Contents lists available at ScienceDirect

Computer Vision and Image Understanding

journal homepage: www.elsevier.com/locate/cviu

Local feature extraction and matching on range images: 2.5D SIFT

Tsz-Wai Rachel Lo *, J. Paul Siebert

Department of Computing Science, University of Glasgow, Sir Alwyn Williams Building, Lilybank Gardens, Glasgow G12 8RZ, UK

ARTICLE INFO

Article history:
Received 23 May 2008
Accepted 10 June 2009
Available online 18 August 2009

Keywords:
Range images
Feature extraction
Point-based matching
SIFT

ABSTRACT

This paper presents an algorithm that extracts robust feature descriptors from 2.5D range images, in order to provide accurate point-based correspondences between compared range surfaces. The algorithm is inspired by the two-dimensional (2D) Scale Invariant Feature Transform (SIFT) 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. We adapt this concept into the 2.5D domain by concatenating the histogram of the range surface topology types, derived using the bounded $[-1,1]$ shape index, 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 are estimated by extracting range surface normals, allowing the three-dimensional (3D) pose of each keypoint to be recovered and used to adapt the descriptor sampling window to provide a more reliable match under out-of-plane viewpoint rotation.

1. Introduction

This paper presents ongoing work to extend Lowe's 2D SIFT [1] algorithm into the 2.5D domain to allow range images to be matched directly using local features that exhibit scale and 3D rotation invariance. Range data acquisition systems have increased in availability and popularity over the last decade [2], primarily due to their ability to capture high quality 2.5D images of surfaces and objects that encode their metric dimensions unambiguously [3,4]. Data captured using this imaging modality also exhibits an inherently greater degree of invariance to illumination variations and 3D pose changes. The ability to capture shape variation within range images, irrespective of illumination variation is discussed by Hesher et al. [5], while Gordon [6] reports the viewpoint rotation invariance properties afforded by range-based imagery. Accordingly, range image analysis has the potential to confer a number of advantages over analyses based on intensity images alone [2] and has therefore motivated us to exploit this potential for machine interpretation.

In [7,8], we introduced our approach to extracting feature descriptors from range images by adapting Lowe's concept of a *keypoint descriptor* to include information about local range surface topology. We have based this new keypoint descriptor on a combination of the histogram of the local surface *shape index* (Eq. (1)) [9], concatenated with the histogram of the local surface orientation gradients, each extracted over the same measurement aperture. By incorporating this descriptor into Lowe's basic SIFT framework and suitably pre-processing the range maps, we demonstrated scale and in-plane 2D rotation invariant range map matching. The principal novel contribution of the work presented here is the extension of the 2.5D SIFT concept to address the effects of local feature foreshortening induced by 3D (out-of-plane) rotations. Since range images can provide explicit knowledge of the local surface direction, we can therefore correct for 3D rotation induced foreshortening and thereby improve the stability of the range features we extract.

Our remit is focused on matching high quality range images, since these have sufficiently different characteristics from mesh-based 3D surface representations and analysis approaches [10], e.g. the regular grid neighbourhood topology of range samples simplifies processing, unlike irregularly polygonised surface tessellations. Furthermore, range images are usually the fundamental representation produced by the triangulation-based ranging devices under consideration here. Due to the plethora of local feature matching approaches reported in the computer vision literature, e.g. [11–14], in this paper we have restricted performance comparisons of our work to that of the basic 2D SIFT algorithm, which we believe to be sufficiently widely recognised and adopted to serve as a standard benchmark. Accordingly, we have devised an experiment to compare the performance of our new *pose-corrected* 2.5D SIFT algorithm, $2.5D_{pc}$ SIFT, with that of 2D SIFT and this indeed demonstrates the potential to achieve improved matching performance in the range domain. The significance of the work presented here is that $2.5D_{pc}$ SIFT can offer improved rigid-body matching,

* Corresponding author.
E-mail addresses: rachel@dcs.gla.ac.uk (T.W.R. Lo), psiebert@dcs.gla. ac.uk (J.P. Siebert).

1077-3142/$ - see front matter © 2009 Elsevier Inc. All rights reserved.
doi:10.1016/j.cviu.2009.06.005

over that currently tackled by 2D SIFT, when significant out-of-plane rotations between compared images are present, e.g. as found in applications such as object recognition, range-map registration, pose estimation for visual servoing, etc.

1.1. Related work

In the range image analysis approach adopted here, a feature descriptor 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. Traditional approaches to machine interpretation of 3D surface manifolds are based on the classification of different types of surface topology, using differential geometry [15], and have been widely used since the 1980s, examples include work by Ittner and Jain [16]. Later, Besl and Jain [17] used the signs of the mean (H) and Gaussian (K) curvatures to segment any smooth and differentiable surface into eight surface types. This forms the basis for 3D shape analysis using differential geometry to achieve classification and segmentation of surfaces according to shape that is now used widely.

In the 1990s, Gordon [6] and Lee and Milios [18] 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, was also widely reported in systems specific to face recognition, and examples include work by [19–22]. Wang et al. [19] reported the use of histograms of shape index and *point signatures* [23] combined in a single feature vector 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. Since this system is based on face sampling using a scheme similar to the Elastic Bunch Graph [24], it is specific to faces. Moreno et al. [20] reported a face recognition system based categorisation of single surface types that requires a specific threshold to be defined in order to segment the object with respect to the H, K and the principal curvatures ($k1$ and $k2$). Since different thresholds are usually required for different types of range image, this process falls short of providing fully automated interpretation. Global pose estimation to improve face recognition in range images with greater rotation invariance has been reported [22] in a scheme that locates the nose and then uses this to normalise the range map orientation. Mouth and eye corner features are then extracted using shape index in the range domain and a corner detector in the intensity domain.

While the above systems represent successful range-based face recognition systems, our goal is to devise a system that is ultimately not restricted to any specific type of subject matter. Systems capable of recognising free-form objects in range images have been reported [23,25–27]. Dorai and Jain [25] used the global shape index histogram as feature to describe the entire visible surface of an object presented in a range image and then relied upon viewpoint clustering to allow pose independent recognition. Point signatures have been reported for free-form object recognition [23] in a scheme that exhaustively sub-samples the signatures over an object surface and then attempts to retain those which are diagnostic of the object. In an initial study that combines histograms of surface normals, shape index and range depth, Hetzel et al. [26] reported good discrimination results on synthetic range image examples, although it appears that the addition of the range depth feature brings only modest additional discrimination power. Chen and Bhanu [27] described a free-form recognition system that uses a 2D histogram comprising shape index combined with surface normal comparisons (the dot product of the central normal of a sampling patch is computed with its neighbouring normals on the patch). Locations exhibiting high shape index variation serve to localise feature sampling points. Good results are reported for a small database of objects, though this scheme (as in most of the above) does not incorporate explicit scale detection and scale invariant sampling.

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* [28,29] and *COSMOS* [30]. Furthermore, little work, until recently [31–34], has been reported on the invariance limits of viewpoint rotational changes achieved using the above techniques.

Despite the above advances, no system reported appears to be able to provide the same type of general purpose scale-invariant matching analysis on range images as SIFT does in the intensity domain, and at the same time also corrects for the effects of 3D pose angle variation. However, it is apparent from many of the reported systems that shape index provides a key ingredient for classifying range surface shape and that it is particularly effective when combined with information regarding the spatial configuration of the surface.

1.2. Approach

To advance the state-of-the-art in matching local range image features, three objectives must be achieved simultaneously: a local feature descriptor extraction mechanism must be devised that can distinguish between the wide variety of pattern types potentially present within the measurement aperture applied to the range surface and this feature extraction mechanism must also be invariant to 3D rotations and relative scale changes. These requirements have motivated our adoption of the overall architecture afforded by the SIFT algorithm to implement feature localisation, scale invariance and rotation invariance. We have observed experimentally that the SIFT mechanism for localising features in (x, y) position and scale, based on detecting inflections in differential scale-space, works well in the range domain primarily due to improved tolerance to illumination variations in this medium. However, a degree of pre-processing is required to mitigate the effects of high-frequency range noise and also to standardise the dynamic signal range to allow use of fixed detection thresholds.

The principal modification required to SIFT to enable it to exploit the potential benefits of operating in the range domain is the formulation of an appropriate feature descriptor sampled from the range surface. In the approach taken in this work, we hypothesise that the sample measurement aperture, i.e. support region, will contain a *mixture* of surface topology types. For example, the descriptor extracted from the *pronasle* (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. Therefore, instead of attempting to segment surfaces into a piece-wise patchwork of single surface types, we propose to extract the distributions of the underlying surface topology types. This concept is illustrated in Fig. 1, 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. We have adopted range images of faces in this work since they both exhibit a rich variety of surface

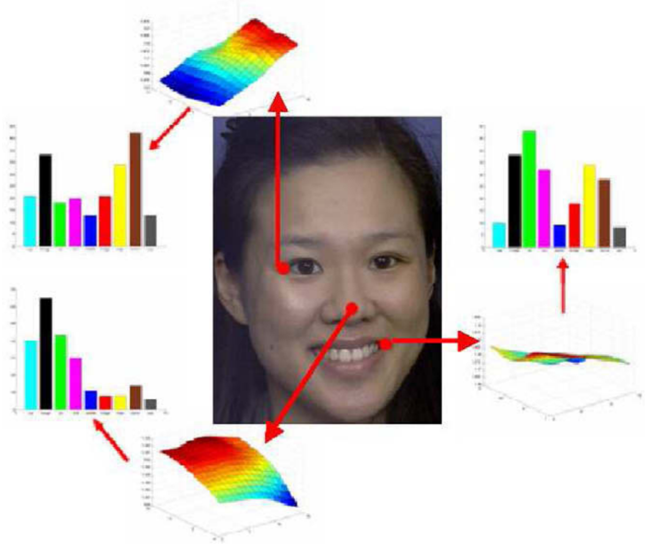


Fig. 1. Surface types and their histograms extracted from three keypoint locations on face range data.

types over a spectrum of spatial scales and are readily available. However, as stated, the use of face data in this case does *not* necessarily imply that our primary interest is face recognition *per se*.

$$s = \frac{2}{\pi}\tan^{-1}\left[\frac{k2 + k1}{k2 - k1}\right], \quad \text{where } k1 \geqslant k2. \tag{1}$$

The question of how to structure a feature descriptor was resolved by examining a number of possible combinations of features reported in the literature based on: *H* and *K* surface type classification, shape index, and *k*1 curvature. Derived from surface curvature, these three popular measures are 3D rotation invariant and can be made scale invariant by appropriate normalisation. In terms of representation power, the *H* and *K* classification quantises the surface types into only eight discrete states. The numeric values generated by the shape index represents a continuum of surface types, ranging from cup (concave) to cap (convex) (see Fig. 2), and represents all intermediate types as the index values smoothly vary between these extremes. The *k*1 curvature only indicates the degree and direction of surface bending and provides no quantification of the local surface in terms of richer shape semantics. We were also keen not to adopt range depth values as features [26] since this limits degree to which the features can be invariant to changes in relative (non-linear) scaling of range values.

Potential weightings of the above types of feature were also investigated and comprised respectively: the degree of curvedness, an isotropic measure of curvature discontinuity [35], and simply uniform weighting. In addition, the ability to classify local surface patches using gradient orientation histograms, as in standard 2D SIFT, was investigated. By measuring the discriminability of the feature descriptors comprising the different permutations of these components, we established that a descriptor comprising a shape index histogram (weighted by the degree of curvedness), concatenated with the orientation histogram (weighted by gradient magnitude), gave the most promising results. A simple KNN classifier was used to determine which feature descriptor gave the highest classification rate when sampling 28 manually-defined (anatomic landmarks) locations on set of 50 range images (of female human faces), compared to in-plane rotated versions comprising nine increments of 10°, details in [36].

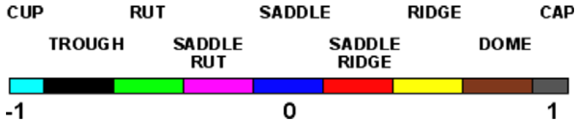


Fig. 2. Shape index scale ranges from −1 to 1.

A potential limitation of using the 1D > shape index, orientation < histogram descriptor described above is that by simply concatenating the distribution histogram describing shape with that recording orientation direction frequency, shape and orientation direction have been de-coupled. Therefore many different combinations of orientation directions and surface shapes, drawn from specific distributions of each feature, could give rise to identical combined histograms. By creating a feature descriptor based on 2D histograms, the co-occurrences of shape and orientation direction feature pairings can be captured and represented uniquely. In this case we examined 2D histograms using either *H* and *K* derived surface classifications or the shape index, coupled with either the *k*1 orientations or the image gradient orientations, using the weighting measures as defined above and weighting individual entries by the products of their respective weights, i.e. the histogram is updated with a value comprising the product of the shape weight and the orientation direction weight. Examining the 2D feature descriptors using the same method as described for the 1D case revealed that the 2D combination of shape index and orientation direction gave the best overall results, but only marginally better than the same combination structured as a 1D descriptor. This result can potentially be explained by examining the 2D descriptor, Fig. 3, which reveals the sparse nature of the captured surface types and orientations. In the data set examined, the complexity of the local surface does not appear to justify the additional representational effort. Furthermore, the increased dimensionality of the 2D descriptor will tend to reduce discriminability, and increase computation and storage costs. Therefore, in this work, the basic feature descriptor comprises the winning <shape index, orientation> 1D histogram combination described above. Note that this feature descriptor is unable to discriminate between different spatial configurations of certain rotationally symmetric stimuli and its modification to overcome this limitation is described in Section 2.2.

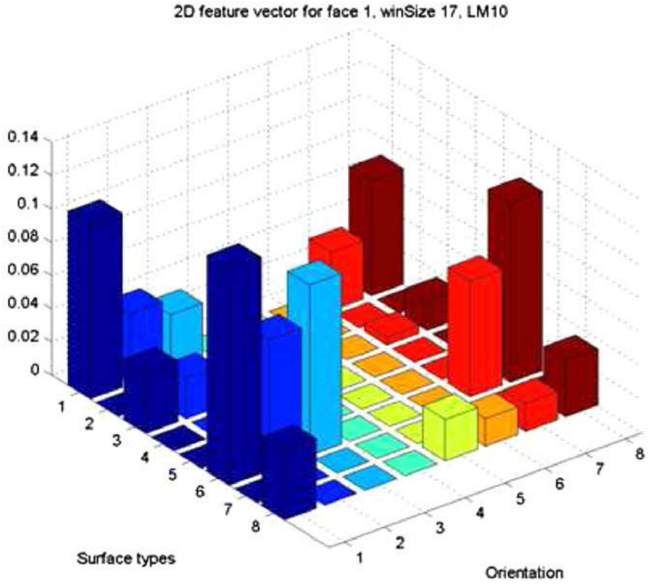


Fig. 3. Example of a 2D feature descriptor. This descriptor was captured from the pronasale keypoint, with sampling aperture of 17 × 17.

In this work, our principal objective is to mitigate the effects of foreshortening induced by 3D rotations in object surface pose. Although the viewpoint orientations of keypoints sampling the range image subject matter are not known a priori, it is possible to recover the 3D poses of these keypoints by estimating their slants and tilts from the local range image surface. An implicit assumption of this approach is that we can approximate the local range surface by a single mean-plane modulated by the underlying shape of the patch. By taking the canonical surface orientation direction of each keypoint into account and correcting this, 2.5D SIFT keypoints representing the same range surface patch captured from different viewpoints can be matched accurately. Having computed the slant and the tilt at each keypoint location over a measurement aperture, the circular Gaussian window normally 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 robust to changes in pose angle.

2. Methodology

This section details the formulation of our 2.5D$_{pc}$ SIFT algorithm for *pose-corrected* range image matching. There are three key stages to our 2.5D$_{pc}$ SIFT implementation: in the first stage the position (x, y), spatial scale σ and canonical in-plane orientation θ of each keypoint is detected using *scale-space* as proposed by Mikolajczyk and Schmid [37]. Since we assume that a single plane can be used to define the 3D orientation of the keypoint support region, single canonical slant ϕ and the canonical tilt τ directions are also extracted for each keypoint. In the second stage, stable feature descriptors are extracted as specified by the $(x, y, \sigma, \theta, \phi, \tau)$ information detected for each keypoint and corrected for 3D foreshortening effects. Finally, in the matching stage, our feature descriptors are compared to those held in a database of trained examples using an approach similar to Lowe's, but extended to cater for the additional 3D orientation information now available.

2.1. Keypoint localisation in range images

First of all, the (x, y) locations 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. 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 z-normalised to have global statistics of $\mu = 0,\ \sigma = 1$ to standardise the (potentially large) dynamic ranges of values present in range images we process. The z-normalised image is then blurred with a factor of 0.5 to suppress aliasing and is up-sampled by a factor of two using linear interpolation. This is followed by the creation of a discrete scale-space representation of the range image using the Gaussian and the Difference-of-Gaussian (DOG) pyramids with sub-interval layers, as proposed by Lowe. The signal maxima and the minima are detected within the DOG scale-space and compared to a *contrast threshold* T of 0.003 to reject potentially unreliable keypoints of low contrast. This T value was tuned by experiment to select a satisfactory number of keypoints from the z-normalised range images. The H (Eq. (2a)), K (Eq. (2b)), $k1$ and $k2$ curvatures are then computed for each sub-level using the first and second Gaussian derivatives $(f_x,\ f_{xx},\ f_{xy},\ f_y,\ f_{yy})$ 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 [38]. By comparing the ratio of the principle curvatures $\frac{k1}{k2}$ to a *curvature threshold* $r = 5$ (again determined by experiment), spatially compact feature locations can be successfully located. As a result, a set of (x, y, σ) values for each keypoint location are detected.

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

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

where (i,j) is the i-th and j-th pixel of the range image. $f_x,\ f_{xx},\ f_{xy},\ f_y$ and f_{yy} denote the first and second Gaussian derivatives at (i,j) position.

Fig. 4a shows a set of keypoints located on the range images using the 2.5D keypoint localisation algorithm and Fig. 4b illustrates the position, scale (shown by the magnitude of the arrows) and the canonical orientation(s) (shown by the directions of the arrows) for each keypoint.

Next, a consistent canonical in-plane orientation θ is assigned to each keypoint location based on the local image gradient orientations (Eq. (3)). Here, multiple canonical orientations can be assigned to a single keypoint, resulting in multiple feature descriptors. An improved version of Lowe's orientation assignment algorithm is used in this work in which a histogram of 360 bins covering the full 360° range of orientations is formulated based on the local image gradients, weighted by their Gaussian weighted magnitudes [7]. The weights of the histogram are stabilised and distributed evenly by convolving the histogram three times using a 1D symmetric Gaussian kernel size = 17, $\sigma = 17$. Peaks are localised within the histogram and the locations of the largest peak, and also any other peak that is within 80% of the magnitude of the largest peak, is interpolated by fitting a quadratic polynomial to the three largest histogram values defining each peak. This process allows keypoint canonical orientations to be estimated to within sub-bin accuracy. The stability and the accuracy of our keypoint orientation estimator is improved over Lowe's standard process by the increased orientation histogram resolution, as well as by the additional smoothing of the histogram. Fig. 5 gives the results from the pilot investigation we conducted in which a patch is synthetically rotated to a known orientation (denoted using black ∘ symbols in the graph) and the orientation is subsequently recovered using our orientation recovery algorithm as described above, shown as red + on the graph, illustrating the stability of our orientation recovery algorithm.

$$\theta_{local} = \tan^{-1}\frac{\partial y}{\partial x} \tag{3}$$

Fig. 6 illustrates the three Euler angle viewpoint rotations and how the 3D orientation of local surface patches can be interpreted in terms of their slant and tilt. In this work, the in-plane (roll) rotation can be estimated by means of the local range image gradient orientations (θ) (also equivalent to the tilt (τ) angles). The out-of-plane (yaw and pitch) rotations can in turn be estimated by interpreting the range surface slant (ϕ) and tilt (τ) angles.

Based on the location (x, y) and the scale σ for each keypoint location, a consistent canonical slant and tilt is assigned based on the surface normals $(Nx,\ Ny,\ Nz)$ [39] and the θ component as described above. The surface normals can be computed from the first Gaussian derivatives of the range images, with the corresponding σ within the scale-space. This is shown in Eq. (4), where f_x and f_y are the first Gaussian derivatives of the range images in the x and y direction respectively. The slant ϕ and tilt τ can then be computed using Eqs. (5) and (6) respectively [32].

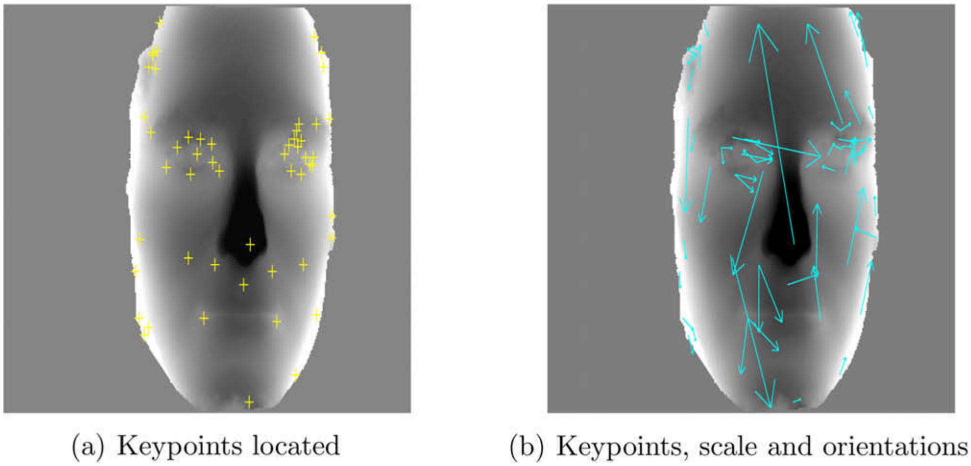


Fig. 4. (a) Location of keypoints (shown as +) extracted using the modified SIFT keypoint localisation algorithm. (b) The scale (demonstrated by the magnitude of the arrows) and the canonical orientation(s) (the directions of the arrows) for each keypoint locations.

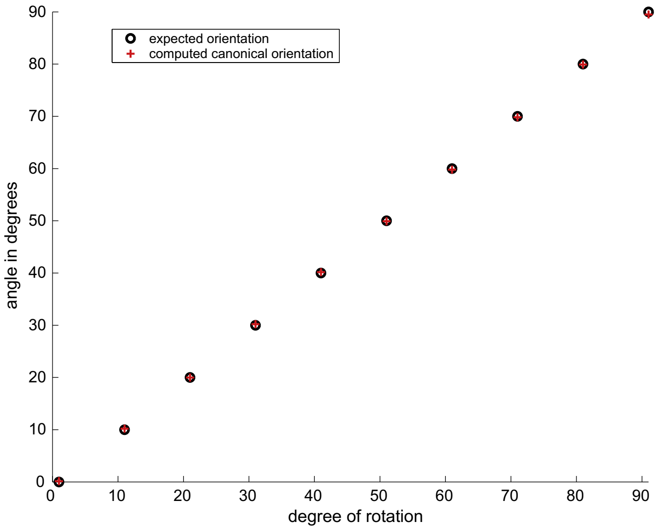


Fig. 5. This graph shows the recovered orientation of a synthetically rotated patch using the orientation recovery algorithm (red +), against known orientation (black ∘). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

$$[Nx, Ny, Nz] = \frac{[-f_x, -f_y, 1]}{\left(1 + f_x^2 + f_y^2\right)^{\frac{1}{2}}} \tag{4}$$

$$\phi = \tan^{-1}\left(\frac{\sqrt{N_x^2 + N_y^2}}{Nz}\right) \tag{5}$$

$$\tau = \tan^{-1}\left(\frac{Nx}{Ny}\right) \tag{6}$$

By comparing the standard equation for orientation, θ, Eq. (3) with that of tilt, τ, reorganised as in Eq. (6), we can observe that θ is in fact a re-expression of τ over a different angle range. Since the circular measurement aperture projects to an ellipse in image space, the slant component determines the aspect ratio of this ellipse, while the tilt component determines the in-plane orientation [32]. Therefore, in the context of 2.5D$_{\text{pc}}$ SIFT, θ and τ would appear at first sight to be mutually redundant. However, as the in-plane orientation component might comprise a mix of competing gradient directions, we have chosen to calculate multiple orientations and express different keypoints assigned to each of these as described above. It is interesting to note that it might be more useful to compute the θ component based on *intensity* information, as albedo variations could be used to detect in-plane rotations on circularly symmetric 3D surfaces that might otherwise be rendered

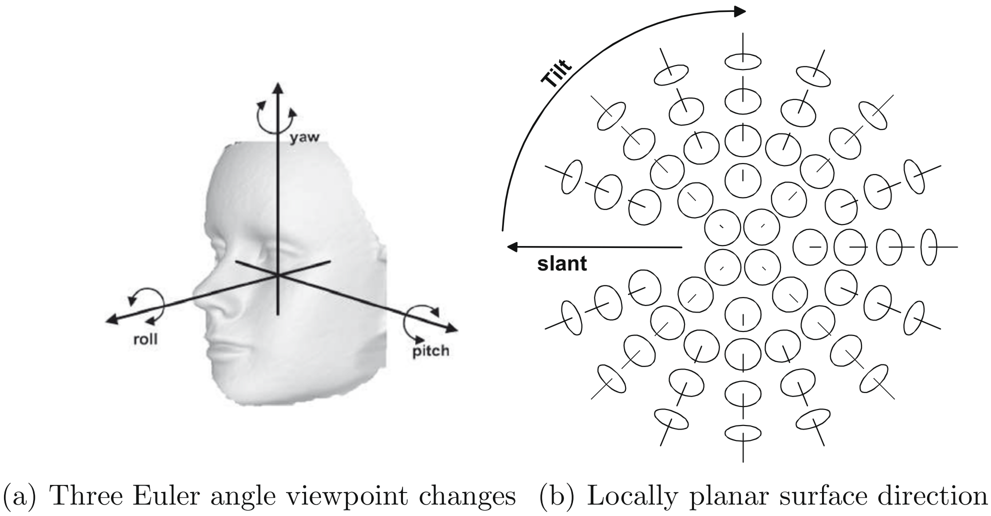

yaw
roll
pitch
Tilt
slant
(a) Three Euler angle viewpoint changes
(b) Locally planar surface direction

Fig. 6. Diagram illustrating (a) 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. (b) Locally planar surface direction encoding by means of slant and tilt angles, reproduced from [32].

invisible in the range domain alone. In this case, the τ component derived from surface normals could take on a different angle from the θ component, derived from intensity gradients.

2.2. Feature descriptor extraction

Based on the $(x, y, \sigma, \theta, \phi, \tau)$ for each keypoint, feature descriptors are extracted at each keypoint location (x, y), over a measurement aperture of σ defining the scale of the keypoint, using the appointed canonical orientation(s) θ and slant ϕ and tilt τ, as follows:

(1) The sampled range image patch is rotated to its canonical in-plane orientation in order to achieve 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 (Eq. (1)). Its magnitude, the degree of curvedness, can also be computed from the first and second Gaussian derivatives.

$$curvedness = \sqrt{2H^2 - K} \tag{7}$$

(2) Nine elliptical Gaussian weighted sub-regions, overlapping by one standard deviation, are placed over each sampling patch, according to the estimated keypoint scale, as shown in Fig. 7. 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 fact, rotation of the sampling patch by θ also renders the τ component to be constant, therefore only the aspect ratio of each sub-sample region then needs be adjusted to accommodate the full Euler's viewpoint rotational changes. Overlapping Gaussian sub-regions minimise the spatial aliasing that occurs during the sampling stages and constitutes a standard approach to ensuring spatial sampling continuity, based on the Gaussian σ. 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 in his *space variant* SIFT implementation [40,41]. Winder and Brown [42] attempt to derive a feature descriptor by learning the underlying structure from a gamut of potential descriptor types and spatial sampling configurations. Since our primary objective had been to determine the utility of range image interpretation in the context of SIFT, we chose to adopt a comparatively uncomplicated sampling configuration that was known to produce viable results. In future studies, it may well be worth investigating a wider range of possible descriptor configurations.

(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 Gaussian weighted degree of curvedness. Similarly, an eight-element histogram, covering the 360° range of orientations, can be formulated, weighted by the Gaussian weighted 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, which is then normalised to unity magnitude.

$$\widehat{LocalHist}_i = (\widehat{H_{i_{surface}}})(\widehat{H_{i_{orientation}}}) \tag{8}$$

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

$$Descriptor_{\theta_{canonical}} = (\widehat{LocalHist}_1, \widehat{LocalHist}_2, \ldots, \widehat{LocalHist}_9) \tag{9}$$

Fig. 8 illustrates four feature descriptors extracted from different keypoint locations on a face range image. The keypoints selected are shown in Fig. 8a, where each corresponding descriptor number in Fig. 8b–e has been labelled accordingly. This demonstrates that the descriptors extracted from different keypoint locations each comprise unique mixes (distributions) of surface types and orientations.

2.3. Matching

In order to determine the similarity between feature descriptors extracted from different range images, we have adopted the keypoint matching scheme proposed by Lowe: the angle (radians) between the descriptors is extracted from different images and ranked in ascending order, thereby allowing a candidate match to be found using the nearest neighbour algorithm. Matches are

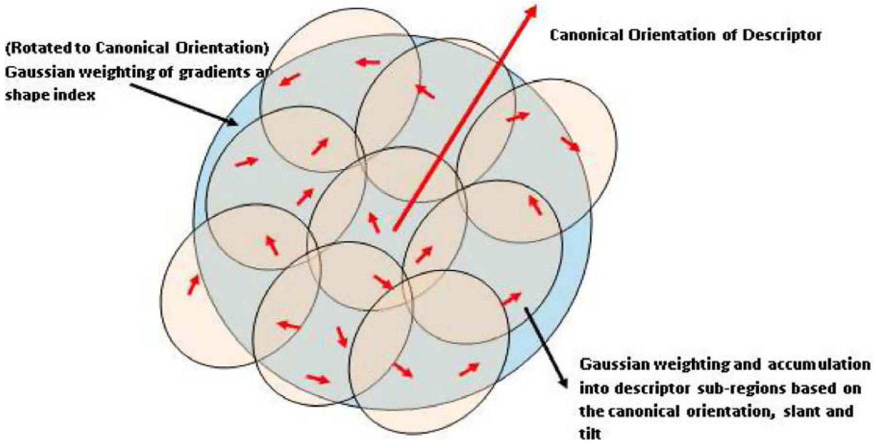
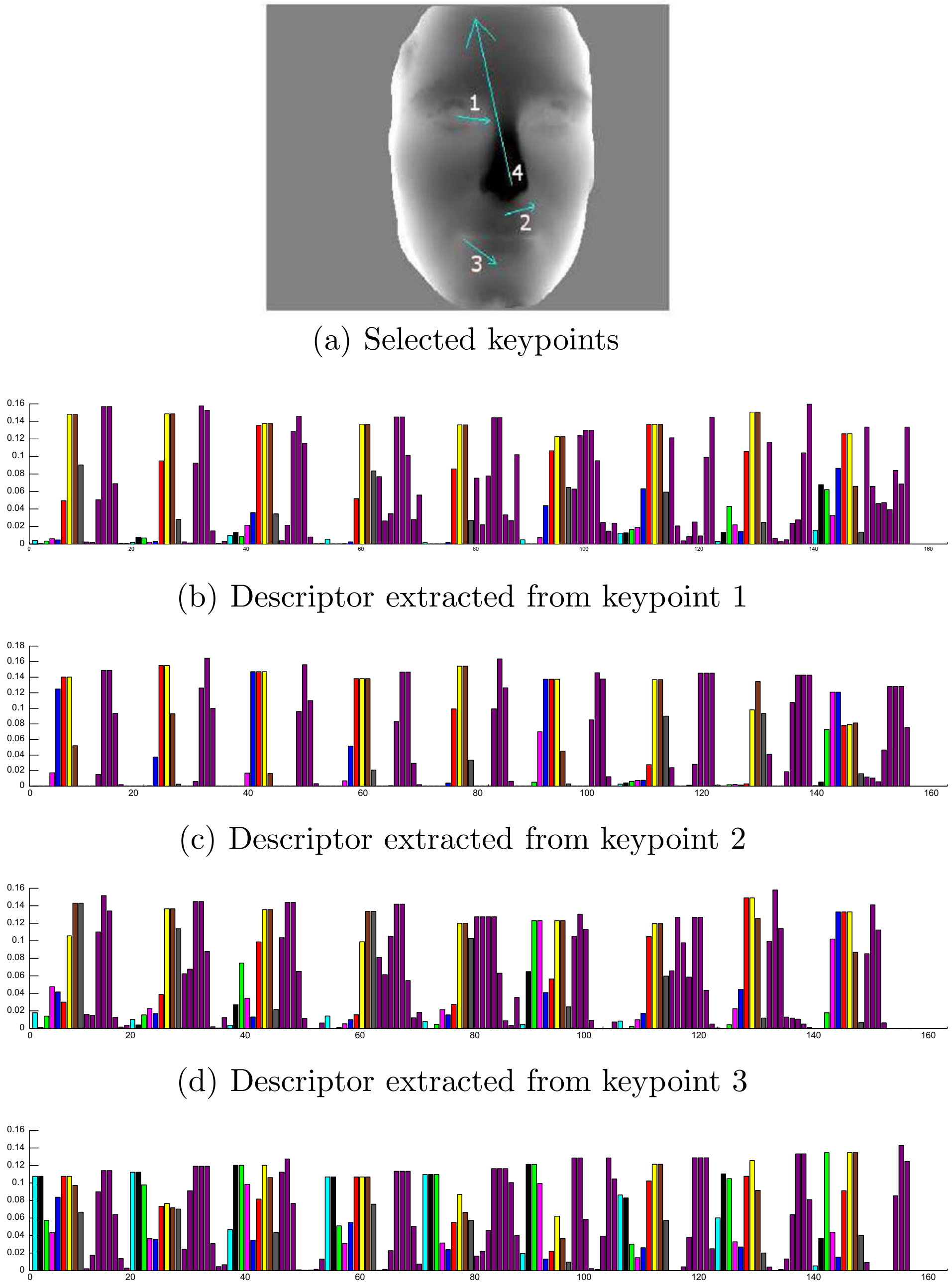

(Rotated to Canonical Orientation)
Gaussian weighting of gradients a
shape index
Canonical Orientation of Descriptor
Gaussian weighting and accumulation into descriptor sub-regions based on the canonical orientation, slant and tilt

Fig. 7. Placement of the nine elliptical sub-regions, with spatial support set at one standard deviation, over the keypoint landmark location is illustrated.



(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

Fig. 8. A selection of descriptors extracted from four different keypoints on the face range image, over nine overlapping sub-regions.

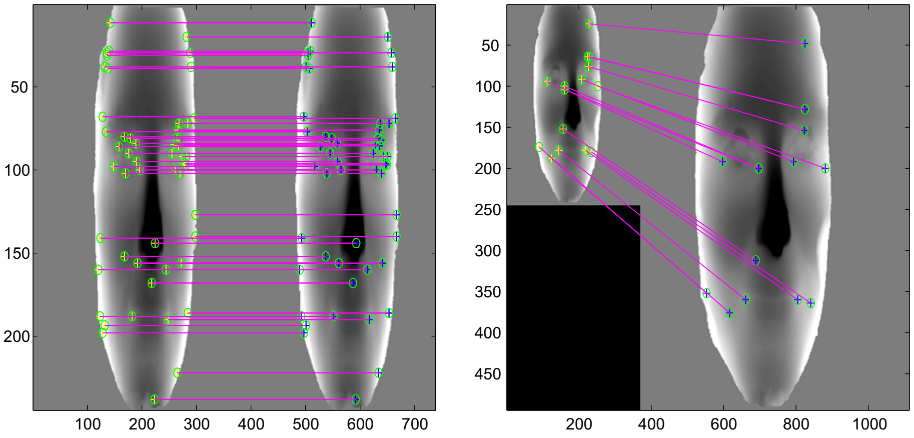
deemed to be unreliable using the ratio test (Eq. (10)), where the ratio between the best matched descriptor and its next best matched descriptor is above a selected threshold, set to be 0.8 in this work.

We have extended the Hough Transform (HT) to verify potential matches between the descriptors, where clusters of matching features having a consistent pose interpretation are identified, i.e. matching keypoints of similar 3D orientation (slant and tilt angles), translation and scale, cluster in Hough space. If there exists three or more entries for each cluster, a robust affine transform fitting algorithm is applied to the cluster in order to recover the affine pose between the matched descriptors and identify outliers. As a result, this enables us to match a set of extracted descriptors to sets of feature descriptors contained in a database. Fig. 9a shows a range image match to an identical copy at the same scale, while Fig. 9b shows the this range image being matched to an enlarged version of itself.

$$Ratio_{distance} = \frac{\theta_{rank_1}}{\theta_{rank_2}} \tag{10}$$

3. Validation

We have conducted an investigation to compare the scale and rotation invariant properties of our proposed 2.5D$_{pc}$ SIFT feature descriptors, when matching range images, with standard 2D SIFT, when matching intensity images. The data set used for this initial



(a) Self-matching, fixed scale

(b) Self-matching, varied scale

Fig. 9. Example of point-to-point matching of range images where (a) shows self-matching a range image and (b) shows this same image being matched to an enlarged version of itself.

investigation comprises a total of 162 range images of a human head. This data set was generated from stereo-pair images captured using two digital cameras each of 13.5-mega pixel resolution and processed using the C3D [4] stereo-photogrammetry package. We captured a baseline image face-on to the cameras, and then at 10° about the yaw axis (to produce out-of-plane rotation of the face in the acquired range images) in both clockwise and anticlockwise directions up to and including 30°. A further data set comprising out-of-plane rotations about the pitch axis was generated by synthetically rotating the frontal face range image up/down at 10° intervals, up to and including ±30°. Pitch rotations are difficult to capture accurately using live subjects, hence these poses were synthesised from real data, as was a set of in-plane rotated data using MATLAB's `imrotate` function. For each range image captured or synthesised (for rotation and scale experiments only), a corresponding 2D intensity image was also captured or synthesised accordingly (85 in total), in order to allow comparisons with standard 2D SIFT matching.

In order to investigate the scale and rotational invariant properties of our $2.5\mathrm{D}_{\mathrm{pc}}$ SIFT, feature descriptors were extracted from each of the above range images using the algorithm described in the previous section. To keep the computational requirements to a minimum, while retaining stable results, the range images have been reduced from their original size of 1498 × 2249 pixel to 244 × 369 pixel. Table 1 shows the data used to validate the $2.5\mathrm{D}_{\mathrm{pc}}$ SIFT system in this paper.

The validation is divided into four parts: we first investigate the viewpoint rotation invariant properties of our system by extracting a set of feature descriptors for each yaw, pitch and roll pose angle sampled. For each rotation axis, all combinations of descriptor sets are then matched against each other. It is then possible to deduce the stability of the feature descriptors by computing a match-matrix showing the percentage of correctly labelled keypoints for each pair of matched images. Furthermore, we use the HT to eliminate matching outliers in Hough space. While the HT itself is not infallible in removing outliers, from visual inspection of the outliers, it does appear to be effective. Although the positive matches detected cannot be guaranteed to be all correct, using this filtered data we can compute a match-matrix that represents more reliably the percentage of correctly labelled keypoints. We can therefore estimate more accurately the discriminability of our 2.5D keypoint descriptors. Using the HT we can also estimate the True Positive Rate (TPR) and the False Positive Rate (FPR) for the descriptors matched between a pair of images. Then for each matched image pair, a point in ROC space can be plotted to allow the reliability of compared sets of images to be gauged.

Secondly, we investigate the scale invariance properties of the feature descriptors by matching images at different scales. This is achieved by enlarging and reducing the size of the range images using a half-octave Gaussian pyramid. Table 2 shows the size of the range images used in this experiment. Using the feature descriptors extracted from the 244 × 369 pixel range images as our baseline, we compare the feature descriptors extracted from the remainder set of data (i.e. same viewpoint angles but of different sizes) to our baseline images, forming a match-matrix that shows the percentage of the correctly labelled keypoints at each scale.

Thirdly, we investigate the behaviour of our $2.5\mathrm{D}_{\mathrm{pc}}$ SIFT with different levels of noise by adding different amounts of noise to the original range images. In this experiment, we add a total of 10% noise to the [0, 1] normalised range images in increments of 1% noise. By comparing each noise-corrupted image to a baseline

Table 1
Validation experiments conducted.

Experiment	System	Images	Rotations
Rotation	$2.5\mathrm{D}_{\mathrm{pc}}$ SIFT	Range images	Out-of-plane (Yaw) Out-of-plane (Pitch) In-plane (Roll)
	2D SIFT	Intensity images	Out-of-plane (Yaw) Out-of-plane (Pitch) In-plane (Roll)
Scale	$2.5\mathrm{D}_{\mathrm{pc}}$ SIFT	Range images	Out-of-plane (Yaw)
	2D SIFT	Intensity images	Out-of-plane (Yaw)
Noise	$2.5\mathrm{D}_{\mathrm{pc}}$ SIFT	Range images	Out-of-plane (Yaw)
Generic	$2.5\mathrm{D}_{\mathrm{pc}}$ SIFT	Range images	Out-of-plane (Yaw)

Table 2
Size of the range images (in pixel) used in this experiment in order to determine the stability of the feature descriptors over scale change.

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

(no noise added) image, it becomes possible to investigate the robustness of our 2.5D$_{\mathrm{pc}}$ SIFT implementation with respect to noise.

Finally, we illustrate the generic utility of our 2.5D $_{\mathrm{pc}}$ SIFT by applying this algorithm to a set of range image scans of manufactured objects (i.e. not comprising human faces), collected by Ohio State University [43]. The results and findings of the experiments outlined above are presented in the next section.

4. Results and analysis

Table 3 sets out the rotation, scale and noise performance comparison tests conducted between 2.5D$_{\mathrm{pc}}$ SIFT and the standard 2D SIFT and summarises the results obtained. The data set used for these experiments comprises range images of a human face, captured at different viewpoint angles. The statistical significance of the observed performance differences between the two different SIFT modalities has been determined by means of the Wilcoxon Matched-Pairs Signed-Rank test, as indicated in Table 3 using 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.

Fig. 10 shows two examples of matching the feature descriptors between different range images captured at different yaw viewpoint angles. The green circles indicate a match whereas the red circles illustrates the outliers identified by the matching algorithm.

Fig. 11a shows the match-matrix obtained by comparing all the combinations of the feature descriptors extracted from range images captured at different yaw angles (from −30° to 30° in both clockwise and anticlockwise directions about the yaw axis) using 2.5D$_{\mathrm{pc}}$ SIFT. The lower axes of this matrix show the viewpoint angles of each pair of the compared range images, while the height of the columns show the percentage of the correctly matched keypoints. Fig. 11b illustrates these results plotted in ROC space. These results show that the majority of combinations of descriptor pairings match. Closer examination of the matrix shows that over 71.4% of the test images are matched correctly at over 60% recognition rate, while 63.2% of the images are matched correctly at over 70% recognition rate. These figures illustrate the rotation invariance properties of 2.5D$_{\mathrm{pc}}$ SIFT feature descriptors for out-of-plane viewpoint changes.

Fig. 12a shows the match-matrix obtained by comparing all the combinations of the feature descriptors extracted from 2D images using standard 2D SIFT [44], captured at different yaw angles (from −30° to 30° in both clockwise and anticlockwise directions about the yaw axis). Fig. 12b illustrates these results in ROC space. These results show that on average, 63.3% of the 2D test images are matched correctly at over 60% recognition rate while only 49% of the images are matched correctly at over 70% recognition rate. In this case the 2.5D$_{\mathrm{pc}}$ SIFT has achieved a higher match performance rate than standard 2D SIFT. Notice also the improved TPR for 2.5D$_{\mathrm{pc}}$ SIFT over that of standard 2D SIFT, indicated by comparing their respective ROC spaces.

Fig. 13a presents the match-matrix obtained by exploring all the combinations of the feature descriptors extracted from the synthetically rotated range images from −30° to 30° about the pitch axis, using 2.5D$_{\mathrm{pc}}$ SIFT. Fig. 13b illustrates these results in a ROC space. Examination of this matrix shows that the average matching rate is 86% with all images matched correctly at over 50% recognition rate.

Fig. 14a presents the match-matrix obtained by exploring all the combinations of the feature descriptors extracted from the synthetically rotated intensity images from −30° to 30° about the pitch axis, using the standard 2D SIFT system. Fig. 13b illustrates these results in a ROC space. Examination of the matrix shows that the average matching rate is 78.8% with 98% of the images matched correctly at over 50% recognition rate.

Again a performance gain in both match rate and ROC space is obtained for 2.5D$_{\mathrm{pc}}$ SIFT over standard 2D SIFT under pitch axis rotation.

Fig. 15a shows the match-matrix obtained by exploring all the combinations of 2.5D$_{\mathrm{pc}}$ SIFT feature descriptors extracted from in-plane rotated range images of a human face, from 0° up to and including 350°. Results show that the average matching rate is 64.9%, with 88% of the images matched at over 50% of keypoint recognition rate.

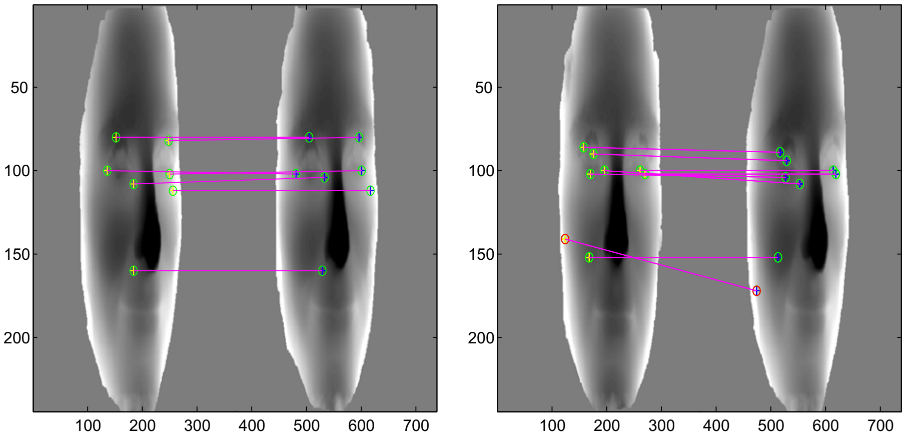
Fig. 16a shows the match-matrix obtained by comparing all the combinations of the feature descriptors extracted from the intensity data of in-plane rotated images of a human face, using standard 2D SIFT. The results illustrate that the feature descriptors are robust to in-plane rotational changes, with on average over 99% of the test images matched at over 50% of keypoint matching rate.

In this case, 2D SIFT achieves a significantly greater match rate than 2.5D$_{\mathrm{pc}}$ SIFT. However, the results plotted within ROC space indicate that the 2D SIFT FPR is much larger than that obtained with 2.5D$_{\mathrm{pc}}$ SIFT, indicating the 2D SIFT results are not as reliable as those of 2.5D$_{\mathrm{pc}}$ SIFT.

Fig. 17 shows the match-matrix obtained by comparing the feature descriptors extracted from the baseline images (of size 244 × 369 pixel and captured at different angles about the yaw axis) to the feature descriptors extracted from the range images of the same viewpoint angle but at a different scale, using 2.5D$_{\mathrm{pc}}$ SIFT. The lower left axis represents the viewpoint angles of each pair of compared range images, while the lower right axis shows

Table 3
Summary of results for the experiments presented in this paper.

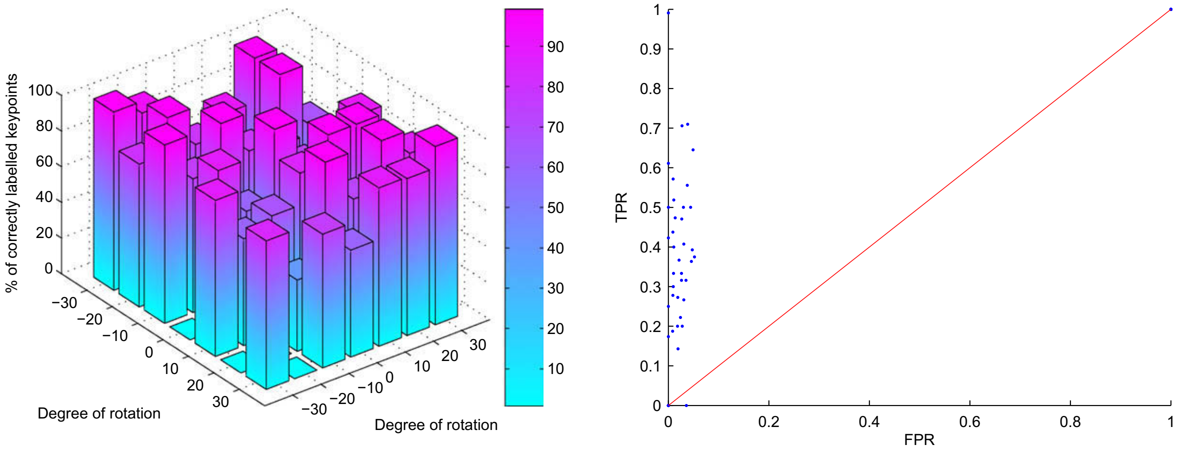
Experiment	System	Rotation	Image pairs	Average matching rate (%)	Image pair match rate > (%)		
					50%	60%	70%
Rotation	2.5D$_{\mathrm{pc}}$	Yaw	49	70.0$\gg$	83.7	71.4	63.3
	2D		49	60.7	79.6	63.3	49.0
	2.5D$_{\mathrm{pc}}$	Pitch	49	86.0$\gg$	100	100	100
	2D		49	78.8	98.0	93.9	81.6
	2.5D$_{\mathrm{pc}}$	Roll	1225	64.9$\ll$	88	63.6	43.7
	2D		1225	84.2	98.9	96.4	88.0
Scale	2.5D$_{\mathrm{pc}}$	Yaw	35	89.5=	100	97.1	97.1
	2D		35	87.0	100	100	97.1
Noise	2.5D$_{\mathrm{pc}}$	Yaw	77	69.4	100	75.4	40.2



(a) -10° compared to 0°

(b) 0° compared to 20°

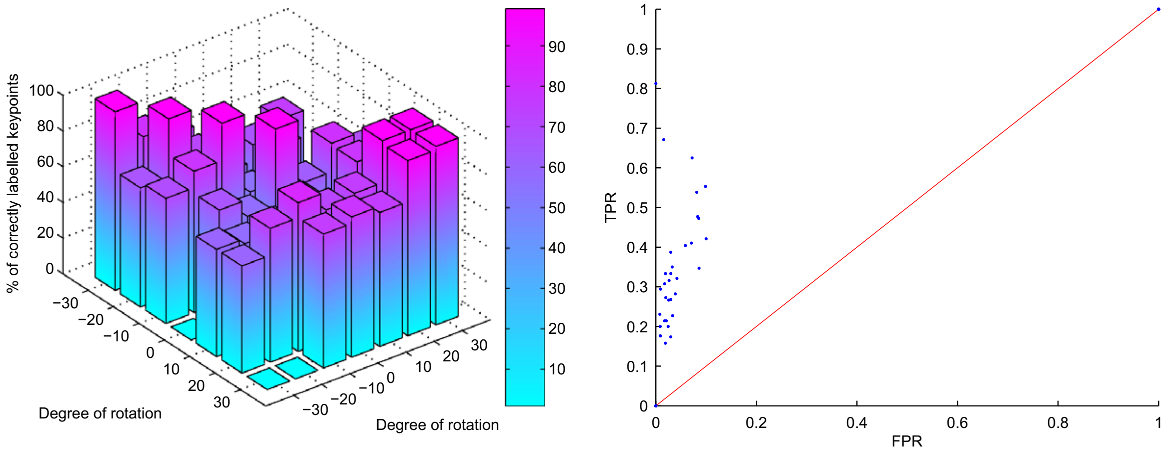
Fig. 10. Matching of the feature descriptors between different range images captured at different yaw viewpoint rotations, using 2.5D$_{pc}$ SIFT, (a) between −10° compared to 0° and (b) between 0° compared to 20°. Green circles illustrate the correct match where red circles show the outliers. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)



(a) Match-Matrix

(b) ROC

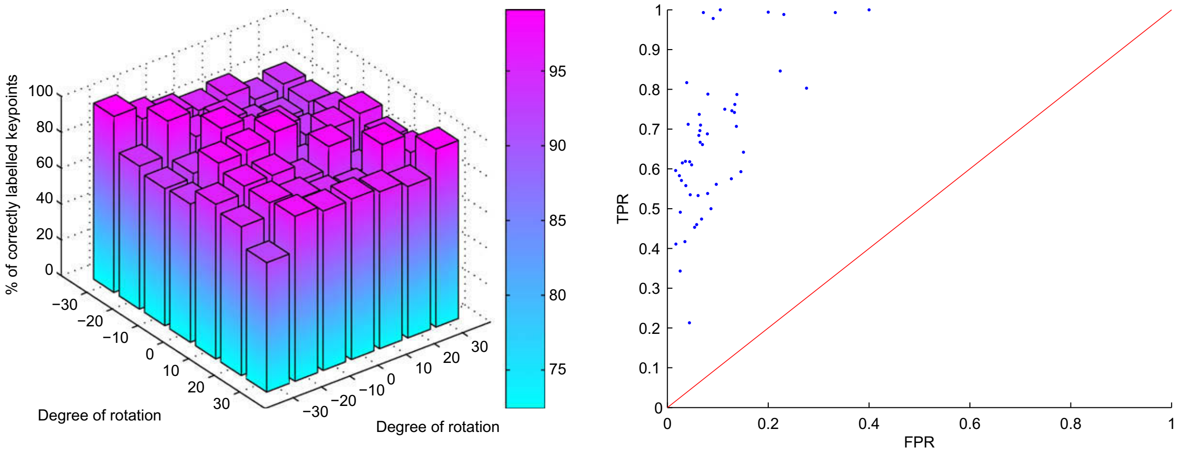
Fig. 11. 2.5D$_{pc}$ SIFT system: (a) The match-matrix results of the percentage of matched keypoints, produced from a range data set of out-of-plane (yaw axis) captured rotations of a human face. (b) The matching results presented in a ROC space.



(a) Match-Matrix

(b) ROC

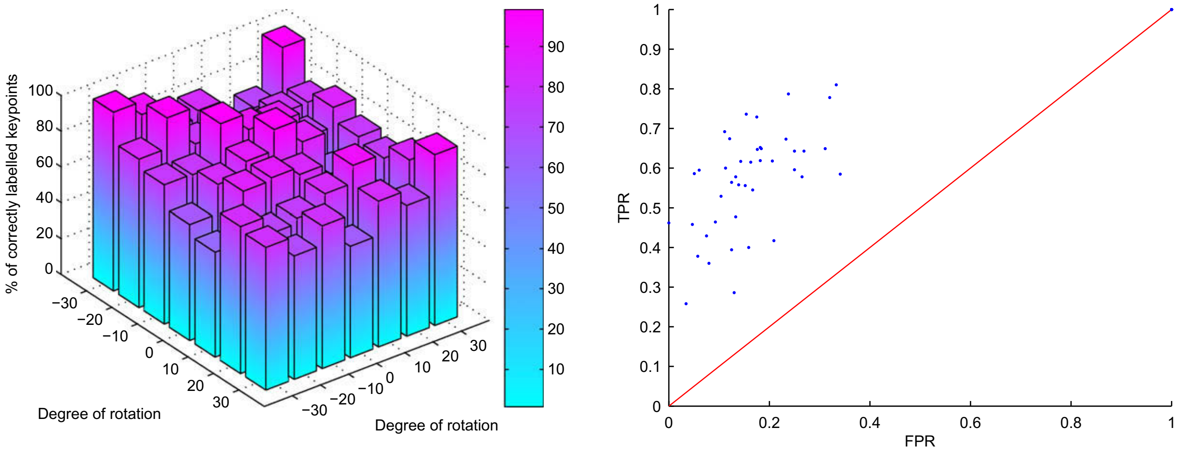
Fig. 12. 2D SIFT system: (a) The match-matrix results of the percentage of matched keypoints, produced from a range data set of out-of-plane (yaw axis) captured rotations of a human face. (b) The matching results presented in a ROC space.



(a) Match-Matrix

(b) ROC

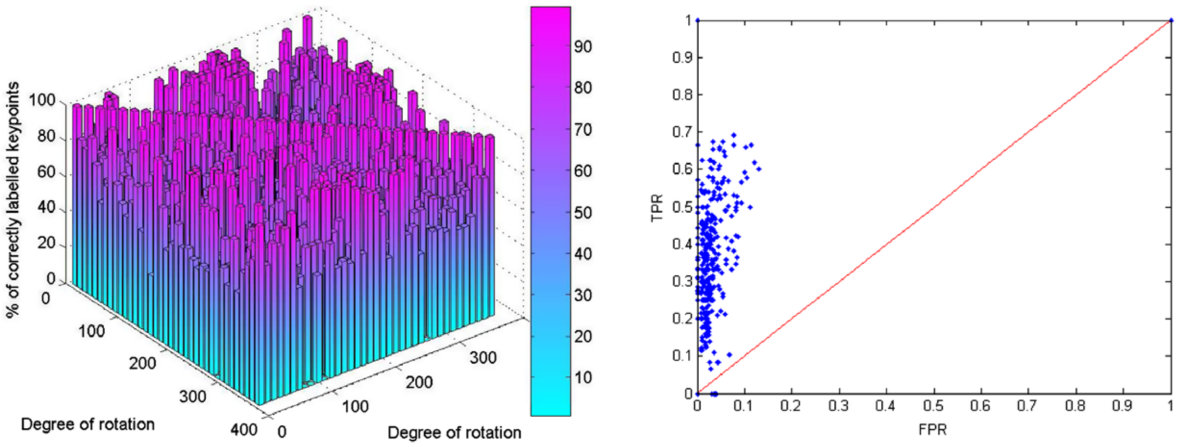
Fig. 13. 2.5D$_{pc}$ SIFT system: (a) The match-matrix results for the percentage of matched keypoints, produced from a range data set of synthetic out-of-plane (pitch axis) rotations of a human face. (b) The matching results presented in a ROC space.



(a) Match-Matrix

(b) ROC

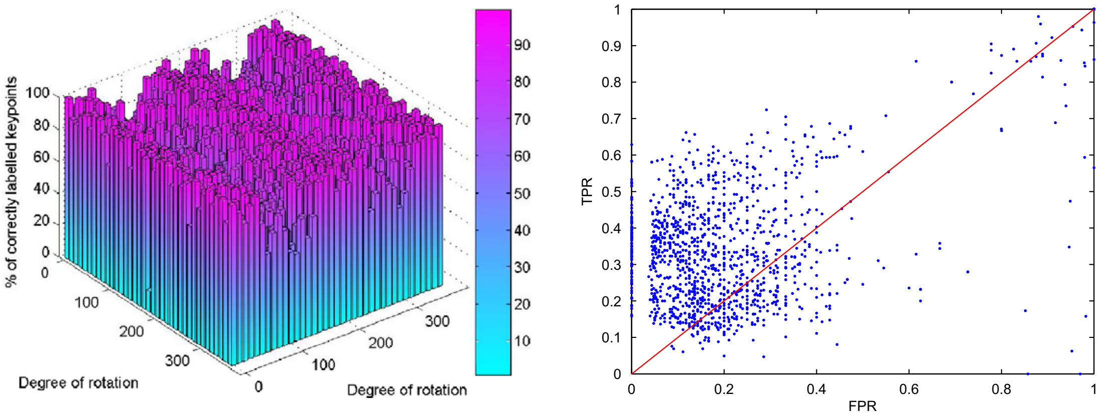
Fig. 14. 2D SIFT system: (a) The match-matrix results for the percentage of matched keypoints, produced from a intensity data set of synthetic out-of-plane (pitch axis) rotations of a human face. (b) The matching results presented in a ROC space.



(a) Match-Matrix

(b) ROC

Fig. 15. (a) The match-matrix results for the percentage of matched keypoints, produced from a set of in-plane rotational *range* 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.



(a) Match-Matrix (b) ROC

Fig. 16. (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.

	1	2	3	4	5
-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

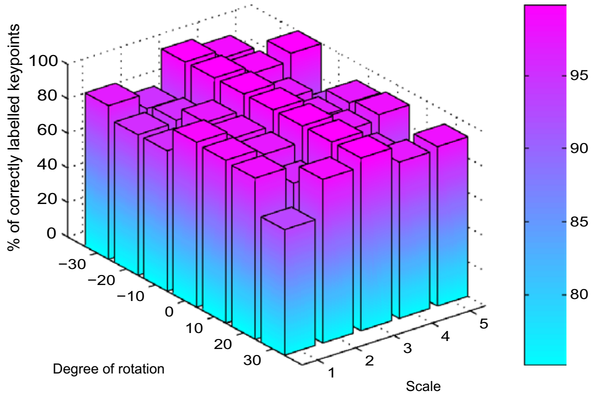


Fig. 17. Match-matrix results for the discriminability of the feature descriptors against scale changes, using 2.5D$_{pc}$ SIFT.

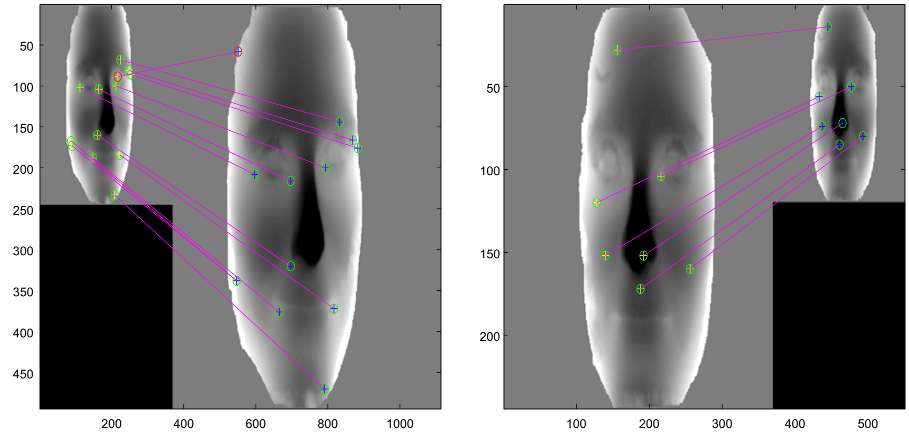
the size of the range images used for the comparison (where the numbers at the axis correspond to the level indicator in Table 2). Similarly, the height of each column indicate the percentage of the correctly matched keypoints at their corresponding scales and angles, compared to the baseline images of 244 × 369 pixel. The results show that the average matching rate is 89.5%, with approximately 97.1% of the range images matched correctly at over 70% recognition rate. This indicates that the feature descriptors demonstrate good invariance to scale changes. Fig. 18 shows two examples of the keypoint matching process. Here, the baseline images (of 244 × 369 pixel) captured at 10° intervals anticlockwise (Fig. 18a) and 10° intervals clockwise (Fig. 18b) are shown on the left image of each respective sub-figure. The compared images, captured at the same angles but taken at different scales, are shown as the right image of each respective sub-figure. The correctly labelled keypoints are illustrated by green circles while outliers are indicated by red circles.

Fig. 19 shows the match-matrix obtained by exploring the feature descriptors extracted from the 2D intensity baseline image (of size 244 × 369 pixel and captured at different angles about the yaw axis) to the feature descriptors extracted from the images of the same viewpoint angle but at a different scale, using the standard 2D SIFT system. Closer examination of this matrix shows that the average matching rate is 87%, with 97.1% of the images matched correctly at 70% recognition rate.

These results illustrate that both the 2.5D$_{pc}$ and 2D SIFT systems exhibit good scale invariance over the majority of the compared pairings.

Fig. 20 shows the match-matrix obtained by comparing the feature descriptors extracted from the baseline images (of size 244 × 369 pixel, captured at different angles about the yaw axis, with no noise added to the images) to the feature descriptors extracted from the range images of the same viewpoint angles and scale but with a known added amount of noise. The lower left axis represents the viewpoint angles of each pairing, while the lower right axis shows the percentage of noise added to the original images. Similar to the previous results, the height of each column indicates the percentage of correctly labelled keypoints. Results show that, on average, over 60% of keypoints are correctly labelled with 10% of noise added to the image. This indicates the degree to which our 2.5D$_{pc}$ SIFT system is robust to noise.

Overall, the results presented indicate that 2.5D$_{pc}$ SIFT outperform 2D SIFT for out-of-plane pose angle changes. Since the degree of performance gain of 2.5D$_{pc}$ SIFT over 2D SIFT is greater for real rotations as opposed to synthesised rotations, this is particularly encouraging as the presence of image noise is assumed to be the primary reason for the performance differential. In both cases, 2.5D$_{pc}$ SIFT produces more compact ROC spaces, indicating a better TPR and hence more reliable matches. Closer examination of the match-matrices for out-of-plane rotation comparisons reveals a tendency for match performance to fall off in the matrix 1st and 3rd quadrants, this is as expected as these regions compare range images presenting *opposing* rotational viewpoint directions. In contrast, the in-plane match rate for 2.5D$_{pc}$ SIFT appears to be much lower than that of 2D SIFT, while the ROC space for 2.5D$_{pc}$ SIFT appears to yield a far better TPR. This result may suggest that the



(a) Baseline image compared to a reduced image, captured at -10° yaw axis rotation

(b) Baseline image compared to an enlarged image, captured at 10° yaw axis rotation

Fig. 18. (a) Feature descriptors matching between the baseline (left) image of size 244 × 369 pixel to a reduced image of size 119 × 181 pixel, captured at 10° anticlockwise rotation in yaw. (b) Feature descriptors matching between the baseline (left) image to an enlarged image of size 494 × 744 pixel, captured at 10° rotation in yaw. The green circles indicate correct matches while red circles denote outliers. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

	1	2	3	4	5
-30°	81	82.1	100	81.5	70
-20°	84.5	84.3	100	88.2	74.1
-10°	80.6	88.5	100	79.4	79.2
0°	83.1	74.5	100	75.6	87.5
10°	78.3	92	100	91.4	84.8
20°	85.5	91.7	100	100	87
30°	87.3	90.3	100	85.7	78.1

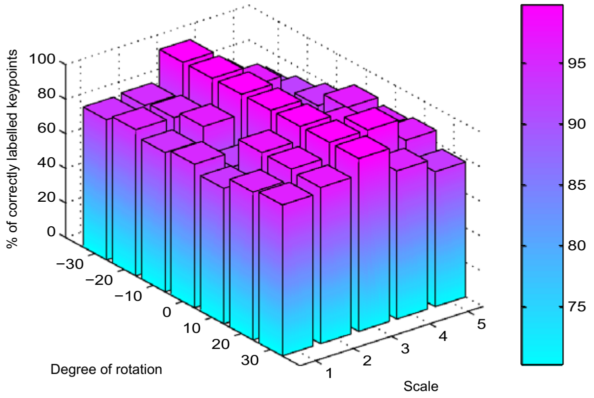


Fig. 19. Match-matrix results for the discriminability of the feature descriptors against scale changes, using 2D SIFT.

	0	1	2	3	4	5	6	7	8	9	10
-30°	100	84.1	71.9	65.2	57.9	64.3	71	64	70	63.2	66.7
-20°	100	79.3	70.2	62.5	58.9	62.8	65.7	54.5	66.7	65.5	70.8
-10°	100	76.6	65.2	54.1	73.7	76.2	77.8	53.1	78.6	65	56.5
-0°	100	73.9	79.2	57.8	71.1	61.3	56.7	60.9	60	56	66.7
10°	100	71.4	71.4	61.9	67.5	75	54.5	57.1	55.2	73.7	82.4
20°	100	75	67.2	63	52.4	65.9	57.7	88.2	85.3	57.1	63
30°	100	76.6	70.6	66.7	62.5	54.2	69.2	63.6	58.8	62.5	57.9

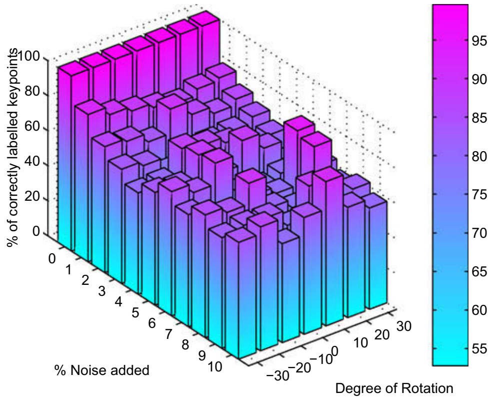
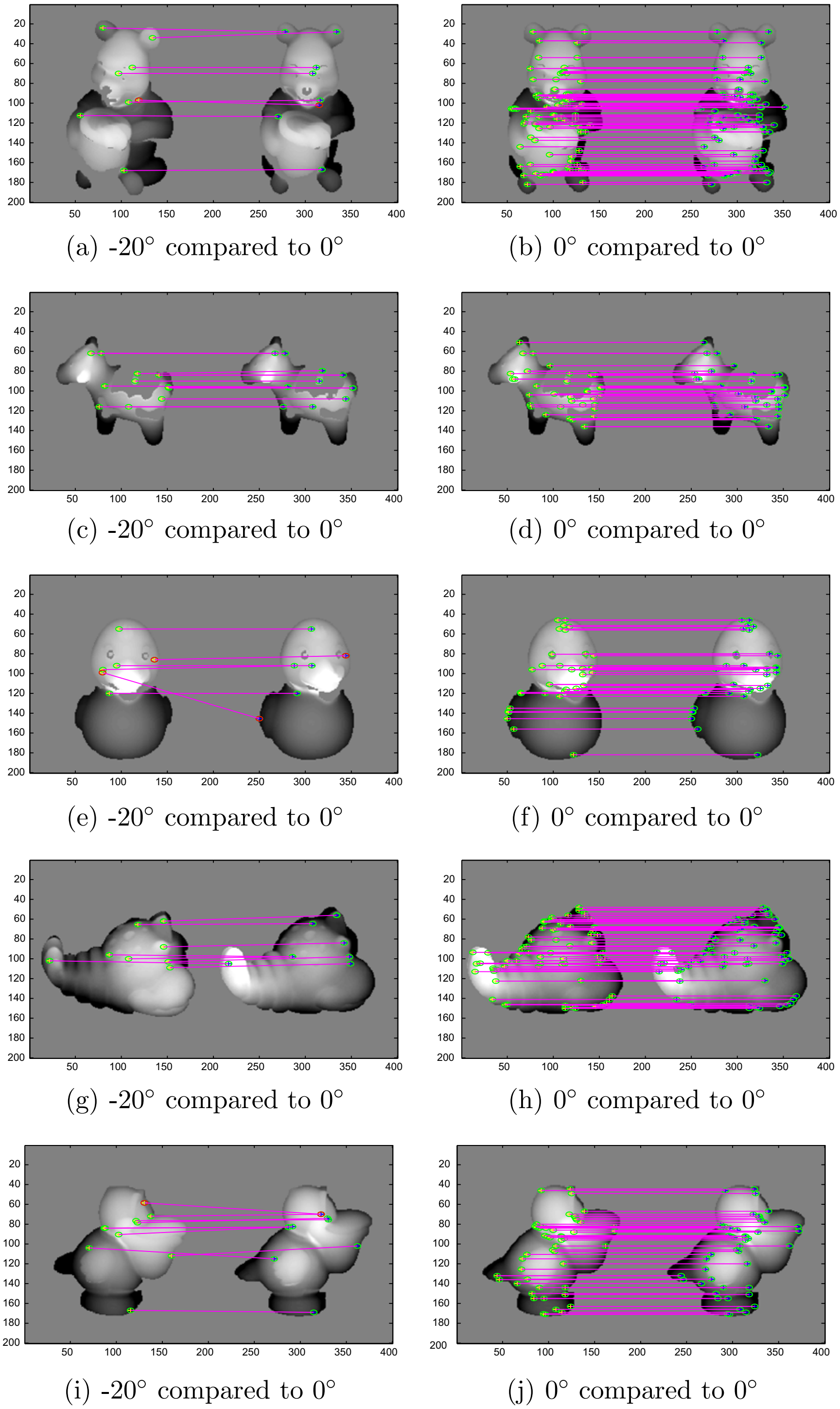


Fig. 20. Match-matrix results for the discriminability of the $2.5D_{pc}$ SIFT feature descriptors against noise.

significantly different feature extraction sampling structure of $2.5D_{pc}$ SIFT is more discriminating, i.e. is more sharply tuned in orientation, resulting in comparatively fewer matches, although these matches are of higher reliability. Further investigation is required to resolve this issue. In terms of scale invariance, both the $2.5D_{pc}$ SIFT and 2D SIFT systems produced essentially the same good scale invariance results, which is less surprising since they share the same scale estimation mechanism. The noise performance of

$2.5D_{pc}$ SIFT indicates that match performance degrades gracefully as the range image noise increases. Even in the presence of severe noise (10%), 60% of the compared images achieve a descriptor match rate above 75%.

Finally, Fig. 21 shows the results of matching a selection of range images (of manufactured objects) captured for either different pose angles (left hand column) or the same pose angle, i.e. matching identical range maps (right hand column), using our $2.5D_{pc}$ SIFT. The range maps have been extracted from the Ohio State University range map collection [43] and comprise 200×200 pixel resolution range images. In Fig. 21, we observe matching results consistent with those obtained for matching identical face range maps (Fig. 9a), where many matches are obtained. We then observe a similar reduction in the number of matches obtained under pose change, as found when matching the face in Fig. 10. Despite this observed reduction in the quantity of matches, sufficient numbers of matches do survive to support subsequent Hough Transform processing. An explanation for the



(a) -20° compared to 0°

(b) 0° compared to 0°

(c) -20° compared to 0°

(d) 0° compared to 0°

(e) -20° compared to 0°

(f) 0° compared to 0°

(g) -20° compared to 0°

(h) 0° compared to 0°

(i) -20° compared to 0°

(j) 0° compared to 0°

Fig. 21. Different examples of matching feature descriptors between range images captured at different pose rotations, using $2.5D_{pc}$ SIFT, left column: between −20° compared to 0°; right column: self-matching. Green circles illustrate correct matches and red circles show outliers. Range images obtained from [43]. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

reduction in the numbers matching keypoints under rotation is as follows: by inspecting Fig. 21, it would appear that for self-matching there are many matches on the occluding boundary of each object (where there are many surface inflections that excite the keypoint detector). Of course, when the object rotates the boundaries between the compared range surfaces are no longer the same, and therefore the number of matches fall. An examination of the surviving matches indicates that these tend to be interior to the object surface. Likewise, many of the interior keypoint-pairs that no longer match under rotation appear to be located on self-occluding boundaries, i.e. where the range surface is discontinuous and therefore unstable in appearance under rotation. This is a well known issue for local matching techniques and is also manifest in 2D SIFT.

Overall, the results of matching range maps depicting manufactured objects are encouraging given that this data was available at a much lower resolution than that employed in our experiments using our own high quality C3D capture facility. Furthermore, the image quality, imaging geometry and noise properties of the specific Minolta capture device used to collect these range images were unknown to us. These results illustrate the generic nature of our $2.5D_{pc}$ SIFT algorithm and therefore its potential to be applied to range images depicting other free-form objects in addition to human faces.

5. Conclusions and future work

In this paper, we present a full adaptation and implementation of the SIFT algorithm in the 2.5D domain. Stable keypoints are located on the range images using the standard search over scale, where the (x, y) locations, along with their appropriate scale σ are defined. Furthermore, the local canonical orientation(s) and the canonical slant and tilt are computed for each keypoint location, recovering the $(x, y, \theta, \phi, \tau)$ values for each keypoint within the range image. Robust 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 is first 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 allows the 3D pose to be estimated and corrected for each keypoint on the range image, and hence provides more stable and representative descriptors than could be achieved using standard SIFT applied to intensity images.

We demonstrate our feature descriptors encapsulating the underlying surface information of the range images at each sampling point, categorising the surface shape in terms of local topology statistics and range surface gradient orientations. By incorporating shape index within our feature descriptor, we have a means of classifying the local range surface shape that can provide strong discrimination between different surface shapes. To obtain good discrimination properties, we observed that the feature descriptor formulation should encode information pertaining to both the distribution of the observed intrinsic image property (surface topology type) and also the distribution of its spatial configuration, in this case orientation direction.

In our pilot study, we demonstrated that by matching the feature descriptors extracted from range images of different sizes and captured at different 3D pose angles, $2.5D_{pc}$ SIFT could match a greater proportion of descriptors than standard 2D SIFT could using the equivalent intensity images under out-of-plane rotation. $2.5D_{pc}$ SIFT and 2D SIFT appeared to be broadly equivalent for scale invariance, although $2.5D_{pc}$ SIFT appeared to yield more reliable matches. An unexpected result was that $2.5D_{pc}$ SIFT matched significantly fewer descriptors than 2D SIFT under in-plane rotation, however, the matches produced by $2.5D_{pc}$ SIFT appear to have a very significantly improved TPR over those produced by 2D SIFT. Moreover, we demonstrated in this work that our $2.5D_{pc}$ SIFT system exhibits good performance under significant noise. The results obtained by matching range images of manufactured objects corroborate with those obtained in the larger investigation of face range images, indicating that the $2.5D_{pc}$ SIFT approach is applicable to free-form object recognition and is not restricted to the face domain. These results also illustrate the importance of range surface topographic features being present that maintain surface continuity, in order to provide appearance stability (and therefore matching stability) under rotational pose change. In the range domain, surface discontinuities can be determined directly (or manifest during stereo-pair image matching as occlusions) and therefore, the potential exists to suppress keypoint detection at these unstable feature locations.

In conclusion, $2.5D_{pc}$ SIFT appears to produce consistently more reliable feature descriptor matches compared to 2D SIFT, although the relative percentages of keypoint matches may be smaller for in-plane rotation. Overall, we believe these represent encouraging results that demonstrate the potential benefits of adopting the range imaging modality.

In the next stages of our work, we propose to explore an architecture that combines the 2D and 2.5D analyses in order to exploit the advantageous properties specific to each modality. Orientation information extracted from the intensity domain has the potential to capture surface rotations revealed by intensity gradients generated from surface albedo discontinuities. On the other hand, the slant and tilt information that can be extracted in the range domain could also serve to correct the sampling window aspect ratio of the 2D descriptors, for example, as in Viewpoint-Invariant Patches (VIP) [45]. The work presented here addresses rigid-body range surface matching, which might serve to initialise close registration methods such as ICP or indeed non-rigid registration techniques, such as the Thin Plate Spline (TPS). We intend to investigate mechanisms for representing local and global biological variability, allowing machine interpretation to be performed on range images and intensity images representing human or animal surface anatomy.

References

[1] D.G. Lowe, Distinctive image features from scale-invariant keypoints, International Journal of Computer Vision 60 (2) (2004) 91–110.

[2] K.W. Bowyer, K. Chang, P. Flynn, A survey of approaches and challenges in 3D and multi-modal 3D + 2D face recognition, Computer Vision and Image Understanding 101 (1) (2006) 1–15.

[3] J.P. Siebert, S.J. Marshall, Human body 3D imaging by speckle texture projection photogrammetry, Sensor Review 20 (3) (2000) 218–226.

[4] X. Ju, T. Boyling, J.P. Siebert, A high resolution stereo imaging system, in: Proceedings of 3D Modelling 2003, 2003.

[5] C. Hesher, A. Srivastava, G. Erlebacher, A novel technique for face recognition using range imaging, in: Seventh International Symposium on Signal Processing and Its Applications, 2003, pp. 201–204.

[6] G.G. Gordon, Face recognition based on depth and curvature features, in: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 1992, pp. 108–110.

[7] T.W.R. Lo, J.P. Siebert, A.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, 2007, pp. 73–82.

[8] T.W.R. Lo, J.P. Siebert, SIFT keypoint descriptors for range image analysis, Annals of the BMVA 2008 (3) (2008) 1–17.

[9] J.J. Koenderink, A.J. van Doorn, Surface shape and curvature scales, Image and Vision Computing 10 (8) (1992) 557–565.

[10] J.W.H. Tangelder, R.C. Veltkamp, A survey of content based 3D shape retrieval methods, in: In Shape Modeling International, 2004, pp. 145–156.

[11] J. Matas, O. Chum, M. Urban, T. Pajdla, Robust wide baseline stereo from maximally stable extremal regions, in: Proceedings of British Machine Vision Conference, 2002, pp. 384–396.

[12] K. Mikolajczyk, C. Schmid, A performance evaluation of local descriptors, IEEE Transactions on Pattern Analysis and Machine Intelligence 27 (10) (2000) 218–226.

[13] H. Bay, T. Tuytelaars, L.V. Gool, Surf: speeded up robust features, in: European Conference on Computer Vision, 2006, pp. 404–417.
[14] E. Tola, V. Lepetit, P. Fua, A fast local descriptor for dense matching, in: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2008, pp. 1–8.
[15] M.P. do Carmo, Differential Geometry of Curves and Surfaces, Prentice-Hall, 1976.
[16] D.J. Ittner, A.K. Jain, 3-D surface discrimination from local curvature measures, in: Proceedings of the IEEE Conference on Computing Vision and Pattern Recognition, 1985, pp. 119–123.
[17] P.J. Besl, R.C. Jain, Three-dimensional object recognition, ACM Computing Surveys 17 (1) (1985) 75–145.
[18] J.C. Lee, E. Milios, Matching range images of human faces, in: Proceedings of the IEEE International Conference on Computer Vision, 1990, pp. 722–726.
[19] Y. Wang, C. Chua, Y. Ho, Facial feature detection and face recognition from 2D and 3D images, Pattern Recognition Letters 23 (2002) 1191–1202.
[20] A.B. Moreno, Á. Sánchez, J.F. Vélez, F.J. Díaz, Face recognition using 3D surface-extracted descriptors, in: Proceedings of Irish Machine Vision and Image Processing Conference, 2003.
[21] C. Xu, Y. Wang, T. Tan, L. Quan, Automatic 3D face recognition combining global geometric features with local shape variation information, in: Proceedings of International Conference on Automated Face and Gesture Recognition, 2004, pp. 308–313.
[22] X. Lu, A.K. Jain, Automatic feature extraction for multiview 3D face recognition, in: Proceedings of International Conference on Automatic Face and Gesture Recognition, 2006, pp. 585–590.
[23] C.S. Chua, R. Jarvis, Point signatures: a new representation for 3D object recognition, International Journal of Computer Vision 25 (1) (1997) 63–85.
[24] L. Wiskott, J.M. Fellous, N. Krűger, C. von de Malsburg, Face recognition by elastic bunch graph matching, IEEE Transactions on Pattern Analysis and Machine Intelligence 19 (7) (1997) 775–779.
[25] C. Dorai, A.K. Jain, Shape spectrum based view grouping and matching of 3D free-form objects, IEEE Transactions on Pattern Analysis and Machine Intelligence 19 (10) (1997) 1139–1146.
[26] G. Hetzel, B. Leibe, P. Levi, B. Schiele, 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, 2001, pp. 394–399.
[27] H. Chen, B. Bhanu, 3D free-form object recognition in range images using local surface patches, in: Proceedings of the International Conference on Pattern Recognition, vol. 3, 2004, pp. 136–139.
[28] A.E. Johnson, Spin-images: a representation for 3D surface matching, Ph.D. thesis, Robotics Institute, Carnegie Mellon University, Pittsburgh, PA , August 1997.
[29] A.E. Johnson, M. Hebert, Using spin images for efficient object recognition in cluttered 3D scenes, IEEE Transactions on Pattern Analysis and Machine Intelligence 21 (5) (1999) 433–449.
[30] C. Dorai, A.K. Jain, COSMOS – a representation scheme for 3D free-form objects, IEEE Transactions on Pattern Analysis and Machine Intelligence 19 (10) (1997) 1115–1130.
[31] S. Pansang, B. Attachoo, C. Kimpan, M. Sato, 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) (2005) 268–277.
[32] J.F. Norman, J.T. Todd, H.F. Norman, A.M. Clayton, T.R. McBride, Visual discrimination of local surface structure: slant, tilt and curvedness, Vision Research 46 (2006) 1057–1069.
[33] E. Akagunduz, I. Ulusoy, 3D object representation using transform and scale invariant 3D features, in: Proceedings of ICCV'07 Workshop on 3D Representation for Recognition (3dRR-07), 2007, pp. 1–8.
[34] X.J. Li, I. Guskov, 3D object recognition from range images using pyramid matching, in: Proceedings of ICCV'07 Workshop on 3D Representation for Recognition (3dRR-07), 2007, pp. 1–6.
[35] T.W.R. Lo, J.P. Siebert, A.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, 2006, pp. 75–82.
[36] T.W.R. Lo, Feature extraction for range image interpretation using local topology statistics, Ph.D. thesis, University of Glasgow, January 2009.
[37] K. Mikolajczyk, C. Schmid, Scale and affine invariant interest point detectors, International Journal of Computer Vision 60 (1) (2004) 63–86.
[38] D. Marr, Vision, W.H. Freeman and Co., 1982.
[39] C.J. Sze, H.Y.M. Liao, H.L. Hung, K.C. Fan, J.W. Hsieh, Multiscale edge detection on range images via normal changes, IEEE Transactions on Circuits and Systems – II: Analog and Digital Signal Processing 45 (8) (1998) 1087–1092.
[40] S. Balasuriya, A computational model of space-variant vision based on a self-organised artificial retina tessellation, Ph.D. thesis, University of Glasgow, March 2006.
[41] L.S. Balasuriya, J.P. Siebert, Hierarchical feature extraction using a self-organised retinal receptive field sampling tessellation, Neural Information Processing – Letters & Reviews 10 (4–6) (2006) 83–95.
[42] S.A.J. Winder, M. Brown, Learning local image descriptors, in: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2007, pp. 1–8.
[43] O.S. University, Osu Range Image Database, 1999. <http://sampl.ece.ohio-state.edu/data/3DDB/RID/minolta/>.
[44] D.G. Lowe, Demo Software: SIFT Keypoint Detector, 2005, Demo available for download at <http://www.cs.ubc.ca/~lowe/keypoints/>.
[45] C. Wu, B. Clipp, X. Li, J. Frahm, M. Pollefeys, 3D model matching with Viewpoint-Invariant Patches (VIP), in: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2008, pp. 1–8.