

Research Article

View Influence Analysis and Optimization for Multiview Face Recognition

Won-Sook Lee¹ and Kyung-Ah Sohn²

¹ School of Information Technology and Engineering, University of Ottawa, Ottawa, Canada K1N6N5

² Computer Science Department, Carnegie Mellon University, Pittsburgh, PA 15213-3891, USA

Received 1 May 2006; Revised 20 December 2006; Accepted 24 June 2007

Recommended by Christophe Garcia

We present a novel method to recognize a multiview face (i.e., to recognize a face under different views) through optimization of multiple single-view face recognitions. Many current face descriptors show quite satisfactory results to recognize identity of people with given limited view (especially for the frontal view), but the full view of the human head has not yet been recognizable with commercially acceptable accuracy. As there are various single-view recognition techniques already developed for very high success rate, for instance, MPEG-7 advanced face recognizer, we propose a new paradigm to facilitate multiview face recognition, not through a multiview face recognizer, but through multiple single-view recognizers. To retrieve faces in any view from a registered descriptor, we need to give corresponding view information to the descriptor. As the descriptor needs to provide any requested view in 3D space, we refer to it as “3D” information that it needs to contain. Our analysis in various angled views checks the extent of each view influence and it provides a way to recognize a face through optimized integration of single view descriptors covering the view plane of horizontal rotation from -90° to 90° and vertical rotation from -30° to 30° . The resulting face descriptor based on multiple representative views, which is of compact size, shows reasonable face recognition performance on any view. Hence, our face descriptor contains quite enough 3D information of a person's face to help for recognition and eventually for search, retrieval, and browsing of photographs, videos, and 3D-facial model databases.

Copyright © 2007 W.-S. Lee and K.-A. Sohn. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

1. INTRODUCTION

Face recognition techniques have started to be used as commercial products in the last few years, especially on the frontal images, but with certain constraints such as indoor environment, controlled illumination, and small degree of facial expression as can be seen in many literatures, for example, in a classic survey paper by Samal and Iyengar [1]. Face recognition is composed of two main steps, registration and retrieval. We register a person's face in a certain form, and we retrieve the person's face out of many people's faces. One problem we want to raise in this paper is what is the optimized way to determine how many views and which angle we need to register the person to retrieve the person in any angle. As an effort to make more practical systems, various researches have been performed to detect and recognize faces in arbitrary poses or views. However, those approaches using statistical learning methods [2–4] reveal limitation to satisfy practically acceptable recognition performance. Novel view generation using 3D morphable model approach [5] shows

quite reasonable success rate in many different views, but it still depends on the database of 3D generic models to build the linear interpolation of a given person and also it needs high computational costs with very complicated algorithms behind. Recently, 3D face model from direct 3D scanning could be used for face recognition [6–8], but the successful reconstruction is not always guaranteed in real time and the recognition rate is not yet as good as the 2D-image-based face recognition. In addition, the acquisition of the data is not always as easy as images and we still need more robust and stable sensing equipment to get meaningful recognition applications. In short, multiview face recognition has still a lot lower recognition rate compared to single view recognition.

As a representative method of the currently available 2D-based face descriptor, MPEG-7 advanced face recognizer [9, 10] shows quite satisfactory results to recognize identity of people with given single view, and it especially shows good performance on the frontal view. However, the single-view-based face descriptor, as it allows only one view to build its

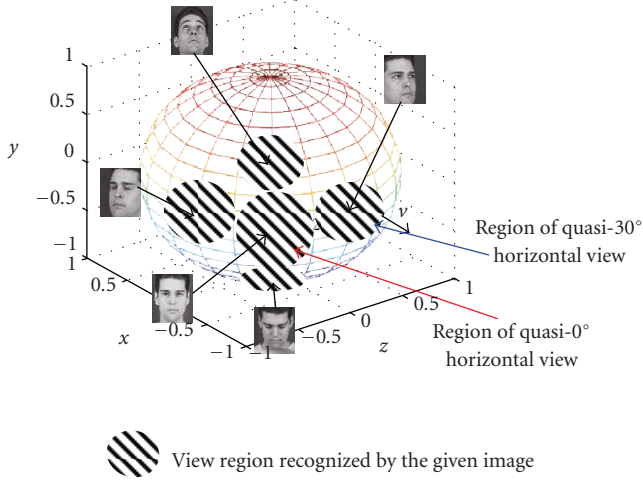


FIGURE 1: Single view recognition of the view-sphere surface.

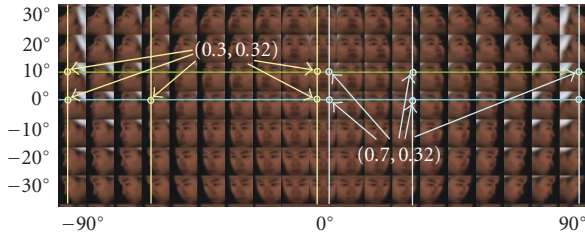


FIGURE 2: Eye positions on view mosaic of faces from 108 rendered images of 3D facial mesh models. Left eye position keeps $(0.7, 0.32)$ for positive horizontal rotation while right eye position does $(0.3, 0.32)$ for negative horizontal rotation when width and height of the image are considered as 1.0.

descriptor, causes problems to recognize other views. Nevertheless, it still allows nearby frontal views recognizable with desirable success rate.

In this paper, we present a novel face descriptor based on multiple single-view recognition, which aims to contain multiview 3D information of a person to help for face recognition in any view. In this scenario, we save or register the face descriptor as unique information of each person, and when we have a query face image in arbitrary view, we can identify the person's identity by comparing the registered descriptors with the one extracted from the query image. To retrieve such 3D information of a face to be recognizable in any view, we propose a method to extend the traditional 2D-image-based face recognition to 3D by combining multiple single views. We take a systematic extension to build 3D information using multiviews and perform optimization of the descriptor in respect of the number and the choice of views to be registered. In the following sections, we first describe the concept of multiview 3D face descriptor, and then show how to optimize multiple single views to build 3D information using our newly proposed “quasiview” concept, an extended term of quasifrontal, which measures the influence power of a certain view to nearby views. Experimental results then follow.

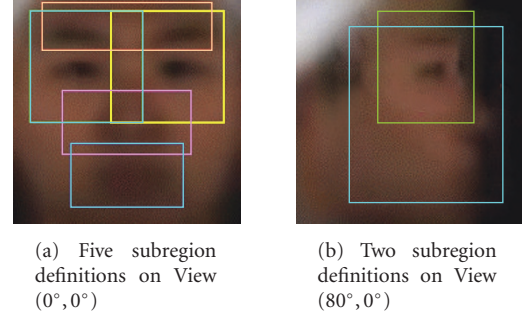


FIGURE 3: Subregion definition depending on views superimposed on the center face of our database.

2. MULTIVIEW 3D FACE DESCRIPTOR

The new descriptor we propose is called multiview 3D face descriptor, which is supposed to have sufficient 3D information of a face by describing the face as a mosaic of many one-views as shown in Figure 1. This multiview 3D face descriptor aims to cover any view between horizontal rotation from -90° to 90° and vertical rotation from -30° to 30° . We note the range of such horizontal and vertical views as $[-90^\circ \dots 90^\circ]$ and $[-30^\circ \dots 30^\circ]$, respectively. The notation of $[\cdot]$ is used to refer the range while (\cdot) used for a position.

There are a few issues we encounter for the extension of the conventional single view descriptor to multiview version.

- (i) *DB collection for training/test*: there are not yet enough data to be used for research on multiview face recognition. Most face database such as PIE, CMU, and YALE has been built mainly for frontal views even though nonfrontal face images are more usual.
- (ii) *Multiview face detector*: to recognize a person from face images, we first need to detect faces on photographs, which is a rough alignment process.
- (iii) *View estimator*: the view of the facial images should be estimated.
- (iv) *Face alignment*: faces are then aligned in predefined location.
- (v) *Feature extraction*: we extract features possibly depending on views.
- (vi) *Descriptor optimization*: we intend to produce efficient descriptor containing views on horizontal rotation $[-90^\circ \dots 90^\circ]$ and vertical rotation $[-30^\circ \dots 30^\circ]$.

For DB generation, we could use 3D facial mesh models and render them to get face images in arbitrary views. For the experiment, the 3D facial mesh models of 108 subjects are used and their rendered images are used for training and test with 50/50 ratio. The database we use for the experiment is described in our previous work [11, 12] as well as pose estimation and feature detection. In this paper, we focus on the last two issues of feature extraction and descriptor optimization considering the various studies about multiview face detections and view estimations. The most naïve idea to create multiview descriptor from a single-view one is the

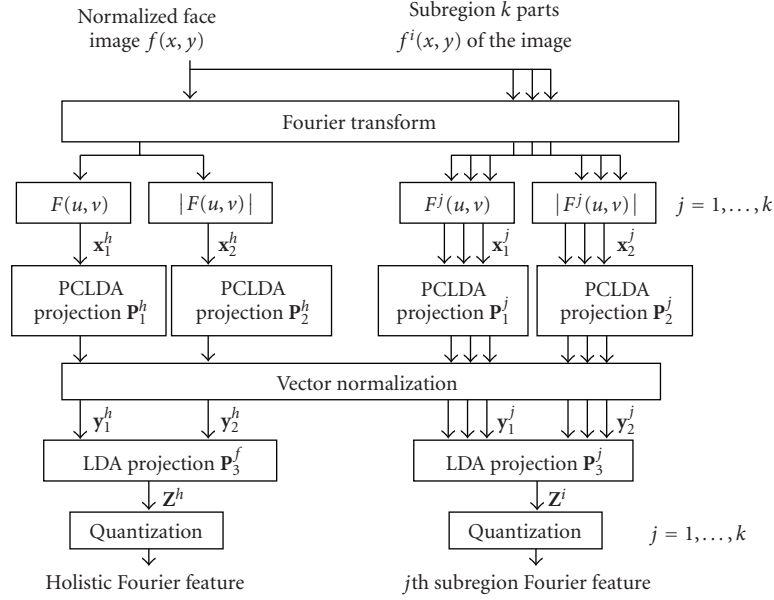


FIGURE 4: Feature extraction used for multiview 3D face descriptor.

simple integration of N uniformly distributed single view descriptors. If we register every 10° apart, that is, if we use every face image 10° apart for our descriptor, we have to register 19×7 views to cover the view space of horizontal rotation $[-90^\circ \dots 90^\circ]$ and vertical rotation $[-30^\circ \dots 30^\circ]$. Then this very naïve descriptor would result in size $133 \times$ single view descriptor size, which becomes too big to be used in practice. Moreover, we could take advantage of the possibility that some view regions might have larger coverage than others so that we may need smaller number of views to describe those regions. While the descriptor optimization is one of the important steps for transition from single view to multiview face descriptor, there has not been until now any published result in this direction to the best of our knowledge. Here, we aim to make use of our learning from frontal-view face descriptors that a registered front view can be used to retrieve nearby frontal views (quasifrontal) with high success rate. Hence, we extend the concept of quasifrontal to quasiview and introduce some useful terms as follows.

- (1) *View mosaic*. Mosaic of views 10° apart covering horizontal rotation $[-X^\circ \dots X^\circ]$ and vertical rotation $[-Y^\circ \dots Y^\circ]$. Here we choose $X = 90$ and $Y = 30$. It can be visualized as shown in Figure 2. This view mosaic is corresponding to any view (i.e., 3D) of a person wherever the face is at least half. It is used later on to check “quasiview” for each view in the view mosaic.
- (2) *Quasiview with error rate K* . It is an extension of quasifrontal, from the frontal view to general views. For instance, quasiview V^q of a given (registered) view V with error rate K means that faces on view V^q can be retrieved using a registered face in view V with expected error rate less than or equal to K . This will be explored in Section 5

3. LOCALIZATION OF FACES IN MULTIVIEW

To use face images for training or as a query, we need to extract and normalize facial region. According to common practice, positions of two eyes are used for normalization such that the normalized image contains enough information of the face but excludes unnecessary background. The detailed localization specification is defined as follows.

- (1) Size of images: 56×56 .
- (2) Positions of two eyes in the front view are on $(0.3, 0.32)$ and $(0.7, 0.32)$ when width and height are considered as 1.0. Here $(,)$ is used for (x, y) coordinates where the numbers are between 0 and 1.
- (3) Left eye position of the positive horizontal rotation keeps $(0.7, 0.32)$ while right eye position of the negative rotation does $(0.3, 0.32)$.
- (4) Vertical rotation has the same eye positions as the ones on zero vertical rotation images.

Figure 2 summarizes the view mosaic of resulting localized images for our view space of horizontal rotation $[-90^\circ \dots 90^\circ]$ and vertical rotation $[-30^\circ \dots 30^\circ]$.

4. FEATURE EXTRACTION

As an example of single view face descriptor, we use the MPEG-7 advanced face recognition descriptor (AFR) [9] which showed best performance in retrieval accuracy, speed and data size as benchmarked by MPEG-7. More details can be found in MPEG document [9]. However, our focus in this paper is to show how to build optimized integration of multiple views to recognize a face in any view based on single-view face recognizers, so any single view face recognizer can be used instead of MPEG-7 AFR.

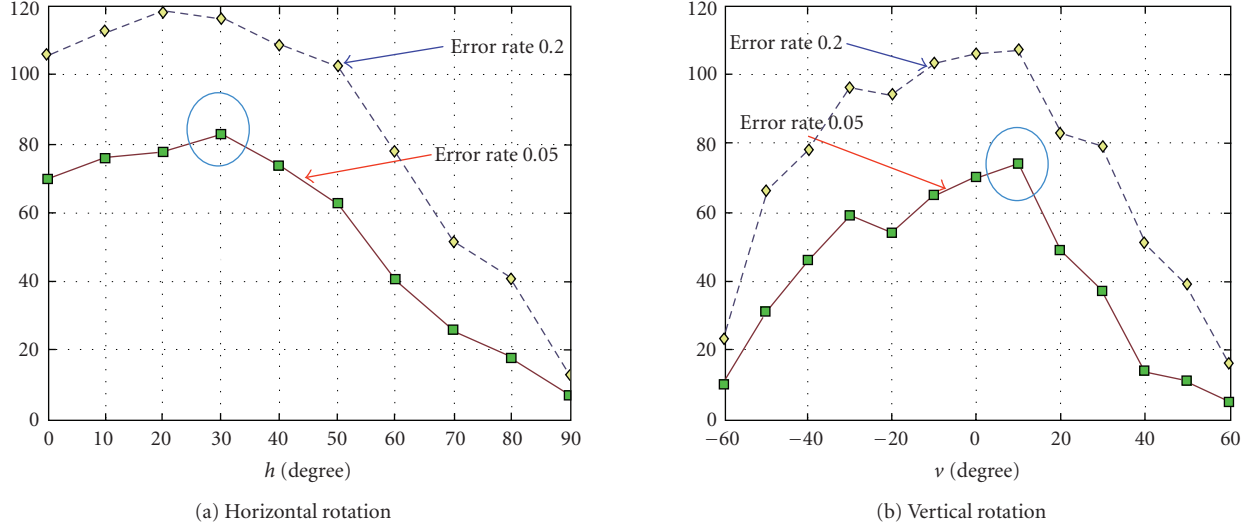


FIGURE 5: Quasiview sizes with horizontal and vertical rotations. The x-axis in (a) and (b) represents the degree of horizontal and vertical rotation, respectively, and the y-axis shows the number of neighboring views which could be recognized by registering the view in x-axis when allowed certain error rate (0.02 for blue plot, and 0.05 for red plot).

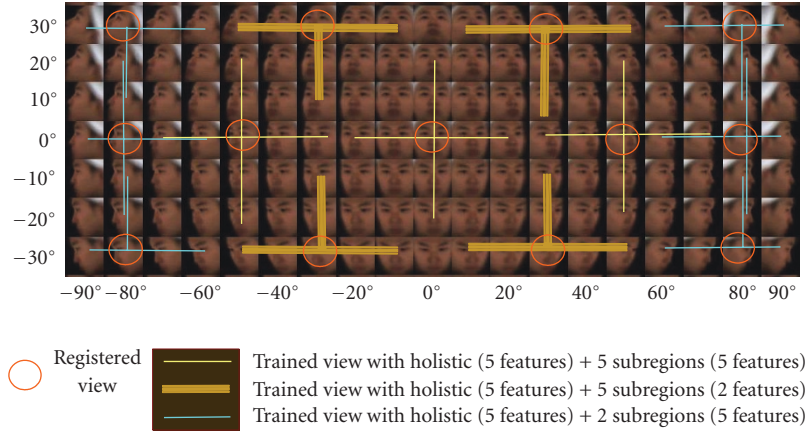


FIGURE 6: Views used for training and registration. 13 representative quasiviews are selected and used for training, and hence for registration. The number of used features (especially, the features for subregions) also varies depending on the view.

For our experiment, MPEG-7 AFR is modified to adapt to be multiview. AFR basically extracts features both in Fourier space and luminance space. In the Fourier space, features are extracted from the whole face, and luminance space extracts features from both the whole face and five subregions on the face as shown in Figure 3(a). We simplify, but also extend, this feature extraction algorithm to our *Subregion-based LDA on Fourier space* for multiview purpose. The biggest differences between the MPEG-7 AFR and our model are (i) feature extraction in luminance space is removed in our model; (ii) the subregion decomposition, which was in luminance space, is now in Fourier space and (iii) the number and positions of subregions are defined depending on a given view, for example, for new frontal views, we use the same five subregions as used in AFR, but for near profile view, we only use two subregions as shown in

Figure 3(b). Figure 4 shows the overall feature extraction diagram. To summarize briefly, we first extract Fourier features from both the whole face image and each subregion of the image, and project all the features and their magnitudes using principle component—linear discriminant analysis (PCLDA) method. After normalizing the resulting vectors, we do additional LDA projection, and finally quantize them for descriptor efficiency. The first two modifications (i) and (ii) give more efficient feature extraction method with smaller descriptor size by extracting the same amount of information on a single space. The third modification (iii) is caused by the multiview extension. If we use the same definition of subregion used in the front view for the profile view, the background may seriously affect for recognition rate. So we define different subregion depending on views as shown in Figure 3.

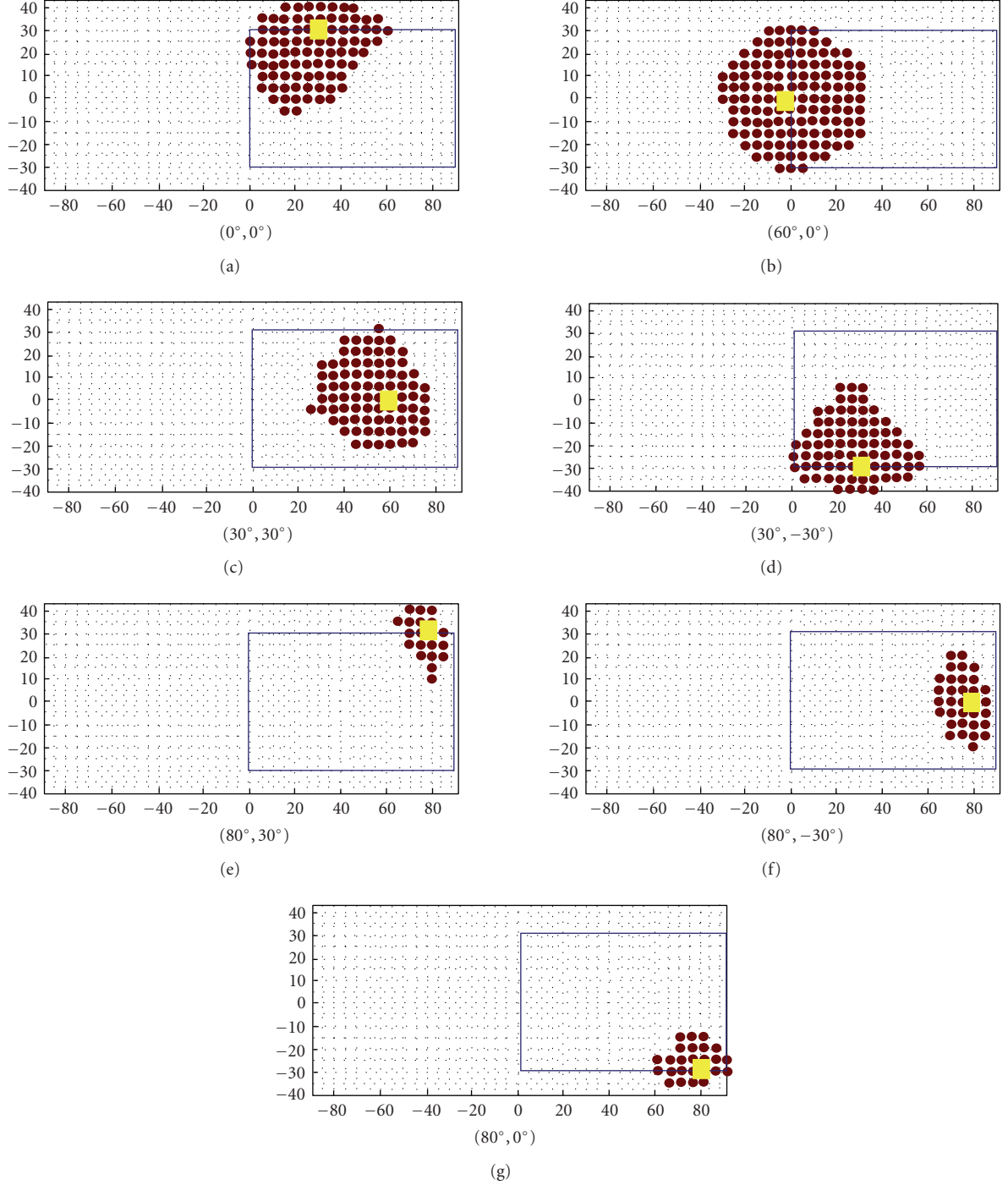


FIGURE 7: Representation of quasiviews. The x -axis and y -axis indicate the horizontal rotation from -90° to 90° and the vertical rotation from -40° to 40° , respectively. Big yellow spots represent the registered views and small red spots indicate corresponding quasiviews with error rate 0.05. The rectangles are the view region of interest in horizontal rotation $[0^\circ \cdots 90^\circ]$ and vertical rotation $[-30^\circ \cdots 30^\circ]$.

5. QUASIVIEW

Graham and Allinson [13] have calculated the distance between faces of different people over pose to predict the pose dependency of a recognition system. Using the average Euclidean distance between the people in the database over the pose angles sampled, they predicted that faces should be

easiest to recognize around the 30° range and consequently, the best pose samples to use for an analysis should be concentrated around this range. Additionally, they expect that faces are easier to recognize at the frontal view $(0^\circ, 0^\circ)$ than the profile $(90^\circ, 0^\circ)$. Here, we use notation of (X°, Y°) to indicate a view with X° horizontal rotation and Y° vertical rotation.

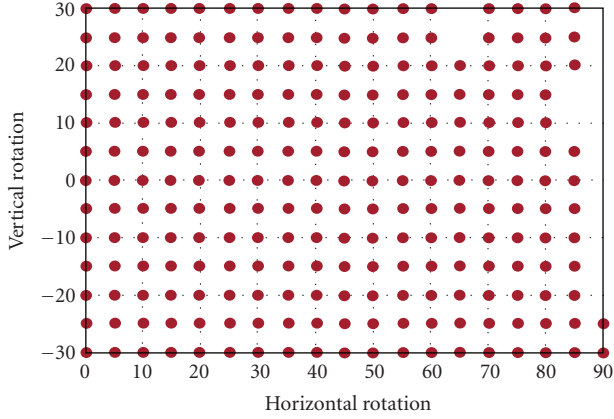


FIGURE 8: The region covered by 7 quasiviews in the view mosaic of horizontal rotation $[0^\circ \cdots 90^\circ]$ and vertical rotation $[-30^\circ \cdots 30^\circ]$ with error rate 0.05. Registration with 7 views covers 93.93% of the view space, which means that we can retrieve faces in any view represented in this plot from the registered descriptor within allowed error rate 0.05.

Note that they have checked only horizontal rotation of human heads.

We use the new concept of “quasiview” corresponding to the conventional “quasifrontal,” which is a measurement of the influence of a registered view for recognition. To prove that quasiview size depends on the view, we performed experiments of quasiview inspection with accepted error rate 0.05, that is, we inspect the range of views that would be recognizable within error rate 0.05 given a view for registration. Figure 5 shows how the quasiview size varies with pure horizontal or vertical rotations of a head. To make fair comparison between different views, we extracted 24 holistic features (without using subregion features) for each view. And images of nearby views are also included in the training of certain view (i.e., in obtaining the PCLDA basis for each view). So for horizontal rotation, 9 views (the view of interest + 8 nearby views) are used for training each view from $(0^\circ, 0^\circ)$ to $(70^\circ, 0^\circ)$, 8 training views for the view $(80^\circ, 0^\circ)$, and 7 training views for the view $(90^\circ, 0^\circ)$. For vertical rotation, 9 training views are used for each view from $(0^\circ, -40^\circ)$ to $(0^\circ, 40^\circ)$, 8 training views for the views $(0^\circ, -50^\circ)$ and $(0^\circ, 50^\circ)$, and 7 training views for the views $(0^\circ, -60^\circ)$ and $(0^\circ, 60^\circ)$. Figure 5 is obtained before adding neighboring images in training. Figure 6 can be helpful to understand which training views are used for each registered view while it reflects our result after optimization.

Figure 5 shows our quasiview measurements with synthetically created (rendered) images of 108 3D facial models by rotating them into various angles. We counted the number of nearby views which could be recognized when we registered a certain view using two kinds of accepted error rates 0.02 and 0.05. The result in Figure 5(a) shows very similar pattern with the graph showing the average distance between faces over view described in Graham and Allinson’s paper [13]. The views $(20^\circ, 0^\circ) \sim (30^\circ, 0^\circ)$ have both the biggest quasiview size and the biggest Euclidean distance between the people in eigenspace among views



FIGURE 9: An example of registration. The views are needed in the registration step to recognize a face in the 93.93% of the view space where horizontal rotation $[0^\circ \cdots 90^\circ]$, vertical rotation $[-30^\circ \cdots 30^\circ]$, and their combined rotation of a head are allowed. It means that we can retrieve a face in various poses within allowed error rate 0.05 when we register only 7 views in a condition that a given face is symmetric.

$(0^\circ, 0^\circ)$, $(10^\circ, 0^\circ)$, \dots , and $(90^\circ, 0^\circ)$. Figure 5(b) shows that the views $(0^\circ, 0^\circ) \sim (0^\circ, 10^\circ)$ have the biggest quasiview size among views $(0^\circ, -60^\circ)$, $(0^\circ, -60^\circ)$, \dots , and $(0^\circ, 60^\circ)$. The views of heading downward have bigger quasiview size than ones of heading upward and it makes us guess that it might be easier to recognize people when they look downward more than they look upward.

6. DESCRIPTOR OPTIMIZATION

Based on our study to check the quasiview size on horizontal and vertical rotated heads, we now optimize the multiview 3D face descriptor by choosing several representative views and recording the corresponding view specific features together. We have used the following selection criteria for registration views: we register views (i) with bigger quasiview size for cost effect; (ii) which appear a lot in practice through target environment analysis, for example, ATM, door access control; (iii) considering efficient integration of quasiviews covering the big region in view-mosaic; (iv) and which are easy to register or easy to obtain. This choice is empirical and we focus on covering the bigger range of face views with more efficient face view registration. Remembering that our features are extracted from PCLDA projections, we can select the dimension of resulting features as we want. Hence, we can also use variable feature numbers depending on the view. If a view is easy to obtain for registration, but not so frequently appear in practice, then we can use a smaller number of features. More important views get bigger feature numbers.

In generating descriptors, training is considered as a step to create space basis and matrix transform for feature extraction and as mentioned in Section 5, many views are trained for one registered view to increase the retrieval ability and reliability. If we can embed more information in the step of training, the registration can be done with smaller information. For example, for the registered view $(30^\circ, 0^\circ)$, we use 9 surrounding views $(10^\circ, 0^\circ)$, $(20^\circ, 0^\circ)$, $(30^\circ, 0^\circ)$, $(40^\circ, 0^\circ)$, $(50^\circ, 0^\circ)$, $(30^\circ, -20^\circ)$, $(30^\circ, -10^\circ)$, $(30^\circ, 10^\circ)$, $(30^\circ, 20^\circ)$ for training. As summarized in Figure 6, for one view registration, the training is done with 6 to 9 views around the registered view. For this experiment, we have given three ways to extract features based on basic feature extraction method described in Section 4. Number of subregions and number of features on subregions vary. So for some views, 5 holistic features and 5 features for each of the five subregions are extracted which results in 30-dimensional view-specific feature vector, and for other views 5, holistic features and 2 features

for 5 subregions are extracted producing 15 dimensional vector. If a view is close to profile, we use 5 holistic features and 5 features for 2. For details for our experiment, see Figures 3 and 6. For one view, one image is selected.

In the experiment for multiview descriptor optimization with rendered images from 3D facial models of 108 individuals, half of the images were used for training and the other half were used for test. We show some examples of quasiview in Figure 7 which shows the influence of each represented view. Big yellow spots are the views for registration and small red spots indicate corresponding quasiviews with allowed error rate 0.05. Therefore, the region covered by small spots surrounding a big spot indicates the influence of the registered view (the big spot). For example, when we register the very front view (the left most one in the middle row in Figure 7), the horizontally 30°-rotated and vertically 20°-rotated views also could be recognized with error rate 0.05.

Through experiments with various combinations of quasiviews, a set of optimal views could be selected to create final multiview 3D descriptor. An example of such possible descriptor from the rendered images contains 13 views with 240-dimensional feature vector as shown in Figure 6. With the allowed error rate 0.05, this descriptor was able to retrieve the rendered images in the test database from 93.93% of the views in view mosaic of horizontal rotation $[-90^\circ \cdots 90^\circ]$ and vertical rotation $[-30^\circ \cdots 30^\circ]$. Figure 8 shows the covered region of the views by the selected 7 views (right half of the view space which corresponds to positive horizontal rotation) considering the symmetry of the horizontal rotation. Figure 9 shows an example which face views are needed for registration to recognize the face in almost any pose. The 7 views are to be registered to recognize a face in the 93.93% of the view space where horizontal rotation $[0^\circ \cdots 90^\circ]$, vertical rotation $[-30^\circ \cdots 30^\circ]$, and their combined rotation of a head are allowed. It means that we can retrieve a face in various poses within allowed error rate 0.05 when we register only 7 views in a condition that a given face is symmetric. For a reference, when we allowed error rate of 0.1, it covers 95.36% of the view space, 97.57% for error rate 0.15, and 97.98% for error rate 0.2. For the experiment, the testing views are situated at intervals 5 degrees while a 10-degree interval is used for training.

As a reference, the MPEG-7 AFR [9, 10] has 48 dimensions with error rate 0.3013 and 128 dimensions with error rate 0.2491 for photograph images. Here we used the error rate of ANMRR (average normalized modified retrieval rank), the MPEG-7 retrieval metric, which indicates how many of the correct images are retrieved as well as how highly they are ranked among the retrieved ones. Details about ANMRR can be found in MPEG related documents like [14].

7. CONCLUSION

We have shown how the single-view face descriptor could be extended to multiview one in efficient way by checking the size of quasiview, which is a measure of the view influence. For the experiment, the 3D facial mesh models of 108 subjects are used and their rendered images are used for training

and test with 50/50 ratio. Only 13 views could be chosen as registered views throughout our optimization. This descriptor in 240 dimensions is able to retrieve images of 93.93% views of total region of view mosaic of horizontal rotation from -90° to 90° and vertical rotation from -30° to 30° within error rate 0.05.

The aim of this new descriptor is to be used to extract a face in any view by containing compact 3D information by optimization for how many and which views are to be registered. The extension to multiview is not very costly in terms of number of registration views thanks to the quasiview analysis. Even though we have used a specific face descriptor for the experiment, the potential of this method enables us to include any available 2D face recognition methods by showing how to combine them in optimized way by checking quasiview size. Ongoing research includes new feature extraction methods for profile views and missing view interpolation in the registration step.

REFERENCES

- [1] A. Samal and P. A. Iyengar, "Automatic recognition and analysis of human faces and facial expressions: a survey," *Pattern Recognition*, vol. 25, no. 1, pp. 65–77, 1992.
- [2] S. Z. Li, L. Zhu, Z. Q. Zhang, A. Blake, H. J. Zhang, and H. Shum, "Statistical learning of multi-view face detection," in *Proceedings of the 7th European Conference on Computer Vision (ECCV '02)*, vol. 4, pp. 67–81, Copenhagen, Denmark, May 2002.
- [3] Y. Li, S. Gong, and H. Liddell, "Support vector regression and classification based multi-view facedetection and recognition," in *Proceedings of the 4th IEEE International Conference on Automatic Face and Gesture Recognition*, pp. 300–305, Grenoble, France, March 2000.
- [4] G. Shakhnarovich, L. Lee, and T. Darrell, "Integrated face and gait recognition from multiple views," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '01)*, vol. 1, pp. 439–446, Kauai, Hawaii, USA, December 2001.
- [5] V. Blanz and T. Vetter, "Face recognition based on fitting a 3D morphable model," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, no. 9, pp. 1063–1074, 2003.
- [6] A. M. Bronstein, M. M. Bronstein, and R. Kimmel, "Expression-invariant 3D face recognition," in *Proceedings of the 4th International Conference on Audio- and Video-Based Biometric Person Authentication (AVBPA '03)*, vol. 2688 of *Lecture Notes in Computer Science*, pp. 62–69, Guildford, UK, June 2003.
- [7] D. M. Gavrila and L. S. Davis, "3-D model-based tracking of humans in action: a multi-view approach," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '96)*, pp. 73–80, San Francisco, Calif, USA, June 1996.
- [8] K. W. Bowyer, K. Chang, and P. Flynn, "A survey of approaches and challenges in 3D and multi-modal 3D + 2D face recognition," *Computer Vision and Image Understanding*, vol. 101, no. 1, pp. 1–15, 2006.
- [9] A. Yamada and L. Cieplinski, "MPEG-7 Visual part of eXperimentation Model Version 17.1," ISO/IEC JTC1/SC29/WG11 M9502, Pattaya, Thailand, March 2003.
- [10] T. Kamei, A. Yamada, H. Kim, W. Hwang, T.-K. Kim, and S. C. Kee, "CE report on Advanced Face Recognition Descriptor,"

- ISO/IEC JTC1/SC29/WG11 M9178, Awaji, Japan, December 2002.
- [11] W.-S. Lee and K.-A. Sohn, "Face recognition using computer-generated database," in *Proceedings of Computer Graphics International (CGI '04)*, pp. 561–568, IEEE Computer Society Press, Crete, Greece, June 2004.
 - [12] W.-S. Lee and K.-A. Sohn, "Database construction & recognition for multi-view face," in *Proceedings of the 6th IEEE International Conference on Automatic Face and Gesture Recognition (FGR '04)*, pp. 350–355, IEEE Computer Society Press, Seoul, Korea, May 2004.
 - [13] D. B. Graham and N. M. Allinson, "Characterizing virtual eigensignatures for general purpose face recognition," in *Face Recognition: From Theory to Applications*, H. Wechsler, P. J. Phillips, V. Bruce, F. Fogelman-Soulie, and T. S. Huang, Eds., pp. 446–456, Springer, Berlin, Germany, 1998.
 - [14] G. Park, Y. Baek, and H.-K. Lee, "A ranking algorithm using dynamic clustering for content-based image retrieval," in *Proceedings of the International Conference Image and Video Retrieval (CIVR '02)*, vol. 2383 of *Lecture Notes in Computer Science*, pp. 328–337, London, UK, July 2002.