

AUTOMATED FACIAL FEATURE DETECTION AND FACE RECOGNITION USING GABOR FEATURES ON RANGE AND PORTRAIT IMAGES

*Sina Jahanbin**, *Hyohoon Choi*†, *Rana Jahanbin*‡, *Alan C. Bovik**

* Department of Electrical and Computer Engineering, The University of Texas at Austin, Austin, TX

† Computer Vision Research Laboratory, Sealed Air Corporation, San Jose, CA

‡ Department of Biomedical Engineering, The University of Texas at Austin, Austin, TX

ABSTRACT

In this paper, we present a novel identity verification system based on Gabor features extracted from range (3D) representations of faces. Multiple landmarks (fiducials) on a face are automatically detected using these Gabor features. Once the landmarks are identified, the Gabor features on all fiducials of a face are concatenated to form a feature vector for that particular face. Linear discriminant analysis (LDA) is used to reduce the dimensionality of the feature vector while maximizing the discrimination power. These novel features were tested on 1196 range images. The same features were also extracted from portrait images, and the accuracies of both modalities were compared. A superior verification accuracy was obtained using the range data, and a highly competitive accuracy to that of other techniques in the literature was also obtained for the portrait data.

Index Terms— Face recognition, Gabor wavelets, Biometry, Range images

1. INTRODUCTION

Biometrics are physiological or behavioral characteristics that can be used by a human being or an automated system to recognize an individual. Examples of researched biometrics include but not limited to: fingerprints, 2D and 3D facial images, iris scans, gait, keystroke dynamics, retina scans, facial thermograms and DNA. The automatic recognition of individual based on their biometric characteristics has attracted researchers' attention in recent years due to the advancement in image analysis methods and the emergence of significant commercial applications.

There is an ever increasing demand for automated personal identification in a wide variety of applications ranging from low to high security such as: human computer interaction, access control, surveillance, airport screening, smart cards, and security. Even though biometrics such as fingerprints and iris scans deliver very reliable performance, the human face remains an attractive biometric because of its advantages over some of the other biometrics. Face recognition

is non-intrusive and does not require aid from the test subjects, whereas other biometrics require a subject's cooperation. For instance, in iris or fingerprint recognition, subjects should look into an eye scanner or place their finger on a fingerprint reader.

The use of portrait images has been heavily researched during the early days of face recognition. Portrait images, also referred to as "2D images", capture texture and facial color information but do not readily provide three dimensional information of faces. More than 30 years of 2D face recognition research has resulted in many sophisticated and mature algorithms. Unfortunately, the performance of 2D face recognition algorithms remains unsatisfactory due to the pose and illumination limitations inherent in 2D images [1].

Recently, quick and low cost acquisition of high quality 3D face scans has become viable due to progresses in three dimensional sensor technology. A "3D image" is a three dimensional representation of the face and can be rendered by a 3D point cloud or a range image. Three dimensional face recognition has become popular among researchers because of its potential to overcome the pose and illumination limitations of 2D images.

Elastic bunch graph matching (EBGM) [2] is a successful 2D face recognition algorithm in which multiple Gabor wavelet coefficients at different scales and orientations are used to model local appearance around fiducial points. In EBGM, the main idea is to encode each 2D portrait image by a set of Gabor coefficients calculated at defined facial landmarks.

Inspired by the EBGM idea, we have employed Gabor wavelets to automatically and accurately detect multiple fiducial points on portrait and range images of faces [3]. Each fiducial point was co-localized using both portrait and range Gabor jets. A jet is a set of Gabor coefficients computed at a pixel. A predetermined search window was placed on a fiducial, and a jet was computed at every pixel in the window. The location of a jet having the maximum similarity to the jets of training data was identified as the precise location of the particular fiducial (please refer to [3] for details). After the automatic fiducial point localization, Gabor jets of all fiducial

points are concatenated in each modality (2D and 3D), creating two independent feature vectors per face. The feature dimensionality of both modalities are separately reduced using LDA.

In this paper a novel personal identification algorithm based on Gabor features extracted from range (3D) representations of the face is evaluated and its performance is compared with its well established portrait-based (2D) counterpart. These identity verification systems are tested using 1196 pairs of range and portrait images and verification accuracies are reported for each modality. To the best of our knowledge this is the first time that Gabor based descriptors are used in face recognition based on 3D representation of faces.

The remainder of this paper is organized as follows: In Section 2 we present an overview of Gabor wavelets and a detailed description of the feature localization method. Section 3 explains the face recognition methods and data. Results are presented in section 4 followed by a conclusion and future work.

2. BACKGROUND

2.1. Gabor Jets

The local appearance around a point, \vec{x} , in a gray scale image $I(\vec{x})$ can be encoded using a set of Gabor coefficients $J_j(\vec{x})$ [4]. Each coefficient $J_j(\vec{x})$ is derived by convolving input image $I(\vec{x})$ with a family of Gabor kernels ψ_j

$$\psi_j(\vec{x}) = \frac{k_j^2}{\sigma^2} \exp\left(\frac{-k_j^2 x^2}{2\sigma^2}\right) \left[\exp(i\vec{k}_j \cdot \vec{x}) - \exp\left(\frac{-\sigma^2}{2}\right) \right] \quad (1)$$

Gabor kernels are plane waves restricted by a Gaussian envelope function of relative width $\sigma = 2\pi$. Each kernel is characterized by a wave vector $\vec{k}_j = [k_v \cos \phi_u \ k_v \sin \phi_u]^T$ where $k_v = 2^{-(v+1)}$ with $v = 0, 1, \dots, 4$ symbolize spatial frequencies and $\phi_u = (\phi/8)u$ with $u = 0, 1, \dots, 7$ are the different orientations of the Gabor kernels used in this paper.

A jet \vec{J} is a set $\{J_j, j = u + 8v\}$ of 40 complex Gabor coefficients obtained from a single image point. Complex Gabor coefficients can be represented in their exponential form $J_j = a_j \exp(i\phi_j)$ where $a_j(\vec{x})$ is the slowly varying magnitude and $\phi_j(\vec{x})$ is the phase of the j th Gabor coefficient at pixel \vec{x} .

The similarity between two jets is defined using the phase sensitive similarity measure:

$$S(\vec{J}, \vec{J}') = \frac{\sum_{i=1}^{40} a_i a'_i \cos(\phi_i - \phi'_i)}{\sqrt{\sum_{i=1}^{40} a_i^2 \sum_{i=1}^{40} a'_i{}^2}} \quad (2)$$

This similarity measure returns real values in the range $[-1, +1]$, where a closer value to $+1$ means a higher similarity between the input jets.

2.2. Gabor Bunch

In order to search for a given feature on a new face image, a general representation of that fiducial point is required. As proposed in [2], the general appearance of each fiducial point can be modeled by bundling the Gabor jets extracted from several manually marked examples of that feature point (e.g. eye corners) in a stack-like structure called a ‘‘Gabor bunch’’.

In order to support a wide range of variations in the appearance of faces caused by subjects’ different gender, race, and facial expression, a comprehensive training set should be selected. For example, the Gabor bunch representing an eye corner should contain jets from open, closed, male, female and other possible eye corners. In this work a training set containing 50 pairs of registered portrait and range images is carefully selected to cover possible variations present in the data set.

The similarity measure between a jet and a bunch is naturally defined to be the maximum of the similarity values between the input jet and each constituent jet of that bunch

$$S_B(\vec{J}, \vec{B}) = \max_{i=1}^{50} S(\vec{J}, \vec{B}_{(i)}) \quad (3)$$

where in (3), \vec{B} represents a bunch and $\vec{B}_{(i)}$ s with $i = 1, \dots, 50$ are its constituent jets.

2.3. Automatic Facial Feature Detection

In order to find the location of facial landmarks on any incoming coregistered range and portrait pairs, we have adapted a multimodal (2D+3D) feature point localization algorithm described in [3].

In the training phase, a set of 50 pairs of registered range and gray scale portrait images was selected in [3] from a database provided by Advanced Digital Imaging Research (ADIR) LLC. (Friendswood, TX). All ADIR’s 2D and 3D face images are roughly aligned with respect to a fixed generic face model using iterative closest point (ICP) algorithm. Hence, all images are frontal face images with nose tip approximately located at the center of the image.

A set of 11 prominent facial features were manually marked only on the portrait images of these 50 training pairs. Since these portrait and range images are perfectly aligned, the location of fiducials on the range image of the pair is exactly the same as the portrait one. Fig. 1 shows a range and portrait pair from the training set with 11 feature points marked with red ‘‘X’’ on the portrait image.

Finally, Gabor jets are calculated from images of each modality (portrait and range) at the manually marked landmarks. All Gabor jets from a specific feature point (e.g. nose tip) are stacked together to create a bunch representation of that fiducial in a given modality. For example the nose tip’s range-bunch describes the nose tip in the range images.

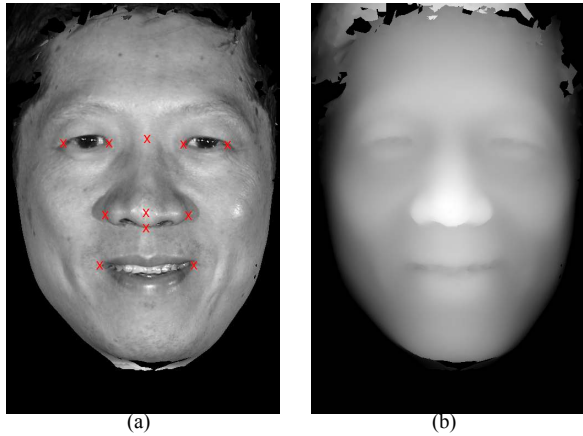


Fig. 1. Example of face images from the ADIR data set a) A portrait image with marked fiducial. b)The corresponding range image.

In the detection phase, each fiducial point is looked for in a search area centered at the average location of that fiducial in the training data. This is because all ADIR faces are coarsely aligned to a frontal view and the prior knowledge about human faces can be used to limit the search areas. For example we only need to look for the left eye in a neighborhood located to the left and above the nose tip. Each search area is a rectangle box with sides selected to be at least 5 times the standard deviation of each fiducial’s coordinates in the training set. In Fig. 2, the search area of the nose tip and inner corners of the eyes are marked with rectangular boxes.

In order to automatically locate a fiducial point on a pair of range and portrait images which have never been seen before, the range and portrait data enclosed by the search area of that feature point are first convolved with the set of 40 Gabor wavelets presented in (1). As a result, each pixel of the search area is represented with two Gabor jets, a “range jet” and a “portrait jet”. Next, the jets of each modality are compared to their corresponding bunch using the similarity measure between a jet and a bunch formulated by (3). Consequently, a similarity map is created for each modality demonstrating the similarity between each pixel in the search area and the appropriate bunch describing the appearance of the target feature point. The final location of our fiducial point is determined by comparing the range and portrait similarity maps and selecting the pixel with the highest similarity score.

3. FACE RECOGNITION MATERIALS AND METHODS

3.1. Data

A collection of 1196 pairs of portrait and range images from 119 subjects has been used in this research. This data base is collected by ADIR LLC., using a stereo imaging system made

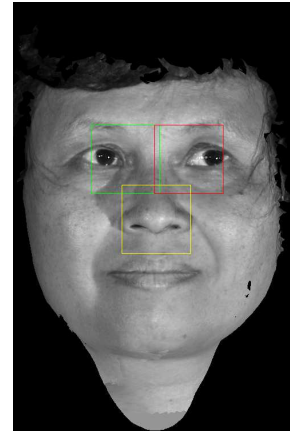


Fig. 2. Search areas of the nose tip and inner eye corners.

by 3Q Technologies Ltd. (Atlanta, GA), and all 2D and 3D images were roughly aligned with respect to a fixed generic face model using the iterative closest point (ICP) algorithm. We have reduced the size of all images by a factor of 3 in each direction and the resulting images are of size 251×167 pixels. Finally, 2D colored portrait images were transformed to gray scale portrait images. No further preprocessing has been applied to the data set.

This data set is partitioned into disjoint training and test sets. The training set contains 390 pairs of gray scale portrait and range images of 10 subjects. The number of image pairs per subject varies between 25 and 47 in the training set. The training set is used to enroll the subjects in a verification experiment. The test set has 382 image pairs from the 10 enrolled subjects plus additional 424 extra images from 109 subjects not enrolled in the verification system.

3.2. LDA Dimensionality Reduction

As explained in section 2.3, 11 landmarks are automatically located on each range or portrait image in the ADIR data set. The magnitude of the complex Gabor jets at these landmarks are concatenated to create a 440 dimensional real-valued feature vector in each modality (range and portrait).

Fisher’s LDA is used to reduced both portrait and range Gabor features to 9 dimensional (9D) spaces that maximize the between-class scatter, S_b , while minimizing the within-class scatter, S_w . Unfortunately in our application the within-class scatter matrix is singular because the number of samples in the training set is smaller than the dimension of the feature space (440). In order to solve this “small sample size problem”, we used the regularization method mentioned in [5]. In this method, the within-class scatter matrix is slightly modified by adding a small positive constant, K , to the diagonal elements of S_w . K is very small compared to the eigen values of S_w .

The projection directions were learned only from the training portion of ADIR date set. The features of the images

in the test set were projected into the lower dimensional space using the projection parameters learned from the training set.

4. RESULTS

To measure the accuracy of our face recognition algorithms, we have arranged an identity verification experiment according to established face recognition evaluation protocols [6]. In a verification scenario, first the individual using the system asserts his identity. The face recognition system compares the retrieved features of the claimed identity (from the database) with the currently captured features of the user and decides whether a match can be declared. This decision is made based on the Euclidean distance between the currently captured features of that individual and the mean value of the features corresponding to claimed identity in the lower dimensional space.

In this paper, the verification performance is evaluated using receiver operating characteristic (ROC) curve and equal error rate (EER) values. Fig. 3 presents the ROC curves for the portrait-based and range-based face recognition algorithms implemented in this work. The EER value for the portrait-based Gabor features is equal to 6.1%. It is observed that by using the range Gabor features, the EER remarkably drops to 2.2%. This remarkable performance of range-based Gabor features compared to one of the most efficient features investigated in '2D' face recognition [6] highlights the important role that Gabor features extracted from range images can play in future face recognition investigations.

The performance of the implemented face recognition algorithm using range-based Gabor features is promising and highly competitive to other well-known 3D face recognition systems in the literature. For example, Pan *et al.* [7] has investigated using a PCA-based approach and a Hausdorff distance approach for 3D face recognition. They have evaluated their algorithms using a database of 360 range images from 30 individuals and reported an EER in the range of 3 – 5% for the Hausdorff distance approach and EER in the range of 5 – 7% for PCA-based approach.

5. CONCLUSION

In this paper we introduced a novel 3D face recognition algorithm based on Gabor responses after automatically identifying fiducial points. The verification performances are reported and compared using a database of 1196 pairs of range and portrait images of expressive and neutral faces. The observed verification accuracies show the highly competitive performance of our algorithm relative to existing face recognition algorithms, and it was shown that the proposed 3D Gabor-based face recognition algorithm outperforms its 2D counterpart. The next logical step to improve the performance is to combine the portrait-based and range-based Gabor features. Another option would be the incorporation

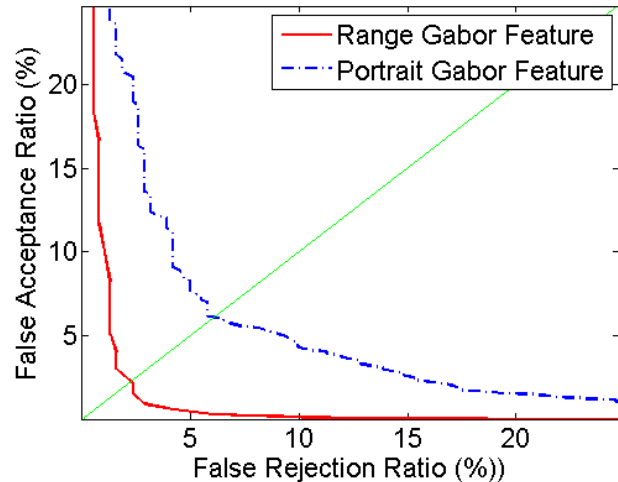


Fig. 3. ROC curves showing the verification accuracies.

of Euclidean or Geodesic distances between these fiducial points.

6. REFERENCES

- [1] K. W. Bowyer, K. Chang, and P. Flynn, "A survey of 3d and multi-modal 3d+2d face recognition," Tech. Rep., Notre Dame Department of Computer Science and Engineering Technical Report., January 2004.
- [2] L. Wiskott, J. M. Fellous, N. Kuiger, and C. von der Malsburg, "Face recognition by elastic bunch graph matching," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 19, no. 7, pp. 775–779, July 1997.
- [3] S. Jahanbin, A. C. Bovik, and H. Choi, "Automated facial feature detection from portrait and range images," *IEEE Southwest Symposium on Image Analysis and Interpretation*, 2008.
- [4] M. Clark, A.C. Bovik, and W.S. Geisler, "Texture segmentation using Gabor modulation/demodulation," *Pattern Recognition Letters*, vol. 6, no. 4, pp. 261–267, 1987.
- [5] W. Zhao, R. Chellappa, and P. Phillips, "Subspace linear discriminant analysis for face recognition," 1999.
- [6] P.J. Phillips, P.J. Grother, R.J. Micheals, D.M. Blackburn, E. Tabassi, and J.M. Bone, "Face recognition vendor test 2002," Tech. Rep., National Institute of Standards and Technology, 2003.
- [7] G. Pan; Z. Wu; Y. Pan, "Automatic 3d face verification from range data," *Acoustics, Speech, and Signal Processing, 2003. Proceedings. (ICASSP '03). 2003 IEEE International Conference on*, vol. 3, pp. III–193–6 vol.3, 6-10 April 2003.