BEARD AND MUSTACHE SEGMENTATION USING SPARSE CLASSIFIERS ON SELF-QUOTIENT IMAGES

T. Hoang Ngan Le, Khoa Luu, Keshav Seshadri and Marios Savvides

{thihoanl, kluu, kseshadr}@andrew.cmu.edu, msavvid@ri.cmu.edu
Department of Electrical and Computer Engineering and CyLab Biometrics Center
Carnegie Mellon University, Pittsburgh, PA, USA

ABSTRACT

In this paper, we propose a novel system for beard and mustache detection and segmentation in challenging facial images. Our system first eliminates illumination artifacts using the self-quotient algorithm. A sparse classifier is then used on these self-quotient images to classify a region as either containing skin or facial hair. We conduct experiments on the MBGC and color FERET databases to demonstrate the effectiveness of our proposed system.

Index Terms— Beard/mustache detection, segmentation, self-quotient image, sparse classifier

1. INTRODUCTION

The detection and segmentation of beards and mustaches plays an important role in gender identification, age estimation, and facial recognition. It is not a trivial task to detect and segment beards and mustaches due to the variation in their shapes, colors, and intensities and because of the effects of shadows and illumination artifacts.

We propose a novel approach for beard and mustache detection and segmentation on challenging images. Our method first pre-processes images using the self-quotient algorithm [1], [2] and subsequently uses a sparse classifier to efficiently detect and segment regions containing facial hair. To the best of our knowledge, this is the first time a sparse classifier has been used for both detection and classification. We demonstrate the accuracy of beard and mustache segmentation by our method on images from the NIST Multiple Biometric Grand Challenge (MBGC) still face database [3] and the color Facial Recognition Technology (FERET) database [4].

The rest of this paper is organized as follows. In section 2 we describe how self-quotient images are generated. In section 3 we describe the Modified Active Shape Model (MASM) that is used to automatically localize facial landmarks. Section 4 describes our sparse classifier while section 5 describes how the previously mentioned tools are used in our beard and mustache detection system. Finally, we demonstrate the effectiveness of our system on images from the MBGC and color FERET databases in section 6 and present some conclusions on this work in section 7.

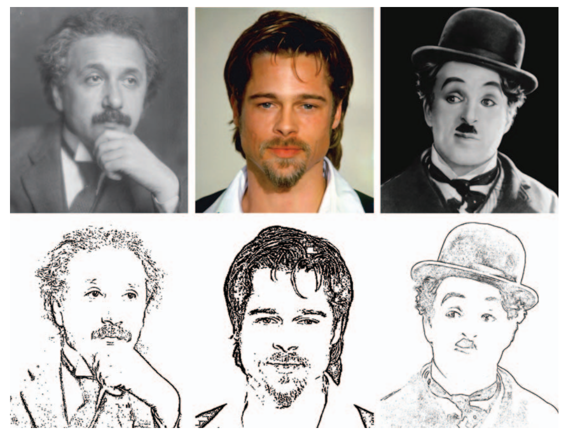

Fig. 1. Examples of self-quotient images. The fist row shows the original images and the second shows the corresponding self-quotient images.

2. THE SELF-QUOTIENT ALGORITHM

The problem of illumination variation in facial images is one that has received a lot of attention in the recent past. An approach using self-quotient images, similar to the quotient images described in [5], is considered a simple and practical method to dealing with facial images under different lighting conditions. The self-quotient image $\mathbf{Q}$ of an image $\mathbf{I}$ can be obtained using the pseudocode given in Algorithm 1. Examples of self-quotient images are shown in Fig. 1.

3. THE MODIFIED ACTIVE SHAPE MODEL (MASM) ALGORITHM

Active Shape Models (ASMs) [6] are deformable templates that aim at modeling the shape of an object using landmark points that lie along its contours. In order to model faces, an ASM requires several images with an identical set of fa-

Algorithm 1 Self-Quotient algorithm

Input: Input image $\mathbf{I}$, Gaussian filter $\mathbf{G}$ of size $s \times s$
Output: Self-Quotient image $\mathbf{Q}$
for all pixel $\mathbf{I}(x, y)$ **do**
  Consider a window $\mathbf{W}$ of size $s \times s$ around $\mathbf{I}(x, y)$
  Compute the anisotropic filter $\mathbf{F}_{\mathbf{W}(x,y)}$ at the location (x, y)
  $\mathbf{F}_{\mathbf{W}(x,y)} = \begin{cases} \mathbf{G}(x,y) & \text{if } \mathbf{W}(x,y) \geq \text{Mean}(\mathbf{W}) \\ 0 & \text{if } \mathbf{W}(x,y) < \text{Mean}(\mathbf{W}) \end{cases}$
  $\mathbf{Z}(x, y) = \sum\sum(\mathbf{F}_{\mathbf{W}(x,y)} \circ \mathbf{W}(x, y))$
  Compute the weight $\mathbf{w}$
  $w = (s \times s) \times \sum\sum \mathbf{F}_{\mathbf{W}}$
  $w = \frac{1}{w}$
end for
Compute self-quotient image $\mathbf{Q}$ and correct singularities $\mathbf{Q} = \frac{\mathbf{I}}{w\mathbf{Z}}$
Adjust histogram and normalize image $\mathbf{Q}$

cial landmarks manually marked in each image. An ASM aligns these set of training faces and uses Principal Component Analysis (PCA) to build a Point Distribution Model (PDM) of facial shape variation. It also builds local texture models of the pixel intensities around each landmark at multiple image resolutions. An ASM can be used to automatically localize these facial landmarks in an unseen image once it is provided with the approximate location of the face in the image using a face detector. This fitting process is a multi-resolution and iterative one that proceeds by moving the landmarks into locations whose surrounding texture best match the training data and subsequently deforming the shape of the face to ensure that it is valid one. The final set of landmark locations is ready once convergence is attained at the finest image resolution.

The Modified Active Shape Model (MASM) described in [7] make a few changes to the above classical ASM to ensure higher landmark fitting accuracy. Chief among these are the use of a new cost function to determine optimal landmark locations during the fitting process and the modeling of local texture around landmarks using a subspace based technique. MASM is able to automatically locate a dense set of 79 facial landmarks in unseen images and is tolerant to variations in illumination, expression and slight variations in pose.

4. SPARSE CLASSIFIER

Sparse representation was first applied to the task of face recognition by Wright et al. [8] and Wagner et al. [9]. In their approach, a sparse classifier is used in a classification problem that is set up as the problem of unseen face identification given a set of training images and the corresponding class labels.

Given a probe sample $\boldsymbol{\nu}_{k,probe}$ of a subject k, the sparse classifier assumes that $\boldsymbol{\nu}_{k,probe}$ can be approximately repre-

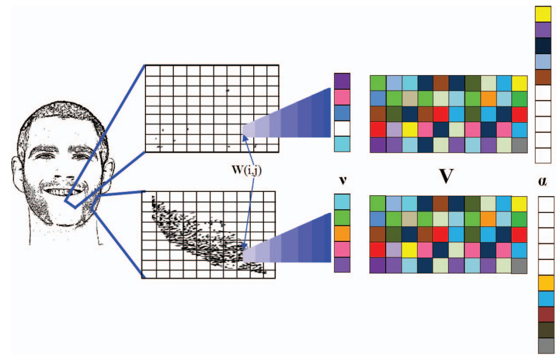

Fig. 2. Sparse classifier used in our proposed approach.

sented by the combination of all samples of the same subject in the gallery set as shown in (1), in which $\boldsymbol{\nu}_{k,i}$ and $\boldsymbol{\alpha}_{k,i}$ are the samples in the gallery corresponding to subject k and the coefficients associated with these samples respectively, N_k is the number of training images of subject k, and $\boldsymbol{\varepsilon}_k$ is the approximation error which is normally distributed.

$$\boldsymbol{\nu}_{k,probe} = \sum_{i=1}^{N_k} \boldsymbol{\alpha}_{k,i}\boldsymbol{\nu}_{k,i} + \boldsymbol{\varepsilon}_k \tag{1}$$

When there are many classes present in the training set, (1) is generally reformulated as shown in (2), in which C is the total number of classes and the total number of samples $N = N_1 + N_2 + \ldots N_C$.

$$\boldsymbol{\nu}_{k,probe} = \sum_{i=1}^{N_1} \boldsymbol{\alpha}_{1,i}\boldsymbol{\nu}_{1,i} + \ldots + \sum_{i=1}^{N_C} \boldsymbol{\alpha}_{C,i}\boldsymbol{\nu}_{C,i} + \boldsymbol{\varepsilon}_k \tag{2}$$

This equation can be compactly represented using matrix-vector notation as shown in (3), in which $\mathbf{V}$ is a matrix of all samples and $\boldsymbol{\alpha}$ is a vector of coefficients whose values correspond to the weights of the different samples in $\mathbf{V}$.

$$\boldsymbol{\nu}_{k,probe} = \mathbf{V}\boldsymbol{\alpha} + \boldsymbol{\varepsilon} \tag{3}$$

The linearity assumption coupled with (3) implies that the elements of the coefficients vector $\boldsymbol{\alpha}_{i,j}$ should be non-zero only when they correspond to the correct class of the probe samples. Based on this assumption, an optimization problem was proposed in [8] and is shown in (4), in which η is related to $\boldsymbol{\varepsilon}$ and is optimized by the MATLAB solver, SPGL1 [1].

$$min_{\boldsymbol{\alpha}} \|\boldsymbol{\alpha}\|_1 \text{ subject to } \|\boldsymbol{\nu}_{k,probe} - \mathbf{V}\boldsymbol{\alpha}\|_2 \leq \eta \tag{4}$$

We use a binary sparse classifier to decide whether a patch consists of skin or facial hair. Fig. 2 illustrates how the sparse classifier works. The gallery is set up as an overcomplete dictionary using 500 skin and 500 hair patches of size $r \times r$ drawn from images in the MBGC database. Each patch is

[1] http://www.cs.ubc.ca/labs/scl/spgl1/

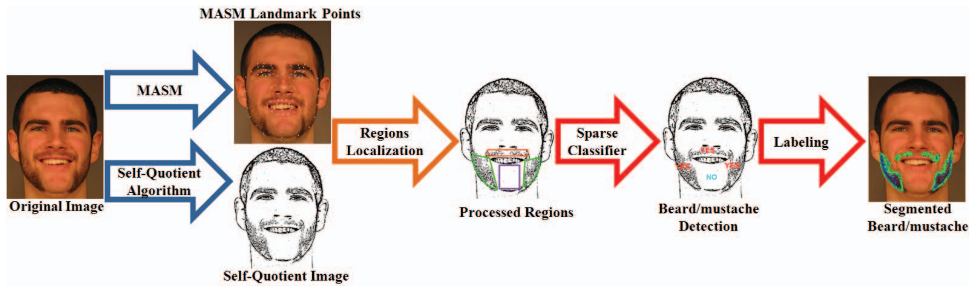

Fig. 3. Our proposed beard/mustache detection and segmentation system.

represented by a set of 28 features as defined in [10]. These features contain information on image textural characteristics such as homogeneity, gray-tone linear dependencies, contrast, number and nature of boundaries present and the complexity of the image.

5. OUR PROPOSED APPROACH

This section describes our proposed beard/mustache detection and segmentation system, which is shown in Fig. 3. First, MASM is used to automatically localize 79 key facial landmarks in an input image. Next, we acquire the self-quotient image of the input image using Algorithm 1. We then use the locations of some of the landmarks to choose regions corresponding to the mustache, left beard, middle beard and right beard. 28 dimensional feature vectors drawn from these regions are used to set up our dictionary and classification scheme.

To summarize, our system is built using the following steps:

1. **Training database**: The overcomplete dictionary (or gallery) $\mathbf{V}$ is built as described in section 4.

2. **Choose regions of interest**: Obtain four regions of interest in the self-quotient image (mustache, right beard, middle beard and left beard) using the landmarks detected by MASM.

3. **Detection**: Each region of interest ({mustache, left beard, middle beard, right beard}), denoted by $\mathbf{Y}$, is processed by considering a window $\mathbf{W}(i,j)$ of size $r \times r$ ($r = 25$) around each pixel $\mathbf{Y}(i,j)$ in the region. Each window is now represented by a feature vector $\boldsymbol{\nu}_{probe}$ of size 28. The sparse vector $\boldsymbol{\alpha}$ is calculated for each feature vector using (4) and a binary value is assigned to pixel $\mathbf{Y}(i,j)$ using (5).

$$\mathbf{Y}(i,j) = \begin{cases} 1 & \text{if} \quad \mathbf{W}(i,j) \text{ is classified as Hair} \\ 0 & \text{if} \quad \mathbf{W}(i,j) \text{ is classified as Skin} \end{cases} \tag{5}$$

The newly created binary region $\mathbf{Y}$ is multiplied (element wise multiplication, represented by $\circ$) by a Gaussian kernel $\mathbf{G}$ and it is decided if the region $\mathbf{Y}$ contains facial hair (YES region) or skin (NO region) using the number of skin and hair patches and their distributions as shown in (6).

$$\mathbf{Y} = \begin{cases} \text{YES} & \text{if} \quad \sum\sum(\mathbf{Y} \circ \mathbf{G}) > \frac{r^2}{2} \\ \text{NO} & \quad \text{otherwise} \end{cases} \tag{6}$$

4. **Segmentation**: Each patch is labeled using the detection results of step 3. A patch is given the label 1 if a detection result of hair is obtained and is given the label 0 if a detection result of skin is obtained.

Test Dataset	Beard Labeling Accuracy (in %)	Mustache Labeling Accuracy (in %)
MBGC	96.2	98.8
FERET	95.8	97.0

Table 1. Beard and mustache classification accuracy of our method on the MBGC and FERET test sets.

6. EXPERIMENTS AND RESULTS

We use images drawn from two databases to evaluate our beard/mustache detection and segmentation system. These are the NIST Multiple Biometric Grand Challenge (MBGC) still face database and the color Facial Recognition Technology (color FERET) database. The MBGC database contains 34,729 still frontal images of 810 subjects presenting varying facial expressions and collected under different illumination conditions. The color FERET database consists of 2413 facial images of 856 subjects with varying facial expressions, illumination, ages, and eyeglass wear. Our test sets consisted of 2500 images from the MBGC database and 989 images from the color FERET database which were manually labeled to indicate whether the subjects in the images had beards or mustaches or not to serve as ground truths. Table 1 shows the percentage of images in the two test sets that were correctly classified as having or not having beards and mustaches while Fig. 4 shows sample results obtained by our method on images from the same test sets. These experimental results

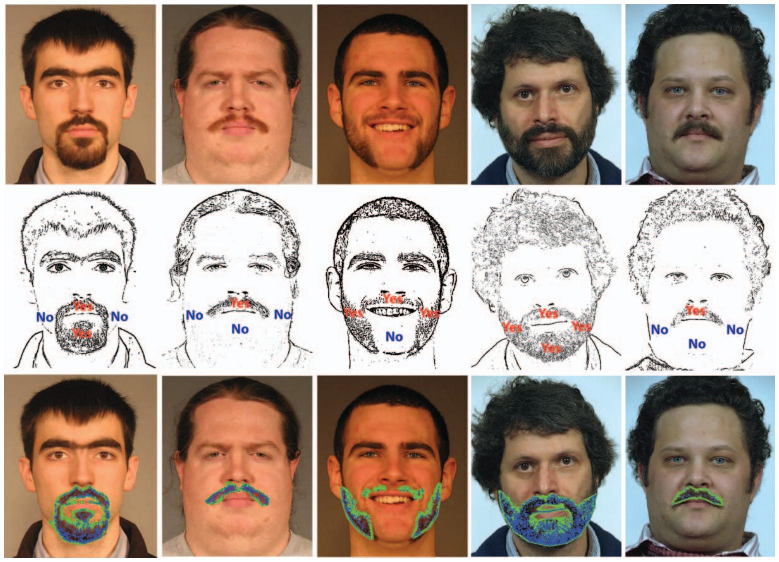

Fig. 4. Sample results produced by our approach on images from the MBGC (first three columns) and color FERET (last two columns) test sets. The first row shows the original images, the second row shows the beard/mustache classification results on the corresponding self-quotient images, and the last row shows the segmented beard and mustache regions.

demonstrate the effectiveness of our method and its ability to deal with facial images exhibiting varying illuminations and expressions.

7. CONCLUSIONS

We have presented a novel beard and mustache detection and segmentation system. Our system first automatically localizes a set of facial landmarks using the Modified Active Shape Model (MASM) and subsequently uses a sparse classifier on features extracted from regions around these landmarks in the self-quotient image. The use of self-quotient images facilitates the removal of illumination artifacts leading to more robust performance. Our method was tested on images from the MBGC and color FERET databases and was found to produce promising results.

8. ACKNOWLEDGMENTS

This research was supported in part by CyLab at Carnegie Mellon under grants DAAD19-02-1-0389 and W911NF-09-1-0273 from the Army Research Office and by cooperative agreement W911NF-10-2-0028 from the Army Research Laboratory.

9. REFERENCES

[1] H. Wang, S. Li, and Y. Wang, "Face Recognition under Varying Lighting Conditions Using Self Quotient Image," in *Proc. of the* 6^{th} *IEEE Intl. Conf. on Automatic Face and Gesture Recognition (FG)*, 2004, pp. 819–824.

[2] S. .U. Jung and J. .H. Yoo, "Robust Eye Detection Using Self Quotient Image," in *Proc. of the Intl. Symposium on Intelligent Signal Processing and Communications (ISPACS)*, 2006, pp. 263–266.

[3] NIST, "National Institute of Standards and Technology: Multiple Biometric Grand Challenge (MBGC) - 2008," `http://face.nist.gov/mbgc`, 2008.

[4] NIST, "National Institute of Standards and Technology: The Color FERET Database Version 2," `http://www.nist.gov/itl/iad/ig/colorferet.cfm`, 2003.

[5] A. Shashua and T. R. Raviv, "The Quotient Image: Class-Based Re-Rendering and Recognition With Varying Illuminations," *IEEE TPAMI*, vol. 23, no. 2, pp. 129–139, Feb. 2001.

[6] T. F. Cootes, C. J. Taylor, D. H. Cooper, and J. Graham, "Active Shape Models - Their Training and Application," *CVIU*, vol. 61, no. 1, pp. 38–59, Jan. 1995.

[7] Keshav Seshadri and Marios Savvides, "Robust Modified Active Shape Model for Automatic Facial Landmark Annotation of Frontal Faces," in *Proc. of the* 3^{rd} *IEEE Intl. Conf. on Biometrics: Theory, Applications and Systems (BTAS)*, 2009, pp. 319–326.

[8] J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry, and Y. Ma, "Robust Face Recognition via Sparse Representation," *IEEE TPAMI*, vol. 31, no. 2, pp. 210–227, Feb. 2009.

[9] A. Wagner, J. Wright, A. Ganesh, Z. Zhou, and Y. Ma, "Towards a Practical Face Recognition System: Robust Registration and Illumination by Sparse Representation," in *Proc. of IEEE CVPR*, 2009, pp. 597–604.

[10] R. M. Haralick, K.Shanmugam, and I.Dinstein, "Textural Features for Image Classification," *IEEE Trans. On Systems, Man, and Cybernertics*, vol. 3, no. 6, pp. 610–621, 1973.