SparCLeS: Dynamic ℓ_1 Sparse Classifiers With Level Sets for Robust Beard/Moustache Detection and Segmentation

T. Hoang Ngan Le, *Student Member, IEEE*, Khoa Luu, *Student Member, IEEE*, and Marios Savvides, *Member, IEEE*

***Abstract*— Robust facial hair detection and segmentation is a highly valued soft biometric attribute for carrying out forensic facial analysis. In this paper, we propose a novel and fully automatic system, called SparCLeS, for beard/moustache detection and segmentation in challenging facial images. SparCLeS uses the multiscale self-quotient (MSQ) algorithm to preprocess facial images and deal with illumination variation. Histogram of oriented gradients (HOG) features are extracted from the preprocessed images and a dynamic sparse classifier is built using these features to classify a facial region as either containing skin or facial hair. A level set based approach, which makes use of the advantages of both global and local information, is then used to segment the regions of a face containing facial hair. Experimental results demonstrate the effectiveness of our proposed system in detecting and segmenting facial hair regions in images drawn from three databases, i.e., the NIST Multiple Biometric Grand Challenge (MBGC) still face database, the NIST Color Facial Recognition Technology FERET database, and the Labeled Faces in the Wild (LFW) database.**

***Index Terms*— Beard/moustache detection, beard/moustache segmentation, dynamic sparse classifier, multiscale self-quotient (MSQ) image, active contour model, active shape model (ASM).**

I. Introduction

FACIAL hairstyle is a critical facial component that has received significant attention from forensic and biometric researchers in the recent past because a person's appearance can be significantly altered by a change in facial hairstyle and this affects our ability to recognize the same person. The detection and segmentation of beards and moustaches also plays an important role in gender identification, age estimation, and even facial recognition, as it can serve as an important soft biometric attribute. In the problem of gender classification [1], beards/moustaches are employed to distinguish males from females with high confidence. Facial hair is also used to effectively support facial age estimation systems [2], [3] in classifying facial images into groups, such as babies, young adults, and senior adults, due to the fact that only males in the last category generally have facial hair. An interesting point in the case of facial hair is that in many men who have beards, there is a lack of small patches of hair under the mouth. This observation of missing facial hair under the mouth could helpful in face recognition as it does not change during the lifetime of a person. These facts motivate the need for carrying out accurate beard and moustache detection and segmentation, which is the aim of this paper.

There is a lack of prior work devoted exclusively to beard and moustache segmentation. Nguyen et al. [4] proposed a method, based on statistical models, for facial beard synthesis and editing. Their paper also discusses facial beard segmentation using a graph cuts based method. However, there are some drawbacks to their approach. It is only effective on high resolution facial images and cannot detect and segment sophisticated facial beard regions. Furthermore, the family of graph cuts based methods usually require manually provided initial landmark points (also referred to as seeds). This is a major handicap as their method can never be deployed in a scenario where automatic segmentation is required. In addition, there is a lack of quantitative results in their work. Pierrard et al. [5] proposed a beard and moustache detection approach using Grabcuts and alpha matting. Their approach was tested and evaluated on the FERET database with face recognition. However, there is again a lack of quantitative results in their work as evaluating the segmentation accuracies obtained by their method. We aim at overcoming the above handicaps in our work and thus propose a novel approach, called SparCLeS, for the completely automatic detection and segmentation of beards and moustaches on challenging facial images.

SparCLeS is composed of three stages. In the first stage, which is a preprocessing one, an input facial image is enhanced using an illumination normalization approach. Typical illumination preprocessing algorithms can be divided into two groups: histogram transformation based approaches and photometric normalization based approaches. Algorithms in the former category try to model light variation, but require a large amount of training data and can not model complicated light sources, while algorithms in the latter category are able to remove lighting effects without any extra domain knowledge. Our experimental results show that an illumination preprocessing approach using the Multiscale Self-

Manuscript received April 15, 2012; revised August 14, 2012, December 11, 2012, and April 9, 2013; accepted October 30, 2012. Date of publication April 24, 2013; date of current version June 4, 2013. This work 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 a Cooperative agreement under Grant W911NF-10-2-0028 from the Army Research Laboratory. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Arun A. Ross.

The authors are with the CyLab Biometrics Center and the Department of Electrical and Computer Engineering, Carnegie Mellon University, Pittsburgh, PA 15213 USA (e-mail: thihoanl@andrew.cmu.edu; kluu@andrew.cmu.edu; msavvid@ri.cmu.edu).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TIP.2013.2259835

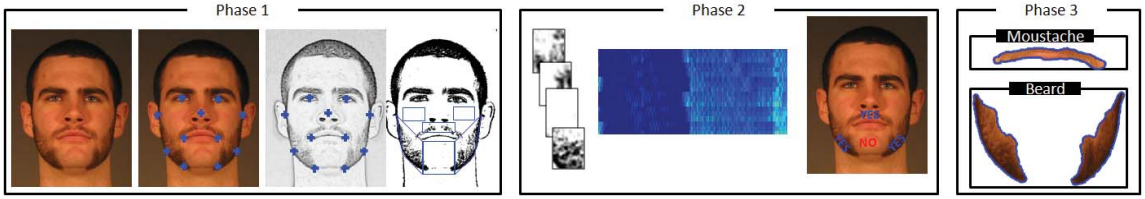


Fig. 1. An outline of our proposed system. There are three key phases: (1) automatic facial landmark detection using MASM, image binarization (illumination normalization), and extraction of feature descriptors from key facial regions (2) beard/moustache detection using a dynamic sparse classifier, and (3) beard/mustache segmentation using our proposed level-set approach.

Quotient (MSQ) algorithm achieves the best performance. Thus, we first normalize an image using the MSQ algorithm and then process the resulting image using Otsu's binarization method [6]. Based on the observation that certain areas (such as portions of the cheeks) of the face do not typically contain any facial hair, we perform a relative area patch analysis in order to determine if hair is present on other areas of the face. It is to be noted that captured facial images usually contain illumination variations, different color tones, different quality due to lossy compression algorithms, etc. These differences are particularly prominent among images that are captured using different devices, such as webcams, scanners, and digital cameras, or under different environmental conditions. Thus, in the classical sparse representation methods [7], [8] it is important to ensure that all training data for the dictionary is obtained under the same illumination, pose, resolution, etc. conditions. These methods are therefore quite limited and are not ideally suited for practical applications, where the training and testing sets are usually acquired from different sources and under different conditions. Our proposed approach addresses this problem by using a dynamic sparse dictionary that allows for the use of training and testing data acquired from different sources with more variability in illumination. In the second stage of SparCLeS, we use a binary decision dynamic sparse classifier to determine if a region contains facial hair or not (i.e. skin). The dictionary of regions that is used for classification is built from Histogram of Oriented Gradients (HOG) feature [9] descriptors extracted from images captured under identical conditions. The third and final stage of our approach uses a new level set based algorithm to accurately and robustly segment the detected beard and moustache regions. Our proposed segmentation approach takes advantages of both local and global information that is modeled by four optimization terms: (1) a local fitting term, (2) a global fitting term, (3) a contour length term, and (4) a regularization term. We use the Modified Active Shape Model (MASM) [10] algorithm to initialize our detection and segmentation stages and thus automate the entire process. An outline of our proposed system is shown in Fig. 1.

To the best of our knowledge, this is the first work that describes the development and thorough evaluation of an automatic facial hair (i.e. beard and moustache) detection and segmentation system. Thus, our contributions in this paper are: (1) the integration of an appropriate preprocessing algorithm (MSQ), a dynamic sparse classifier, and a novel segmentation algorithm based on level sets to form SparCLeS - a new beard/moustache detection and segmentation algorithm, (2) the automation of SparCLeS by using MASM for initialization, and (3) a comprehensive qualitative and quantitative analysis of the detection and segmentation results produced by our algorithm across challenging images drawn from three diverse databases.

The rest of this paper is organized as follows. Section II describes the preprocessing stage of SparCLeS that uses the MSQ algorithm while section III introduces the MASM landmarking algorithm to automatically detect a predefined set of facial landmarks that are subsequently used to initialize the detection and segmentation steps in SparCLeS. Section IV describes our dynamic sparse classifier while section V describes how active contours models, composed of local and global energy functions, are used to segment beards and moustaches. In section VI, we demonstrate the beard/moustache detection and segmentation accuracy of SparCLeS on images from the NIST Multiple Biometric Grand Challenge - 2008 (MBGC) still face database [11], [12], the NIST Color Facial Recognition Technology (FERET) database [13], and the Labeled Faces in the Wild (LFW) [14] database. Finally, section VII offers some conclusions on this work.

II. Illumination Normalization by the Multiscale Self-Quotient (MSQ) Algorithm

Variations in illumination pose a major problem to facial recognition systems and as a result, have received much attention in recent years. Recently, Quotient-Image (QI) based methods were reported to be a simple and efficient solution to the problem. It has been proven that the Quotient-Image, which is the obtained as the image ratio of a test image to a linear combination of three unknown independent illumination images, depends only on the relative surface texture information, and is thus, illumination free. However, such a QI based method method makes a number of assumptions about facial shape, the absence of shadows, and the alignment of the images, which significantly limits its application.

To address some of these issues, Wang et al. [15] proposed the Self-Quotient Image (SQI). The SQI $\mathbf{Q}$ of an image $\mathbf{I}$ is defined by (1), in which $\mathbf{S}$ is a smoothing kernel, w is the weight associated with the kernel, $\bigotimes$ is the convolution or filtering operation and the division operation refers to element by element division of two matrices.

$$\mathbf{Q} = \frac{\mathbf{I}}{w(\mathbf{S} \bigotimes \mathbf{I})} \tag{1}$$

Pseudocode for obtaining the SQI $\mathbf{Q}$ of an image $\mathbf{I}$ is given in Algorithm 1.

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 $\mathbf{Q}$

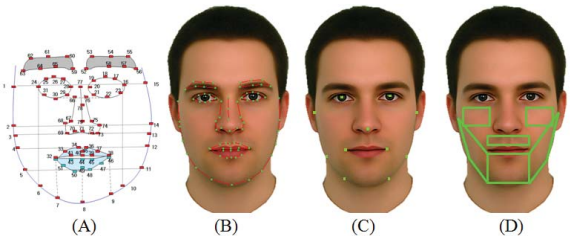


Fig. 2. Using MASM as a tool in our detection and segmentation system. (A) The landmarking scheme used by MASM consisting of 79 facial landmarks, (B) an example of an image automatically landmarked using MASM, (C) the 11 key facial landmarks finally used in our system, (D) the six facial crop regions used in our method.

A detailed analysis to prove that the SQ algorithm is robust against illumination variation under three cases: 1) in regions without shadows and with small surface normal variations, 2) in regions without shadows but with large surface normal variations, and 3) in regions with shadows [15]. Although the analysis is based on the Lambertian model of point illumination, it is also valid for other types of illumination sources. This is because any illumination can be expressed as a linear combination of several illumination sources. It is also shown that the SQI algorithm improves on the QI algorithm by not requiring an image alignment procedure.

The SQI algorithm is very useful for illumination normalization. However, the properties of image Q are dependent on the kernel size of the filter S (as shown in 1). If the kernel size is too large, halo effects begin to appear. On the other hand, if the kernel size is too small, all reflection information will be lost. In this system, we use the Multiscale Self-Quotient (MSQ) algorithm which is an extension of the SQI algorithm that adds up SQI images obtained using kernels of varying dimensions $s_1, s_2, \ldots, s_n$ for illumination normalization.

We also compared the performance of the MSQ and SQI algorithms in our proposed beard/moustache detection and segmentation. We observed that they obtained almost similar results on controlled illumination images. However, the MSQ algorithm outperformed the SQI algorithm on images with uncontrolled illumination variations.

III. The Modified Active Shape Model (MASM) Algorithm

Active Shape Models (ASMs) [16] are deformable templates that model the shape of an object using landmark points that lie along its contours. In order to model faces, an ASM requires several images which are manually annotated in a consistent manner with the same set of landmarks. During the training stage, the ASM first aligns the facial shapes into a common reference frame using Generalized Procrustes Analysis [17] and then builds a shape model or a Point Distribution Model (PDM) to capture the variation in shape exhibited by its training set. This PDM is obtained by carrying out Principal Component Analysis (PCA) on the n aligned shape vectors $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n$ and retaining the eigenvectors corresponding to the directions of maximum variance. It is now possible to approximately represent a shape $\mathbf{x}$, which is a vector of the x and y coordinates of its landmarks, using the mean shape $\overline{\mathbf{x}}$, a matrix of eigenvectors $\mathbf{P}$, and a vector of the projection coefficients (PCA coefficients) of the basis vectors $\mathbf{b}$ using the shape model equation that is given in (2).

$$\mathbf{x} \approx \overline{\mathbf{x}} + \mathbf{Pb} \tag{2}$$

The next stage of training involves building statistical models of the local appearance (grayscale pixel intensities) of regions surrounding each landmark. This is also referred to as construction of profiles. Profiles can either be built by sampling pixels along a line through the landmarks (1D profiling) or by sampling small square regions around the landmarks (2D profiling). Most recent ASM implementations, such as those in [18], [10], use 2D profiles to capture more texture information.

There have been several modifications to ASMs over the past few years to improve fitting accuracy. We use a recent ASM implementation, called the Modified Active Shape Model (MASM), in our work for automatically localizing facial landmarks in unseen images. MASM uses only 2D profiles and searches in a 2D region around a landmark in order to determine the best location for it. The use of 2D profiles results in a better modeling local texture around a landmark, compared to the 1D profiles used by the classical ASM approach described by Cootes et al. in [19]. MASM also goes a step further than conventional ASMs by building a PCA subspace of the pixel variation around each landmark. At the test stage, a 2D profile around a potential landmark location is projected onto this subspace to obtain a vector of projection coefficients and is then reconstructed using these coefficients. The reconstruction error between this reconstructed profile and the original profile is calculated and the candidate location with the lowest reconstruction error is the one that is used as the new location for the landmark in question. The use of these improvements contribute to the fairly high fitting accuracy of MASM and enable it to outperform classical ASM, as demonstrated in [10].

The 79 point landmarking scheme used by MASM is shown in Fig. 2(A) while Fig. 2(B) shows a face which was automatically fitted, i.e., the 79 facial landmarks in the image were automatically detected, by MASM. However, for our

purposes, we only require a subset of 11 landmarks that are located in the typical regions where a beard or moustache may be found. The locations of these 11 landmarks that are used by us in our beard/moustache detection and segmentation system are shown in Fig. 2(C) while the six regions around these landmarks, the left cheek, right cheek, moustache, middle beard, left beard, and right beard, that are of interest to us are shown in Fig. 2(D).

IV. Our Dynamic Sparse Classifier for Beard/Moustache Detection

Sparse representation was first applied to a face recognition problem by Wright et al. [7] and Wagner et al. [8]. In their approach, a sparse classifier was designed and used to classify a face, given a set of training images (the subject was assumed to be in the training set) and the corresponding class labels.

Given a test sample $\boldsymbol{v}_{k,test}$ of a subject (or class) k, it is assumed that $\boldsymbol{v}_{k,test}$ can be approximately represented as a linear combination of the training samples belonging to its correct class as shown in (3), in which $\boldsymbol{v}_{k,i}$ and $\boldsymbol{\alpha}_{k,i}$ are training samples that belong to class k and and the coefficients associated with these samples respectively, N_k is the number of training images for subject k and $\boldsymbol{\varepsilon}_k$ is the approximation error, which is normally distributed.

$$\boldsymbol{v}_{k,test} = \sum_{i=1}^{N_k} \boldsymbol{\alpha}_{k,i} \boldsymbol{v}_{k,i} + \boldsymbol{\varepsilon}_k \tag{3}$$

When there are C classes presented in the training set, (3) is generally reformulated as shown in (4), in which C is the total number of classes.

$$\boldsymbol{v}_{k,test} = \sum_{i=1}^{N_1} \boldsymbol{\alpha}_{1,i} \boldsymbol{v}_{1,i} + \cdots + \sum_{i=1}^{N_C} \boldsymbol{\alpha}_{C,i} \boldsymbol{v}_{C,i} + \boldsymbol{\varepsilon} \tag{4}$$

This equation can be compactly represented using matrix-vector notation as shown in (5), in which $\mathbf{V}$ is a matrix of all training samples (stacked as vectors) and $\boldsymbol{\alpha}$ is a vector of coefficients whose values correspond to the weights of the different samples in $\mathbf{V}$.

$$\boldsymbol{v}_{k,probe} = \mathbf{V}\boldsymbol{\alpha} + \boldsymbol{\varepsilon} \tag{5}$$

The linearity assumption coupled with (5) implies that the elements of the coefficients vector $\boldsymbol{\alpha}_{i,j}$ should be non-zero only when they correspond to the correct (the i^{th}) class of the test sample in question. Ideally, sparse optimization takes the form of (6), in which $\|\|_0$ is the l_0 norm and is defined as the number of non-zero terms in the vector $\boldsymbol{\alpha}$.

$$min_{\boldsymbol{\alpha}} \|\boldsymbol{\alpha}\|_0 \text{ subject to } \|\boldsymbol{v}_{k,test} - \mathbf{V}\boldsymbol{\alpha}\|_2 \leq \varepsilon \tag{6}$$

The problem setup by (6), that involves minimizing the l_0 norm of a vector, is an NP hard one and is converted into a simpler problem that provides an approximate solution by either using convex optimization or a greedy search algorithm. In the first approach, the l_0 norm is replaced by its nearest convex surrogate, the l_1 norm, and is solved by quadratic programming. The l_1 norm solution can be obtained by reframing the original problem, given by (6), as a Basis Pursuit Denoising (BPDN) problem, as shown in (7), or as a Least Angle Shrinkage and Selection Operator (LASSO) problem, given in (8), in which η is related to $\boldsymbol{\varepsilon}$.

$$min_{\boldsymbol{\alpha}} \|\boldsymbol{\alpha}\|_1 \text{ subject to } \|\boldsymbol{v}_{k,test} - \mathbf{V}\boldsymbol{\alpha}\|_2 \leq \varepsilon \tag{7}$$

$$min_{\boldsymbol{\alpha}} \|\boldsymbol{v}_{k,test} - \mathbf{V}\boldsymbol{\alpha}\|_2 \text{ subject to } \|\boldsymbol{\alpha}\|_1 \leq \eta \tag{8}$$

In the second approach, that uses a greedy search algorithm to solve (6), a solution can be usually be found using Matching Pursuit or Orthogonal Matching Pursuit [20]. Our system uses the convex optimization framework to reframe (6) as (7) and then find a solution using a convex solver.

Our problem can be thought of as a binary classification problem i.e. the determination of the presence or absence of facial hair in different parts of the face. Therefore, two classes ($C = 2$), a skin class and a hair class, are modeled in the dictionary $\mathbf{V}$. Based on our observations across a large amount of, we have found that there is typically no hair found in the facial regions right below the eyes. Thus, these cheek regions are used for extracting image patches and building the skin model class in the dictionary. Similarly, a number of patches containing hair samples are used to build the hair model. An overcomplete dictionary is thus synthesized for both the classes.

In order to discriminate between skin and facial hair, we have used HOG features. It has been observed that the difference between skin and facial hair regions is that the latter contains edges (high frequency information) in different orientations, which depend on whether the hair is straight, curly, or wavy, whereas the skin region is much more smooth. Based on this observation, we used HOG features, which reflect occurrences of gradients in different orientations. The process by which our dynamic overcomplete dictionary is generated and used to detect the presence or absence of beards and moustaches beard/non-beard is expressed as follows:

Step 1: Creating the fixed hair class.

Step 1a: Build a gallery with about $N_P = 10$ facial images of different people who have beards or moustaches. Obtain and then binarize the MSQ images.

Step 1b: Crop the beard/moustache regions and divide them into overlapping or semi-overlapping patches (or regions) of size $w \times w$. In our experiments, we choose w to be one tenth of the interocular distance.

Step 1c: Represent each patch by a Histogram of Oriented Gradients (HOG) feature vector. HOG features are calculated by obtaining the distribution of intensity gradients or edge directions (orientations) in local regions.

Step 2: Creating the dynamic skin class and set up the sparse classifier.

Step 2a: For a given test image, use MASM to find eleven landmarks (as discussed in section III) and crop six regions corresponding to left-cheek, right-cheek, beard, left-moustache, middle-moustache, and right-moustache from around these landmarks.

Step 2b: For the left-cheek and right-cheek regions; divide the regions into overlapping or semi-overlapping patches of size $w \times w$ and represent each patch by its HOG feature vector. These feature vectors from cheek regions represent the

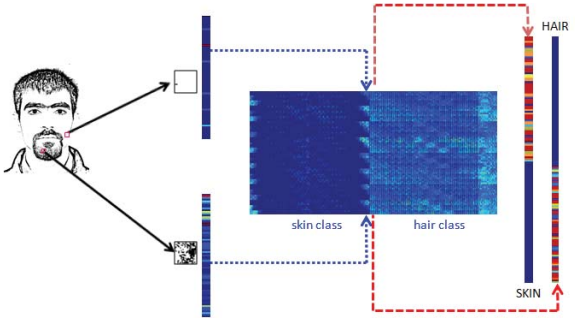


Fig. 3. Binary sparse classifier (with hair and non hair classes) used in our proposed approach.

skin class class and are different for each individual image (for this reason, it is represented dynamically for the test image as well). This approach takes care of illumination artifacts and provides a customizable model that learns the skin texture model of each person, rather than trying to model the skin of every person in the world. This is much harder than modeling hair. The dynamic dictionary (**V**) is formed by appending the hair class samples from Step 1 to the skin class samples. **Step 2c**: To classify the remaining (beard, left-moustache, middle-moustache, right-moustache) regions as either skin or facial hair regions, we firstly split the region into overlapping or semi-overlapping patches. Each patch is represented by a coefficient vector by projecting it onto the dictionary **V** using (7). As shown in Fig. 3, the patch is then classified as either containing hair or skin, based on the number of non-zero values in the coefficient vector. Our detection system produces highly accurate results and even detects remnants of beards and moustaches on nearly clean-shaven faces. We demonstrate the effectiveness of our detection algorithm on widely varying images from three different databases and its ability to generalize across them in section VI.

V. Our Level Set Based Method for Beard/Moustache Segmentation

Level set based active contours are superior to other segmentation methods due to the sub-pixel accuracy that they achieve during the fitting process. There are two types of active contour methods: global active contour methods and local methods. Global segmentation techniques use statistical information to guide the segmentation. Chan and Vese [21] is one of the best known global active contour technique that uses level sets. Global active contours are guided by a global force, which makes them robust to presence of noise. Additionally, they are more efficient in segmenting discontinuous regions which have weaker edges. Moreover, their global attribute allows these methods to be less sensitive to the localization of the initial contour in the image.

Despite their efficiency and robustness in segmenting uniform intensity regions in images, global segmentation techniques have difficulties with non-uniform intensity regions. This can be explained by the fact that global techniques are based on the hypothesis that each region must have homogeneous intensities. In such cases, over-segmentation problems can occur. On the other hand, local segmentation techniques are guided by local image information. Local active contour based segmentation methods show accurate results in segmenting inhomogeneous intensity region based images. However, due to their local specificity, those techniques are sensitive to noise. Moreover, they can be affected by local minima, particularly when the initialization contour location is far away from the structure of interest. A good example of a local active contour is described by Li et al. [22]. They proposed a region-based active contour model using a level set formulation based on region scalable forces using a Gaussian kernel.

In this paper, we propose to make use of the advantages of both global and local active contours and we define the energy function using four terms: (1) a local fitting term, (2) a global fitting term, (3) a contour term, and (4) a regularization term as shown in (9).

$$\begin{aligned} F(\phi, f_1, f_2, c_1, c_2) = {} & (1-\omega)F^{LIF}(\phi, f_1, f_2) \\ & + \omega F^{GIF}(\phi, c_1, c_2) + \mu P(\phi) \\ & + \nu \int_{\Omega} \delta(\phi)|\nabla\phi| \mathrm{d}x\mathrm{d}y \end{aligned} \tag{9}$$

In (9), $\omega \in [0, 1]$ is a constant that weighs the global and local forces. When the images has inhomogeneous regions, the value of ω should be small. $\mu > 0$ and $\nu > 0$ are respectively the level set regularization term and the length term of the contour. ϕ is the level set function. In (9), the first term representing the local force of the active contour is expressed using (10), in which f_1 and f_2 are the internal and external forces respectively determined by the local information inside and outside the contour. These forces are weighted by $\lambda_{11} > 0$ and $\lambda_{21} > 0$. F^{LIF} is a region-scalable fitting energy function defined by a Gaussian kernel function K_σ defined by $\frac{1}{2\pi\sigma^2} e^{\frac{-|u|^2}{2\sigma^2}}$, in which the standard deviation σ can be seen as a scale parameter that controls the region-scalability from small neighborhood to the whole image domain.

$$\begin{aligned} F^{LIF}(\phi, f_1, f_2) = {} & \lambda_{11} \int_{\Omega} K_\sigma (f_1 - u_0)^2 H(\phi) \mathrm{d}x\mathrm{d}y \\ & + \lambda_{21} \int_{\Omega} K_\sigma (f_2 - u_0)^2 (1 - H(\phi)) \mathrm{d}x\mathrm{d}y \end{aligned} \tag{10}$$

The second term in (9) represents the global force and it is formulated as shown in (11), in which c_1 and c_2 represent the average intensities inside and outside the contour respectively and $\lambda_{12} > 0$ and $\lambda_{22} > 0$ are the parameters that control the force inside and outside the contour.

$$\begin{aligned} F^{GIF}(\phi, c_1, c_2) = {} & \lambda_{12} \int_{\Omega} (c_1 - u_0)^2 H(\phi) \mathrm{d}x\mathrm{d}y \\ & + \lambda_{22} \int_{\Omega} (c_2 - u_0)^2 (1 - H(\phi)) \mathrm{d}x\mathrm{d}y \end{aligned} \tag{11}$$

To preserve the regularity of the level set function, which is necessary for an accurate computation and a stable evolution of the level set, a regularization term $P(\phi)$ is used as part of the third term in (9) and is defined as shown in (12).

$$P(\phi) = \int_{\Omega} \frac{1}{2} |\nabla\phi - 1|^2 \mathrm{d}x\mathrm{d}y \tag{12}$$

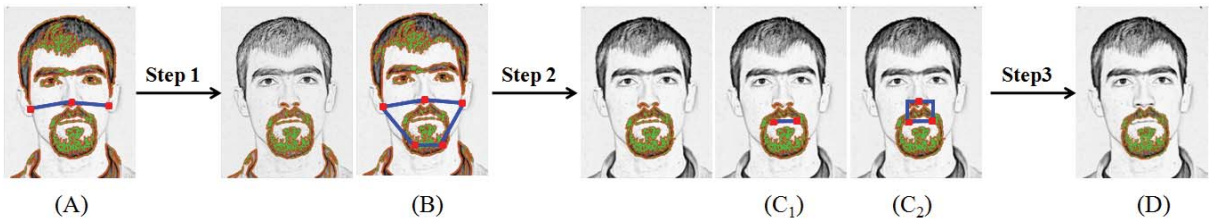


Fig. 4. Post processing used in our segmentation approach. (A) The limit line, (B) the beard and moustache regions, (C_1, C_2) the nostril and mouth respectively, (D) the final segmentation result produced after connected component analysis.

The last term in (9) is used to regularize the zero level set and is thus to derive a smooth contour and defined by $H_\epsilon(\mathbf{x}) = \frac{1}{2}\left(1 + \frac{2}{\pi}\arctan\left(\frac{\mathbf{x}}{\epsilon}\right)\right)$, in which H is the Heaviside function and δ is the Dirac function defined by $\delta_\epsilon(\mathbf{x}) = \frac{1}{\pi}\frac{\epsilon}{\epsilon^2+\mathbf{x}^2}$, in which ϵ is a positive constant.

To minimize the function $F(\phi, f_1, f_2, c_1, c_2)$, standard gradient descent is used. First, ϕ is fixed and $F(\phi, f_1, f_2, c_1, c_2)$ is minimized with respect to $c_1(\phi) = \frac{\int_\Omega u_0 H(\phi)\,\mathrm{d}x\mathrm{d}y}{\int_\Omega H(\phi)\,\mathrm{d}x\mathrm{d}y}$, $c_2(\phi) = \frac{\int_\Omega u_0(1-H(\phi))\,\mathrm{d}x\mathrm{d}y}{\int_\Omega (1-H(\phi))\,\mathrm{d}x\mathrm{d}y}$, $f_1(x) = \frac{K_\sigma * [u_0 H(\phi)]}{K_\sigma * H(\phi)}$, and $f_2(x) = \frac{K_\sigma * [u_0(1-H(\phi))]}{K_\sigma * (1-H(\phi))}$.

In the second step, c_1, c_2, f_1, and f_2 are fixed and $F(\phi, f_1, f_2, c_1, c_2)$ is minimized with respect to ϕ to obtain (13).

$$\frac{\partial\phi}{\partial t} = \delta(\phi)(F^{LIF} + F^{GIF}) + v\delta(\phi)\,div\left(\frac{\nabla\phi}{|\nabla\phi|}\right) + \mu\left(\nabla^2\phi - div\left(\frac{\nabla\phi}{|\nabla\phi|}\right)\right) \tag{13}$$

In (13), F^{GIF} and F^{LIF} are the global and local intensity fitting forces respectively ,and are complementary forces. When the contour is near the boundaries, F^{LIF} is stronger than F^{GIF}. When the contour is far away from the object boundaries F^{GIF} is stronger. Those two forces are expressed using (14) and (15) respectively.

$$F^{LIF} = (1-\omega)[-\lambda_{11}\int K_\sigma(f_1-u_0)^2\,\mathrm{d}x\mathrm{d}y + \lambda_{21}\int K_\sigma(f_2-u_0)^2)\,\mathrm{d}x\mathrm{d}y] \tag{14}$$

$$F^{GIF} = \omega\left(-\lambda_{12}(c_1-u_0)^2 + \lambda_{22}(c_2-u_0)^2\right) \tag{15}$$

In our proposed segmentation method, the initial mask is defined by what we refer to as a skin image. We use an ad hoc skin color detection model in our work. Chai and Ngan [23] determined that the Cb and Cr images of the YCbCr color space are the useful in distinguishing skin color and that in these images values in the range $min_{Cb} \leq Cb \leq max_{Cb}$ and $min_{Cr} \leq Cr \leq max_{Cr}$ could be used to differentiate skin regions from other regions. The values of the parameters used in their work were: $min_{Cb} = 77$, $max_{Cb} = 127$, $min_{Cr} = 133$, and $max_{Cr} = 173$.

Given an input image, we first process it to obtain its MSQ representation. Our proposed active contour using level sets is then applied on the MSQ image with the initial contour determined by using the skin image. Besides beard and moustaches, other facial hair such as present in eyebrows, are also present inside the segmented contour. To remove (prune out) these segmentation results, we perform a post processing step based on the connect-component analysis with three observations:

1) Beards and moustaches (if present) are usually found below the limit line, which is decided by three landmarks as illustrated in Fig. 4(A)
2) If a beard and moustache exist, they touch or connect to the beard/moustache region decided by five landmarks shown in Fig.4(B)
3) The beard/moustache region also contains the mouth and nostril, besides facial hair, as shown in Fig.4(C_1) and Fig.4(C_2)

Thus, post processing is performed using the following three steps:

Step 1: Eliminate parts inside the contour and above the limit line from the segmented image.

Step 2: Extract all the connected components (CC). If a CC has no connection to the beard/moustache region, take it out in accordance with the previously stated second observation.

Step 3a: The CC placed between the two mouth landmarks (i.e. that connects these two landmarks), as shown in (Fig.4(C_1)), is removed in four cases:

1) If neither a beard nor a mustache is detected
2) If a moustache exists but a beard does not and this CC does not touch the moustache region
3) If beard exists but a moustache does not and this CC does not touch the beard region
4) If both a moustache and a beard are detected and this CC does not touch either the beard or moustache regions

Step 3b: The CC connected to or in the nose-mouth region (Fig.4 (C_2)), is removed in two cases:

1) If neither a beard nor a mustache is detected
2) If a moustache exists and the CC does not touch the moustache region

Applying the above rules in conjunction with our proposed segmentation method produces highly accurate results on images across varying databases, as we will show in section VI.

VI. Experiments and Results

Our proposed beard/moustache segmentation system combines the steps we have described in sections II - V. MASM

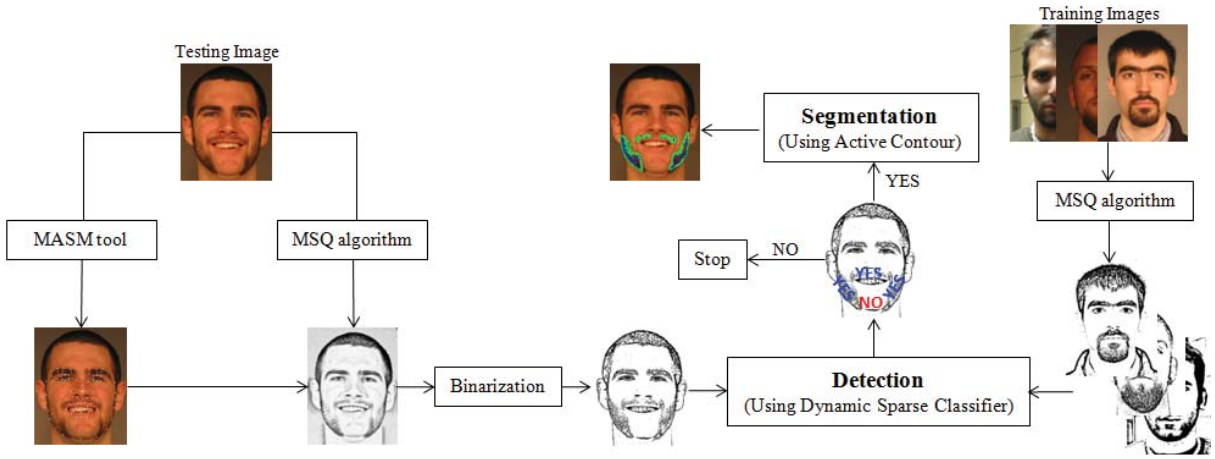


Fig. 5. Flowchart of our proposed beard/mustache detection and segmentation system, called SparCLeS.

was trained on images across several poses obtained from Session 1 of the Multi-PIE (MPIE) database [24]. A commercial face detector is used to initialize MASM so that it can automatically localize 79 key facial landmarks in a given facial image. From these 79 points, 11 key points (that are shown in Fig. 2) are extracted. Next, the MSQ image is obtained using Algorithm 1 and binarized using Otsu's thresholding. Six crop regions corresponding to the mustache, left beard, middle beard, right beard, and cheeks are then determined using the 11 key landmarks. HOG feature vectors drawn from these regions are used to set up our dynamic dictionary and binary classification scheme. If a beard or a moustache is found, the proposed energy function defined in (9) is used to segment the beard and moustache. A flowchart of our proposed beard/moustache detection and segmentation system is shown in Fig. 5.

The rest of this section provides details on the performance (using various metrics) of our system (SparCLeS) on images drawn from varying databases. We evaluate SparCLeS both qualitatively and quantitatively and demonstrate its effectiveness in accurately segmenting facial hair in several challenging images.

A. Datasets and Performance Metrics

We conducted our experiments on subsets of large and diverse datasets to evaluate our beard/mustache detection and segmentation system. The databases we tested on were the NIST Multiple Biometric Grand Challenge - 2008 (MBGC) still face challenge database, the NIST Color Facial Recognition Technology (Color FERET) database, and the Labeled Faces in the Wild (LFW) database. The MBGC database contains 34,729 still frontal images of 810 subjects with varying facial expressions and illumination conditions. The color FERET database consists of 2,413 facial images of 856 subjects with varying facial expressions, illumination, ages, and eyeglass wear. On the MBGC database, our experimental results were conducted on three subsets which contained 8,000 indoor controlled images, 2,000 indoor uncontrolled, and 2,500 outdoor uncontrolled images respectively. Each subset contained 810 subjects with multiple images per subject.

TABLE I

DETAILS ON THE IMAGES FROM THE MBGC (MBGC[1], MBGC[2], MBGC[3]), FERET, AND LFW DATASETS USED IN OUR EXPERIMENTS. NUMBER OF IMAGES WITH (1) BEARD ONLY, (2) MOUSTACHE ONLY, (3) BOTH A BEARD AND A MOUSTACHE, (4) NO BEARD OR A MOUSTACHE (MALE), (5) NO BEARD OR A MOUSTACHE (FEMALE)

	MBGC[1] 8 000 Indoor Controlled	MBGC[2] 2 000 Indoor Uncontrolled	MBGC[3] 2 500 Outdoor Uncontrolled	FERET 989 Images	LFW 335 Images
(1)	251	109	88	5	1
(2)	165	97	60	76	38
(3)	496	174	59	70	23
(4)	3088	720	893	440	188
(5)	4000	900	1400	398	85

TABLE II

DETAILS ON THE NUMBER OF GROUND TRUTH IMAGES USED TO EVALUATE OUR SEGMENTATION ALGORITHM ACROSS THREE DATASETS.

Dataset	Beard Only	Moustache Only	Beard & Moustache	Total
MBGC	100	70	80	250
FERET	0	25	15	40
LFW	0	13	7	20

We also tested our approach on 989 images of 989 subjects from the color FERET database and 335 images of 177 subjects from LFW database to prove that it generalizes across databases and that it can handle challenging illumination conditions from real world images that are present in the LFW database.

Each dataset was as a collection of four types of images: those with facial images containing only a beard, only a moustache, both a beard and a moustache, and neither a beard nor a moustache. The details and the distribution of images that we used in our experiments are provided in Table I.

To evaluate our segmentation algorithm, we manually labeled 310 images from the MBGC, FERET and LFW databases with ground truths. The number of ground truth images corresponding to each dataset are listed in Table II.

TABLE III

BEARD AND MOUSTACHE DETECTION ACCURACY ON (1) 8000 CONTROLLED INDOOR IMAGES OF 810 SUBJECTS FROM THE MBGC DATABASE, (2) 2000 UNCONTROLLED INDOOR IMAGES OF 810 SUBJECTS FROM THE MBGC DATABASE, (3) 2500 UNCONTROLLED OUTDOOR IMAGES OF 810 SUBJECTS FROM THE MBGC DATABASE, (4) 989 IMAGES OF 989 SUBJECTS FROM THE FERET DATABASE, AND (5) 335 IMAGES OF 177 SUBJECTS FROM THE LFW DATABASE

	Beard Labeling Accuracy (in %)					Moustache Labeling Accuracy (in %)				
	Beard Only	Moustache Only	Both Beard and Moustache	No Beard or Moustache Female	No Beard or Moustache Male	Beard Only	Moustache Only	Both Beard and Moustache	No Beard or Moustache Female	No Beard or Moustache Male
(1)	97.6	94.0	95.3	98.4	98.0	93.2	89.0	91.1	98.4	97.7
(2)	90.8	91.7	90.8	92.0	91.5	86.2	88.7	83.3	89.4	88.5
(3)	85.2	85.0	83.0	90.4	90.9	80.7	76.7	79.7	88.1	86.8
(4)	100	94.7	91.4	97.7	97.0	100	90.8	88.6	94.7	95.2
(5)	0	78.9	82.6	81.1	79.3	100	76.3	73.9	75.3	77.1

Segmentation performance is separately considered for three types of images, faces with only beards, moustaches only, and both beards and moustaches. It is to be noted that for the FERET and LFW datasets, there are only two categories used - only moustache present and beard and moustache present, as there were no images in these datasets with only beards present. We consider the segmentation problem as a binary classification one and use eight performance metrics, (1) F-Measure (FM), (2) Peak Signal to Noise Ratio (PSNR), (3) Negative Rate Metric (NRM), (4) Misclassification Penalty Metric (MPM), (5) Sensitivity (Sen), (6) Specificity(Spec), (7) Balanced Classification Rate (BCR), and (8) $_\beta$F Measure ($_\beta$FM) (as detailed by Sokolova and Lapalme in [25]), to evaluate our proposed beard/moustache segmentation algorithm. Sokolova and Lapalme showed that the correctness of a classification can be evaluated by: (1) computing the number of correctly recognized class samples (true positives - TP), (2) calculating the number of correctly recognized samples that do not belong to the class (true negatives - TN), and (3) finding the samples that are either incorrectly assigned to the class (false positives - FP) or that are not recognized as class samples (false negatives - FN).

1) **F-measure (FM)**: Evaluates how well an algorithm can retrieve the desired pixels and is defined by $\text{FM} = \frac{2\times\text{precision}\times\text{recall}}{\text{precision}+\text{recall}}$, in which the precision is a positive value calculated as $\text{precision} = \frac{\text{TP}}{\text{TP+FP}}$ and the recall is is the proportion of actual positives that are predicted to be positive and is determined as $\text{recall} = \frac{\text{TP}}{\text{TP+FN}}$.
2) **Peak Signal to Noise Ratio (PSNR)**: It is a measure of how close an image (I) of size $M \times N$ is to another (I') and is defined by $\text{PSNR} = 10\log\frac{C^2}{\text{MSE}}$, in which C is a constant that denotes the difference between the foreground and the background and is set equal to 1 in our work, and MSE is the mean square error, defined by $\text{MSE} = \frac{\sum_{i=0}^{M-1}\sum_{j=0}^{N-1}(I(i,j)-I'(i,j))^2}{M\times N}$.
3) **Negative Rate Metric (NRM)**: It is also called the Balanced Error Rate and is based on the pixel-wise mismatches between the ground truth image and the original one and defined by $\text{NRM} = \frac{1}{2}\frac{\text{FN}}{\text{TP+FN}}\frac{\text{FP}}{\text{TN+FP}}$.
4) **Misclassification Penalty Metric (MPM)**: It is a measure of how well the resulting image represents the contour of the ground truth image and is defined by $\text{MPM} = \frac{1}{2D}(\sum_i^{\text{FN}} d_{\text{FN}}^i + \sum_j^{\text{FN}} d_{\text{FP}}^j)$, in which d_{FN}^i and d_{FP}^j respectively denote the distance of the i^{th} false negative and the j^{th} false positive pixel from the contour of the text in the ground truth image. The factor D is the sum of all the pixel to contour distances of the ground truth object. A low MPM indicates that the algorithm is good at detecting the boundary.
5) **Sensitivity (Sens)**: It is the proportion of actual positives which are predicted to be positive and is defined by $\text{Sens} = \frac{\text{TP}}{\text{TP+FN}}$.
6) **Specificity (Spec)**: It is the proportion of actual negatives which are predicted as negative and is given by $\text{Spec} = \frac{\text{TN}}{\text{TPN+FP}}$.
7) **Balanced Classification Rate (BCR)**: Gives balanced assessments on the two classes that have to be adopted and is given by $\text{BCR} = \frac{\text{Sens+Spec}}{2}$.
8) **$_\beta$F Measure ($_\beta$FM)**: Is the weighted harmonic mean between sensitivity and specificity and is given by $_\beta\text{FM} = \frac{2\times\text{Sens}\times\text{Spec}}{\text{Sens+Spec}}$.

In contrast to FM, $_\beta$FM, PSNR, and BCR, the segmentation accuracy is higher for lower values of NRM and MPM.

B. Experimental Results

The performance of our proposed algorithm on beard/moustache detection and segmentation was evaluated on the five datasets mentioned in Table I. Detection of beards and moustaches was accomplished using our proposed dynamic sparse classifier. To study the impact of the sparse coefficient ε on beard/moustache detection, we randomly selected 200 images from each dataset and changed the value of ε from 10^{-7} to 0.9. We chose $\varepsilon = 0.1$ for our remaining experiments. The detection accuracies corresponding to four types of images (beard only, moustache only, both beard/moustache, no beard or moustache) of our individual test datasets is reported in Table III while Table IV shows the overall percentage of images in the datasets that are correctly classified as having or not having beards and moustaches.

Sample results of beard and moustache detection on images with different illumination conditions from the MBGC database are shown in Fig. 6(A,B,C). Similarly, detection results

TABLE IV

OVERALL BEARD AND MOUSTACHE DETECTION ACCURACIES ON THE MBGC, FERET, AND LFW DATASETS.

	MBGC[1] 8000 Indoor Controlled	MBGC[2] 2000 Indoor Uncontrolled	MBGC[3] 2500 Outdoor Uncontrolled	FERET 989 Illumination	LFW 335 Low-Dimension
Beard accuracy (in %)	98.0	91.7	90.1	96.8	79.7
Moustache accuracy (in %)	97.4	88.4	86.9	94.2	76.4

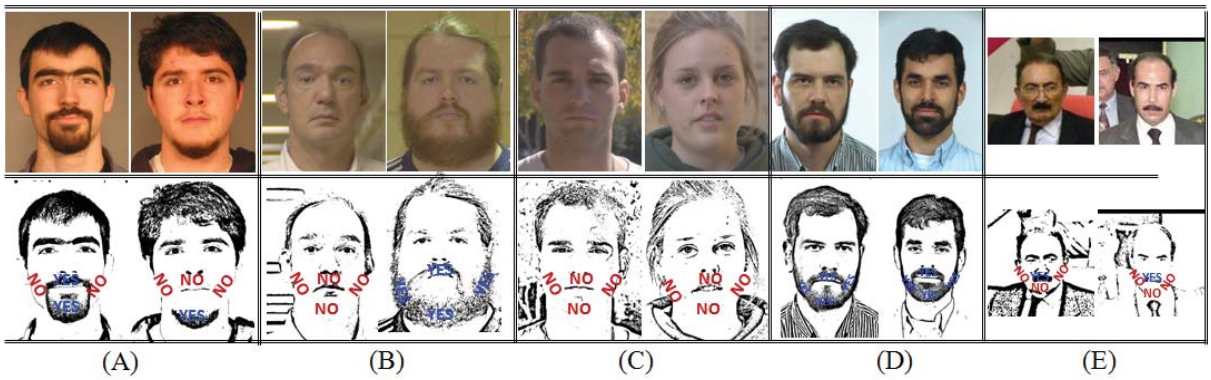


Fig. 6. Sample detection results produced by our proposed dynamic sparse classifier on images from the (A) Indoor and controlled illumination samples from MBGC database, (B) indoor and uncontrolled illumination samples from MBGC database, (C) outdoor and uncontrolled illumination samples from MBGC database, (D) FERET database, (E) LFW database.

TABLE V

RESULTS OF EVALUATING OUR SEGMENTATION APPROACH USING EIGHT PERFORMANCE METRICS

		FM	PSNR	NRM ($\times 10^{-2}$)	MPM($\times 10^{-3}$)	Sens	Spec	BCR	$_{\beta}$FM
MBGC	Beard	99.2	22.4	14	0.4	0.99	0.73	0.86	81.8
MBGC	Moustache	99.4	21.6	8	0.22	0.99	0.85	0.92	91
MBGC	Beard and moustache	97.7	17.8	15	1.6	0.98	0.73	0.85	83
FERET	Moustache	99.7	23.1	5.2	0.086	0.99	0.90	0.95	94.3
FERET	Beard and moustache	98.4	18.9	8.3	0.88	0.98	0.85	0.91	91.1
LFW	Moustache	99.7	22.7	7.9	0.09	0.99	0.84	0.92	91.3
LFW	Beard and moustache	98.7	20.3	12.4	0.57	0.98	0.76	0.87	86.1

obtained on some images from the FERET and LFW databases are shown in Fig. 6(D, E). In both figures, the top shows the original images while the lower row shows the MSQ images with the sparse classifier results superimposed. As can be seen from 6, our detection system is quite accurate even when uncontrolled illumination is present because of the use of the illumination invariant MSQ representation.

The cases when our detection algorithm fails can be divided into two categories:

1) Errors caused by nearly clean-shaven faces: Sometimes our classifier detects the presence of moustaches and beards in faces that at first glance, appear to be clean-shaven. This is because of the presence of light stubbles etc. Ground truths for these images differ from our algorithm detection results in such cases due to some inconsistencies in the ground truth creation due to the subjectivity involved in assessing the presence or absence of facial hair in some cases. In such cases, detection results produced by our system disagree with the ground truths. It is important to note that in such cases, our approach provides a way of determining where a beard or moustache is likely to grow on the face of the person in question and this can serve as a valuable soft biometric.
2) Errors caused by very strong cast-shadows: Our proposed beard/moustache detection system sometimes fails in the presence of strong shadows and illumination artifacts.

Fig. 7, show segmentation results produced by our approach on some images from the MGBC, FERET, and LFW databases respectively. These are purely qualitative results that visually show the effectiveness of our segmentation approach. Quantitative results produced using each of the eight performance measures on 310 images from the MGBC, FERET, and LFW databases are reported in Table V. The segmentation accuracies (low values of NRM and MPM together

Fig. 7. Segmentation results produced by SparCLeS on images from the MBGC, FERET, and LFW dataset. The first row shows the original images. The second row shows the corresponding beard/moustache segmentation results.

with high values of FM, $_{\beta}$FM, PSNR, and BCR) obtained on images with only moustaches present were higher than those obtained on images with both beards and moustaches present. All these results demonstrate the effectiveness of our segmentation algorithm and the overall ability of SparCLeS to generalize across images from varying databases and still accurately detect and segment beards and moustaches in each of them.

VII. Conclusion

We have proposed a novel and completely automatic beard and moustache detection and segmentation system called SparCLeS. SparCLeS first automatically localizes a set of facial landmarks using the Modified Active Shape Model (MASM) algorithm and subsequently extracts HOG features from regions in the binarized Multiscale Self-Quotient (MSQ) image. The use of the MSQ algorithm facilitates the removal of illumination artifacts, leading to more robust performance. Beard and moustache detection is carried out using a dynamic sparse classifier on HOG features while segmentation of facial hair is accomplished using a level set based approach. SparCLeS was qualitatively and quantitatively evaluated on the NIST MBGC, NIST Color FERET, and LFW databases and was found to obtain highly accurate results even on challenging images with widely varying illumination and environmental conditions drawn from these databases. To the best of our knowledge, our work is the first that describes the development and thorough evaluation of an automatic beard/moustache detection and segmentation system and it is our hope that it contributes to the advancement and benchmarking of future algorithms that aim at accomplishing this challenging task.

Since the dictionary is built using high resolution images, the performance of our proposed system on low quality images, such as those from the LFW database, is not as good as the performance on high quality ones, such as those from the MBGC or FERET databases. We are currently investigating how to improve the performance and generalization of the system in order to better deal with high/low resolution images, varying illumination conditions etc. Furthermore, we are working towards customizing the code in C++ and also improving the system in order to obtain faster performance and possibly real-time performance.

Acknowledgment

The authors would like to thank all members of the CMU Biometrics Center for help with different modules and numerous valuable discussions. They would also like to thank Mr. K. Seshadri for helping organizing and proofreading the paper.

References

[1] S. Y. D. Hu, B. Jou, A. Jaech, and M. Savvides, "Fusion of Region-based representations for gender identification," in *Proc. IEEE Int. Joint Conf. Biometrics*, Oct. 2011, pp. 1–7.

[2] G. G. Y. Fu and T. S. Huang, "Age synthesis and estimation via faces: A survey," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 11, pp. 1955–1976, Nov. 2010.

[3] K. Luu, K. Seshadri, M. Savvides, T. D. Bui, and C. Y. Suen, "Contourlet appearance model for facial age estimation," in *Proc. IEEE Int. Joint Conf. Biometrics*, Oct. 2011, pp. 1–8.

[4] M. H. Nguyen, J. F. Lalonde, A. A. Efros, and F. de la Torre, "Image-based shaving," *Comput. Graph. Forum J. (Eurograph. 2008)*, vol. 27, no. 2, pp. 627–635, 2008.

[5] J. S. Pierrard, "Skin segmentation for robust face image analysis," Ph.D. dissertation, Faculty Sci., Univ. Basel, Basel, Switzerland, 2008.

[6] N. Otsu, "A threshold selection method from gray-level histogram," *IEEE Trans. Syst., Man, Cybern.*, vol. 9, no. 1, pp. 62–66, Jan. 1979.

[7] J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry, and Y. Ma, "Robust face recognition via sparse representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 2, pp. 210–227, Feb. 2009.

[8] A. Wagner, J. Wright, A. Ganesh, Z. Zhou, and Y. Ma, "Toward a practical face recognition system: Robust registration and illumination by sparse representation," in *Proc. IEEE Int. Conf. Comput. Vis. Pattern Recognit.*, Jun. 2009, pp. 597–604.

[9] N. Dalal and B. Triggs, "Histogram of oriented gradients for human detection," in *Proc. IEEE Int. Conf. Comput. Vis. Pattern Recognit.*, Jun. 2005, pp. 886–893.

[10] K. Seshadri and M. Savvides, "Robust modified active shape model for automatic facial landmark annotation of frontal faces," in *Proc. 3rd IEEE Int. Conf. Biometrics, Theory, Appl. Syst.*, Sep. 2009, pp. 319–326.

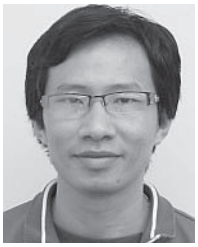
[11] NIST. (2008). *Multiple Biometric Grand Challenge (MBGC)*, Gaithersburg, MD, USA [Online]. Available: http://face.nist.gov/mbgc

[12] P. J. Phillips, P. J. Flynn, J. R. Beveridge, W. T. Scrugs, A. J. O'Toole, D. Bolme, K. W. Bowyer, B. A. Draper, G. H. Givens, Y. M. Lui, H. Sahibzada, A. Joseph, and S. Weimer, "Overview of the multiple biometrics grand challenge," in *Proc. 3rd IAPR/IEEE Int. Conf. Biometrics*, Jun. 2009, pp. 705–714.

[13] NIST. (2003). *The Color FERET Database Version 2*, Gaithersburg, MD, USA [Online]. Available: http://www.nist.gov/itl/iad/ig/colorferet.cfm

[14] G. B. Huang, M. Ramesh, T. Berg, and E. Learned-Miller, "Labeled faces in the wild: A database for studying face recognition in unconstrained environments," Univ. Massachusetts, Amherst, MA, USA, Tech. Rep. 07-49, Oct. 2007.
[15] H. Wang, S. Z. Li, Y. Wang, and J. Zhang, "Self quotient image for face recognition," in *Proc. IEEE Int. Conf. Pattern Recognit.*, Oct. 2004, pp. 1397–1400.
[16] T. F. Cootes, C. J. Taylor, D. H. Cooper, and J. Graham, "Active shape models-their training and application," *Comput. Vis. Image Understand.*, vol. 61, no. 1, pp. 38–59, Jan. 1995.
[17] J. C. Gower, "Generalized procrustes analysis," *Psychometrika*, vol. 40, no. 1, pp. 33–51, Mar. 1975.
[18] S. Milborrow and F. Nicolls, "Locating facial features with an extended active shape model," in *Proc. 10th Eur. Conf. Comput. Vis.*, Oct. 2008, pp. 504–513.
[19] T. F. Cootes, C. J. Taylor, and A. Lanitis, "Active shape models: Evaluation of a multi-resolution method for improving image search," in *Proc. 5th Brit. Mach. Vis. Conf.*, 1994, pp. 327–336.
[20] Y. C. Pati, R. Rezaiifar, and P. S. Krishnaprasad, "Orthogonal matching pursuit: Recursive function approximation with applications to wavelet decomposition," in *Proc. 27th Annu. Asilomar Conf. Signals, Syst., Comput.*, 1993, pp. 1–3.
[21] T. F. Chan and L. A. Vese, "Active contours without edges," *IEEE Trans. Image Process.*, vol. 10, no. 2, pp. 266–277, Feb. 2001.
[22] C. Li, C. Y. Kao, J. C. Gore, and Z. Ding, "Minimization of region-scalable fitting energy for image segmentation," *IEEE Trans. Image Process.*, vol. 17, no. 10, pp. 1940–1949, Oct. 2008.
[23] D. Chai and K. N. Ngan, "Face segmentation using skin colour map in videophone aplications," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 9, no. 4, pp. 551–564, Jun. 1999.
[24] R. Gross, I. Matthews, J. Cohn, T. Kanade, and S. Baker, "Multi-PIE," *Image Vis. Comput.*, vol. 28, no. 5, pp. 807–813, May 2010.
[25] M. Sokolova and G. Lapalme, "A systematic analysis of performance measures for classification tasks," *J. Inf. Process. Manage.*, vol. 45, no. 4, pp. 427–437, Jul. 2009.

T. Hoang Ngan Le (S'10) received the Bachelors and Masters degrees from the Department of Computer Sciences, Faculty of Information Technology, University of Natural Sciences, HCM City, Vietnam National University, Ho Chi Minh City, Vietnam, in 2005 and 2009, respectively. He is currently pursuing the Ph.D. degree in electrical and computer engineering with Carnegie Mellon University (CMU), Pittsburgh, PA, USA, and a Research Assistant in CMU CyLab Biometrics Center since 2011. She was a Ph.D. student and Research Assistant with the Centre for Pattern Recognition and Machine Intelligence (CENPARMI), Concordia University, Montreal, ON, Canada, from May 2010 to September 2011. Her current research interests include image segmentation, facial recognition, classification, sparse representation, and compressed sensing.



Khoa Luu (SM'08) received the Bachelors degree from the Department of Computer Sciences, Faculty of Information Technology, University of Natural Sciences, HCM City, Vietnam National University, Ho Chi Minh City, Vietnam, in 2005. He has been a Post-Doctoral Research Scientist with the CMU Cylab Biometrics Center since April 2011. He was a Doctoral Student and Research Assistant with the Centre for Pattern Recognition and Machine Intelligence (CENPARMI), Concordia University, Montreal, ON, Canada, from January 2008 to March 2011. His current research interests include biometrics, image processing, multifactor analysis, compressed sensing and machine learning.

Marios Savvides (M'00) is the Founder and Director of the Biometrics Center at Carnegie Mellon University, Pittsburgh, PA, USA, and is an Associate Research Professor with the Electrical & Computer Engineering Department and CMU CyLab. He is also one of the tapped researchers to form the Office of the Director of National Intelligence (ODNI) 1st Center of Academic Excellence in Science and Technology. His current research interests include developing algorithms for robust face and iris biometrics as well as pattern recognition, machine vision and computer image understanding for enhancing biometric systems performance. He is on the program committee on several Biometric conferences, such as IEEE BTAS, ICPR, SPIE Biometric Identification, IEEE AutoID and others and has also organized and co-chaired Robust Biometrics Understanding the Science & Technology (ROBUST 2008) conference. He is an annually invited speaker at IDGA's main conference on Biometrics for National Security and Defense (now IDGA Biometrics for National Security and Law Enforcement). He has authored and co-authored over 170 journal and conference publications, including several book chapters in biometrics and served as the area editor of the *Springer's Encyclopedia of Biometrics*. He helped develop the IEEE Certified Biometrics Professional (CBP) program and served on the main steering committee of the IEEE CBP program. His achievements include leading the R&D in CMU's past participation at NIST's Face Recognition Grand Challenge 2005 and also in NIST's Iris Challenge Evaluation. He is listed in *Marquis Who's Who in America* and in *Marquis' Who's Who in Science & Engineering*. He has filed over 20 patent applications in biometrics and was the recipient of the CMU's 2009 Carnegie Institute of Technology (CIT) Outstanding Faculty Research Award.