Active Shape Models with Invariant Optimal Features: Application to Facial Analysis

Federico M. Sukno, Sebastián Ordás, Constantine Butakoff, Santiago Cruz, and Alejandro F. Frangi, *Senior Member*, *IEEE*

Abstract—This work is framed in the field of statistical face analysis. In particular, the problem of accurate segmentation of prominent features of the face in frontal view images is addressed. We propose a method that generalizes linear Active Shape Models (ASMs), which have already been used for this task. The technique is built upon the development of a nonlinear intensity model, incorporating a reduced set of differential invariant features as local image descriptors. These features are invariant to rigid transformations, and a subset of them is chosen by Sequential Feature Selection for each landmark and resolution level. The new approach overcomes the unimodality and Gaussianity assumptions of classical ASMs regarding the distribution of the intensity values across the training set. Our methodology has demonstrated a significant improvement in segmentation precision as compared to the linear ASM and Optimal Features ASM (a nonlinear extension of the pioneer algorithm) in the tests performed on AR, XM2VTS, and EQUINOX databases.

Index Terms—Face and gesture recognition, feature evaluation and selection, invariants, shape model, statistical image analysis.

1 Introduction

In many automatic systems for face analysis, following the stage of face detection and localization and before face recognition is performed, prominent facial features must be extracted. This process currently occupies a large area within computer vision research.

A human face is part of a smooth 3D object mostly without sharp boundaries. It exhibits an intrinsic variability (due to identity, gender, age, hairstyle, and facial expressions) that is difficult, if not impossible, to characterize analytically. Artifacts such as make-up, jewelery, and glasses cause further variation. In addition to all these factors, the observer's viewpoint (in-plane or in-depth rotation of the face), the imaging system, the illumination sources, and other objects present in the scene, may affect the overall appearance. All these intrinsic and extrinsic variations make the segmentation task difficult and hamper a search for fixed patterns in facial images. To overcome these limitations, statistical learning from examples is becoming popular in order to characterize, model, and segment prominent features of the face.

An Active Shape Model (ASM) is a flexible methodology that has been used for the segmentation of a wide range of objects, including facial features, e.g., [1]. In the seminal approach of Cootes et al. [2], shape statistics were computed from a training set of shapes and local gray-level profiles (normalized first-order derivatives) were used to capture the local intensity variations at each landmark point. In [3], Cootes et al. introduced another powerful approach to deformable template models, namely, the Active Appearance Model (AAM). In AAMs, a combined PCA of the landmarks and pixel values inside the object is performed. The AAM handles a full model of appearance, which represents both shape and texture variation.

While AAM has its own benefits, like the ability to model texture variation, ASM is fast, mainly due to the simplicity of its texture model. The latter is constructed with just a few pixels around each landmark whose distribution is assumed to be Gaussian and unimodal. This simplicity, however, turns into weakness when complex textures must be analyzed. In practice, local gray levels around the landmarks can have large variations and pixel profiles around an object boundary are not very different from those in other parts of the image. To provide a more elaborated intensity model, van Ginneken et al. [4] proposed the Optimal Features ASM (OF-ASM). It is nonlinear and allows for multimodal distribution of intensities, since it uses k-nearest neighbor (kNN) classification of local texture descriptors (*jets*), based on image derivatives. Wiskott et al. [5] also use local jets to construct their Elastic Bunch Graph. The latter are based on Gabor kernels and constrain the shape variations by an elastic model rather than by a statistical point distribution model as in ASMs. Although this method is mainly oriented toward an in-class recognition task, it can be also applied to obtain segmentations of facial images, achieving satisfactory results.

The main contribution of the OF-ASM was an increased accuracy in the segmentation task, which has been shown to be particularly useful in segmenting objects with textured boundaries in medical images. However, its application to facial images is not straightforward: Facial images have a more complex geometry of embedded shapes and present large texture variations for the same *region* across different individuals. In this work, we will

- *F.M. Sukno, S. Ordás, C. Butakoff, and A.F. Frangi are with the Department of Technology, Pompeu Fabra University, Passeig de Circumvallacio 8, (08003) Barcelona, Spain. E-mail: {federico.sukno, sebastian.ordas, constantine.butakoff, alejandro.frangi}@upf.edu.*
- *S. Cruz is with the Aragon Institute of Engineering Research, University of Zaragoza, Maria de Luna 3, (50018) Zaragoza, Spain. E-mail: cruzll@unizar.es.*

Manuscript received 30 Dec. 2005; revised 31 July 2006; accepted 2 Oct. 2006; published online 16 Jan. 2007.
Recommended for acceptance by S. Baker.
For information on obtaining reprints of this article, please send e-mail to: tpami@computer.org, and reference IEEECS Log Number TPAMI-0744-1205.
Digital Object Identifier no. 10.1109/TPAMI.2007.1041.

discuss those problems and develop modifications to the model in order to make it deal with facial complexities. The OF-ASM derivatives will also be replaced so that the intensity model is invariant to rigid transformations. The new method, coined Invariant Optimal Features ASM (IOF-ASM) [6], also tackles the problem of the segmentation speed, which was a drawback in OF-ASM [4]. It will be shown that IOF-ASM offers the possibility to trade off between segmentation accuracy and speed. The performance of our method will be compared against both the original ASM and the OF-ASM, using the AR [7], XM2VTS [8], and Equinox [9] databases as test beds. Experiments were split into segmentation accuracy and identity verification tests, based on the Lausanne protocol [8].

The remainder of this paper is organized as follows: In Section 2, we briefly describe the underlying theory of ASM and OF-ASM approaches, while in Section 3, the proposed IOF-ASM is presented. In Section 4, we describe the materials and methods for the evaluation and show the results of our experiments, which are discussed in Section 5. Section 6 summarizes and concludes the paper.

2 Previous Approaches

2.1 Linear ASM

In its original form [2], ASM is built from sets of prominent points known as *landmarks* [2] by computing a Point Distribution Model (PDM) and a local image intensity model around each of those points.

The PDM is constructed by applying PCA to the aligned set of shapes, each represented by landmarks. The original shapes $\mathbf{u}_i$ and their model representation $\mathbf{b_i}$ $(i = 1, \ldots, N)$ are related by the mean shape $\overline{\mathbf{u}}$ and the eigenvector matrix $\mathbf{\Phi}$:

$$\mathbf{b}_i = \mathbf{\Phi}^T(\mathbf{u}_i - \overline{\mathbf{u}}), \qquad \mathbf{u}_i = \overline{\mathbf{u}} + \mathbf{\Phi}\mathbf{b}_i. \tag{1}$$

To decrease the dimensionality of the representation, it is possible to use only the eigenvectors corresponding to the largest eigenvalues. In that case, (1) becomes an approximation, with an error depending on the magnitude of the excluded eigenvalues. Furthermore, under the assumption of Gaussianity, each component of the $\mathbf{b}_i$ vectors is constrained to ensure that only *valid shapes* are represented:

$$|b_i^m| \leq \beta\sqrt{\lambda_m} \qquad 1 \leq i \leq N, \quad 1 \leq m \leq M, \tag{2}$$

where β is a regularization constant, usually set between 1 and 3, according to the degree of flexibility desired in the shape model, M is the number of retained eigenvectors, and λ_m are the eigenvalues of the covariance matrix.

The intensity model is constructed by computing second-order statistics for the normalized image gradients, sampled at each side of the landmarks, perpendicularly to the shape's contour, hereinafter, the profile. In other words, the profile is a fixed-size vector of values (in this case, pixel intensity values) sampled along the perpendicular to the contour such that the contour passes right through the middle of the perpendicular. The matching procedure is an alternation of image driven landmark displacements and statistical shape constraining based on the PDM, as shown in (2). It is usually performed in a multiresolution fashion in order to enhance the capture range of the algorithm. The landmark displacements are individually determined using the intensity model, by minimizing the Mahalanobis distance between the candidate gradient and the model's mean.

2.2 Optimal Features ASM

As an alternative to the construction of normalized gradients and the use of the *Mahalanobis* distance as a cost function, van Ginneken et al. [4] proposed a nonlinear intensity model constructed from local image descriptors. Each pixel on the image was assigned a set of such descriptors (or *features*). Then, during matching, the landmark points are displaced along the perpendiculars to the current shape estimates to find the contours of the object of interest. However, the best displacement now will be the one for which everything on one side of the profile is classified as being outside the object, and everything on the other side, as inside of it. Optimal Features ASMs (OF-ASMs) use image derivatives as local image descriptors. The idea behind the choice of such descriptors is the fact that a function can be locally approximated by its Taylor series expansion provided that the derivatives at the point of expansion can be computed up to a sufficient order. The number of image features is reduced by sequential feature selection [10] and interpreted by a kNN classifier with weighted voting [11], to cope with the non-linearity of the texture.

3 Invariant Optimal Features ASM

This work concentrates on a generalization of OF-ASM called IOF-ASM. The modifications that we introduce in our method can be summarized as follows:

- The structure of the sampling points, as well as their interpretation, has been completely revisited (Section 3.3). As it will be explained in Section 3.4, this provides robustness with respect to shape complexities of the face.
- The local texture descriptors are replaced by irreducible Cartesian differential invariants, making the intensity model invariant to rigid transformations (Section 3.1).
- The kNN classifiers are replaced by multivalued neurons (MVN) [12], [13]. The MVN is a very fast classifier whose speed is independent of the number of training samples, as opposed to kNNs (Section 3.2).
- During the construction of the model, different feature selection strategies can be used to tune the model toward segmentation accuracy or speed, or a compromise thereof (Section 3.6).

3.1 Irreducible Cartesian Differential Invariants

A limitation of using the derivatives in a Cartesian framework, as features in the OF-ASM approach, is the lack of invariance with respect to translation and rotation (rigid transformations). Consequently, these operators can only cope with textured boundaries of the same orientations as those seen in the training set. To overcome this issue, we

TABLE 1
Tensor and Cartesian Formulation of Invariants

Tensor Formulation	2D Cartesian Formulation
L	L
L_{ii}	$L_{xx} + L_{yy}$
$L_i L_i$	$L_x^2 + L_y^2$
$L_i L_{ij} L_j$	$L_x^2 L_{xx} + 2L_{xy} L_x L_y + L_y^2 L_{yy}$
$L_{ij} L_{ji}$	$L_{xx}^2 + 2L_{xy}^2 + L_{yy}^2$

introduce a multiscale feature vector that is invariant under 2D rigid transformations.

Cartesian differential invariants describe the differential structure of an image independently of the chosen Cartesian coordinate system [14], [15], [16]. The term irreducible is used to indicate that any other algebraic invariant can be reduced to a linear combination of elements of this minimal set. Table 1 shows the (linear) Cartesian invariants up to second order. Notice that, for the tensor formulation, the Einstein convention is used. Hence, when an index variable appears twice in a single term, a summation over all of its possible values takes place. In our context, the index variables are i and j, which can take values x or y each, the latter corresponding to the image axes.

The use of these invariants as the basis for texture description makes IOF-ASM invariant to rigid transformations. In this work, we will use first and second-order linear invariants at three different scales, $\sigma = 1$, 2, and 4 pixels. The zero-order invariants (which correspond to the raw images seen at different Gaussian blurred scales) were not used since the differential images are expected to provide a more accurate and stable information about facial contours (edges). For example, the zero-order invariant could make the texture model dependant on undesirable features, such as the color of the background surrounding the face.

3.2 Multivalued Neural Network

In our approach, we used a nonlinear classifier in order to label image points near a boundary or contour. Among the many available options, we have chosen the Multivalued Neurons (MVNs) mainly based on the need to improve segmentation speed. These are very fast classifiers, since their decision is based only on a vector multiplication in the complex domain. Furthermore, a single neuron is enough to deal with nonlinear problems [12], [13], which avoids the need for carefully tuning the number of layers (and neurons in each of them) that characterizes multilayer perceptron networks.

In our approach, a MVN is assigned to each landmark, with as many inputs as the number of features selected for that landmark. All of the inputs are mapped to a complex number on the unit circle, and the argument of their weighted sum is the activation function, after appropriate scaling in order to deal with different magnitudes of the features. The number of output sectors will depend on the chosen profile size (see Section 3.3). The reader is referred to [13] for further details.

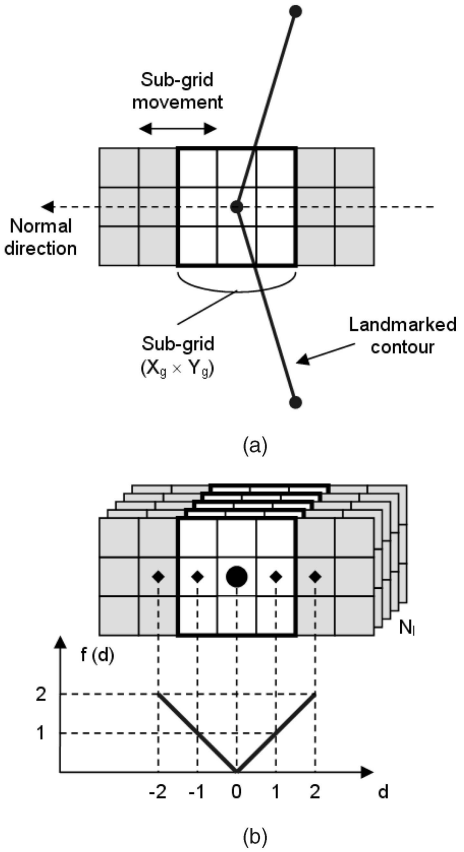

Fig. 1. (a) Example of sampling with five 3 × 3 subgrids and (b) the corresponding labeling, based on the displacement (d) of their center from the landmark. Notice that the labeling function is $f(d) = |d|$.

3.3 Construction of the Intensity Model

The intensity model of IOF-ASM is based on image invariants. During the construction of the model, every image of the training set is processed in order to generate a set of *invariant images*. By an invariant image, we mean that one of the invariants in Table 1 is computed at a specific scale for the whole image. Using these precomputed invariant images, an individual intensity model is constructed for every landmark and resolution level.

The construction of the intensity model for each landmark starts by placing a rectangular grid centered at the landmark position. All invariant images are sampled at the positions defined by this grid, generated by displacing a smaller grid (subgrid) a predefined number of positions toward each side of the landmark. Fig. 1 illustrates the concept: The $X_g \times Y_g$ (3 × 3 in the plot) subgrid can take five positions, since we allow its center to depart from the landmark up to two pixels on each side, along the normal to the object's contour. We call the positions taken by the centers of the subgrids as *main grid*, of size $X_G \times Y_G$ (5 × 1 in the drawn example). Notice that the total sampling region covered by the subgrids is $(X_G + X_g - 1) \times Y_g$ (resulting 7 × 3 pixels in Fig. 1). The sampled values are normalized to zero mean and unit variance across the whole sampling region to reduce the influence of global illumination.

So far, for a number of positions in the neighborhood of each landmark, we have the pixel intensities sampled by the subgrid across all the (preprocessed) invariant images. That

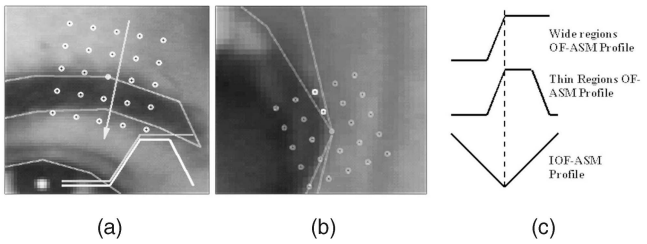

Fig. 2. (a) A typical eyebrow image and a 5 × 5 grid with the arrow indicating the normal to the contour. (b) The same grid over the mouth corner, where only three points lie inside the lip. (c) The plots of the typical profiles for OF- and IOF-ASM for the eyebrow image showed in (a).

is, the extracted data for each position of the main grid can be thought of as a cuboid of size $X_g \times Y_g \times N_I$, where X_g, Y_g are the dimensions of the subgrid and N_I is the number of invariant images. Each of the cuboids is labeled according to the distance from its center to the landmark, as shown in Fig. 1b. The typical plot of the labels as a function of the subgrids displacements will take a shape of letter "V." Its vertex will be located at the landmark position.

After labeling all of the training images, the labeled cuboids are used to train their corresponding MVN *texture classifier*. Note that for each landmark, there is an independent MVN with $\frac{1+X_G}{2}$ output sectors.[1] When used for classification, the MVN will return the distance to the most likely position for the landmark (according to its training) in the (continuous) interval defined by $[-\frac{1}{2}; 1 + \frac{X_G}{2})$ (notice that the interval of interest is $[0; \frac{1+X_G}{2}]$ and there is a $\frac{1}{2}$ extension to each side because integer-valued labels are mapped to the centers of the MVN sectors [13]).

3.4 Shape Complexities

Let us revisit for a moment the OF-ASM. Its training is based on a landmarked set of images for which all derivative images were computed and described by local (histogram) statistics. Once the texture classifiers are trained, they would be able to classify a point as inside or outside the region of interest based on the texture descriptors (the features). Therefore, labeling inside pixels with 1 and outside pixels with 0 and plotting the labels corresponding to the profile pixels, the classical step function is obtained, and the transition will correspond to the landmark position.

Nevertheless, there are a couple of reasons why this may not happen. The first one is that certain regions of the object can be thinner than the size of the grid, and then the correct labeling of the points would look more like a bar rather than like a step function. An illustrative example arises when the square grid is placed over the eye or eyebrow contours (Fig. 2). Moreover, in a multiresolution framework, image subsampling contributes to "step over" these structures. Another problem is that the classifiers will not make a perfect decision, so the labeling will look much noisier than the ideal step or bar. Additionally, Fig. 2 illustrates how, for certain landmarks where there is a high contour curvature

1. Since the subgrids are made symmetric with respect to their center, X_G is an odd integer and, therefore, the resulting number of output sectors is also integer.

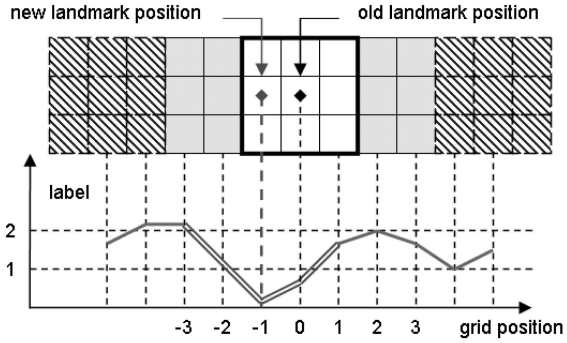

Fig. 3. Looking for new landmark positions during model matching. A number of subgrids is sampled over the image and classified by the MVN. The bottom plot shows the labels assigned to the subgrids for each position. The landmark point will be displaced to the position that best fits to the V-shape.

(i.e., mouth/eyes/eyebrows corners), most of the grid points would lie outside the contour, promoting quite an unbalanced training of the classifiers.

The IOF-ASM has been designed to deal with these problems: Once the learning process is completed, the MVNs should be able to tell the distance of a given cuboid with respect to the correct landmark position. Therefore, the typical plot of the profiles will be a "V," with its minimum (the vertex) located at the landmark position, irrespective of which region is sampled and its width relative to the grid size.

3.5 Model Matching

During matching, the best fit to the "V" shape is searched for at every landmark. The subgrids are moved over a more extensive area than during training, to allow selecting the best V from several *candidate profiles*. An example of this is shown in Fig. 3, where the main grid is allowed to move three pixels to each side of the current landmark position. Thus, there are seven main grids in the plot, and each of them contains five subgrids (so, the profile to search for is a five-pixel wide V). The outputs of the texture classifier (the labels for each subgrid) show that the group of points that best fits the V is centered at -1 (indicated with a double line in Fig. 3), one pixel away from the current landmark position. That would be the updated position for that landmark.

The fact that all subgrids are labeled with the distance to the landmark allows for introducing a robust estimation metric [17]. The best position for the landmark is now the one which minimizes the profile distance to the ideal V, excluding the *outliers*. In this context, an outlier is a point on the profile whose distance to corresponding point on the ideal V is greater than one. The later can be easily understood by noticing that such a point is suggesting a different position to place the landmark (i.e., its distance would be smaller if the V is adequately displaced). If the number of outliers exceeds 1/3 of the profile size, then the image model is regarded as not trustworthy and the distance for that position is set to infinity. Otherwise, preference is given to the profiles with fewer outliers. The objective of matching the intensity model is to find such a k that minimizes

$$f(k) = N_{OL}(k) + \frac{\sum_{i=1}^{N_P - N_{OL}(k)} |p_i(k) - v_i|}{N_P - N_{OL}(k)}, \quad (3)$$

where k are the different candidate positions for the landmark, $N_{OL}(k)$ is the number of outliers, N_P is the profile size, and p_i and v_i are the ith components of the kth input profile, $\mathbf{p}(k)$, and ideal (V) profile, respectively.

3.6 Feature Selection

The computational load of the segmentation process can be associated to two main tasks: the computation of invariant images and the iterations of the matching procedure. The invariants computation time is proportional to the image size: Its complexity is roughly $O(N_I \times I_H \times I_W)$, where I_H and I_W are the height and width of the image, respectively. On the other hand, the iterative process complexity is proportional to the number of landmarks, the number of features, and the number of iterations. Notice that the (total) number of features is the size of the subgrids multiplied by the number of invariants. Thus, it is clear that selecting a subset of the available features will speed-up the segmentation. Additionally, if this selection determines that some invariant images will not be used at all by any landmark, then there will be a further speed-up by skipping their precomputation.

Among feature selection methods, wrapper greedy search seems to be particularly computationally advantageous and robust to overfitting [18]. In IOF-ASM, we use a Sequential Backwards Selection (SBS) [10] without retraining the classifier while evaluating the features to exclude. This is combined with a voting strategy to determine the exclusion of the same invariants for all landmarks at once, thus avoiding their calculation at the preprocessing stage of matching. The algorithm, which is executed independently for each resolution level, is outlined below. The iterations start with all of the available invariants, N_I, excluding one invariant at each step until it reaches N_{IF}, the desired (*final*) number of invariants:

Algorithm 1 Joint feature selection with no retraining

```
1: n = N_I
2: while n > N_IF do
3:   Initialize voting array to zero
4:   for l = 1 to number_of_landmarks do
5:     Train texture classifier (with n invariants)
6:     Compute classifier score
7:     for i = 1 to n do
8:       Compute score without ith invariant
9:     end for
10:    Update voting array based on saved scores
11:  end for
12:  Eliminate most voted invariant
13:  n = n − 1
14: end while
```

Let us concentrate on lines 6 to 10. The learning process for a MVN can involve some thousands of passes through all training samples while the classification score is computed in a single pass, so its computational load can be considered negligible. Then, if the texture classifiers are not retrained while deciding which invariant to discard, the solution is suboptimal but much faster. While the feature selection does not constitute a step of the matching algorithm, it is of practical necessity to speed-up this process. The improvement due to the absence of retraining at line 8 is:

$$\eta \simeq \frac{SBS_{NR} \quad speed}{SBS \quad speed} = \sum_{n=N_{IF}}^{N_I} n - 1, \quad (4)$$

where SBS_{NR} stands for SBS with no retraining. That is, at each elimination step, there is only one retraining of the classifier (with the current set of features, at line 5) instead of one retraining per feature.[2] It can be seen that the speed improvement grows easily by one or two orders of magnitude even for small sets of invariants.

The other key point of the algorithm is at line 10. The invariant image to be discarded at each step will be the same for all landmarks, selected according to a weighted voting strategy [19]. Let $s_{0,l}$ be the initial classifier score for landmark l (computed at line 6) and $s_{k,l}$ the same score after excluding the kth invariant (line 8). The classifier scores range from 0 (complete failure) to 1 (perfect success). Then,

$$\Delta_{k,l} = s_{0,l} - s_{k,l} \quad (5)$$

measures how much the kth invariant affects texture classification for landmark l. Actually, we define the index k such that the invariants are sorted in ascending order of $\Delta_{k,l}$. Then, every landmark l assigns $\nu_{k,l}$ votes to each invariant k according to:

$$\nu_{k,l} = (1 - \Delta_{k,l})2^{-k}. \quad (6)$$

It can be seen that the lower $\Delta_{k,l}$, the less important is the kth invariant for the lth landmark and, therefore, the more votes are assigned to the exclusion of that invariant. The negative exponential balances the voting privileges among all landmarks (i.e., each landmark will at most influence just a few invariants, regardless of the values of $\Delta_{k,l}$). The invariant eliminated at each step is then the most voted one:

$$\underset{k}{\operatorname{argmax}} \sum_{l} (1 - \Delta_{k,l})2^{-k}. \quad (7)$$

4 Experimental Evaluation

The performance of the proposed method was compared to that of the ASM and OF-ASM schemes. Data sets from three different databases were used, namely, the AR database [7], XM2VTS database [8], and Equinox database (visible band images from [9]).

The performance was tested in terms of segmentation accuracy and identity verification scores. The Configuration II of the Lausanne protocol [8] was used for the XM2VTS database, while AR and Equinox data sets were divided accordingly to make verification scores comparable. The individuals in each group were randomly

2. Under the above considerations, each elimination cycle skips $n - 1$ retrainings of the classifier. Since the time for computing the classification score is negligible, this leads immediately to (4).

TABLE 2
Composition of the Employed Data Sets and the Different Groups into Which They Were Divided

Database	AR	Equinox	XM2VTS
Total identities	133	91	295
Images per person	4	6	8
Total images	532	546	2360
Landmarks per image	98	98	64
Users	90	62	200
· training images	180	186	800
· test images	90	124	400
· eval. images	90	62	400
Test Impostors	32	21	70
· images	128	126	560
Eval Impostors	11	8	25
· images	44	48	200

chosen, making sure to have the same proportion of facial expression in all of them. Table 2 summarizes the resulting number of images on each group as well as the templates[3] used to landmark each data set.

The Equinox images were enlarged by a factor of 2.2:1 such that the average distance between the centers of the eyes matched that of AR data set (approximately 115 pixels). The XM2VTS images were kept unchanged since the needed resizing factor was less than 1.3:1. In all of the experiments that will be presented, only luminance information has been used.

The following sections evaluate the algorithm in terms of segmentation accuracy, rotation invariance, and identity verification, as well as the influence of the joint features selection strategy explained in Section 3.6.

4.1 Segmentation Accuracy

We constructed an ASM, an OF-ASM[4] and an IOF-ASM model for each of the three data sets, and tested their performance on all data sets. The models were built always from the training images of the *users* group (see Table 2), such that they can also be used in identity verification tests. In all cases, the image model was allowed to search within ± 3 pixels along the profiles (on each iteration) and β was set to 1.5 (see (2)). The rest of the parameters are shown in Table 3. They were chosen to obtain the best ASM results over the AR data set and were kept unchanged for the other databases and algorithms, whenever possible. For specific OF-ASM and IOF-ASM parameters, segmentation speed was given more importance than accuracy. It should be noted that, due to the different structure of the sampling grids in OF-ASM and IOF-ASM, the parameters of Table 3 make their sampling regions (per landmark) coincident for the smallest scale of OF-ASM.[5]

TABLE 3
Parameters Used to Build the Statistical Models

Parameter	ASM	OF-ASM	IOF-ASM
Main grid size	n/a	5×5	$X_G = 7, Y_G = 1$
Sub-grids size	n/a	$\alpha = 2\sigma$	$X_g = 7, Y_g = 5$
Profile length (p)	17	9	7
Resolutions	4	5	5
Iterations	12	12	12
Image properties	n/a	L, L_x, L_y L_{xx}, L_{xy}, L_{yy}	$L_{ii}, L_i L_i$ $L_i L_{ij} L_j, L_{ij} L_{ji}$
Blurring scales (pix)	n/a	$\sigma = 1, 2, 4$	$\sigma = 1, 2, 4$

TABLE 4
Point-to-Curve Segmentation Error, Normalized to the Distance between Eye-Centers

Training with	Model	Segmenting AR	Segmenting Equinox	Segmenting XM2VTS
AR	ASM	2.42 ± 0.06	3.74 ± 0.07	9.62 ± 0.20
	OF-ASM	2.55 ± 0.12 **(+5.4 %)**	12.2 ± 0.41 **(+227 %)**	6.04 ± 0.06 **(-37.2 %)**
	IOF-ASM	1.63 ± 0.03 **(-33.2 %)**	3.59 ± 0.07 **(-4.2 %)**	5.22 ± 0.10 **(-45.8 %)**
Equinox	ASM	4.72 ± 0.09	2.56 ± 0.05	13.9 ± 0.27
	OF-ASM	7.24 ± 0.21 **(+53.4 %)**	2.17 ± 0.07 **(-15.2 %)**	11.3 ± 0.09 **(-18,7 %)**
	IOF-ASM	4.16 ± 0.10 **(-11.8 %)**	1.92 ± 0.03 **(-25.2 %)**	8.35 ± 0.17 **(-39.7 %)**
XM2VTS	ASM	5.17 ± 0.14	4.45 ± 0.10	3.07 ± 0.08
	OF-ASM	7.92 ± 0.28 **(+53.2 %)**	13.9 ± 0.34 **(+169 %)**	2.13 ± 0.03 **(-30.6 %)**
	IOF-ASM	4.21 ± 0.12 **(-18.6 %)**	4.06 ± 0.11 **(-8.8 %)**	2.03 ± 0.02 **(-33.8 %)**

The face location was assumed to be roughly known from a (previous) detection step[6] and all available features were used for the segmentation. Feature selection will be covered in Section 4.4.

Table 4 shows the point-to-curve segmentation error, based on the distance from the landmarks obtained by the segmentation, measured perpendicularly to the curve defined by the manual annotations. The displayed values correspond to the average over all landmarks and all segmented images, with their corresponding standard error. The errors for each face were normalized dividing by 0.01 of the distance between the centers of the eyes (from the manual annotations) to make segmentation error comparable among faces of different sizes.

Below the segmentation errors for OF-ASM and IOF-ASM, we show the difference with respect to the ASM results as a percentage. It can be seen that IOF-ASM performed the best in all cases. We also observe a high degree of consistency in the results over the diagonal of the table, that is, when the segmented images belong to the same database used to construct the models. In these cases, the IOF-ASM approach always outperformed ASM by about 30 percent. In Fig. 4, we show the average segmentation errors over the table's diagonal, divided into the

3. The AR and Equinox data sets have been landmarked with our own 98-points template [20], while XM2VTS landmarks were obtained from http://www.isbe.man.ac.uk/~bim/data/xm2vts/xm2vts_markup.html with a 68-points template.

4. For OF-ASM, the profiles to search for were modified according to the training set statistics, since the ones proposed originally in [4] performed too poorly to allow for comparison of results (see Section 3.4).

5. The sampling region of OF-ASM is $(p + 2\alpha) \times (2\alpha + 1)$, so its minimum size is 13×5, when $\sigma = 1$. The sampling region of IOF-ASM is $(X_G + X_g - 1) \times Y_g$, that is 13×5 pixels independently of the blurring scale.

6. The model is always initialized at around 90 percent of the size of the average face in the corresponding database. In this way, the initialization usually falls inside the face region, thus reducing the background effects which regard mostly to the detection step. The average point-to-point error of the initialization was of 13.4 and 17.4 pixels for the AR and Equinox databases, respectively, while the greater size variations of the XM2VTS database made this initialization error grow up to 47 pixels (on average) in this database.

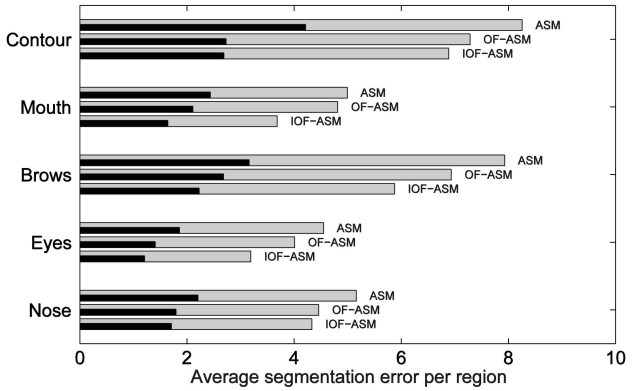

Fig. 4. ASM, OF-ASM, and IOF-ASM point-to-point (light) and point-to-curve (dark) segmentation errors per region. The values are in pixels, normalized with respect to the intereye distances and averaged over AR, Equinox, and XM2VTS data sets together, each segmented with their own models.

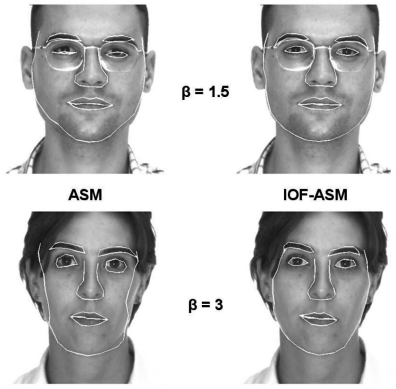

Fig. 6. Typical segmentation results of ASM and IOF-ASM on images from the AR database for different values of the PDM regularization constant (see (2)).

different facial regions. It can be seen that IOF-ASM always performed better, with both point-to-point and point-to-curve error metrics.

Fig. 5 shows further comparison of ASM and IOF-ASM accuracy when varying the PDM regularization constant β. It can be seen that, as β increases, the difference between the error of both models tends to grow. At the same time, the PCA reconstruction error introduced by the PDM decreases, which means the segmentation relies more on the image model precision. This behavior enforces the hypothesis of performance improvement in favor of IOF-ASM.

The value of the regularization constant greatly influences the segmentation performance. As shown in the curves displayed for ASM, if the PDM is given too much freedom, the segmentation error could exceed even the error obtained at $\beta = 0$ when only a similarity transformation of a fixed shape (the mean) is allowed. Examples of segmentation results are shown in Fig. 6. When β is increased to 3, the shape is clearly less restricted by the PDM, and the result achieved by the ASM is not a plausible face.

Away from the diagonal of Table 4, when different databases were used for training and testing, the behaviors were more random, but IOF-ASM was still the best, achieving statistically significant improvements with respect to both ASM and OF-ASM in all cases.[7] An underlying conclusion from these experiments, however, is that none of the models was able to preserve the same accuracy when running the segmentation on a database different to the one it has been trained with.

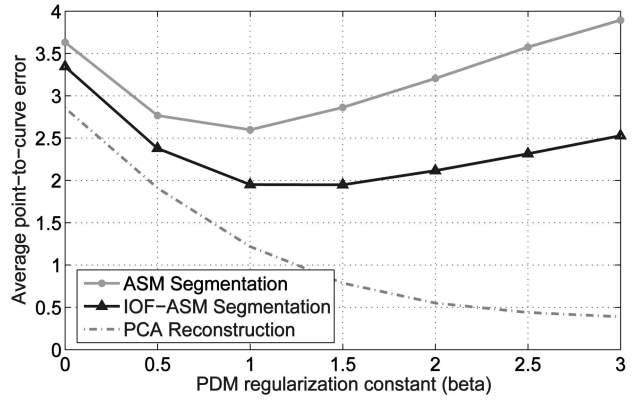

Fig. 5. ASM and IOF-ASM point-to-curve segmentation errors while varying the regularization constant of the PDM. The IOF-ASM improvement progressively grows from 8 percent at $\beta = 0$ to more than 35 percent at $\beta = 3$. The reconstruction error due to the regularization constraints of the PDM is also shown.

4.2 Invariance to Image Rotations

It was emphasized in Section 3.1 that the IOF-ASM features extracted from the images are invariant to rigid transformations. ASM exhibits the same invariance, but OF-ASM does not. To verify this fact, we performed the segmentation experiments at the finest resolution on the AR data set by rotating the images from -150 to $+150$ degrees. The PDM was constructed from the rotated images, such that the starting shape (based on the *mean shape*) was also rotated. However, the image models were not changed (i.e., they were based on the nonrotated images) so that their invariance was the only thing to be tested.

The results of the experiment are presented in Fig. 7. For each method, all segmentation errors were divided by the value obtained when segmenting the nonrotated images. Therefore, the three methods were scaled differently according to their respective accuracy, and the plot demonstrates only the relative influence of the rotation angle on each of them.

As expected, there is a clear increase of the segmentation error in the OF-ASM as the rotation angle departs from zero. On the other hand, the ASM and IOF-ASM performances are only affected by the numerical approximations due to the discrete nature of the image. That is, the samples for the rotated versions of the image were computed by interpolation, except for ± 90 degrees, where the error attains again its minimum value. The lack of invariance can be substantial due to numerical approximations when doing multiresolution, since the chosen low-pass filters usually depend on the orientation of the axis. Care must be taken both when implementing and testing invariant methods in that context.

7. The t-test showed confidence larger than 99 percent in all cases except when comparing IOF-ASM and ASM on the Equinox database training with the AR database, where the confidence was around 97 percent.

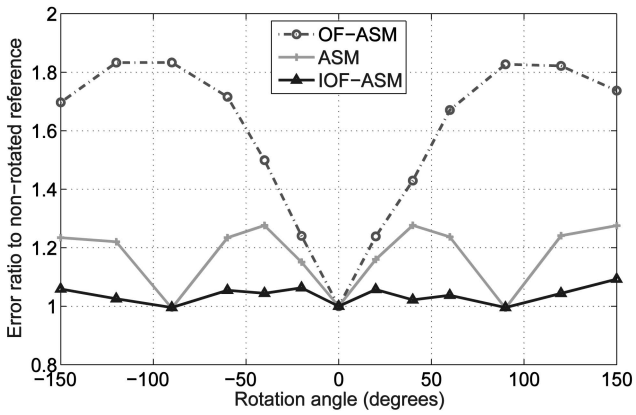

Fig. 7. ASM, OF-ASM, and IOF-ASM point-to-curve segmentation error for different rotation angles on the AR database. The models were constructed with the nonrotated images, whose segmentation error was taken as reference. The plot shows the ratio of the segmentation errors (for the different rotation angles) to the (nonrotated) reference.

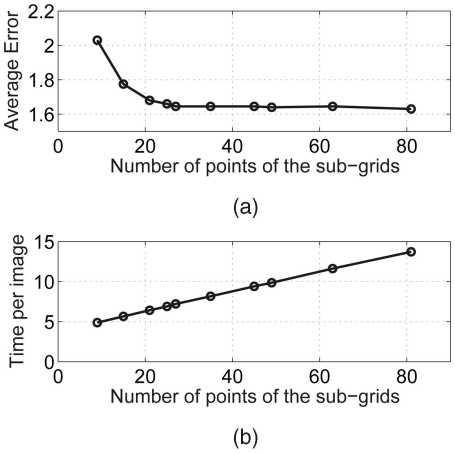

Fig. 8. (a) Normalized point-to-curve segmentation error and (b) segmentation time, in seconds, while varying the number of points of the subgrids. The displayed curves show the average over the 532 images from the AR data set.

4.3 Subgrid Size

The size of the subgrids is an important parameter of IOF-ASM. It can affect both the accuracy and the speed of the segmentation process, as demonstrated in Fig. 8. As explained in Section 3.3, X_g and Y_g can take only odd values. Hence, we repeated the segmentation test on the AR database for all the possible subgrid sizes from three to nine pixels in both dimensions (i.e., 3×3, 3×5, etc.).

The first conclusion from these experiments was that accuracy and speed changed as a function of the total number of points of the subgrids, being almost insensitive to the swap between X_g and Y_g. For this reason, the horizontal axis of Fig. 8 is labeled with the product $X_g \times Y_g$.

Analyzing the segmentation error, it can be seen that very small subgrid sizes degrade the accuracy. However, for subgrids of 25 or more points, the differences are very subtle. Statistical significance was observed only for the subgrids smaller than 25 points.

Regarding the segmentation time, its variation was linear with respect to the number of pixels.[8] The strong linearity of the curve allows us to clearly distinguish between the time consumed by the iterative fitting and by the preprocessing (the extrapolation of the line to intersect the vertical axis for a hypothetical zero-size subgrid). This preprocessing time, of approximately 3.5 seconds, is mainly due to the computation of the invariants.

The size of 7×5 points chosen for the experiments in Table 3 is within the flat region of accuracy, and allows for a direct comparison with OF-ASM, since the total region analyzed for each landmark will match the one of OF-ASM for $\sigma = 1$ (see Section 4.1). However, the displayed curves suggest that smaller subgrid sizes could accelerate the segmentation up to 15 percent without compromising the accuracy.

4.4 Feature Selection

The price paid for the accuracy improvement of IOF-ASM reported in Section 4.1 is a decrease in segmentation speed due to the more complex image processing compared to ASM. However, the method can be accelerated by reducing the number of features.

Using the training images (from the *users* group) of AR database, we performed the feature selection as explained in Section 3.6, and reduced the number of invariant images from 12 to 4. We constructed several models with a different number of invariants and compiled the segmentation results in Table 5. From left to right, the columns show the number of invariants used to build the model, the point-to-curve errors averaged over the AR database, over all images from the three available data sets (a total of 3,438), and the segmentation time per image. The ASM and OF-ASM results are also shown and the percentages are computed by taking the IOF-ASM with all the features as a baseline.

It can be seen that the smaller the number of invariants (N_{IF}) we use to build the model, the higher the segmentation error and the lower the segmentation time. This behavior was very clear in segmenting the AR database, since part of its images were used to guide the feature selection process. When moving to other data sets, where the best set of features may, in general, be different, there

TABLE 5
Segmentation Accuracy and Speed for Different Number of Invariants, over 3,438 Images from AR, Equinox, and XM2VTS Data Sets

Model	N_{IF}	Avg error AR	Avg error ALL	Time per image
ASM	n/a	2.42 ± 0.06	7.58 ± 0.16	0.70 s
OF-ASM	n/a	2.55 ± 0.12	6.47 ± 0.13	> 100 s
IOF-ASM	12	1.63 ± 0.03	4.40 ± 0.09	8.13 s
	10	1.66 ± 0.03 **(+1.8 %)**	4.43 ± 0.09 **(+0.7 %)**	7.31 s **(-10 %)**
	8	1.71 ± 0.03 **(+4.9 %)**	4.59 ± 0.09 **(+4.3 %)**	6.39 s **(-21 %)**
	6	1.82 ± 0.04 **(+11.7 %)**	4.49 ± 0.09 **(+2.1 %)**	5.62 s **(-31 %)**
	4	1.87 ± 0.04 **(+14.7 %)**	5.79 ± 0.11 **(+31.6 %)**	4.09 s **(-50 %)**

8. The reported segmentation times were measured on a nonoptimized implementation of the algorithms in C++, on an AMD Athlon running at 2 GHz. Significantly better times should be possible, especially if optimizing the calculation of invariants.

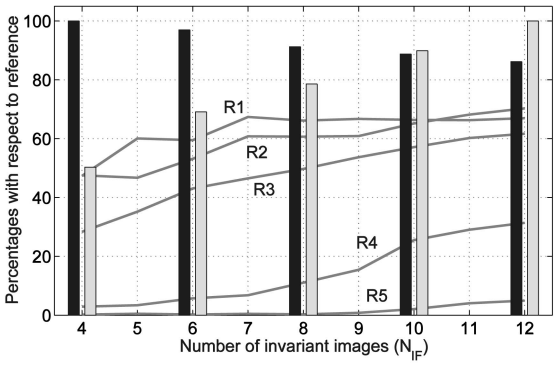

Fig. 9. Feature selection statistics from the AR model reducing N_{IF} from 12 to 4. The continuous lines show the texture classifiers scores after training, averaged over all landmarks at each resolution level. The bars display the segmentation error (black) and the segmentation time (gray) for the AR data set (see Table 5). The maximum values of the bars are used as reference (fixed to 100 percent) and the rest are rescaled accordingly.

was some unexpected decrease of the average error with six invariants, but the tendency was clearly the same.

Therefore, N_{IF} can be chosen to make the model faster or more accurate. Furthermore, the influence of excluding invariants from the model can be evaluated prior to the segmentation experiments, during training. The continuous lines in Fig. 9 show the classifier scores averaged over all the landmarks for each resolution level. Notice the clear correlation between them and the pairs of bars showing the segmentation accuracy (left) and the segmentation time (right).

The results displayed suggest that all of the used invariant images brought valuable information to the texture classifiers. The most discriminant feature in our experiments was $\{L_{ij}L_{ji}, \sigma = 1\}$, which was among the four nonexcluded features in all the five resolution levels.

4.5 Step-by-Step Analysis

It is interesting to go back to the hypothesis made at the beginning of Section 3. It was stated that IOF-ASM would provide a number of improvements with respect to OF-ASM, which were attributed to specific components (or changes) of the new method. To verify this, we constructed intermediate versions between OF-ASM and IOF-ASM, which are summarized in Table 6. Three measures were computed to compare their performance, by using the AR data set:

- e: Point-to-curve segmentation error, averaged over all landmarks and all images, with their corresponding standard error. The error values are normalized as in Table 4.
- r: Rotation variance ratio. The experiments of Section 4.2 were repeated for the original images (no rotation) and rotated versions at ± 60 and ± 120 degrees. Then, r was computed as the ratio from the average segmentation error of the rotated images to the error of the nonrotated images. Hence, as in Fig. 7, if a model is rotation-invariant, then r approaches 1.
- t: Segmentation time per image (on average), under the same conditions of Sections 4.3 and 4.4.

TABLE 6
Intermediate Steps from OF to IOF-ASM

Model	Classif.	Features	Profiles	Results
OF-ASM	kNN	Derivatives	OF-ASM	$e = 2.55 \pm 0.12$ $r = 1.76 \pm 0.03$ $t > 100s$
n/a	**MVN**	Derivatives	OF-ASM	$e = 2.72 \pm 0.11$ $r = 1.69 \pm 0.04$ $\mathbf{t = 9.78s}$
n/a	MVN	**Invariants**	OF-ASM	$e = 2.68 \pm 0.10$ $\mathbf{r = 1.03 \pm 0.01}$ $t = 8.54s$
n/a	MVN	Invariants	**V-shape**	$\mathbf{e = 1.83 \pm 0.04}$ $r = 1.06 \pm 0.01$ $t = 8.15s$
IOF-ASM	MVN	Invariants	**Robust-V**	$\mathbf{e = 1.63 \pm 0.03}$ $r = 1.05 \pm 0.02$ $t = 8.13s$

On each row, the change with respect to the previous one is highlighted in bold letters, as well as the most affected measure. The results are very consistent with the expectations drawn in Section 3. The exchange of the kNN classifier by the MVN (second row) reduced the segmentation time in one order of magnitude, with a small accuracy loss (no statistically significant in this case). The use of invariants instead of derivatives (third row) did not affect the segmentation error, but clearly achieved rotation invariance.

The remaining change at this step is the replacement of the OF-ASM profiles by the V-shape of IOF-ASM, including the modification of the grids structure (see Section 3.4). The resulting model is the IOF-ASM proposed in this paper (row five). However, an intermediate step is shown in row four, which does not use the outliers concept of (3), but a simple absolute distance. It can be seen that both steps considerably reduced the segmentation error (the results are statistically significant with confidence greater than 99 percent).

4.6 Identity Verification

Having demonstrated that IOF-ASM is more accurate in segmenting facial images, there remains the question of whether or not it will improve recognition as well. We tested verification scores using both shape and texture classifiers in the three data sets with the segmentation performed by their corresponding model. The development of a state-of-the-art classifier was beyond the scope of this paper. In both cases, we applied Z-normalization [21] and then used a whitened correlation classifier, known to be a good choice for PCA-based metrics [22].

In Table 7, the landmark points found during the model matching were used to construct a mesh by means of a Delaunay triangulation of the whole face. This mesh permits the establishment of a mapping for the texture of the face from this specific shape to the mean shape of the training set. Using this mapping, the texture was warped onto the mean shape and sampled into a texture vector [1]. In order to parameterize the texture, a model was constructed by applying PCA to all the vectors from the training images of the database. Then, the biometric

TABLE 7
Identity Verification Scores Using Texture Parameters

Database	Metric	ASM	OF-ASM	IOF-ASM
AR	EER (Eval)	3.8%	5.6%	2.8%
	HTER (Test)	4.5% ±1.9%	5.8% ±1.8%	4.2% ±1.8%
Equinox	EER (Eval)	0.1%	0.6%	$< 0.1\%$
	HTER (Test)	1.2% ±1.6%	1.6% ±1.6%	0.5% ±0.9%
XM2VTS	EER (Eval)	2.7%	2.4%	1.0%
	HTER (Test)	2.6% ±0.8%	2.3% ±0.7%	1.3% ±0.6%

parameters were obtained by projecting the texture vector onto the subspace of this model.

The results in Table 8 were obtained using the PCA parameters of the shape model. Only the most significant modes were taken into account, such that 95 percent of the total variance was explained. The error rates were computed according to the protocols described in Table 2, using the *evaluation* sets Equal Error Rate (EER) [23] to fix the working point of the classifier and compute the False Acceptance (FAR) and False Rejection (FRR) rates from the *test* sets. The tables show the average of both metrics (HTER, for Half Total Error Rate) plus a 90 percent confidence interval, computed using the *test of two proportions* as in [24].

The results show that IOF-ASM outperforms the other methods in all cases, except when using shape parameters on the Equinox data set. In that case, a lower EER is achieved by using ASM, but that behavior does not generalize to the test set (HTER), where the scores are the same as those of IOF-ASM. It must be pointed out that, in most cases, the differences in error rates are not statistically significant, due to the limited number of images available. However, the trend in the three data sets is consistent and indicates an improvement in the verification task possibly due to the more accurate segmentation. Additionally, the work by Kang et. al. [25] allows for the comparison of IOF-ASM with similar approaches. In that work, they outperform the best distance measures using the eigenfaces approach [26] obtaining a 2.6 percent EER on the XM2VTS database. This is similar to the ASM performance shown in Table 7, but clearly worse than IOF-ASM. In the same work, however, more sophisticated classification schemes demonstrate better identity verification rates.

TABLE 8
Identity Verification Scores Using Shape Parameters

Database	Metric	ASM	OF-ASM	IOF-ASM
AR	EER (Eval)	16.6%	20.0%	15.6%
	HTER (Test)	17.7% ±3.4%	19.5% ±3.3%	11.9% ±2.5%
Equinox	EER (Eval)	6.4%	14.5%	9.7%
	HTER (Test)	11.5% ±3.9%	16.2% ±4.1%	11.6% ±3.6%
XM2VTS	EER (Eval)	13.8%	13.7%	10.0%
	HTER (Test)	14.0% ±1.4%	13.4% ±1.4%	9.1% ±1.1%

TABLE 9
Comparison with the Segmentation Errors Reported by Other Researchers

Method and Paper	Data Sets Details	Reported error (in pixels)	Normalized error (estimated)
IOF-ASM	AR-XM2-EQX 3438 images see Section IV	1.95 point-curve 4.82 point-point	1.95 point-curve 4.82 point-point
AAM [30]	XM2VTS 1817 images 720×576 pix	3.9 to 5.4 (point-curve)	3.9 to 5.4 (point-curve)
AAM [32]	400 images	4.0 pixels (point-point)	> 5.2 pixels (point-point)
EBGM [5]	Bochum [35] 432 images 128×128 pix	1.6 pixels (point-point)	> 5.3 pixels (point-point)
AAM [33]	IMM 37 images 640×480 pix	2.78 point-curve 5.57 point-point	2.21 point-curve 4.42 point-point

5 Discussion

The results presented in the previous section show a significant accuracy improvement of the IOF-ASM with respect to its predecessors, namely, the ASM and OF-ASM. To put these results into a more general context, a wider comparison with other methods would be desirable. The task, however, is not easy. On one hand, segmentation accuracy is usually tested against manual annotations, which are subjective and not widely available. Only recently have some large-sized facial data sets been annotated and made freely available.[9] This fact has led researchers to evaluate their methods on different databases, with different image sizes, annotation templates, and/or number of samples, hampering consistent comparisons.

Moreover, there is no universally accepted standard for the metric to be used. Researchers have reported segmentation results as the percentage of pixels correctly identified inside a region [27], or the fraction of test images correctly detected within a threshold [28], [29], to cite some. The most popular measurement is the Euclidean distance, in its point-to-point or point-to-curve variants. Despite the fact that the former is the most intuitive, the latter usually provides greater correlation between the error value and the failure of the model when the objective is to determine the boundaries between regions. A clear example is the points on the contour of the face, where the exact location of the landmarks does not really matter. What matters is the way the curve fits into the face boundary. As opposed to that, we can also find points in the face whose exact location is important, such as the corners of the eyes or the mouth.

Considering the above restrictions, we have gathered in Table 9 a list of the results reported in the literature that are possible to compare with our experiments. The error values (given in pixels) were *corrected* dividing by the average distance between the eyes, obtaining an estimation of the *normalized error*. In this way, we made these results comparable to the ones presented in the previous section. The most direct comparison is probably a work of Scott et al. [30]. They tested several AAM approaches on (almost) the whole XM2VTS database, with the same kind of

9. Some annotated data sets can be found at http://www.isbe.man.ac.uk/~bim/.

annotation template as the ones used here. Depending on the technique, their errors range from 3.9 to 5.4 pixels,[10] which are significantly higher, even than our implementation of ASM. This result was somehow expected [31], although the difference should not be that high. The explanation is probably the different choice of the parameters and the more challenging initialization used in [30].

In [32], the accuracy of AAMs is tested with a better initialization, resulting from a previous estimation of the correct pose. An average point-to-point error of 4.0 pixels is obtained, on approximately 200 pixels wide faces. To determine the correction factor we computed the ratio between the intereye distance and the face's width on our data sets, which was around 2.6 (on average). In this way, the normalized error would be 5.2 pixels, but this value does not include a 1.6 percent of the test data for which the algorithm did not converge (average error greater than 7.5 pixels). Therefore, 5.2 pixels is the (estimated) lower bound for the segmentation error.

In [5], Elastic Bunch Graph Matching (EBGM) is used to segment facial images with a less dense template that includes the whole head. An average point-to-point error of 1.6 pixels is reported on a 432 images database, in which the distance between the center of the eyes is not higher than 30 pixels. Therefore, the normalized error in this case becomes greater than 5.3 pixels, which suggests a slightly lower performance than our IOF-ASM. However, the much lower resolution of their images and the presence of slightly in-depth faces rotation hampers any direct comparison. On the other hand, their less dense template implies some benefit when using point-to-point distance, since this allows for choosing mainly clearly defined points. As opposed to that, more dense templates are meant to define curves along the boundaries of different regions. It is often the case that the intermediate points along the boundaries have a very ambiguous location (i.e., the face silhouette) and the automatic segmentation differs from the manual annotations along the boundary, wrongly increasing the computed distance.

Tamminen and Lampinen [33] segmented images from the IMM database [34], using a variant of AAM based on Gabor filters. The average intereye distance on this database is almost 126 pixels, so the normalized error suggests that their results are very good. Unfortunately, they report tests on only 37 images, a much smaller data set than the other works compared here.

6 Summary and Conclusions

In this paper, a new segmentation method has been presented to solve some limitations of its predecessor, the OF-ASM approach. The main contributions introduced here are an increased segmentation accuracy, an invariance to rigid transformations, the ability to deal with shape complexities (such as multiple embedding), and the speed-up of the segmentation process.

The IOF-ASM was compared with its two predecessors over three different data sets (almost 3,500 images). Moreover, we gathered data from different segmentation methods to put our results in a more global context. It was shown that the achieved accuracy is comparable to other state-of-the-art algorithms, and that differences become smaller when similar techniques are employed (i.e., the Gabor-based AAM in [33]). That is, by means of using more elaborated descriptions of the texture, it is possible to increase the accuracy of the segmentation. In this regard, our method provides a generic framework, since it can be extended to any new set of local descriptors. It was shown that using just linear invariants up to the second order is enough to obtain a very good performance.

The IOF-ASM has been designed to provide segmentations by means of dense annotated templates. Thus, the different regions of an object can be identified and analyzed by further processing. We have demonstrated this by performing identity verification experiments, obtaining results comparable to the fully automatic methods reported in [36], although we used only basic techniques to classify the identities.

The price paid to increase accuracy is, in general, a higher computational load. The banks of filters applied to the image significantly slow down the matching process. To reduce this effect, we have shown that feature selection can save up to 50 percent of computational time while degrading accuracy by only about 15 percent. Different trade-offs between speed and accuracy are also possible.

10. No correction is needed here, since the XM2VTS images have an average intereye distance of approximately 100 pixels.

Acknowledgments

This work was partially funded by grants TIC2002-04495-C02 and TEC2006-03617/TCM, from the Spanish Ministry of Education & Science, and grant FIT-360000-2006-55 from the Spanish Ministry of Industry. Federico M. Sukno is supported by a BSCH grant. Sebastián Ordás is supported by an FPU grant from the Spanish Ministry of Education & Science. Alejandro F. Frangi holds a Ramón y Cajal Research Fellowship. The Computational Imaging Lab at Pompeu Fabra University is a member of the Biosecure (IST-2002-507534) European Network of Excellence.

References

[1] A. Lanitis, C.J. Taylor, and T.F. Cootes, "Automatic Interpretation and Coding of Face Images Using Flexible Models," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 19, no. 7, pp. 743-756, July 1997.

[2] T.F. Cootes, C.J. Taylor, D.H. Cooper, and J. Graham, "Active Shape Models—Their Training and Application," *Computer Vision and Image Understanding,* vol. 61, no. 1, pp. 38-59, 1995.

[3] T.F. Cootes, G. Edwards, and C.J. Taylor, "Active Appearance Models," *Proc. European Conf. Computer Vision,* vol. 2, pp. 484-498, 1998.

[4] B. van Ginneken, A.F. Frangi, J.J. Staal, B.M. ter Har Romeny, and M.A. Viergever, "Active Shape Model Segmentation with Optimal Features," *IEEE Trans. Medical Imaging,* vol. 21, no. 8, pp. 924-933, 2002.

[5] L. Wiskott, J-M Fellows, N. Krüger, and C. von der Malsburg, "Face Recognition by Elastic Bunch Graph Matching," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 19, no. 7, pp. 775-779, July 1997.

[6] F. Sukno, S. Ordas, C. Butakoff, S. Cruz, and A.F. Frangi, "Active Shape Models with Invariant Optimal Features IOF-ASMs," *Proc. Fifth Int'l Conf. Audio and Video-Based Biometric Person Authentication,* pp. 365-375, 2005.

[7] A. Martínez and R. Benavente, "The AR Face Database," technical report, Computer Vision Center, Barcelona, Spain, 1998.

[8] K. Messer, J. Matas, J. Kittler, J. Luettin, and G. Maitre, "XM2VTSDB: The Extended M2VTS Database," *Proc. Second Int'l Conf. Audio and Video-Based Biometric Person Authentication,* pp. 72-77, 1999.
[9] A. Selinger and D. Socolinsky, "Appearance-Based Facial Recognition Using Visible and Thermal Imagery: A Comparative Study," Technical Report 02-01, Equinox Corp., 2002.
[10] M. Kudo and J. Sklansky, "Comparison of Algorithms that Select Features for Pattern Classiers," *Pattern Recognition,* vol. 33, pp. 25-41, 2000.
[11] S. Arya, D.M. Mount, N.S. Netanyahu, R. Silverman, and A.Y. Wu, "An Optimal Algorithm for Approximate Nearest Neighbor Serching in Fixed Dimensions," *Int'l J. Computer Vision,* vol. 45, no. 6, pp. 891-923, 1998.
[12] I. Aizenberg, C. Butakoff, V. Karnaukhov, N. Merzlyakov, and O. Milukova, "Blurred Image Restoration Using the Type of Blur and Blur Parameters Identification on the Neural Network," *SPIE Proc. Image Processing: Algorithms and Systems,* pp. 460-471, 2002.
[13] I. Aizenberg, N. Aizenberg, and J. Vandewalle, *Multi-Valued and Universal Binary Neurons: Theory, Learning, Applications.* Kluwer Academic, 2000.
[14] L. Florack, "The Syntactical Structure of Scalar Images," PhD thesis, Utrecht Univ., Utrecht, The Netherlands, 2001.
[15] K.N. Walker, T.F. Cootes, and C.J. Taylor, "Correspondence Using Distinct Points Based on Image Invariants," *Proc. British Machine Vision Conf.,* vol. 1, pp. 540-549, 1997.
[16] C. Schmid and R. Mohr, "Local Greyvalue Invariants for Image Retrieval," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 19, no. 5, pp. 530-535, May 1997.
[17] P.J. Huber, *Robust Statistics.* Wiley, 1981.
[18] I. Guyon and A. Elisseeff, "An Introduction to Variable and Feature Selection," *J. Machine Learning Research,* vol. 3, pp. 1157-1182, 2003.
[19] E. Alpaydin, "Multiple Neural Networks and Weighted Voting," *Proc, 11th IAPR Int'l Conf. Pattern Recognition Methodology and Systems,* pp. 29-32, 1992.
[20] F. Sukno, J. Guerrero, and A.F. Frangi, "Homographic Active Shape Models for View-Independant Facial Analysis," *Proc. SPIE Biometric Technologies for Human Identification II,* pp. 152-163, 2005.
[21] K.P. Li and J.E. Porter, "Normalizations and Selection of Speech Segments for Speaker Recognition Scoring," *Proc. IEEE Int'l Conf. Acoustics, Speech, and Signal Processing,* pp. 595-598, 1988.
[22] V. Perlibakas, "Distance Measures for PCA-Based Face Recognition," *Pattern Recognition Letters,* vol. 25, no. 6, pp. 711-724, 2004.
[23] R.M. Bolle, N.K. Ratha, and S. Pankanti, "Error Analysis of Pattern Recognition Systems—The Subsets Bootstrap," *Computer Vision and Image Understanding,* vol. 93, pp. 1-33, 2004.
[24] S. Bengio and J. Mariéthoz, "A Statistical Significance Test for Person Authentication," *Proc. Odyssey 2004: The Speaker and Language Recognition Workshop,* pp. 237-244, 2004.
[25] H. Kang, T.F. Cootes, and C.J. Taylor, "A Comparison of Face Verification Algorithms Using Appearance Models," *Proc. British Machine Vision Conf.,* vol. 2, pp. 477-486, 2002.
[26] M. Turk and A. Pentland, "Eigenfaces for Recognition," *J. Cognitive Neuroscience,* vol. 3, no. 1, pp. 71-86, 1991.
[27] A.W.C. Liew, S.H. Leung, and W.H. Lau, "Segmentation of Color Lip Images by Spatial Fuzzy Clustering," *IEEE Trans. Fuzzy Systems,* vol. 11, no. 4, pp. 542-549, 2003.
[28] H. Demirel, T.J.VV. Clarke, and P.Y.K. Cheung, "Adaptive Automatic Facial Feature Segmentation," *Proc. Second Int'l Conf. on Face and Gesture Recognition,* pp. 277-282, 1996.
[29] D. Cristinacce and T.F. Cootes, "A Comparison of Shape Constrained Facial Feature Detectors," *Proc. Sixth IEEE Int'l Conf. Automatic Face and Gesture Recognition,* 2004.
[30] I.M. Scott, T.F. Cootes, and C.J. Taylor, "Improving Appearance Model Matching Using Local Image Structure," *Proc. Information Processing in Medical Imaging,* pp. 258-269, 2003.
[31] T.F. Cootes, G.J. Edwards, and C.J. Taylor, "Comparing Active Shape Models with Active Appearance Models," *Proc. 10th British Machine Vision Conf.,* vol. 1, pp. 173-182, 1999.
[32] T.F. Cootes and C.J. Taylor, "Statistical Models of Appearance for Computer Vision," technical report, Wolfson Image Analysis Unit, Univ. of Manchester, 2000.
[33] T. Tamminen and J. Lampinen, "A Bayesian Occlusion Model for Sequential Object Matching," *Proc. British Machine Vision Conf.,* 2004.
[34] M.M. Nordstom, M. Larsen, J. Sierakowski, and M.B. Stegmann, "The IMM Face Database: An Annotated Dataset of 240 Face Images, Informatics and Mathematical Modelling," Technical Univ. of Denmark, DTU, Denmark, http://www2.imm.dtu.dk/pubdb/p.php?3160, 2004.
[35] M. Lades, J.C. Vorbrüggen, J. Buhmann, J. Lange, C. von der Malsburg, R.P. Würtz, and W. Konen, "Distortion Invariant Object Recognition in the Dynamic Link Architecture," *IEEE Trans. Computers,* vol. 42, no. 3, pp. 300-311, Mar. 1993.
[36] K. Messer et al. "Face Verification Competition on the XM2VTS Database," *Proc. Fourth Int'l Conf. Audio and Video-Based Person Authentication,* 2003.

Federico M. Sukno graduated in electrical engineering from the Universidad Nacional de La Plata, Argentina (2000). After graduation, he worked for two years in mobile telecommunications (Ericsson) and industrial control (ABB), both in Argentina. He is currently working on facial biometrics in the Computational Imaging Lab at Pompeu Fabra University. He is enrolled in the Biomedical Engineering PhD program at the University of Zaragoza, with a grant from the university in collaboration with the Santander Central Hispano Bank (SCH).

Sebastián Ordás studied biomedical engineering at the University of Entre Ríos, Argentina (2000). Before starting the PhD program, he worked for three years as a software developer in medical-related applications. He is currently with the Computational Imaging Lab, Department of Technology, University Pompeu Fabra. His thesis topics and research interests are on statistical model-based cardiac image analysis. His work is funded by an FPU grant from the Spanish Ministry of Education.

Constantine Butakoff received the MSc degree in mathematics from the Uzhgorod National University in 1999. Since 1998, he has worked in the field of image processing and recognition. In August 2003, he began a PhD program at the University of Zaragoza, and he is now a member of Computational Imaging Lab at University Pompeu Fabra in Barcelona, Spain.

Santiago Cruz studied telecommunication engineering at the Universitad Politecnica de Madrid from 1991-1996. In 1997, he was at Ecole Nationale Superieure des Telecommunication, where he received a master's degree in image processing and television systems. He works as an assistant professor at the Universidad de Zaragoza. His main topics of reseach are biometrics (it is the subject of his PhD, defended at Universidad Politécnica de Madrid, 2005), statistical pattern recogntion, and image and speech processing. Previous activities regarding biometrics were developed at Biometrics research Lab ATVS (Universidad Autonoma de Madrid).

Alejandro F. Frangi received the undergraduate degree in telecommunications engineering from the Technical University of Catalonia, Barcelona, in 1996. He subsequently carried out research on electrical impedance tomography for image reconstruction and noise characterization at the same institution under a CIRIT grant. In 1997, he obtained a grant from the Dutch Ministry of Economic Affairs to pursue his PhD degree in the Image Sciences Institute at the University Medical Center Utrecht on model-based cardiovascular image analysis under the sponsorship of Philips Medical Systems, Nederland BV. After graduation in 2001, he was an assistant professor at the University of Zaragoza until 2003. Subsequently, Dr. Frangi was awarded a Ramón y Cajal Research Fellowship from 2003 until 2008, a national program of the Spanish Ministry of Science and Technology for promoting outstanding young investigators. In September 2004, he was invited to join the Department of Technology at Pompeu Fabra University in Barcelona, Spain where he leads the Computational Imaging Lab (www.cilab.upf.edu). His main research interests are in computer vision and medical image analysis with particular emphasis in model and registration-based techniques and statistical methods. He has been twice a guest editor of two special issues of the *IEEE Transactions on Medical Imaging* and one of *Medical Image Analysis*. He is a senior member of the IEEE, and an associate editor of the *IEEE Transactions on Medical Imaging* and *Medical Image Analysis*. He was awarded the 2006 IEEE EMBS Early Career Award.