IDIAP RESEARCH REPORT

Robust-to-Illumination Face Localisation using Active Shape Models and Local Binary Patterns

Sébastien Marcel [a] Jean Keomany [a]
Yann Rodriguez [a]

IDIAP–RR 06-47

July 2006

SUBMITTED FOR PUBLICATION

[a] IDIAP Research Institute

IDIAP Research Institute **www.idiap.ch**
Rue du Simplon 4 P.O. Box 592 1920 Martigny – Switzerland
Tel: +41 27 721 77 11 Fax: +41 27 721 77 12 Email: info@idiap.ch

IDIAP Research Report 06-47

ROBUST-TO-ILLUMINATION FACE LOCALISATION USING ACTIVE SHAPE MODELS AND LOCAL BINARY PATTERNS

Sébastien Marcel Jean Keomany Yann Rodriguez

JULY 2006

SUBMITTED FOR PUBLICATION

Abstract. This paper addresses the problem of locating facial features in images of frontal faces taken under different lighting conditions. The well-known Active Shape Model method proposed by Cootes *et al.* is extended to improve its robustness to illumination changes. For that purpose, we introduce the use of Local Binary Patterns (LBP). Three different incremental approaches combining ASM with LBP are presented: profile-based LBP-ASM, square-based LBP-ASM and divided-square-based LBP-ASM. Experiments performed on the standard and darkened image sets of the XM2VTS database demonstrate that the divided-square-based LBP-ASM gives superior performance compared to the state-of-the-art ASM. It achieves more accurate results and fails less frequently.

1 Introduction

Active Shape Model (ASM) is a popular statistical tool for locating examples of known objects in images. It was first introduced by Cootes *et al.* [5] in 1995 and has been developed and improved for many years. ASM is a model-based method which makes use of a prior model of what is expected in the image. Basically, the Active Shape Model is composed of a deformable shape model and a set of local appearance models. The shape model describes the typical variations of an object exhibited in a set of manually annotated images and the local appearance models give a statistical representation of the gray-level structures around each model point. Given a sufficiently accurate starting position, the ASM search attempts to find the best match of the shape model to the data in a new image using the local appearance models. ASM is thus fundamentally similar to Active Contour Model, or snake, proposed by Kass *et al.* [13]. However, ASM has global constraints that allow the shape model to deform only in ways found in the training set. A direct extension of the ASM approach has lead to the Active Appearance Model (AAM) [1]. Besides shape information, the textural information, i.e. the pixel intensities across the object, is included into the model. The AAM algorithm seeks to match both the position of the model points and a representation of the texture of the object to an image.

Although ASM and AAM can be used to find any object in an image, we focus this paper on the detection of facial features such as eyes, nostrils, nose and mouth. Locating such features is an important stage in many facial image interpretation tasks such as face recognition, face tracking or face expression recognition. However, facial feature detection is a challenging task because human faces vary greatly between individuals. Faces can also appear at a wide range of sizes in images and face shape, hair style or glasses can cause the facial features to be occluded. Although good results for facial feature localisation using ASM and AAM have been reported [2], [14], [7], the ability of the model to perform well in different lighting conditions is still limited.

We propose in this paper three incremental approaches combining ASM with Local Binary Patterns (LBPs). LBP is a powerful and computationally simple descriptor of local texture patterns. It expresses the difference of intensity between a given pixel and its neighborhood. LBPs are therefore more robust to illumination changes.

This paper is organized as follows. First, we introduce the original ASM technique proposed by Cootes *et al.*, we describe the LBP approach and present a state-of-the-art method. Then, we present the different approaches proposed to combine ASM with LBP. Finally, experiments and results are presented and a conclusion is drawn.

2 Active Shape Models and Local Binary Patterns

Active Shape Model is a powerful tool for face localisation and alignment. However, the features used to model the local gray-level structures are very sensitive to illumination, particularly when the lighting conditions during search are significantly different from the lighting conditions used to train the shape model.

In this section, we also introduce Local Binary Patterns (LBP) as features for local appearance representations. LBPs are powerful texture descriptors which are much more robust to illumination changes. As far as we know, only Huang *et al.* [11] proposed an improved ASM method based on this idea but they used extended local binary patterns which encode not only the original image but also the gradient magnitude image.

2.1 Active Shape Models

Faces may vary from one image to another due to the identity of the individual, his facial expression, the lighting conditions and the 3D pose (both in plane and out of plane head rotation, scale variation, face location). To locate facial features in an image using Active Shape Models, we must (1) build a model that can describe shapes and typical variations of a face, (2) build local appearance models that represent local gray-level structures and (3) perform the search in the image.

2.1.1 Statistical model of face shape

To build a statistical shape model of a face, we need a set of training images reflecting all possible variations.

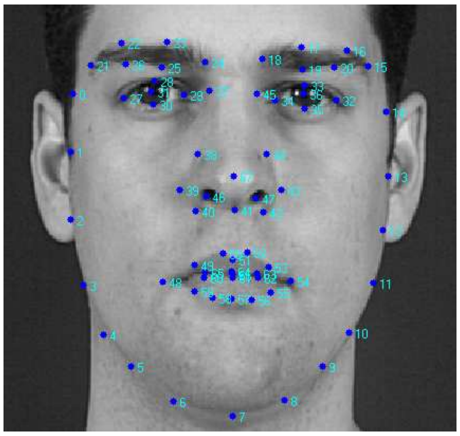


Figure 1: Face image example annotated with 68 landmarks from T.F. Cootes

The shape of a face is represented by a set of n landmark points or landmarks, which may be in any dimension d. Figure 1 shows a face image from the XM2VTS database manually labelled with 68 landmarks. The reader can refer to [4] for more explanations on good choices for landmarks. To have a mathematical representation, the coordinates of each point are concatenated to form a single vector of length $n \times d$. For instance, the n points of a planar shape ($d = 2$) can be represented by the vector **x**:

$$\mathbf{x} = (x_1, x_2, \ldots, x_n, y_1, y_2, \ldots, y_n)^T \quad (1)$$

where (x_j, y_j) are the coordinates of the j^{th} landmark. Given N training images, N such vectors $\mathbf{x_i}$ are then available.

The shape of an object is normally considered to be independent of the scale, orientation and position of that object. Therefore, before any statistical analysis of the training shapes can be performed, variation due to scale, rotation and translation must be removed from the shapes by aligning them into a common coordinate frame. This is achieved using Procrustes Analysis [10].

Each aligned shape can be considered as a single point in the nd dimensional space and the whole training data as a cloud of points in this space. To capture the statistics of the shape variations, we apply Principal Components Analysis (PCA) to the data. PCA computes the eigenvectors $\mathbf{\Phi}= \{\phi_1 \cdots \phi_i \cdots \phi_{nd}\}$ of the covariance matrix of the data. Each eigenvector gives a "mode of variation", a way in which the landmarks tend to move together as the shape varies. Since the landmarks positions are always partially correlated, most of the variation exhibited in the training set can usually be explained by a small number of modes, t. Hence, the dimension of the model can be reduced. The number of eigenvectors to retain can then be chosen so that the model represents a certain percentage (e.g. 98%) of the total variance given by the sum of all the eigenvalues.

PCA allows then each shape **x** in the training set to be approximated (Equation 2) using the mean shape $\bar{\mathrm{x}}$ and a small number of parameters **b**.

$$\mathbf{x} \approx \bar{\mathbf{x}} + \mathbf{\Phi}\mathbf{b} \tag{2}$$

where **b** is a vector of dimension t ($t < nd$), obtained by projecting **x** into the subspace defined by the eigenvectors.

2.1.2 Statistical Model of Local Appearance

It is necessary to have a local appearance model of the gray-level structures around each landmark. This model will be used during the image search to find the best movement in each region around each point. More specifically, the region examined by the ASM is the normal profile to the shape boundary, passing through each landmark. The best approach according to Cootes is to learn this model from the training set.

2.1.3 Image Search

An initial shape model which is generally the mean shape model is first projected into the image being searched. We assume that we know roughly the position in which the model should be placed. We then use the iterative method proposed by Cootes *et al.* [4] (Algorithm 1). This involves finding the set of shape parameters (**b**) and pose parameters (position, orientation and scale) which best match the model to the image. Shape and pose parameters are altered such that the model moves and evolves in the image plane, hopefully converging to the best possible match of the model to the face image.

Later, Cootes *et al.* [6] proposed a multi-resolution approach which consists in computing and exploring a pyramid of coarse to fine images. The search is performed by first searching at the top level of the pyramid (coarse). When the position of a certain percentage of landmarks does not change significantly, the algorithm is declared to have converged at

Algorithm 1 Active Shape Model Algorithm [4]

1. Examine a region of the image around each point (x_i,y_i) to find the best nearby match for the point,
2. Update the shape and pose parameters to best fit the new found points,
3. Apply constraints to the parameters **b** to ensure plausible shapes,
4. Repeat until convergence.

---

that resolution. For instance, when 95% of the new points are within the central 50% of the search profile, the current shape model is projected into the next image and run to convergence again. The search is stopped when convergence is reached on the lowest level of the pyramid (fine). Multi-resolution improves the efficiency, the robustness and the speed of the algorithm while making it less likely to get stuck on wrong image structures.

2.2 Local Binary Patterns

The LBP operator, first introduced by Ojala *et al.* [16], is a powerful method for analyzing textures. The operator labels the pixels of an image by thresholding the 3×3 neighborhood of each pixel with the center value and considering the result as a binary number (Figure 2).

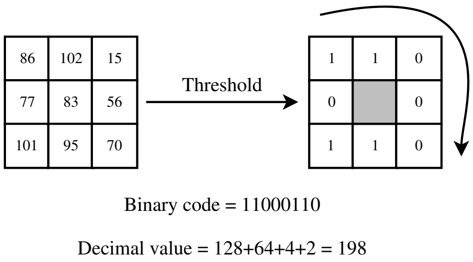


Figure 2: Calculating the original LBP code

At a given pixel position (x_c, y_c), the decimal form of the resulting 8-bit word can be expressed as follows:

$$LBP(x_c, y_c) = \sum_{n=0}^{7} s(i_n - i_c)2^n \tag{3}$$

where i_c corresponds to the gray value of the center pixel (x_c, y_c), i_n to the gray values of the 8 surrounding pixels and function $s(x)$ is defined as:

$$s(x) = \begin{cases} 1 & \text{if } x \geq 0 \\ 0 & \text{if } x < 0 \end{cases} \tag{4}$$

The operator is therefore invariant to monotonic changes in gray-scale and can resist illumination variations as long as the absolute gray value differences are not badly affected. However, the limitation of the original LBP operator comes from its small 3×3 neighborhood which can not capture features with large scale structures.

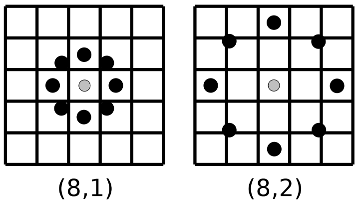


Figure 3: Examples of extended LBP operators

Hence, Ojala *et al.* [17] extended their original LBP operator to a circular neighborhood of different radius size. Figure 3 illustrates examples of extended LBP operators where (P, R) refers to *P* equally spaced pixels on a circle of radius *R*. The value of neighbors that do not fall exactly on pixels, are estimated by bilinear interpolation.

Further extension of LBP is to use uniform patterns [17]. A Local Binary Pattern is called uniform if it contains at most two bitwise transitions from 0 to 1 or vice versa when the binary string is considered circular. For instance, 00000000, 11111001 or 00011111 are uniform patterns. It has been observed that uniform patterns contain most of the texture information. They mainly represent primitive micro-patterns such as spots, lines, edges, corners. The notation $LBP_{P,R}^{u2}$ denotes the extended LBP operator in a (P, R) neighborhood. The superscript ^{u2} indicates that only uniform patterns are used, labelling all remaining patterns with a single label.

Since each bit of the LBP resulting code has the same significance level, two successive bit values may have a totally different meaning. That is the reason why histograms of the labels are used to describe textures.

2.3 Extended Local Binary Patterns ASM

Recently, Huang *et al.* proposed in [11] an ASM method, ELBP-ASM, in which local appearance models of facial landmarks are represented using *Extended* Local Binary Patterns (ELBP).

Huang *et al.* pointed out in their paper that LBP can only reflect the first derivation information of images, but could not represent the velocity of local variations. To solve this problem, they proposed an extended version of Local Binary Patterns that encodes the gradient magnitude image in addition to the original image. Moreover, to retain spatial information, sub-images of landmarks are divided into small regions from which the LBP histograms are extracted and concatenated into a single feature histogram representing the local appearance models (Figure 4). Finally, the mean ELBP histogram $\bar{H}_i$ of each landmark i is computed by summing over all individual LBP histograms (kernels $LBP_{8,1}^{u2}$, $LBP_{8,2}^{u2}$ and $LBP_{8,3}^{u2}$ applied both to the original image and the gradient image).

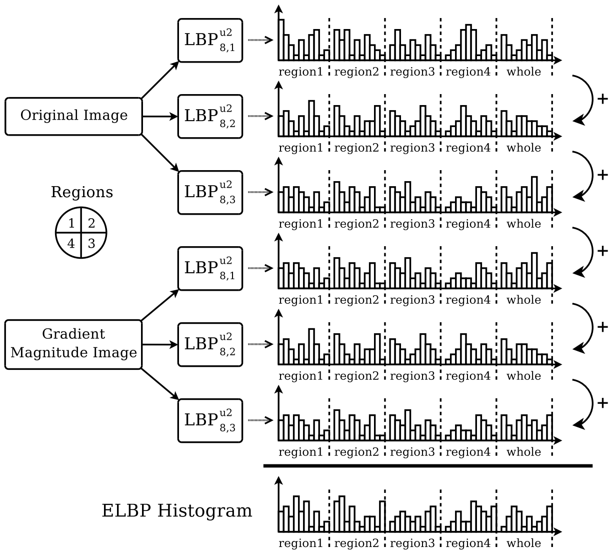


Figure 4: Building an ELBP histogram

During search, ELBP histograms H_j corresponding to each position j located on the normal profile i, are built. They are then compared to the mean histogram $\bar{H}_i$. The dissimilarity between the testing point's histogram H and the mean histogram $\bar{H}$ is calculated using Chi square measure (Equation 5).

$$\chi^2(H, \bar{H}) = \sum_k \frac{(H(k) - \bar{H}(k))^2}{(H(k) + \bar{H}(k))} \tag{5}$$

The smaller the distance is, the more similar the histograms are. The landmark is thus moved to the profile position j^* whose ELBP histogram is the closest to the mean histogram (Equation 6). Similarly to the original ASM, the pose and shape parameters of the shape model are then adjusted to fit the new suggested points, before starting a new iteration.

$$j^* = \arg\max_j \chi^2(H_{ji}, \bar{H}_i) \tag{6}$$

Huang *et al.* reported that ELBP-ASM achieves more accurate results than the original ASM. However, they didn't provide any explicit results on the robustness to illumination using for instance a distinct darken dataset. We also believe that summing up the original image and the gradient magnitude image histograms might not be the most efficient way to take advantage of all the available information. Indeed, the features specific to each image histogram are lost when they are summed together. Moreover, using multi-scale LBP allows to capture the gray-level structures at different scales but it also adds computational load.

Consequently, we believe that even better results can be achieved using simpler methods.

3 The Proposed Approaches

We investigate new methods for modelling the local structures using LBPs. The following sections describe the three incremental approaches.

3.1 Profile-based LBP-ASM

We first propose to use a local appearance descriptor based on the LBP values extracted from the normal profile of each landmark. In this method, $LBP^{u2}_{8,2}$ operator is used. During training, we extract a profile of length n for every point of every training image and build the associated histogram of LBP values. We then compute the mean histogram of each landmark.

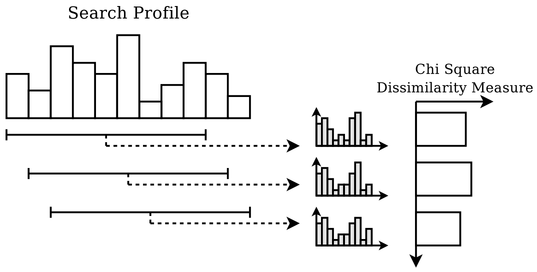


Figure 5: Search using histograms extracted from a profile

During search, we extract for each landmark, a search profile which is longer than the training profile. For each sub-profile of length n contained in the search profile, we build a histogram. The obtained histograms are compared to the corresponding mean histogram using the Chi square dissimilarity measure (Equation 5). The landmark is then moved to the center of the sub-profile which produces the most similar LBP histogram.

This approach is very simple but limited. First, the training profile has to be long enough to provide a sufficient number of points to build a reliable histogram. Since the $LBP^{u2}_{8,2}$ operator produces 59 different labels, the profile has to be at least 59 pixels long to fill the histogram with in average one pixel per bin. However, this condition can hardly be satisfied. Second, comparing the histograms of two consecutive points along the profile does not make any sense since only one point has been replaced from one histogram to the other. These histograms can be considered to be almost identical. To cope with these problems, we propose to build the histogram with the points contained in a square centered at the landmark.

3.2 Square-based LBP-ASM

The local appearance models are complex and it is hard to represent them well only using simple profiles. To capture more information on the local gray-level structures, we use the points which are located within a square centered at a given landmark to build the LBP histogram.

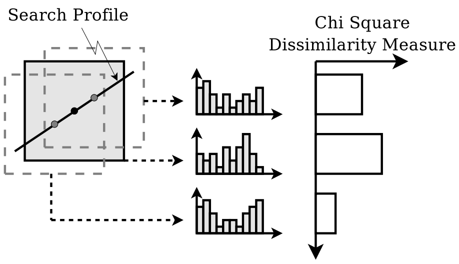


Figure 6: Search using histograms extracted from a square

Basically, the training part is very similar to the one described before but sampling the points in a square region instead of a profile. During search, a LBP histogram is computed in the same manner for each point located on the search profile (Figure 6). The length of the search profile depends in this case only on the distance we allow the landmark to move at each iteration (a few points). The dissimilarity between the testing point's histograms and the mean histogram is also measured using Equation 5.

Hence, this method allows us to model larger structures and fills the histograms with much more LBP values. However, this approach still suffers from the lack of spatial information. Indeed, the main structure we want to detect could be anywhere in the square, the resulting histogram will always look similar. To retain spatial information, we divide the square into small regions as Huang *et al.* did in their work.

3.3 Divided-Square-based LBP-ASM

The square used in the previous method is divided into four regions from which the LBP histograms are extracted and concatenated into a single feature histogram representing the local appearance models (Figure 7).

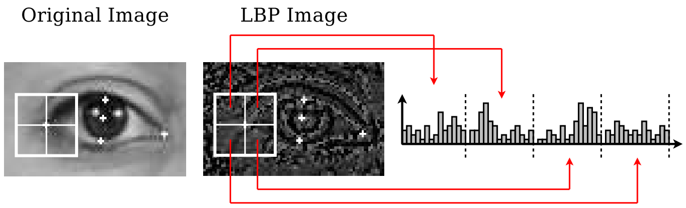


Figure 7: Local appearance representation using a divided square

This representation uses information on three different levels: LBP labels describe the pixel-level patterns, histograms extracted from the small regions provide more spatial information and the concatenated histogram gives a global description of the gray-level structures around each landmark. And last but not the least, this representation is easy to compute.

4 Experiments and Results

This section describes the experiments we performed to compare the performances of the different approaches presented in the previous sections: original ASM, ELBP-ASM, profile-based LBP-ASM, square-based LBP-ASM and divided-square-based LBP-ASM. Each algorithm has been implemented using Torchvision[1] which is an open source machine vision library. The tests have been carried out using the standard and darkened image sets of the XM2VTS database.

4.1 Database

The XM2VTS database [15] consists in face images of 295 subjects collected over four sessions, at one month intervals. It was originally designed for the research and the development of identity verification systems but it has been used also to evaluate performances of facial feature detection algorithms. In this work, we use the frontal face images from the standard set and from the darkened set. For the standard set, two frontal images were recorded for each of the 295 subjects and for each of the four sessions. The 2360 images are at resolution 720×576 pixels. They have been taken under controlled conditions against a flat blue background. The face is large in the image and there is no background clutter. The subjects were volunteers of both sexes and several ethnical origins. Since the data acquisition was distributed over a long period of time, significant variability of appearance of individuals is present in the recordings, such as changes of hair style, facial shape and presence or absence of glasses. Some examples are shown in Figure 8.

Figure 8: Sample images from the standard set

The darkened set contains four frontal views for each of the 295 subjects. In two of the images, the studio light illuminating the left side of the face was turned off. In the other two images, the light illuminating the right side of the face was turned off (Figure 9).

The standard and darkened sets are both supplied with manually located eye center positions. However to enable more detailed testing and model building, the XM2VTS markup (available on Tim Cootes' web site[2]) has been expanded to landmarking 68 facial features on each face of the standard image set. The 68 points chosen are shown in Figure 1. Since

[1] torch3vision.idiap.ch
[2] www.isbe.man.ac.uk/~bim

Figure 9: Sample images from the darkened set

the ground-truth position of these landmarks are not available for the darkened set, tests on this dataset will essentially be based on the eye locations.

4.2 Experimental Setup

From the standard set, training set, evaluation set and test set are built according to the Lausanne protocol [15]. The Lausanne protocol was originally defined for the task of person verification. The standard set is divided into 200 clients, 25 evaluation impostors and 70 test impostors. It exists two configurations that differ in the distribution of client training and client evaluation data. For our experiments, we use the configuration 1 (3 client images both in the training set and in the evaluation set, and 2 client images in the test set).

The training set is used to build the face shape model and the local gray-level structures models. The evaluation set is then used to find the optimal search parameters. Finally, the test set is selected to evaluate the performance of the facial feature detection algorithms. To test the robustness to illumination changes, the detection is performed also on the darkened set using the shape model and search parameters obtained with the standard set.

We assume that the facial feature detection follows a face detection step. The shape model is then initialized according to the estimated eye positions provided by a face detector. For our experiments we implemented the face detector proposed by [9].

4.3 Model Training

From the training set, we build a statistical model for each method described previously: original ASM, ELBP-ASM, profile-based LBP-ASM, square-based LBP-ASM and divided-square-based LBP-ASM. The building process of each model requires the choice of three parameters:

- the number of landmarks,
- the number of modes to use,
- the size of the local appearance descriptors.

The number of landmarks is equal to 68 and the number of modes is chosen so that the model represents 98% of the variance. As a result, 58 modes are retained. For the original

ASM and the profile-based LBP-ASM, 12 pixels along the normal profile are sampled either side of the landmark to build the local appearance model. To simplify the implementation, the ELBP histogram is built using the LBP values contained within a square instead of a disk. The size of the square used in the ELBP-ASM, the square-based LBP-ASM and the divided-square-based LBP-ASM, is set to 25 pixels (12 pixels from the landmark to each side).

4.4 Optimal Search Parameters Estimation

The evaluation set is then used to find the optimal search parameters. Each algorithm requires the choice of four parameters:

- L, the coarsest level of the multi-resolution image pyramid to search,
- n_s, the longest displacement the landmark can make along the search profile,
- it_{max}, the maximum number of iterations allowed at each level,
- q, the proportion of points found determining when to change pyramid level (see Section 2.1.3).

However, we noticed during experiments that the choice of parameters it_{max} and q does not significantly affect the final shape compared to parameters L and n_s. Therefore, in the following tests, the maximum number of iterations allowed at each level is set to 20 and the shape model is projected to a lower level when 95% of points are found within the central 50% of the search profiles.

To measure the quality of fit of the resulting shapes to the ground-truth model, we compute the *Mean Square Error* and estimate the *Point Location Accuracy*.

4.4.1 Mean Square Error

The mean square error (MSE) is given by:

$$MSE = \frac{1}{2n}\sum_{i=0}^{2n}(x_i - gt_i)^2 \tag{7}$$

where n is the number of landmark points ($n = 68$), **x** is the search vector and **gt** is the ground-truth vector.

Figure 10 shows the mean and median of the MSE computed for each algorithm on all images from the evaluation set given different combinations of L and n_s. The median is the value in the middle of the MSE distribution: half the MSE measures are above the median and half are below it. The variances have also been calculated but are not represented on the graphs due to their large values.

We observe that the median is always much smaller than the mean. This indicates that the MSE distribution is highly skewed. MSEs are typically close to 10 when the system converges to a good solution, whereas they can go up to 2000 when the detection fails. Therefore, a small median MSE indicates that the facial feature detection succeeded in most images of the evaluation set. On the other hand, a mean value greater than the median, involves that some large values caused by detection failures, have affected the mean MSE. The median is therefore more appropriate to evaluate the algorithm performances since it is less sensitive to extreme values. The optimal search parameters are consequently given by the combination which produces the smallest median MSE. To validate the choices, we also measure the point location accuracy.

4.4.2 Point Location Accuracy

After search, we measure the distance between the found points and their associated ground-truth position. We then build a frequency histogram for the resulting point-to-target errors. The histograms show the proportion of found points whose point-to-target error lies from 0 (perfect match) to 14 pixels. Any point located further than 14 pixels from its corresponding ground-truth position is considered as a failure. Therefore, we want to maximize the proportion of points close to the target while minimizing the number of detection failures. For each algorithm, the frequency histograms of the four best configurations suggested by MSE statistics are compared with this frequency histogram. This method is more reliable than the median MSE since it provides more information on the whole set of shapes and it is not influenced by convergence failures. As a result, the optimal parameters are also chosen based on this method. Most of the time, they correspond to the combination selected with the median MSE.

Table 1 summarizes the parameters (L and n_s) selected for each algorithm. it_{max} and q are fixed to 20 and 0.95 respectively.

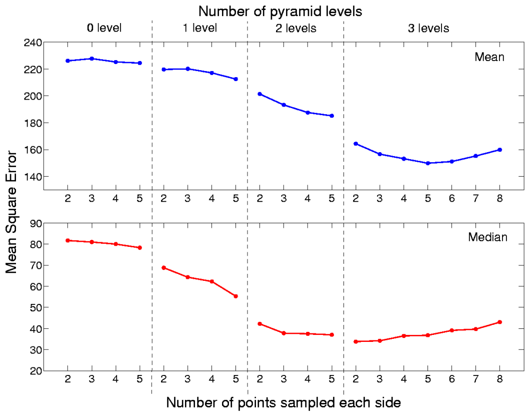
Method	L	n_s
original ASM	3	3
ELBP-ASM	3	4
profile-based LBP-ASM	2	5
square-based LBP-ASM	1	4
divided-square-based LBP-ASM	2	2

Table 1: Optimal search parameters

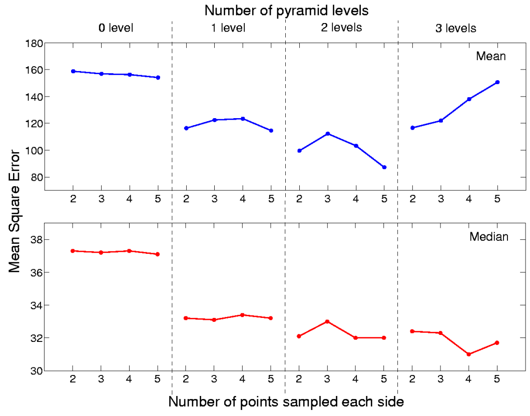
4.5 Evaluation on the Test Set

4.5.1 Mean Square Error

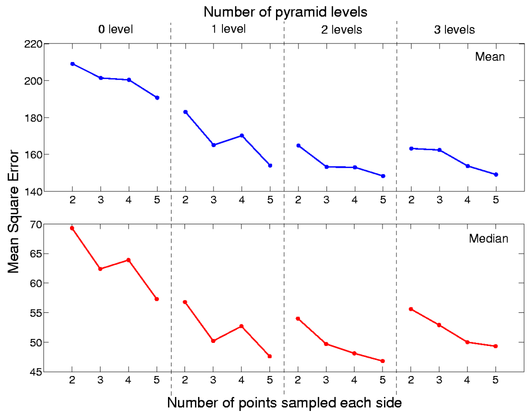
The image search is performed on each image of the test set using the parameters chosen in the evaluation part. Figure 11 shows the mean MSE and median MSE obtained with each algorithm.



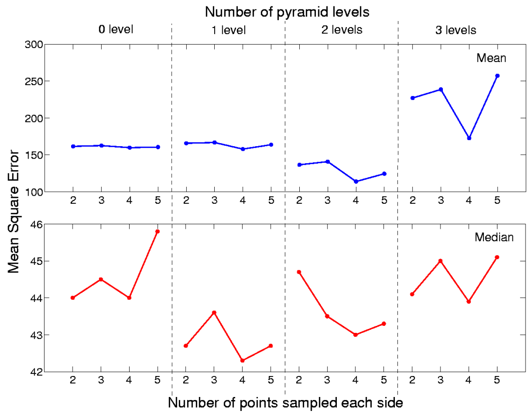
(a)



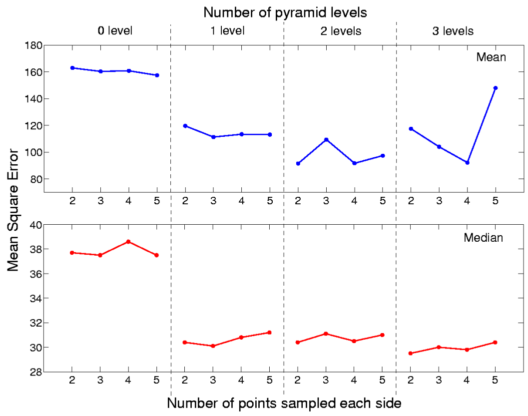
(b)



(c)



(d)



(e)

Figure 10: mean MSE and median MSE on the evaluation set (a) original ASM, (b) ELBP-ASM, (c) profile-based LBP-ASM, (d) square-based LBP-ASM and (e) divided-square-based LBP-ASM

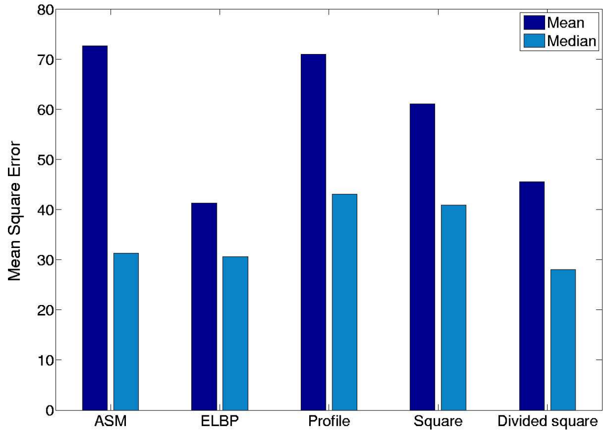


Figure 11: mean MSE and median MSE on the standard test set

The divided-square-based LBP-ASM seems to give better results than the other approaches since it has the smallest median. However, due to the reasons explained in Section 4.4.1, this test cannot be used to draw any final conclusion on the performances of each algorithm. It only gives a first insight.

4.5.2 Point Location Accuracy

The frequency histograms of the point-to-target errors described in Section 4.4.2 are compared in Figure 12.

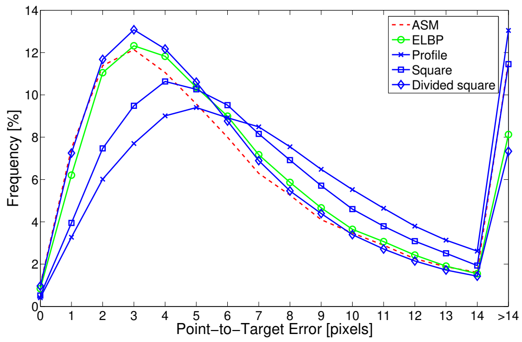


Figure 12: Frequency histograms of point-to-target errors of the standard test set

As expected, the performance of the profile-based LBP-ASM is very limited. LBP histograms extracted from a profile are not reliable local appearance descriptors due to the small number of points they are made of. Using a square region instead of a profile is a good idea but the results of the square-based LBP-ASM show the relevance of retaining spatial information. Indeed, we observe that our proposed method based on a divided square gives much more accurate results and less detection failures than the other approaches. The ELBP-ASM locates the points slightly less accurately than the original ASM but fails less frequently. The small failure rate of the divided-square-based LBP-ASM and the ELBP-ASM is due to the good ability of a square to catch the target gray-level structure within it. Theses two algorithms are then less likely to diverge. We can also notice from Figure 12 the difference of accuracy between our approach and Huang *et al.*'s one. The ELBP histogram gathers too much information that can not be totally exploited during search. As a result, this affects the ELBP-ASM's performance.

4.6 Robustness to Illumination

To test the robustness of each algorithm to illumination changes, the detection is performed on the darkened set using the shape model and search parameters obtained with the standard set. Facial feature localisation is particularly difficult in this case because the lighting conditions during search are considerably different from the lighting conditions used to train the shape model.

Since only the ground-truth eye center positions are available for this set of images, the quality of fit is assessed using the eye location accuracy and the Jesorsky's measure [12]. Let C_l (respectively C_r) be the true left (resp. right) eye coordinate position and let $\tilde{C}_l$ (resp. $\tilde{C}_r$) be the left (resp. right) eye position estimated by the facial feature detector. Jesorsky's measure can be written as

$$d_{eye} = \frac{max(d(C_l, \tilde{C}_l), d(C_r, \tilde{C}_r))}{\|C_r - C_r\|} \tag{8}$$

where $d(a, b)$ is the Euclidean distance between positions a and b. A successful localisation is accounted if $d_{eye} < 0.25$ (which corresponds approximately to half the width of an eye).

Figure 13 presents the mean and the median of the Jesorsky's measure derived from the standard test set and the darkened set. Figure 14 shows the frequency histogram of the point-to-target errors corresponding to the eye center positions computed on the darkened images.

In Figure 13 and 14, the detector's values correspond to the measures obtained after the face detection stage (before facial feature detection). As expected, the original ASM, the ELBP-ASM and the divided-square-based LBP-ASM improve significantly the Jesorsky's measure for the standard test set. However, we can see that ELBP-ASM completely fails on the darkened set. The ELBP histogram is based on 6 images: the $LBP^{u2}_{8,1}$, $LBP^{u2}_{8,2}$, $LBP^{u2}_{8,2}$ of the original image and the $LBP^{u2}_{8,1}$, $LBP^{u2}_{8,2}$, $LBP^{u2}_{8,2}$ of the gradient magnitude image. When lighting conditions change, each image is degraded in a different way. Therefore, the ELBP histogram obtained by summing up the six LBP histograms is considerably different from the mean histogram trained on the standard set. Then, the ELBP algorithm diverges more frequently, which does not happen with the proposed approaches. We observe in Figure 14, that the square-based LBP-ASM and the divided-square-based LBP-ASM are more robust to illumination changes than the original ASM. Indeed, the eye localisation failure rates are much lower.

When the facial feature localisation is used for face recognition, it is important to accurately locate the eye center positions. However, in other applications, minimizing the

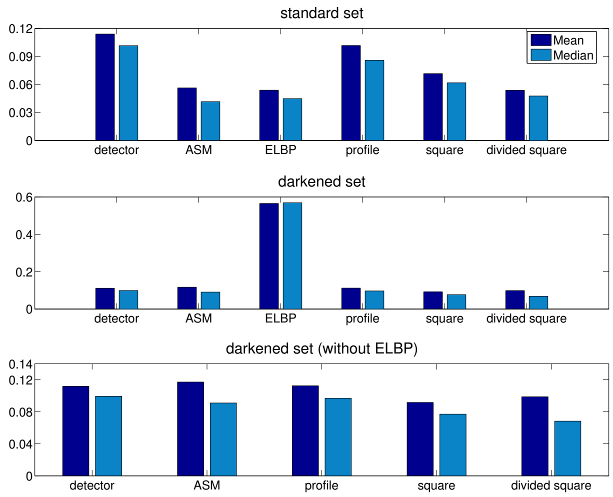


Figure 13: Mean and median of the Jesorsky's measure on the standard test set and the darkened set

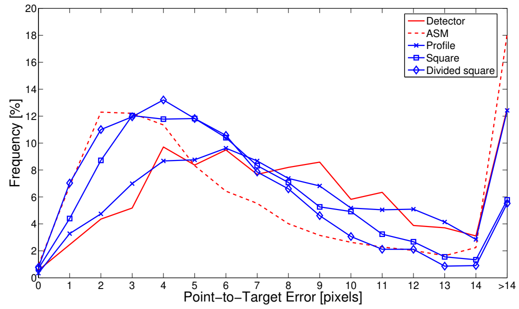


Figure 14: Frequency histograms of point-to-target errors corresponding to the eye center positions computed on the darkened set

Jesorsky's measure is not sufficient. Indeed, the Jesorsky's measure expresses only partially the quality of fit. The system can properly locate the eye center and fail on the other facial features. To perform more detailed tests, it would have been useful to annotate the 1160 darkened images with the same 68 landmark points. Unfortunately, it could not be done during this work due to time constraints. Figure 15 shows examples of search on a darkened image using the original ASM, the ELBP-ASM and the divided-square-based LBP-ASM. We can observe that the facial feature localisation performed by the divided-square-based LBP-ASM is the most accurate whereas the Jesorsky's measure is not the lowest.

Method	Computation time (s)	# of iterations
original ASM	2.3	12.6
ELBP-ASM	29	13.4
profile-based LBP-ASM	5.3	38.9
square-based LBP-ASM	4.4	14.6
divided-square-based LBP-ASM	7.4	23.4

Table 2: Computation times and average numbers of iterations

4.7 Computation Times

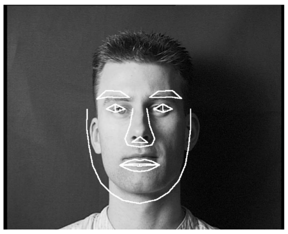
Table 2 summarizes the computation times and the average number of iterations that the five algorithms need to converge. Experiments were performed on a 1GHz PC with 1GB memory.

5 Conclusion and Future Work

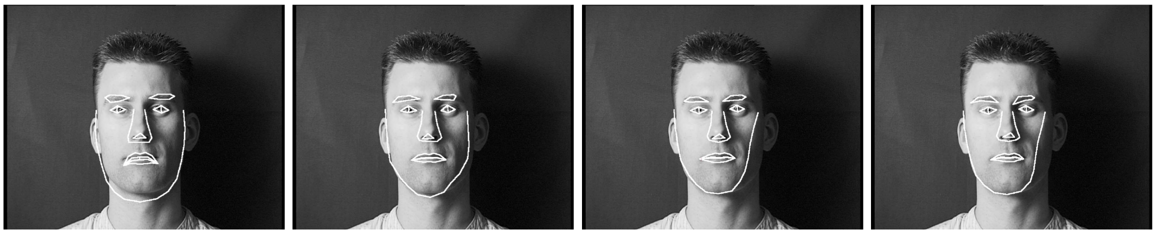
In this paper, we extended the Active Shape Model method proposed by Cootes *et al.* to improve its robustness to illumination changes. Three incremental approaches using Local Binary Patterns to model the structures around each landmark point were proposed.

In the **profile-based LBP-ASM** method, the local appearance models are described using LBP histograms extracted from the normal profile of each landmark. Similar to the original ASM, this method suffers from the limited ability of normal profiles to describe complex structures. In the **square-based LBP-ASM** method, the local structures are modelled using LBP histograms extracted from a square region around each landmark. This method captures more visual information but does not retain spatial information. With the **divided-square-based LBP-ASM**, the square region used in the square-based LBP-ASM is divided into four regions from which the LBP histograms are extracted and concatenated into a single feature histogram representing the local appearance models.

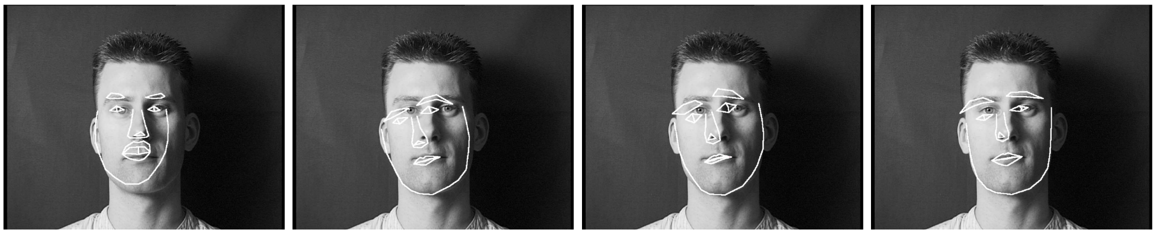
Experiments were performed to compare those three approaches with the original ASM and the only method combining ASM and LBP exiting so far, ELBP-ASM. The tests were carried out using the standard and darkened sets of the XM2VTS database. Experiments



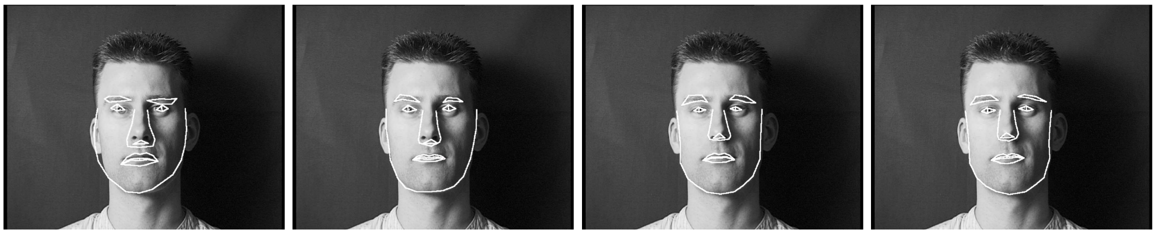
(a) Initial Condition. Jesorsky's measure before facial feature detection = 0.181623



(b) ASM: iteration 1, 4, 8 and 13. Jesorsky's measure = 0.023976



(c) ELBP: iteration 1, 16, 25 and 32. Jesorsky's measure = 0.241385



(d) divided-square-based LBP-ASM: iteration 1, 5, 10 and 19. Jesorsky's measure = 0.039618

Figure 15: Example of search on a darkened image using the original ASM, the ELBP-ASM and the divided-square-based LBP-ASM

on the standard set demonstrated that the divided-square-based LBP-ASM achieves more accurate results and fails less frequently than the other approaches. The accuracy can still be improved by using more landmarks. Indeed, 68 landmarks were used whereas facial feature localisation is usually performed using at least 133 landmarks.

Experiments on the darkened set have shown clearly the robustness to illumination changes of most of the proposed algorithms. Since only the eye center ground-truth positions were available, tests were based on the Jesorsky's measure. As expected, the divided-square-based LBP-ASM is the most robust to illumination changes. However, we showed through an example that a large Jesorsky's measure does not mean that the facial feature detection failed completely.

Therefore, although the results look very promising, more experiments still have to be done before drawing any final conclusion. A straightforward continuation of this work would be to extend the LBP divided-square-based method to Active Appearance Model.

References

[1] T.F. Cootes and G.J. Edwards and C.J. Taylor, "Active Appearance Models", *European Conference on Computer Vision*, Vol.2, pp.484–498, 1998.

[2] T.F. Cootes and G.J. Edwards and C.J. Taylor, "Comparing Active Shape Models with Active Appearance Models", *British Machine Vision Conference*, Vol.1, pp.173–182, 1999.

[3] T.F. Cootes and A. Hill and C.J. Taylor and J. Haslam, "The Use of Active Shape Models for Locating Structures in Medical Images", *International Conference on Information Processing in Medical Imaging*, pp.33–47, 1993.

[4] T.F. Cootes and C.J. Taylor, "Statistical models of appearance for computer vision", *Technical Report – University of Manchester*, March 2004.

[5] T.F. Cootes and C.J. Taylor and D. Cooper and J. Graham, "Active Shape Models – their training and applications", *Computer Vision and Image Understanding*, Vol.61, No.1, pp.38–59, 1995.

[6] T.F. Cootes and C.J. Taylor and A. Lanitis, "Active Shape Models: evaluation of a multi-resolution method for improving image search", *British Machine Vision Conference*, pp.327-336, 1994.

[7] D. Cristinacce, "Automatic Detection of Facial Features in Grey Scale Images", *PhD. Thesis – University of Manchester*, 2004.

[8] I.L. Dryden and K.V. Mardia, "Statistical Shape Analysis", *Wiley Series in Probability and Statistics*, 1998.

[9] B. Froba and A. Ernst, “Face detection with the Modified Census Transform”, *Proceedings of the Automatic Face and Gesture Recognition Conference*, pp.91–96, 2004.

[10] C. Goodall, “Procrustes methods in the statistical analysis of shape”, *Journal of the Royal Statistical Society Series B*, Vol.53, No.2, pp.285–339, 1991.

[11] X. Huang and S. Li and Y. Wang, “Shape Localization based on Statistical Method using Extended Local Binary Pattern”, *International Conference on Image and Graphics*, pp.184–187, 2004.

[12] O. Jesorsky and K.J. Kirchberg and R.W. Frischholz, “Robust Face Detection Using the Hausdorff Distance”, *International Conference on Audio and Video-Based Biometric Person Authentication*, Vol.2091, pp.90–95, 2001.

[13] M. Kass and A. Witkin and D. Terzopoulos, “Snakes: Active Contour Models”, *International Conference on Computer Vision*, pp.259–268, 1987.

[14] A. Lanitis and C.J. Taylor and T.F. Cootes and T. Ahmed, “Automatic Face Identification System Using Flexible Appearance Models”, *Image and Vision Computing*, Vol.13, No.5, pp.392–401, 1995.

[15] K. Messer and J. Matas and J. Kittler and J. Luettin and G. Maitre, “XM2VTSDB: The Extended M2VTS Database”, *International Conference on Audio and Video-based Biometric Person Authentication*, pp.72–77, 1999.

[16] T. Ojala and M. Pietikäinen and D. Harwood, “A comparative study of texture measures with classification based on feature distributions”, *Pattern Recognition*, vol.29, pp.51–59, 1996.

[17] T. Ojala and M. Pietikäinen and T. Mäenpää, “Multiresolution gray-scale and rotation invariant texture classification with local binary patterns”, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol.24, pp.971–987, 2002.