Facial Landmarks Detection Using Extended Profile LBP-Based Active Shape Models

Nelson Méndez, Leonardo Chang,
Yenisel Plasencia-Calaña, and Heydi Méndez-Vázquez

Advanced Technologies Application Center 7ma A ♯ 21406, Playa, Havana, Cuba
{nllanes,lchang,yplasencia,hmendez}@cenatav.co.cu

Abstract. The accurate localization of facial features is an important task for the face recognition process. One of the most used approaches to achieve this goal is the Active Shape Models (ASM) method and its different extensions. In this work, a new method is proposed for obtaining a Local Binary Patterns (LBP) based profile for representing the local appearance of landmark points of the shape model in ASM. The experimental evaluation, conducted on XM2VTS and BioID databases, shows the good performance of the proposal.

Keywords: facial landmarks, facial features detection, ASM, LBP.

1 Introduction

A fundamental step in the face recognition process is to model the face shape in an accurate manner; this allows one to find a correspondence of landmark points in face images for their posterior normalization or affine warping. However, the problem of accurately find these landmarks remains an open issue. The Active Shape Models (ASM) method [1] is one of the main approaches to automatically detect these points. ASM uses an appearance model to represent the image appearance around each landmark. In the original approach, this model is based on the so-called "gray-level profile", defined by the differences in the intensity values of adjacent pixels on a line centered at the landmark point. This is very sensitive to illumination changes and noise, not providing a sufficiently good description and discrimination of the local appearance.

Some extensions have been proposed in order to improve or replace the gray-level profile in ASM [2,3,4,5]. The Local Binary Patterns (LBP) operator [6] is a popular local appearance descriptor that can be considered for this task [5]. When using LBP, robustness to monotonic illumination changes and noise is achieved. LBP have been used in [4] and [5] for representing the local appearance in ASM, obtaining very good results in comparison with the original approach. In this work we propose a new method that uses LBP with ASM, aiming at better describing the local appearance of possible locations to which the landmark can shift. The rest of this paper is organized as follows. Section 2 describes the original ASM method and some ASM extensions related to this work. In

J. Ruiz-Shulcloper and G. Sanniti di Baja (Eds.): CIARP 2013, Part II, LNCS 8259, pp. 407–414, 2013.

Section 3, after a short introduction on LBP, the new proposal named EP-LBP ASM is presented, as well as some aspects related to its computation. Section 4 presents the experiments followed by concluding remarks in Section 5.

2 Active Shape Models (ASM)

ASM was first introduced by Cootes *et al.* [1] in 1995. It can be considered a deformable model that attempts to locate landmark points of known objects in images using a shape model, which describes the typical variations of the object shape, and a set of profile models that give a statistical representation of the image appearance around each point of the model.

Using ASM, the shape of an object in a 2D image is represented by a set of n landmarks, concatenated into a single vector of dimension $2 \times n$. Then, given M training samples, M such vectors are generated and aligned to a common coordinate frame before performing the statistical analysis [1]. The aligned shapes can be then considered to form a distribution in a $2 \times n$ dimensional space, where the relationships between positions of every point can be modeled and learned. Finally, Principal Component Analysis (PCA) is used to approximate each shape in the training set, obtaining the so-called statistical shape model.

Given the statistical shape model, the ASM searches along profiles of each point the best match to the data in a new image. It is then necessary to have a model of the local appearance surrounding each landmark. In the original method proposed by Cootes *et al.* [1], the gray-level profile is represented by the normal to the shape boundary, passing through each landmark. This model is also learned from the training images. For searching and fitting the model in a new image, the learned mean shape is used as the initial shape. Then, each region is examined iteratively for searching the best shape and position parameters which best match the model to the image, until convergence is achieved.

2.1 ASM Extensions

There are several ASM extensions in the literature, most of them try to obtain a more discriminative and robust representation of landmarks appearance, which is also the aim of this paper.

The Combined-ASM method [2] is a combination of the original ASM gray-level profile with SIFT descriptor. Combined-ASM represents the local appearance of inner landmarks of the face using SIFT and uses the gray-level profile to represent the appearance of face border landmarks. This method provides a more discriminative description to inner points while maintains a description of face border that better describes edges. In Reg-ASM [3], regression is used to describe landmarks local appearance, in order to learn from false displacements in rectangular regions centered at landmarks. The use of regression allows one to infer causal relations, but may lead to false relations. Besides, it imposes a detailed labeling of true and false examples on the training dataset. A multi-resolution detector based on Multiple Kernel Learning (MKL) is proposed in [7],

combining kernels from different resolutions in order to use more information. The Active Appearance Models (AAM) method [8] is another extension of ASM, which introduces a more detailed description of the appearance. However, ASM has shown to be more accurate than AAM for locating landmarks [9]. But there are some recent extensions of AAM, such as CLM [10], SOS [11] and TST [12] that outperform both AAM and ASM.

An extension of ASM based on the LBP descriptor was first proposed in [4]. The method, named ELBP-ASM, extracts different radius LBP descriptors [6] over circular subimages (gray scale image and gradient magnitude image) centered at each landmark. In order to retain spatial information, block-based LBP is used. Later, Marcel *et al.* [5] propose three LBP-based ASM approaches: profile-based LBP-ASM, square-based LBP-ASM and divided-square-based LBP-ASM. The profile-based LBP-ASM approach extracts LBP values from the normal profile of each landmark. Square-based LBP-ASM builds LBP histograms from a squared region centered at a given landmark. In divided-square-based LBP-ASM, the same square is used, but partitioned into four regions from which the LBP histograms are extracted and concatenated into a single feature histogram. The best results were achieved with the divided-square-based LBP-ASM [5].

In this work we propose a new Extended Profile LBP-based ASM (EP-LBP ASM) method, aiming to improve ASM landmark points localization. Although our proposal is also based on LBP, it differs substantially from previous approaches. In order to describe the appearance of not only the neighborhood of each landmark, but of the regions in the face image which define possible fittings in each iteration, we extract several LBP histograms of squared regions equally separated over profile normal.

3 ASM Using a New Extended Profile Based on LBP

In this section, we first briefly introduce the LBP operator. Next, we describe our proposal for modeling and fitting the shape model using a new Extended Profile based on LBP operator (EP-LBP).

3.1 Local Binary Patterns (LBP)

The LBP operator is a texture descriptor introduced in [6]. The operator and its different extensions have been widely applied in many computer vision applications, and particulary for face analysis. The original operator labels each pixel of an image by thresholding its 3×3 neighborhood with reference to its intensity value, and considering the result as a binary number. Since the operator only encodes the ordinal comparison (darker or brighter) between pixels intensities, it is considered to be invariant to monotonic illumination variations. On the other hand, when using this operator, the local appearance of an image is usually represented by histograms of the image regions, which makes the representation more robust to different kinds of noise in the image.

3.2 Landmarks Local Appearance Description with EP-LBP

In this paper we propose the EP-LBP as a more distinctive landmarks local appearance descriptor. Unlike other LBP-based approaches in literature (refer to Section 2.1) that describe circular or squared regions centered in the landmark point, we aim at describing those regions of the face image which define possible landmarks fittings in each iteration.

The profile $\varrho(p, k, d)$ of a landmark point p, is defined as the set of k points over p normal, centered at p and separated by d pixels. Figure 1 shows an example of the profile $\varrho(p, 5, 2)$, defined by a set of $k = 5$ points in the line over p normal, separated by $d = 2$ pixels.

Once the profile is defined, for each point $\varrho_i \in \varrho(p, k, d)$, a LBP histogram, $HLBP_{\varrho_i,m}$, over a squared sized region of width m and centered at point ϱ_i is extracted. Then, the EP-LBP descriptor of a landmark point p, EP-LBP$_{(p)}$, is obtained by concatenating these k histograms. In this work we use the original $LBP^{u2}_{(8,1)}$, so the number of LBP labels is 59 [6] and then the size of our EP-LBP descriptor for a given landmark is $59 \cdot k$ bins.

3.3 Shape Fitting Using EP-LBP

For shape fitting using the proposed EP-LBP descriptor, in each iteration, for every point p, a set $C_p = \{c_1, c_2, \ldots, c_N\}$ of N candidate points of landmark point p is determined. Every point $c_i \in C_p$ is on the line defined by p normal and it is separated by d pixels from c_{i-1} and c_{i+1}. Point p corresponds with candidate point $c_{\lceil N/2 \rceil}$. This is graphically explained in Figure 2.

The descriptor EP-LBP$_{(c_i)}$ associated to each candidate point $c_i \in C_p$ is calculated and the new point p^*, will be then the candidate point with the smallest distance between its EP-LBP descriptor and the trained EP-LBP for its corresponding landmark:

$$p^* = \arg\min_{c_i} \chi^2(\text{EP-LBP}_{(c_i)}, \overline{\text{EP-LBP}}_{(p)}), \tag{1}$$

where χ^2 is the Chi-Squared histogram distance and $\overline{\text{EP-LBP}}_{(p)}$ is the mean EP-LBP for landmark point p obtained in the training step.

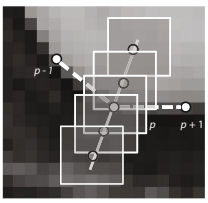

Fig. 1. Example of profile $\varrho(p, 5, 2)$

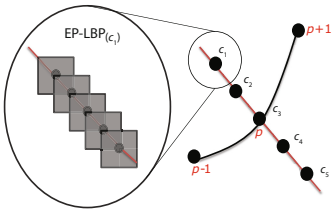

Fig. 2. Example of candidate points for landmark point p

3.4 EP-LBP Efficiency Improvements

In the fitting step, in each iteration, for every landmark point p, due to the intersection between C_p and $\varrho(p, k, d)$, $O(N \cdot k)$ re-calculations of $HLBP_{\varrho_i,m}$ histograms should be done when obtaining the EP-LBP descriptor for every candidate point in C_p. In order to avoid unnecessary processing we propose to compute a single EP-LBP$_{(p)}$ over a profile $\varrho(p, N + k - 1, d)$. Then, the process of selecting the new point p^* is simplified to selecting the sub-EP-LBP that best match the extracted EP-LBP$_{(p)}$. The previously mentioned improvement reduces the time complexity of this process from $O(m \cdot N \cdot k)$ to $O(m \cdot (N + k))$.

In practice, we have realized that when dealing with face frontal images and faces with little expression variations, the algorithm converges in the first ten iterations. Based on that fact, we propose another stop criteria for the fitting stage. The number of points which displacement in the current iteration was less than 2 pixels is determined and if this number is greater than the 95% of the total landmarks, the process is stopped.

4 Experimental Evaluation

In this section two experiments are described. First, we determine the best parameters for our proposal and then compare it with some existing approaches. Two standard face databases were used for this purpose: the BioID and the XM2VTS. The BioID Face Database (http://www.bioid.com) consists of 1,521 gray level images from 23 different persons, captured with large variations in expression, illumination, background and face size. The XM2VTS [13] contains 2,360 images captured during four recordings of 295 subjects over a period of four months, with different variations in expression, occlusions and appearance. In both cases the manual annotations of landmark points are given, 20 landmarks in the case of BioID and 68 for images in the standard sets of XM2VTS.

4.1 Model Parameters

To obtain the model parameters we used BioID, as it contains large variability in illumination and facial expressions. The images were randomly divided into two equal parts, one for training and the other for testing. The three different parameters were changed in the following way: $k = \{5, 7, 9, 11, 13\}$, $m = \{3, 5, 7, 17, 31\}$, and $d = \{1, 2, 3, 4\}$; all combinations of these values were tested.

To measure the quality of the fit we use the mean error in landmarks localization compared to ground-truth, given by:

$$m_{e_P} = \frac{1}{P * d_{eyes}} \sum_{j=1}^{P} d_j, \tag{2}$$

where P represents the number of points in the model, d_{eyes} is the distance between labeled eyes, and d_j is the distance between every detected point and its corresponding ground-truth position.

Figure 3 shows the results obtained for each parameter in all conducted experiments. When analyzing the graphics we found that the size of the region used, m, and the separation, d, are not as important as the number of points considered in the profile, k. Therefore, when computational time is a critical issue, minimum values for these parameters ($m = 3$ and $d = 1$) can be selected. Nevertheless, slightly better results were obtained for $m = 7$, and $d = 2$, so we used these values on the rest of our experiments.

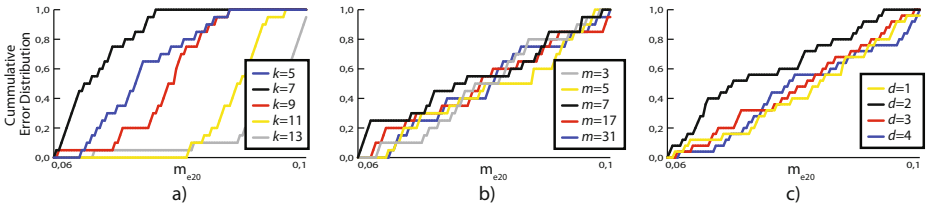

Fig. 3. Results for different values of EP-LBP ASM parameters: a) k, b) m, and c) d

4.2 Facial Landmarks Detection Accuracy and Timing

First, we compare our proposal with other existing ASM extensions based on LBP [4,5] and with the eyes detector used in [5]. This experiment is conducted on XM2VTS database, using the configuration I of Laussane Protocol [13]. Under this configuration, the shape model is trained with the parameters obtained in Section 4.1 using 3 images of 200 clients. The evaluation is performed on the standard test set defined in the protocol, as well as on the darkened set, where the robustness to illumination changes is evaluated. For the standard test set, the Mean Square Errors of all 68 landmarks is computed. In the case of the darkened set, similar to [5], we used the Jesorsky's measure for evaluation, that only takes into account the error in the center of the eyes points, since the annotations for all landmarks are not provided. The obtained results are shown in Table 1. As it can be seen in the table, our method obtained the best results in both, the standard and darkened sets of the XM2VTS database.

In order to compare the proposed EP-LBP ASM method with previously reported results of other ASM and AAM extensions described on Section 2.1, we used only 17 of the 68 landmarks to compute m_{e_P} following Equation 2. These landmarks correspond to the inner regions of the face: eyes, nose, eyebrows and

Table 1. Mean and median values of Mean Square Errors for the standard test set of XM2VTS, and Jesorsky's error measure for the darkened set

	Mean square error for the standard test set						Jesorsky's measure (error) for the darkened set					
	ASM	ELBP	Profile	Square	Divided	EP-LBP	detector	ASM	Profile	Square	Divided	EP-LBP
mean	73.0	41.0	71.0	61.0	46.0	21.0	0.11	0.11	0.11	0.09	0.10	0.06
median	32.0	31.0	43.0	41.0	28.0	18.0	0.10	0.09	0.09	0.08	0.07	0.05

mouth, which are actually the most difficult for fitting [10]. In Figure 4 a) the results are shown for our proposed EP-LBP ASM method compared to original ASM [1], Reg-ASM [3], CLM [10], SOS [11], and TST [12]. It can be seen that our method outperforms all the other methods.

Most of the above methods have been also tested on BioID database. In this case, similar to [10], we used the 20% of the images for training and the rest of them for testing, and the m_{e_P} with the 17 inner points is computed. The obtained results for our method compared to ASM [1], Reg-ASM [3], CLM [10] and MKL-based method [7], are shown on Figure 4 b). It can be seen that also in this database the results achieved by the proposed EP-LBP are better than or comparable to the other methods.

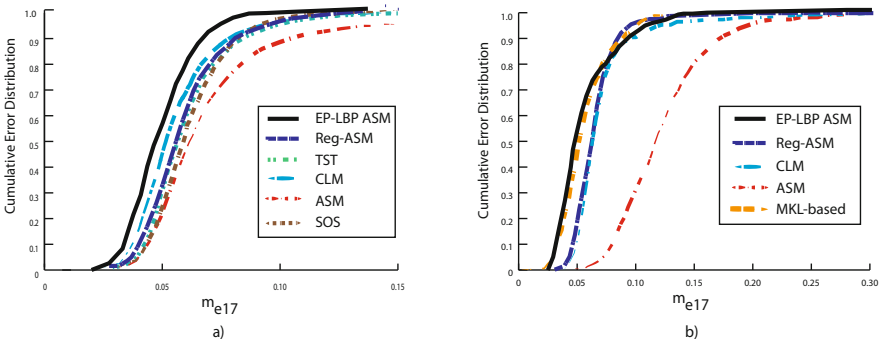

Fig. 4. Results in terms of $m_{e_{17}}$ on a) the standard test set of XM2VTS database and b) the BioID database

In the case of the Combined-ASM method [2], the results on BioID database are reported in terms of the Jesorsky's measure. They reported that about the 86% of the images have less than 0.10 of error, and about the 90% of them an error less than 0.15. Using this measure we have obtained a 92% of the images with less than 0.10 of error, and 98% with an error less than 0.15.

In terms of computation time, our approach, on average, needs 320 ms and the ASM needs 97 ms to detect landmarks in an image. This is comparable to other extensions [5] that use more information than the original ASM, which are also 2 or 3 times slower than ASM. The analysis was performed on a 2.5 GHz with 4 GB of RAM computer.

5 Conclusion and Future Work

In this paper a new Extended Profile based on LBP operator was introduced for representing the appearance of landmarks regions in the ASM method. The proposal, unlike other LBP-based approaches in the literature that describe circular or squared regions centered in the landmark point, describes regions of the face

image which better define possible shape fittings in each iteration. Experimental results on two well-known databases frequently used for this purpose, showed the good performance of the method and its superiority with respect to other state-of-the-art ASM and AAM extensions.

It should be noticed that EP-LBP ASM and the other methods described in Section 2.1, extend classical ASM by proposing better landmark appearance representations. However, there are other approaches (e.g. [14]) that improve ASM by enhancing the shape constraint, i.e. the correlation between landmarks. Our future work will focus on extending our method by exploiting this idea.

References

1. Cootes, T.F., Taylor, C.J., Cooper, D.H., Graham, J.: Active Shape Models - their Training and Application. Computer Vision and Image Understanding 61(1), 38–59 (1995)
2. Zhou, D., Petrovska-Delacrétaz, D., Dorizzi, B.: Automatic landmark location with a Combined Active Shape Model. In: 3rd IEEE International Conference on Biometrics: Theory, Applications and Systems. BTAS 2009 (2009)
3. Cristinacce, D., Cootes, T.F.: Boosted Regression Active Shape Models. In: British Machine Vision Conference, pp. 79.1–79.10. BMVA Press (2007)
4. Huang, X., Li, S.Z., Wang, Y.: Shape localization based on statistical method using extended Local Binary Pattern. In: Third International Conference on Image and Graphics. ICIG 2004, IEEE Computer Society, USA (2004)
5. Marcel, S., Keomany, J., Rodriguez, Y.: Robust-to-illumination face localisation using Active Shape Models and Local Binary Patterns. Tech Report, Idiap-RR-47-2006, IDIAP (2006)
6. Ojala, T., Pietikäinen, M., Mäenpää, T.: Multiresolution Gray-scale and Rotation Invariant Texture Classification with Local Binary Patterns. IEEE Transactions on Pattern Analysis and Machine Intelligence 24(7), 971–987 (2002)
7. Rapp, V., Senechal, T., Bailly, K., Prevost, L.: Multiple kernel learning svm and statistical validation for facial landmark detection. In: FG, pp. 265–271. IEEE (2011)
8. Cootes, T.F., Edwards, G.J., Taylor, C.J.: Active Appearance Models. IEEE Transactions on Pattern Analysis and Machine Intelligence 23(6), 681–685 (2001)
9. Cootes, T.F., Edwards, G.J., Taylor, C.J.: Comparing active shape models with active appearance models. In: British Machine Vision Conference (1999)
10. Cristinacce, D., Cootes, T.: Automatic feature localisation with constrained local models. Pattern Recogn. 41(10), 3054–3067 (2008)
11. Cristinacce, D., Cootes, T.F.: A comparison of shape constrained facial feature detectors. In: 6th International Conference on Automatic Face and Gesture Recognition 2004, Seoul, Korea, pp. 375–380 (2004)
12. Cristinacce, D., Cootes, T.F.: Facial feature detection and tracking with automatic template selection. In: FG, pp. 429–434. IEEE Computer Society (2006)
13. Messer, K., Matas, J., Kittler, J., Jonsson, K.: XM2VTSDB: The extended M2VTS database. In: Second International Conference on Audio and Video-based Biometric Person Authentication, pp. 72–77 (1999)
14. Sun, J., Wei, Y., Wen, F., Cao, X.: Face alignment by Explicit Shape Regression. In: IEEE Conf. on Computer Vision and Pattern Recognition, pp. 2887–2894 (2012)