75	57.14	0	60	66.67	90	72.73	100	66.67	66.67	66.67
120°	55.56	66.67	66.67	66.67	66.67	62.5	75	75	66.67	62.5	66.67	0	57.14	62.5	78.57	63.64	57.14	56.25
130°	52.94	57.14	66.67	58.33	100	71.43	60	0	0	75	71.43	66.67	75	60	100	62.5	100	100
140°	87.5	66.67	66.67	0	60	57.14	0	0	57.14	66.67	60	57.14	81.82	54.55	100	62.5	100	33.33
150°	96.3	83.33	81.82	57.14	62.5	55.56	0	0	60	57.14	60	62.5	91.67	76.92	57.14	100	90.91	90.91
160°	78.38	76.92	66.67	62.5	80	63.64	60	0	60	66.67	0	0	57.14	100	77.78	66.67	100	71.43
170°	84.91	83.33	80	87.5	66.67	80	77.78	75	14.29	0	60	66.67	60	100	100	92.86	100	66.67
180°	100	90.24	82.14	75	57.14	55.56	0	66.67	55.56	75	0	55.56	55.56	0	83.33	100	100	94.44
190°	86.67	100	80.39	76	61.54	58.82	66.67	100	75	40	60	55.56	100	75	100	80	90.91	85.71
200°	91.3	80.43	100	86	69.7	89.47	100	60	0	66.67	83.33	0	62.5	66.67	55.56	57.14	75	62.5
210°	93.75	81.82	86.05	100	76	72.41	58.82	63.64	55.56	0	66.67	60	0	60	75	57.14	57.14	60
220°	58.33	69.23	79.49	70.73	100	71.93	77.78	60	57.14	0	60	66.67	57.14	100	60	62.5	75	100
230°	62.5	80	66.67	70.83	81.63	100	80.56	71.43	66.67	66.67	0	57.14	0	75	66.67	87.5	57.14	57.14
240°	61.54	70	100	59.09	76.32	80.43	100	76.92	78.26	100	57.14	60	60	60	53.85	55.56	60	83.33
250°	54.55	70	84.62	85.71	61.9	67.5	77.14	100	88.46	94.44	78.57	75	60	75	66.67	55.56	57.14	75
260°	0	57.14	75	62.5	57.14	58.82	78.26	70	100	80.77	78.26	70.59	0	55.56	66.67	100	60	58.33
270°	71.43	80	60	60	57.14	56.25	76.92	86.36	83.33	100	94.44	81.48	87.5	66.67	72.73	71.43	57.14	20
280°	77.78	66.67	40	55.56	62.5	50	57.14	92.31	94.44	84	100	87.88	82.61	70	88.89	57.14	66.67	87.5
290°	81.82	83.33	87.5	0	55.56	60	75	83.33	73.33	73.68	86.21	100	78.43	74.07	78.57	72.73	66.67	75
300°	90.91	66.67	62.5	62.5	77.78	78.57	0	0	62.5	83.33	75	83.33	100	77.78	71.88	66.67	66.67	66.67
310°	70	55.56	83.33	58.33	57.14	57.14	71.43	0	0	57.14	80	65.22	85.42	100	80	81.25	64.71	64.71
320°	66.67	66.67	57.14	57.14	75	57.14	60	60	60	0	66.67	71.43	77.78	81.63	100	73.81	74.19	95.65
330°	80	69.23	66.67	46.15	60	54.55	83.33	60	60	0	55.56	62.5	68.75	78.79	80.43	100	85.29	82.86
340°	91.67	100	62.5	75	85.71	57.14	0	100	75	0	75	60	62.5	68.42	67.65	73.91	100	92.59
350°	90	54.55	55.56	60	66.67	66.67	60	71.43	71.43	60	75	33.33	66.67	71.43	70.59	72.73	84.44	100

Table 8.6: Match-matrix results of the percentage of matched keypoints, produced from the in-plane rotational data of a **human face** (from 0° at 10° clockwise increments up to 350°), using the 2.5D$_{pc}$ SIFT system.



(a) Match-Matrix (b) ROC

Figure 8.2: (a) The match-matrix results of the percentage of matched keypoints, produced from a set of in-plane rotational data of a **human face** (from 0° at 10° increments in the clockwise direction up to 350°), using the 2.5D$_{pc}$ SIFT system. (b) The matching results presented in a ROC space. This figure can be compared directly with Figure 6.7, where the results were obtained from the same data using the 2.5D SIFT.

The matching results obtained by employing the 2.5D$_{pc}$ SIFT system on in-plane range images are comparable to the matching results obtained by using the original 2.5D SIFT as shown in Table 6.3 of Chapter 6. Moreover, when the results are plotted in a ROC space, the majority of the results are above the line of no discrimination, indicating the obtained matching results are reliable.

8.3.1.2 Out-of-plane Rotations on Real Data

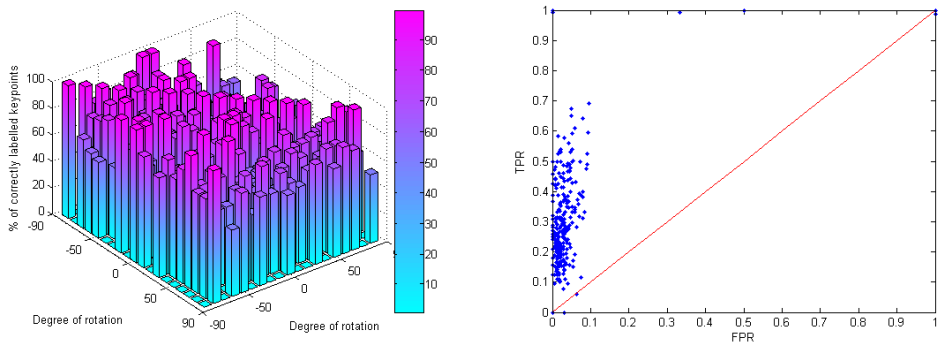
This section presents the results obtained by comparing all pairwise combination of the feature descriptors extracted from the out-of-plane rotated images of a mannequin head and a human face respectively, using the 2.5D$_{pc}$ SIFT system. The images are captured from -90° to 90° at 10° intervals, i.e. full lateral, using the methodology described in Section 6.2.1.2. Table 8.7 summarises the results obtained from this investigation. Statistically significant improvements from the baseline (2.5D SIFT) are shown using one of these five symbols to denote significance: $\ll$, $<$, $=$, $>$, $\gg$. The test of significance was conducted using Wilcoxon Matched-Paired Signed-Ranked test.

System	Subject	Pairings	Average Matching Rate	Images Achieving Matching Rate > 50%	60%	Match-Matrix	ROC
2.5D$_{pc}$	Mannequin head	361	54.9%$^{\gg}$	72.7%	55.7%	Table 8.8 Figure 8.3	Figure 8.3
	Human face	361	37.0%$^{=}$	48.5%	35.5%	Table 8.9 Figure 8.4	Figure 8.4

Table 8.7: Summary of results obtained by exploring all pairwise combinations of the feature descriptors extracted from captured out-of-plane range images of a mannequin head and a human face, using the 2.5D$_{pc}$ SIFT system.

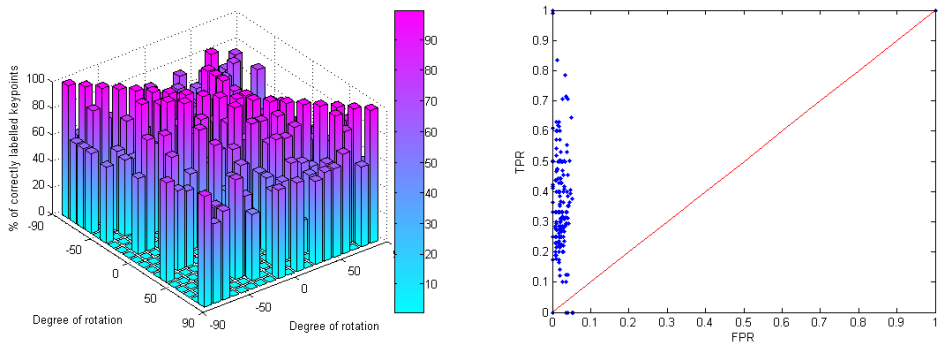
	-90°	**-80°**	**-70°**	**-60°**	**-50°**	**-40°**	**30°**	**-20°**	**-10°**	**0°**	**10°**	**20°**	**30°**	**40°**	**50°**	**60°**	**70°**	**80°**	**90°**
-90°	100	0	62.5	57.1	57.1	0	71.4	62.5	100	100	0	0	0	0	0	0	0	0	0
-80°	0	100	85	91.3	69.6	85.7	94.4	75	23.1	62.5	0	0	0	0	0	66.7	0	60	0
-70°	66.7	80	100	77.8	81.3	100	69.6	60	92.9	61.5	53.3	75	75	60	100	66.7	40	0	0
-60°	60	81.8	90.9	99.4	75	92.9	84.2	56.3	82.4	25	100	0	0	0	60	60	0	0	0
-50°	57.1	85.7	77.8	94.7	68.5	79.2	87.1	100	76.9	66.7	60	0	60	57.1	62.5	0	0	0	0
-40°	0	53.8	100	73.7	82.8	100	88.4	82.8	100	100	75	12.5	71.4	0	54.5	57.1	60	60	0
-30°	0	71.4	66.7	53.8	58.3	82.4	100	78.6	94.4	75	66.7	66.7	57.1	0	62.5	66.7	83.3	75	0
-20°	100	75	80	60	93.5	76.9	86.8	60.4	83.3	61.5	68.8	87.5	54.5	62.5	72.7	0	0	0	0
-10°	0	100	62.5	0	61.5	76.5	67.7	88.6	100	70.4	72.2	66.7	69.2	0	70	87.5	0	0	66.7
0°	100	100	57.1	100	28.6	88.9	83.3	70	71.4	100	76	60	77.8	70	57.1	55.6	0	60	0
10°	83.3	0	0	60	55.6	57.1	69.2	63.6	87.5	82.4	100	78.1	72	60	70	61.5	0	0	0
20°	0	57.1	0	62.5	57.1	0	70	77.8	54.5	95.2	89.7	100	77.8	75	57.1	75	0	0	0
30°	75	85.7	75	0	57.1	0	61.5	100	100	89.5	82.4	71.4	100	81.3	57.1	0	66.7	0	0
40°	0	75	0	0	100	60	87.5	100	53.3	58.3	58.3	60	66.7	100	55.6	71.4	55.6	0	0
50°	0	85.7	66.7	66.7	75	60	54.5	81.3	59.1	61.5	100	63.6	81.3	68.8	100	76	77.8	63.6	0
60°	0	100	75	0	75	100	66.7	60	57.1	60	78.6	88.9	76.5	76.9	79.2	73.3	78.8	62.5	0
70°	0	75	0	57.1	57.1	66.7	75	60	80	55.6	70	78.6	72.7	60	68.4	91.2	100	94.1	60
80°	0	75	0	0	0	66.7	50	83.3	60	72.7	62.5	57.1	66.7	0	77.8	77.8	71.4	99.3	0
90°	0	100	0	50	75	0	66.7	0	0	60	0	0	75	0	66.7	80	75	0	51.5

Table 8.8: Match-matrix results of the percentage of matched keypoints using the 2.5D$_{pc}$ SIFT system, produced from the out-of-plane rotational data of a **mannequin head**.



(a) Match-Matrix (b) ROC

Figure 8.3: 2.5D$_{pc}$ SIFT system: (a) The match-matrix results of the percentage of matched keypoints, produced from a set of out-of-plane rotational data of a **mannequin head** (captured from -90° at 10° increments up to 90°). (b) The matching results presented in a ROC space. This figure can be compared with Figure 6.8, where the results were obtained from the same data set using 2.5D SIFT.



(a) Match-Matrix (b) ROC

Figure 8.4: 2.5D$_{pc}$ SIFT system:(a) The match-matrix results of the percentage of matched keypoints, produced from a set of out-of-plane rotational data of a **human face**. (b) The matching results presented in a ROC space. This figure can be compared with Figure 6.9, where the results were obtained from the same data set using 2.5D SIFT.

	-90°	**-80°**	**-70°**	**-60°**	**-50°**	**-40°**	**30°**	**-20°**	**-10°**	**0°**	**10°**	**20°**	**30°**	**40°**	**50°**	**60°**	**70°**	**80°**	**90°**
-90∘	100	60	75	66.7	0	0	55.6	0	66.7	66.7	62.5	60	75	0	0	57.1	0	66.7	75
-80∘	60	100	78.6	62.5	0	0	0	0	80	20	57.1	71.4	40	66.7	75	0	66.7	75	0
-70∘	60	85.7	100	75	55.6	66.7	60	75	75	0	0	0	0	0	0	0	0	0	0
-60∘	60	0	0	100	71.4	75	0	0	0	57.1	55.6	66.7	0	100	0	0	0	0	75
-50∘	0	0	60	80	100	81.8	72.7	83.3	0	0	71.4	0	0	0	0	0	0	0	0
-40∘	57.1	0	57.1	55.6	75	100	92.9	75	0	60	0	100	0	57.1	0	0	0	0	0
-30∘	0	0	0	100	66.7	94.4	100	93.3	66.7	60	57.1	100	0	83.3	60	0	0	0	0
-20∘	0	0	0	0	0	100	80	100	78.9	90	54.5	100	66.7	57.1	0	0	60	0	0
-10∘	0	57.1	83.3	0	66.7	0	100	75	100	78.6	75	0	0	66.7	0	0	0	0	0
0∘	0	0	0	87.5	0	0	0	88.9	60	100	73.3	66.7	90	75	0	0	75	0	0
10∘	0	0	0	75	0	57.1	87.5	60	66.7	84.6	100	100	80	60	0	0	0	0	0
20∘	0	0	0	25	0	0	0	0	40	100	73.3	100	87	93.3	90	0	0	0	60
30∘	0	0	60	57.1	0	0	83.3	0	75	60	88.9	88	100	72.2	81.8	0	57.1	0	0
40∘	75	0	0	0	0	0	0	60	66.7	0	55.6	75	92.9	100	75	0	62.5	0	0
50∘	0	0	0	0	0	0	0	0	0	0	55.6	81.8	80	85.7	100	71.4	60	0	0
60∘	0	0	0	0	0	0	0	0	0	0	0	0	0	62.5	57.1	100	71.4	75	75
70∘	0	0	0	0	0	75	60	0	0	55.6	60	71.4	25	57.1	0	60	100	87.5	0
80∘	0	66.7	66.7	0	0	0	50	0	0	0	0	0	0	0	66.7	100	85.7	100	57.1
90∘	83.3	60	66.7	0	0	0	0	0	66.7	0	66.7	0	62.5	66.7	66.7	75	0	60	100

Table 8.9: Match-matrix results of the percentage of matched keypoints using the 2.5D$_{pc}$ SIFT system, produced from the out-of-plane rotational data of a **human face** (from -90° to 90° of rotations).

Comparing the results obtained using the original 2.5D SIFT (summarised in Table 6.7 in Chapter 6) and the results obtained using the 2.5D$_{pc}$ SIFT system, it is noted that the average percentages of matched keypoints for the out-of-plane rotated range images of a mannequin head has increased by approximately 25% at 50% keypoint matching rate. The average percentages of the matched and filtered keypoints of the out-of-plane rotated images of a human face has decreased by approximately 2%. However, examining the results plotted in a ROC space reveals that the matches are more stable and reliable when using the extended version of the 2.5D$_{pc}$ SIFT, with a smaller FPR.

8.3.1.3 Synthetic Out-of-plane Rotations

This section presents the results obtained by matching synthetically rotated range images, where the images are rotated about the yaw and pitch axis ($\pm 40°$)respectively, at 10° intervals.

Table 8.10 summaries the results obtained by exploring all pairwise combinations of the feature descriptors extracted from the synthetically rotated range images of a mannequin head and a human face respectively, about the yaw and pitch axes.

The results obtained from synthetically rotated range images about both the yaw and pitch axis shows that the feature descriptors extracted from the 2.5D$_{pc}$ SIFT

Subject	Axis	Pairings	Average Matching Rate	Images Achieving Matching Rate > 50%	60%	70%	Match-Matrix	ROC
Mannequin head	yaw	81	80.0%	100%	96.3%	82.7%	Table 8.11 Figure 8.5	Figure 8.5
Human head	yaw	81	81.2%	100%	95.1%	81.5%	Table 8.12 Figure 8.6	Figure 8.6
Mannequin head	pitch	81	83.7%	100%	96.3%	86.4%	Table 8.13 Figure 8.7	Figure 8.7
Human head	pitch	81	84.2%	100%	98.8%	92.6%	Table 8.14 Figure 8.8	Figure 8.8

Table 8.10: Results obtained by exploring all pairwise combinations of the feature descriptors extracted using the 2.5D$_{pc}$ SIFT system on synthetically rotated out-of-plane range images of a mannequin head and a human face.

	-40°	**30°**	**-20°**	**-10°**	**0°**	**10°**	**20°**	**30°**	**40°**
-40°	100	82.1	76.9	75	65	58.8	62.5	100	91.7
-30°	71.4	99.1	63.6	78.9	67.9	79.3	69.2	73.7	95.7
-20°	66.7	65.9	100	84.4	74.4	86.8	73.1	78.6	80
-10°	75	77.1	82.2	100	75	78.4	89.7	68.4	71.4
0°	86.4	71	70.7	77.1	97.1	86	61.3	70.8	86.7
10°	73.7	80.8	78.8	79.5	84.6	100	76.9	64.3	65.6
20°	81.3	73.9	79.2	80.8	82.1	85.3	98.4	85.3	80.8
30°	91.7	91.7	77.3	60	83.3	75.6	86.7	98.5	73.7
40°	100	95.2	75	57.1	76.5	81.3	73.1	87.5	99.3

Table 8.11: Match-matrix results of the percentage of matched keypoints, produced from the synthetically rotated out-of-plane (about yaw axis) data of a **mannequin head**, from -40° to 40° of rotations, using the 2.5D$_{pc}$ SIFT system.

system show excellent invariance to Euler's out-of-plane rotational changes. The next section will address the scale invariant properties of the 2.5D SIFT.

	-40°	**30°**	**-20°**	**-10°**	**0°**	**10°**	**20°**	**30°**	**40°**
-40°	100	75.7	70.5	67.9	78.3	68	73.9	100	100
-30°	69	99.3	79.7	76.8	72.2	94.5	70.7	100	85.7
-20°	77.1	76.6	96.8	84.1	84.9	73.7	65.8	100	63.2
-10°	79.3	86.8	80.6	97.7	88.2	70.2	73.3	88.9	100
0°	75	70	85.2	79.8	99.2	77.3	69.1	70.8	92.3
10°	68.4	71.4	72.1	80	88.5	98.1	73.3	80.8	64.5
20°	61.5	73.9	54.1	72.2	62.5	80.4	98.7	89.3	88.9
30°	100	91.7	92	96.8	95.8	84.2	60	98.6	71.2
40°	55.6	53.8	85.7	91.7	91.7	82.9	84.6	78.7	98.7

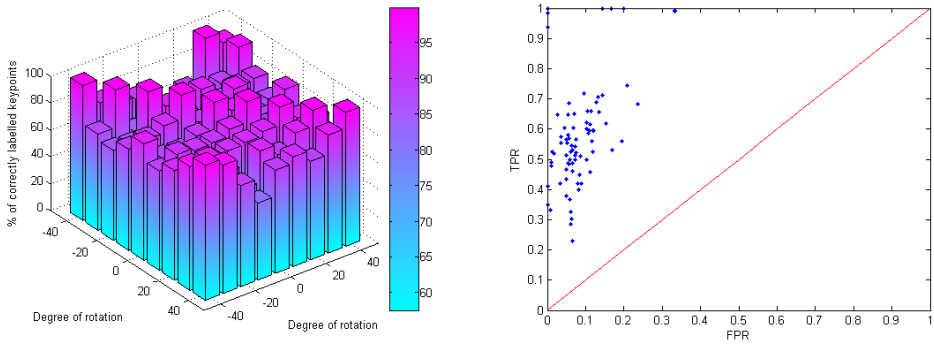
Table 8.12: Match-matrix results of the percentage of matched keypoints, produced from the synthetically rotated out-of-plane (about yaw axis) data of a **human face**, from -40° to 40° of rotations, using the 2.5D$_{pc}$ SIFT system.

	-40°	**30°**	**-20°**	**-10°**	**0°**	**10°**	**20°**	**30°**	**40°**
-40°	99.2	66.7	67.7	81	62.5	100	100	91.7	84.2
-30°	70.4	97.3	83.9	91.3	68.8	72.7	80	73.3	68.8
-20°	76.9	84.8	100	93.3	80	75.9	83.3	77.3	81
-10°	86.4	92	90.5	100	75.8	87.8	89.3	100	70
0°	77.8	57.1	75	76.5	97.1	76.7	90.5	92.3	57.1
10°	62.5	94.7	88.5	80.5	85.7	98.3	89.1	95	80.6
20°	92.3	73.3	80.8	86.2	77.8	91.1	98.1	91.7	78.4
30°	100	78.6	81.8	82.6	75.8	90.9	84.8	99.1	68
40°	100	55.6	94.4	78.9	84.2	96.4	82.4	79.1	97.6

Table 8.13: Match-matrix results of the percentage of matched keypoints, produced from the synthetically rotated out-of-plane (about pitch axis) data of a **mannequin head**, from -40° to 40° of rotations, using the 2.5D$_{pc}$ SIFT system.

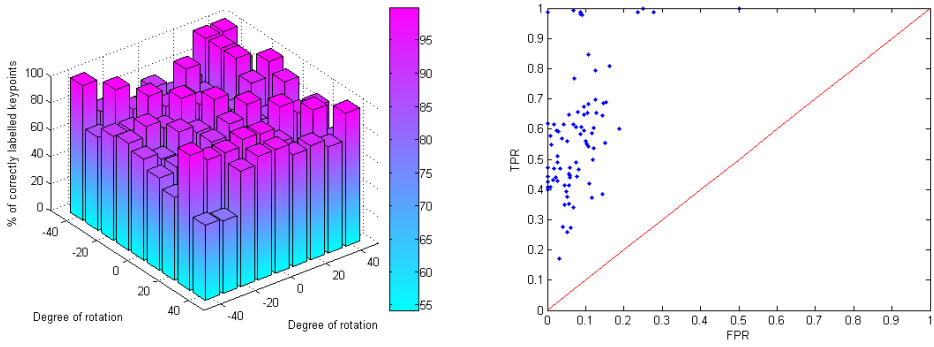
	-40°	**30°**	**-20°**	**-10°**	**0°**	**10°**	**20°**	**30°**	**40°**
-40°	98.7	75	74.3	87	94.1	87.5	77.3	60	77.3
-30°	85.1	98.2	90.7	86.7	83.3	85.7	77.4	80	88.9
-20°	68.3	79.7	98.8	84.1	86.7	77.8	75.7	79.3	86.4
-10°	88.5	76.4	78.8	97.9	87.5	86.3	76.2	82.1	85.7
0°	80	77.1	93.9	95.1	99.2	74.4	89.1	91.4	71
10°	81.3	86.2	90.3	90.2	81.5	94.3	74.5	86.7	88.9
20°	63.2	82.9	90.3	86.4	81.4	79.3	98.6	88.1	87.2
30°	68.4	72.2	92	88.9	89.5	87.5	84.4	99.3	83.7
40°	100	72.2	94.1	64.3	81.6	88.6	83	67.3	95.2

Table 8.14: Match-matrix results of the percentage of matched keypoints, produced from the synthetically rotated out-of-plane (about pitch axis) data of a **human face**, from -40° to 40° of rotations, using the 2.5D$_{pc}$ SIFT system.



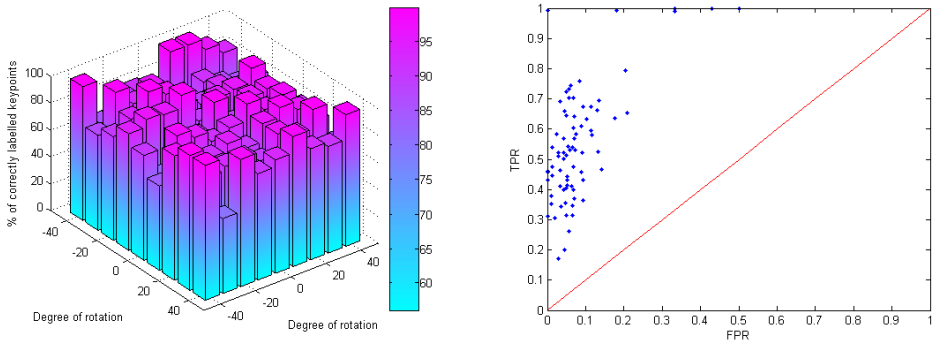
(a) Match-Matrix (b) ROC

Figure 8.5: 2.5D$_{pc}$ SIFT: (a) The match-matrix of matched keypoints, produced from a set of synthetically rotated out-of-plane (about the yaw axis) range images of a **mannequin head**. (b) The matching results presented in a ROC space.



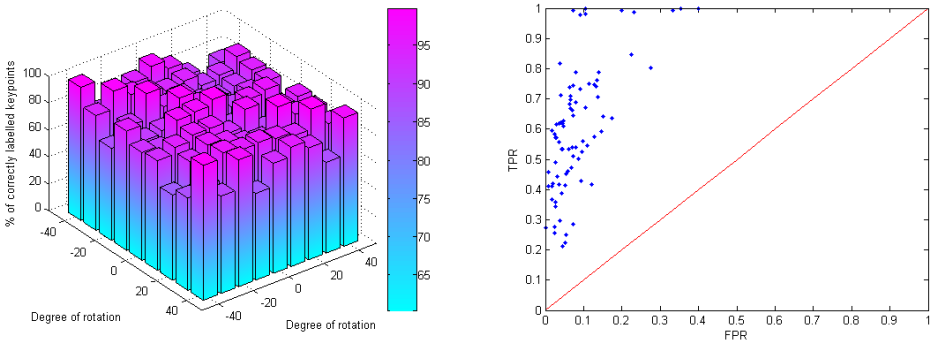
(a) Match-Matrix (b) ROC

Figure 8.6: 2.5D$_{pc}$ SIFT: (a) The match-matrix of matched keypoints, produced from a set of synthetically rotated out-of-plane (about the yaw axis) range images of a **human face**. (b) The matching results presented in a ROC space.



(a) Match-Matrix (b) ROC

Figure 8.7: 2.5D$_{pc}$ SIFT: (a) The match-matrix of matched keypoints, produced from a set of synthetically rotated out-of-plane (about the pitch axis) range images of a **mannequin head**. (b) The matching results presented in a ROC space.



(a) Match-Matrix (b) ROC

Figure 8.8: 2.5D$_{pc}$ SIFT: (a) The match-matrix of matched keypoints, produced from a set of synthetically rotated out-of-plane (about the pitch axis) range images of a **human face**. (b) The matching results presented in a ROC space.

8.3.2 Scale Invariant 2.5D$_{pc}$ SIFT

This section presents the results obtained by extracting feature descriptors from range images of different sizes using the 2.5D$_{pc}$ SIFT. By using the feature descriptors extracted from the baseline image of size 244×369 pixels and comparing to the feature descriptors extracted from the remainder set of data (i.e. of the same viewpoints but of different sizes), a match-matrix that shows the percentage of matched keypoints at each scale can be formulated. The match-matrix can be plotted in a bar graph, illustrating the results visually. The lower left axis of the graph represents the viewpoint angles of each pair of the compared range images, while the lower right axis shows the size of the range images used for the comparison (where the numbers at the axis correspond to the level indicator in Table 8.15.). Similar to the investigation conducted on the single-scale 2.5D SIFT feature descriptors, the results can also be presented in terms of a ROC space, providing a visual aid as to the reliability of the matches.

A total of 360 range images of different sizes simulating in-plane rotations (2×180), 190 range images of captured out-of-plane images of different sizes (2×95) and 180 range images of different sizes simulating out-of-plane rotations (2×90) rotated images about the pitch axis and (2×45) rotated images about the yaw axis are used in this experiment.

Levels	**Size of images**
1	494×744
2	370×557
3	**244×369 (Baseline)**
4	182×275
5	119×181

Table 8.15: Size of the range images (in pixels) used in this experiment in order to determine the stability of the feature descriptors against scale issues.

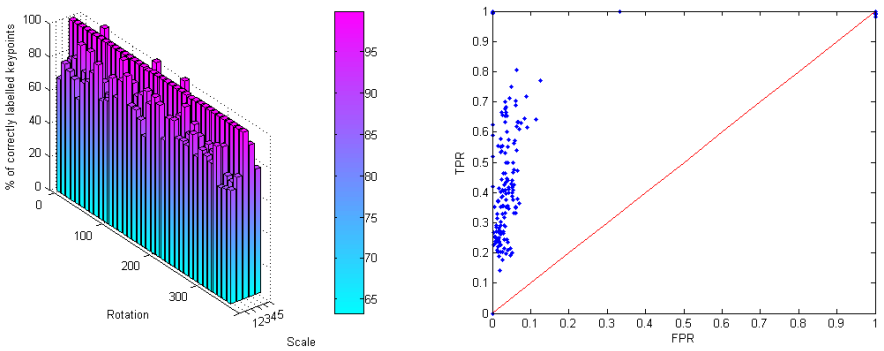
8.3.2.1 Synthetic In-plane Rotations

This section presents the results obtained by comparing the feature descriptors extracted from the unrotated and in-plane rotated images of size 244×369 pixels and the feature descriptors extracted from the range images of different sizes (but of the same rotational changes), using the 2.5D$_{pc}$ SIFT system.

Table 8.16 summaries the results obtained by exploring all pairwise combinations of the feature descriptors extracted from the baseline images (of size 249x336 pixels and at different angles about the roll axis) to the feature descriptors extracted from the range images of the same viewpoint angle but at a different scale.

System	Subject	Pairings	Average Matching Rate	Images Achieving Matching Rate >			Match-Matrix	ROC
				70%	80%	90%		
2.5D$_{pc}$	Mannequin head	185	82.7%	87.0%	58.9%	34.6%	Table 8.17 Figure 8.9	Figure 8.9
	Human head	185	85.0%	96.2%	93.5%	67.6%	Table 8.18 Figure 8.10	Figure 8.10

Table 8.16: Results obtained by matching the feature descriptors extracted from different sized synthetically rotated in-plane range images of a mannequin head and a human face and feature descriptors extracted from the baseline images, using the 2.5D$_{pc}$ SIFT system.



(a) Match-matrix (b) ROC

Figure 8.9: (a) The match-matrix results of the percentage of matched keypoints, produced by comparing the feature descriptors extracted from a set of in-plane rotational data of a **mannequin head** of size 244×369 pixels and the range images captured at a different scale, using the 2.5D$_{pc}$ SIFT system. (b) The matching results illustrated in ROC space.

The results obtained illustrate the feature descriptors extracted using the 2.5D$_{pc}$ SIFT show good invariance to scale changes.

8.3.2.2 Out-of-plane Rotations on Real Data

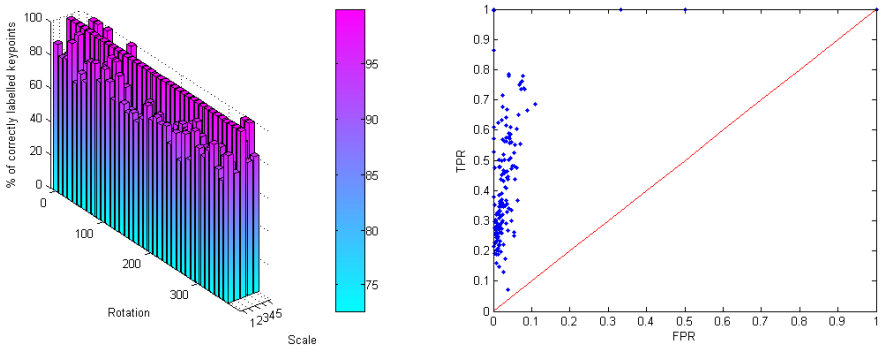
The results obtained by exploring the feature descriptors extracted from out-of-plane rotated range images of a human face and a mannequin head of size 244×369 pixels and

	494×744	**370×557**	**244×369**	**182×275**	**119×181**
0°	67.9	74.4	100	90.9	75
10°	78.9	80	100	83.9	78.6
20°	76.2	82.5	100	75	85.7
30°	73.3	92.3	100	83.8	90
40°	63	77.8	100	80.6	100
50°	75	92.1	100	82.9	75
60°	77.3	76.7	100	90.3	90.9
70°	84.4	78.6	100	80	83.3
80°	91.3	79.3	100	75	0
90°	71.4	73.7	100	86.7	66.7
100°	73.5	72	100	85.7	0
110°	88.9	76.9	100	94.4	66.7
120°	78.9	100	100	91.7	57.1
130°	90.9	81.5	100	85	66.7
140°	95.7	90	100	93.8	84.6
150°	79.3	83.9	100	71.4	100
160°	79.1	86.7	99.4	73	80
170°	75	90.6	100	90.9	83.3
180°	66.7	67.6	100	80	91.7
190°	88	73.7	100	90.9	88.9
200°	94.1	77.4	100	88.9	75
210°	91.3	67.7	100	80.6	100
220°	72.2	78.9	100	85.2	85.7
230°	86.7	78.8	100	89.3	57.1
240°	87.5	88.9	100	92.6	60
250°	78.3	96.2	99.4	77.3	60
260°	78.9	81.5	100	86.4	75
270°	77.8	89.3	100	83.3	66.7
280°	86.7	84	100	90.9	0
290°	75	73.3	100	91.7	75
300°	80.8	82.9	100	91.2	0
310°	79.2	80.6	100	83.9	80
320°	76.9	80.8	100	88.2	80
330°	88.9	87.1	100	80.8	57.1
340°	65.5	71.4	100	86.1	70
350°	66.7	69.7	100	75	57.1
360°	67.9	74.4	100	90.9	75

Table 8.17: Match-matrix results of the percentage of matched keypoints by comparing the feature descriptors obtained from the in-plane rotational data of a **mannequin head** (from 0° at 10° clockwise increments up to and including 360°) of size 244×369 pixels and the feature descriptors extracted from the same viewpoint but of different scales, using the 2.5D$_{pc}$ SIFT system.

	494×744	**370×557**	**244×369**	**182×275**	**119×181**
0°	88.5	76.4	100	80	81.8
10°	82.1	82.8	100	91.2	78.6
20°	83.3	87.5	100	82.1	100
30°	94.7	97.9	100	92.1	100
40°	72.4	76.7	100	93	83.3
50°	80	84.1	100	87.1	100
60°	76.7	83.3	99.6	76.9	58.3
70°	90.6	78.6	100	88.6	66.7
80°	91.7	74.5	100	85	66.7
90°	82.6	91.2	100	79.4	92.3
100°	91.9	97	100	90.3	100
110°	86.2	96.8	100	81.3	75
120°	77.1	77.1	100	84.6	57.1
130°	93.5	85.4	100	88.1	77.8
140°	78.7	73.3	100	85.2	72.7
150°	75	71.8	100	79.6	72.2
160°	76.9	72.7	100	83	71.4
170°	73.3	72.9	100	91.1	90.9
180°	76.9	85.4	99.6	81.8	83.3
190°	86.7	71.4	99.6	82.8	73.3
200°	80	81.4	100	88.9	91.7
210°	83.3	75.5	100	80.6	80
220°	81	76.3	100	84.8	63.6
230°	82.1	71.7	100	72.7	72.7
240°	73	82.9	100	85.3	75
250°	81.5	85.4	100	90.6	25
260°	66.7	83.8	100	90.3	66.7
270°	84.6	89.2	100	90.3	80
280°	71	92.3	100	85.3	55.6
290°	90	85.3	100	86.8	60
300°	81.8	80.5	100	88.9	83.3
310°	78.3	82	100	86.5	90.9
320°	84	74.6	100	83	60
330°	89.7	86.5	99.6	83.7	55.6
340°	76.5	82.2	100	74	100
350°	71.4	80.4	100	90	100
360°	88.5	76.4	100	80	81.8

Table 8.18: Match-matrix results of the percentage of matched keypoints by comparing the feature descriptors obtained from the in-plane rotational data of a **human face** (from 0° at 10° clockwise increments up to and including 360°) of size 244×369 pixels and the feature descriptors extracted from the same viewpoint but of different scales, using the 2.5D$_{pc}$ SIFT system.



(a) Match-matrix (b) ROC

Figure 8.10: (a) The match-matrix results of the percentage of matched keypoints, produced by comparing the feature descriptors extracted from a set of in-plane rotational data of a **human face** (from 0° at 10° increments in the clockwise direction up to 350°) of size 244×369 pixels and the range images captured at a different scale, using the 2.5D$_{pc}$ SIFT system. (b) The matching results illustrated in ROC space.

the feature descriptors extracted from the same viewpoint rotation but of a different sized images are presented in this section. Table 8.19 summaries the results.

System	Subject	Pairings	Average Matching Rate	Images Achieving Matching Rate > 70%	80%	90%	Match-Matrix	ROC
2.5D$_{pc}$	Mannequin head	95	83.2%	80%	65.3%	42.1%	Table 8.20 Figure 8.11	Figure 8.11
	Human head	95	87.8%	91.6%	77.9%	53.7%	Table 8.21 Figure 8.12	Figure 8.12

Table 8.19: Results obtained by matching the feature descriptors extracted from different sized rotated out-of-plane range images of a mannequin head and a human face to their baseline images, using the 2.5D$_{pc}$ SIFT system.

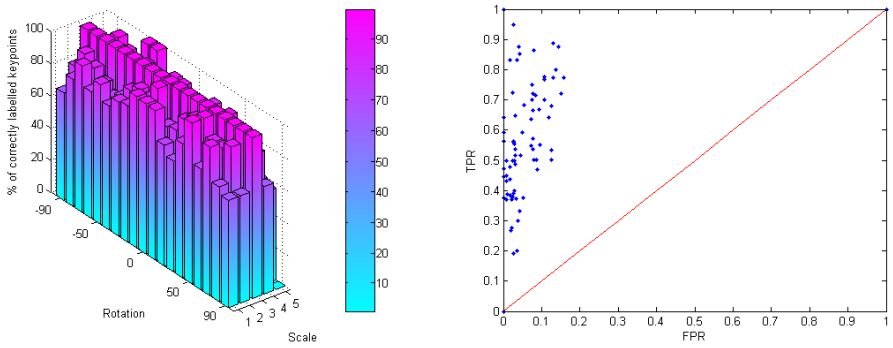
The results obtained in this section indicate that the feature descriptors demonstrate good invariance to scale changes, with respect to out-of-plane rotational changes.

	494×744	**370×557**	**244×369**	**182×275**	**119×181**
-90°	66.7	80	100	70	0
-80°	78.9	91.7	100	88.5	0
-70°	90.9	86.7	100	100	60
-60°	78.6	81	100	94.7	57.1
-50°	86.7	75.9	100	85.7	100
-40°	76.9	79.3	100	93.8	100
-30°	83.3	87.5	100	85	69.2
-20°	84.6	91.7	100	82.6	70
-10°	94.1	90	100	96.2	77.8
0°	93.3	100	100	95.5	57.1
10°	92.9	80	100	85.7	81.8
20°	75	81.5	100	78.3	88.2
30°	68.8	91.3	100	87.5	77.8
40°	83.3	95.8	100	84.6	75
50°	100	100	100	95.8	70
60°	75	88.9	100	85	66.7
70°	92.3	88.9	100	66.7	60
80°	71.4	93.3	100	83.3	57.1
90°	66.7	66.7	100	66.7	0

Table 8.20: Match-matrix results of the percentage of matched keypoints by comparing the feature descriptors obtained from the out-of-plane rotational data of a **mannequin head** (from -90° at 10° clockwise increments up to and including 90°) of size 244×369 pixels and the feature descriptors extracted from the same viewpoint but of different scales, using the 2.5D$_{pc}$ SIFT system.

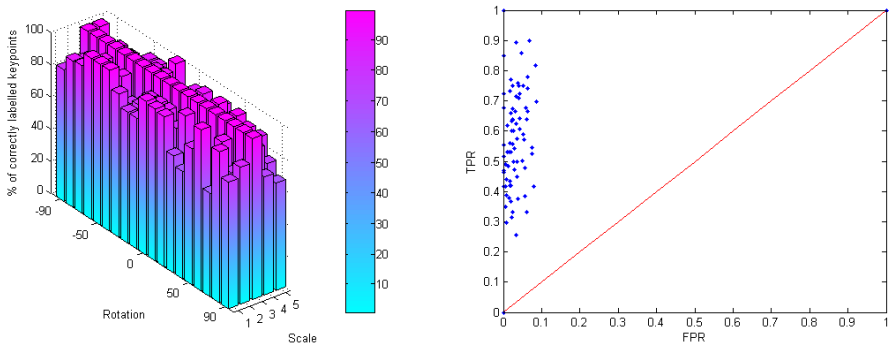
	494×744	**370×557**	**244×369**	**182×275**	**119×181**
-90°	81.8	75	100	100	81.8
-80°	90.9	88.9	100	81.8	0
-70°	91.7	87.5	100	94.1	85.7
-60°	100	95.5	100	92.6	66.7
-50°	100	90	100	87	90.9
-40°	100	93.1	100	91.7	90.9
-30°	88.9	80	100	95.2	75
-20°	81.3	82.6	100	95.5	100
-10°	81.3	90.6	100	89.3	75
0°	95.5	90.9	100	87.5	90
10°	94.4	90.6	100	89.3	92.3
20°	93.3	83.3	100	85.2	55.6
30°	72.7	94.7	100	90.5	92.3
40°	66.7	76.7	100	100	75
50°	84.6	75	100	87.5	60
60°	100	76.9	100	100	66.7
70°	64.3	93.3	100	95.7	87.5
80°	93.8	77.8	100	87	75
90°	78.6	85	100	73.3	66.7

Table 8.21: Match-matrix results of the percentage of matched keypoints by comparing the feature descriptors obtained from the out-of-plane rotational data of a **human face** (from -90° at 10° clockwise increments up to and including 90°) of size 244×369 pixels and the feature descriptors extracted from the same viewpoint but of different scales, using the 2.5D$_{pc}$ SIFT system.



(a) Match-matrix (b) ROC

Figure 8.11: 2.5D$_{pc}$ SIFT: (a) The match-matrix results of the percentage of matched keypoints, produced by comparing the feature descriptors extracted from a set of out-of-plane rotational data of a **mannequin head** (from -90° at 10° increments in the clockwise direction up to 90°) of size 244×369 pixels and the range images captured at a different scale. (b) The matching results illustrated in ROC space.



(a) Match-matrix (b) ROC

Figure 8.12: 2.5D$_{pc}$ SIFT: (a) The match-matrix results of the percentage of matched keypoints, produced by comparing the feature descriptors extracted from a set of out-of-plane rotational data of a **human face** (from -90° at 10° increments in the clockwise direction up to 90°) of size 244×369 pixels and the range images captured at a different scale. (b) The matching results illustrated in ROC space.

The next section will presents the results obtained by comparing feature descriptors extracted from different sizes of range images simulating synthetically about the out-of-plane axes.

8.3.2.3 Synthetic Out-of-plane Rotations

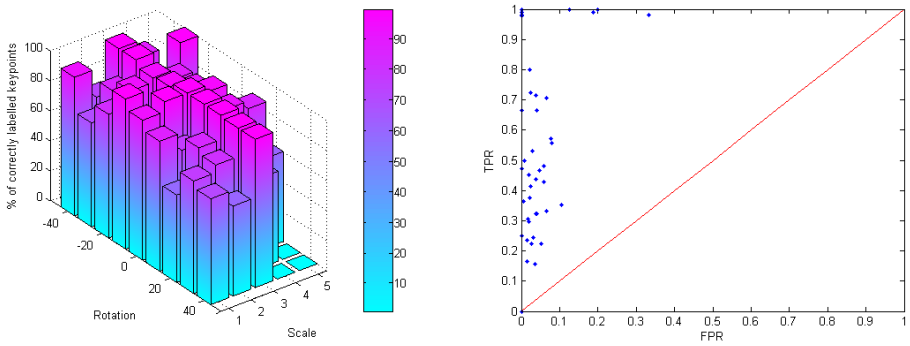
This section investigates the matching performances of synthetically rotated range images of different sizes against out-of-plane rotations. Here, the feature descriptors extracted from the baseline images (of size 244×369 pixels) are matched to the feature descriptors extracted from images of the same viewpoint angle but of a different size. Table 8.22 summarises the results.

Subject	**Axis**	**Pairings**	**Average Matching Rate**	**Images Achieving Matching Rate >**			**Match-Matrix**	**ROC**
				70%	**80%**	**90%**		
Mannequin head	yaw	45	76.6%	75.6%	55.6%	37.8%	Table 8.23 Figure 8.13	Figure 8.13
Human head	yaw	45	74.9%	68.9%	44.4%	33.3%	Table 8.24 Figure 8.14	Figure 8.14
Mannequin head	pitch	45	85.0%	77.8%	68.9%	46.7%	Table 8.25 Figure 8.15	Figure 8.15
Human head	pitch	45	77.4%	75.6%	60%	33.3%	Table 8.26 Figure 8.16	Figure 8.16

Table 8.22: Results obtained by matching the feature descriptors extracted from different sized synthetically rotated out-of-plane range images of a mannequin head and a human face and the feature descriptors extracted from the baseline images, using the 2.5D$_{pc}$ SIFT system.

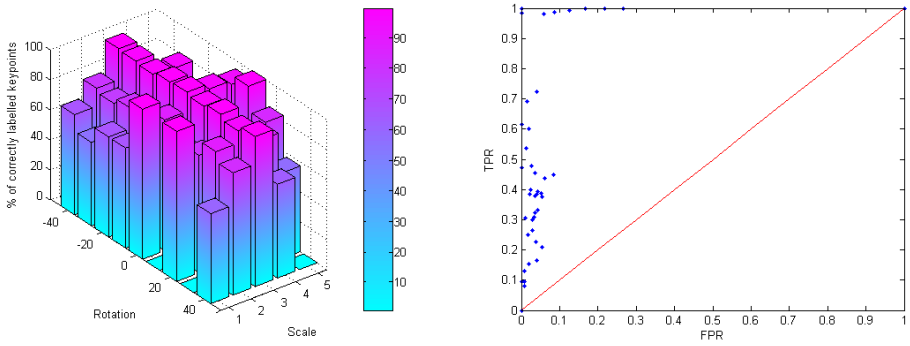
	494×744	**370×557**	**244×369**	**182×275**	**119×181**
-40°	92.3	73.3	100	88.9	0
-30°	69.2	84.6	100	80	100
-20°	82.4	92.3	95.3	88.9	60
-10°	100	70.6	96.7	89.5	71.4
0°	93.3	100	100	95.5	57.1
10°	86.7	75	100	90.9	85.7
20°	58.3	84	100	80	60
30°	75	80	98.9	60	0
40°	70.6	60	98.9	0	0

Table 8.23: Match-matrix results of the percentage of matched keypoints by comparing the feature descriptors obtained from synthetically out-of-plane rotational data (about the yaw axis) of a **mannequin head** (from -40° at 10° clockwise increments up to and including 40°) of sized 244×369 pixels and the feature descriptors extracted from the same viewpoint but of different scales, using the 2.5D$_{pc}$ SIFT system.



(a) Match-matrix (b) ROC

Figure 8.13: 2.5D$_{pc}$ SIFT system: (a) The match-matrix results of the percentage of matched keypoints, produced by comparing the feature descriptors extracted from a set of synthetically out-of-plane rotational data (about the yaw axis) of a **mannequin head** (from -40° at 10° increments in the clockwise direction up to 40°) of size 244×369 pixels and the range images captured at a different scale. (b) The matching results illustrated in ROC space.



(a) Match-matrix (b) ROC

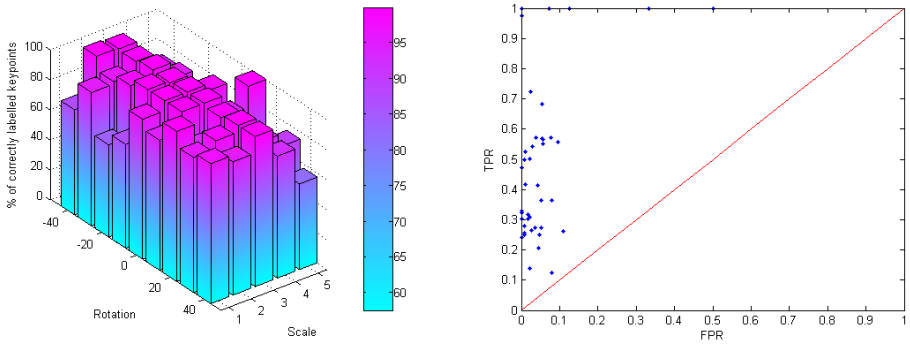
Figure 8.14: 2.5D$_{pc}$ SIFT system: (a) The match-matrix results of the percentage of matched keypoints, produced by comparing the feature descriptors extracted from a set of synthetically out-of-plane rotational data (about the yaw axis) of a **human face** (from -40° at 10° increments in the clockwise direction up to 40°) of size 244×369 pixels and the range images captured at a different scale. (b) The matching results illustrated in ROC space.

	494×744	**370×557**	**244×369**	**182×275**	**119×181**
-40°	66.7	78.6	99.3	76.9	0
-30°	55.6	76	98.4	75	60
-20°	66.7	80	97.1	95.7	71.4
-10°	66.7	79.3	99.3	78.9	60
0°	100	87.2	100	96.4	93.8
10°	0	73.7	98	70	100
20°	100	78.3	100	78.9	83.3
30°	0	88.2	96.4	84.6	57.1
40°	60	81.5	100	62.5	0

Table 8.24: Match-matrix results of the percentage of the matched keypoints by comparing the feature descriptors obtained from synthetically out-of-plane rotational data (about the yaw axis) of a **human face** (from -40° at 10° clockwise increments up to and including 40°) of size 244×369 pixels and the feature descriptors extracted from the same viewpoint but of different scales, using the 2.5D$_{pc}$ SIFT system.

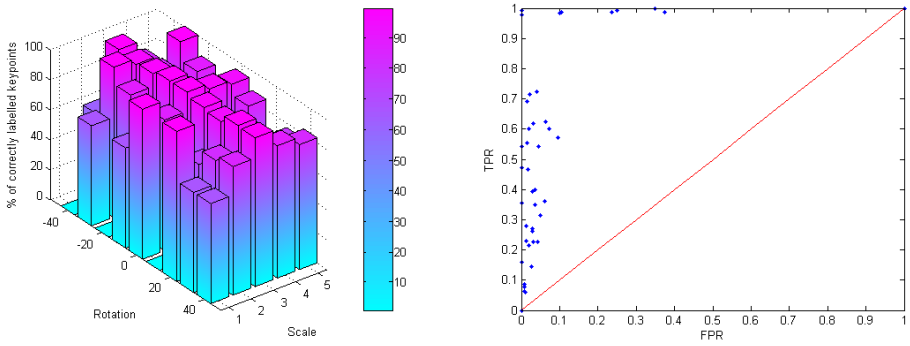
	494×744	**370×557**	**244×369**	**182×275**	**119×181**
-40°	70	100	99.2	80	66.7
-30°	88.9	90.5	98.2	86.7	60
-20°	61.5	95.2	100	88.9	60
-10°	69.2	100	98.9	78.6	60
0°	93.3	100	100	95.5	57.1
10°	86.7	100	100	85.7	100
20°	100	83.3	97.8	70.6	40
30°	90	95.5	100	83.3	75
40°	93.3	88.9	100	81.3	57.1

Table 8.25: Match-matrix results of the percentage of matched keypoints by comparing the feature descriptors obtained from synthetically out-of-plane rotational data (about the pitch axis) of a **mannequin head** (from -40° at 10° clockwise increments up to and including 40°) of size 244×369 pixels and the feature descriptors extracted from the same viewpoint but of different scales, using the 2.5D$_{pc}$ SIFT system.



(a) Match-matrix (b) ROC

Figure 8.15: 2.5D$_{pc}$ SIFT: (a) The match-matrix results of the percentage of matched keypoints, produced by comparing the feature descriptors extracted from a set of synthetically out-of-plane rotational data (about the pitch axis) of a **mannequin head** (from -40° at 10° increments in the clockwise direction up to 40°) of size 244×369 pixels and the range images captured at a different scale. (b) The matching results illustrated in ROC space.



(a) Match-matrix (b) ROC

Figure 8.16: 2.5D$_{pc}$ SIFT: (a) The match-matrix results of the percentage of matched keypoints, produced by comparing the feature descriptors extracted from a set of synthetically out-of-plane rotational data (about the pitch axis) of a **human face** (from -40° at 10° increments in the clockwise direction up to 40°) of size 244×369 pixels and the range images captured at a different scale. (b) The matching results illustrated in ROC space.

The results obtained indicates the feature descriptors extracted using the pose corrected version of the 2.5D SIFT show a good invariance to scale changes.

	494×744	**370×557**	**244×369**	**182×275**	**119×181**
-40°	0	61.5	98.7	83.3	0
-30°	66.7	100	96.2	84.2	100
-20°	0	88.9	97.5	83.3	87.5
-10°	66.7	70.6	100	78.3	75
0°	100	87.2	100	96.4	93.8
10°	0	66.7	97.1	76.5	86.7
20°	100	72.2	95.2	82.8	66.7
30°	66.7	82.4	100	74.1	76.9
40°	66.7	85.7	99.2	88.2	83.3

Table 8.26: Match-matrix results of the percentage of matched keypoints by comparing the feature descriptors obtained from synthetically out-of-plane rotational data (about the pitch axis) of a **human face** (from -40° at 10° clockwise increments up to and including 40°) of size 244×369 pixels and the feature descriptors extracted from the same viewpoint but of different scales, using the 2.5D$_{pc}$ SIFT system.

The next section will repeat the experiments conducted in this section using standard 2D SIFT on 2D images. A comparison can then be made on the performance rate of the 2D SIFT and 2.5D SIFT.

8.4 Comparison between 2D and 2.5D$_{pc}$ Frameworks

A comparison between the performance of standard 2D SIFT on 2D images and the performance of the 2.5D$_{pc}$ SIFT on range images is essential in order to validate the 2.5D SIFT system. This section addresses the performance rate of standard 2D SIFT on 2D images against rotational changes. A MATLAB version of standard 2D SIFT is available online (Lowe, 2005) where stable keypoints are located on the 2D images, along with their appropriate scale σ. The output of the standard 2D SIFT comprises the image itself, the (x, y, σ, θ) information and the feature descriptors extracted for each (x, y, σ, θ). Therefore, the resulting $(x, y, \sigma, \theta, \text{descriptor})$ can be stored within a database for each 2D image of different rotational changes.

By comparing all the combinations of the feature descriptors extracted from 2D images captured at different angles using the standard 2D SIFT algorithm, it is possible to determine the performance rate of standard 2D SIFT. The next section presents the results obtained for the performance rate against both in-plane and out-of-plane rotational changes.

8.4.1 2D SIFT on Intensity Images: Fixed Scale

This section investigates the matching performance of standard 2D SIFT on 2D intensity images at a fixed scale. In order to be able to compare the performance rates between the standard 2D SIFT and the proposed 2.5D SIFT, the data and the size of the images used are identical. To this end, the 2D images used for this experiments are of size 244×369 pixels.

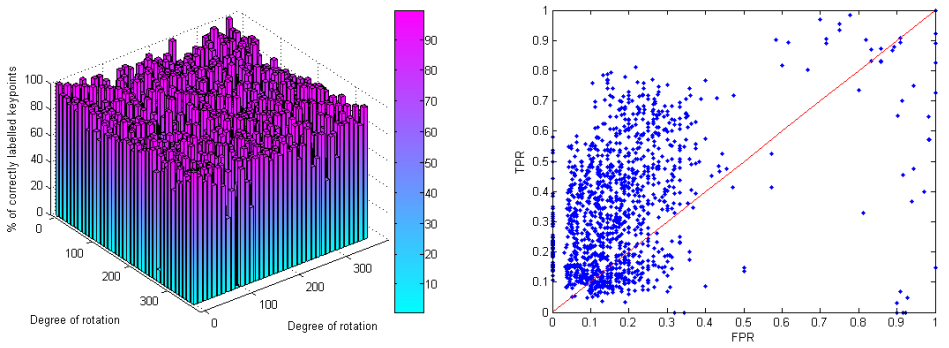
A total of 72 2D images simulating in-plane rotations (2×36), 38 images of captured out-of-plane rotations (2×19) and 36 images simulating the out-of-plane rotations about the yaw (2×9) and pitch (2×9) axes respectively are used in this experiment.

8.4.1.1 Synthetic In-plane Rotations

The performance rate of the feature descriptors extracted from in-plane rotated 2D images using standard 2D SIFT is addressed in this section. Table 8.27 summarises the results obtained by exploring all combinations of the feature descriptors extracted from the synthetically in-plane rotated 2D images of a mannequin head and a human face respectively, from 0° at increments of 10° in the clockwise direction up to 350°. Beside each average matching rate is two symbols, denoting the statistically significant, using the Wilcoxon Matched-Paired Signed-Ranked test, when compared to the two baseline methods (2.5D SIFT and 2.5D$_{pc}$ SIFT respectively). Each symbol can be one of $\ll$, $<$, $=$, $>$, $\gg$. For example, consider the average matching result of the mannequin head using 2D SIFT: 83.7%$^{\gg\gg}$, the first symbol indicates that the average matching result is significantly better ($p < 0.01$) than the average matching result obtained using 2.5D SIFT; and the second symbol indicates that the average matching result is again significantly better ($p < 0.01$) than the average matching rate obtained using the 2.5D$_{pc}$ SIFT.

System	Subject	Pairings	Average Matching Rate	Images Achieving Matching Rate > 50%	60%	Match-Matrix	ROC
2D	Mannequin head	1369	83.7%$^{\gg\gg}$	98.4%	94.1%	Table 8.28 Figure 8.17	Figure 8.17
	Human face	1369	84.2%$^{\gg\gg}$	98.97%	96.4%	Table 8.29 Figure 8.18	Figure 8.18

Table 8.27: Results obtained by exploring all pairwise combinations of the feature descriptors extracted from rotated in-plane 2D images of a mannequin head and a human face, using 2D SIFT.



(a) Match-matrix (b) ROC

Figure 8.17: (a) Match-matrix obtained by exploring all pairwise combinations of the feature descriptors extracted from rotated in-plane 2D images of a **mannequin head**, using standard 2D SIFT. (b) Results presented in a ROC space. This figure can be compared with Figure 6.6 and Figure 8.1, where the results were obtained from the same data set using 2.5D SIFT and 2.5D$_{pc}$ SIFT respectively.

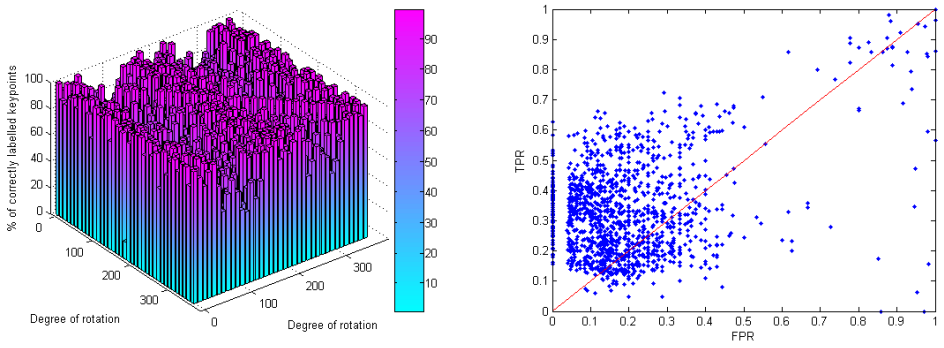
	0°	10°	20°	30°	40°	50°	60°	70°	80°	90°	100°	110°	120°	130°	140°	150°	160°	170°
0°	100	95.7	90.2	93.3	84.8	96.7	81.3	55.6	75	18.2	88.9	77.8	57.1	92.9	94.1	91.3	82.1	91.3
10°	93.5	100	84.8	87.1	90.6	93.1	75	66.7	75	87.5	57.1	83.3	87.5	90	76.9	83.3	86.2	83.3
20°	93.5	88.1	100	97.2	91.2	86.7	100	100	80	80	55.6	75	87.5	63.6	66.7	82.4	88	91.9
30°	90.5	92.7	86.7	100	82.1	85.4	91.3	92.9	77.8	72.7	87.5	55.6	77.8	75	81.3	89.5	85.7	87.9
40°	92.3	88.9	82	86.5	100	85.4	94.1	81	87.5	55.6	75	83.3	57.1	90	90	85.7	94.7	100
50°	88.9	92.9	83.8	87.8	85.4	100	88.4	89.7	87.5	100	66.7	66.7	88.9	62.5	68.8	100	89.5	84.6
60°	96.9	96.9	93.9	84.8	85.7	90.7	100	86.1	93.8	70.6	80	87.5	75	83.3	100	94.1	78.6	95.5
70°	96.4	92.6	93.8	87.1	86.5	88.1	90.5	100	93.5	81.8	62.5	100	88.9	61.5	82.4	94.1	85.7	95
80°	80	89.3	90	87.1	79.4	86.2	86.5	92.3	100	94.7	53.8	66.7	62.5	92.3	73.3	92.3	68.8	94.4
90°	94.4	89.7	86.7	85.7	90.9	96.9	96.3	100	96.7	90.5	88	76.9	76.9	85.7	94.4	100	87.5	100
100°	82.8	70.8	93.3	88	87.5	87	82.6	93.8	95.2	90.9	0	90.9	83.3	72.2	90.5	66.7	93.8	80
110°	77.8	90.5	88	92	87	84.2	64.3	83.3	92.9	68.4	95.5	100	87.5	80	80	77.3	84.2	91.3
120°	91.2	87	92.3	95.2	85	72.2	81.8	80	80	70	80	75	100	79.3	83.3	86.4	83.3	80
130°	85.7	84.6	77.8	85	90.5	76.2	80	63.6	84.6	72.7	80	69.2	82.6	95.8	94.3	82.6	88.9	88.2
140°	85	93.1	93.3	96	81.8	81.8	76.9	57.1	75	90.9	57.1	72.7	89.5	96.8	100	85.4	89.7	82.1
150°	88.2	96.2	92.9	93.3	91.3	95	100	63.6	88.9	63.6	55.6	66.7	100	86.4	84.6	100	89.1	87.5
160°	86.4	100	80.6	93.3	93.1	94.7	64.3	57.1	57.1	55.6	62.5	77.8	70	88.9	78.6	93.5	100	91.5
170°	87.8	83.3	93.9	96.3	96.3	96	92.3	55.6	77.8	60	81.8	87.5	57.1	75	75	77.4	90.2	94
180°	91	95.7	87.8	87.1	85.7	96	80	55.6	57.1	100	88.9	77.8	75	61.5	86.4	85	90	93.9
190°	89.8	92.1	80	84.8	87.5	90	75	72.7	60	70	83.3	57.1	77.8	83.3	94.1	82.4	85.2	84.2
200°	86.7	86	87.7	91.9	84.8	84.8	82.4	63.6	70	77.8	55.6	60	77.8	100	87.5	80	95.5	85
210°	93	86.8	85.1	90.4	79.5	86.1	81	83.3	91.7	72.7	55.6	77.8	66.7	75	100	76.5	94.7	91.7
220°	89.7	92.1	83.3	89.2	5.71	90.2	94.1	83.3	88.9	55.6	62.5	66.7	57.1	80	81.8	83.3	87.5	92.3
230°	92.3	100	84.8	95	88.1	44.6	89.1	95.5	94.4	100	70	87.5	66.7	88.9	70.6	71.4	80	90.5
240°	94.1	100	84.2	84.4	85	88.6	90.6	91.9	94.7	81.8	100	75	85.7	100	100	80	100	95.5
250°	96.8	96.3	87.9	90.6	86.5	87.5	88.1	43.1	95.2	81.8	55.6	75	87.5	92.3	93.8	82.4	89.5	88.9
260°	81.5	92.6	90.3	86.2	79.4	86.2	84.8	84.2	52.2	84.2	58.3	77.8	66.7	66.7	84.6	100	76.9	100
270°	9.09	89.3	96.6	85.7	91.7	100	96.6	100	95	96.6	90.5	75	83.3	92.9	94.7	87.5	81.3	100
280°	86.2	73.9	89.7	92	80	82.6	84.2	53.3	95.2	100	93.7	90.9	82.4	88.2	85.7	88.9	94.1	81.8
290°	89.7	78.3	91.7	95.7	90.9	100	53.3	91.7	100	76.5	84.2	92.6	91.3	84.2	73.7	95	94.7	91.3
300°	96.9	87.5	89.7	95.5	90	72.2	63.6	92.3	100	72.7	71.4	75	13	81.5	95.8	90	90.9	89.7
310°	77.1	92	82.1	87	90.9	90.5	81.8	70	76.9	72.7	77.8	75	90	8.33	97	92.3	81	88
320°	84.6	82.1	89.7	96.2	83.3	85	85.7	66.7	77.8	91.7	75	70	87	82.8	87.9	84.1	85.4	82.1
330°	85.7	92.3	93.1	88.9	91.3	95.2	100	66.7	88.9	63.6	55.6	88.9	83.3	94.7	85.7	85.5	91.5	84.6
340°	93.3	94.9	81.8	93.1	92.6	95.2	69.2	66.7	62.5	70	57.1	90	90	87.5	88.5	95.1	91.5	89.4
350°	91.1	88.9	86	96.4	96.4	91.7	91.7	62.5	87.5	66.7	75	66.7	71.4	72.7	78.3	81.3	92.7	90.3
	180°	**190°**	**200°**	**210°**	**220°**	**230°**	**240°**	**250°**	**260°**	**270°**	**280°**	**290°**	**300°**	**310°**	**320°**	**330°**	**340°**	**350°**
0°	91	95.7	88.6	90	80	88.9	100	70	75	0	55.6	87.5	80	92.9	93.3	95.7	90.7	87.5
10°	95.7	91.9	84.4	88.6	93.5	85.2	76.5	0	62.5	62.5	66.7	75	55.6	90.9	100	77.8	84.8	75.6
20°	90.5	88.6	29.6	92.1	88.9	83.3	87.5	63.6	90	75	75	62.5	77.8	78.6	66.7	71.4	75.9	89.1
30°	94.7	92.5	89.1	2.82	85	87.5	68.4	93.3	90	70	75	62.5	90	75	92.9	93.3	87.5	82.9
40°	94.4	91.7	84.1	86.8	90	87	93.8	83.3	91.7	66.7	75	83.3	85.7	100	90	93.8	82.1	82.8
50°	87.5	93.1	88.2	89.7	85.7	92.1	93.2	96.2	100	100	66.7	77.8	80	81.8	73.3	73.3	82.6	87.1
60°	94.1	87.5	90.6	83.9	87.5	89.8	90.8	88.6	87	75	76.9	87.5	100	88.9	100	93.3	95.5	89.7
70°	90.3	88.9	88.6	82.1	84.4	82.9	88.9	70.3	84.8	93.3	88.9	80	91.7	90.9	93.8	95	88.5	88.9
80°	85.7	72	89.7	84.8	75	89.3	92.6	91.9	92.8	90.5	81.3	81.8	75	84.6	78.6	84.6	84.6	78.1
90°	97.1	96.2	87.1	96	90	96.4	95.8	91.3	93.5	95.1	88.9	84.6	78.6	92.9	87.5	93.3	100	90.3
100°	87.5	85	85.7	87	90.5	91.3	95.2	61.1	90	91.7	1.56	85.2	83.3	83.3	88.9	80	85	80.8
110°	76	85.7	75	94.1	94.4	73.3	58.3	100	94.1	88.2	91.7	35.8	92	84.2	70.6	78.9	85.2	79.3
120°	90.3	84	89.3	95	81.3	64.3	90	80	80	100	80	83.3	91.5	78.6	81.5	87.5	85.7	84.6
130°	79.3	90.9	80	80	88.2	82.4	90.9	70	72.2	81.8	75	81.3	79.2	83.6	89.2	86.7	80	87.1
140°	83.3	92.9	77.8	85.7	82.4	83.3	54.5	83.3	72.7	84.6	66.7	92.9	94.1	93.1	89.4	82	81.4	81
150°	93.5	92.6	92.9	92.9	88.9	94.1	100	83.3	87.5	70	90	66.7	100	93.8	94.3	15.9	91.7	81.8
160°	88.6	94.7	96.7	90	91.7	100	75	57.1	62.5	87.5	75	90	81.8	88.9	85.3	91.5	52.2	88.9
170°	91.5	86.5	80	100	86.7	87	92.3	33.3	60	90	80	66.7	80	66.7	86.4	84.2	85.4	86.7
180°	100	95.8	93.3	94.1	86.1	88.9	100	57.1	71.4	90.9	85.7	71.4	55.6	91.7	87.5	89.3	88.9	86
190°	89.6	100	90.7	82.4	91.2	85.2	70.6	71.4	57.1	66.7	83.3	60	88.9	83.3	100	76	84.2	71.8
200°	91.5	90.7	100	91.7	89.5	81.8	73.3	62.5	63.6	80	66.7	75	90	70	87.5	83.3	85.3	86.7
210°	89.7	88.4	80	100	86.5	87.5	90.9	85.7	72.7	72.7	60	75	88.9	75	66.7	73.7	83.3	86.1
220°	90.6	91.2	87	89.2	0	89.7	94.1	85.7	77.8	55.6	62.5	40	100	66.7	83.3	71.4	81.5	84.8
230°	86.1	88.9	85	90.5	90.2	56.6	92.1	90	85	92.3	75	87.5	87.5	81.8	73.3	78.6	95.5	84.2
240°	92.9	93.5	90.9	82.4	87.5	90.5	100	93.9	91.7	76.5	90	75	100	88.9	100	87.5	96.4	86.7
250°	89.7	85.2	85.3	81.8	82.9	86.4	92.1	100	82.4	100	55.6	62.5	77.8	90.9	100	78.6	85.7	90.3
260°	96.2	75	89.3	89.7	78.9	86.2	93.5	86.1	100	90	94.7	90	75	92.3	84.6	100	86.4	87
270°	87.9	96.7	84.4	91.7	86.7	93.3	95.8	100	94.4	98.7	90	83.3	78.6	92.9	93.3	64.3	100	96.6
280°	86.2	70.8	80.6	100	78.3	88	95.5	95.5	95	88	96.6	84	94.4	77.8	72.2	75	73.7	74.1
290°	81.8	90	87	95.8	90	88.9	61.5	54.5	80	88.9	92	100	95.7	85	88.9	81.8	90.5	95.8
300°	93.3	84.6	92.3	95	87.5	64.3	63.6	66.7	90.9	100	75	85.7	100	78.8	90.9	96.2	90.9	87.1
310°	80	81.5	80	85.7	69.2	66.7	91.7	63.6	78.6	72.7	72.7	77.8	79.2	100	92.5	89.7	90.9	91.7
320°	86.1	93.3	90.3	91.3	78.3	83.3	100	57.1	77.8	92.9	63.6	54.5	95	90.9	100	81.3	84.4	83.9
330°	96.9	88.9	87.1	92	85	88.2	66.7	75	57.1	60	55.6	75	86.7	93.3	84.6	100	93.9	84.4
340°	88.1	94.3	73.7	87.1	85.7	100	100	0	77.8	87.5	55.6	77.8	75	83.3	89.7	95.8	100	83
350°	91.8	84.2	90.9	96.7	88.9	85	92.3	85.7	66.7	88.9	70	100	57.1	66.7	83.3	85.7	90.7	100

Table 8.28: Match-matrix results of the percentage of matched keypoints, produced from the in-plane rotational 2D data of a **mannequin head** (from 0° at 10° clockwise increments up to 350°), using 2D SIFT.

	0°	10°	20°	30°	40°	50°	60°	70°	80°	90°	100°	110°	120°	130°	140°	150°	160°	170°
0°	100	93.9	86.1	95.5	100	81.8	61.5	73.3	82.4	0	64.3	54.5	75	61.5	73.7	93.9	96.6	100
10°	88.9	100	91.9	90.3	95	82.4	72.7	80	76.9	53.3	58.3	62.5	87.5	76.9	88.2	76	96.4	93.1
20°	87.5	91.7	100	94.1	90	92.9	73.3	76.9	78.9	70.6	73.3	68.8	100	83.3	61.5	95.5	92.6	97
30°	93.3	93.8	93.3	98.8	82.1	71.4	70.6	82.4	73.3	64.7	69.2	58.3	84.6	55.6	100	95.7	96.3	96.2
40°	96.7	93.3	82.9	91.4	0	78	74.1	68.4	77.3	71.4	85.7	73.3	81.8	76.9	91.7	80	95.7	96.4
50°	90.3	96.3	87.1	94.3	89.2	100	84.6	82.6	79.2	68.2	93.3	92.3	90.9	83.3	93.3	81.8	91.3	100
60°	100	92	96.3	94.7	80.6	88.9	35.6	89.2	89.7	72.7	81.8	64.7	81.3	76.9	76.9	100	95.2	100
70°	100	81	86.4	90.5	80	88	81.6	100	72.7	80	72.7	71.4	86.7	72.7	100	100	85	95.7
80°	91.7	85	89.5	70.6	88.2	88.2	86.4	78.1	97.6	69.2	80	90	87.5	76.9	100	80	83.3	96.2
90°	75	95	89.5	93.3	90	93.3	85	87.5	77.4	100	91.7	92.3	72.2	69.2	87.5	86.4	86.4	88.9
100°	100	81.3	95.5	100	77.8	71.4	66.7	72.7	92.3	85.2	100	82.4	81.5	94.7	95	96	91.7	100
110°	96.2	88.2	93.8	94.1	66.7	75	73.3	60	69.2	75	89.3	79	89.6	86.1	88.5	88.9	96.4	96
120°	95.8	100	82.4	87.5	69.2	72.7	83.3	64.3	78.9	69.6	84	87.5	100	81.8	83.9	77.3	92.9	96
130°	100	91.3	100	87.5	60	80	83.3	71.4	80	81	80	90.6	80	100	85.4	94.1	93.1	100
140°	82.1	90.9	85.7	100	88.9	81.8	72.7	80	72.2	75	75	59.1	81.8	89.4	100	95.5	90.6	96.4
150°	100	87.5	95.7	95.5	77.8	54.5	66.7	66.7	68.4	68.4	80	68.4	87	80.5	89.2	100	93	94.7
160°	97.1	96.3	96	80	92.9	66.7	75	66.7	66.7	64.7	68.8	76.5	78.9	79.2	86.8	87	100	93.8
170°	94.3	96.3	92	90.9	83.3	81.8	55.6	80	73.3	76.5	70.6	85.7	92.9	78.9	95.7	90.2	93.2	100
180°	13.1	94.6	86.1	95.8	93.3	72.7	66.7	91.7	66.7	84.2	64.3	64.3	75	75	90.9	87.5	94.4	94.9
190°	90	5.08	87.2	93.9	89.5	92.3	90	90.9	57.9	70.6	58.3	60	73.3	66.7	83.3	91.7	87.1	91.9
200°	97.4	89.5	0	82.9	88.5	81.3	53.3	82.4	77.8	55.6	72.7	73.3	75	75	94.1	100	96.3	100
210°	100	93.1	88.2	80.6	82.1	68.2	80	82.4	70	68.4	66.7	58.3	72.7	66.7	93.8	96.2	100	96.2
220°	100	90	90.6	92.5	76.8	82.5	74.1	81.8	75	76.5	78.6	92.9	81.8	90.9	100	100	95.5	96.3
230°	87.9	95.8	88.5	89.7	86.4	34.6	86.5	79.2	88.5	63.6	71.4	66.7	66.7	83.3	86.7	90.9	90.5	96.6
240°	100	95.7	92.3	84.2	91.7	92.1	47.1	88.6	85.2	72.7	86.4	61.1	92.3	73.3	60	94.1	94.7	100
250°	91.3	91.3	90.5	95.8	90	85	82.4	75.4	69.7	69.2	77.8	64.3	86.7	90	93.8	100	80	96
260°	95.2	87	89.5	78.9	87.5	82.4	81.8	77.4	78.9	84.6	81.8	89.5	89.5	71.4	100	90.9	92.6	96
270°	70.4	89.5	94.4	100	90.9	86.7	89.5	78.3	83.9	79.2	96.6	85.7	80	66.7	94.7	100	95	100
280°	100	94.1	100	100	81.8	72.7	66.7	72	78.6	86.2	27.9	85.3	95.7	88.9	100	91.7	96.3	100
290°	84.2	87.5	92.9	66.7	80	92.3	69.2	81	66.7	65.4	80	80.3	91.7	83.3	91.3	87	91.7	90.9
300°	95.7	100	94.4	87.5	60	66.7	73.3	70.6	68.2	68.2	84	86.5	18.1	81.3	88.5	87.5	92.6	96.2
310°	100	85	100	66.7	77.8	66.7	81.8	75	87	81	75	82.1	77.1	74.7	84.8	87.5	90.3	96.2
320°	92.6	90.9	90.5	100	55.6	58.3	77.8	85.7	81	81	81.3	55.6	79.3	84.3	81.1	92.7	93.1	90
330°	100	88.9	92	100	100	66.7	69.2	80	66.7	76.5	73.3	75	95.5	74.1	89.2	75	97.4	97.2
340°	100	96.6	95.8	79.2	91.7	66.7	63.6	72.7	81.8	60	80	64.7	93.8	88.9	90	92.7	77.5	97.8
350°	87.2	92.9	88.5	100	100	91.7	75	62.5	73.3	72.2	68.8	58.3	69.2	73.3	92.3	96.9	85.1	75

	180°	190°	200°	210°	220°	230°	240°	250°	260°	270°	280°	290°	300°	310°	320°	330°	340°	350°
0°	84.2	89.2	86.1	93.5	100	91.7	63.6	73.3	72.2	76.2	72.2	70.6	66.7	80	95.2	96.9	97.3	93.9
10°	85.7	71.7	92.1	91.4	93.1	100	68.4	80	73.3	73.3	60	71.4	78.6	71.4	77.8	81	92.9	90.3
20°	88.6	89.7	80.9	87.2	94.1	70	87.5	76.2	90.5	65	75	72.2	80	76.9	81.8	95.7	96.2	100
30°	92.9	93.9	88.9	72.5	89.1	88.9	75	90	71.4	70	81.3	57.1	76.9	80	89.5	100	96	96
40°	89.7	93.1	83.9	83.7	67.1	85.7	77.4	73.9	81	76.2	81	83.3	83.3	80	85.7	76.2	100	100
50°	87.1	96.3	93.5	88.6	84.4	38.2	84.4	76	76	78.3	72.7	83.3	93.3	90.9	90.9	100	95.7	100
60°	100	96.2	96	88	93.5	89.5	68.6	86.1	85.7	84	88	81	100	75	91.7	100	100	100
70°	86.4	87	95	70.8	80.8	78.6	80.5	77	75	75	72	76.5	80	84.6	76.9	100	90	100
80°	90	83.3	90.5	88.5	94.1	91.7	85.7	70.6	78.6	82.8	85.2	73.9	88.2	78.6	91.7	85	88	90.5
90°	100	95.2	92.6	83.3	95.5	70	73.1	86.7	87.1	32.7	86.7	84.6	70	63.2	75	85.7	95.7	87
100°	95.5	90.5	100	94.1	100	90.5	85	83.3	82.6	80.6	53.7	80	86.4	83.3	76.2	89.3	96.4	100
110°	95.7	90.9	95	90	86.7	92.9	85.7	65.2	73.1	80	85.3	79.7	87.2	74.2	83.3	95.7	90.3	96
120°	91.3	100	81	86.7	85.7	83.3	81.3	72.7	73.9	79.2	88	94.4	77.1	77.1	81.8	96.2	89.3	92.6
130°	96.6	83.3	100	87.5	100	81.8	68.8	81	88.5	83.3	88.5	84.4	82.9	74.1	83.3	93.8	91.4	96
140°	96.6	92.3	83.3	90	84.6	85.7	87.5	81.8	90.9	85.7	79.2	85.2	79.3	87.5	79.2	90	94.1	77.8
150°	96.8	96.2	100	95.8	76.9	100	68.8	68.8	73.7	72.7	76.2	72.7	95.2	80.6	88.4	78.1	95.2	94.1
160°	93.9	84.4	96.4	84	60	81.8	75	58.8	73.7	68.4	78.9	76.2	84.2	92	93.1	97.6	74	88.6
170°	97.1	90.9	96.6	92	87.5	92.3	69.2	58.3	81.3	75	65	94.1	100	86.7	84	93.8	95.6	76.2
180°	100	91.7	92.3	89.7	100	69.2	66.7	76.5	82.4	78.9	72.2	66.7	69.2	78.6	77.8	93.9	94.1	91.4
190°	86.5	100	92.7	94.7	93.8	75	52.9	66.7	70	71.4	76.2	68.8	92.3	83.3	82.4	92.6	91.4	93.5
200°	97.2	91.7	74.4	89.5	92.5	89.5	80	77.8	72.2	71.4	66.7	76.5	71.4	83.3	84.6	96	100	93.3
210°	86.2	94.4	84.4	100	80	79.2	77.3	78.3	75	70	80	80	83.3	60	93.8	95	92	96.2
220°	82.8	93.9	92.3	82.9	100	86	74.3	71.4	78.3	82.6	83.3	82.4	83.3	70	82.4	100	96.2	100
230°	84.4	93.5	93.5	80.5	80.9	100	80.4	76.9	82.6	80	75	78.9	84.6	78.6	70	82.4	92	96.4
240°	100	92.3	96.4	84	83.8	82.9	100	84.6	85.7	82.6	72	60.9	88.2	75	90.9	100	100	89.5
250°	90.9	87.5	91.7	87.5	81.5	85.7	84.6	100	74.2	71.4	78.3	70	81.3	83.3	81.3	95	79.2	100
260°	96	92.3	95.8	90	96.2	94.7	87.1	69.7	100	85.7	88.5	81.8	88.9	75	86.7	90.9	66.7	100
270°	78.6	95.7	96.2	88.2	100	83.3	91.7	81.8	89.7	100	90	70.4	78.9	75	87.5	100	95.2	95.7
280°	81.8	95.2	100	88.2	92.3	78.9	77.8	76	89.7	80	100	86.8	92	65	84.2	91.7	100	100
290°	95	89.5	92.3	92.3	70	92.9	73.7	87	70.4	73.1	87.9	100	93.6	81.6	84	95.5	87.5	95.2
300°	91.7	100	95.5	86.7	83.3	83.3	85.7	70	94.7	70.4	90	91.7	98.9	81.6	85.7	84.6	90.9	96
310°	96.4	96	95.2	87.5	83.3	90	64.7	82.4	87.5	84	72	87.9	75.5	100	83.6	94.1	93.1	96.3
320°	96.4	91.7	95.8	88.9	83.3	92.9	76.9	86.7	90.9	72.4	70.4	80.8	79.4	80.7	100	82.6	91.9	85.2
330°	96.9	96.3	100	100	72.7	75	64.3	78.6	66.7	76.2	66.7	71.4	91.3	86.1	85.1	100	95.7	97.1
340°	93.9	79.3	100	83.3	93.8	63.6	71.4	64.7	73.7	70.6	73.7	70	73.7	91.3	90	100	100	87.8
350°	91.7	96.9	89.7	95.8	84.6	60	76.5	86.7	81.3	70	61.9	85.7	94.1	82.4	80	96.9	80.9	100

Table 8.29: Match-matrix results of the percentage of matched keypoints, produced from the in-plane rotational 2D data of a **human face** (from 0° at 10° clockwise increments up to and including 360°), using 2D SIFT.



(a) Match-matrix (b) ROC

Figure 8.18: (a) Match-matrix obtained by exploring all pairwise combinations of the feature descriptors extracted from rotated in-plane 2D images of a **human face**, using standard 2D SIFT. (b) Results presented in a ROC space. This figure can be compared with Figure 6.7 and Figure 8.2, where the results were obtained from the same data set using 2.5D SIFT and 2.5D$_{pc}$ SIFT respectively.

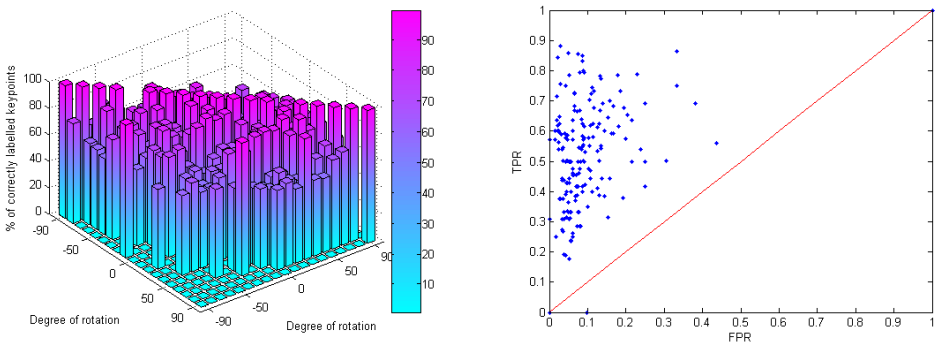
The results obtained from the feature descriptors extracted from in-plane rotated 2D images using standard 2D SIFT illustrate that they are robust to in-plane rotational changes, with over average of 90% of the test images matched at over 50% of keypoint matching rate. However, it is noted that when the results are plotted within a ROC space, the FPR are larger than that obtained with 2.5D$_{pc}$ SIFT for range images, indicating the results are not as reliable as 2.5D SIFT.

8.4.1.2 Out-of-plane Rotations on Real Data

This section investigates the invariance properties of the feature descriptors extracted from the out-of-plane rotated 2D images captured using the stereo-pair system. The feature descriptors were extracted from standard 2D SIFT and Table 8.30 shows the results obtained from a set of 2D human face data and a set of 2D mannequin head data. Beside each average matching rate is two symbols, denoting the statistically significant, using the Wilcoxon Matched-Paired Signed-Ranked test, when compared to the two baseline methods (2.5D SIFT and 2.5D$_{pc}$ SIFT respectively). Each symbol can be one of $\ll$, $<$, $=$, $>$, $\gg$. The results show that the performance of 2D SIFT on intensity images is significantly worse ($p < 0.01$) than the performance of the 2.5D$_{pc}$ SIFT on range images (i.e. 2.5D$_{pc}$ SIFT outperforms 2D SIFT).

System	Subject	Pairings	Average Matching Rate	Images Achieving Matching Rate > 50%	60%	Match-Matrix	ROC
System	Mannequin head	361	41.6%>≪	55.1%	43.8%	Table 8.31 Figure 8.19	Figure 8.19
	Human face	361	27.7%≪≪	35.5%	30.1%	Table 8.32 Figure 8.20	Figure 8.20

Table 8.30: Results obtained by exploring all pairwise combinations of the feature descriptors extracted from captured 2D out-of-plane range images of a mannequin head and a human face, using 2D SIFT.



(a) Match-Matrix (b) ROC

Figure 8.19: (a) The match-matrix results of the percentage of matched keypoints, produced from a set of out-of-plane rotational 2D data of a **mannequin head** (captured from -90° at 10° increments up to 90°), using 2D SIFT. (b) The matching results presented in a ROC space. This figure can be compared to Figure 6.8 and Figure 8.3, where the results were obtained from the same data set using 2.5D SIFT and 2.5D$_{pc}$ SIFT.

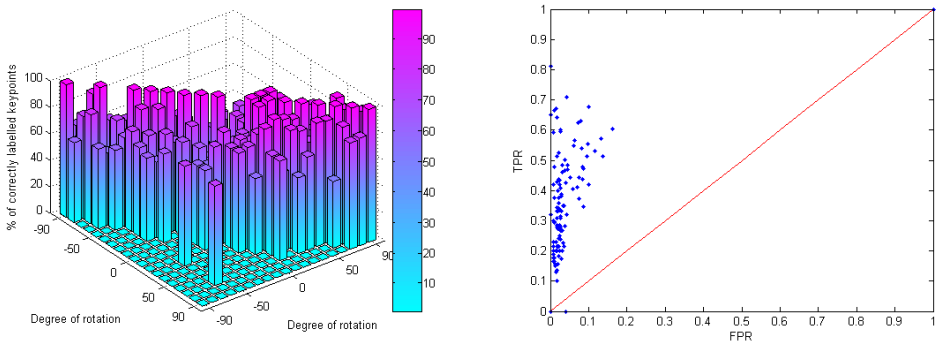
By comparing the matching results obtained by extracting the feature descriptors from out-of-plane rotated 2D images using standard 2D SIFT, and by extracting feature descriptors from out-of-plane rotated range images using the 2.5D$_{pc}$ SIFT shown in Table 8.7, it can be deduced that the performance of 2.5D$_{pc}$ SIFT is, on average, approximately 15% better than the performance of standard 2D SIFT with out-of-plane 2D rotated images, at 50% keypoint matching rate.

	-90°	**-80°**	**-70°**	**-60°**	**-50°**	**-40°**	**30°**	**-20°**	**-10°**	**0°**	**10°**	**20°**	**30°**	**40°**	**50°**	**60°**	**70°**	**80°**	**90°**
-90°	100	60	0	57.1	75	0	0	0	0	0	0	0	60	0	0	0	0	0	0
-80°	75	100	68.8	61.5	60	57.1	0	0	0	0	0	0	0	0	0	0	0	0	0
-70°	0	61.1	100	81.5	77.8	55.6	55.6	57.1	0	0	0	0	0	0	0	0	0	0	0
-60°	0	60	86.4	100	66.7	72.2	77.8	62.5	75	63.6	0	0	0	0	0	0	0	0	0
-50°	0	57.1	66.7	73.3	84.3	76.5	96.4	87.5	77.8	83.3	75	62.5	0	0	0	0	0	0	0
-40°	0	75	62.5	80	86.8	100	97.4	86.7	86.7	64.7	90	60	75	0	0	75	0	0	0
-30°	0	0	0	76.9	65.4	90.3	100	95.5	84.8	82.1	78.6	66.7	72.7	75	0	0	0	0	0
-20°	0	0	57.1	63.6	83.3	82.8	77.6	100	95.2	74.1	80	78.6	60	66.7	0	0	0	0	0
-10°	100	0	0	72.7	84.6	77.8	75	95.3	100	96.9	84.2	61.5	66.7	66.7	0	0	0	0	0
0°	0	0	0	71.4	58.3	68.8	80.6	80.8	93.3	100	96.4	64.7	73.3	77.8	87.5	83.3	0	0	0
10°	0	0	0	75	77.8	62.5	88.9	93.3	89.5	91.2	100	72.4	72.7	92.9	55.6	75	57.1	0	0
20°	0	60	0	0	60	62.5	58.8	66.7	71.4	69.6	84.4	100	78.4	85.7	83.3	88.9	70	66.7	0
30°	0	0	0	0	0	55.6	63.6	62.5	66.7	60	68.4	82.4	100	76.2	91.7	62.5	80	75	0
40°	0	0	60	66.7	57.1	57.1	62.5	77.8	55.6	72.7	84.6	79.2	76	93.9	65	66.7	55.6	0	0
50°	0	0	0	0	0	0	0	0	0	88.9	90	88.2	94.1	85	100	83.3	80	57.1	0
60°	0	0	0	0	66.7	0	62.5	85.7	57.1	54.5	60	60	85.7	95.2	84.2	100	66.7	70	0
70°	0	0	0	0	0	0	100	0	60	57.1	66.7	58.3	56.3	55.6	66.7	73.7	100	55.6	0
80°	0	0	0	0	0	0	0	0	0	60	62.5	55.6	90.9	57.1	55.6	66.7	70	100	0
90°	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100

Table 8.31: Match-matrix results of the percentage of matched keypoints, produced from the out-of-plane rotational 2D data of a **mannequin head**, from -90° to 90° of rotations, using 2D SIFT.

	-90°	**-80°**	**-70°**	**-60°**	**-50°**	**-40°**	**30°**	**-20°**	**-10°**	**0°**	**10°**	**20°**	**30°**	**40°**	**50°**	**60°**	**70°**	**80°**	**90°**
-90°	100	66.7	60	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
-80°	60	77.2	91.3	66.7	60	0	0	0	0	0	0	0	0	0	0	0	0	0	0
-70°	0	82.1	100	72.7	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
-60°	0	60	54.5	76	73.9	81.3	66.7	66.7	0	0	0	0	0	0	0	0	0	0	0
-50°	0	0	60	66.7	100	80	66.7	62.5	0	0	0	0	0	0	0	0	0	0	0
-40°	0	0	0	75	88.9	100	66.7	63.6	0	0	0	0	0	0	0	0	0	0	0
-30°	0	0	0	66.7	73.1	90.3	100	80	75	0	60	0	0	0	0	0	0	0	0
-20°	0	0	0	62.5	57.1	75	66.7	99.2	72.7	57.1	66.7	75	0	75	0	0	0	0	0
-10°	0	0	0	0	66.7	0	70	79.2	100	55.6	75	0	0	0	0	0	0	0	0
0°	0	0	0	0	0	0	0	66.7	57.1	100	64	80	69.2	72.7	60	0	0	0	0
10°	0	0	0	0	0	0	60	57.1	60	68	6.25	78.8	80.6	70.8	77.8	0	0	0	0
20°	0	0	0	0	0	0	60	75	83.3	73.3	76.5	100	100	100	92.9	66.7	0	0	0
30°	0	0	0	75	0	0	0	0	75	78.6	75	98	100	72.7	81	70	75	0	0
40°	0	0	0	0	0	0	0	0	75	76.9	84.2	71.4	66.7	100	82.9	75	90	0	0
50°	0	0	0	0	0	0	0	0	0	57.1	91.7	96	93.3	90.7	100	77.8	71.4	0	0
60°	0	0	0	0	0	0	0	0	0	0	75	60	88.9	85.7	75	100	85	100	0
70°	0	0	0	75	0	0	0	0	0	0	75	0	57.1	70	92.9	91.7	100	95	81.8
80°	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	50	88.2	100	94.7
90°	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	80	81	100

Table 8.32: Match-matrix results of the percentage of matched keypoints, produced from the out-of-plane rotational 2D data of a **human face**, from -90° to 90° of rotations, using 2D SIFT.



(a) Match-Matrix (b) ROC

Figure 8.20: (a) The match-matrix results of the percentage of matched keypoints, produced from a set of out-of-plane rotational 2D data of a **human face** (captured from -90° at 10° increments up to 90°), using 2D SIFT. (b) The matching results for presented in a ROC space. This figure can be compared to Figure 6.9 and Figure 8.4, where the results were obtained from the same data set using 2.5D SIFT and 2.5D$_{pc}$ SIFT.

8.4.1.3 Synthetic Out-of-plane Rotations

This section presents the results obtained by matching synthetically rotated 2D images, where the images are rotated about the yaw and pitch axis (±40°) respectively, at 10° intervals.

Table 8.10 summaries the results obtained by exploring all pairwise combinations of the feature descriptors extracted from the synthetically rotated 2D images of a mannequin head and a human face respectively, about the yaw and pitch axes, using standard 2D SIFT. Shown next to the average matching rate is a symbol (one of $\ll$, $<$, $=$, $>$, $\gg$), showing the statistical significance difference, using the Wilcoxon Matched-Pair Signed-Rank test, to the average matching rate of 2.5D$_{pc}$ SIFT.

The performance rate of the feature descriptors extracted from synthetically rotated range images (about the yaw and pitch axes) using 2.5D$_{pc}$ SIFT is better than the performance rate of the feature descriptors extracted from the equivalent 2D images using 2D SIFT. Examination of the two sets of results (by taking the averages between the subjects) show that at 70% keypoint recognition rate, the obtained 2.5D$_{pc}$ results for out-of-plane images rotated about the yaw axis (Table 8.10) were approximated 1.5% better than the results obtained from 2D SIFT. The results obtained for 2.5D$_{pc}$

Subject	Axis	Pairings	Average Matching Rate	Images Achieving Matching Rate > 50%	60%	70%	Match-Matrix	ROC
Mannequin head	yaw	81	81.7%$^{<}$	98.8%	92.6%	86.4%	Table 8.34 Figure 8.21	Figure 8.21
Human head	yaw	81	74.9%$^{\ll}$	97.5%	93.8%	66.7%	Table 8.35 Figure 8.22	Figure 8.22
Mannequin head	pitch	81	82.4%$^{=}$	100%	97.5%	87.7%	Table 8.36 Figure 8.23	Figure 8.23
Human head	pitch	81	75.9%$^{\ll}$	97.5%	87.6%	69.1%	Table 8.37 Figure 8.24	Figure 8.24

Table 8.33: Results obtained by exploring all pairwise combinations of the feature descriptors extracted from synthetically rotated out-of-plane 2D images of a mannequin head and a human face, using standard 2D SIFT.

	-40°	**30°**	**-20°**	**-10°**	**0°**	**10°**	**20°**	**30°**	**40°**
-40°	100	97.7	75.7	92.7	80.6	79.3	82.6	75	54.5
-30°	89.8	100	85.1	83	79.6	72.2	81.3	74.1	57.9
-20°	76.2	85.7	100	67.3	69.6	75.6	81.6	74.2	57.9
-10°	83.8	85.4	69.8	46.8	55.9	71.7	91.3	80.6	73.9
0°	87.9	82.2	72.5	56.7	100	75.9	82.6	77.8	78.3
10°	81.3	74.4	69.6	81.5	83.3	100	82.2	83.7	62.5
20°	78.6	87.9	89.5	91.8	90	84.1	100	95.1	74.1
30°	87.5	81.5	76.7	85.7	95.5	84.4	91.1	100	75.9
40°	100	94.4	100	73.9	82.1	75.9	93.3	87.1	100

Table 8.34: Match-matrix results of the percentage of matched keypoints, produced from the synthetically rotated out-of-plane (about yaw axis) 2D data of a **mannequin head**, from -40° to 40° of rotations, using 2D SIFT.

out-of-plane range images rotated about the pitch axis (Table 8.10) were approximately 18% better than the results obtained from 2D images using 2D SIFT.

The next section presents the results obtained by extracting feature descriptors from 2D images of different scales using 2D SIFT.

	-40°	**30°**	**-20°**	**-10°**	**0°**	**10°**	**20°**	**30°**	**40°**
-40°	100	68.2	73.7	70	70	73.1	53.3	63.6	92.3
-30°	71.1	100	71.4	70.7	71.4	86.1	65.2	80	64.3
-20°	70	67.3	10.1	70	80	85.7	69	85.7	62.5
-10°	69.7	71.1	68.6	2.7	70.9	73.3	71	91.3	70
0°	70.4	76.7	77.8	68.6	100	76.8	87.2	78.6	69.6
10°	61.9	80.6	81.1	79.5	77.6	53.8	89.5	75	76.9
20°	68.8	69.6	72.7	68.8	91.1	90.2	100	71	78.6
30°	57.1	73.7	77.3	88.5	79.4	73.7	68.8	100	87.5
40°	92.9	68.8	76.5	71.4	74.1	93.5	78.8	77.8	100

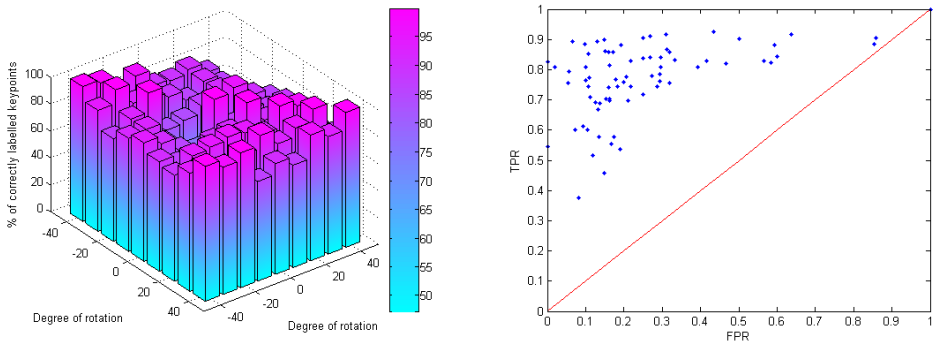
Table 8.35: Match-matrix results of the percentage of matched keypoints, produced from the synthetically rotated out-of-plane (about yaw axis) 2D data of a **human face**, from -40° to 40° of rotations, using 2D SIFT.

	-40°	**30°**	**-20°**	**-10°**	**0°**	**10°**	**20°**	**30°**	**40°**
-40°	94.5	89.2	81.8	100	84.6	80	75	75	85.7
-30°	93.5	100	85.7	90.6	81.5	79.2	80	81.3	78.9
-20°	78.6	89.1	100	76.5	88.2	81.3	80.6	70.4	77.3
-10°	89.5	85.7	82.7	100	81.3	76.7	77.8	80.6	84.4
0°	54.5	100	75	80.9	100	67.7	77.8	74.4	89.7
10°	57.1	87.5	80.6	81.4	68.9	100	66.7	69.8	80
20°	76.9	83.3	97	75.6	72	72	100	91.5	75
30°	72.7	87.5	92.9	91.9	77.5	66.7	74.5	100	89.2
40°	100	66.7	69.2	91.7	85.7	76.9	71.7	67.6	100

Table 8.36: Match-matrix results of the percentage of matched keypoints, produced from the synthetically rotated out-of-plane (about pitch axis) 2D data of a **mannequin head**, from -40° to 40° of rotations, using 2D SIFT.

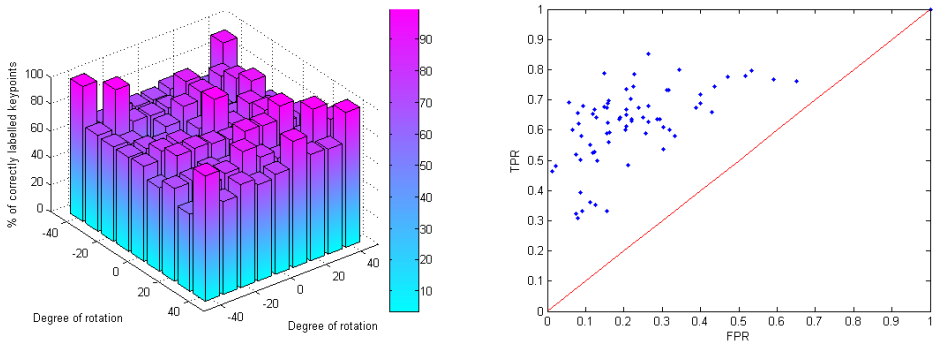
	-40°	**30°**	**-20°**	**-10°**	**0°**	**10°**	**20°**	**30°**	**40°**
-40°	100	75	72	90	90	65	70	66.7	75
-30°	60	100	92.6	78.6	82.8	66.7	72.2	100	66.7
-20°	88.9	82.9	100	79.4	77.8	74.3	78.3	78.6	66.7
-10°	100	77.8	77.4	100	90.7	88.6	74.1	85	58.3
0°	69.2	65	81.3	88.1	100	87	63.2	77.8	68.8
10°	85.7	58.8	86.7	88.2	84	0	63.2	75	81.3
20°	55.6	82.4	72	75	82.5	92.3	57.3	83.3	80
30°	0	80	69.2	80	62.5	81.5	73	95.5	79.3
40°	60	71.4	61.5	64.3	58.8	55.6	77.4	74.1	100

Table 8.37: Match-matrix results of the percentage of matched keypoints, produced from the synthetically rotated out-of-plane (about pitch axis) 2D data of a **human face**, from -40° to 40° of rotations, using 2D SIFT.



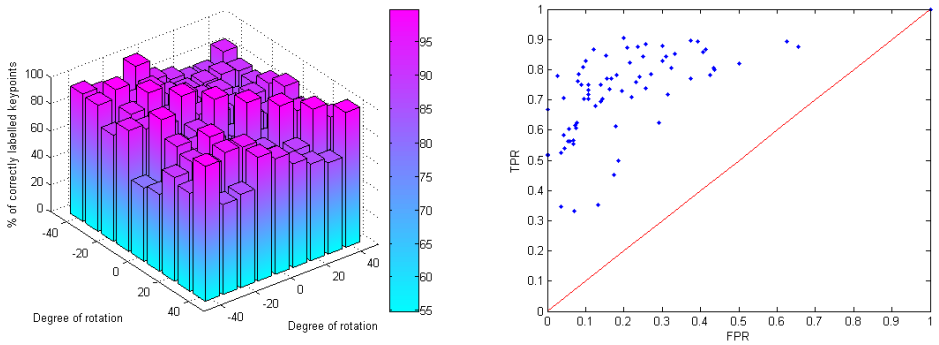
(a) Match-matrix (b) ROC

Figure 8.21: (a) Match-matrix of matched keypoints, derived from a set of synthetically rotated 2D images of a **mannequin head**, about the yaw axis, using 2D SIFT. (b) The matching results presented within a ROC space. This figure can be compared with Figure 8.5, where the results were obtained from the same data set using 2.5D$_{pc}$ SIFT.



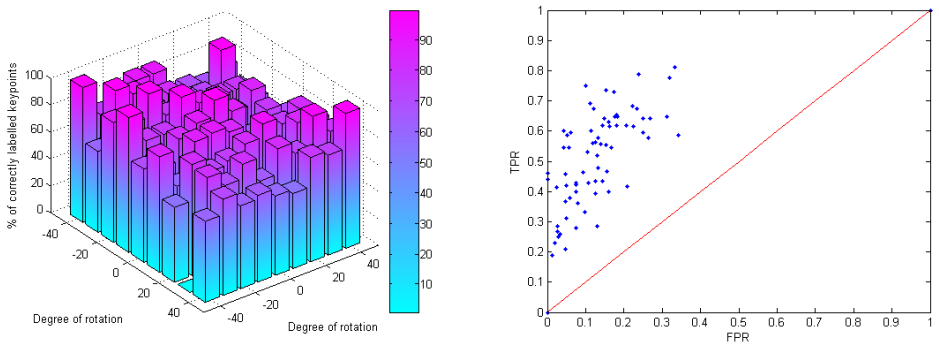
(a) Match-matrix (b) ROC

Figure 8.22: (a) Match-matrix of the matched keypoints, derived from a set of synthetically rotated 2D images of a **human face**, about the yaw axis, using 2D SIFT. (b) The matching results presented within a ROC space. This figure can be compared with Figure 8.6, where the results were obtained from the same data set using 2.5D$_{pc}$ SIFT.



(a) Match-matrix (b) ROC

Figure 8.23: (a) Match-matrix of matched keypoints, derived from a set of synthetically rotated 2D images of a **mannequin head**, about the pitch axis, using 2D SIFT. (b) The matching results presented within a ROC space. This figure can be compared with Figure 8.7, where the results were obtained from the same data set using 2.5D$_{pc}$ SIFT.



(a) Match-matrix (b) ROC

Figure 8.24: (a) Match-matrix of matched keypoints, derived from a set of synthetically rotated 2D images of a **human face**, about the pitch axis, using 2D SIFT. (b) The matching results presented within a ROC space. This figure can be compared with Figure 8.8, where the results were obtained from the same data set using 2.5D$_{pc}$ SIFT.

8.4.2 2D SIFT on Intensity Images: Vary Scale

Similar to the experimental setup outlined in Section 8.3.2, the performance rate of 2D SIFT on 2D images of different scales will be investigated and addressed in this section. Here, the feature descriptors extracted from the baseline 2D images (of size 244×369 pixels) are compared to the feature descriptors extracted from 2D images of the same viewpoint but of a different scale (See Table 8.15 for the sizes of images used). The results are presented in a match-matrix, illustrating the percentage of matched and filtered keypoints. Moreover, the results are also presented in a ROC space, demonstrating the reliability of the matches.

In each summary table, there is a symbol next to the average matching rate, indicating the statistical significance using the Wilcoxon Matched-Pair Signed-Rank test when compared to the results obtained using the 2.5D$_{pc}$ SIFT system. Each symbol can be one of the following: $\ll$, $<$, $=$, $>$ and $\gg$.

A total of 360 2D images of different sizes simulating the in-plane rotations (2×180), 190 captured out-of-plane images of different sizes (2×95) and 180 images of different sizes simulating the out-of-plane rotations about the yaw (2×45) and pitch (2×45) axes respectively are used in this experiment.

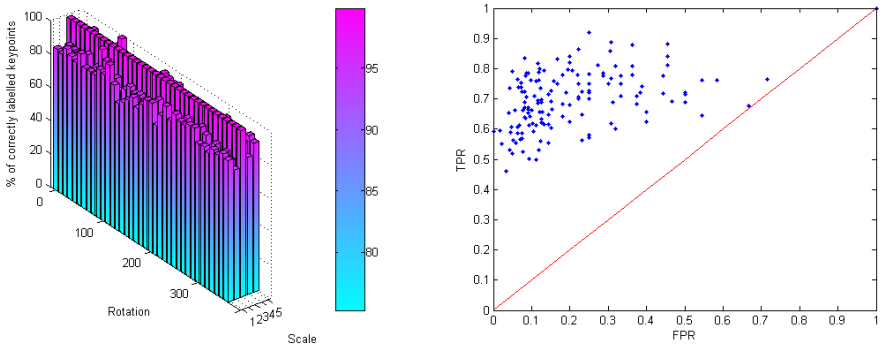
This section is separated into three parts, where the invariance of the feature descriptors extracted from 2D images using 2D SIFT against scale changes will be tested against a) in-plane rotated images, b) real out-of-plane images and finally c) synthetically rotated images.

8.4.2.1 Synthetic In-plane Rotations

The results obtained by comparing the feature descriptors extracted from the baseline images and the feature descriptors extracted from the same viewpoint but of a different sized images are presented in this section where Table 8.38 summarises the results.

System	Subject	Pairings	Average Matching Rate	Images Achieving Matching Rate > 70%	80%	90%	Match-Matrix	ROC
2D	Mannequin head	185	88.1%$\gg$	100%	84.9%	32.4%	Table 8.39 Figure 8.25	Figure 8.25
	Human head	185	84.7%$=$	97.3%	67.6%	22.7%	Table 8.40 Figure 8.26	Figure 8.26

Table 8.38: Results obtained by matching the feature descriptors extracted from different sized synthetically rotated in-plane 2D images of a mannequin head and a human face, to the feature descriptors extracted from the baseline images, using 2D SIFT.



(a) Match-matrix (b) ROC

Figure 8.25: 2D SIFT: (a) The match-matrix results of the percentage of matched keypoints, produced by comparing the feature descriptors extracted from a set of 2D in-plane rotational data of a **mannequin head** (from 0° at 10° increments in the clockwise direction up to 350°) of size 244×369 pixels and the range images captured at a different scale. (b) The matching results illustrated in ROC space. This figure can be compared to Figure 8.9, where the results were obtained from the same data set using 2.5D$_{pc}$ SIFT.

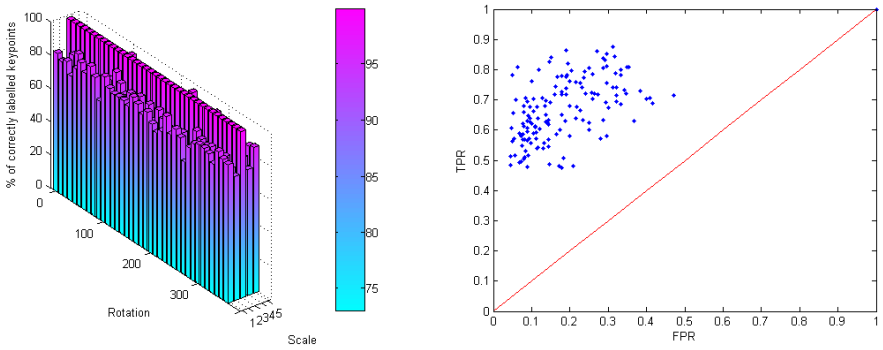
Comparing these results with the results obtained by matching the feature descriptors extracted from in-plane rotational 2.5D images of different sizes using the 2.5D$_{pc}$ SIFT (Table 8.16), the performance of 2.5D$_{pc}$ SIFT is, on average, approximately 23.6% better than the performance of 2D SIFT at 90% of matched keypoint recognition rate. This shows that the 2.5D$_{pc}$ SIFT is more stable to scale changes (w.r.t. in-plane rotations).

	494×744	**370×557**	**244×369**	**182×275**	**119×181**
0°	85	77.8	100	82.4	89.5
10°	83.9	87.3	100	88.9	76.5
20°	88.7	86.7	100	75	94.1
30°	86.2	86.2	100	86.1	88.2
40°	84.6	75.4	100	85.7	76.2
50°	91.4	89.5	100	83.9	81.3
60°	84.6	80.4	100	85.2	80
70°	84.4	85.7	100	81.3	93.8
80°	84.3	98	100	88	100
90°	89.3	93	100	88.5	86.7
100°	88.2	84.9	100	80.6	85.7
110°	95.7	86.4	100	81.5	91.7
120°	85.7	80.8	100	72.4	87.5
130°	76.5	83.3	100	87.5	77.8
140°	78.9	86.8	100	88	87.5
150°	82.1	84.6	100	72	78.6
160°	85.5	85.7	100	76.9	85.7
170°	81	83.8	100	83.3	95.8
180°	85.1	79.6	100	88.2	94.4
190°	89.5	86	100	90.9	93.3
200°	89.7	90	100	77.4	94.1
210°	78	82.1	100	87.9	88.2
220°	85.9	73	100	72.2	72.7
230°	93.3	82.8	100	90	77.8
240°	86.5	83.7	100	88.9	85.7
250°	92.2	87.8	100	78.1	88.2
260°	89.6	90	100	83.3	91.7
270°	89.2	81.6	100	96	81.3
280°	91.1	87.5	100	82.4	87.5
290°	91.5	90.5	100	77.8	91.7
300°	83.6	87.5	100	76.9	93.3
310°	83.9	81.8	100	74.1	85.7
320°	82	85.7	100	85.7	75
330°	85.7	78.2	100	78.3	81.3
340°	84.5	84.7	100	85.2	84.6
350°	87.1	81.9	100	89.5	92
360°	85	77.8	100	82.4	89.5

Table 8.39: Match-matrix results of the percentage of matched keypoints by comparing the feature descriptors obtained from the 2D in-plane rotational data of a **mannequin head** (from 0° at 10° clockwise increments up to and including 360°) of size 244×369 pixels and the feature descriptors extracted from the same viewpoint but of different scales, using 2D SIFT.

	494×744	**370×557**	**244×369**	**182×275**	**119×181**
0°	83.1	74.5	100	75.6	87.5
10°	80	73.6	100	88.4	80
20°	82	84.2	100	86.3	85.2
30°	75	86.7	100	82.9	82.9
40°	81.3	80.6	100	88.4	72.7
50°	86.7	77.2	100	88	81.3
60°	82.3	76.3	100	83.7	88.5
70°	88.2	76.5	100	81.6	76.5
80°	83.6	75.9	100	81.6	78.3
90°	71.2	70.9	100	83.8	63.2
100°	87.5	84.3	100	89.7	94.4
110°	82.4	90.6	100	87.8	90.9
120°	80.6	85.7	100	77.3	80.8
130°	80.9	80.9	100	84.4	80.8
140°	80.7	83.6	100	88.4	78.9
150°	79.1	81	100	80	81.8
160°	86.3	85.2	100	78	69.2
170°	84.1	76.8	100	77.8	88.5
180°	79.7	84.2	100	88.1	75
190°	83.1	76.3	100	79.5	71.4
200°	77.8	84.5	100	81.6	80.6
210°	74.3	81	100	79.5	84.4
220°	77.1	77.9	100	76	76.5
230°	87.9	78	100	83.3	96.2
240°	78.8	72.7	100	77.1	84
250°	83.9	80.3	100	84.2	69
260°	80.7	70.2	100	86.5	78.3
270°	67.9	72.5	100	80.5	70.8
280°	78.3	86.8	100	83.8	93.3
290°	82.4	83.6	100	84.2	84.2
300°	83.8	81.8	100	70.2	78.6
310°	83.6	82.9	100	78.7	77.8
320°	82.1	74.6	100	88.6	63.2
330°	77.9	80.6	100	83	79.3
340°	87.2	82.7	100	78.9	73.1
350°	84.4	77.3	100	81	86.4
360°	83.1	74.5	100	75.6	87.5

Table 8.40: Match-matrix results of the percentage of matched keypoints by comparing the feature descriptors obtained from the 2D in-plane rotational data of a **human face** (from 0° at 10° clockwise increments up to and including 360°) of size 244×369 pixels and the feature descriptors extracted from the same viewpoint but of different scales, using 2D SIFT.



(a) Match-matrix (b) ROC

Figure 8.26: 2D SIFT: (a) The match-matrix results of the percentage of matched keypoints, produced by comparing the feature descriptors extracted from a set of 2D in-plane rotational data of a **human face** (from 0° at 10° increments in the clockwise direction up to 350°) of size 244×369 pixels and the range images captured at a different scale. (b) The matching results illustrated in ROC space. This figure can be compared to Figure 8.10, where the results were obtained from the same data set using 2.5D$_{pc}$ SIFT.

8.4.2.2 Out-of-plane Rotations on Real Data

Table 8.41 summarises the results obtained by matching the feature descriptors extracted from the baseline 2D images (of size 244×369 pixels) with the feature descriptors extracted from the 2D images of the same out-of-plane rotations but of different scales.

System	Subject	Pairings	Average Matching Rate	Images Achieving Matching Rate > 70%	80%	90%	Match-Matrix	ROC
2D	Mannequin head	95	85.7%$^{=}$	95.8%	65.3%	31.6%	Table 8.42 Figure 8.27	Figure 8.27
	Human head	95	84.6%$^{\ll}$	91.6%	62.1%	29.5%	Table 8.43 Figure 8.28	Figure 8.28

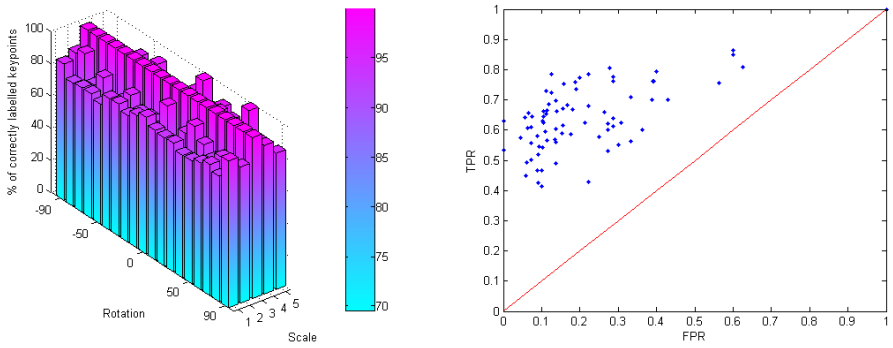
Table 8.41: Results obtained by exploring matching the feature descriptors extracted from different sized rotated out-of-plane 2D images of a mannequin head and a human face, compared to the baseline images, using 2D SIFT.

	494×744	**370×557**	**244×369**	**182×275**	**119×181**
-90°	84.8	88.6	100	90.5	78.6
-80°	76.5	92	100	83.3	75
-70°	77.3	76.3	100	85.2	69.2
-60°	77.4	84.4	100	72.5	88.9
-50°	74.6	72.7	100	89.5	75
-40°	81.7	77.3	100	78	91.7
-30°	82.1	84.6	100	80.4	76
-20°	78.6	77.6	100	88.1	77.3
-10°	82.3	78.6	100	87.9	73.9
0°	85	79.6	100	82.4	89.5
10°	80.5	92.9	100	89.3	100
20°	79.6	69.4	100	83.8	85.7
30°	80	83.3	100	86.2	90.9
40°	76.1	91.3	100	94.7	86.7
50°	76.4	83.6	100	77.8	69.2
60°	78.7	78	100	77.8	100
70°	81.6	83	100	85	84.6
80°	78	82.5	100	66.7	75
90°	91.2	84.6	100	92.3	84.6

Table 8.42: Match-matrix results of the percentage of matched keypoints by comparing the feature descriptors obtained from a set of 2D out-of-plane rotational data of a **mannequin head** (from -90° at 10° clockwise increments up to and including 90°) of size 244×369 pixels and the feature descriptors extracted from the same viewpoint but of different scales, using 2D SIFT.

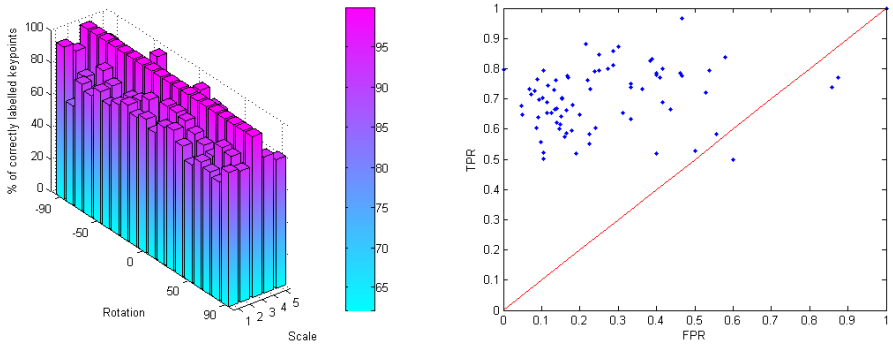
	494×744	**370×557**	**244×369**	**182×275**	**119×181**
-90°	81.8	75	100	100	81.8
-80°	90.9	88.9	100	81.8	0
-70°	91.7	87.5	100	94.1	85.7
-60°	100	95.5	100	92.6	66.7
-50°	100	90	100	87	90.9
-40°	100	93.1	100	91.7	90.9
-30°	88.9	80	100	95.2	75
-20°	81.3	82.6	100	95.5	100
-10°	81.3	90.6	100	89.3	75
0°	95.5	90.9	100	87.5	90
10°	94.4	90.6	100	89.3	92.3
20°	93.3	83.3	100	85.2	55.6
30°	72.7	94.7	100	90.5	92.3
40°	66.7	76.7	100	100	75
50°	84.6	75	100	87.5	60
60°	100	76.9	100	100	66.7
70°	64.3	93.3	100	95.7	87.5
80°	93.8	77.8	100	87	75
90°	78.6	85	100	73.3	66.7

Table 8.43: Match-matrix results of the percentage of matched keypoints by comparing the feature descriptors obtained from a set of 2D out-of-plane rotational data of a **human face** (from -90° at 10° clockwise increments up to and including 90°) of size 244×369 pixels and the feature descriptors extracted from the same viewpoint but of different scales, using 2D SIFT.



(a) Match-matrix (b) ROC

Figure 8.27: 2D SIFT: (a) Match-matrix showing the results obtained by comparing the feature descriptors extracted from a set of 2D out-of-plane data of different sizes to the feature descriptors extracted from their respective baseline images of a **mannequin head**. (b) The results shown in a ROC space. This figure can be compared with Figure 8.11, where the results were obtained from the same data set using 2.5D$_{pc}$ SIFT.



(a) Match-matrix (b) ROC

Figure 8.28: 2D SIFT: (a) Match-matrix showing the results obtained by comparing the feature descriptors extracted from a set of 2D out-of-plane data of different sizes to the feature descriptors extracted from their respective baseline images of a **human face**. (b) The results shown in a ROC space. This figure can be compared with Figure 8.12, where the results were obtained from the same data set using 2.5D$_{pc}$ SIFT.

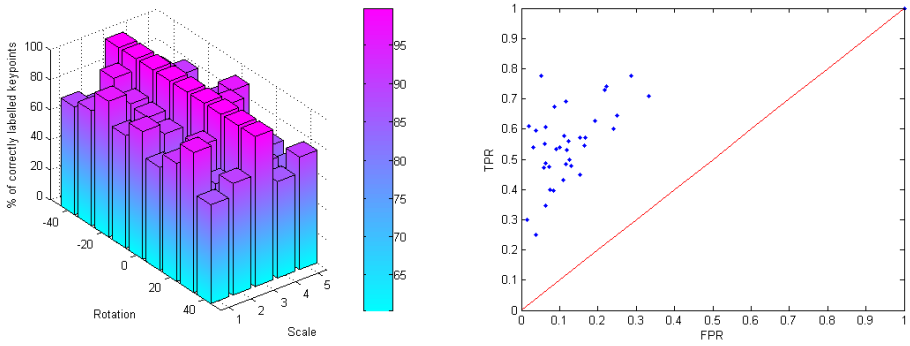
By comparing the performance rate of the 2D SIFT to that of the 2.5D$_{pc}$ SIFT against scale (Table 8.19), on average, 2.5D$_{pc}$ outperform 2D SIFT by approximately 17.4% at 90% of keypoint recognition rate. This demonstrates that the feature descriptors extracted from 2.5D SIFT show a higher invariance to scale changes, w.r.t. out-of-plane rotational images.

8.4.2.3 Synthetic Out-of-plane Rotations

Table 8.44 summarises the results obtained by matching the feature descriptors extracted from different sized synthetically rotated (about yaw and pitch axes respectively) 2D images of a mannequin head and a human face to the feature descriptors extracted from the baseline 2D images.

Subject	Axis	Pairings	Average Matching Rate	Images Achieving Matching Rate >			Match-Matrix	ROC
				70%	80%	90%		
Mannequin head	yaw	45	80.3%=	71.1%	44.4%	28.9%	Table 8.45 Figure 8.29	Figure 8.29
Human head	yaw	45	83.2%>	88.9%	60%	28.9%	Table 8.46 Figure 8.30	Figure 8.30
Mannequin head	pitch	45	84.8%=	88.9%	71.1%	42.2%	Table 8.47 Figure 8.31	Figure 8.31
Human head	pitch	45	85.8%=	91.1%	66.7%	40%	Table 8.48 Figure 8.32	Figure 8.32

Table 8.44: Results obtained by matching the feature descriptors extracted from different sized synthetically rotated out-of-plane range images of a mannequin head and a human face and the feature descriptors extracted from the baseline images, using 2D SIFT.



(a) Match-matrix (b) ROC

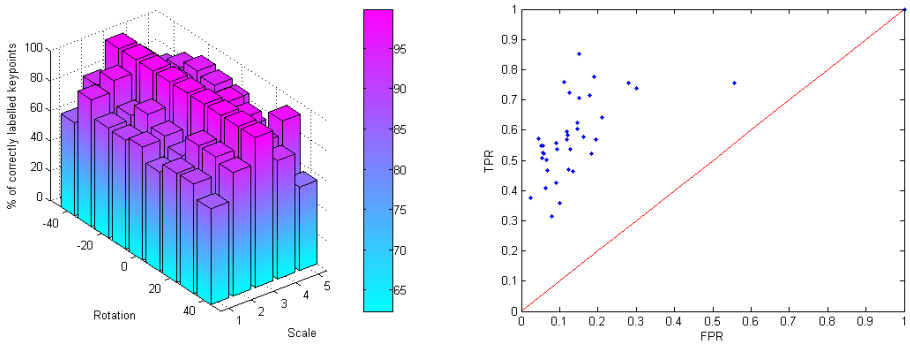
Figure 8.29: 2D SIFT: (a) Match-matrix showing the percentage of matched keypoints, produced by comparing the feature descriptors extracted from a set of different sized 2D synthetically out-of-plane rotational data (about the yaw axis) of a **mannequin head** (from -40° at 10° increments in the clockwise direction up to 40°) and the feature descriptors extracted from their baseline images. (b) The matching results illustrated in ROC space. This figure can be compared with Figure 8.13, where the results were obtained from the same data set using 2.5D$_{pc}$ SIFT.

	494×744	**370×557**	**244×369**	**182×275**	**119×181**
0°	71.8	66.7	100	65.4	60
0°	76.7	92.6	100	85.3	77.8
0°	90.4	83	100	79.4	66.7
0°	75	81.4	100	79.3	75
0°	85	77.8	100	82.4	89.5
0°	68.6	59.1	100	93.1	60
0°	78.7	84.4	100	71.4	66.7
0°	93.3	66.7	100	68.8	60
0°	65.9	74.4	100	64.3	75

Table 8.45: Match-matrix results of the percentage of matched keypoints by comparing the feature descriptors obtained from 2D synthetically out-of-plane rotational data (about the yaw axis) of a **mannequin head** (from -40° at 10° clockwise increments up to and including 40°) of sized 244×369 pixels and the feature descriptors extracted from the same viewpoint but of different scales, using 2D SIFT.

	494×744	**370×557**	**244×369**	**182×275**	**119×181**
-40°	62	81.4	100	85.2	75
-30°	84.6	91.8	100	77.4	83.3
-20°	73.3	76	100	85.7	88.5
-10°	72.6	84.1	100	82.9	86.4
0°	75.4	76.9	100	80	84
10°	66.1	71.2	100	95.5	73.9
20°	72.7	85.4	100	92.6	66.7
30°	73	85.3	100	83.3	92.3
40°	63.9	82.4	100	78.9	55.6

Table 8.46: Match-matrix results of the percentage of matched keypoints by comparing the feature descriptors obtained from 2D synthetically out-of-plane rotational data (about the yaw axis) of a **human face** (from -40° at 10° clockwise increments up to and including 40°) of size 244×369 pixels and the feature descriptors extracted from the same viewpoint but of different scales, using 2D SIFT.



(a) Match-matrix (b) ROC

Figure 8.30: 2D SIFT: (a) The match-matrix results of the percentage of matched keypoints, produced by comparing the feature descriptors extracted from a set of 2D synthetically out-of-plane rotational data (about the yaw axis) of a **human face** (from -40° at 10° increments in the clockwise direction up to 40°) of size 244×369 pixels and their 2D images of a different scale. (b) The matching results illustrated in ROC space. This figure can be compared to Figure 8.14, where the results were obtained from the same data set using 2.5D$_{pc}$ SIFT.

By comparing the performance rates between the 2D SIFT and 2.5D$_{pc}$ SIFT against scale changes using synthetically rotated images about the yaw axis (Table 8.22) show that the performance rates of the two systems are comparable. However, closer examination of the matrices reveal that 2.5D$_{pc}$ SIFT outperforms 2D SIFT marginally, by approximately 6.8% at 90% keypoint recognition rate.

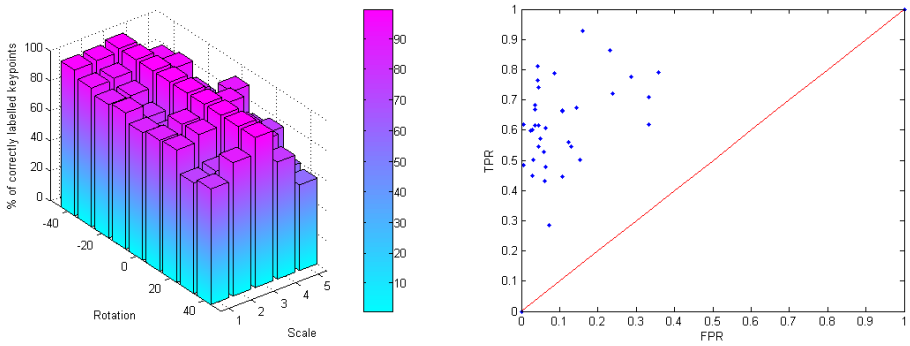
The performance of the feature descriptors extracted from synthetically 2D rotated (about the pitch axis) images using 2D SIFT against scale changes are comparable to the performance of the feature descriptors extracted from the range data using 2.5D$_{pc}$ SIFT. Close examination shows that, on average, 2D SIFT outperforms 2.5D$_{pc}$ SIFT (Table 8.22) by 0.4% at 90% keypoint recognition rate.

	494×744	**370×557**	**244×369**	**182×275**	**119×181**
-40°	97.1	97.7	100	82.1	0
-30°	92.5	93.3	100	93.8	66.7
-20°	88.9	89.1	100	96.3	71.4
-10°	90.7	82.1	100	87.2	68.4
0°	85	77.8	100	82.4	89.5
10°	87.3	92.3	100	94.9	72.2
20°	88.2	92.9	100	86.1	76.5
30°	73.7	75.9	100	84.6	60
40°	77.1	88.9	100	78	57.1

Table 8.47: Match-matrix results of the percentage of matched keypoints by comparing the feature descriptors obtained from 2D synthetically out-of-plane rotational data (about the pitch axis) of a **mannequin head** (from -40° at 10° clockwise increments up to and including 40°) of size 244×369 pixels and the feature descriptors extracted from the same viewpoint but of different scales, using 2D SIFT.

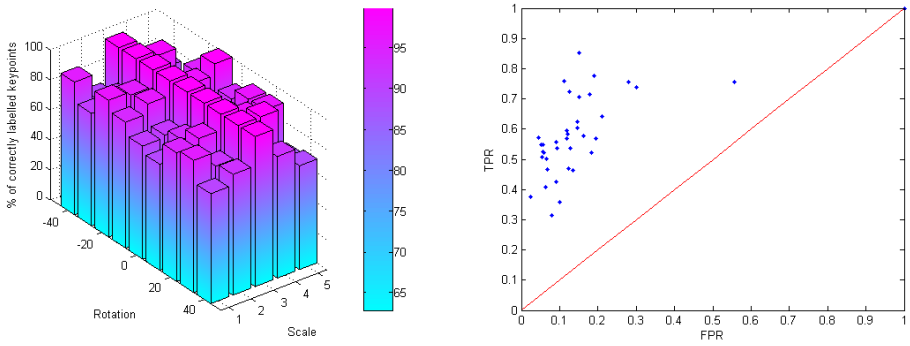
	494×744	**370×557**	**244×369**	**182×275**	**119×181**
-40°	88.4	71.7	100	65.2	75
-30°	75	79.1	100	96.6	62.5
-20°	91.1	88.9	100	72.5	86.7
-10°	83.9	90.4	100	92.5	100
0°	75.4	76.9	100	80	84
10°	70.7	81.4	100	85.4	91.3
20°	85.7	89.6	100	92.3	91.7
30°	88.1	78.6	100	100	66.7
40°	73.3	80.5	100	81.1	69.2

Table 8.48: Match-matrix results of the percentage of matched keypoints by comparing the feature descriptors obtained from 2D synthetically out-of-plane rotational data (about the pitch axis) of a **human face** (from -40° at 10° clockwise increments up to and including 40°) of size 244×369 pixels and the feature descriptors extracted from the same viewpoint but of different scales, using 2D SIFT.



(a) Match-matrix (b) ROC

Figure 8.31: 2D SIFT: (a) Match-matrix showing the percentage of matched keypoints, produced by comparing the feature descriptors extracted from a set of different sized 2D synthetically out-of-plane rotational data (about the pitch axis) of a **mannequin head** (from -40° at 10° increments in the clockwise direction up to 40°) and the feature descriptors extracted from their baseline images. (b) The matching results illustrated in ROC space. This figure can be compared with Figure 8.15, where the results were obtained from the same data set using 2.5D$_{pc}$ SIFT.



(a) Match-matrix (b) ROC

Figure 8.32: 2D SIFT: (a) The match-matrix results of the percentage of matched keypoints, produced by comparing the feature descriptors extracted from a set of 2D synthetically out-of-plane rotational data (about the pitch axis) of a **human face** (from -40° at 10° increments in the clockwise direction up to 40°) of size 244×369 pixels and their 2D images of a different scale. (b) The matching results illustrated in ROC space. This figure can be compared with Figure 8.16, where the results were obtained from the same data set using 2.5D$_{pc}$ SIFT.

8.5 Summary and Discussions

The validation process and results of the pose estimated and corrected version of 2.5D SIFT are presented in this chapter, in which the aim of the validation was to investigate the invariance properties of the feature descriptors to rotational and scale changes. If any planar surface patch is rotated to a different viewpoint orientation, the area representing this rotated patch will be reduced (or enlarged, depending on the degree of orientation), compared to the original patch. As a result, by warping the circular Gaussian measurement aperture into an elliptical shape, according to the local slant and tilt of the keypoint location, it allows the same underlying surface area to be sampled.

The validation of the $2.5D_{pc}$ SIFT was separated into two main categories: firstly the invariance properties of the extracted feature descriptors against in-plane and out-of-plane rotational data was addressed. Here, the data sets employed were of fixed size (244×369 pixels), where each range images was rotated about all three Euler's axes. Secondly, the invariance properties of the extracted feature descriptors against scale changes was addressed, in which the feature descriptors extracted from different sized range images were matched to the feature descriptors extracted from the baseline (244×369 pixels) images of the same viewpoint orientation. Each pairwise combination of the extracted feature descriptors were examined in this validation, providing a match-matrix of results presenting the percentage of correctly matched keypoints for each pairing. Moreover, these results were presented in a ROC space, thereby providing a visual measurement of the reliability of the matches.

Apart from the validation of the $2.5D_{pc}$ SIFT, the performance rate between this system on range images and the performance rate of standard 2D SIFT on 2D images were compared. Similar to the validation of the $2.5D_{pc}$ SIFT, the standard 2D SIFT was applied to the 2D images captured at different viewpoint orientations, as well as of different sizes, allowing comparison of the two platforms to be made.

Table 8.49 summarises the results obtained in this chapter in which the average results obtained from the human face and the mannequin head data are presented. Here, the symbols next to the average matching rate obtained from the $2.5D_{pc}$ SIFT system denotes the statistical significance differences to the 2D SIFT and 2.5D SIFT (where applicable) respectively. The statistical significance tests were conducted using the

Wilcoxon Matched-Pair Signed-Rank test and the symbols denote one of the following: $\ll$, $<$, $=$, $>$ and $\gg$.

This table illustrates that $2.5D_{pc}$ SIFT outperforms the standard 2D SIFT, with respect to out-of-plane viewpoint rotational and scale changes. Moreover, when the results were plotted in a ROC space, it can be deduced that the performance of $2.5D_{pc}$ SIFT is more reliable than the performance of 2D SIFT, where the FPR of the $2.5D_{pc}$ SIFT is smaller, compared to the FPR of the 2D SIFT.

By taking averages of the match-matrices of the matched keypoints, the performance rate of the standard 2D SIFT against in-plane rotational changes outperforms the performance of $2.5D_{pc}$ SIFT against in-plane rotational changes. However, closer examination of the results plotted in a ROC space illustrates that the FPR rate is higher in the 2D SIFT.

		Fixed Scales			**Vary Scales**			
Rotation	**Modality**	**Average Matching Rate**	**Images Achieving Matching Rate >**		**Average Matching Rate**	**Images Achieving Matching Rate >**		
			50%	**60%**		**70%**	**80%**	**90%**
Synthetic In-Plane	2D	84.0%	98.7%	95.6%	84.6%	98.7%	76.3%	27.6%
	2.5D	53.4%	74.3%	48.5%	N/A	N/A	N/A	N/A
	$2.5D_{pc}$	55.1%$\ll\gg$	76.1%	53.3%	83.9%$\ll$	91.6%	76.2%	51.1%
Real Out-Plane	2D	34.6%	45.3%	36.6%	85.2%	93.7%	63.7%	30.6%
	2.5D	37.0%	49.3%	35.7%	N/A	N/A	N/A	N/A
	$2.5D_{pc}$	46.0%$\gg\gg$	60.6%	85.5%$=$	45.6%	85.8%	71.6%	47.9%
Synthetic Out-Plane (about yaw axis)	2D	78.3%	98.2%	93.2%	81.7%	80%	52.2%	28.9%
	$2.5D_{pc}$	80.6%$=$	100%	95.7%	75.7%$=$	72.3%	50%	35.7%
Synthetic Out-Plane (about pitch axis)	2D	79.2%	98.8%	92.6%	85.3%	90%	68.9%	41.4%
	$2.5D_{pc}$	83.9%$\gg$	100%	97.55%	81.2%$=$	75.6%	60%	33.3%

Table 8.49: Summary of the results obtained of the performance rate of the pose estimated and corrected version of the 2.5D SIFT and standard SIFT, against rotational and scale changes.

These results illustrate the feature descriptors extracted from the $2.5D_{pc}$ SIFT are more stable to out-of-plane rotational changes than 2D SIFT whereas the feature descriptors extracted from 2D SIFT are more stable to in-plane rotational changes. This suggests that combining 2D and $2.5D_{pc}$ methodologies together could potentially provide a feature descriptor that is robust to both in-plane and out-of-plane rotational changes. For instance, the global pose could be estimated from the local pose by noting the common rotation *difference* between matching keypoints. As a result, the global pose could be determined prior to the feature descriptors extraction, thereby allowing

the system to choose between the use of 2D SIFT or $2.5D_{pc}$ SIFT for the feature descriptor extraction process accordingly.

The next chapter will draw conclusions from this research project and give suggestions as how to improve and extend this work.

Chapter 9

Conclusions and Future Work

This chapter serves to summarise the research conducted in this thesis and indicate the significance of this work in light of the current literature. The achievements and limitations of this work are addressed and the chapter concludes with potential directions for future work in which this thesis initiated.

9.1 Objectives and Thesis Hypothesis Revisited

The goal of this work is to build and validate a system to interpret range images by matching local features. Initially, no assumptions were made as to what types of features are required for the representation. The broader goal of this work is to progress towards working in the context of medical surface anatomy assessment, thereby potentially providing a system applicable to 3D clinical photography.

The following objectives were set for this work:

- To determine what local surface representation to adopt for the range manifold, in order to progress automatic analysis of the surface anatomy based on local descriptors matching.
- To improve the stability of the matches obtained by exploiting the potential for invariance to illumination variations and pose that the use of range maps might confer.
- To construct a complete engine in which the above concepts have been embodied, where the performance can be measured and validated against 2D approaches.

To this end, this thesis presents a methodology for extracting stable feature descriptors from 2.5D range images in order to provide point-based correspondences between range surfaces. Lowe's 2D SIFT (Lowe, 2004) has been adapted into the 2.5D domain in which feature descriptors comprising local topological information are extracted. Figure 9.1 illustrates the 2.5D SIFT system presented in this thesis.

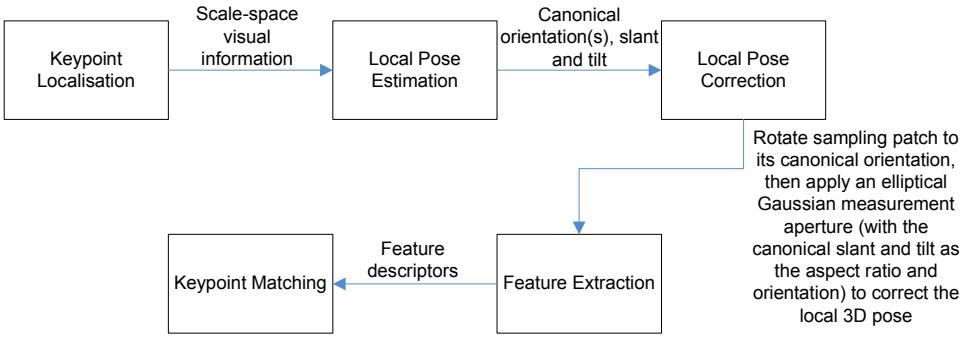


Figure 9.1: Implemented 2.5D SIFT.

At the beginning of the this thesis, the following hypothesis was made:

> "This thesis argues that by exploiting statistical representations of local surface properties, range images can be represented and matched based on direct recovery of the local surface topology sampled at discrete locations on the range manifold. The use of range data allows surface normals to be recovered directly, such that the pose of the locally sampled descriptors can be estimated and corrected. The above advantages, along with the lighting and pose invariant properties of range images, imply that 2.5D image interpretation techniques can potentially offer improvements over standard 2D techniques."

The objectives have been achieved and the hypothesis has been verified in the course of this thesis. These are summarised in the next section.

9.2 Contributions

The following contributions of this work addressed the initial objectives and hypothesis:

- **Major Contributions:**
  - Formulation of a feature descriptor, comprising local surface topologies using the shape index and the local surface orientations using image gradients, for range image analysis;
  - Formulation and investigation of a 2.5D SIFT range image matching algorithm, where the aforementioned feature descriptor is embedded;
  - Development of a pose estimated and corrected version of 2.5D SIFT system, thereby potentially improving the invariance of the feature descriptors to range images exhibiting out-of-plane viewpoint rotational changes.
- **Minor Contributions:**
  - Improvement of the accuracy of the keypoint localisation of the SIFT system;
  - Improvement of the in-plane orientation estimates of the SIFT system.

9.2.1 A Novel Feature Descriptor for Range Image Analysis

Traditionally the extraction of 3D surface types, derived by either the signs of H and K curvatures or the shape index, are used to categorise the underlying surface for a range image. This is an effective approach since measurements can be made directly on the range images. Moreover, since the shape of an imaged surface remains constant under viewpoint changes, the local distribution of the curvatures observed on its surface should therefore remain constant, thereby providing a feature descriptor that is potentially invariant to viewpoint changes. In this work, the traditional feature extraction technique was extended by taking the local distributions of the surface types, along with the local distributions of the orientations, into account simultaneously.

Initially, the signs of H and K curvatures were employed to categorise the surface into eight surface types. However, it was noted that these eight surface types were not exploited fully, since no surface on a human face is truly "flat" unless an user-defined threshold is introduced to the assignment of the H and K curvatures. As a result, alternative methodology in extracting the surface types was required. The shape index

was considered, where the surfaces were classified into nice surface types. This is more suitable for face analysis, due to the continuity typical of facial surfaces.

The second component in the feature descriptor takes the orientation of the range images into account. Initially, the direction estimates obtained from the $k1$ curvature were employed, however, these estimates were based on the second order Gaussian derivatives and thus the error rate introduced is higher than the direction estimates obtained from lower order derivatives. Therefore, the orientation estimates, derived from the image first order gradients, were employed. In order to achieve viewpoint rotational invariance, the orientation component histogram was normalised to its canonical form (i.e. rotating the orientation histogram until the peak is located at the first histogram position).

By capturing the underlying information of the surface types and their orientations simultaneously, a basic feature descriptor was formulated. Different formulations of the feature descriptor were investigated, including a coupled 2D feature descriptor. However, it was noted that the performance of the 2D descriptor was only marginally better than the basic 1D feature descriptor. This could be explained by the "curse of dimensionality" (Bellman, 1957; Duda and Hart, 1973). As a result, the final structure of the feature descriptor employed in this work is based on histogramming the shape index of the underlying surface and is weighted by the degree of local surface curvedness. A second component of the keypoint descriptor comprises the histogrammed local orientations within the same patch, weighted by the local gradient magnitude values. Therefore the keypoint signatures based on surface shape, degree of curvature and the relative orientations have been captured simultaneously. This form of feature descriptor was validated on range images of human faces. However, it has not been validated on range images of other types of subjects except for synthetic data created from orthogonal spatial sine waves.

9.2.2 Development of a 2.5D SIFT System

Given the availability of the 2D SIFT system, where feature descriptors comprising the image gradients of the image are extracted from each keypoint location, the development of a 2.5D SIFT system by incorporating the 2.5D feature descriptor comprising the local topological information of the surfaces was discussed in this thesis. This development is believed to be the first in the current literature, since no other such system

has been reported to date. Figure 9.2 recapitulates the stages involved in the proposed 2.5D SIFT system.

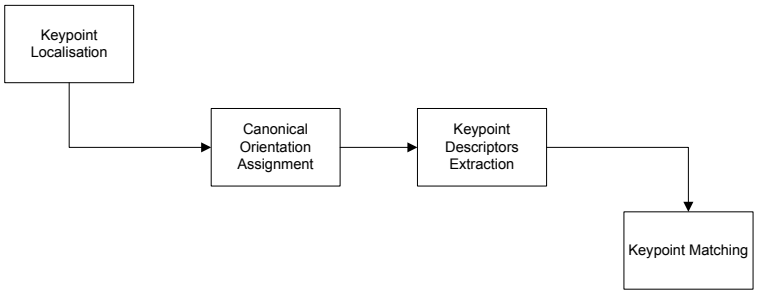


Figure 9.2: Pipeline of the 2.5D SIFT system.

By adopting statistical normalisation of the input range images, it becomes possible to set a consistent set of parameters appropriate to detecting stable keypoint locations and their appropriate scales (independently of the dynamic range of the input range maps or their content).

A canonical orientation (to the nearest degree) at each keypoint sampling location can then be extracted, and by rotating the sampling patch to this canonical orientation, viewpoint rotation can be achieved. Therefore, each individual histogram of the orientation component will no longer be required to be normalised to its canonical form.

In order to capture a perceptually significant description of the local surface patches sampled at keypoint locations, the 2.5D keypoint descriptors had been formulated to sample the underlying relative frequencies of surface types present. This form of feature descriptor is capable of capturing not only the topological information of a surface, but also capable of differentiating between mirror keypoints on a range image, thereby providing an unique identification to represent the underlying surface at each keypoint location. Furthermore, potential sampling effects caused by spatial aliasing within the standard SIFT keypoint descriptor have been minimised by placing nine Gaussian circular sub-regions, with spatial support of one standard deviation, over each sampled keypoint location at the detected scale. The aim of this feature descriptor is to increase the invariance properties of the feature descriptor to both Euler's in-plane and out-of-plane rotations.

This basic 2.5D SIFT system was validated using both in-plane and out-of-plane rotated range images. Each pairwise combinations of the images were compared, thereby

formulating a match-matrix used to interpret the obtained results. The results obtained showed that the feature descriptors show good invariance to in-plane and out-of-plane rotations.

9.2.2.1 The Development of a Pose Estimated and Corrected Version of 2.5D SIFT

The standard 2D SIFT relies only on the in-plane orientation to estimate and correct object pose. However, with the availability of 2.5D range images, it is possible to estimate the 3D pose based on computing the local slant and tilt. Since the relative viewpoint orientation at the keypoint sampling location is not known a priori, it is necessary to take the degree of out-of-plane rotation exhibited at each keypoint into account. Moreover, if an approximately planar surface patch is rotated to a different viewpoint orientation, the area representing this rotated patch will be reduced (or enlarged, depending on the degree of orientation), compared to the original patch. Hence, if no estimation of the local slant and tilt is taken into account, the keypoint matching will result in comparing different regions of the surface. To this end, the thesis also developed a pose estimated and corrected version of the 2.5D SIFT based on the local slant and tilt information extracted at each sampling location.

By warping the circular Gaussian measurement aperture into an elliptical shape based on the local slant and tilt, it ensured the same consistent patches were extracted from different viewpoint orientated images of the same surface, resulting in a more reliable matching between feature descriptors extracted from out-of-plane rotated range images.

The pose estimated and corrected version of the 2.5D SIFT was validated using both in-plane and out-of-plane rotational images. The out-of-plane invariance of the feature descriptors were noticed to have improved. The scale invariance properties were addressed as well. Moreover, a comparison between the performance of the standard 2D SIFT and the pose corrected version of the 2.5D SIFT was made. Whilst the performance of feature descriptors extracted from this pose estimated and corrected version of 2.5D SIFT outperforms the feature descriptors extracted from the standard 2D SIFT with respect to out-of-plane viewpoint rotated images, the performance of 2D SIFT on in-plane rotated images slightly outperformed the performance of the 2.5D SIFT.

9.2.3 The Improvement of the Keypoint Localisation Technique

In order to extend the 2D SIFT scale-space keypoint localisation stage into the 2.5D domain, necessary adaptation was made to accommodate the range imaging modality. Furthermore, the accuracy of the curvatures extraction within the keypoint localisation stage was improved by using the first and second Gaussian derivatives parameterised with the corresponding σ, instead of using the Hessian matrix. This provided a more stable range surface gradient estimates by employing the Gaussian smoothing in the calculation of the derivatives.

However, the computational cost of deriving the curvatures using the Gaussian derivatives is slighter higher than that computed from the Hessian matrix, since it requires a 2D image filtering process using convolution.

9.2.4 The Improvement of In-plane Orientation Estimates of SIFT

The accuracy of the in-plane orientation estimates of SIFT was improved in the 2.5D SIFT by employing 360 bins in the formulation of the orientation histogram, instead of using 36 bins as proposed by Lowe (2004). This allows the estimated orientation to be made accurate to within $\approx \pm 1.0^{\circ}$. Moreover, the values of the orientation histogram were stabilised, in terms of orientation continuity, by distributing each entry over a number of bins by means of a 1D symmetric Gaussian convolution kernel of $\sigma \approx$7-bins. This step anti-aliased the orientation histogram and stabilised the keypoint canonical orientation allocation process by providing estimates of orientation that change smoothly as the input visual stimulus changes in orientation and is confirmed by experiment.

9.3 Future Work

This work provides an initial investigation into the development of a 2.5D SIFT system to present and analyse range images of human faces. Further development could be made to improve the stability and accuracy of the system and also to widen its application, as proposed below:

9.3.1 Global Pose Estimator

The latter part of this work has been focused on local pose estimation and correction. However, it is possible to recover the global pose of an object depicted within a range

image w.r.t. a set of exemplars . The global pose could be estimated from the local pose by noting the common rotation *difference* between the matching keypoints of the unknown instance and those of the exemplar. In other words, each keypoint recovered from a range image might each have a different local slant and tilt, however, the dominant difference in the slant and tilt between matching keypoints between range images would give the relative pose between them, hence the global pose could be recovered, with the use of the Hough Transform.

9.3.2 Invariance of the Feature Descriptors to Noise

It was outwith the scope of this work to address the sensitivity of the feature descriptors w.r.t. noise. In order to deduce the susceptibility of the feature descriptor to noise, a known amount of random noise could be added to the range images. The 2.5D SIFT system could then be applied to the images to determine the keypoint matching rate by plotting ROC curves w.r.t. the signal-to-noise ratio (SNR). Thereby a comparison between the baseline (original, no noise added) image and the noise-induced images could be made.

Moreover, the robustness of the 2.5D matching process could be determined by collecting a data set comprising images captured under different lighting settings, resulting in 2D images that are affected by lighting whereas the range images should remain unchanged.

9.3.3 Improving the 2.5D SIFT System

This section provides a few suggestions to improve the current 2.5D SIFT system.

9.3.3.1 Multiple Canonical Slant and Tilt

One of the limitations of the current version of the 2.5D SIFT is that it contains only one canonical slant and one canonical tilt for each keypoint location by the assumption of the approximate local planarity in this work. Further investigation is therefore required to address this issue, where multiple canonical slants and canonicals tilt may potentially improve the matching performances of the feature descriptors. The multiple canonical slants and canonical tilts would result in multiple feature descriptors for each keypoint location.

Multiple canonical slants and tilts could be assigned to each keypoint location by means of iterating each slant and tilt histograms peak to within 80% (similar to the assignment of the canonical orientation).

9.3.3.2 Independent Means of Verifying Keypoint Matches

Currently, the keypoint matches are verified by computing a similarity transform between the two sets of descriptors by means of the Hough Transform. However, the HT itself is not the only methodology for verifying keypoint matches. Therefore an investigation could be conducted using other means of verifying the keypoint matches. For example, the location of the keypoints on the rotated images could be computed (by rotating the baseline image), thereby the potential matches between the baseline and rotated images could be verified (Lowe, 2004). If the matched keypoint locations are within 1σ of the predicted keypoint locations, then the matches are verified. However, this methodology was not applied in this work, since one of the data sets employed in this work was captured using the stereo-pair system, where the orientation were only an estimate, hence an accurate rotation estimate was unavailable.

9.3.3.3 Extend the Hough Transform to include 3D Pose

Investigation is required in order to extend the HT used in this work to include the slant and tilt information. Currently, the Hough accumulator space is parameterised by the underlying degrees of freedom, namely translation (in plane), rotation (in plane) and scale in size. In order to parameterise the underlying degrees of freedom accurately, the slant and tilt rotation would need to be taken into account as well as the range estimate scaled by translation.

9.3.3.4 Grouping of 2D and 2.5D Feature Descriptors

Based on this work, it is possible to extend the architecture to combine 2D and 2.5D analyses by grouping keypoints from each modality within the Hough Transform. Keypoints from 2D and 2.5D images could be grouped together, resulting in the contributions from each imaging modality being combined automatically. A weight which indicates the contribution of their confidence could be included in some manner, for instance in range images, the scale properties could be included, whereas in the intensity images the contrast properties could be considered. This could potentially improve the

significance of the contributions from each modality and could provide a more stable interpretation than using single imaging modality alone. The trade-off between the 2.5D keypoints being more stable while 2D SIFT yielding more keypoints could be reached.

Moreover, this multi-modal integration strategy could potentially switch modes as they become available to the system by accumulating Hough votes from keypoints of whichever modality is present.

It might even be possible to incorporate different cues, such as colour, at the keypoint detection stage to cue keypoint sampling diagnostic of a particular object or anatomic feature (determined by training samples).

9.3.4 Extend 2.5D SIFT to Accommodate Different Biological Forms

Finally, this work has the potential to be extended to allow different biological forms to be interpreted by taking both the local and global variability of keypoint descriptors into account. For example, local variabilities could be taken into account in PCA SIFT (Ke and Sukthankar, 2004) where natural variabilities have been projected onto the PCA space. Alternatively, a different yet effective approach, the *bag of words* model (Sivic et al., 2005), could be applied where similar keypoints are clustered together using k-means clustering. Variations in the locations of the keypoints themselves might be accommodated by PCA projection into the Hough space (Li and Zhao, 2001).

Appendix A

Settings Used for Model Matching and Building in C3D2

Shown in this appendix is the settings used for model matching and building in C3D2 for the data set. The algorithm used was direct range mesh which produces a more uniform 3D polygon model for this data set, compared to using the marching cube algorithm. These settings were derived especially for the building of face models for the cameras set up described in Section 3.2 and were used to produce the data set used for this work, as well as the data set derived as part of a longitudunal study to investigate the properties of collagen (Project ID: TSL PIN6001).

A.1 Settings for Models Matching

Match Size:	2250
Normal Iteration:	15
Hard Iteration:	15
Match Smoothing:	5
Minimum Pyramid Size:	32
Use default MSSM98 Matching:	false

A.2 Settings for Models Building

Voxel Size:	0.001
Merge:	false
Mesh Reduction Fraction:	1
Fill Holes:	true
Reverse Fill:	false
Smoothing Iteration:	1
Dot Render:	false
Direct Range Mesh:	true
Range Mesh Decimator:	10
Tolerated for Overlap:	0.002
Export Raw XYZ:	false

Appendix B

Relating Landmarks on Range Images using Barycentric Coordinates

The per-vertex assignment of texture coordinates is the key to mapping a texture image to rendered geometry. Each landmark point (L) lies within a certain triangle with vertices p_1, p_2 and p_3 as shown in Figure B.1 below.

The Facial Analysis Tool records and saves the x, y, and z coordinates (L_x, L_y, L_z) of each landmark to a text file as soon as the landmarks have been assigned on the 3D model. This text file can be viewed or used to load the landmarks on the appropriate 3D model at a later time point. The algorithm in Facial Analysis Tool searches all of the triangles in the 3D model in order to locate the appropriate triangle in which the landmark belongs to, hence providing the vertices information including the (x, y, z) coordinates of the vertices (p_1, p_2, p_3) and their corresponding Texture Coordinates. The Barycentric Coordinates of the landmark (L) can be obtained once the vertices coordinates are known. Having obtained the Barycentric Coordinates of the landmark, it is possible to compute the Texture Coordinates of the landmark and as a result, landmarks can be related on the range image directly and successfully.

B.0.0.1 Barycentric Coordinates

Triangles are an important primitive in computer graphics since they can represent a planar face effectively. Barycentric coordinates are a fundamental tool for dealing with

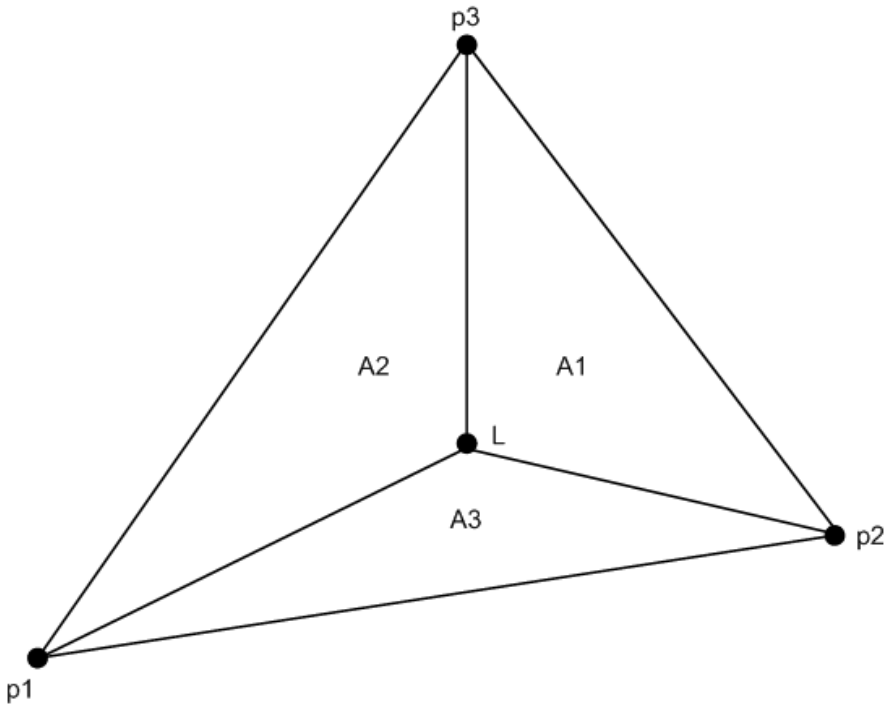


Figure B.1: Barycentric Coordinates of Landmark L, with respect to a triangle with vertices p_1, p_2 and p_3.

triangles. A Barycentric combination of three points (in a triangle) takes the form:

$$L = up_1 + vp_2 + wp_3 \tag{B.0.1}$$

where p_1, p_2, p_3 are the vertices of the triangle and $u + v + w = 1$.

Therefore (u, v, w) is the Barycentric coordinates of point L in which L lies within a triangular plane. Equation B.0.2 represents three equations, and hence can be formed as a linear system for the unknown u, v and w.

$$\begin{bmatrix} p_1 & p_2 & p_3 \end{bmatrix} \begin{bmatrix} u \\ v \\ w \end{bmatrix} = L \tag{B.0.2}$$

In order to obtain (u, v, w), the areas of the triangles A, A_1, A_2 and A_3 must be

calculated.

$$\begin{array}{rcl} A & = & \begin{vmatrix} p_1 & p_2 & p_3 \end{vmatrix} \\ A_1 & = & \begin{vmatrix} L & p_2 & p_3 \end{vmatrix} \\ A_2 & = & \begin{vmatrix} p_1 & L & p_3 \end{vmatrix} \\ A_3 & = & \begin{vmatrix} p_1 & p_2 & L \end{vmatrix} \end{array} \tag{B.0.3}$$

As a result, the Barycentric representation can be presented as follows:

$$u = \frac{A_1}{A}, \quad v = \frac{A_2}{A}, \quad w = \frac{A_3}{A} \tag{B.0.4}$$

B.0.0.2 Texture Coordinates

Following from the previous section, the Texture Coordinates of the landmark (L'_x, L'_y) can be computed from the Barycentric Coordinates of the landmark and the vertices coordinates:

$$\begin{array}{rcl} L'_x & = & x'_1 u + x'_2 v + x'_3 w \\ L'_y & = & y'_1 u + y'_2 v + y'_3 w \end{array} \tag{B.0.5}$$

where (x_1, y_1) is the texture coordinates of vertice p_1 etc.

Using the texture coordinates, it is possible to locate the landmark on the range image. This can be achieved by multiplying the width of the range image by L'_x and the height by L'_y. This would provide the coordinates of the landmark on the range image correctly.

Appendix C

Rotating Range Image Synthetically

The section presents a methodology in generating synthetically rotated range images about the Euler's out-of-plane yaw and pitch axes, where the aliasing of the images due to rotational changes are minimised. It is possible to simulate out-of-plane (about the pitch and yaw axes) rotations of a range image by re-projecting the data into a new image, thereby providing a data set of range images with known rotation relative to the camera. The steps are detailed here.

Based on the assumption that the range images used in this work are planar, the range image can therefore be re-written as a vector:

$$V = \begin{bmatrix} x & y & z & 1 \end{bmatrix} \tag{C.0.1}$$

where z is the normalised range image (to $[0, 100]$).

A scaling matrix can also be defined as:

$$S = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & Sy & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{C.0.2}$$

where Sy is the scaling factor. In other words, if $Sy = 1$, the output image size remains the same.

A combined rotational matrices for the out-of-plane rotational changes (Pansang

et al., 2005) can be formed, as shown in Equation C.0.3 below:

$$R = \begin{bmatrix} cos(\beta) & 0 & -sin(\beta) & 0 \\ 0 & cos(\alpha) & sin(\alpha) & 0 \\ sin(\beta) & -sin(\alpha) & cos(\alpha)cos(\beta) & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{C.0.3}$$

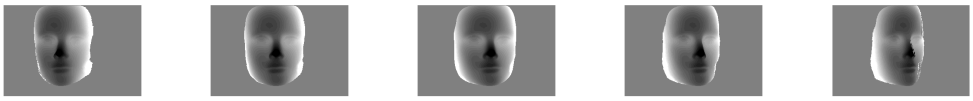
where α corresponds to the rotation in the pitch axis (i.e. up/down) whereas β corresponds to the rotation in the yaw axis (i.e. left/right, lateral).

The rotated version of the range image can then be defined as follows:

$$V1 = R * S * V \tag{C.0.4}$$

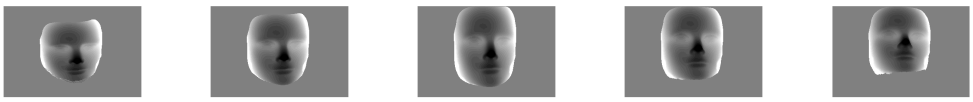
where R is the rotation matrix, S is the scaling matrix and V is the original range image vector.

Figure C.1 illustrates a selection of range images of a mannequin head, synthetically rotated about the yaw axis, while Figure C.2 shows a selection of range images of a mannequin head, synthetically rotated about the pitch axis.



(a) -40° (b) -20° (c) 0° (d) 20° (e) 40°

Figure C.1: A selection of synthetically rotated out-of-plane range images (about the yaw axis) of a mannequin head, generated at (a) -40°, (b) -20°, (c) 0°, (d) 20° and (e) 40°.



(a) -40° (b) -20° (c) 0° (d) 20° (e) 40°

Figure C.2: A selection of synthetically rotated out-of-plane range images (about the pitch axis) of a mannequin head, generated at (a) -40°, (b) -20°, (c) 0°, (d) 20° and (e) 40°.

This methodology is effective in rotating the original range image about the pitch and yaw axis to ±40° of the original input image. Moreover, the 2D intensity images

can be warped onto the range image to achieve the same rotated view. However, the resulting rotated range images are not equivalent to re-projecting the true range maps from the real world coordinates system. Nevertheless, the rotations are consistent and therefore the resulting images are believed to be sufficient to the purposes of this work.

References

Achermann, B. and Jiang, X. (2000), Classifying range images of human faces with hausdorff distance, *in* 'Proceedings of International Conference on Pattern Recognition', Vol. 2, pp. 809–813. 2.6

Achermann, B., Jiang, X. and Banke, H. (1997), Face recognition using range images, *in* 'Proceedings of International Conference on Virtual Systems and Multimedia', pp. 129–136. 2.5

Akagündüz, E. and Ulusoy, I. (2007), 3D object representation using transform and scale invariant 3D features, *in* 'Proceedings of ICCV'07 Workshop on 3D Representation for Recognition (3dRR-07)', pp. 1–8. 1.2, 2.7.2, 8.1

Al-Osaimi, F., Bennamoun, M. and Mian, A. S. (2007), Interest-point based face recognition from range images, *in* 'Proceedings of British Machine Vision Conference'. 2.6

Altman, D. G. and Bland, J. M. (1994), 'Statistics Notes: Diagnostic tests 3: receiver operating characteristic plots', *British Medical Journal* **309**(6948), 188. 6.2.3.2

Ansari, A., Abdel-Mottaleb, M. and Mahoor, M. H. (2008), 'A multimodal approach for 3D face modeling and recognition using 3D deformable facial mask', *To appear in Machine Vision and Applications* . 2.8

Balasuriya, L. S. and Siebert, J. P. (2006), 'Hierarchical feature extraction using a self-organised retinal receptive field sampling tessellation', *Neural Information Processing - Letters & Reviews* **10**(4-6), 83–95. 2

Balasuriya, S. (2005), A Computational Model of Space-Variant Vision Based on a Self-Organised Artificial Retina Tessellation, PhD thesis, University of Glasgow. 2

Ballard, D. H. (1981), 'Generalizing the Hough transform to detect arbitrary patterns', *Pattern Recognition* **13**(2), 111–122. 5.5.1

Belew, R. K. (2000), *Finding Out About: Search Engine Technology from a Cognitive Perspective*, Cambridge University Press. 6.2.3.2

Bellman, R. E. (1957), *Dynamic Programming*, Princeton University Press. 4.5, 9.2.1

BenAbdelkadera, C. and Griffin, P. A. (2005), 'Comparing and combining depth and texture cues for face recognition', *Image and Vision Computing* **23**(3), 339–352. 2.8

Besl, P. F. and Jain, R. C. (1986), 'Invariant surface characteristics for 3D object recognition in range images', *Computer Vision, Graphics, and Image Processing* **33**(1), 33–80. 2.1

Besl, P. J. (1998), *Surfaces in Range Image Understanding*, Springer-Verlag. 1.1

Besl, P. J. and Jain, R. C. (1985), 'Three-dimensional object recognition', *ACM Computing Surveys* **17**(1), 75–145. 1.2, 2.4.1

Besl, P. J. and Jain, R. C. (1988), 'Segmentation through variable-order surface fitting', *IEEE Transactions on Pattern Analysis and Machine Intelligence* **10**(2), 167–192. 2.4.1

Beumier, C. and Acheroy, M. (1997), Automatic profile identification, *in* 'Proceedings of International Conference on Audio- and Video-based Biometric Person Authentication', pp. 145–152. 2.8

Beumier, C. and Acheroy, M. (2001), 'Face verification from 3D and grey level cues', *Pattern Recognition Letters* **22**(12), 1321–1329. 2.8

Bewick, V., Cheek, L. and Ball, J. (2004), 'Statistics review 13: Receiver operating characteristic curves', *Critical Care* **8**, 508–512. 6.2.3.2

Bhanu, B. and Zhou, X. (2004), Face recognition from face profile using dynamic time warping, *in* 'Proceedings of International Conference of Pattern Recognition', pp. 499–502. 2.4.2

Bookstein, F. L. (1997), *Morphometric Tools for Landmark Data, Geometry and Biology*, Cambridge University Press. 2.2

Bowyer, K. W., Chang, K. and Flynn, P. (2006), 'A survey of approaches and challenges in 3D and multi-modal 3D+2D face recognition', *Computer Vision and Image Understanding* **101**(1), 1–15. 1.1, 2.1.1

Brady, M., Ponce, J., Yuille, A. and Asada, H. (1985), 'Describing surfaces', *Computer Vision, Graphics and Image Processing* **32**(1), 1–28. 2.4

Bronstein, A. M., Bronstein, M. M. and Kimmel, R. (2003), Expression-invariant 3D face recognition, *in* 'Proceedings of International Conference on Audio- and Video-based Biometric Person Authentication', pp. 62–70. 2.8

Bronstein, A. M., Bronstein, M. M. and Kimmel, R. (2005), 'Three-dimensional face recognition', *International Journal of Computer Vision* **64**(1), 5–30. 2.6

Burt, P. J. and Adelson, E. H. (1983), 'The laplacian pyramid as a compact image code', *IEEE Transactions on Communications* **31**(4), 532–540. 3.3.1

Campbell, R. J. and Flynn, P. J. (2002), Recognition of free-form objects in dense range data using local features, *in* 'Proceedings of the International Conference on Pattern Recognition', Vol. 3, pp. 607–610. 2.4.1

Cartoux, J. Y., LaPreste, J. T. and Richetin, M. (1989), Face authentication or recognition by profile extraction from range images, *in* 'Proceedings of the Workshop on Interpretation of 3D Scenes', pp. 194–199. 2.4.1

Chang, K., Bowyer, K. and Flynn, P. (2003), Face recognition using 2D and 3D facial data, *in* 'Proceedings of Multimodal User Authentication Workshop', pp. 25–32. 2.5, 2.8

Chang, K. I., Bowyer, K. W. and Flynn, P. J. (2005), 'An evaluation of multimodal 2D+3D face biometrics', *IEEE Transactions on Pattern Analysis and Machine Intelligence* **27**(4), 619–624. 2.6

Chang, K. I., Bowyer, K. W. and Flynn, P. J. (2006), 'Multiple nose region matching for 3D face recognition under varying facial expression', *IEEE Transactions on Pattern Analysis and Machine Intelligence* **28**(10), 1695–1700. 2.4.2, 2.13

Chellapa, R., Wilson, C. and Sirohey, S. (1995), 'Human and machine recognition of faces: a survey', *Proceedings of the IEEE* **83**(5), 705–740. 2.1.1

Chen, H. and Bhanu, B. (2004), 3D free-form object recognition in range images using local surface patches, *in* 'Proceedings of International Conference on Pattern Recognition', Vol. 3, pp. 136–139. 1.2, 2.7.2

Chen, H. and Bhanu, B. (2007), '3D free-form object recognition in range images using local surface patches', *Pattern Recognition Letters* **28**(10), 1252–1262. 2.7.2

Chua, C., Han, F. and Ho, Y. K. (2000), 3D human face recognition using point signature, *in* 'Proceedings of IEEE International Conference on Automatic Face and Gesture Recognition', pp. 233–238. 2.6, 2.7.1

Chua, C. S. and Jarvis, R. (1997), 'Point signatures: a new representation for 3D object recognition', *International Journal of Computer Vision* **25**(1), 63–85. 2.6, 2.7.1

Colombo, A., Cusano, C. and Schettini, R. (2006), '3D face detection using curvature analysis', *Pattern Recognition* **39**(3), 444–455. 2.4.2, 2.11

Coombes, A. M. (1993), Shape Classification: Towards a Mathematical Description of the Face, PhD thesis, University of College London. 2.1.1

Davis, J. and Goadrich, M. (2006), The relationship between Precision-Recall and ROC curves, *in* 'Proceedings of the 23rd International Conference on Machine Learning', pp. 233–240. 6.2.3.2

do Carmo, M. P. (1976), *Differential Geometry of Curves and Surfaces*, Prentice-Hall. 1.2, 2.3.8

Dorai, C. and Jain, A. K. (1997*a*), 'COSMOS – A representation scheme for 3D free-form objects', *IEEE Transactions on Pattern Analysis and Machine Intelligence* **19**(10), 1115–1130. 2.7, 2.7.1

Dorai, C. and Jain, A. K. (1997*b*), 'Shape spectrum based view grouping and matching of 3D free-form objects', *IEEE Transactions on Pattern Analysis and Machine Intelligence* **19**(10), 1139–1146. 1.2

Downie, J., Mao, Z., Lo, T. W. R., Barry, S., Bock, M., Siebert, J. P., McConnachie, A., Yap, C., Bowman, A. and Ayoub, A. (2008), 'Double-blind, clinical evaluation of facial augmentation treatments: a comparison of PRI 1, PRI 2, Zyplast® and Perlane®', *To appear in Journal of Plastic, Reconstructive & Aesthetic Surgery* . 3.2

Dryden, I. L. and Mardia, K. V. (1998), *Statistical Shape Analysis*, John Wiley & Sons Inc. 2.2

Duda, R. and Hart, P. (1973), *Pattern Classification and Scene Analysis*, Wiley. 2.5.0.1, 4.5, 9.2.1

Duda, R. O. and Hart, P. E. (1972), 'Use of the Hough transformation to detect lines and curves in pictures', *Communications of the ACM* **15**(1), 11–15. 5.5.1

Duda, R. O., Hart, P. E. and Stork, D. G. (2000), *Pattern Classification*, second edn, Wiley Interscience. 2.5.0.1

Egan, J. P. (1975), *Signal detection theory and ROC analysis*, Academic Press, New York. 6.2.3.2

El-Maraghi, T. F. (2004), 'Sift tutorial in matlab'. 4.5

Fan, T. G., Medioni, G. and Nevatia, R. (1986), Description of surfaces from range data using curvature properties, *in* 'Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition', pp. 86–91. 2.4.1

Fawcett, T. (2006), 'An introduction to ROC analysis', *Pattern Recognition Letters* **27**(8), 861–874. 6.2.3.2

Ferrario, V. F., Sforza, C., Poggio, C. E., Cova, M. and Tartaglia, G. (1998), 'Preliminary evaluation of an electromagnetic three-dimensional digitizer in facial anthropometry', *Cleft Palate-Craniofacial Journal* **35**(1), 9–15. 2.2.1, 2.5

Forsyth, D. and Ponce, J. (2003), *Computer Vision: A Modern Approach*, Prentice Hall. 2.1, 2.5.0.1

Fukunaga, K. (1990), *Introduction to Statistical Pattern Recognition*, second edn, Academic Press. 2.5.0.1

Godil, A., Ressler, S. and Grother, P. (2004), Face recognition using 3D facial shape and color map information: comparison and combination, *in* 'Proceedings of the SPIE - The International Society for Optical Engineering', Vol. 5404, pp. 351–361. 2.8

Gökberk, B., Írfanoğlu, M. O. and Akarun, L. (2006), '3D shape-based face representation and feature extraction for face recognition', *Image and Vision Computing* **24**(8), 857–869. 2.4.2

Gonzalez, R. C., Woods, R. E. and Eddins, S. L. (2004), *Digital Image Processing using MATLAB*, Prentice Hall. 2.5.0.1

Gordon, G. G. (1991), 'Face recognition from depth maps and surface curvature', *SPIE Geometric Methods in Computer Vision* **1570**, 234–247. 2.4.1

Gordon, G. G. (1992), Face recognition based on depth and curvature features, *in* 'Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition', pp. 808–810. 1.1, 1.2, 2.1.1, 2.3, 2.4.1

Gordon, G. G. and Vincent, L. (1992), Application of morphology to feature extraction for face recognition, *in* 'Proceedings of SPIE Nonlinear Image Processing III', Vol. 1658, pp. 151–164. 2.8

Haralick, R. M. (1983), 'Ridges and valleys on digital images', *Computer Vision, Graphics and Image Processing* **22**(10), 28–38. 2.4.1

Heseltine, T. D. (2005), Face Recognition: Two-Dimensional and Three-Dimensional Techniques, PhD thesis, University of York. 2.8

Heseltine, T., Pears, N. and Austin, J. (2004), Three-dimensional face recognition: an eigensurface approach, *in* 'Proceedings of International Conference on Image Processing', Vol. 2, pp. 1421–1424. 2.5

Hesher, C., Srivastava, A. and Erlebacher, G. (2003), A novel technique for face recognition using range imaging, *in* 'Seventh International Symposium on Signal Processing and Its Applications', pp. 201–204. 1.1, 2.1.1, 2.5

Hetzel, G., Leibe, B., Levi, P. and Schiele, B. (2001), 3D object recognition from range images using local feature histograms, *in* 'Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition', Vol. 2, pp. 394–399. 1.2, 2.7.1

Hilbert, D. and Cohn-Vossen, S. (1952), *Geometry and the Imagination*, Chelsea, New York. 2.4

Huang, Y., Wang, Y. and Tan, T. (2006), Combining statistics of geometrical and correlative features for 3D face recognition, *in* 'Proceedings of British Machine Vision Conference', Vol. 3, pp. 879–888. 2.7.2

Huang, Y., Wang, Y. and Tan, T. (2007), Discriminating 3D faces by statistics of depth differences, *in* 'Proceedings of Asian Conference on Computer Vision', pp. 690–699. 2.7.2

Hüsken, M., Brauckmann, M., Gehlen, S. and von der Malsburg, C. (2005), Strategies and benefits of fusion of 2D and 3D face recognition, *in* 'Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition', p. 174. 2.8

Ittner, D. J. and Jain, A. K. (1985), 3-D surface discrimination from local curvature measures, *in* 'Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition', pp. 119–123. 1.2, 2.4.1

J. Cook, C. M., Chandran, V. and Sridharan, S. (2006), Combined 2D / 3D face recognition using log-gabor templates, *in* 'Proceedings of IEEE International Conference on Video and Signal Based Surveillance', p. 83. 2.8

Jain, R., Kasturi, R. and Schunck, B. G. (1995), *Machine Vision*, MIT Press and McGraw-Hill Inc. 2.4.1, 1, 2

Johnson, A. E. (1997), Spin-Images: A Representation for 3-D Surface Matching, PhD thesis, Robotics Institute, Carnegie Mellon University, Pittsburgh, PA. 2.7.1

Johnson, A. E. and Hebert, M. (1999), 'Using spin images for efficient object recognition in cluttered 3D scenes', *IEEE Transactions on Pattern Analysis and Machine Intelligence* **21**(5), 433–449. 2.7.1

Ju, X., Boyling, T. and Siebert, J. P. (2003), A high resolution stereo imaging system, *in* 'Proceedings of 3D Modelling 2003'. 1.3, 3.1, 3.2.2

Ke, Y. and Sukthankar, R. (2004), PCA-SIFT: A more distinctive representation for local image descriptors, *in* 'Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition', pp. 506–513. 9.3.4

Kim, I., Lee, Y. and Shim, J. (2005), An automated facial pose estimation using surface curvature and tetrahedral structure of a nose, *in* 'Proceedings of International Conference on Advanced Concepts for Intelligent Vision Systems', pp. 276–283. 2.4.2

Kim, T. K., Kee, S. C. and Kim, S. R. (2001), Real-time normalization and feature extraction of 3D face data using curvature characteristics, *in* 'Proceedings of 10th IEEE International Workshop on Robot and Human Interactive Communication', pp. 74–79. 2.4.1, 2.9

Koenderink, J. J. (1990), *Solid Shape*, MIT Press. 2.3.8

Koenderink, J. J. and van Doorn, A. J. (1992), 'Surface shape and curvature scales', *Image and Vision Computing* **10**(8), 557–565. 2.3.7, 4.2.2.1

Landy, M. S. and Sperling, G. (1984), 'HIPS: Image processing under UNIX. Software and applications', *Behaviour Research Methods, Instrumentation and Computers* **16**(2), 199–216. 3.2.1, 3.3

Lee, J. C. and Milios, E. (1990), Matching range images of human faces, *in* 'Proceedings of International Conference on Computer Vision', pp. 722–726. 1.2, 2.4.1

Lee, Y., Kim, I., Shim, J. and Marshall, D. (2006), 3D facial image recognition using a nose volume and curvature based eigenface, *in* 'Proceedings of Geometric Modeling and Processing', pp. 616–622. 2.4.2

Lee, Y., Park, K. and Shim, J. (2003), 3D face recognition using statistical multiple features for the local depth information, *in* 'Proceedings of 16th International Conference on Vision Interface'. 2.7.2

Lee, Y. and Shim, J. (2004), Curvature-based human face recognition using depth-weighted hausdorff distance, *in* 'Proceedings of International Conference on Image Processing', pp. 1429–1432. 2.4.2

Lee, Y., Song, H., Yang, U., Shin, H. and Sohn, K. (2005), Local feature based 3D face recognition, *in* 'Proceedings of International Conference on Audio- and Video-based Biometric Person Authentication', Vol. 3546, pp. 909–918. 2.7.2

Li, S. and Zhao, D. (2001), Three-dimensional object pose estimation for robotic assembly applications, *in* 'Proceedings of Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series', Vol. 4553, pp. 61–71. 9.3.4

Li, X. J. and Guskov, I. (2007), 3D object recognition from range images using pyramid matching, *in* 'Proceedings of ICCV'07 Workshop on 3D Representation for Recognition (3dRR-07)', pp. 1–6. 1.2, 2.7.2

Li, Y. and Su, G. (2006), Face pose estimate and multi-pose synthesize by 2D morphable model, *in* 'Proceedings of 2006 International Conference on Computational Intelligence and Security', Vol. 2, pp. 1857–1860. 8.1

Lindeberg, T. (1994*a*), 'Scale-space theory: A basic tool for analysing structures at different scales', *Journal of Applied Statistics* **21**(2), 225–270. 1.3, 5.4

Lindeberg, T. (1994*b*), *Scale-Space Theory in Computer Vision*, Kluwer Academic Publishers. 1.3, 2.6.1, 5.4

Lo, T. W. R., Siebert, J. P. and Ayoub, A. F. (2006), Robust feature extraction for range images interpretation using local topology statistics, *in* 'Proceedings of MICCAI 2006 Workshop on Craniofacial Image Analysis for Biology, Clinical Genetics, Diagnostics and Treatment', pp. 75–82. 1.2

Lo, T. W. R., Siebert, J. P. and Ayoub, A. F. (2007), An implementation of the scale invariant feature transform in the 2.5d domain, *in* 'Proceedings of MICCAI 2007 Workshop on Content-based Image Retrieval for Biomedical Image Archives: Achievements, Problems, and Prospects', pp. 73–82. 1.2

Lowe, D. G. (1999), Object recognition from local scale-invariant features, *in* 'Proceedings of International Conference on Computer Vision', pp. 1150–1157. 2.6.1

Lowe, D. G. (2001), Local feature view clustering for 3D object recognition, *in* 'Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition', Vol. 1, pp. 682–688. 2.6.1

Lowe, D. G. (2004), 'Distinctive image features from scale-invariant keypoints', *International Journal of Computer Vision* **60**(2), 91–110. 1, 1.3, 2.6, 2.6.1, 2.14, 2.8, 4.5, 5, 3, 5.4, 5.5, 5.5.2, 6.1, 9.1, 9.2.4, 9.3.3.2

Lowe, D. G. (2005), 'Demo Software: SIFT Keypoint Detector'. Demo available for download at http://www.cs.ubc.ca/∼lowe/keypoints/. 4.5, 8.1, 8.4

Lu, X., Colbry, D. and Jain, A. K. (2004), Three-dimensional model based face recognition, *in* 'Proceedings of International Conference on Pattern Recognition', Vol. 1, pp. 362–366. 2.4.2

Lu, X. and Jain, A. K. (2005), Deformation analysis for 3D face matching, *in* 'Proceedings of 7th IEEE Workshop on Applications of Computer Vision', pp. 99–104. 2.4.2, 2.8

Lu, X. and Jain, A. K. (2006), Automatic feature extraction for multiview 3D face recognition, *in* 'Proceedings of International Conference on Automatic Face and Gesture Recognition', pp. 585–590. 2.12

Lu, X., Jain, A. K. and Colbry, D. (2006), 'Matching 2.5D face scans to 3D models', *IEEE Transactions on Pattern Analysis and Machine Intelligence* **28**(1), 31–43. 2.4.2

Malassiotis, S. and Strintzis, M. G. (2005), 'Robust face recognition using 2D and 3D data: Pose and illumination compensation', *Pattern Recognition* **38**(12), 2537–2548. 2.8

Mao, Z. (2005), Computer assisted methods to support the clinical assessment of human surface anatomy in 3D images, PhD thesis, University of Glasgow. 3.1, 3.2.3

Marr, D. (1982), *Vision*, W. H. Freeman and Co. 4.2.1

Maurer, T., Guigonis, D., Maslov, I., Peseni, B., Tsaregorodtsev, A., West, D. and Medioni, G. (2005), Performance of geometrix Active^{TM} 3D face recognition engine on the FRGC data, *in* 'Proceedings of IEEE Workshop on Face Recognition Grand Challenge Experiment'. 2.8

Mavridis, N., Tsalakanidou, F., Pantazis, D., Malassiotis, S. and Strintzis, M. G. (2001), The HISCORE face recognition application: Affordable desktop face recognition

based on a novel 3D camera, *in* 'Proceedings of International Conference on Augmented Virtual Environments and 3D Imaging', pp. 157–160. 2.5

McFarlane, N. J. B., Wu, J., Tillett, R. D., Schofield, C. P., Siebert, J. P. and Ju, X. (2005), 'Shape measurements of live pigs using 3D image capture', *Animal Science, British Society of Animal Science* **1**, 383–391. 2.4.2

Medioni, G. and Waupotitsch, R. (2003), Face modeling and recognition in 3-D, *in* 'IEEE International Workshop on Analysis and Modeling of Faces and Gestures', pp. 232–233. 2.1.1

Mian, A., Bennamoun, M. and Owens, R. (2006*a*), Face recognition using 2D and 3D multimodal local features, *in* 'Proceedings of 2nd International Symposium on Visual Computing', pp. 860–870. 2.8

Mian, A. S., Bennamoun, M. and Owens, R. (2008), 'Keypoint detection and local feature matching for textured 3D face recognition', *International Journal of Computer Vision* **79**, 1–12. 2.8

Mian, A. S., Bennamoun, M. and Owens, R. A. (2006*b*), 'A novel representation and feature matching algorithm for automatic pairwise registration of range images', *International Journal of Computer Vision* **66**(1), 19–40. 2.8

Mian, A. S., Bennamoun, M. and Owens, R. A. (2007), Keypoint identification and feature-based 3D face recognition, *in* 'Proceedings of International Conference on Biometrics', pp. 163–171. 2.7.2

Mikolajczyk, K. (2002), Detection of local features invariant to affine transformations, PhD thesis, Institute National Polytechnique de Grenoble, France. 2.6.1, 5.4

Mikolajczyk, K. and Schmid, C. (2004), 'Scale and affine invariant interest point detectors', *International Journal of Computer Vision* **60**(1), 63–86. 2.6.1, 5.4

Moreno, A. B., Sánchez, Á., Vélez, J. F. and Díaz, F. J. (2003), Face recognition using 3D surface-extracted descriptors, *in* 'Proceedings of Irish Machine Vision and Image Processing Conference'. 1.2, 2.4.1, 2.10, 2.7.1

Moses, Y., Adini, Y. and Ullman, S. (1997), 'Face recognition: the problem of compensating for illumination changes', *IEEE Transactions on Pattern Analysis and Machine Intelligence* **19**(7), 721–732. 2.1.1

Mowforth, P. H. and Jin, Z. P. (1986), 'Implementation for noise suppression in images', *Image and Vision Computing* **4**(1), 29–37. 3.3.1

Mustafa, A. A. Y., Shaprio, L. G. and Ganter, M. A. (1999), '3D object identification with color and curvature signatures', *Pattern Recognition* **32**(3), 339–355. 2.7.1

Nagamine, T., Uemera, T. and Masuda, I. (1992), 3D facial image analysis for human identification, *in* 'Proceedings of International Conference on Pattern Recognition', Vol. 1, pp. 324–327. 2.6

Norman, J. F., Todd, J. T., Norman, H. F., Clayton, A. M. and McBride, T. R. (2006), 'Visual discrimination of local surface structure: Slant, tilt and curvedness', *Vision Research* **46**(6–7), 1057–1069. 1.2, 7.2.1.1, 7.3.1

Oja, E. (1983), *Subspace methods of pattern recognition*, Research Study Press. 2.5.0.1

Pan, G., Han, H., Wu, Z. and Wang, Y. (2005), 3D face recognition using mapped depth images, *in* 'Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition', p. 175. 2.5

Pan, G., Wu, Z. and Pan, Y. (2003), Automatic 3D face verification from range data, *in* 'Proceedings of International Conference on Acoustics, Speech and Signal Processing', pp. 193–196. 2.6

Pansang, S., Attachoo, B., Kimpan, C. and Sato, M. (2005), 'Invariant range image multi-pose face recognition using gradient face, membership matching score and 3-layer matching search', *IEICE Transactions on Information and Systems* **E88-D**(2), 268–277. 1.2, 8.1, C

Papatheodorou, T. and Reuckert, D. (2004), Evaluation of automatic 4D face recognition using surface and texture registration, *in* 'Proceedings of International Conference on Automated Face and Gesture Recognition', pp. 321–326. 2.8

Park, S. H., Goo, J. M. and Jo, C. H. (2004), 'Receiver Operating Characteristic (ROC) Curve: practical review for radiologists', *Korean Journal of Radiology* **5**(1), 11–18. 6.2.3.2

Phillips, P. J., Moon, H., Rizvi, S. A. and Rauss, P. J. (2000), 'The FERET evaluation methodology for face recognition algorithms', *IEEE Transactions on Pattern Analysis and Machine Intelligence* **22**(10), 1090– 1104. 2.6.2

Rao, R. P. N. (1994), Top-down gaze targeting for space-variant active vision, *in* 'Proceedings of ARPA Image Understanding Workshop', pp. 1049–1058. 2.9

Riccio, D. and Dugelay, J. (2007), 'Geometric invariants for 2D/3D face recognition', *Pattern Recognition Letters* **28**(14), 1907–1914. 2.8

Russ, T., Boehnen, C. and Peters, T. (2006), 3D face recognition using 3D alignment for PCA, *in* 'Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition', pp. 1391–1398. 2.5

Shan, Y., Sawhney, H. S., Matei, B. and Kumar, R. (2006), 'Shapeme histogram projection and matching for partial object recognition', *IEEE Transactions on Pattern Analysis and Machine Intelligence* **28**(4), 568–577. 2.7.2

Siebert, J. P. and Marshall, S. J. (2000), 'Human body 3D imaging by speckle texture projection photogrammetry', *Sensor Review* **20**(3), 218–226. 1.3, 3.2.2

Siebert, J. P. and Urquhart, C. W. (1994), C3D: a novel vision-based 3-D data acquisition system, *in* 'Proceedings of the European Workshop on Combined Real and Synthetic Image Processing for Broadcast Video Production', pp. 170–180. 1.3, 3.2.2

Sivic, J., Russell, B., Efros, A., Zisserman, A. and Freeman, W. (2005), Discovering objects and their location in images, *in* 'Proceedings of International Conference on Computer Vision', pp. 370–377. 9.3.4

Spackman, K. A. (1989), Signal detection theory: Valuable tools for evaluating inductive learning, *in* 'Proceedings of Sixth International Workshop on Machine Learning', pp. 160–163. 6.2.3.2

Sun, Y. and Yin, L. (2006), Evaluation of 3D facial feature selection for individual facial model identification, *in* 'Proceedings of International Conference on Pattern Recognition', Vol. 4, pp. 562–565. 2.4.2

Sze, C. J., Liao, H. Y. M., Hung, H. L., Fan, K. C. and Hsieh, J. W. (1998), 'Multiscale edge detection on range images via normal changes', *IEEE Transactions on Circuits and Systems - II: Analog and Digital Signal Processing* **45**(8), 1087–1092. 7.2.1.1

Tanaka, H. T., Ikeda, M. and Chiaki, H. (1998), Curvature-based face surface recognition using spherical correlation principal directions for curved object recognition, *in* 'Proceedings of International Conference on Automated Face and Gesture Recognition', pp. 372–377. 2.4.1

Trucco, E. and Verri, A. (1998), *Introductory Techniques for 3-D Computer Vision*, Prentice Hall. 2.1.1

Tsalakanidou, F., Tzocaras, D. and Strintzis, M. (2003), 'Use of depth and colour eigenfaces for face recognition', *Pattern Recognition Letters* **24**(9-10), 1427–1435. 2.8

Urquhart, C. (1997), The active stereo probe: the design and implementation of an active videometrics system, PhD thesis, University of Glasgow. 3.2.2.1

van Rijsbergen, C. J. (1979), *Information Retrieval*, second edn, London: Butterworths. 6.2.3.2

Wang, Y., Chua, C. and Ho, Y. (2002), 'Facial feature detection and face recognition from 2D and 3D images', *Pattern Recognition Letters* **23**(10), 1191–1202. 1.2, 2.8

Wang, Y., Peterson, B. S. and Staib, L. H. (2000), Shape-based 3D surface correspondence using geodesics and local geometry, *in* 'Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition', Vol. 2, pp. 2644–2651. 2.4.1

Wiskott, L., Fellous, J. M., Krűger, N. and von de Malsburg, C. (1997), 'Face recognition by elastic bunch graph matching', *IEEE Transactions on Pattern Analysis and Machine Intelligence* **19**(7), 775–779. 1.1.1, 2.6, 2.6.2, 2.15, 4.5

Xu, C., Wang, Y., Tan, T. and Quan, L. (2004*a*), Automatic 3D face recognition combining global geometric features with local shape variation information, *in* 'Proceedings of International Conference on Automated Face and Gesture Recognition', pp. 308–313. 1.2, 2.6, 2.7.2

Xu, C., Wang, Y., Tan, T. and Quan, L. (2004*b*), Depth vs. intensity: Which is more important for face recognition?, *in* 'Proceedings of IEEE International Conference on Pattern Recognition', Vol. 1, pp. 342–345. 2.1.1

Yokoya, N. and Levine, M. D. (1989), 'Range image segmentation based on differential geometry: A hybrid approach', *IEEE Transactions on Pattern Analysis and Machine Intelligence* **11**(6), 643–649. 2.4.1

Zhao, W., Chellappa, R., Phillips, P. J. and Rosenfeld, A. (2003), 'Face recognition: A literature survey', *ACM Computing Surveys* **35**(4), 399–458. 2.1

Zou, K. H. (2002), 'Receiver operating characteristic (ROC) literature research'. Bibliography available online at http://www.spl.harvard.edu/archive/spl-pre2007/pages/ppl/zou/roc.html. 6.2.3.2