Interest-point Based Face Recognition from Range Images

F. R. Al-Osaimi, M. Bennamoun and A. Mian
The University of Western Australia
35 Stirling Highway, Crawley, WA 6009, Australia

Abstract

We present a novel approach to interest-point detection tailored to range images. A range image is represented by two images with blob-like patterns that have easily detectable peaks and can be efficiently extracted using convolution kernels. These kernels were designed to produce repeatable and independent blob-like patterns when convolved with the range image. The interest-points correspond to peaks of the patterns after dropping the unstable ones and performing Non-Maximal Suppression (NMS) on their union. The approach was applied to facial range images from the FRGC V2.0 dataset and about 88% repeatability was achieved. Face recognition was also performed by matching the local range regions around the interest-points. An approach based on three levels of matching combined with RANSAC algorithm was used to increase the correct matches and reduce the false ones. Preliminary recognition results for a database of 466 subjects and 1765 probes were 96.33% identification rate and 90% verification rate at 0.1% False Accept Rate (FAR) for faces under neutral expression.

1 Introduction

In recent years, the paradigm of object recognition by matching local regions around interest-points (point features) has been the focus of research in computer vision, especially in 2D recognition. This paradigm has many vital advantages over the classical recognition approaches. For example, it is robust to occlusions and does not require object/background segmentation. We believe that it can be also advantageous in the context of 3D face recognition as it has been shown that recognition by matching local regions [1] or point features [5] is more robust to makeup and facial expressions. However, applying this paradigm to 3D face recognition requires a suitable interest-point detection approach.

In spite of that, some approaches to 3D face recognition are based on matching local regions. In the approach by Moreno et al. [3] local regions are segmented according to the mean and the Gaussian curvatures and the segmented regions are then matched against each other. Errors in the curvatures (which are sensitive to noise) may affect the segmentation of the local regions. Elastic Bunch Graph Matching (EBGM) which matches local regions was extended to face recognition from integrated 2D and 3D images [4]. Although EBGM is a successful face matcher, it is based on a fixed small number of points. Consequently, it is not as robust as interest-point based recognition. In the approach by Mpiperis et al. [5] local features are computed around all the 3D points.

BMVC 2007 doi:10.5244/C.21.34

It is tempting to apply the existing interest-point detection approaches that are designed for 2D images to range images (e.g. the SIFT [6]). However, the nature of 2D images is different from range images and this difference can affect the applicability of these approaches to range images. Firstly, these existing approaches usually rely on salient features which are usually induced by texture such as edges, corners and/or blobs. On the other hand, pixels (range values) of range images smoothly vary. Consequently, range images may not have sufficient easily detectable interest-points. For example, there is a very limited number of such features in a range image of a human face. Applying a 2D image interest-point detection technique to such range image may result in either an insufficient number of interest-points or unreliable ones depending on the selection of the tuning parameters. Secondly, under rigid transformations, the appearance of an object in range images vary differently from its appearance in 2D images. In a 2D image the value of a pixel representing a point on the object generally remains constant with rigid transformations (apart from the effects of illumination) whilst the pixel value in range images varies accordingly.

In our approach to interest-point detection for range images, we represent the range image by multiple images of blob-like patterns from which a sufficient number of repeatable interest-points can easily be detected at the peaks of these patterns. The patterns in the representing images are independent from each other in the sense that they can define different interest-points based on different 3D surface information. The representing images are efficiently extracted by convolving the range image with kernels of sufficiently large sizes and non-overlapping spatial spectrums. Such kernel sizes suit range images especially when the range image lacks sufficient salient features as they cover larger pixel neighborhoods and generate response based on more surface information. In addition, the kernels are robust to object translations and rotations within $\pm 15^{\circ}$.

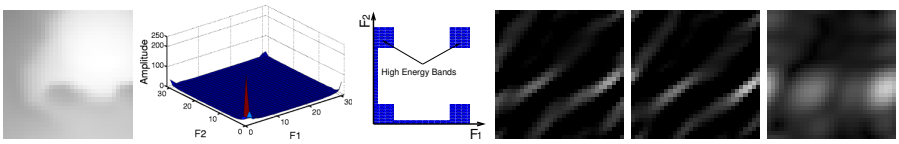
2 Input Range Images

The input range images are computed from the frontal facial 3D pointclouds of the FRGC v.02 dataset [7]. The data is in the form of three matrices x, y and z. The spikes are removed by dropping the outlier points from the three matrices based on local statistics. The matrices are then smoothed using a mean filter which neglects the missing points. After that the holes are filled using bi-cubic interpolation of the missing points. The range image was computed from the three matrices by interpolating for integral x and y coordinates and storing the corresponding z coordinates in the range image matrix using x as horizontal index and y as a vertical index. Finally, the range image is smoothed using a Gaussian filter.

3 Interest-points Detection for Range Images

The summation of a group of adjacent spatial frequencies forms spatial beats in a similar manner to the well-known sound beats phenomenon, as the differences in the spatial frequencies also cause repeated patterns of constructive and destructive interferences over spatial distances. Depending on the phase and amplitude values of the spatial frequencies, the spatial beats can shift, intensify and/or merge. From these spatial beats, the representing 2D blob-like images can be computed (see Fig. 1).

According to 2D Discrete Fourier Transform (DFT), a range image $I(x,y)$ of size $N \times M$ can be represented as the summation of complex exponents as in the well-known



(a) (b) (c) (d) (e) (f)

Figure 1: (a) is a window from a facial range image (part of the nose appears in the widow). (b) Discrete Fourier transform (DFT) of the range window. (c) shows the frequency bands which have high energy. (d) and (e) are real and imaginary spatial beats generated by taking inverse DFT of a subset of 3×3 adjacent frequencies. (f) a window of blob-like patterns that was generated by taking the absolute values of (d) and (e).

Eqn. 1.

$$I(x,y) = \sum_{m=0}^{M-1}\sum_{N=0}^{N-1} a_{mn} e^{j2\pi(mf_1x+nf_2y+\phi_{mn})}$$
$$= \sum_{m=0}^{M-1}\sum_{N=0}^{N-1} a_{mn}\cos(2\pi(mf_1x+nf_2y+\phi_{mn})) + j\sum_{m=0}^{M-1}\sum_{N=0}^{N-1} a_{mn}\sin(2\pi(mf_1x+nf_2y+\phi_{mn})) \quad (1)$$

where $f_1 = \frac{1}{M}$ and $f_2 = \frac{1}{N}$ are the fundamental horizontal and vertical frequencies, respectively. The factors a_{mn} and ϕ_{mn} are the amplitudes and the phase shifts of the spatial frequencies.

By taking square windows of different sizes (from 21×21mm to 31×31mm) from many facial range images and examining their spatial spectrum using DFT, only certain frequency bands have high energy (See Fig. 1.c). These frequency bands are the frequencies which are low in both directions (LB), the ones which are horizontally low and vertically high (LH), the ones which are horizontally high and vertically low (HL), the ones which are high both vertically and horizontally (HB), the stripe of frequencies that are horizontally low (HS), the stripe of frequencies that are vertically high (VS). We are interested in generating a blob-like image from a band of adjacent frequencies (a window 3×3 frequencies) that have sufficient energies and are less affected by rigid transformations. The HB band and the frequency subset of the LB band which are highest horizontally and vertically seem more appealing as it can be shown empirically that the frequencies which are lowest either horizontally or vertically are generally more affected by the orientation of a given surface (pitch and yaw rotations).

3.1 Kernel Design

The selected 3×3 window of the adjacent spatial frequencies $\mathscr{F}$ produces a corresponding signal R that has real and imaginary spatial beats according to Eqn. 2.

$$R(x,y) = \sum_{(a_i,\phi_i)\in\mathscr{F}} a_i\cos(2\pi(m_if_1x+n_if_2y+\phi_i)) + j\sum_{(a_i,\phi_i)\in\mathscr{F}} a_i\sin(2\pi(m_if_1x+n_if_2y+\phi_i)) \quad (2)$$

The blob-like image can be extracted from R by simply taking the absolute value of R (see Fig. 1.f). An equivalent but more efficient way to achieve that is to convolve the range image with a corresponding kernel r. The kernel r was computed by setting the selected frequencies set $\mathscr{F}$ to 1 and all the other frequencies to 0 then taking the inverse DFT. In a similar way, the blob-like image can be computed by taking the absolute value of the convolution of the range image and r.

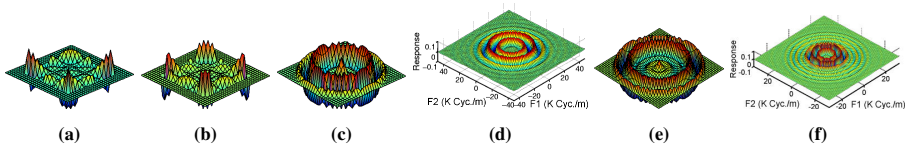


Figure 2: (a) and (b) are the real and imaginary components of the kernel that passes the selected adjacent frequencies. Their corners have high weights which affect their invariance to rotation around z axis. (c) and (e) are rotationally invariant kernels computed from the real kernel (a) by rotating it in small steps from angle 0° to 360°. The interpolation of all the rotations is averged and the DC component is removed. (d) and (f) are the frequency responses of the kernels (c) and (e), respectively

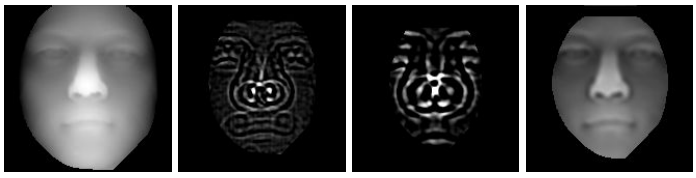


Figure 3: (a) is a facial range image. (b) and (c) are blob-like image which are computed by convolving the range image (a) with the kernels shown in Fig. 2.c and Fig. 2.e, respectively. The interest-points are extracted from the peaks in (b) and (c) which are easily detectable compared to the LoG in (e).

Unfortunately, the kernel has revealed to be sensitive to rotations around z axis (roll rotation). Although, some peaks in the blob-like image are repeatable some other points can go wrong. To circumvent this problem, we take the real part of r and rotate it around the central point from 0° to 360° in steps of 1°. At each step the kernel was bi-linearly interpolated at the integral values of x and y indices. The 360° interpolated kernels were then averaged and the DC value was subtracted. The resulting kernel h has complete invariance to roll rotations and can produce suitable blob-like patterns (see Fig. 2 and Fig. 3). The frequency response of h shows that it alternatively passes and suppresses consecutive rings of frequencies (their phase also alternates from 0° to 180°) that overlap with the initially selected frequencies (see Fig. 2.d and 2.f). Note that the resulting filter differs from LoG filters which are widely used for 2D interest-point detection. It has more oscillations and less weight at the center. Our filters can generate patterns with much more prominent peaks (see Fig. 3).

Four kernels of different sizes h_1 (21mm$\times$21mm), h_2 (19mm$\times$19mm), h_3 (31mm$\times$31mm) and h_4 (29mm$\times$29mm) were computed to produce two independent representing blob-like images (one image from h_1 and validated by h_2 and the other one from h_3 and validated by h_4). The $\mathscr{F}$ frequencies which were used to extract the kernels are the highest 3×3 adjacent frequencies in the LB band (the band of frequencies which are low horizontally and vertically) namely, the $3rd$, $4th$ and $5th$ frequencies in both horizontal and vertical directions. These frequencies are formally described in Eqn. 3

$$\mathscr{F} = \{(mf_1, nf_2) \in \mathscr{F} \mid 3 \leq m \leq 5 \text{ and } 3 \leq n \leq 5\} \tag{3}$$

There are no sufficient frequencies with high energy in the LB band for another non-

overlapping 3×3 window to produce another independent representing image using a kernel of the same size. The aforementioned kernel sizes of h_1 and h_3 are chosen so that using the same frequency window, their frequency responses are non-overlapping. The frequencies of the first kernel h_1 range from $\frac{3}{21}$ to $\frac{5}{21}$ K cycle/m (horizontally and vertically) but in the case of h_3 the range is $\frac{3}{31}$ to $\frac{5}{31}$ K cycle/m.

The kernel h_2 is very close in size and frequency response to h_1 but h_3 is close to h_4. The two representing images are computed using h_1 and h_3 and the other two kernels are used to increase the reliability of the detected interest-points. The peaks of the patterns in the first and the second representing images are validated using the patterns generated by h_2 and h_4, respectively. If the peaks of the patterns are still detectable at the same locations or within a small distance from their actual locations in the representing images, they are deemed reliable and considered as interest-points (their locations are less sensitive to minor changes in the spatial frequencies). Note that kernels of different sizes (scales) are designed to produce different patterns. Hence, the approach of maxima across a large-range of scales which is used in the SIFT [6] for scale-invariance (not required for 3D) is not applicable here.

3.2 Steps of the Proposed Approach

The approach combines interest-points from the two representing blob-like images as follows

1. The range image is convolved with each one of the four kernels h_1, h_2, h_3 and h_4 to produce the representing and validating images R_1, V_1, R_2 and V_2, respectively.

2. The peaks of the patterns in the four blob-like images are found by detecting the pixels which are the largest in their 7×7mm neighborhoods.

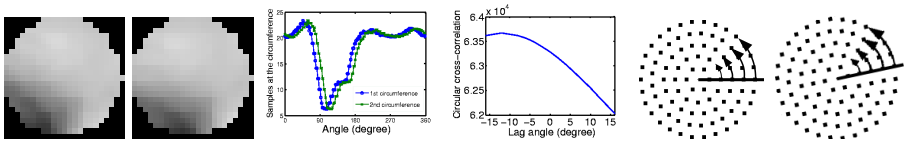
3. Each peak is assigned a strength measure s as in Eqn. 4. The peaks with low s are dropped.
$$s = R(x,y) \prod_{b(u,v)\in\mathscr{B}} R(x,y) - R(x+u,y+v) \tag{4}$$
where $\mathscr{B}$ is the set of border pixels of the 7×7 local neighborhood which have u and v offsets from the x and y location of the peak.

4. The set of peaks in R_1 that have corresponding peaks in V_1 within a distance of 2mm are called the first set of interest-points $\mathscr{I}_1$ and combined with the second interest-point set $\mathscr{I}_2$ which is extracted in a similar way from R_2 and validated by V_2 to form the total set of interest-points, $\mathscr{I}_t = \mathscr{I}_1 \cup \mathscr{I}_2$.

5. The non-maximum suppression (NMS) technique [10] is performed on the total set $\mathscr{I}_t$ to produce a filtered set $\mathscr{I}_f$. The points with maximal strength s suppress the inferior interest-points within a certain radius. See Section 5.1 for interest-points repeatability tests.

4 Face Recognition

For matching two facial range images, the interest-points in both images are detected as described in Section 3. Then, the local regions around the interest-points are matched against each other (Section 4.1). From the local region matches a similarity measure S between the matched facial images is computed.



(a) (b) (c) (d) (e) (f)

Figure 4: (a) and (b) are two local regions with a roll rotation between them. The curves in (c) are depth samples at their circumferences from which the roll angle is estimated using cross-correlation (d). The first region is sampled on concentric circles and the samples are vectorized staring from the angle 0 (the samples on the circles are appended in a depth vector starting from the largest circle to the smallest one) while the second local region is sampled and vectorized in accordance to the roll angle to achieve correspondence between the two sample vectors.

The shape of a human face deforms with expression, age and many other causes. Given such deformations, we may need to accept weak local region matches. The best match approach to the local regions produces some false matches resulting in a suboptimal total similarity measure. Instead of the best match approach to matching local regions, the local regions are putatively matched as described in Section 4.2.

4.1 Feature Extraction

The local region of each interest-point is sub-sampled on concentric circles as shown in Fig. 4.e and 4.f. This circular sampling facilitates the association between the depth values of the samples in the matched local regions. Under roll rotations we can achieve one-to-one correspondences between the depth values simply by vectorizing the sampling circles staring from an angle θ_s that equals the roll angle. It is worth mentioning that this circular sampling in the range image (only x and y coordinates are considered) is similar to finding the intersection between a sphere and a 3D surface [5, 2] (z coordinate is also considered). However, finding the relative positions of the samples to the interest-point position in case of circular sampling is more efficient computationally and can be computed offline in a look-up table while in the other case the intersection has to be computed online. On the other hand, the circular depth samples vary with pitch and yaw rotations. For small local regions (the sampling circle with largest radius is 15mm) and small pitch and yaw rotation angles the change is insignificant.

The vectorization starting angle θ_s provides invariance to roll rotations. To account for pitch and yaw rotations (surface orientations) of the matched local regions, the sampled depth values from the two local regions are linearly fitted to each other using an efficient least squares fitting technique. Then, the sum of absolute errors between the fitted depth values is used as a matching measure. The linear fitting gives invariance to pitch and yaw rotations. In addition, it mitigates the affect of difference in surface orientations on the circular sampling and minor errors in the locations of the detected interest-points. However, it does not have the flexibility to over-fit the depth values of the local regions as over-fitting dampens the error between dissimilar local regions leading to false matches.

The sum of absolute errors is computed from the local regions of two interest-points as follows:

1. The circumference of the first local region is sampled in steps of 1° and unfolded in

the counter clock-wise direction starting from angle zero ($\theta_s = 0$) into a sequence of depth values called p. In a similar manner, the circumference of the second local region is sampled and unfolded into q.

2. The roll rotation angle γ between the two local regions is found using the circular cross-correlation C between p and q as given in Eqn. 5

$$C(m) = \sum_{n=0}^{N-1} p_s q_n \quad \text{where} \quad s = \begin{cases} n+m & \text{for} \quad 0 \leq n+m < N \\ N+n+m & \text{for} \quad n+m < 0 \\ n+m-N & \text{for} \quad n+m \geq N \end{cases} \tag{5}$$

There is no zero padding to the front or the end of the sequence p as in ordinary cross-correlation. Instead, the elements of the sequence are shifted from one end and inserted into the other end. The correlation sequence C is only computed over the range of lags m from -8 to 8 sampling angle steps. The roll angle is computed from the lag that yields maximum cross-correlation m_{max} by multiplying by minus the sampling angle step, $\gamma = -m_{max}$ (giving invariance within $\pm 8^\circ$, see Fig. 4.c and 4.d for an illustration example).

3. The first local region is sub-sampled on five concentric circles (see Fig. 4.e and 4.f). The radii of the circles are 15, 12, 9, 6, and 3 mm and the number of samples on the circles are 30, 25, 18, 12 and 6 respectively. The placement of the samples is based on the a starting angle $\theta_s = 0$. Each circle of samples is vectorized starting from the angle γ and appended to a vector $\mathbf{z_1}$, starting from the largest to the smallest circle. Similarly, the second local region is sub-sampled and vectorized into $\mathbf{z_2}$ but with a starting angle $\theta_s = \gamma$.

4. The depth vectors $\mathbf{z}_1$ and $\mathbf{z}_2$ are linearly fitted to each other. First, the depth at the center of the first local region (interest-point) z_{c1} is subtracted from $\mathbf{z}_2$, $\mathbf{z}'_2 = \mathbf{z}_2 - z_{c1}$. Then we adjust $\mathbf{z}'_2$ by adding a plane so that it fits $\mathbf{z}_1$ as in Eqn. 6.

$$\mathbf{z}''_2 = \mathbf{z}'_2 + \mathbf{L}_2 [u\, v]^\top \tag{6}$$

where $\mathbf{L}_2$ is a matrix of two columns: the first one is the vectorization of the x coordinates of the samples and the second one is the vectorization of their y coordinates. The u and v parameters that define the adjusting plane is computed by the following Eqn. 7.

$$[u\, v]^\top = (\mathbf{L}_2^\top \mathbf{L}_2)^{-1} \mathbf{L}_2^\top (\mathbf{z}_1 - \mathbf{z}'_2) \tag{7}$$

5. Finally, the sum of the absolute errors e between $\mathbf{z}_1$ and $\mathbf{z}''_2$ is computed (Eqn. 8).

$$e = \sum_{n=1}^{n=N} |\mathbf{z}_1(n) - \mathbf{z}''_2(n)| \tag{8}$$

4.2 Matching a probe to Face Gallery

During the offline phase, the interest-points are detected in the gallery faces as described in Section 3.2. Then their local regions are sub-sampled and vectorized (Section 4.1). We also, compute a difficulty factor σ_i for every interest-point in the gallery face. These difficulty factors are used in weighting the contributions of the local region matches. In the situation when the local regions are flat, the local regions can strongly but falsely

match. Even if the match is correct, the information in a flat surface is low and may not be proportional to the matching measure. While in some other situations the local regions may be complex and rich in information but may have smaller matching measure. The difficulty factor is heuristically defined for an interest-point based on the sum of absolute errors of fitting the local region to a planar region (in other words, the local region is matched to a planar region as described in Section 4.1). The difficulty factor ($\sigma = \log(e) - 2$) is expected to be high for complex local regions. Then the interest-points with difficulty factors less than a threshold $\sigma_t = 250$ are dropped.

The local regions with difficulty factors $\sigma_i > \sigma_t$ in the gallery faces are putatively matched to the local regions in the probe face. Firstly, we randomly select a subset from them. Then this subset is matched against all the local regions in the probe face using the best match approach. After that, we take the matches with fitting error e less than a strong threshold $t_s = 85$ (more likely to be correct matches) and use the RANSAC algorithm [9] to find the rotation $\mathbf{R}$ and the translation $\mathbf{t}$ between the probe and the gallery faces. In case the number of the strong matches are not sufficiently large (in our case we used a threshold of seven strong matches), we include matches with medium strength to find $\mathbf{R}$ and $\mathbf{t}$ (the matches with e less than a medium threshold $t_m = 110$). If the number of the matches is still not enough (less than seven) or RANSAC has failed to find $\mathbf{R}$ and $\mathbf{t}$ that fit the matching points, the two faces are considered dissimilar and a value of zero is assigned to the similarity measure.

The rotation $\mathbf{R}$ and the translation $\mathbf{t}$ that relate the gallery face to the probe face are used to restrict the search scope for interest-points in the probe face corresponding to the remaining ones in the gallery face. $\mathbf{R}$ and $\mathbf{t}$ are applied to x, y and z coordinates of the interest-point to give an estimate of the location of the corresponding interest-point in the probe image. The local regions around the interest-points in the proximity of that location are matched to the local region around the interest-point in the gallery image. The best match among them is accepted as a match if it is less than a weak threshold $t_w = 165$.

The RANSAC algorithm [9] can robustly fit a model (in our case, the rotation $\mathbf{R}$ and the translation $\mathbf{t}$) to data (the x, y and z coordinates of the matching interest-points) in the presence of outliers (false matches). From four randomly selected matches, $\mathbf{R}$ and $\mathbf{t}$ are computed (four is the minimum number from which $\mathbf{R}$ and $\mathbf{t}$ can be computed in a least square fashion [11]). The mean location of the four interest-points in the gallery face $\mathbf{m}_g$ is subtracted from their locations. Then their x, y and z coordinates are stored in the rows of a 3×4 matrix $\mathbf{P}_g$ and similarly the matrix $\mathbf{P}_p$ is computed from their corresponding interest-points in the probe face. A matrix $\mathbf{A}$ that transforms $\mathbf{P}_g$ to $\mathbf{P}_\mathbf{p}$ is found, $\mathbf{A} = \mathbf{P}_p\mathbf{P}_g^\top(\mathbf{P}_g\mathbf{P}_g^\top)^{-1}$. The rotation matrix $\mathbf{R}$ is the nearest orthonormal matrix to $\mathbf{A}$. The singular value decomposition of $\mathbf{A}$ is computed, $\mathbf{A} = \mathbf{USV}^\top$. The rotation matrix $\mathbf{R}$ is computed, $\mathbf{R} = \mathbf{UDV}^\top$, where the matrix $\mathbf{D}$ is diagonal and the elements on the diagonal are $\{1, 1, 1/\mathbf{det}(\mathbf{UV}^\top)\}$. Then the translation is found $\mathbf{t} = \mathbf{m}_p - \mathbf{Rm}_g$. Then RANSAC finds the number of matches that fit $\mathbf{R}$ and $\mathbf{t}$ (the matches with fitness function $f = |\mathbf{Rp}_g + \mathbf{t} - \mathbf{p}_p|$ less than a distance threshold D_t, where $\mathbf{p}_g$ and $\mathbf{p}_p$ are the locations of the interest-point in the gallery image and the matching interest-point in the probe image respectively). If the number of matches that fit $\mathbf{R}$ and $\mathbf{t}$ is high (more than a percentage threshold) they are accepted as the fitting model. Otherwise RANSAC iterates until a number of trials is exhausted and takes the best model fitting the data.

The similarity measure S is extracted from the local region matches as given in Eqn.9. The weights of the matches in S depend on their strengths and difficulty factors σ_i. $\mathscr{W}$,

$\mathscr{M}$ and $\mathscr{S}$ are the sets of weak, medium and strong matches, respectively. t_w and t_m are the weak and the medium thresholds and e_i is the fitting error.

$$S = \sum_{e_i \in \mathscr{W} \cup \mathscr{M} \cup \mathscr{S}} |(t_w - e_i)|\sigma_i + \sum_{e_i \in \mathscr{M} \cup \mathscr{S}} (t_m - e_i)^2 \sigma_i \tag{9}$$

5 Experiments and Results

5.1 Repeatability Test of Interest-points

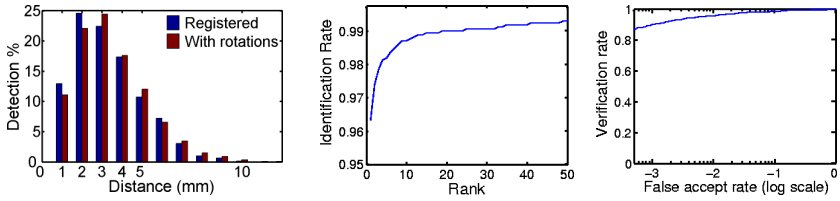
To test the repeatability and accuracy of the interest points, the approach was applied on a number of facial range images under neutral expression of many subjects. The detected interest-points that belong to the same subject were registered to each other using the ICP algorithm [8]. At a time, the ICP (3D) was applied on two sets of interest-points from a pair of range images. For each point in the first set, the 2D Euclidean distance (z is dropped as the error in z is not relevant to pixel accuracy) to the nearest point in the second set was found (distance error). A small distance indicates that the interest-point was accurately detected in the second range image. Fig. 5.a shows a histogram of error distance computed from all the range images in the test set. It shows that 60% of the interest-points were detected within a distance of 3mm and about 88% of the points are within a distance of 5mm.

The performance of the approach with respect to rotations was tested. The underling pointclouds of one of the range images of each pair were rotated using the pitch, yaw and roll angles in the range of $\pm 15^\circ$. After that, the range image was recalculated as in Section 2 (without the spike removal part). Then, the same repeatability test was performed on the range images (see Fig. 5.a).

In comparison to the performance without rotations, some degradation can be noticed in the accuracy of interest-point detection. However, generally the performance did not degrade significantly. The interest-points which were detected within an error range of 0 to less than 1mm has decreased from 13% to 12%. Also, the interest-points which were detected in the range of 1 to less than 2mm has decresed from 24.5% to 22% (see Fig. 5.a for more ranges). Staring from the range of 2 to less than 3 and onward, the interest-point detection has shows generally an increase in the detection percentage rather than a decrease which is an indication that some of the missed points in the smaller ranges were detected in the less accurate ones. The overall detection percentage was about 87.5% within 5mm. This repeatablity is comparable to typical 2D interest-point detection approaches when applied on 2D images [12, 6].

5.2 Recognition Results

The face recognition approach (Section 4) was applied to the FRGC V2.0 dataset [7], the largest publicly available dataset. The number of subjects in the test data is 466. Our approach to interest-point based face recognition was applied to the near frontal view 3D facial pointclouds with neutral expression. The range images are extracted from the pointclouds as described in Section 4. Each subject is represented by a single range image. The remaining range images are matched against the 466 gallery images. Fig. 5 shows the recognition performance of 1765 probes. The first rank recognition is 96.3% and increases to about 99% at the tenth rank (Fig. 5.b). The verification rate is about 90% at 0.1% FAR (Fig. 5.c).



(a) (b) (c)

Figure 5: (a) A histogram showing the repeatability of the detected interest-points in registered range images (no rotations) and in rotated images in the range of $\pm 15^{\circ}$. The width of the histogram bins is 1mm. (b) and (c) are the Cumulative Matching Characteristics (CMC) and the Receiver Operational Characteristics (ROC) showing face recognition performance.

6 Conclusions

An approach to interest-point detection in range images is presented and applied to facial range images. Experiments showed that the proposed technique can detect interest-points in facial range images with 88% repeatability at an accuracy of less than 5mm. We also presented novel algorithms for local feature extraction and feature matching for 3D face recognition. The proposed algorithms extract features around the interest points on a 3D face and match them for recognition under small pose variations. Experiments were performed on the FRGC v2 dataset and a rank one recognition rate of 96.3% was achieved.

References

[1] A. Mian, M. Bennamoun and R. Owens. *Region-based Matching for Robust 3D Face Recognition* . Proc. of BMVC, 2005.

[2] C. Chua and R. Jarvis. *Point signatures: A new representation for 3d object recognition*. IJCV, 1997.

[3] A. Moreno, A. Sanchez, J. Fco, V. Fco and J. Diaz . *Face recognition using 3D surface-extracted descriptors*. IMVIP, 2003.

[4] Y. Wang, C. Chua, Y. Ho and Y Ren. *Integrated 2D and 3D images for face recognition*. IEEE IAP, 2001.

[5] I. Mpiperis, S. Malasiotis, and M Strintzis. *3D Face Recognition by Point Signatures and Iso-contours*. Proc. of SPPRA , 2007.

[6] D. Lowe. *Distinctive Image Features from Scale-Invariant Keypoints* . IJCV, 2004.

[7] P. Phillips, P. Flynn, T. Scruggs, K. Bowyer, J. Chang, K. Hoffman, J. Marques, J. Min and W. Worek. *Overview of the Face Recognition Grand Challenge*. IEEE CVPR, 2005.

[8] Y. Chen and G. Medioni. *Object modeling by registration of multiple range images*, IEEE PAMI, 1991.

[9] M. Fischler and R. Bolles. *Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography*. Comm. of the ACM, 1981.

[10] D. Forsyth and J. Ponce. *Computer Vision: A Modern Approach*. Prentice-Hall, 2003.

[11] S. Umeyama. *Least-squares estimation of transformation parameters between two-point patterns*. IEEE PAMI, 1991.

[12] M. Brown, R. Szeliski and S. Winder. *Multi-Image Matching using Multi-scale Oriented Patches*. IEEE CVPR, 2005.