

Automated Facial Feature Detection from Portrait and Range Images

Sina Jahanbin, Alan C. Bovik
Department of Electrical and Computer Engineering
The University of Texas at Austin, USA
{jahanbin, bovik}@ece.utexas.edu

Hyohoon Choi
Computer Vision Research Laboratory
Sealed Air Corp., San Jose, USA
hyohoon@alumni.utexas.net

Abstract

We propose a novel technique to detect feature points from portrait and range representations of the face. In this technique, the appearance of each feature point is encoded using a set of Gabor wavelet responses extracted at multiple orientations and spatial frequencies. A vector of Gabor coefficients, called a jet, is computed at each pixel in the search window on a fiducial and compared with a set of jets, called a bunch, collected from a set of training data on the same type of fiducial. The desired feature point is located at the pixel whose jet is the most similar to the training bunch. This is the first time that Gabor wavelet responses were used to detect facial landmarks from range images. This method was tested on 1146 pairs of range and portrait images and high detection accuracies are achieved using a small number of training images. It is shown that co-localization using Gabor jets on range and portrait images resulted in better accuracy than using any single image modality. The obtained accuracies are competitive to that of other techniques in the literature.

1. Introduction

Automatic detection of facial feature points (fiducials) plays an important role in applications such as facial feature tracking, human-machine interaction, and face recognition.

Many face recognition algorithms operate based on geometrical features like the distances and angles defined between several prominent facial landmarks [3]. Other researchers have used local texture or shape descriptors extracted around one or more fiducial points as desired features for face recognition purposes [4].

The main landmarking assumption in [4] is that the nose tip is the closest point to the 3D sensor and hence in the resulting range image, the nose tip has the highest gray scale value. This method may simply fail when the face registration step erroneously detects a streak of hair or tip of a protruding chin as the nose tip.

The above examples highlight the important role that a robust automatic facial feature detection can play in automation of those face recognition algorithms that currently require human intervention [3]. A robust feature detection method can also improve the performance of face recognition algorithms that are reliant on fiducial detection algorithms founded on poor heuristics [4].

Recently, extensive work has been focused on automatic feature localization from portrait images of the face. A portrait image also referred to as a “2D image,” contains facial texture and color information. The active appearance model (AAM) by Cootes *et al.* [1] is one of the most effective facial landmark detection algorithms on 2D images. An iterative search algorithm seeks the best location for each feature using a texture model describing that feature’s surrounding. These feature locations are then fine-tuned using the spatial distribution of feature points encoded by a shape model. In a later work, Cristinacce *et al.* [2] improved the AAM algorithm and showed that their new shape optimized search (SOS) algorithm outperforms the AAM.

There have been very few techniques proposed in the literature that use 3D facial information for fiducial detection. The existing ones are mainly based on mean and Gaussian curvatures extracted from range images. Curvature features are very sensitive to 3D acquisition noise; therefore, they require extensive preprocessing. Recent studies [5] show that these techniques suffer from a large number of false positives and thus result in low accuracies.

Elastic bunch graph matching (EBGM) [6] 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 this paper, we have extended the concept of using Gabor wavelets to represent local information around feature points to detect features on both 2D portrait and 3D range images of the face. To the best of our knowledge this is the first time that Gabor-based appearance models are used to detect feature points from range images of the face.

The remainder of this paper is organized as follows: In Section 2 we present an overview of Gabor wavelets and

similarity measures defined to compare Gabor wavelet responses. A detailed description of our feature localization methods is given in Section 3. In section 4 our algorithm is evaluated on a dataset of 1146 pairs of co-registered range and portrait images from 119 subjects. The conclusion and future works are presented in section 5.

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 $J_j(\vec{x})$. 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 [6], the general appearance of each fiducial point can be modeled by bundling the Gabor jets extracted from several manually marked examples of that feature points (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 selected carefully 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 equation (3), \vec{B} represents a bunch and $\vec{B}_{(i)}$ s with $i = 1, \dots, 50$ are its constituent jets.

3. Materials and Methods

3.1. Data

In this study we have used a collection of face images provided by Advanced Digital Imaging Research (ADIR) LLC. (Friendswood, TX). This data set contains 1196 pairs of portrait and range images from 119 subjects captured using a stereo imaging system made by 3Q Technologies Ltd. (Atlanta, GA). 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.

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 pre-processing has been applied to the data set.

3.2. Training

A set of 50 pairs of registered range and gray scale portrait images covering a variety of facial appearances from subjects with different ages, races, and genders was selected for the training and Gabor bunch extraction. This training set contains neutral and expressive faces among which many examples have open mouth or closed eyes.

We have manually marked 11 prominent facial features 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. Figure 1 shows a range and portrait pair from the training set with 11 feature points marked with red ‘‘X’’ on the portrait image.

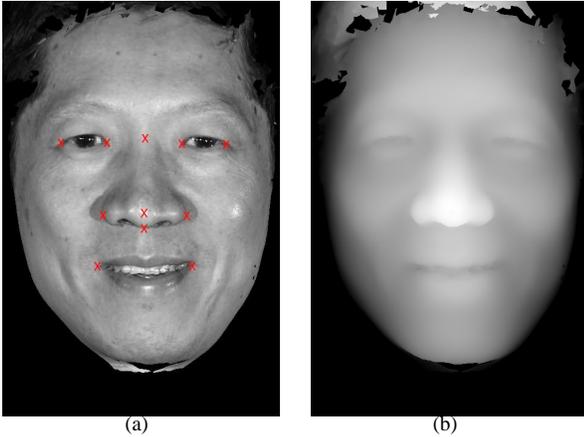


Figure 1. Example of face images from the ADIR data set a) A portrait image with marked fiducial. b)The corresponding range image.

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) and modality are stacked together to create a bunch representation of that fiducial in that modality. For example the nose tip’s range-bunch describes the nose tip in the range images.

3.3. Localization Method

In the elastic bunch graph matching [6], the “search area” of each feature point is constrained by penalizing the deformation of a graph connecting the fiducials in an optimization algorithm. Since all of the face images in ADIR data set are coarsely aligned to a frontal view, our prior knowledge about the human face can be used to limit the search area of each feature point. For example, the nose tip is expected to be located at the center of the image and the left eye’s inner corner is always located above and to the left of the nose tip.

In this work, each fiducial point is looked for in a search area centered at the average location of that fiducial in the training data. Each search area is a rectangle box of size 40×40 pixels. The sides of these rectangular areas are at least 5 times the standard deviation of each fiducial’s coordinates in the training set. Our results show that the search window is reasonably large and all fiducials are located in their expected search area. In figure 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. In each modality’s similarity map, the pixel with the highest similarity value is the target fiducial. Ultimately, the range and portrait information are combined by selecting the modality whose localized feature point has higher similarity score.

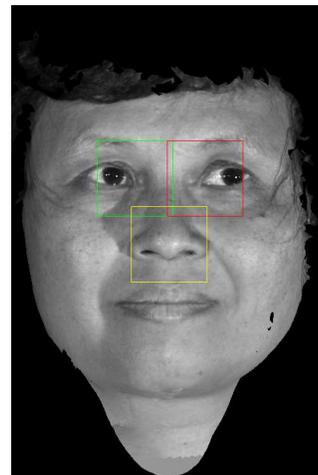


Figure 2. Search areas of the nose tip and inner eye corners.

4. Results

The proposed landmark detection algorithm is tested on the remaining 1146 pairs of facial images available in ADIR data set. The correct location of the feature points are manually marked on all testing images. To evaluate the detection accuracy, the positional error of each detected fiducial is normalized by dividing the error distance with the interocular distance of that face.

The average of the normalized positional errors is denoted as m_e , adopting the same notation as in [2]. Figures 3(a) shows the cumulative probability distribution of m_e averaged over all feature points. This figure indicates that when the acceptable detection error is $m_e \leq 0.06$, the search algorithm based only on range information is successful for 91% of faces whereas the success rate of a localization algorithm based only on portrait information is as high as 96%. Eventually, the combination of range and portrait information works on 98% of faces.

Figure 3(b) demonstrate the same distribution for 4 nasal fiducials (i.e. nose tip, subnasal, and nose left/right corners). This figure indicates that when only portrait information is used by our proposed detection algorithm, nasal fiducials are successfully located on more than 98% of faces given that $m_e \leq 0.06$. From figure 3(b), it is apparent that the range information outperforms the portrait information in finding nasal fiducials for any given value m_e .

The performance of our novel feature point localization algorithm is promising and competitive to well known algorithms like AAM [1] and SOS [2]. Cristinacce *et al.* [2] has compared the performance of AAM versus SOS in detecting 17 features on 1521 facial images. Cristinacce *et al.* used a very large training set containing 1055 face images compared to our algorithm which needs only 50 image pairs. They reported that when the acceptable normalized displacement error is $m_e \leq 0.1$, AAM is successful on 70% and SOS works for 85% of faces. Whereas, with $m_e \leq 0.1$, the success rates of our proposed method are more than 99% for any fiducial using any combination of range or portrait modalities (see figure 3(a)).

Our detection algorithm employs a simple search method with a remarkable performance in finding fiducial points on expressive and neutral facial images. Our proposed localization algorithms require a small set of representative faces for training and because the computational cost is low, it is suitable for real time applications.

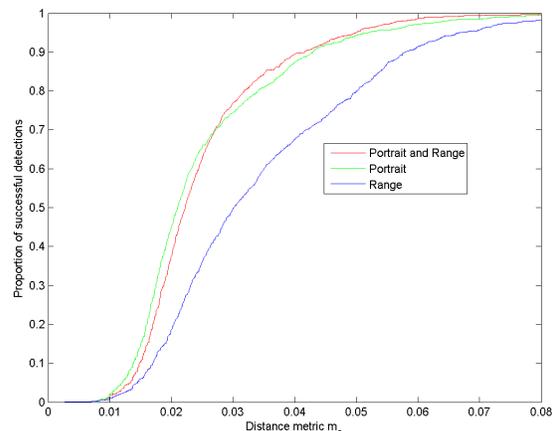
5. Conclusion

In this paper we have presented novel algorithms for 2D and 3D facial landmarking. Gabor coefficients are used to encode the appearance of facial landmarks. Coefficient extracted from pixels of the search window are compared to those of the training data to automatically locate fiducials on images of each modality. The Gabor coefficients have been employed for the first time to detect feature points on range images. It was shown that the proposed 3D Gabor-based feature detection algorithm can outperform its 2D counterpart in detecting nasal feature points. Experimental results on a database of 1146 portrait and range images indicate the highly competitive performance of our method compared to well-known methods in the literature. In the future, it is desirable to adopt a coarse-to-fine landmarking strategy to increase the computational efficiency.

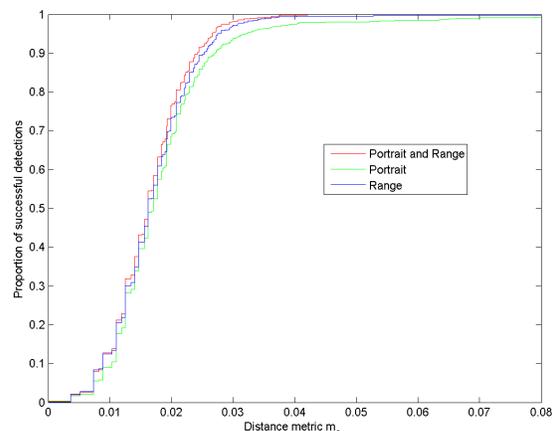
References

[1] T. Cootes, G. Edwards, and C. Taylor. Active appearance models. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 23(6):681–685, June 2001.

[2] D. Cristinacce and T. Cootes. A comparison of shape constrained facial feature detectors. *Automatic Face and Ges-*



(a) All fiducial points



(b) Nasal fiducial points

Figure 3. Average positional error (m_e) for: a) All fiducial points. b) Only nasal fiducial points.

ture Recognition, 2004. Proceedings. Sixth IEEE International Conference on, pages 375–380, 17-19 May 2004.

[3] S. Gupta, M. K. Markey, J. K. Aggarwal, and A. C. Bovik. Three dimensional face recognition based on geodesic and euclidean distances. volume 6499, page 64990D. SPIE, 2007.

[4] S. Jahanbin, H. Choi, A. C. Bovik, and K. R. Castleman. Three dimensional face recognition using wavelet decomposition of range images. *IEEE International Conference on Image Processing*, 2007.

[5] A. Salah and L. Akarun. 3d facial feature localization for registration. *Multimedia Content Representation, Classification and Security*, pages 338–345, 2006.

[6] 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*, 19(7):775–779, July 1997.