



HAL
open science

Automatic landmark location with a combined active shape model

Dianle Zhou, Dijana Petrovska-Delacrétaz, Bernadette Dorizzi

► **To cite this version:**

Dianle Zhou, Dijana Petrovska-Delacrétaz, Bernadette Dorizzi. Automatic landmark location with a combined active shape model. BTAS 2009: IEEE 3rd International Conference on Biometrics: Theory, Applications and Systems, Sep 2009, Washington, Dc, United States. 10.1109/BTAS.2009.5339037 . hal-01360767

HAL Id: hal-01360767

<https://hal.science/hal-01360767>

Submitted on 6 Sep 2016

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Automatic Landmark Location with a Combined Active Shape Model

Dianle Zhou, Dijana Petrovska-Delacrétaz and Bernadette Dorizzi

Abstract—Automatic facial landmark location is a difficult challenge for realistic face recognition applications, where the face is recorded under variable illumination conditions including indoor and outdoor recordings and also with some pose and scale variability. Moreover, the image distortion and complex background also bring some difficulty both for landmark location and face recognition.

The proposed landmark detection method, called Combined Active Shape Models, is robust to illumination, translation, and rotation. It exploits the Scale Invariant Feature Transform (SIFT) [1] and the Active Shape Model (ASM) [2]. In order to have a better representation of face images, the landmarks on the face region and the face contour are modeled and processed separately. The performance of the proposed Combined-ASM algorithm is tested on the BioID and FRGCv2.0 face image databases.

I. INTRODUCTION

Finding the correct position of facial landmarks (key-points) is a crucial step for many face processing algorithms such as face recognition, modeling or tracking. It is also needed for a variety of statistical approaches in which a model is built from a set of labeled examples. Many 2D face recognition algorithms depend on a careful normalization, by the location of landmarks such as eyes and mouth centers, that is necessary for the global feature extraction step. With more reliable and more precise landmarks better face recognition performance is obtained. The number and position of these landmarks are not unique and depend on applications and algorithms. For 2D face recognition with global methods, usually eyes centers and nose and mouth positions are used. Those landmarks can be labeled by hand, but for realistic applications it is necessary to have automated methods. Due to the variety of human faces and their variability related to expressions, pose, accessories, or lighting and acquisition conditions, fully automatic landmark localization remains an unsolved problem.

This paper focuses on automatic facial landmark location for face recognition, in situations where mainly illumination, pose and scale variabilities are present. We are mostly interested in automatically detecting the two eyes and mouth centers which are used for our face normalization step for our face recognition system [3].

The rest of the paper is organized as follows: a brief literature review about facial landmark location is given in Section II. A reminder of the original Active Shape Model (ASM) and Scale Invariant Feature Transform (SIFT), on which our proposed combined model is based on, are given in Section III. The proposed combined model is explained in Section IV. The databases and experimental protocols used for the training and evaluation phases are presented in Section V, whereas the results are reported in Section VI. Finally, the conclusions and perspectives can be found in Section VII.

II. BRIEF LITERATURE REVIEW ABOUT AUTOMATIC LANDMARK LOCATION

A lot of algorithms have been proposed for facial landmark location. As suggested by Hamouz et al. [4], they can be classified in two categories: image-based and structure-based methods.

In **image-based methods**, faces are treated as vectors in a large space and these vectors are furthermore transformed. The most popular transformations are Principal Components Analysis, Gabor Wavelets [5], [6], [7], [8], Independent Components Analysis [9], Discrete Cosine Transform [7], and Gaussian derivative filters [5], [10]. Through these transforms, the variability in facial features is captured, and machine learning approaches like boosted cascade detectors [11], [8], Support Vector Machines [9], [12] and Multi-layer Perceptions are used to learn the appearance of each landmark. Some examples of such methods are proposed by Viola and Jones [11], Jesorsky et al. [13], and Hamouz et al. [4].

Structure-based methods use prior knowledge about facial landmark positions, and constrain the landmark search by heuristic rules that involve angles, distances, and areas. The face is represented by a complete model of appearance consisting of points and arcs connecting these points [14]. For each point of this model, a description of these features is associated. Typical methods include Active Shape Models (ASM) [2], [15], [12], Active Appearance models (AAM) [16] and Elastic Bunch Graph Matching [17], [18]. These methods are well suited for precise localization [19]. However, few

extensive evaluation results have been published on face recognition experiments.

III. REMINDER ABOUT ASM AND SIFT

A. Original Active Shape Model (ASM)

In the original Active Shape Model (ASM) [2] there are two statistical models that exploit the global shape and the local texture prior knowledge in the segmentation process. The Point Distribution Model (PDM) [2] represents the mean geometry of a shape and its statistical variations from the training set of shapes. While the Local Texture Models (LTM) [2] are used to describe the texture variations at each landmark position of the Point Distribution Model.

1) *Point Distribution Model*: The Point Distribution Model is constructed by applying Principal Component Analysis (PCA) to the aligned set of shapes, which are presented by landmarks on the training face database. Assume there are N training images, so the i^{th} shape S_i and the model parameters P_i in the shape space can be represented as follows:

$$P_i = \Phi^T (S_i - \bar{S}), S_i = \bar{S} + \Phi P_i, \quad (1)$$

where \bar{S} is the mean shape and Φ is the eigenvector matrix. Briefly, the Point Distribution Model describes heuristic rules of the face shape. During the fitting, this model helps in the interpretation of noisy and low-contrasted pixels.

2) *Local Texture Models*: A typical image structure that describes the local texture around each landmark is the Grey-Level Profile (GLP) [2], calculated from the fixed-length pixels sampled around each landmark. The direction of the profile is perpendicular to the contour. The first derivative of the profile is calculated and used as the feature vector. Those vectors are extracted from all the training images, and represent the normalized derivatives profiles, denoted as g_1, g_2, \dots, g_N . The mean profile \bar{g} and the covariance matrix C_g are computed for each landmark. The Mahalanobis distance measure is used to compute the difference between the new profile and the mean profile \bar{g} , defined as follows:

$$f(g_{new}) = (g_{new} - \bar{g})C_g^{-1}(g_{new} - \bar{g})^T. \quad (2)$$

Some previous work exist that try to choose better Local Texture Models [15], [19]. For example, Ordas et al. [15] replace the 1D normalized first derivative profiles of the original ASM with local texture descriptors calculated from “locally orderless images”, for reliable segmentation for cardiac Magnetic Resonance data. In [19], the authors use the 2D profile in the square region around the landmark for a more precise

fitting result. Actually different Local Texture Models are adapted to different conditions. In our application, we use the SIFT feature descriptor, which is robust to degraded conditions, such as illumination or small pose variations.

B. Scale Invariant Feature Transform (SIFT) Features

In 2004, David Lowe presented a method to extract distinctive invariant features from images [1]. He named them Scale Invariant Feature Transform (SIFT). The process consists of four major stages. In the first stage, potential interest points are identified by scanning the image over location and scale by constructing a Gaussian pyramid and searching for local peaks (termed keypoints). In the second stage, candidate keypoints are localized to sub-pixel accuracy and eliminated if found to be unstable. The third step identifies the dominant orientations for each keypoint based on its local image patch. Finally, local image descriptors are built for each keypoint. Local gradient data is used to create keypoint descriptors. The gradient information is rotated to line up with the orientation of the keypoint and then weighted by a Gaussian with variable scale. This data is then used to create a set of histograms over a window centered on the keypoint. In our application we use this SIFT local image descriptor as our Local Texture model.

IV. A COMBINED FACE MODEL FOR AUTOMATIC LANDMARK LOCATION

A. Advantages of the SIFT Feature Descriptor

Using SIFT features for object matching is very popular, and seems to be a reliable choice for solving the problem of illumination and pose variability. Since it is based on the local gradient histograms around the landmark, the SIFT descriptor is highly distinctive and partially invariant to variations, like illumination or 3D view point. In our application, we use the SIFT descriptor to replace the Grey-Level profiles. In order to make the ASM shape model rotation invariant, the gradient orientations of the descriptor are always computed relative to the edge normal vector at the landmark point which could be obtained by interpolation of neighboring landmarks, as depicted in Fig.1.

There are a two main advantages of the SIFT feature descriptor. The first advantage is that SIFT descriptors encode the internal gradient information of a patch around the landmark, thus capturing essential spatial position and edge orientation information of the landmark while Grey-level Profile only capture the one-dimensional pixel information that is perpendicular to the contour. Though the Mahalanobis distance measure assumes a normal multivariate unimodal distribution of

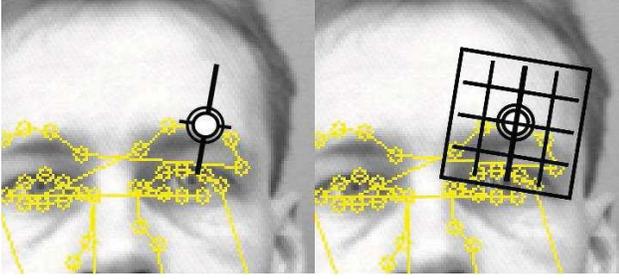


Fig. 1. Left: The Grey-Level Profile (GLP) is extracted from the neighborhood pixels perpendicular to the contour. Right: The SIFT descriptor is computed over a patch along the normal vector at the landmark (the original image is from the BioID database [13])

Grey-level Profile, in practice, they can be any statistical distribution. The SIFT descriptors have a more discriminative likelihood model which is distinctive enough to differentiate between landmarks.

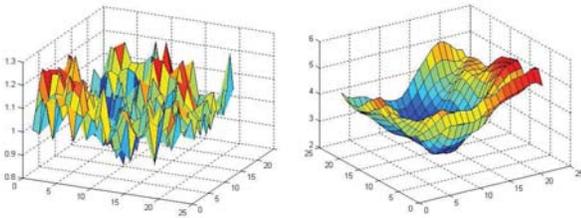


Fig. 2. Left: Gradient profile matching cost of the landmark highlighting in Fig. 1 over a window of size 21×21 . Notice the multiple minima resulting in poor alignment of shapes. Right: SIFT descriptor matching cost for the same landmark point

In Fig. 2, we calculated the Mahalanobis distance of the neighborhood points over a 21×21 window around the landmark highlighted in Fig. 1. The SIFT descriptor has a unambiguous minimal point in the center of the neighborhood region. Also the SIFT descriptors are invariant to affine changes in illumination and contrast by quantizing the gradient orientations into discrete values in small spatial cells and normalizing these distributions over local blocks. Such features are important in challenging real-life situations presenting illumination variabilities.

The second advantage of the SIFT descriptors is that they are more stable to changes that occur due to changes of pose, that can occur when dealing with faces. As shown in Fig. 3, when using SIFT features, there is a correspondence of the detected landmarks, even when the position of the face is nonfrontal.



Fig. 3. Example showing the robustness of SIFT features to pose variation. When in the ASM model we replace the the original Grey-level Profile with SIFT features, the corresponding landmark points are detecting correctly, despite the pose variation. The original image is from the IMM database [20])

B. Combined Active Shape Model Based on Facial Internal Region Model and Facial Contour Model

As shown above, using the SIFT features, we can find correspondences between landmarks in two images that have small pose variability, even when the landmarks used to train the ASM are in 2D. The points in the face region that we denote as “internal” (such as eyes’ corners), could be considered as the perspective projection of the 3D face on the image plan. While the contour points are different, and are more dependent on the 3D view point. In that case the SIFT descriptor doesn’t work when the acquisition angle of testing images is different from the training images.

So in our proposed approach, two models are used to represent the human face. One of the model represents the landmarks of what we call “internal region”, including the landmarks on the eyes, nose, eyebrows and mouth. Those points could be considered as 3D position invariant during the perspective projection. So we use the SIFT descriptor for this model, and we name it **facial internal region model**. The other one models the contour point on the face only. For those points using SIFT representations will result in wrong matches. The gradient of the profile is more suited for the contour points, so we use Grey-Level Profile to describe them, and we name it **facial contour model**.

The facial internal region model is represented with 45 points, while 13 points are used for the facial contour model, as depicted in Fig. 4. Each of them has its own shape model and shape variability. In order to cope with the out-plan rotation case, we also use some training images which contain faces with some pose variability. In the fitting step, for each iteration the two models are matched to the face image separately. After matching,

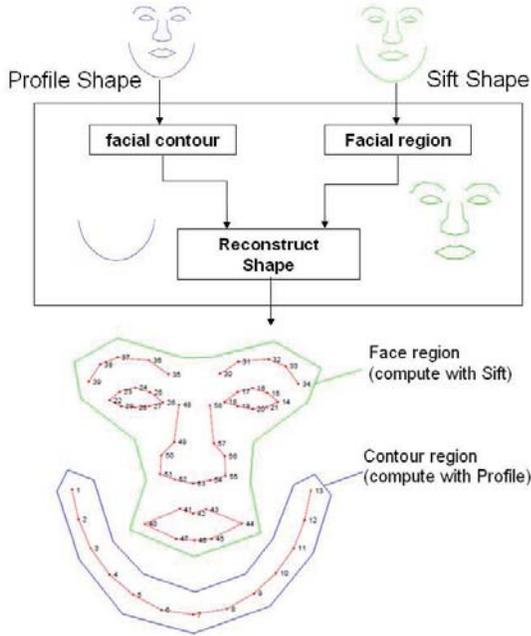


Fig. 4. Combined landmark detection model: 45 landmarks define the facial internal region model (represented with SIFT features) and 13 landmarks define the facial contour model (represented with Grey-Level Profile features)

we combine them into a new face model and use the Point Distribution Model to constrain it to a plausible shape in the shape space, as shown in Fig. 4. This is repeated iteration by iteration at each resolution until convergence is reached.

Fig. 5 (bottom row), shows some examples of the fitting results of the proposed method, denoted as Combined Active Shape Model (Combined-ASM). While in the original Active Shape Model the texture is represented only by Gray-level Profile features, we propose to use two ASM models with SIFT and Gray-level Profile features. For comparison purposes, we show on the bottom row the results of the automatic landmark detection with the original ASM model (using only profile features), the ASM model that is based on SIFT features (middle row), and the result of our Combined-ASM. We can observe that the Grey-Level Profile has better performance on the contour points of side-view images, while SIFT features seem to be more adapted for the internal face region points.

V. DATABASES

Different face databases are needed for our experiments. Some of them are needed to train the ASM models, some of them to tune different parameters, and



Fig. 5. Typical fitting result of non-frontal face achieved by original ASM (top row), SIFT-ASM (middle row) and Combined-ASM (bottom row), (the original images are from the IMM database [20])

distinct datasets are necessary for the evaluation of the proposed method.

For the training phase of the Combined-ASM, we have used a subset of IMM database [20], composed of 240 annotated monocular images of 40 different human subjects. And Points of correspondence are placed on each image.

The BANCA database [21] is used to tune the parameters of the Combined-ASM (the size of the square region to calculate the SIFT descriptors and the length of the profile). It is a challenging dataset designed for audio-visual biometric experiments. The subjects were recorded in three different scenarios-controlled, degraded and adverse-over 12 different sessions, with a variable time span between sessions. We used the English subset of this database for tuning purposes.

In order to evaluate our automatic landmark detection algorithms, we have used the BioID [13] and the FRGCv2.0 databases [22]. The BioID database was chosen because there are already published results on that database for facial landmark detection, while the FRGCv2.0 was chosen because it includes a huge number of subjects (around 500) with some variabilities including illumination, pose and expression, and because the ground truth position of eyes, and mouth is available.

The BioID dataset consists of 1521 gray level images

with a resolution of 384×286 pixels. Each one shows the frontal view of a face of one out of 23 different test persons. The number of images per subject is variable, as is the background (usually cluttered like in an office environment). The FRGCv2.0 face database consists of 50,000 recordings. The validation partition consists of data from 4,003 subject sessions. A subject session is the set of all images of a person taken each time a person's biometric data is collected and consists of four controlled still images, two uncontrolled still images. The controlled images were taken in a studio setting, are full frontal facial images taken under two lighting conditions and with two facial expressions (smiling and neutral). The uncontrolled images were taken in varying illumination conditions; e.g., hallways, atrium, or outside. Each set of uncontrolled images contains two expressions, smiling and neutral.

VI. EXPERIMENTAL RESULTS

For the evaluation of our algorithms we use two databases BioID and FRGCv2.0. Because we are interested in face recognition scenarios, we evaluate in this paper only the points which we use for our face normalization step[3]. These points are the centers of the eyes and the mouth center. For the eyes we use the measure that was proposed by Jesorsky et al. [13], where the localization criterion is defined in terms of the eyes center positions:

$$d_{eye} = \frac{\max(d_{lefteye}, d_{righteye})}{\|C_l - C_r\|} \quad (3)$$

where C_l, C_r are the ground truth eye center coordinates and $d_{lefteye}, d_{righteye}$ are the distances between the detected eye centers and the ground truth ones. In the evaluation, we treat localizations with d_{eye} above 0.05 as unsuccessful. Mouth center is evaluated in the same way but normalized with the distance between the average point of two eyes C_{eyeC} and mouth center C_{mouth} from ground truth.

$$C_{eyeC} = \frac{C_l + C_r}{2}, d_{mouth} = \frac{d_{mouthRaw}}{\|C_{eyeC} - C_{mouth}\|} \quad (4)$$

where $d_{mouthRaw}$ is the distances between the detected mouth centers and the ground truth.

As explained in Section IV, for the Combined-ASM model, we use the 58 landmarks (present in the IMM database), divided into 45 landmarks for the internal facial region and 13 points that belong to the facial contour regions. As eyes' and mouth centers are not present among these 58 landmarks (see Fig.4), we calculate them by averaging the landmarks detected around the eyes and mouth. For training the combined-ASM model,

we use a subset of the IMM database [20], ignoring some nonfrontal images. We use coarse to fine search over 2 levels of Gaussian scale pyramid. The SIFT block contain 4×4 cells with 4×4 pixels and 8 gradient orientation bins thus having descriptor size of 128. The length of the profile is set to be 17. It has to be noted that as prior to the landmark location step, we apply a face detection algorithm in order to have a rough location of where the face is located. We use the AdaBoost approach proposed by Viola and Jones [11], freely available from the OpenCV library [23].

For comparison purposes, we also use the publicly available STASM software [19], with already trained models for face landmark location. There are published results of this software for automatic landmark detection on the BioID database. So we can compare the performances of the proposed Combined-ASM method for facial landmark detection with the published results on the BioID database and also with the existing STASM software. We also used the STASM available code, to train a new model, using different training databases, in order to evaluate the influence of the training data used to train the statistical models.

A. Evaluation on the BioID Database

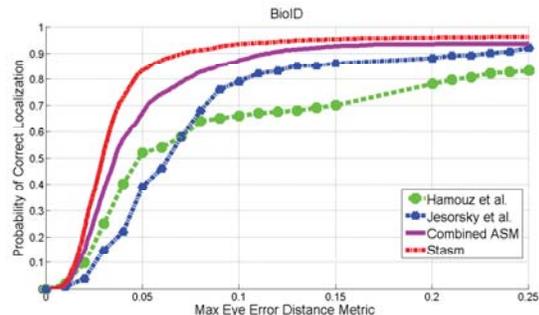


Fig. 6. Comparison of the proposed Combined-ASM with already published results for eyes detections on the BioID database

For comparison purposes, in Fig.6, we reproduce three published result related to the d_{eye} measurements including results of Jesorsky et al. [13], Hamouz et al. [4], and the results of the STASM software by Milborrow [19]. These results are compared with our Combined-ASM model implementation. The first two methods are image-based methods, while the last two ones are structure-based methods. It is obvious that structure-based methods have better performance than image-based methods even at the error level of Error < 0.1 . Our Combined-ASM method performs better than the two image-based methods, but worse than the available STASM software.

The STASM method extended the original ASM by using among other points, 2D profile and more landmarks, during the training step (using annotated images from the XM2VTS database). The effect of the number of landmarks on the detection performance is out of scope of this paper. Generally speaking, using more landmarks will give an overall better fitting [24]. The results of the STASM software that we trained with a different training data (the IMM database) are also presented in Table VI-B.

B. Evaluation on the FRGC Database

From the FRGCv2.0 database, we used the subpart called **spring2003** which contains 11,204 images, to evaluate our landmark location precision. There are not published results available on the FRGCv2.0 related to landmark location. Therefore we can only compare our results with the results of the STASM software.

Because with the STASM software (that also uses a face detection part as a first step) there are about 39 % of the above mentioned **spring2003** set of the FRGCv2.0 database images where the STASM face detection algorithm fails, we applied our landmark location software on the same set, for sake of comparison.

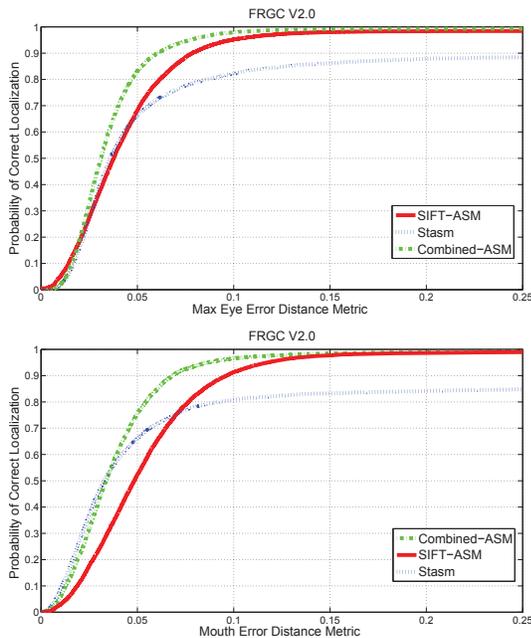


Fig. 7. Cumulative histograms on FRGCv2.0 database with Maximum eyes and mouth error

In Fig.7 we compare the results of the STASM software, with the Combined-ASM model. In order to evaluate the contribution of using different models for

different parts of the face, we also report results of using the ASM model with the SIFT features instead of the originally proposed features (denoted as SIFT-ASM). The result show that the combined-ASM method gives better performance for the eyes and mouth locations then the STASM software and the ASM method with new SIFT features.

The Combined-ASM highly improves the precision of the mouth center, because the SIFT feature descriptor works more inaccurately in the contour points, and those points will affect the mouth region landmarks during the fitting phase.

TABLE I
SPATIAL MEAN ERROR RATE (AT 10 %) OF EYES AND MOUTH CENTERS DETECTION, OF VARIOUS LANDMARK DETECTION ALGORITHMS ON BIOID AND FRGCv2.0, FROM VARIOUS LEARNING DATABASES: IMM AND XM2VTS(XM2 FOR SHORT) (IN %)

Testing data	BioID		FRGCv2.0	
	IMM	XM2	IMM	XM2
Stasm	78.5	95	77.4	81.6
SIFT-ASM	75	-	86.8	-
Combined-ASM	86	-	95	-

C. Discussion

The above presented experiments show that structure-based methods have better performance than image-based methods for facial landmark location. The STASM software which uses the 2D profile and an extended set of landmarks for the training phase, presents better results on the BioID database but not as good as in the FRGCv2.0 Database. One possible reason is due to the different characteristics of the databases. In the BioID database, all the images are captured when the person is near to the camera, so the face is the largest part of the image. In the FRGCv2.0 database, there are uncontrolled images where the human face occupies a smaller area in the image. When the face area is small in the image, the initialization from the face detector will not be as precise as it works for "passport style" photographs. Actually the ASM is an iteration strategy whose performance highly depends on the initialization. Using 2D profile as Local Texture Models will increase the precision, while using the SIFT descriptor will increase the robustness when bad initialization happens. Because the SIFT descriptor is scale and rotation invariant, even if the face area detected by the face detector is enlarged and decreased or distorted, it will not affect the Local Texture Models matching phase.

In the BioID database the average distance between two eyes is about fifty pixels, one pixel costs two percent error rate, and that error can be ignored when normalizing the face for face recognition. In that case the Combined-ASM algorithm seems to be robust without losing much of its accuracy for facial landmark detection for face recognition in cases when illumination, scale and small pose variation is present in the recording conditions (such as in the FRGCv2).

VII. CONCLUSIONS AND PERSPECTIVES

We present a new algorithm that successfully localizes facial landmarks for face recognition experiments. We assess the localization performance of the proposed method on two benchmarking datasets (BioID and FRGCv2). A preliminary version of the proposed method (using only the ASM model with SIFT features) was experimented on the portal challenge experiment of the MBGCv1 evaluations, that took place in December 2008, and that is described in [3], where an automatic landmark detector was needed, in order to find the position of the normalization points in the video frames.

In the proposed Combined-ASM we extend the original ASM by using the SIFT descriptor as a new local texture model and split the facial landmarks in facial internal region and facial contour landmarks. The proposed Combined Active Shape Model algorithm is more robust for eyes and mouth center localization in more challenging lighting conditions, and also where some pose and expressions variabilities are presented.

Our future work is oriented towards working on an extended set of landmarks which are not only limited to eyes and mouth locations, and that can be used for face processing algorithms, like face modeling and tracking.

REFERENCES

- [1] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, vol. 60, pp. 91–110, 2004.
- [2] T. F. Cootes, C. J. Taylor, D. H. Cooper, and J. Graham, "Active shape models—their training and application," in *Computer Vision and Image Understanding*, vol. 61, no. 1. New York, NY, USA: Elsevier Science Inc., 1995, pp. 38–59.
- [3] A. Mayoue, A. M. Mellakh, D. Zhou, D. Petrovska-Delacrétaz, and B. Dorizzi, "Utilisation de séquences vidéo avec critres de qualité pour la reconnaissance faciale," in *TRAITEMENT ET ANALYSE DE L'INFORMATION : Méthodes et Applications*, 2009.
- [4] M. Hamouz, J. Kittler, J. K. Kamarainen, P. Paalanen, H. Kalvinainen, and J. Matas, "Feature-based affine-invariant localization of faces," vol. 27, no. 9, pp. 1490–1495, Sept. 2005.
- [5] S. Arca, P. Campadelli, and R. Lanzarotti, "A face recognition system based on automatically determined facial fiducial points," *Pattern Recogn.*, vol. 39, no. 3, pp. 432–443, 2006.
- [6] I. R. Fasel, M. S. Bartlett, and J. R. Movellan, "A comparison of gabor filter methods for automatic detection of facial landmarks," in *Proc. Fifth IEEE International Conference on Automatic Face and Gesture Recognition*, 21–21 May 2002, pp. 242–246.
- [7] A. A. Salah, H. Inar, L. Akarun, and B. Sankur, "Robust facial landmarking for registration," *Annals of Telecommunications*, vol. 62, pp. 1–2, 2006.
- [8] D. Vukadinovic and M. Pantic, "Fully automatic facial feature point detection using gabor feature based boosted classifiers," in *Proc. IEEE International Conference on Systems, Man and Cybernetics*, vol. 2, 10–12 Oct. 2005, pp. 1692–1698.
- [9] G. Antonini, V. Popovici, and J. Thiran, "Independent Component Analysis and Support Vector Machine for Face Feature Extraction," in *4th International Conference on Audio- and Video-Based Biometric Person Authentication*, Guildford, UK, ser. Lecture Notes in Computer Science, vol. 2688. Berlin: IEEE, 2003, pp. 111–118.
- [10] N. Gourier, D. Hall, and J. L. Crowley, "Facial features detection robust to pose, illumination and identity," in *Proc. IEEE International Conference on Systems, Man and Cybernetics*, vol. 1, 2004, pp. 617–622.
- [11] P. Viola and M. Jones, "Robust real-time object detection," in *International Journal of Computer Vision*, 2001.
- [12] C. Du, Q. Wu, J. Yang, and Z. Wu, "Svm based asm for facial landmarks location," in *Proc. 8th IEEE International Conference on Computer and Information Technology CIT 2008*, 8–11 July 2008, pp. 321–326.
- [13] O. Jesorsky, K. J. Kirchberg, and R. Frischholz, "Robust face detection using the hausdorff distance," in *AVBPA '01: Proceedings of the Third International Conference on Audio- and Video-Based Biometric Person Authentication*. London, UK: Springer-Verlag, 2001, pp. 90–95.
- [14] T. Shakunaga, K. Ogawa, and S. Oki, "Integration of eigen-template and structure matching for automatic facial feature detection," in *Proc. Third IEEE International Conference on Automatic Face and Gesture Recognition*, 14–16 April 1998, pp. 94–99.
- [15] S. Ordas, L. Boisrobert, M. Huguet, and A. F. Frangi, "Active shape models with invariant optimal features (iof-asm) application to cardiac mri segmentation," in *Proc. Computers in Cardiology*, 21–24 Sept. 2003, pp. 633–636.
- [16] T. F. Cootes, C. Taylor *et al.*, "Statistical models of appearance for computer vision," 2004.
- [17] D. Monzo, A. Albiol, and J. Sastre, "Hog-ebgm vs. gabor-ebgm," in *Proc. 15th IEEE International Conference on Image Processing ICIP 2008*, 12–15 Oct. 2008, pp. 1636–1639.
- [18] L. Wiskott, J. M. Fellous, N. Kruger, and C. von der Malsburg, "Face recognition by elastic bunch graph matching," in *Proc. International Conference on Image Processing*, vol. 1, 26–29 Oct. 1997, pp. 129–132.
- [19] S. Milborrow and F. Nicolls, "Locating facial features with an extended active shape model," *ECCV*, 2008.
- [20] M. B. Stegmann, B. K. Ersbøll, and R. Larsen, "FAME – a flexible appearance modelling environment," *IEEE Trans. on Medical Imaging*, vol. 22, no. 10, pp. 1319–1331, 2003.
- [21] E. Bailly-Bailliere *et al.*, "The banca database and evaluation protocol," in *Audio- and Video-Based Biometric Person Authentication: Proceedings of the 4th International Conference*, 2003.
- [22] P. J. Phillips, P. J. Flynn, T. Scruggs, K. W. Bowyer, J. Chang, K. Hoffman, J. Marques, J. Min, and W. Worek, "Overview of the face recognition grand challenge," *Computer Vision and Pattern Recognition, IEEE Computer Society Conference on*, vol. 1, pp. 947–954, 2005.
- [23] G. Bradski, A. Kaehler, and V. Pisarevski, "Learning-based computer vision with intel's open source computer vision library," *Intel Technology Journal*, vol. 9, no. 2, pp. 119–130, May 2005.
- [24] G. M. Beumer, Q. Tao, A. M. Bazen, and R. N. J. Veldhuis, "A landmark paper in face recognition," in *Proc. 7th International Conference on Automatic Face and Gesture Recognition FGR 2006*, 2–6 April 2006, pp. 6pp.–78.