Fast and Robust Self-Training Beard/Moustache Detection and Segmentation

T. Hoang Ngan Le, Khoa Luu and Marios Savvides
CyLab Biometrics Center, Carnegie Mellon University, Pittsburgh, PA, USA
thihoanl@andrew.cmu.edu, kluu@andrew.cmu.edu, msavvid@ri.cmu.edu

Abstract

Facial hair detection and segmentation play an important role in forensic facial analysis. In this paper, we propose a fast, robust, fully automatic and self-training system for beard/moustache detection and segmentation in challenging facial images. In order to overcome the limitations of illumination, facial hair color and near-clear shaving, our facial hair detection self-learns a transformation vector to separate a hair class and a non-hair class from the testing image itself. A feature vector, consisting of Histogram of Gabor (HoG) and Histogram of Oriented Gradient of Gabor (HOGG) at different directions and frequencies, is proposed for both beard/moustache detection and segmentation in this paper. A feature-based segmentation is then proposed to segment the beard/moustache from a region on the face that is discovered to contain facial hair. Experimental results have demonstrated the robustness and effectiveness of our proposed system in detecting and segmenting facial hair in images drawn from three entire databases i.e. the Multiple Biometric Grand Challenge (MBGC) still face database, the NIST color Facial Recognition Technology FERET database and a large subset from Pinellas County database.

1. Introduction

Facial hair analysis has recently received significant attention from forensic and biometric researchers because of three important observations as follows. Firstly, changing facial hairstyle can modify a person's appearance such that it effects facial recognition systems. Secondly, most females do not have beard or moustache. Therefore, detecting facial hair helps to distinguish male against female with high confidence in the gender classification problem. Finally, opposed to babies and young adults, only male senior adults generally have beard or moustache. The facial hair detection can help to improve the accuracy of an age estimation system. In addition to beard and moustache detection, segmentation also plays an crucial role, especially in facial recognition systems due to the following biometric observation. There is lack of small patches under a human mouth and these features does not change during the lifetime of a person [5]. Accurate beard and moustache detection and segmentation enables us to perform occlusion removal, gender and ethnicity classification, age estimation, facial recognition, etc.

In 2008, Nguyen, et al.[6] proposed a method for facial beard synthesis and editing. However, this method only works on high resolution images with the assumption that facial hair already existed. Also, there was the requirement of initial seeds for Graphcuts-based method. Their proposed method was lack of experimentally quantitative results. To overcome those weaknesses, Pierrard, et al. [9] proposed facial hair detection using GrabCut and alpha matching. Although their proposed method was evaluated in the application of face recognition on FERET database [8], there was a lack of quantitative results in their work. Recently, Le, et al. [4] [5] proposed a new approach SparCLeS for automatically detecting and segmenting beard/moustache with full measurement. Their method is shown to be robust against illumination and obtains high accuracy of detection. However, it is time consuming and highly complex because of the Multiscale Self-Quotient algorithm and the dictionary learning approach. Furthermore, their facial hair detection depends on the training data used to build the binary decision dynamic dictionary.

In this paper, we aim at overcoming the above handicaps by proposing a fast, robust and automatic detection and segmentation of beard/moustache on challenging facial images without the need of a training dataset. The proposed method is based on three observations as follows:

- Certain areas, such as cheeks, do not typically contain any facial hair
- Certain areas, such as brows, often contain facial hair
- Unlike a skin region, a facial hair region contains high frequency information in different orientations, which depends on whether the hair is straight, wavy or curly

Given an image, the Region Of Interest (ROI), consisting of moustache, left beard, middle beard and right beard,

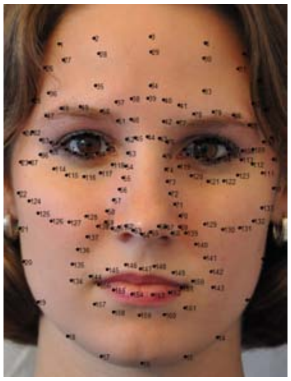


Figure 1. The dense facial landmarking scheme consisting of 161 landmarks.

is first located using a set of facial landmarking points processed by the Modified Active Shape Model (MASM) approach [10]. The ROI is then projected on to the transformation vector computed using the Local Fisher Discriminant Analysis (LFDA) method [11]. Far apart form [4, 5] where training set is required to build a binary dynamic dictionary, in this paper, we propose to use patches extracted from the brows and the cheeks of a given image to construct the transformation vector. Therefore, there is no need to use a training dataset in our proposed method. In order to extract high frequency information, there are two features proposed to used in our method, i.e. Histogram of Gabor (HoG) and Histogram of Oriented Gradient of Gabor (HOGG). Those features will demonstrate their efficiency in facial hair feature extraction in both beard/moustache detection and segmentation in Section 4.

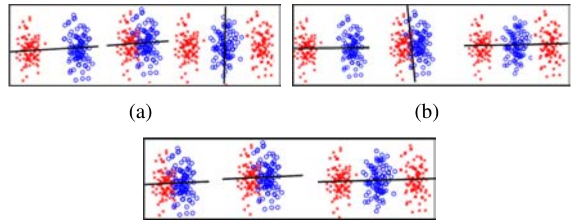
The rest of this paper is organized as follows. Section 2 introduces the background of the main techniques employed in our method, including: the MASM landmarking algorithm, the 2-D Gabor feature extraction method, and the Local Fisher Discriminant Analysis approach. Section 3 describes our proposed algorithm that consists of three key parts, i.e. feature extraction, detection and segmentation. In Section 4, we demonstrate the performance of our proposed algorithms on images collected from three databases, i.e. the Multiple Biometric Grand Challenge (MBGC) database, the color FERET database, and a large subset from the Pinellas County face database.

2. Related Works

2.1. Facial Landmarking

An Active Shape Model (ASM) [1] aims to fit a set of landmarks specified in its training onto an unseen instance of the object. For automatically locating facial landmarks, the ASM must be initialized so that its mean shape lies approximately on top of the face in the given image. In order to receive the best fitting in ASM, an iterative process of moving the landmarks into locations is followed by a shape based constraining step that ensures that the group of landmarks satisfy the structure of a typical face.

In this paper, we propose to use the Modified Active Shape Model (MASM) approach [10] to robustly find facial landmarking points in our system. We employ a dense facial landmarking scheme consisting of 161 landmark points. Figure 1 shows an example of our proposed landmarking scheme. In this MASM approach, it uses 2D texture descriptor and searches for candidate locations of a landmark in a 2D region around the initial locations. At the test stage, a 2D profile around a potential landmark is projected onto a subspace to obtain a vector of projection coefficients and is then reconstructed with these coefficients. The reconstruction error between this reconstructed and the original profiles is calculated. The best candidate location is the one gives the lowest reconstruction error.



(a) (b) (c)

Figure 2. An illustration of discrimination performance by (a)FDA, (b)LPP, (c)LFDA under three different data distributions, i.e. simple, label-mixed and multimodal

2.2. Local Fisher Discriminant Analysis(LFDA)

Fisher Discriminant Analysis (FDA) [2] is a well-known linear discriminant analysis method which maximizes the distance between classes while making each class cluster more compact in this lower dimensional space. In Figure 2(a), it is evident that FDA works well if samples in each class are formed in a Gaussian distribution and the covariance matrices of the classes are equal. However it gives undesired results if samples in each class from multimodal, i.e. several separate clusters. Locality Preserving Projection (LPP) [3] is able to meet requirement of embedding multimodal and is able to help decreasing the distance between nearby samples in the original space. However, is does not take between class separability into account as shown in Figure 2(b). To overcome the disadvantages of FDA which does not take within-class multimodality into account and only extract $(C-1)$ features and LPP which does not take between class separability into account, LFDA [11] is presented to combine the properties of FDA and LPP in order to maximize the between-class separability while preserving the within-class local structure as shown in Figure 2(c).

Given a set of N samples $\underline{\boldsymbol{x}}_i \in \mathcal{R}^M$ belongs to C classes, the class label of $\underline{\boldsymbol{x}}_i$ is denoted as L_i and the number of samples in the class l is denoted as N_l. The within-class scatter $\widehat{\boldsymbol{S}}_w$ and between-class scatter $\widehat{\boldsymbol{S}}_b$ is defined as follows:

$$\widehat{\boldsymbol{S}}_w = \frac{1}{2}\sum_{i,j=1}^{N} \boldsymbol{A}_{i,j}\boldsymbol{W}_{i,j}(\underline{\boldsymbol{x}}_i - \underline{\boldsymbol{x}}_j)(\underline{\boldsymbol{x}}_i - \underline{\boldsymbol{x}}_j)^T \quad (1)$$

$$\widehat{\boldsymbol{S}}_b = \frac{1}{2}\sum_{i,j=1}^{N} \boldsymbol{A}_{i,j}\boldsymbol{B}_{i,j}(\underline{\boldsymbol{x}}_i - \underline{\boldsymbol{x}}_j)(\underline{\boldsymbol{x}}_i - \underline{\boldsymbol{x}}_j)^T \quad (2)$$

$$\boldsymbol{A}_{i,j} = \begin{cases} e^{-\|\underline{\boldsymbol{x}}_i - \underline{\boldsymbol{x}}_j\|^2} & \text{if } ||\underline{\boldsymbol{x}}_i \text{ - } \underline{\boldsymbol{x}}_j|| < \varepsilon \\ 0 & \text{otherwise} \end{cases} \quad (3)$$

where ε is a tuning parameter.

$$\boldsymbol{W}_{i,j} = \begin{cases} 1/N_l & \text{if } (L_i = L_j = l) \\ 0 & \text{if } (L_i \neq L_j) \end{cases} \quad (4)$$

$$\boldsymbol{B}_{i,j} = \begin{cases} 1/N - 1/N_l & \text{if } (L_i = L_j = l) \\ 1/N & \text{if } (L_i \neq L_j) \end{cases} \quad (5)$$

The transformation vector $\underline{\boldsymbol{\omega}}_{\boldsymbol{LFDA}}$ is chosen by the following objective function:

$$\underline{\boldsymbol{\omega}}_{\boldsymbol{LFDA}} = \arg\max_{\underline{\boldsymbol{\omega}}} (\underline{\boldsymbol{\omega}}^T \widehat{\boldsymbol{S}}_w \underline{\boldsymbol{\omega}})^{-1} (\underline{\boldsymbol{\omega}}^T \widehat{\boldsymbol{S}}_b \underline{\boldsymbol{\omega}}) \quad (6)$$

Solving Eqn. (6) is equivalent to finding the solution of the generalized eigenvalue problem of $\widehat{\boldsymbol{S}}_b\underline{\boldsymbol{\omega}} = \lambda\widehat{\boldsymbol{S}}_w\underline{\boldsymbol{\omega}}$. The eigenvector $\underline{\boldsymbol{\omega}}_{\boldsymbol{i}}$ is associated to the eigenvalue λ_i and $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_k$.

3. Proposed Algorithms

This section is composed of three parts. First, we present our proposed feature extraction approach in Section 3.1. These features are employed in both facial hair detection and segmentation. Second, Section 3.2 details the facial hair detection algorithm on challenging beard/moustache images. Finally, the detection results will be used as the inputs in the facial hair segmentation that is presented in Section 3.3.

3.1. Feature Extraction

We propose two kinds of 2D Gabor based features for both facial beard/moustache detection and segmentation. In general, a complex Gabor filter can be formulated as a product between Gaussian function and a complex exponential Sinusoid function. In spatial domain, a complex Gabor filter includes two separated components, i.e. real and imaginary, defined as in Eqn. (7).

$$h(x,y) = K\exp\left(-\frac{R_1^2}{2\sigma_x^2} - \frac{R_2^2}{2\sigma_y^2}\right)\exp\left(j2\pi F_0 R_1\right) \quad (7)$$

where $R_1 = (x-\mu_x)\cos\theta + (y-\mu_y)\sin\theta$ and $R_2 = -(x-\mu_x)\sin\theta + (y-\mu_y)\cos\theta$. The other parameters, i.e. μ_x, μ_y, σ_x, σ_y, denote the means and standard deviations along each corresponding x and y direction respectively, and $K = \frac{1}{2\pi\sigma_x\sigma_y}$. The frequency F_0 and the direction ω_0 define the sinusoid function in the polar coordinates.

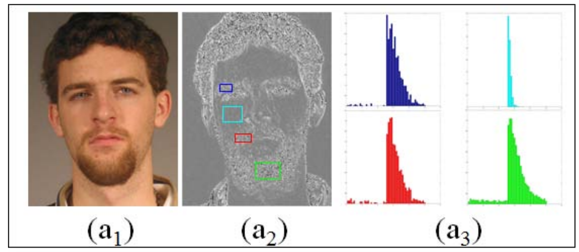


Figure 3. An illustrate of HoG: original image (a_1), Gabor filtering image (a_2), HoG of four ROIs corresponding to brow (in blue), cheek (in cyan), moustache (in red) and beard (in green) (a_3)

In this work, our first feature, Histogram of Gabor (HoG), is defined as the histogram of a 2D Gabor feature constructed using a Gabor filter defined as in Eqn. (7) at different frequencies F_0 and orientations θ.

Given a testing image, its corresponding HoG feature is computed in four steps as follows:

- Apply the 2D Gabor filter at different frequencies and orientations to obtain a set of Gabor filtering images
- Compute the histogram of each filtering image
- Concatenate the histograms of all Gabor filtering images
- Normalize the set of histograms using the l_2-norm normalization.

Given an image, a demonstration of HoG feature in four different ROIs, i.e. brow (in blue), cheek (in cyan), moustache (in red) and beard (in green) can be shown as in Figure 3.

In the second proposed feature in this paper, the Histogram of Orientated Gradient of Gabor (HOGG) method is employed to extract robust facial hair features. Given an image, the 2D Gabor filter is first applied to obtain the Gabor features. Both the phase and the magnitude of the Gabor filtered image are then taken into account in our method. The HOGG feature is computed in four steps as follows:

- Apply the 2D Gabor filter at different frequencies and orientations to obtain a set of Gabor filtering images
- For each Gabor filtered image:
  - Divide the image into a small blocks (cells)
  - Compute the phase and the magnitude of each block
  - Calculate the histogram of edge orientation of each block, where the gradient direction is based on the phase image and gradient is based on the magnitude image
  - Represent the histogram of orientation of the Gabor filtering image as a set of these block histogram

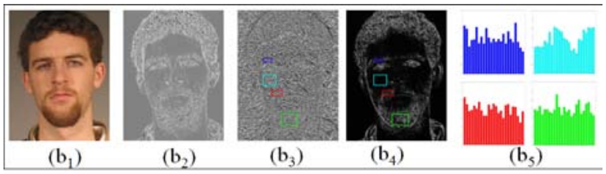


Figure 4. An illustrate of HOGG: original image (b_1), Gabor filtering image (b_2), phase of Gabor filtering image(b_3) and magnitude of Gabor filtering image(b_4), HOGG of four ROIs corresponding to brow (in blue), cheek (in cyan), moustache (in red) and beard (in green)(b_5)

- Concatenate histograms of all Gabor filtering images
- Normalize the set of histogram using l_2 norm

A demonstration of HOGG feature in four different ROIs, i.e. brow (in blue), cheek (in cyan), moustache (in red) and beard (in green), can be seen in Figure 4(b).

3.2. Beard/Moustache Detection

In this subsection, our proposed self-training beard/moustache detection is presented in detail. The method is fully automatic and does not require any training dataset. Given an unknown testing image, the landmarking is first applied to obtain 161 key points. In this method, there are eight ROIs, i.e. left brow, right brow, left cheek, right cheek, moustache, left beard, middle beard and right beard, defined using 161 landmarking points extracted from the MASM component as presented in Section 2.1. An example of those proposed regions are illustrated in Figure 5. Even the MASM is quite capable of handling images with and without beard and moustache in challenging indoor and outdoor illuminations [5], still, locating the keypoints along facial boundary with beard is a very difficult problem and has some limitations as shown by empirical results. In order to overcome these weaknesses, the ROIs of the beard, i.e. left beard, right beard and middle beard, are defined using a reference distance between the two eyes. As such, the key points are able to cover the boundary in the case that MASM results underfit as shown in the Figure 5. The overall flowchart of our proposed self-training facial hair detection method is given in Fig. 6.

Our proposed self-training beard/moustache detection algorithm contains two main stages, i.e. building a projection subspace and classification. In the first stage, there are four ROIs used to present a hair class and a non-hair class. Particularly, the left and right brow regions are delegated for the hair class. Meanwhile the left and right cheek regions are assigned for the non-hair class. They are set up according to the following assumptions. There are certain areas, such as cheeks, not typically containing facial hair. Meanwhile, the other certain areas, such as brows, often contain facial hair.

In order to avoid the problem of under-fitting, due to not having enough training data, each ROI is first divided into

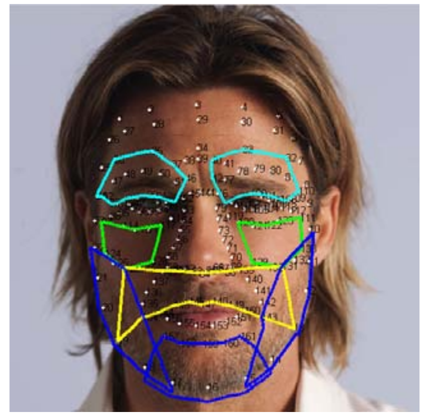


Figure 5. An illustration of eight ROIs based on the landmarking scheme MASM with 161 key points

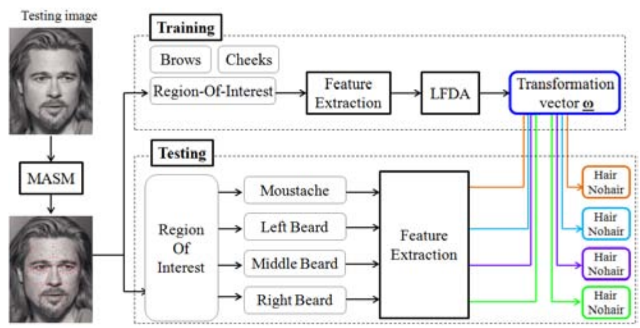


Figure 6. Flowchart of our proposed beard/moustache detection method

a set of patches. Let n_1, n_2, n_3 and n_4 be the number of patches divided from the left brow, the right brow, the left cheek and the right cheeks respectively. A feature vector consists of HoG and HOGG at different frequencies and orientations are then extracted from the patches. Each feature vector is considered as a sample x_i which belongs to either a hair class or a non-hair class. There are $N_1 = n_1 + n_2$ samples in the hair class and $N_2 = n_3 + n_4$ samples in the non-hair class. A transformation vector $\underline{\omega}$ defined by Eqn. (6) that separates the hair and non-hair classes is then determined. The goal of the second stage is to determine if a given testing image contains beard/moustache.

In the testing stage, in order to determine if facial hair occurs in a particular ROI, we first divide this ROI into a set of patches. The feature vectors comprising HoG and HOGG is then extracted from the patches and projected onto the transformation vector $\underline{\omega}$. Each patch will be classified as either '1' if containing facial hair or '-1' if having no hair. Based on the enclosed number of patches classified as either hair or non-hair classes, the ROI is classified as containing hair or not.

3.3. Facial Hair Segmentation

In this subsection, we present the proposed feature-based segmentation to extract the beard/moutache from a ROI which contains facial hair, according to the result from the beard/moustache detection. The overall flowchart of our proposed facial hair segmentation is given in Fig. 7.

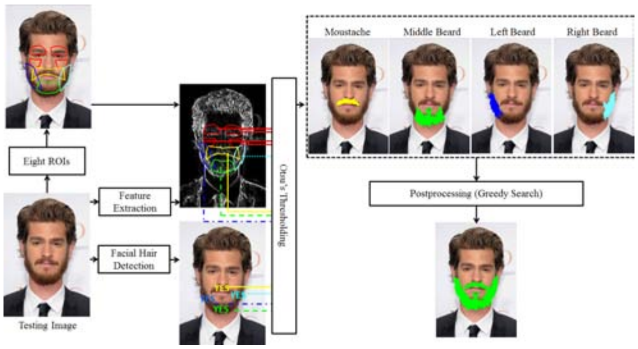


Figure 7. Flowchart of our proposed facial hair segmentation algorithm

Given an unknown testing image, we first perform facial hair detection in order to discover the presence of facial hair. Based on the detection result, the proposed facial hair segmentation is separately applied if relevant. Let X be a ROI which is detected to have facial hair. In order to extract the facial hair from X, the feature vector consisting of HoG and HOGG of each pixel from the region X and the four ROIs of brows and cheeks is first extracted and **reduced in dimension** by projecting onto the subspace $\underline{\omega}$. Using 161 key points of MASM to crop the cheeks and brows regions, it is shown that the cheeks region contain skin only and the brows region contain both skin and hair whereas the region X has hair and might have skin as illustrated in Fig. 5. Thus, each feature vector belongs to either hair or non-hair (skin) and is represented by one value in the projection subspace. The Otsu's binarization [7] which automatically searches for a threshold by minimizing the within class variance, defined as the weighted sum of the variances of each cluster, is then applied to the set of values. In our algorithm, the region X can be either moustache or left bread or middle beard or right beard and the facial hair of each region is separately extracted by thresholding. Even MASM works well on the facial landmarking, though it still goes off track in fitting the beard of an unseen image which presents facial hair. In order to extract the full beard, we use a greedy search strategy as a postprocessing to find all the facial hair outside the boundary keypoints. The postprocessing starts with facial hair pixels extracted by the threshholding and then considering if eight-neighbor also contains facial hair. Each neighbor pixel is projected on the subspace and the similarity between the neighbor pixel and the facial hair is computed by cosine distance. Fig. 7 shows the flowchart of our proposed segmentation algorithm.

4. Datasets and Results

We conduct the experiments on large and diverse databases to evaluate our proposed self-detecting and segmenting facial hair algorithm. The images used for our experimental results are from the NIST Multiple Biometric Grand Challenge - 2008 (MBGC) still face challenge database which contains 34,696 still frontal images of 810 subjects with varying facial expressions and illumination conditions. There are 989 color images of 989 subjects under different illuminations from the NIST Color Facial Recognition Technology (FERET) database [8]. We test our proposed algorithm on 10,000 images with JPEF "artifact" from the first folder DIR-000 in the Pinellas County database. The detailed distribution of images used in our experiments is shown in Table 1 corresponding to MBGC, FERET, Pinellas. Each database is divided into three categories: having facial hair (either moustache or beard), male without facial hair and female without facial hair.

Table 1. Images distribution of MBGC, FERET and Pinellas databases used in our experiments

	MBGC	FERET	Pinellas
Facial Hair	2,050	157	4,782
Non Facial hair (Male)	16,860	434	3,296
Non Facial hair (Female)	15,786	398	1,922
Total	34,696	989	10,000

The performance of our proposed self-training facial hair detection algorithm evaluated on the three databases in Table 1 and is provided in Fig.8. The accuracy is computed as the ratio of number of correct detections and the sum of number of correct and incorrect detections for each category. To compare with the SparCLeS algorithm proposed by Le, et al.[5], we run the proposed facial hair detection on the same dataset which was used in [5] and the comparison is shown in Fig.9. Regarding time consumption, we tested our approach on the MBGC database and on a CPU with Intel Core i7 processor, 2.93 GHz, 8GB of RAM and Matlab 2011a, our proposed algorithm consumed 5.6 seconds per image and is much faster than [5] which consumed 38.6 seconds per image.

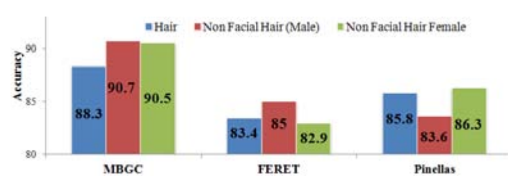
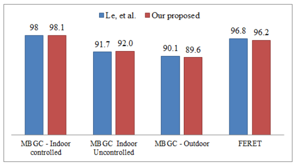
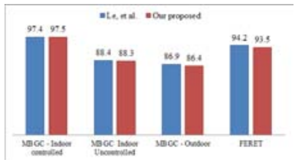


Figure 8. Accuracy of our proposed facial hair detection

Fig.10, Fig.11, Fig.12 show some results produced by our proposed segmentation algorithm on MBGC, FERET



(a)



(b)

Figure 9. Performance comparison of our proposed algorithm and Let, et al.[5]

and Pinellas databases respectively. These are purely qualitative results that visually indicate the effectiveness of our segmentation approach. To provide the quantitative results, we used the same images from the MBGC and FERET which were used in [5] to conduct the experiment. The average F-measure of our proposed segmentation algorithm is 98.5 while [5] achieved slightly better segmentation result with 98.8 of F-measure. However, our method consumes less than 2 seconds per image, whereas the time consumption of [5] is more than 15 seconds per image relying upon the number of iterations.

Figure 10. Examples of FHS on MBGC database

Figure 11. Examples of FHS on FERET database

Figure 12. Examples of FHS on Pinellas database

5. Conclusions

In this paper, we have proposed a fast, robust, fully automatic and self-training beard/moustache detection and segmentation method. In order to emphasize the high frequency feature of facial hair, we first proposed two features, namely, HoG and HOGG. The proposed beard and moustache detection is able to self-learn the transformation vector which discriminates the hair and non-hair classes by extracting HoG and HOGG features from the ROIs of brows and cheeks. Based on the detection results, the facial hair segmentation is then proposed by feature-based thresholding algorithm with the postprocessing to handle the situation where the landmarking points are inside the actual facial boundary. Our algorithms have been evaluated on the entire MBGC, FERET database and a large subset from Pinellas database and found to obtain highly accurate results of detection and segmentation with small time consumption.

References

[1] T. F. Cootes, C. J. Taylor, D. H. Cooper, and J. Graham. Active shape models - their training and application. *Computer Vision and Image Understanding*, 61(1):38–59, 1995.

[2] R. A. Fisher. The Use of Multiple Measurements in Taxonomic Problems. *Annals of Eugenics*, 7:179–188, Sept. 1936.

[3] X. He and P. Niyogi. Locality Preserving Projections. In *Neural Information Processing Systems Foundation (NIPS)*, pages 1–8, 2003.

[4] T. H. N. Le, K. Luu, K. Seshadri, and M. Savvides. Beard and Mustache Segmentation Using Sparse Classifiers on Self-quotient Images. *IEEE International Conference on Image Processing (ICIP)*, pages 165–168, 2012.

[5] T. H. N. Le, K. Luu, K. Seshadri, and M. Savvides. Sparcles: Dynamic l_1 Sparse Classifiers with Level Sets for Robust Beard/Moustache Detection and Segmentation. *IEEE Transactions on Image Processing (TIP)*, 22(8):3097–3107, 2013.

[6] M. H. Nguyen, J. F. Lalonde, A. A. Efros, and F. de la Torre. Image-based Shaving. *Computer Graphics Forum Journal (Eurographics 2008)*, 27(2):627–635, 2008.

[7] N. Otsu. A Threshold Selection Method from Gray-level Histogram. *IEEE Transactions on Systems, Man, and Cybernetics*, 9(1):62–66, Jan. 1979.

[8] P. J. Phillips, H. Moon, P. J. Rauss, and S. Rizvi. The FERET Evaluation Methodology for Face-Recognition Algorithms. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 22(10):1090–1104, Oct. 2000.

[9] J. S. Pierrard. Skin Segmentation for Robust Face Image Analysis. PhD Thesis, University of Basel, Faculty of Science, 2008.

[10] K. Seshadri and M. Savvides. Robust Modified Active Shape Model for Automatic Facial Landmark Annotation of Frontal Faces. In *Biometrics: Theory, Applications and Systems (BTAS)*, pages 319–326, Sept. 2009.

[11] M. Sugiyama. Dimensionality Reduction of Multimodal Labeled Data by Local Fisher Discriminant Analysis. *Journal of Machine Learning Research*, 8:1027–1061, May 2007.