Facial Template Synthesis based on SIFT Features

Ajita Rattani, D. R. Kisku
Department of Computer Science and Engineering
Indian Institute of Technology Kanpur
Kanpur, India
{ajita,drkisku}@iitk.ac.in

Andrea Lagorio and Massimo Tistarelli
Computer Vision Laboratory
University of Sassari
Alghero, Italy
{lagorio,tista}@uniss.it

Abstract— **This paper proposes a procedure for facial template synthesis based on features extracted from multiple facial instances with varying pose.**

The proposed system extracts the rotation, scale and translation invariant SIFT features, also having high discrimination ability, from the frontal and half left and right profiles of an individual face images. These affine invariant features obviate the need of ad-hoc algorithms to register features of side profiles against frontal profiles for feature-set augmentation. An augmented feature set is then formed from the fusion of features from frontal and side profiles of an individual, after removing feature redundancy. The augmented feature sets of database and query images are matched using the Euclidean distance and Point pattern matching techniques. The experimental results are compared with the system using only frontal face images for both the matching strategies. The reported results prove the efficacy of the proposed system.

Keywords-Face;Fusion; SIFT; Template Synthesis

I. INTRODUCTION

In recent years, biometrics improved considerably in reliability and accuracy, with some specific traits showing very good identification and verification performance. Among all biometric traits, face recognition is the most natural physiological characteristic to recognize each other. The ability shown by humans to recognize known faces is present since birth, thus making face recognition a very interesting and challenging field.

But despite of being a research topic for more than 30 years with several classifiers developed [1-3], the threats like change in illumination, pose and facial expression have not been solved yet [4,5]. In order to cope with these limitations, information from multiple sources are concatenated in multibiometric systems [6,7] at various levels, such as sensor level, feature level, match score level and decision level. Thus using the concept of multibiometrics, efforts have been made in the literature to improve the performance of face recognition in the form of multiple classifiers [8,9], fusion of 2D and 3D datata [10] and the use of morphable face models [11].

However, of current interest is mosaicing where multiple images of an individual can be combined together at both image and feature level. At image level (image mosaicing) multiple snapshots of an individual are combined together to generate an enhanced image. This image is then subject to feature extraction and matching. On the other hand, at the feature level (feature mosaicing or template synthesis) the features, extracted from multiple images, are combined together to generate a composite feature set thus accounting for incomplete information. Both schemes depend on the accurate alignment and on the selection of a proper transformation algorithm for a pair of images (or corresponding feature sets) before the integration process. As addressed in [12] and [13], multiple biometric templates obviate the need to store multiple images of the same individual, it also obviates the need of a query image to be compared with multiple instances. The composite template also reduces the chances of high false rejection rates, while the template selection is not required [13]. The method proposed in [13] proposes an image and feature mosaicing strategy applied to fingerprints, where two different transformation algorithms are used. The Elastic Matching algorithm is applied for the detection of corresponding points within two instances and the Iterative Closest point (ICP) algorithm is used for the estimation of the final transformation matrix to register different instances of fingerprint images. The proposed system reports an increase in accuracy of more than 4% with respect to the use of a single-image template.

For face recognition, at image mosaicing level, few examples are reported in the literature. In [14] a procedure to create panoramic face mosaics by acquiring different views from five cameras is proposed. Corresponding points in multiple face views are determined explicitly by placing ten colored markers on the face. This procedure utilizes the control points to compute a series of linear transformations to generate a face mosaic. In [15] the side profile images are aligned with the frontal image using a terrain transform exploiting neighborhood properties to determine the transformation relating the two images. Multiresolution splining is then used to blend the side profiles with the frontal image, thereby generating a composite face image of the user.

However, to the best of our knowledge, no work has been reported in the literature relating to template synthesis at feature or feature mosaicing level, applied to face recognition. The analysis of the current state of the art, highlights the importance of transformation algorithms and selection of control points to accurately align multiple instances, whether at image or feature level.

1-4244-1300-1/07/$25.00 2007 IEEE

The primary focus of this paper is to use affine invariant SIFT features for facial template synthesis from frontal and side profiles of an individual face images. These features obviate the need of a transformation algorithm for registering side profiles with frontal profiles. The correspondence between features of side and frontal profiles is easily computed without requiring an additional transformation. SIFT features also provide a good discrimination ability for face recognition as already proposed in [17]. The composite templates are matched using the Euclidean distance and point pattern matching techniques.

The paper is organized as follows: Section II, briefly describes the Scale Invariant Feature Transform (SIFT features); the procedure of Facial template synthesis and matching techniques are explained in section III; experimental results are presented in section IV.

II. SCALE INVARIANT FEATURE TRANSFORM (SIFT)

The *Scale Invariant Feature Transform* (SIFT) [16] has recently emerged as a cut edge methodology in general object recognition as well as for other machine vision applications. One of the interesting features of the SIFT approach is the capability to capture the main gray level features of an object's view by means of local patterns extracted from a scale-space decomposition of the image.

These features are invariant to image scaling, translation, and rotation, and partially invariant to illumination changes and affine or 3D projections. Features are efficiently detected through a staged filtering approach that identifies stable points in scale space and are highly distinctive, in the sense that a single feature can be correctly matched with high probability against a large database of features from many images. The first filtering stage identifies key locations in scale space by looking for locations that are maxima or minima of a difference-of-Gaussian function. Each point is used to generate a feature vector that describes the local image region sampled relative to its scale-space coordinate frame. By blurring the image gradient, the features achieve partial invariance to local variations, such as affine or 3D projections. The resulting feature vectors are called SIFT keys.

Following are the major stages of computation used to generate the set of image features or keys.

A. *Scale-space extrema detection*

The first processing stage searches over all scales and image locations. It is implemented efficiently by using a difference-of-Gaussian function to identify potential interest points, invariant to scale and orientation.

B. *Keypoint localization*

At each candidate location, a detailed model is fit to determine the location and scale. Keypoints are selected based on the measurement of their stability.

C. *Orientation assignment*

One or more orientations are assigned to each keypoint location based on local image gradient directions. All subsequent operations are performed on image data that has been transformed relative to the assigned orientation, scale, and location for each feature, thereby providing invariance to these transformations.

D. *Keypoint descriptor*

The local image gradients are measured at the selected scale in the region around each keypoint. These are transformed into a representation that allows for a significant amount of local shape distortion and change in illumination [16].

The employment of SIFT for face analysis was systematically investigated in [17] where the matching was performed using three techniques: (a) minimum pair distance, (b) matching of the eyes and mouth, and (c) matching on a regular grid. However, none of these techniques considered the spatial position and orientation but they were solely dependent on the values of the keydescriptors.

The input to the system is the face image and the output is the set of extracted SIFT features $s=(s_1, s_2, \ldots \ldots, s_m)$ where each feature $s_i=(x, y, \theta, k)$ consist of x, y spatial location, θ as local orientation and k as keydescriptor of size *1x128* as shown in Fig. 1

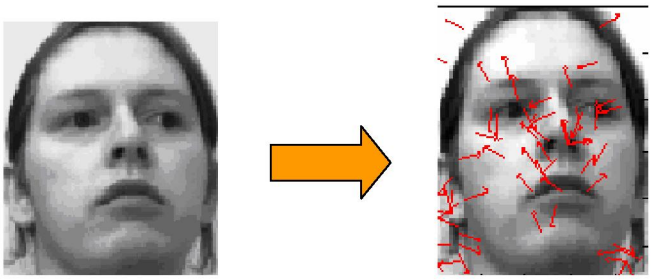

Figure 1. Example face image and the extracted SIFT features depicted as red arrows

III. FACIAL TEMPLATE SYNTHESIS

The synthesized face template is created by extracting the SIFT features from frontal, left and right profile images of an individual. The extracted feature set is concatenated to form an augmented feature set containing the complete information of an individual derived from frontal and side views. The whole process consist of the following modules.

A. *Feature extraction*

SIFT features are extracted from the frontal, left and right side head poses. The side profiles are chosen such that they have an angle of more than 25° from the frontal face image. The SIFT features are extracted separately from frontal (F_S), left (L_S) and right profiles (R_S). The correspondence between the features is then determined using the keypoint descriptor (k). In Fig. 2, three sample views of a subject are shown.

Figure 2. Three example images from the same subject, used for facial template synthesis

B. Feature point correspondence

After the feature set has been extracted, redundant features, extracted from overlapping regions between frontal and side profiles, must be removed to handle the problem of curse of dimensionality. This requires finding the corresponding features between the frontal and side profiles and retaining the average of them. Given the affine invariance property of SIFT features, there is no need to register the side profiles with respect to the base (frontal) face image. The pointwise correspondence between features of frontal and side views is easily established by finding the difference between the keydescriptors (*k*) of the SIFT features in both the views. The corresponding points are identified by keydescriptors with minimum euclidean distance within the two view. The correspondence is calculated in a pairwise manner between the frontal and two side profiles. Keeping the average of the features, i.e., spatial locations, orientation and keypoint descriptor (*x, y, θ, k*) of the corresponding points between different views in the composite template, redundancy is removed. The correspondence between two views of an individual is shown in Fig. 3. The six red lines connect the corresponding points between the two views calculated in a pairwise manner.

Figure 3. The point correspondence is detected using keypoint descriptor as shown with red lines within two instances of an individual

C. Feature set concatenation

At the last step of the process all redundant features, i.e. the corresponding points between the frontal (F_S) and side profiles images (L_S) (R_S), were identified and averaged in a pairwise manner. All non-corresponding feature points, extracted from different views, and the averaged corresponding points are now put together (*concat*) to form the composite feature set encompassing the complete information as in (1):

$$concat = (F_S) \cup (L_S) \cup (R_S) - \Sigma((F_S) \cap (L_S) \cap (R_S)) \quad (1)$$

Two types of templates are formed in this experiment. The former is based only on the keypoint descriptors which are retained from the extracted SIFT features of all the instances. The latter is based on the spatial location, orientation and the keypoint descriptors, in which all the information pertaining to the SIFT features are retained. Different classifiers are applied accordingly.

D. Feature Matching

Once the feature reduction strategy is applied and the feature sets are concatenated together, the concatenated feature sets (*concat and concat'*) from the database and the query images are processed to compute the proximity between the two pointsets. In this study two different matching techniques are applied.

1) Euclidean distance: This measure is used for templates containing only keypoint descriptors extracted and augmented from all the instances. The match score is computed on the basis of the number of keypoint descriptors matched between the database and the query image. A matching is established between two keypoint descriptors if their Euclidean distance lies within some predetermined threshold as in (2):

$$kd(concat'_j, concat_i) = \sqrt{\sum_i (k^i_j - k^{i'}_j)^2} \le k_0 \quad (2)$$

where k^i_j is the ith element of keydesriptor *j* within a composite template.

2) Point pattern matching: This technique is used to match the templates containing the spatial location, orientation and the keypoint descriptor. The aim of this method is to find the number of points "paired" between the concatenated feature sets from the database and the query images. Two points are considered paired only if the spatial distance (S_d), the direction distance (D_d) and the Euclidean distance (k_d) between the corresponding feature points are all within some threshold as in (3) (4)and (5) [18]:

$$S_d(concat'_j, concat_i) = \sqrt{(x'_j - x_i)^2 + (y'_j - y_t)^2} \le r_0 \quad (3)$$

$$D_d(concat'_j, concat_t) = \min(|\theta'_j - \theta_t|, 360^\circ - |\theta'_j - \theta_t|) \le \theta_o \quad (4)$$

$$kd(concat'_j, concat_i) = \sqrt{\sum_i (k^i_j - k^{i'}_j)^2} \le k_0 \quad (5)$$

where the points *i* and *j* are represented by (*x, y, θ,* $\boldsymbol{k}$) with $\boldsymbol{k} = k^1, k^2, ..., k^{128}$ of the concatenated database and the query pointsets *concat'* and *concat.*

The final matching score for both the Euclidian distance and the Point pattern matching technique is based on the number of matched pairs found in the two concatenated sets, and computed from (6):

$$MS = \frac{100 * MPQ^2}{M * N} \quad (6)$$

MPQ is the number of paired points between the database and the query concatenated pointsets, while M and N are the number of points in the concatenated feature sets of the database and the query images.

IV. EXPERIMENTAL RESULTS

The proposed approach has been tested on the UMIST face database [19], which consists of 564 images of 20 people. A range of poses from profile to frontal views, are covered for each subject, as shown in Fig. 4. The subjects in the database cover a wide range of mixed races and genders with different appearances. The matching results are computed and compared with the system trained and tested on only frontal snapshots for both the matching techniques. The False Aceptance Rate (FAR), False Rejection Rate (FRR) and Accuracies are duly recorded.

The following protocol has been established for training and testing the system:

Training: one image per person is used for enrollment based on frontal images only; one frontal image and two side profiles are used for the template synthesis procedure.

Testing: five frontal samples per person are used for testing and generating client scores for system based on frontal image only. Impostor scores are generated by testing the client against the five samples of the rest of the eleven individuals. In case of template synthesis procedure, the client is tested against the five genuine composite template containing features from frontal, left and right profiles on an individual. The client is subjected to five imposter attacks of rest of eleven candidates. Thus in total 12x5=60 client scores and 12*11*5=660 imposters scores for each of the systems are generated and evaluated. A snapshots from the database are shown in Fig. 4.

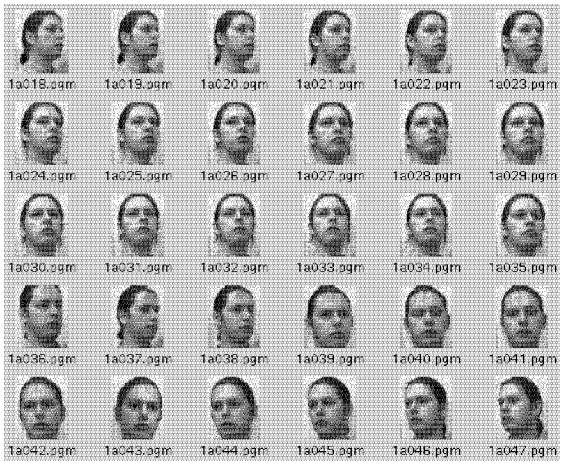

Figure 4. Images recorded for one subject in the UMIST database. The set covers a wide range of views from profile to frontal.

Table 1 shows the performance obtained by the matching (Euclidean distance) applied to template mosaicing against the frontal images. The representation in based only on the values of the keydescriptor extracted from the SIFT features.

Table 2 shows the performance obtained by using the template mosaicing against the use of frontal images only for the matching scheme (Point pattern matching). In this case the spatial location, orientation and keydescriptor information of the SIFT features was used.

TABLE 1. FFR, FAR AND ACCURACY VALUES

ALGORITHM	FRR(%)	FAR(%)	Accuracy
Face SIFT	5.38	10.97	91.82
Feature Synthesis	3.66	6.78	94.77

TABLE 2. FFR, FAR AND ACCURACY VALUES FOR THE TWO FACE REPRESENTATIONS. THE MATCHING IS BASED ON THE COMPLETE INFORMATION ASSOCIATED TO THE SIFT FEATURES

ALGORITHM	FRR(%)	FAR(%)	Accuracy
Face SIFT	5.0	8.98	92.94
Feature Synthesis	2.24	5.85	95.95

Fig. 5. shows ROC Curve for the two methods, using only the frontal image and the frontal and side profiles. The "Face SIFT (k)" and the "Template synthesis (k)" curves are obtained from the Euclidean distance classifier applied to the values of the keypoint descriptors alone. The "Face SIFT" and the "Template synthesis" curves are obtained from the point pattern matching technique applied to all the data obtained from the SIFT Features.

It is evident from both the tables and the ROC curve, that the matching based on the feature synthesis outperforms the system based on frontal images. Consequently, the features extracted from multiple instances, captured from different views, provide complementary information of an individual. The enhanced information content reduces the chances of both high FAR and FRR and also combats the threat to face recognition system due to variation in poses.

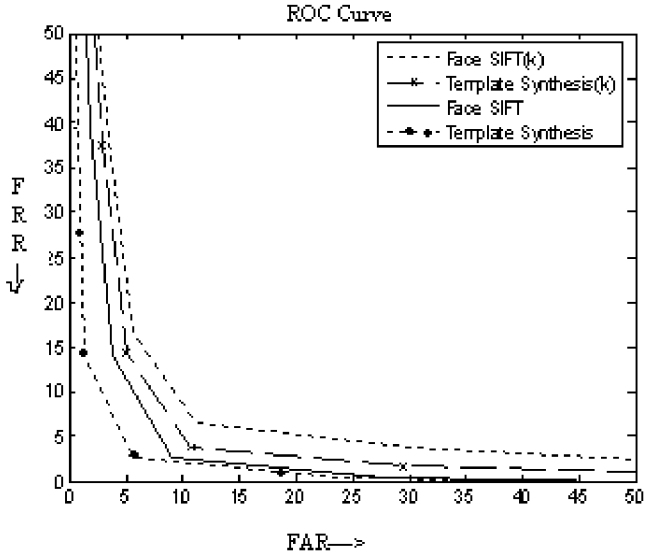

Figure 5. shows the ROC curves for the two matching techniques using respectively frontal images and the template synthesis.

V. Conclusion and Future Work

The work proposed a facial feature synthesis scheme based on SIFT features. As reported in the literature, the existing works related to mosaicing are dependent very much on the selection of the transformation scheme to align the features belonging to multiple instances of an individual. Hence, the need to use affine invariant features, which apart from obviating the need to explicitly model the transformation, provides a better discrimination capability. In the proposed approach affine and illumination invariant SIFT features are used to synthesize a complex face template. The corresponding features between different views of an individual face images are easily detected without using any transformation matrix also reducing time complexity.

The results obtained from the tests show that multiple poses provide complementary information pertaining to an individual which drastically improves the performance of the identification system. The same process, accumulating evidence on the subject's identity obtained from more views, may be extended by using more than three instances for the template synthesis. This extension to the present work is currently under investigation together with a more comprehensive testing on a face database including more subjects. This work will further investigate whether SIFT features can be still employed as control points for image mosaicing between multiple views.

Acknowledgment

This work has been partially supported by grants from the Italian Ministry of Research, the Ministry of Foreign Affairs and the Biosecure European Network of Excellence.

References

[1] W. Zhao, R. Chellappa, A. Rosenfeld, and P.J. Phillips, "Face Recognition: A Literature Survey", ACM Computing Surveys, vol 35 (4), pp. 399 – 458, December 2003.

[2] M. Turk and A. Pentland, "Eigenfaces for face recognition", Journal of Cognitive Neuroscience, vol. 3 (1), pp. 71-86, 1991.

[3] L. Wiskott, J.M. Fellous, N. Krüger, and C. von der Malsburg, "Face recognition by elastic bunch graph matching", IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 19 (7), pp. 775-779, 1997.

[4] P.J. Phillips, P. Grother, R.J Micheals, D.M. Blackburn, E. Tabassi, and J.M. Bone, "FRVT 2002: Evaluation Report", 2003.

[5] P. Jonathon Phillips, Patrick J. Flynn, Todd Scruggs, Kevin W. Bowyer, Jin Chang, Kevin Hoffman, Joe Marques, Jaesik Min, and William Worek, "Overview of the face recognition grand challenge", Computer Vision and Pattern Recognition (CVPR 2005), San Diego, pp. 947-954, June 2005.

[6] A. K. Jain and A. Ross, "Multibiometric systems", Communications of the ACM, vol. 47 (1), pp. 34-40, 2004,.

[7] A. Ross and A.K. Jain, "Information fusion in biometrics", Pattern Recognition Letters, vol. 24, pp. 2115- 2125, 2003.

[8] F. Roli and J. Kittler Eds, "Multiple classifier systems", Springer Verlag, LNCS 2364, 2002.

[9] X. Lu, Y. Wang, and A.K. Jain, "Combining classifiers for face recognition", Proceedings of IEEE International Conference on Multimedia & Expo, pp. 13-16, 2003.

[10] K.I. Chang, K.W. Bowyer, and P.J. Flynn, "An evaluation of multi-modal 2D+3D face biometrics", IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 27(4), pp. 619-624, 2005,.

[11] V. Blanz, S. Romdhami, and T. Vetter, "Face identification across different poses and illuminations with a 3D morphable model", Proceedings of International Conference on Automatic Face and Gesture Recognition, pp. 202-207, May 2002.

[12] A. Jain and A. Ross, "Fingerprint mosaicking", Proceedings of ICASSP, vol. 4, pp. IV-4064 – IV-4067, 2002.

[13] A. Ross, S. Shah and J. Shah, "Image versus feature mosaicing: A case study in fingerprint", Proceedings. of SPIE Conference on Biometric Technology for Human Identification III, Orlando, USA, pp. 620208-1 - 620208-12, 2006.

[14] F. Yang, M. Paindavoine, H. Abdi, and A. Monopoly, "Development of a fast panoramic face mosaicing and recognition system", Optical Engineering, vol. 44, 2005.

[15] R Singh, M. Vasta, A. Ross and A. Noore, " Performance enhancement of 2D face recognition via mosaicing", Proceedings. Of fourth IEEE workshop on automatic identification advanced technologies, buffalo, USA, pp. 63-68, 2005.

[16] Lowe and G. David, "Object recognition from local scale invariant features," International Conference on Computer Vision, Corfu, Greece, pp. 1150–1157, September 1999.

[17] M. Bicego, A. Lagorio, E. Grosso and M. Tistarelli,"On the use of SIFT features for face authentication", Proceedings. of Int Workshop on Biometrics, in association with CVPR 2006.

[18] D.M. Mount, N.S. Netanyahu and J. Le Moigne, "Efficient algorithms for robust point pattern matching", Pattern Recognition, vol. 32, pp. 17-38, 1999.

[19] http://images.ee.umist.ac.uk/danny/database.html