Face Class Modeling based on Local Appearance for Recognition

Mokhtar Taffar, Serge Miguet

Face Class Modeling based on Local Appearance for Recognition

Mokhtar Taffar[1] and Serge Miguet[2]
[1]*Computer Sc. Dpt, University of Jijel, BP 98, Ouled Aissa, 18000, Jijel, Algeria*
[2]*LIRIS, Université de Lyon, UMR CNRS 5205, 5 av. Pierre Mendès-France, 69676, Bron, France*
mokhtar.taffar@gmail.com, serge.miguet@univ-lyon2.fr

Keywords: Invariant Descriptors, Local Binary Patterns, Features Matching, Probabilistic Matching, Model Learning, Appearance Modeling, Object Class Recognition, Facial Detection

Abstract: This work proposes a new formulation of the objects modeling combining geometry and appearance. The object local appearance location is referenced with respect to an invariant which is a geometric landmark. The appearance (shape and texture) is a combination of Harris-Laplace descriptor and local binary pattern (LBP), all is described by the invariant local appearance model (ILAM). We applied the model to describe and learn facial appearances and to recognize them. Given the extracted visual traits from a test image, ILAM model is performed to predict the most similar features to the facial appearance, first, by estimating the highest facial probability, then in terms of LBP Histogram-based measure. Finally, by a geometric computing the invariant allows to locate appearance in the image. We evaluate the model by testing it on different images databases. The experiments show that the model results in high accuracy of detection and provides an acceptable tolerance to the appearance variability.

1 INTRODUCTION

The facial image analysis remains an active domain of study (Agarwal et al., 2004; Fei-Fe et al., 2003) due to the difficulty to model and learn a wide range of intra-class appearance variability characterizing the face objects. The face detection is a subjacent problem to recognition where detect face can be considered as a two-class recognition problem in which a pattern is classified as being a face or non-facial appearance.

Thus, developing systems for facial detection has mainly two challenges: facial appearance modeling and probabilistic classifier design. The aim of facial modeling is to choice a set of the most discriminative local features extracted from face images and to construct a model, across these instances of facial features. The model should represent a large range of facial appearance by minimizing the intra-class variations and maximizing the extra-class ones. Obviously, if inadequate facial features are adopted, even the most performant classifiers will fail to accomplish the given recognition task of facial appearance (Hadid et al., 2004). Therefore, it is important to derive local features which should verify some properties like: prompt and easy extraction from images for an efficient processing, coding in a small size descriptor vector (low dimensionality of appearance space) to avoid a high computational cost of classifier, and a best classes discrimination with tolerance to within-class variations. But, it is not obvious to find features which simultaneously meet all these criteria because of the large variability in appearances due to different factors such as scale, face pose, facial expressions, lighting conditions, etc.

The basic LBP (local binary patterns) (Ojala et al., 2002) features have been performed very well in various applications, including texture classification and segmentation, image retrieval and surface inspection. By this work, we adapted a discriminative feature space which will be suitable to use for facial appearance recognition. The proposed approach based on the local descriptions consists of extracting a set of independent facial regions using Harris-Laplace detector. For each region, LBP feature histogram (representing texture contents within region) is computed and combined with the Harris-Laplace descriptor to build descriptor code of the region. That is this code that allows features matching and objects recognition in scene.

We present an invariant model based on local appearance, denoted ILAM, which is useful to detect/recognize faces in images. The learned model is based on similarity of appearances to recognize the

facial patches, then it become possible to predict their presence on new image. The ILAM model is defined across instances of a face; it is a geometric referential that links features over appearance changes. We boosted our LBP representation by a convenient probabilistic formulation to learn appearance variations. By this manner, the model captures well the multimodal nature of facial appearances in the cases of illumination and viewpoint changes. With new features a classifier is trained to capture the facial appearance of any viewpoint (frontal, profile, ...) in cluttered images. The facial LBP-appearance approach proposed here is suitable for any resolution images and has a short feature vector necessary for fast processing.

Experimentation proves that the ILAM learned and boosted leads to accurate face localization even when the appearance variation and intra-class variability occur (i.e., beard, ethnicity, etc.). The developed appearance algorithm is simple and has acceptable cost. Experiments with detecting low-resolution faces from images are also carried out to demonstrate that the same facial modeling can be reliable and efficiently used for such tasks.

In the following, a summary is given on works related to object class appearance modeling and recognition based on local descriptors. Section 3 provides a presentation of the new objects appearance description based on local traits useful for both learning process described in section 4 and facial detection process presented in section 5. We give, in section 6, some experimental results obtained on facial appearance recognition. Finally, at section 7, a conclusion finishes this paper.

2 RELATED WORKS

Due to difficulties to capture the large appearance variability of objects through the local features, despite the invariance of the last ones to different variations such as illumination, viewpoint, partial occlusion, etc., many models (Fergus et al., 2003; Toews and Arbel, 2006; Taffar and Benmohammed, 2011; Lindeberg, 1998) and features (Lowe, 2004; Mikolajczyk and Schmid, 2004; Kadir and Brady, 2001; Ojala et al., 2002) have been proposed and applied to facial appearance analysis. For instance in face detecting, the normalized pixel values (Heisele et al., 2000; Yang et al., 2002) and Haar-like features (Viola and Jones, 2001) are the most considered ones. Heisele et *al.* (Heisele et al., 2000) reported that normalized pixel values perform better than the gradient and wavelet features. Viola and Jones (Viola and Jones, 2001) used Haar-like features to form integral image characteristics and boosted them by AdaBoost algorithm for fast learning, this results an efficient face detection system.

Some features, such as those using PCA (Turk and Pentland, 1991) and LDA (Etemad and Chellappa, 1997) subspaces in face recognition, have also been considered. Such features are simple to compute, but their discriminative power is limited (Phillips et al., 2000). To overcome the main limitation of the PCA representation, Local Feature Analysis (LFA) is developed in (Penev and Atick, 1996). A good results have been obtained with Gabor wavelet features used in the elastic bunch graph matching algorithm (EBGM) (Wiskott et al., 1997). Unfortunately, the algorithm performs a complex analysis to extract a large set of Gabor wavelet coefficients. In (Ahonen et al., 2004), authors have obtained good performances in face recognition using an LPB-based method in which the face image was divided into many small non-overlapping blocks, but the representation cannot be used for small-sized face images common in many face detection and recognition problems. In (Taffar et al., 2012) the authors present a model which combines SIFT (Lowe, 2004) local features and a face invariant used as a geometric landmark. The model have a detection performance highly invariant to face viewpoints.

In (Hadid and Pietikinen, 2004), authors introduced a representation which consists of dividing the face image into several (e.g. 49) non-overlapping blocks from which the local binary pattern histograms are computed (using the $LBP^{u2}_{8,2}$ operator) and concatenating them into a single histogram. In such a representation, the texture of facial regions is encoded by the LBP while the shape of the face is recovered by the concatenation of different local histograms. However, this representation is more adequate for larger sized images (such as the FERET images) and leads to a relatively long feature vector typically containing thousands of elements. Therefore, to overcome this effect, they proposed in (Hadid et al., 2004) a new facial representation which is efficient for low-resolution images.

The emerging paradigm tries to model the objects as a collection of parts (Pope and Lowe, 2000; Fergus et al., 2003; Bart et al., 2004). Many contributions (Nanni et al., 2012; Déniz et al., 2011; Yu et al., 2013) used a combination of features from local regions looking for growth the performance of the detectors and recognition systems. However, the difficulty lies in learning the parameters for the model because we do not want to explore a huge space to know which parts are best for recognition. We overcome this problem by designing a convenient appearance

representation. The approach adopted learns simultaneously the facial LBP-appearance, its geometry and co-occurrence of features. The preselected facial features, through an histogram-based matching using a linear measure, are used in a probabilistic matching to predict facial appearance and to localize and recognize it with accuracy even in the presence of viewpoint changes and a rich multimodal appearance (i.e., expression, race, glasses).

3 LOCAL FACIAL APPEARANCE FORMULATION

In several LBP approaches of the literature, the representation of the whole face by dividing the face image (either or not by overlapping blocks) is effective and appropriate whether for images of high or low resolutions, but never for both. In addition, a LBP description computed over the whole face image encodes only the occurrences of the micro-patterns without any indication about their locations with respect to faces on images. The new appearance representation described here tries to overcome these limits. It will be suitable to deal with facial images of any sizes and where faces can be anywhere on image.

During learning, we compute a facial LBP-feature at each keypoint detected on face by using a scale and affine invariant detector, such Extended Harris-Laplacian detector (Mikolajczyk and Schmid, 2004). A scale and an affine invariant interest point detector combines the Harris detector with the Laplacian-based scale selection. The Harris-Laplace detector is then extended to deal with significant affine transformations.

In affine scale-space the second moment matrix μ, at a given point $\mathbf{x}$, is defined by:

$$\mu(\mathbf{x},\Sigma_I,\Sigma_D) = det(\Sigma_D)g(\Sigma_I) * ((\nabla_L)(\mathbf{x},\Sigma_D)(\nabla_L)(\mathbf{x},\Sigma_D)^T) \quad (1)$$

where Σ_I and Σ_D are the covariance matrices which determine the integration and differentiation Gaussian kernels.

These interest keypoints have invariant properties and are reputed to be tolerant to affine transforms, in-plane changes of scale and rotation. Each keypoint location corresponds to a central pixel of LBP region. By this manner, a geometric information enriches the structure of the local facial appearance with respect to an invariant. The keypoints detected on face are located anywhere in the image and computed on the 16×16 neighborhood. In this way, they can be at different locations, nearest or farest, from each other. In other terms, the LBP-blocks can be overlapped or not, as shown in Fig. 1. Thus, the representation which consists of dividing the face image into several overlapping blocks or not, and from which the local binary pattern histograms are computed then concatenated into a single histogram has been abandoned.

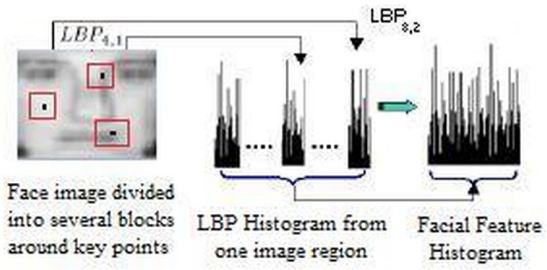

Figure 1: Local representation of facial appearance: in addition to Extended Harris-Laplacian descriptor, a 15×15 facial region around the keypoint is described by a concatenation of a set of local ($LBP_{4,1}$ and $LBP_{8,2}$ operators) LBP histograms.

The proposed facial LBP-appearance representation consists to define two kinds of blocks (overlapping and non overlapping ones) around each detected keypoint from which the local binary pattern histograms are computed (using two LBP operators: $LBP^{u2}_{8,2}$ and $LBP_{4,1}$) then concatenated into a single (two dimension) histogram. In such a representation, the texture of facial regions is encoded by the LBP while their shape is recovered by the concatenation of different local histograms. Therefore, we will propose here a general facial modeling which is not only efficient for low-resolution images, but also more adequate for larger sized images (such as FERET images). A strong point of this approach is that it requires no pretreatment of the face images (such standardization, background suppression, face mask extraction, etc.) and no geometric constraints (such as size of face and its location in image), so it is independent of the nature and size of the used image.

The first step uses the overlapping regions which are scrutinized by 8-neighborhood LBP operator ($LBP_{8,2}$), where the overlapping size is set to 2 pixels, this allows to avoid statistical unreliability due to long histograms computed over small regions. At each keypoint, in total 36 regions can be used to generate LBP code and construct histograms. The second process uses the non-overlapping blocks exploited by 4-neighborhood LBP operator ($LBP_{4,1}$) where the corresponding histogram is sampled in the same way as the $LBP_{8,2}$ operator. In total 25 regions are used to generate LBP code and construct histograms. Thus, each face is modeled by a set of local features. A local facial characteristic is defined by two LBP codes

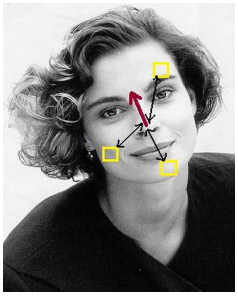

Figure 2: Facial invariant, represented by a red arrow on nose, is a geometric landmark for the local representation of facial appearance represented by yellow squares on face. The double-headed arrow describes affine transformation of the appearance region f_i to face invariant *inv*.

and concatenated histograms.

Finally, in the learning model, each local facial appearance is described by an Extended Harris-Laplace descriptor enhanced by a global LBP histogram computed over a 15×15 facial region around the keypoint by a concatenation of a set of local ($LBP_{4,1}$ and $LBP_{8,2}$ operators) LBP histograms. In addition, we assigned to each facial feature f_i the geometric parameters which correspond to in-plan transformations of the facial region with respect to (*wrt*) the landmark located on the nose and schematized as an arrow, it represents the face invariant (FI), denoted *inv*, as shown in Fig. 2. By this manner, during detection process, from the learning model, it will be possible to deduce the presence of a new facial region from all LBP histograms computed over a combination of detected features that strongly match to some model traits belonging to different learning images. This approach is also very useful for recognition. Thus, the face invariant *inv* in the test image could be easily predicted and localized from the geometric parameters (e.g., position, scale, and orientation) of a detected facial region (which has similar facial appearance in the model) with respect to invariant in the learning model.

In our experiments, we considered 15×15 as the minimal standard resolution region around a detected keypoint and we derived the facial LBP-appearance representation as follows:

At first, we divide a 15×15 facial image region around the keypoint into 36 overlapping regions of 5×5 pixels (overlapping size=2 pixels). From each region, we compute a 256-bin histogram using the $LBP_{8,2}$ operator which is sampled into 16-bin histogram with a sampling step of 16. Then, we concatenate the results into a single 576-bin histogram. In the second step, we divide the same 15×15 face region around the same keypoint into 25 non-overlapping blocks of 3×3 pixels. From each region, we compute a 16-bin histogram using the $LBP_{4,1}$ operator and concatenate the results into a single 400-bin histogram. Additionally, we apply $LBP^{u2}_{8,1}$ to the whole 15×15 facial region and derive a 59-bin histogram which is added to the 976 bins previously computed. Thus, we obtain a $(59 + 976 = 1,035)$-bin histogram as a local face representation at the detected point of interest (see Fig. 1). Finally, a face is defined by a set of independent local representation of facial appearance who is none other than a set of Extended Harris-Laplace descriptor and 1035-bin histogram. Thus, each facial feature, denoted $f_i = \{f_i^p, f_i^g, f_i^a\}$, has three parameters: presence f_i^p, geometric f_i^g, and appearance f_i^a.

The model is based on the assumptions which are the presence parameter f_i^p follows a discrete binomial distribution in the presence space and the appearance parameter $f_i^a = (D_{f_i}^{EHL}, LBP_{f_i})$ modeled by Extended Harris-Laplacian descriptor (denoted $D_{f_i}^{EHL}$) and LBP_{f_i} representation of feature f_i follows a normal distribution with mean μ^a and covariance $\sum^a$ in the appearance space. The geometric parameter f_i^g of the feature when with him is determined with respect to face invariant *inv* in the image.

4 LEARNING PROCESS

Given a set of N visual traits $\{f_i\}$ extracted from the training image, the model learns to detect if each f_i^a whether or not a facial appearance. In probability term, from a set of facial appearance $\{f_i^a\}$ of the subwindows extracted from the training image, the model quantifies the likelihood term of each $f_i^a = (D_{f_i}^{EHL}, LBP_{f_i})$ feature which can be expressed as

$$p(f_i^{p=1}|f_i^a) = \frac{p(f_i^{p=1})p(f_i^a|f_i^{p=1})}{p(f_i^a)} \qquad (2)$$

where p is the binary presence parameter of facial appearance, e.g., $p = 0$ for non face or background sample. LBP_{f_i} is the LBP facial representation of the training sample f_i extracted around Extended Harris-Laplace keypoint described by $D_{f_i}^{EHL}$ appearance, $D_{f_i}^{EHL}$ and LBP_{f_i} parts of f_i^a are statistically independents. Thus, it is important to accomplish the learning model under the following assumptions:

- f_i^g and f_i^a are statistically independent given presence f_i^p.
- $D_{f_i}^{EHL}$ and LBP_{f_i} are geometrically independents.
- f_i^g parameter is related to the geometry of $D_{f_i}^{EHL}$ appearance parameter of f_i.

Depending on whether f_i is a positive or negative sample (face or non-face), the model exhibits the

quantity of the probability ratio $R(f_i^a) = \frac{p(f_i^{p=1}|f_i^a)}{p(f_i^{p=0}|f_i^a)}$, e.g., f_i is face if $R(f_i^a) > 1$.

5 RECOGNITION PROCESS

In facial appearance recognition, given a set of visual observations $\{f_i\}$ extracted from test image, each f_i is defined by its appearance values $f_i^a = (D_{f_i}^{EHL}, LBP_{f_i})$ and geometric values $f_i^g = \{f_i^{g:(x,y)}, f_i^{g:\sigma}, f_i^{g:\theta}\}$ of the extracted image feature f_i. Thus, the learning model attempts to confirm if each f_i is or not a facial appearance $(f_i^a, f_i^{p=1})$. The classifier decides on the facial appearance of the subwindow according to the likelihood value of the following expression:

$$p(f_i^a|f_i^{p=1}) = \frac{p(f_i^a)p(f_i^{p=1}|f_i^a)}{p(f_i^{p=1})} \quad (3)$$

where $p(f_i^a|f_i^{p=1})$ is a posterior value to affirm the facial appearance of the feature, $p(f_i^a)$ is a prior over facial appearance in the learning model, $p(f_i^{p=1}|f_i^a)$ is the likelihood value of feature presence f_i given its facial appearance, and $p(f_i^{p=1})$ is the evidence that the feature is facial in the learning model.

Thus, it is interesting to perform facial appearance recognition using the learning model under the assumptions that are:

- $D_{f_i}^{EHL}$ and LBP_{f_i} are geometrically independents given f_i^g.
- f_i^g parameter is defined by geometric parameters of $D_{f_i}^{EHL}$ descriptor in the image.
- $D_{f_i}^{EHL}$ and LBP_{f_i} are appearance independents given facial appearance f_i^a.
- f_i^p presence parameter depends on presence of the local facial appearance f_i^a in the image.

From these hypothesis, given an appearance $D_{f_i}^{EHL}$ detected in the test image, it becomes easier to deduce multiple shapes and patterns of the facial appearance given by a combination of $D_{f_j}^{EHL}$ and LBP_{f_k} parts of different traits in the learning model, where f_j and f_k are the model traits, and $D_{f_i}^{EHL}$ is the appearance part that matches to $D_{f_j}^{EHL}$ in the model.

In addition, before to perform the EM classifier, a set of similarity measures is applied over the LBP-histograms (three different histograms-based metrics: Histogram Intersection, *Log*-Likelihood Statistic, Chi Square Statistic) in order to confirm the facial presence detections and remove the false ones. The threshold values are fixed empirically for each distance. For a given detected window, we count the number of recognitions by matching the histograms within a neighborhood of 15×15 pixels (each detected window is geometrically localized by its central trait). The detections are removed if no matching occurs at this region. Otherwise, we keep them, the regions for which the matching occurs have a high outcome of EM classification. The LBP traits extracted from new image are expected as facial features under different lightning and viewpoints variations. The idea is to find a cluster of features that have appearance agreement with a face. This set of data observations $\{f_i^{p=1}\}$ is formed by estimating appearance distance $d(f_i^a, f_j^a|f_j^b)$ result of a true face or background relatively to an appearance threshold Γ^a.

For each feature f_i, when appearance matching occurs, the facial appearance probability of f_i, denoted $p(f_i|f_i^{p=1}, f_j^a)$, can be computed, where p is the presence parameter, e.g., *p=0* for background or face appearance absence. Features with facial appearance are retained and they are reputed to belong to face, e.g., $p(f_i|f_i^{p=1}, f_j^a) > p(f_i|f_i^{p=0}, f_k^a)$, where f_j and f_k are the features of best probabilistic matching with f_i in facial and background spaces respectively. Moreover, we calculated the number of model features $v_i^{p=1}$ (resp. $v_i^{p=0}$) that have voted for image feature f_i as facial (resp. background). Thus, this appearance probabilistic classification of f_i allows deciding one by one if the image features have facial appearance or not.

Furthermore, once all the facial features are known in the test image, a hierarchical clustering based on the geometrical classification algorithm is performed. This makes it possible to group them according to their geometrical positions in the image. Using a geometric clustering threshold Γ^c, the algorithm provides one or more clusters of facial features. This allows to generate one invariant *inv* for each cluster. Each *inv* localizes a possible facial presence. Thus, a multiple facial appearance can be recognized in image. This procedure tries to confirm the appearance statistical dispersion on test image with respect to the appearance in the learning model.

6 EXPERIMENTATION

6.1 Data and Experimental Considerations

Because we assume that our appearance-based detection scheme captures very well the variability of

Figure 3: Examples of face images from CMU-Profiles database (CMU-Database, 2009) where faces present pose changes.

facial appearances, a low supervised learning with a training set of some hundreds images is sufficient to build the facial model. For this purpose, we collected a set of 300 face images belonging to a part of FERET (FERET-Database, 2009), CMU-profile (CMU-Database, 2009), PIE (PIE-Database, 2009), and AT&T (AT&T-Database, 1994) databases. Then, we increase the number of $f_i = \{f_i^p, f_i^g = (f_i^{g:(x,y)}, f_i^{g:\sigma}, f_i^{g:\theta}), f_i^a = (D_{f_i}^{EHL}, LBP_{f_i})\}$ features in the learning model by adding a negative samples $\{f_i^{p=0}\}$ from some natural images from the net to obtain a set of 960 face and non-face appearances. Additionally, to enable the system to also detect faces from any viewpoint (in-plane rotation), we used a training set of face images of the CMU-Profile database. Fig. 3 shows the examples of face images and the different rotated face samples. Overall, we obtained a training set of 360 faces. The faces are divided into three categories of views: frontal, view angle 20°-70°, and profile.

To collect more consistent and reliable (face and nonface) appearances (patterns and shapes), we used the *boostrap* strategy in five iterations [15]. First, we randomly extracted 200 appearances from a set of MIT-CMU Profile database which contain faces and 100 appearances from a set of natural images which do not contain faces. Then, at each iteration we trained the system, run the face detector, and collected all those face (resp. nonface) appearances that were wrongly classified as nonfaces (resp. faces) and used them for training. Overall, we obtained $1080 + 132$ facial appearances as positive training examples and $60 + 47$ nonface as negative training examples. The learning model involved $1,319$ features ($1,212$ facial features and 107 negatives); they are chosen well, deemed to be informative with respect to (*wrt*) invariant, and not redundant.

Some parameters have been experimentally fixed by testing their impact on accurate localization of the face invariant. The appearance threshold Γ^a is empirically set at different values for the corresponding histogram distances, this allows to preselect only features with coherent facial appearance.

To check the geometric agreement of predicted invariants and a symmetric consistency of facial features, the value of threshold Γ^g is set on 3-parameters that correspond to position, scale and rotation. The position parameter of threshold $\Gamma^{g:(x,y)}$ is a pixel distance that must be less than inverse of invariant scale. The scale parameter of threshold $\Gamma^{g:\sigma}$ is limited to a ratio of 5% in scale with respect to the scale variation in image, and orientation threshold $\Gamma^{g:\theta}$ enables 10 degrees of tolerance.

A clustering threshold Γ^c allows aggregating the geometry of predicted invariants for accurate localization of the best cluster of invariants. An invariant is clustered if its minimal distance to any invariant (either clustered or not) is less than Γ^c relative to the mean of scales.

6.2 Facial Recognition from Viewpoint

The performance of ILAM model is evaluated for different values of appearance threshold Γ^a and for different histogram distances. Figure 4 plots the result of experiments which allow us to set the best Γ^a value for each distance. For intersection and likelihood metrics the best detection rates (resp. 83.27% and 79.6%) are obtained for the values 0.4 and 0.9 of Γ^a respectively but less than for Chi Square distance. For $\Gamma^a = 0.6$, ILAM model gives a good precision result and the Chi Square distance is clearly best. The face detection grows quickly to reach the rate of 92.4% since the chosen threshold Γ^a is cut as a precision factor to predict the facial appearance presence.

The Precision-Recall curves (PRC) in figure 5 drawn the performance comparison of ILAM model for different distances of similarity on the same protocol. It depicts that histogram distance of *Chi Square* provides an accuracy quality to the model than the *Log*-likelihood Statistic similarity. The *Chi Squared* distance is slightly better than the *Histogram Intersection* metric and no need to impose geometric or appearance constraints on faces in cluttered image before detection.

From the collected training sets, we extracted the proposed facial representations (as described in section 5). Then, we used these features as inputs to the classifier and trained the face detector. Thus, the system is run on several images from different sources in order to prove its acceptable performance of facial appearance recognition. Fig. 6 shows some detection examples. It can be seen that not only most of the upright frontal faces are detected but also the faces which present viewpoint changes and an appearance variability. For instance, Fig. 6.A shows perfect detections. In Fig. 6.D no face is missed and one face is detected by the system even when the severe occlu-

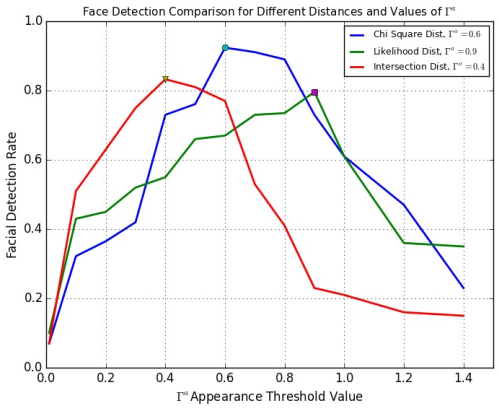

Figure 4: The evaluation results of facial appearance detection for different values of Γ^a, on the protocol of 300 face images from the CMU-Profile database (CMU-Database, 2009), allows to set the best Γ^a value for each distance. The results illustrate the rate of 92.4% of ILAM model to infer faces which present viewpoint changes.

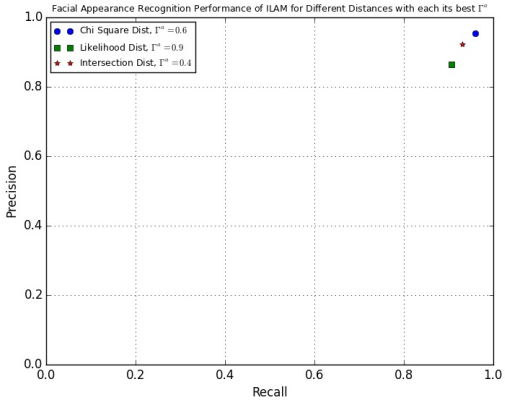

Figure 5: PRC curves of facial appearance model, on the protocol of 180 face images from the ORL database (AT&T-Database, 1994) for a face localization task, illustrates the rate of 95.6% of ILAM model to infer faces in images for threshold appearance $\Gamma^a = 0.6$, since *Chi Squared* distance is proved the best in this context.

sion occurs. A similar case is shown in Fig. 6.G in which the face is not missed despite a large in-plane rotation.

Since the system is trained to detect faces at any degree of in-plane rotation, from the face view to the profile view (i.e., up to $\pm 90^\circ$), it succeeded to find the strongly rotated faces in Fig. 6.G, 6.H and Fig. 6.I, and failed to detect slightly rotated ones (as those in Fig. 6.C) due to the large appearance variability. A false negative is shown in Fig. 6.J while a false positive is shown in Fig. 6.C, 6.G and 6.H but their numbers in the statistics detector are very low. Notice that this false positive is expected since the face is

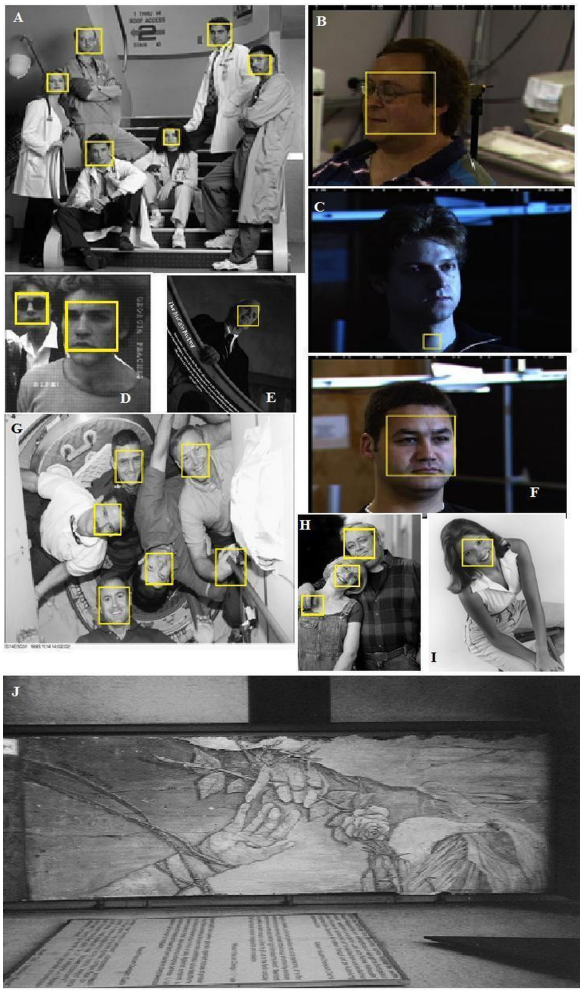

Figure 6: Facial appearance recognition examples in several images from different databases. The images A, D, E, G, H and I are from the subset of MIT-CMU tests. They belong to the 120 images considered for comparison. The images B, C and F are from PIE database. We notice the excellent detections of upright faces in A, D, F and G; detections under slight in-plane rotation in G and H; even with glass occlusion a right detection in D, missed face and false detection in C because of high appearance variability, detected faces correctly in E, G, H and I even with a large in-plane rotation; no missing face in B caused by a profile view of face; and false detections in G and H due to similarity to facial appearance.

pose-angled from the range angle of $45^\circ \pm 5$ worsened when the facial appearance variability is drastic and the detector performs well for the profil view. These examples summarize the main aspects of our detection system using images from different sources. The detected faces are marked by a rectangle encompassing or covering one or several facial regions each characterized by the detected facial LBP-appearance, as shown in Figure 7.

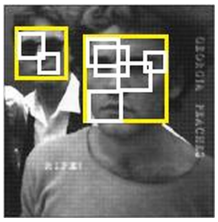

Figure 7: The yellow englobing rectangles of detected faces include one or several facial LBP-appearance regions represented by white rectangles.

Table 1: Comparative performance of ILAM facial detector based LBP-appearance representation with some detectors on 80 images containing 227 faces.

Method	Face Det.	False Det.	Det. rates
BDF Method	221	1	97.4%
Schneiderman-Kanade(1.0, 1.0)	218	41	96.0%
Normalized Pixel features	213	6	93.8%
LBP feature $LBP^{u2}_{8,1}$ (59 bins)	47	184	20.7%
LBP feature $LBP_{4,1}$ (144 bins)	211	9	92.9%
$LBP_{4,1}$+$LBP^{u2}_{8,1}$ trait(203 bins)	222	13	97.8%
ILAM based f_i trait	225	3	99.1%

In order to report quantitative results and compare them against those of the state-of-the-art algorithms, we considered the test images from the MIT-CMU sets (CMU-Database, 2009) that are used principally with the Bayesian Discriminating Features (BDF) method (Liu, 2003), Schneiderman-Kanade approach (Schneiderman and Kanade, 1998), Normalized Pixel features (Heisele et al., 2000), and LBP representation as in (Hadid et al., 2004). There are 80 images containing 227 faces. Some of these images are shown in Fig. 6.(A, D, E, G, H, and I).

Table 1 presents the performance of our facial appearance recognition system and those of other approaches like: BDF (Liu, 2003), Schneiderman-Kanade (Schneiderman and Kanade, 1998), and LBP-feature used in (Hadid et al., 2004). We can see (from the 1st, 2nd, 5th and 6th rows of Table 1) that our approach has a performance slightly higher to the comparative approaches. The proposed ILAM detector using LBP-based approach (where LBP features are extracted around Extended Harris-Laplace keypoint) succeeded in detecting 225 faces with few false positives caused by the similar appearance of the detected objects to faces. Some missing faces are mainly due to severe facial appearance variability added to a large in-plane rotation (as shown an example in Fig. 6.C) and sometimes to occlusion. We notice that ILAM system has an high performance, moreover, it is more general and not only dedicated to frontal faces but also to faces in different poses and even when occlusion occurs.

Additionally, if the detected faces are to be fed to a recognition step, then, no raison to tolerate some false detections even if it is likely that these images will be rejected (therefore they will not be accepted as those of an individual). In such a context even if our face detector performs slightly better as it succeeded in detecting 225 faces among 227 (the system detected the tilted faces in Fig. 6.*G* despite the large in-plane rotation) it tolerates only 3 false detections. The $8th$ row of Table 1 presents this performance.

Analyzing the ILAM representations and investigating the usefulness of dividing the facial images into regions around detected keypoints, we noticed that calculating the LBP traits from these regions yielded a good result (see the 8th row in Table 1). This is expected since such a representation encodes only the occurrences of the micro-patterns without any indication about their locations. Combining both representations further enhances the detection performance. However, computing the LBP traits only from the whole images (59 bins) yields a low detection rate of 20.7% (see 5th row in Table 1).

In order to further investigate the discriminative power of facial appearance of ILAM model, we used a similar face detector combined with an EM classifier and using different features as inputs, then compared the results to those obtained using the proposed f_i traits. We trained the system using the same training samples as described in Section 6.1 and we perform the tests on 238 images containing 300 faces from the subset of CMU-Profile (CMU-Database, 2009), CMU-PIE (PIE-Database, 2009), and AT&T (AT&T-Database, 1994).

We chose, for experimental purpose, the HoG (histograms of oriented gradients) features (Déniz et al., 2011), LBP/LTP representation (Nanni et al., 2012), and ones based on patch-based SIFT-LBP integration (Yu et al., 2013) as inputs, even if, it has been shown in (Nanni et al., 2012) that such texture descriptors (LBP/LTP patterns and local phase quantization) for describing region and a bag-of-features approach for describing object performs comparably well to HoG and SIFT-LBP based ones when using a SVM classifier.

Table 2 (5th row) shows the performance of ILAM model based on $LBP_{4,1} + LBP_{8,2} + LBP^{u2}_{8,1}$ traits computed over the local region. Although the results are quite good as 294 faces among 300 were detected, still the proposed approach using the f_i visual features, where $f^a_i = (D^{EHL}_{f_i}, LBP_{f_i})$ is the facial appearance part of f_i, computed over regions around keypoints

- performed better (comparison between the $5th$ row and $6th$ row in Table 2);
- used a combination of well know features which results on simple descriptor and histogram vectors

Table 2: Comparative performance of ILAM modeling combined with an EM classifier to different features used as inputs.

Method	Face Det.	False Det.	Det. rates
HOG features	293	8	97.6%
LBP/LTP representation	294	6	98.1%
Patch-based SIFT-LBP	296	6	98.6%
$LBP_{4,1} + LBP_{8,2} + LBP^{u2}_{8,1}$ trait	294	9	98.0%
$f_i = \{f_i^p, f_i^g, f_i^a\}$ trait	298	5	99.3%

and thus more faster to compute over the little regions;

- did not need to impose anyone constraint like histogram equalization; and
- principally needs a simple EM classifier to estimates the latent data, than using a series of SVM classifiers (Hadid et al., 2004; Vapnik, 1998).

7 CONCLUSION

The appearance representation of face class presented in this paper offers robust properties such as tolerance to geometric transforms and illumination changes. It captures well the viewpoints variations and especially intra-class variability. It has a geometric localization sufficiently accurate and its magnitude remains roughly constant with respect to size of object in image. The ILAM model based on combination of local appearance of Extended Harris-Laplace descriptor and texture of LBP feature provides a low degree of supervision. The experimentation reveals that the facial formulation is useful and has high capability to classify new face instances, of course this representation can be applied to another object class.

REFERENCES

Agarwal, S., Awan, A., and Roth, D. (2004). Learning to detect objects in images via a sparse, part-based representation. In *PAMI, 26(11), pp. 1475–1490*.

Ahonen, T., Hadid, A., and Pietikinen, M. (2004). Face recognition with local binary patterns. In *Proc. of the 8th ECCV Conference*.

AT&T-Database (1994). At&t: The database of faces. In *Cambridge University, http://www.cl.cam.ac.uk/*.

Bart, E., Byvatov, E., and Ullman, S. (2004). View-invariant recognition using corresponding object fragments. In *ECCV, pp 152-165*.

CMU-Database (2009). Cmu face group and face detection project, frontal and profile face images databases. In *http://vasc.ri.cmu.edu/idb/html/face/*.

Déniz, O., Bueno, G., Salido, J., and la Torre, F. D. (2011). Face recognition using histograms of oriented gradients. In *Pattern Recognition Letters, vol.32, pp:1598-1603*.

Etemad, K. and Chellappa, R. (1997). Discriminant analysis for recognition of human face images. In *Journal of the Optical Society of America, vol.14, pp:1724-1733*.

Fei-Fe, L., Fergus, R., and Perona, P. (2003). A bayesian approach to unsupervised one-shot learning of object categories. In *ICCV, Nice, France, pp. 1134–1141*.

FERET-Database (2009). Color feret face database. In *www.itl.nist.gov/iad/humanid/colorferet*.

Fergus, R., Perona, P., and Zisserman, A. (2003). Object class recognition by unsupervised scale-invariant learning. In *CVPR, Madison, Wisconsin, pp. 264–271*.

Hadid, A. and Pietikinen, M. (2004). Selecting models from videos for appearance-based face recognition. In *Proc. of the 17th International Conference on Pattern Recognition (ICPR)*.

Hadid, A., Pietikinen, M., and Ahonen, T. (2004). A discriminative feature space for detecting and recognizing faces. In *CVPR Proceedings, Vol. 2, pp. 797–804*.

Heisele, B., Poggio, T., and Pontil, M. (2000). Face detection in still gray images. In *Technical Report 1687, Center for Biological and Computational Learning, MIT*.

Kadir, T. and Brady, M. (2001). Saliency, scale and image description. In *IJCV, 45(2), pp. 83–105*.

Lindeberg, T. (1998). Feature detection with automatic scale selection. In *International Journal of Computer Vision, vol. 30(2), pp. 79-116*.

Liu, C. (2003). A bayesian discriminating features method for face detection. In *IEEE Trans. on PAMI, vol. 25, pp:725-740*.

Lowe, D. (2004). Distinctive image features from scale-invariant keypoints. In *IJCV, 60(2), pp. 91–110*.

Mikolajczyk, K. and Schmid, C. (2004). Scale and affine invariant interest point detectors. In *IJCV, 60(1), pp. 63-86*.

Nanni, L., Brahnam, S., and Lumini, A. (2012). Random interest regions for object recognition based on texture descriptors and bag of features. In *Expert Systems with Applications, Elsevier Journal, vol.39, pp:973-977*.

Ojala, T., Pietikinen, M., and Menp, T. (2002). Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. In *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI), vol.24, pp:971-987*.

Penev, P. and Atick, J. (1996). Local feature analysis: a general statistical theory for object representation. In *Network: Computation in Neural Systems, vol.7, pp:477-500*.

Phillips, P., Moon, H., Rizvi, S. A., and Rauss, P. J. (2000). The feret evaluation methodology for face recognition algorithms. In *IEEE Trans. on PAMI, vol.22, pp:1090-1104*.

PIE-Database (2009). Cmu pose, illumination, and expression (pie) database. In *http:/www.ri.cmu.edu/projects/project_418.html*.

Pope, A. and Lowe, D. (2000). Probabilistic models of appearance for 3-d object recognition. In *IJCV, 40(2), pp. 149–167*.

Schneiderman, H. and Kanade, T. (1998). Probabilistic modeling of local appearance and spatial relationships for object recognition. In *CVPR Proceedings, pages 45-51*.

Taffar, M. and Benmohammed, M. (2011). Generic face invariant model for face detection. In *Proc. IP&C Conference Springer, pp 39-45*.

Taffar, M., Miguet, S., and Benmohammed, M. (2012). Viewpoint invariant face detection. In *Networked Digital Technologies, Communications in Computer and Information Science, Springer Verlag, pp:390-402*.

Toews, M. and Arbel, T. (2006). Detection over viewpoint via the object class invariant. In *Proc. Int'l Conf. Pattern Recognition, vol. 1, pp. 765-768*.

Turk, M. and Pentland, A. (1991). Eigenfaces for recognition. In *Journal of Cognitive Neuroscience, vol. 3, pp:71-86*.

Vapnik, V. (1998). Statistical learning theory. In *Wiley Edition, New York*.

Viola, P. and Jones, M. (2001). Rapid object detection using a boosted cascade of simple features. In *Proc. Computer Vision and Pattern Recognition (CVPR), pages 511-518*. Springer.

Wiskott, L., Fellous, J.-M., Kuiger, N., and der Malsburg, C. V. (1997). Face recognition by elastic bunch graph matching. In *IEEE Transactions on PAMI, vol.19, pp:775-779*.

Yang, M.-H., Kriegman, D. J., and Ahuja, N. (2002). Detecting faces in images: A survey. In *IEEE Trans. on PAMI, vol.24, pp:34-58*.

Yu, J., Qin, Z., Wan, T., and Zhang, X. (2013). Feature integration analysis of bag-of-features model for image retrieval. In *Neurocomputing, vol.120, pp:355-364*.