 CMES, vol.36, no.3, pp.213-241, 2008

Methods to Automatically Build Point Distribution Models for Objects like Hand Palms and Faces Represented in Images

Maria João M. Vasconcelos[1] **and João Manuel R. S. Tavares**[1]

Abstract: In this work we developed methods to automatically extract significant points of objects like hand palms and faces represented in images that can be used to build Point Distribution Models automatically. These models are further used to segment the modelled objects in new images, through the use of Active Shape Models or Active Appearance Models. These models showed to be efficient in the segmentation of objects, but had as drawback the fact that the labelling of the landmark points was usually manually made and consequently time consuming. Thus, in this paper we describe some methods capable to extract significant points of objects like hand palms and compare the segmentation results in new images.

Keyword: Feature Points; Image Analysis; Point Distribution Models; Segmentation and Object Recognition.

1 Introduction

Thanks to their powerful vision system, human beings can describe and understand images in an easy way. The value vision system has been recognized by many researchers, leading them to attempt to reproduce the human vision system computationally. Thus, automatic systems of Computational Vision are continually being developed to perform some of its functions.

Segmentation and analysis of objects represented in images are areas of great interest in Computational Vision; many approaches capable of characterizing objects and later recognizing them are included in these areas. Point Distribution Models (PDMs), which we consider in this work, are a particular case of deformable models used to extract the most representative characteristics of the objects represented in images.

[1] FEUP – Faculdade de Engenharia da Universidade do Porto / INEGI – Instituto de Engenharia Mecânica e Gestão Industrial, Rua Dr. Roberto Frias S/n, 4200-465 Porto, Portugal, Email: {maria.vasconcelos, tavares}@fe.up.pt

Other segmentation techniques are based on level set methods [Wang, Lim et al., (2007)] and physical principles [Gonçalves, Tavares et al., (2008)]. In [Zhang, Cheng et al., (2008)] and [Tavares, Carvalho et al., (2008)] some application examples of these techniques are shown.

The use of deformable models in image analysis and image interpretation was first introduced in [Kass, Witkin et al., (1987)], where the authors present the concept of Snake. Snake is an active contour model which adjusts itself to the object involved through a combination of internal and external forces: the internal forces translate the flexibility and stretch and the external forces pull the contour toward to relevant areas of the image. In this method, the adjustments of the active contour stop when a minimal energy state is reached, typically when it finds the structure's borders.

In order to improve active contour models, researchers have suggested some developments, such as the addition of an inner force ("balloon force") which can be used to inflate the snake like a balloon so that the front move towards mesh features. So far, improvements were made, mostly, in order to allow the use of active contours in more complex images and also to use it automatically.

Other types of deformable models are the deformable templates which use templates described by parametric functions. For instance, in [Yuille, Cohen et al., (1992)] the authors build a model to detect eyes in images, where the eye is represented by a circle describing iris, two parabolic curves describing eyelids and also the intensity in turn of these regions. One of the disadvantages of this method is the fact that the construction of deformable templates is complex and greatly depends on the object being modelled [Tian, Kanade et al., (2000)].

The models which we consider in this work are also included in the category of deformable models: Point Distribution Models (PDMs) were initially proposed by [Cootes, Taylor et al., (1992)] to model objects based on its statistical analysis. These models are obtained through the analysis of the statistics of the coordinates of the landmarks that represent the deformable object in study: after aligning the object shapes, a Principal Component Analysis is made and the mean shape of the object and the main modes of its variation are obtained.

Active Shape Models (ASMs) [Cootes and Taylor, (1992)] and Active Appearance Models (AAMs) [Cootes, Edwards et al., (1998)] use PDMs to segment (identify) and recognize the modelled objects in new images. Both models use a combination of the statistical shape model with the grey levels of the object. In [Sung, Kanade et al., (2007)] the authors propose a new fitting method that combines the functions of both ASM and AAM into a single function in a gradient based optimization framework with the aim to reduce the average fitting error.

These statistical models are very useful for image analysis in different applications

of Computational Vision. For instance, they can be used in areas like: medicine, for locating bones and organs in medical images [Rijsdam, (1999); Schaap, (1999)]; industry, for industrial inspection [Aixut, Meneses et al., (2003)]; and security, for face recognition [Lanitis, Taylor et al., (1995)].

The main goals of this work consisted in describing new methods capable of labelling automatically objects like hand palms and faces represented in images, testing the efficiency of the presented methods and also turning the construction of Point Distribution Models automatic for the referred objects. The automation of these models allows the reduction of the computation time and also the possibility to enlarge easily the size of the training set without spending too much time.

In the next section of this paper, we present our methods to automatically extract the landmarks points of objects like hand palms and faces represented in images; briefly explain how to build PDMs, ASMs and AAMs; and then, present some applications of these models on the referred type of objects; in the last section, we draw some conclusions and suggestions of future work.

2 Automatic Extraction of Landmark Points

In [Hill and Taylor, (1994)], a two-stage algorithm to automatically generate PDMs from a training set of example shapes is described. This method starts with a pairwise matching to establish an approximate set of landmarks on each of the example boundaries and then uses a non-linear optimiser to refine the locations of the landmarks generated in the first stage. [Hicks, Marshall et al., (2002)] presented a method for automatic landmark extraction from the contours of biological specimens to automatically build ASMs. The authors [Bailleul, Ruan et al., (2003)] present a method based on a priori knowledge provided by anatomical atlases to build almost automatically a PDM of intern brain structures. Later, in [Angelopoulou and Psarrou, (2004)] a method is proposed for automatic landmark detection on the contour of hand shapes, based on Freeman chain code. [Baker and Matthews, (2004)] address the problem of automatic construction of AAM as a image coding problem, where the authors consider landmarks as nodes of a regular mesh and build an AAM.

In [Caunce and Taylor, (1998)], it is presented a statistical shape model of the cortical surface of the brain and developed an automated method to facilitate the marking process. The authors outline and discuss methods including simple point matching and more complex curve matching in order to solve the problem of identifying correspondences between examples. In [Lohmann and Cramon, (2000)], the authors presented a method that automatically detects and attributes neuroanatomical names to the cortical folds using image analysis methods applied to

magnetic resonance data of human brains. The former method starts to segment sucal basins using morphological and region growing techniques and later uses model matching procedure based on a volumetric shape representation and on a model of spatial variation for automatically attribute anatomical labels to the segmented basins. Later, [Jaume, Macq et al., (2002)] proposed to progressively match an atlas labelled mesh to the patient brain mesh from the largest folds to the smallest folds and then transfer the labels from the matched mesh to label the patient mesh without manual intervention.

Taking into account all the construction processes of active shape models and active appearance models, the most time consuming step is the landmark annotation, because is usually manually made. So, considering this fact, we choose to develop algorithms capable of automating this step for the objects studied in this work: hand palms and faces. In the following sections we will describe our methods.

2.1 Hand labelling

The first step of this method consists in the segmentation of the considered object; considering that the images of the training set normally do not have a homogeneous background, the usual threshold application followed by an edge detector revealed not to be adequate to segment the hand. An alternative solution is to first extract the object area using an algorithm to detect skin regions [Carvalho and Tavares, (2007) ; Gonçalves, Tavares et al., (2008); Tavares, Carvalho et al., (2008)].

This algorithm uses a representative skin model, based on the training set of skin samples from the hand. Images from the training set are originally represented in *RGB* format, with the components (R,G,B) meaning the colour and luminosity of each pixel. Studies like [Zheng, Daoudiy et al., (2004)] show that skin colour has a great variability in *RGB* format, but its variation is smaller than the luminosity; this means that the skin of different people is similar varying only the intensity. Due to this fact, the representative skin model is obtained in the *YCbCr* (*luminance-chrominance*) format, where Y represent the intensity, Cb the blue component and Cr the red component of the image. Chromatic colours can be obtained from the *RGB* colour space using the following transformation:

$$Cr = \frac{R}{R+G+B}, \quad Cb = \frac{B}{R+G+B}, \tag{1}$$

where R, G and B correspond to red, green and blue components in *RGB* colour space.

The skin distribution function can be represented by a Gaussian model $N(\mu,C)$, where μ is the mean of each component and C the covariance between them. So,

we can define the probability density function of the Gaussian distribution as:

$$f(Cr,Cb) = e^{-\frac{1}{2}(x-\mu)^T C^{-1}(x-\mu)}, \quad (2)$$

where $x = (Cr\ \ Cb)^T$. Fig.1 shows the probability density function $f(Cr,Cb)$ considered for the hand. The mean values obtained for variables Cr and Cb were 152 and 111, respectively; these values represent the places where the probability of a pixel belonging to a skin region is higher in our examples. In order to segment all the images of the training set, we apply an adaptive threshold to find skin regions in each image (Fig.2 a-c).

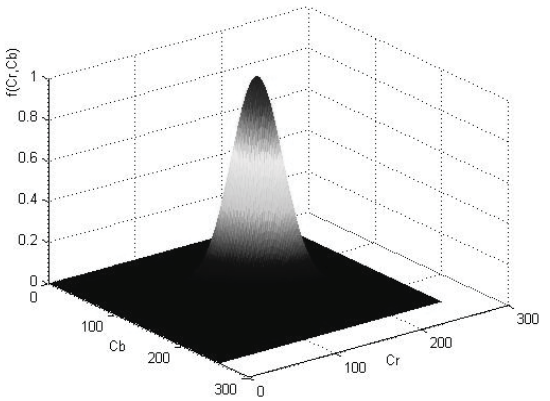

Figure 1: Probability function obtained in the construction of the representative skin model of the hand.

After obtaining the binary image, we apply a function that traces the exterior boundary of the hand and then finds the contour zones with highest curvature, through the k-curvature method [Lim, (1990)]. The strategy used to find these zones is as follows: first we identify the high curvature zones and after that locate the point between the thumb and wrist or between the wrist and little finger with higher curvature; finally, we search for the corresponding point in the opposite site [Vasconcelos and Tavares, (2006)]. Later, the contour of the hand is obtained and the other contours associated with the wrist and the arm are deleted.

In the last step, we choose the number of landmarks points to be considered on high curvature zones and also the number of landmarks points between them; in order to get the landmarks points desired, we make a sample of the extracted contour.

In Fig.2, we present an example of a training image (Fig.2a), the result of applying the probability function on it (Fig.2b), the segmentation result using the skin

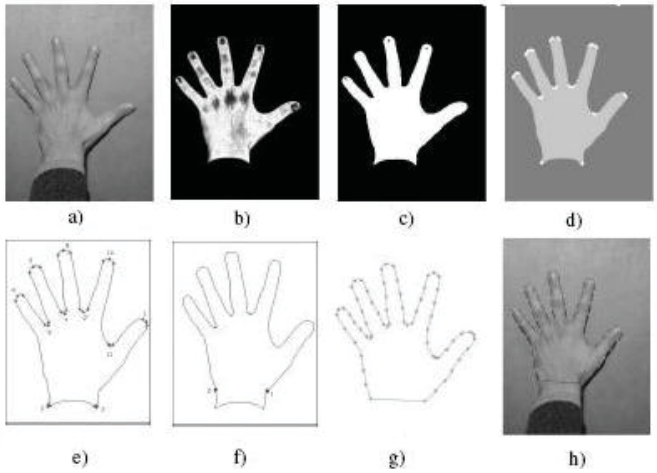

Figure 2: a) Training image, b) image obtained with the representation of the probability of the existence of skin, c) segmentation result using the skin algorithm, d) hand contour and high curvature zones found (white), e) identification of the high curvature zones, f) contour points (1 and 2) that delimit hand contour, g) landmark points extracted using this method and h) final contour of the training image.

algorithm (Fig.2c), the contour found with the high curvature zones detected and correspondent identification (Fig.2d and 2e), the contour points of the hand contour (Fig.2f), an example of the landmarks points extracted using our method (Fig.2g) and the final contour of the training image (Fig.2h).

2.2 Face contour extraction

In order to obtain good segmentation results, the authors of Point Distribution Models [Cootes, Edwards et al., (1998); Cootes and Taylor, (1992); Cootes, Taylor et al., (1992)] suggested that the landmark points used to train the models were key points of the objects, such as important zones of the boundary or significant internal locations of the object. Thus, this method was developed to extract significant points of faces represented in images; namely, on chin, eyes, eyebrows and mouth.

Like the method presented in 2.1, this one starts to use the skin detection algorithm to find the face region in the images. The representative skin model of the face has mean values obtained for variables Cr and Cb were 149 and 106, respectively. After applying the skin detection algorithm and converting the obtained grey images to binary images with the adaptive threshold algorithm, we still need to know which region belongs to the face, so we choose the region by studying the number of

holes that each region has; face region probably will have holes representing the eyes, eyebrows, nose and mouth that will not be identified as skin regions.

Studies, for example [Campadelli, Cusmai et al., (2003); Hsu, Abdek-Mottaleb et al., (2002)], showed that the use of chrominance maps are useful for eyebrows and eyes localization in images. Usually, eyes are characterized in *CbCr* plane by low values on the red component, *Cr*, and high values on the blue component, *Cb*, so the chrominance map for eyes can be defined by the following equation:

$$EyeMap = \frac{1}{3}\left\{ \left(Cb^2\right) + \left(\hat{Cr}\right)^2 + \left(\frac{Cb}{Cr}\right) \right\}, \tag{3}$$

where Cb^2, $\hat{Cr}^2$ and Cb/Cr are normalized to the range $[0,255]$ and $\hat{Cr}$ is the negative of Cr (ie, $\hat{Cr} = 255 - Cr$). In our work, the *EyeMap* was also used to effectively identify the eyebrows region.

In our method, the mouth region is identified using the *HSV* space, where *H*, *S*, *V* represent hue, saturation and value, respectively; since the mouth is habitually characterized by having high values on the saturation component.

Through the congregation of contours of the face, eyebrows, eyes and mouth, it is finally possible to extract landmark points from each of these zones. Considering that the zone of the chin is the most important segment of the face contour, we only use the inferior part between ears.

In Fig.3 we show some results of a training image example using this method to automatically extract landmark points of faces represented in images.

2.3 Face regular mesh

The method presented in this section was developed with the intention of verifying the importance of the use of significant points to model an object. So, we describe an algorithm able of automatically built active appearance models, through considering the construction problem as image coding problem, with the landmarks used being the nodes of a mesh defined on the object.

The former method built the active appearance models using landmark points as the nodes of a regular rectangular mesh placed over the face in analysis, considering that landmarks do not need to represent significant locations of the object.

In the method developed by us, the mesh used is correctly adjusted over the face to model. Our method starts to identify face and eye regions using the same algorithms described in the previous section, adjusts a regular rectangular mesh to the detected face region, rotating it according to the angle given by the centroids of the eyes. The mesh nodes obtained are then considered as landmark points of the object and used to build the active appearance model.

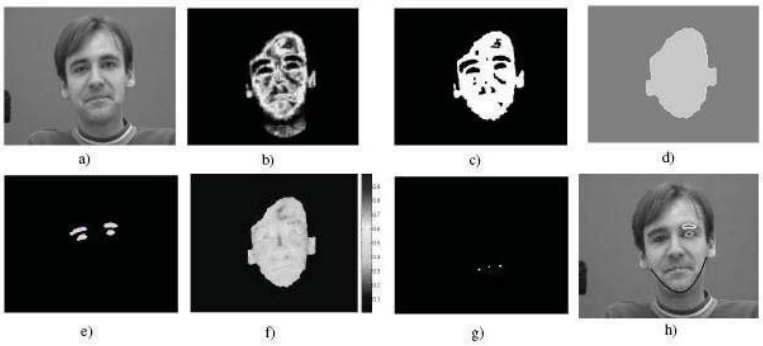

Figure 3: a) Training image, b) image obtained with the representation of the probability of the existence of skin, c) segmentation result using the skin algorithm, d) face contour extracted, e) identification of eyes and eyebrows using the chrominance map, f) Saturation component of the training image, g) mouth identification and h) contours extracted using the automatic face contour extraction.

In the implementation of this method, the user should choose the number of landmark points to be used in the x axis and y axis, so that he can control the total number of landmarks generated. Fig.4 shows the face mesh result obtained in a training image example using this method to automatically extract landmark points of faces represented in images.

Figure 4: a) Training image with face contour, face centroid and eyes centroids overlapped, b) face regular mesh adapted to the face region detected (face contour) and rotated according to the eyes direction.

2.4 Face adaptative multiresolution mesh

Finally, and after comparing the segmentation results of the previous two methods, we thought that the combination of the philosophy of the first method described for

faces represented in images, using face, eyes and mouth localization with the previous method that considers landmark points as the nodes of the defined mesh should give interesting results. Hence, in this new method we build a multiresolution mesh considering the positions of face, eyes and mouth.

Fig.5a presents a scheme explaining the construction of the multiresolution mesh. After locating the face, eyes and mouth regions in the input image using the method described before, this method constructs adaptive meshes, detecting the eye and mouth regions according to their localization; and then adds additional nodes in the larger mesh defined by the external edges and the bounds of the sub-meshes used in the regions of eyes and the mouth. In Fig.5b and 5c we present two examples of the resulting final mesh using our method for faces represented in images. In this method, it is possible to choose between two types of mesh (type 1 or 2), with the second type including more landmarks in the exterior mesh (Fig.5a).

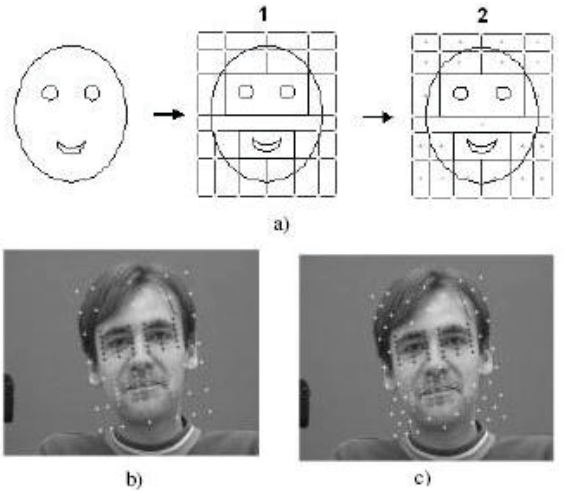

Figure 5: a) Construction scheme of the face adaptive multiresolution meshes (type 1 and type 2), b) and c) examples of the face adaptive multiresolution mesh of each type.

In the implementation developed for the last two methods to automatically extract landmark points of faces represented in images, the user can choose the parameters that define the resulting mesh; that is, the number of landmark points defined in each interesting zone of the object to be modelled, and in the multiresolution method with the preferred type of the mesh.

3 Point Distribution Models and Their Variants

The point distribution model is the basis for the construction of active shape models and active appearance models. Full details of the PDM can be found in [Cootes, Taylor et al., (1992)] but the following gives a brief description. The methodology used to build PDM allows obtaining a model that represents the mean shape of an object, as well as the admissible deviations to its shape starting with a set of images from the object in analysis (Fig.6).

Figure 6: Training images used to build the a) hand model and b) face model.

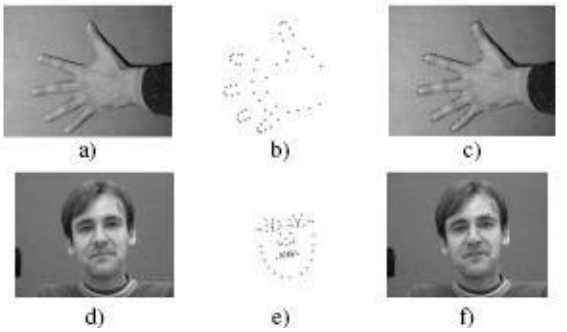

Figure 7: a) and d) Training images, b) and e) landmarks points considered and c) and f) images labelled with landmark points considered.

Each shape from the training set is represented by a set of labelled landmark points (Fig.7). In order to study the variation of the coordinates of the landmark points of the training shapes it is necessary that they are aligned. An example of an alignment method that can be considered is given in [Oliveira and Tavares, (2008)], which is based on dynamic programming.

Given co-ordinates (x_{ij}, y_{ij}) at each feature i of object j, the shape vector is:

$$x_i = (x_{i1}, x_{i2}, \ldots, x_{in}, y_{i1}, y_{i2}, \ldots, y_{in})^T,$$

where $i = 1, \ldots, N$, with N representing the number of shapes in the training set and n the number of landmark points. Once the shapes are aligned, the mean shape and variability can be found. The mean shape $\bar{x}$ is calculated through:

$$\bar{x} = \frac{1}{N} \sum_{i=1}^{N} x_i.$$

The modes of variation characterize the ways that landmarks of the shape tend to move together and can be found applying principal component analysis to the deviations from the mean. So, each vector x_i can be rewritten as:

$$x_i = \bar{x} + P_s b_s, \tag{4}$$

where x represents the n landmark points of the new shape of the modelled object, (x_k, y_k) is the position of landmark point k, $\bar{x}$ is the mean position of landmark points, $P_s = (p_{s1}\ p_{s2}\ \ldots\ p_{st})$ is the matrix of the first t modes of variation, p_{si} correspond to the most significant eigenvectors in a Principal Component Analysis of the position variables, and $b_s = (b_{s1}\ b_{s2}\ \ldots\ b_{st})^T$ is a vector of weights for each variation mode of the shape. The equation above represents the Point Distribution Model of an object and can be used to generate new shapes of it.

The local grey-level environment of each landmark point can also be considered in the modelling of an object [Cootes and Taylor, (1993)]. Thus, statistical information is obtained about the mean and covariance of the grey values of the pixels around each landmark point. This information is used in the PDMs variations: to evaluate the match between landmark points in Active Shape Models and to construct the appearance models in Active Appearance Models, as we will explain next.

3.1 Active Shape Model

The combination of PDM and the grey level profiles for each landmark of an object can be used to segment this object in new images through the Active Shape Models, an iterative technique for fitting flexible models to objects represented in images [Cootes and Taylor, (1992)].

The referred technique is an iterative optimisation scheme for PDMs allowing initial estimates of pose, scale and shape of an object to be refined in a new image. The used approach is summarized on the following steps: 1) at each landmark point of

the model calculate the necessary movement to displace that point to a better position; 2) calculate changes in the overall position, orientation and scale of the model which best satisfy the displacements; 3) finally, through calculating the required adjustments to the shape parameters, use residual differences to deform the shape of the model.

In [Cootes, Taylor et al., (1994)], the authors presented an improved active shape model which uses multiresolution. So, the proposed method first constructs a multiresolution pyramid of the input images by applying a Gaussian mask, and then studies the grey level profiles on the various levels of the pyramid built, making active models faster and more reliable.

3.2 Active Appearance Model

This approach was first proposed in [Cootes, Edwards et al., (1998)] and allows the building of texture and appearance models. These models are generated by combining a model of shape variation (a geometric model), with a model of the appearance variations in a shape-normalised frame. The statistical model of the shape that uses it is also described by Eq. (4).

To build a statistical model of the grey level appearance, we deform each example image so that its landmark points match the mean shape of the object, by using a triangulation algorithm. We then sample the grey level information, g_{im} from the shape-normalised image over the region covered by the mean shape. In order to minimize the effect of global light variation, we normalize this vector to obtain g. After applying a Principal Component Analysis to this data, we obtain a linear model called the texture model:

$$g = \bar{g} + P_g b_g, \tag{5}$$

where $\bar{g}$ is the mean normalised grey level vector, P_g is a set of orthogonal modes of grey level variation and b_g is a set of grey level model parameters. Therefore, the shape and appearance of any example of the modelled object can be defined by vectors b_s and b_g.

Since there may exist correlation between the shape and grey levels variations, we apply a further Principal Component Analysis to the data. Thus, for each training example we generate the concatenated vector:

$$b = \begin{pmatrix} W_s b_s \\ b_g \end{pmatrix} = \begin{pmatrix} W_s P_s^T (x - \bar{x}) \\ P_g^T (g - \bar{g}) \end{pmatrix}, \tag{6}$$

where W_s is a diagonal matrix of weights for each shape parameter, allowing the adequate balance between the shape and the grey models. Then, we apply a Principal

Component Analysis on these vectors, giving a further model:

$$b = Qc, \tag{7}$$

where Q are the eigenvectors of b and c is the vector of appearance parameters controlling both the shape and the grey levels of the model. In this way, an example object can be obtained for a given c by generating the shape-free grey level object, from the vector g, and be deformed it using the landmark points described by x.

4 Results

The methods described in this paper were used to automatically build active shape and active appearance models for objects like hand palms and faces represented in images.

The new methods described to automatically extract landmarks from objects like hand palms and faces were implemented in MATLAB. We developed an application in MATLAB to build active shape models, using the *Active Shape Models software* [Hamarneh, (1999)] as basis. For the appearance models, we used the *Modelling and Search Software* available in [Cootes, (2004)]. The images of hand palms and faces used in this analysis are, respectively, available in [Stegmann and Gomez, (2002)] and [Cootes, (2004)].

4.1 Hand palm models analysis

To analyse the hand palms models, we used a training set of 25 images and other 5 test images [Stegmann and Gomez, (2002)], not contained in the training set, each one was automatically labelled with 79 landmark points along its boundary using the hand labelling method previously mentioned in section 2.1 (Fig.7).

The training set had images from 4 different Caucasian subjects, 3 male subjects and 1 female subject. The images were acquired with similar background and approximately the same light conditions. The images used had hand palms with fingers in different positions, but always without touching it selves.

A shape model was built and trained on this data set, and through Tab. 1, we can see that the first 4 modes of variation can explain 90% of all the shape variance of the modelled object and the first 11 modes of variation can explain 99% of all the shape variance.

The effects of varying the first three modes of variation are shown in Fig.8 and basically consist in the combination of global transformations (first mode) and local transformations as the independent movement of each finger (second and third modes).

Table 1: The first 5 eigen values of the obtained model of the hand and their cumulative retained percents.

	Eigen value	Retained Percent
λ_1	26863	70.73%
λ_2	3123.6	78.95%
λ_3	2866.4	86.50%
λ_4	1521.7	90.50%
λ_5	1322.3	93.99%

With the purpose of analysing the differences between models in the case that the retained percentage varies, we built two types of models: one that explains 95% of all shapes variance and the other that explains 99% of all the shapes variance.

We also built models with different grey level profile dimensions to analyse the methods sensibility through this variation: so, we create models with reduced profiles (7 pixels, 3 to each side of the landmark point), medium profiles (15 pixels) and large profiles (33 pixels).

The hand images size is 800×600 pixels and the active search runs at most 10 iterations on each resolution level, starting from level 3. The maximum number of iterations chosen was the number that presented better results, an inferior number was not sufficient and a superior one was not necessary.

In order to measure the correct segmentation of the modelled object in new images we calculate the minimum and maximum values, mean and standard deviation of the Euclidean distances of both the landmarks of the object and the obtained result.

Fig.9 presents the segmentation result using the automatic active shape model in a test image. The grey level profile used for the construction of this active shape model is 7 pixels long and the explained percent was 95%. The segmentation results using the same model on the others four test images are shown in Fig.10.

In Tab. 2, we show the obtained results of the mean and standard deviation which translates the adequate segmentation of each built model in each test image. In this table the blank cells refers to cases that the segmentation was not successful. In Fig.11, we represent graphically the values of Tab. 2, we show de mean errors (and standard deviations) obtained in each test image for all active shape models built; when the method fails we consider the mean error infinite.

Through the analysis of Tab. 2 and Fig.11, we can see that our models, in general, obtain good segmentation results and the ones that use longer profiles (p16) have

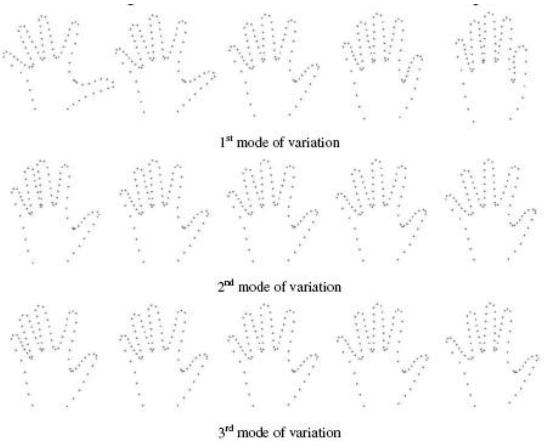

Figure 8: Effects of varying each of the first three parameters of the hand shape model ($\pm$2sd).

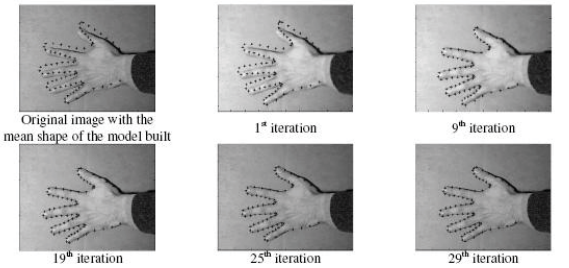

Figure 9: Test image with initial position of the mean shape model built overlapped, and after the 1^{st}, 9^{th}, 19^{th}, 25^{th} and 29^{th} iteration of the segmentation process using the active shape model built.

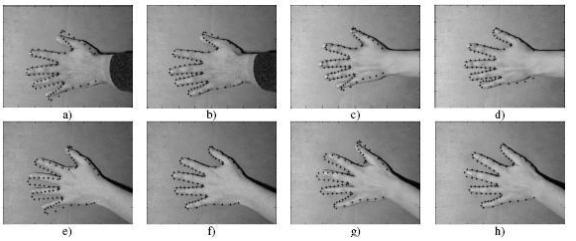

Figure 10: Test images with initial position of the mean shape model built overlapped (a, c, e, g) and the final results of the segmentation process (b, d, f, h) using the active shape model.

Table 2: Obtained errors of the active shape models built to the hand in the test images (mean and sd).

Models	Image 1	Image 2	Image 3	Image 4	Image 5
Hand_95_p3	5.38±3.18	-	10.91±4.32	3.64±3.19	5.88±3.77
Hand_95_p7	6.13±3.67	4.41±3.15	-	4.25±3.17	6.05±3.25
Hand_95_p16	5.09±2.96	10.01±5.50	-	-	-
Hand_99_p3	7.52±7.49	-	10.47±3.76	4.10±3.41	4.88±3.23
Hand_99_p7	-	-	-	3.39±2.69	11.73±5.56
Hand_99_p16	-	-	-	-	-

worse performances. In the models built, the ones who retain 95% of the variance achieve better results than the ones who retain 99%, which goes against our expectations. This situation can be explained, however, by the fact that using 11 modes of variation (99% of all shapes variance) the active shape model tends to over deform in the first iterations of the search, leading to worse results when compared with the models that explain 95% of all shapes variance.

As mentioned before, active appearance models can also identify objects in new images. We built active appearance models using the same images as the previous models and with the same landmark points. Models were built using 95% and 99% of all the shape variance and using different number of pixels in the texture model (5000, 10000 and 50000 pixels).

Using 95% of all the shape variance and 10000 pixels in the construction of the texture model, we extracted 6 variation shape modes, 17 texture modes and 11 appearance modes. On the other side, if we consider the active appearance model built using 99% of all the shape variance and considering the same number of pixels, we obtain 11 shape modes, 23 texture modes and 19 appearance modes.

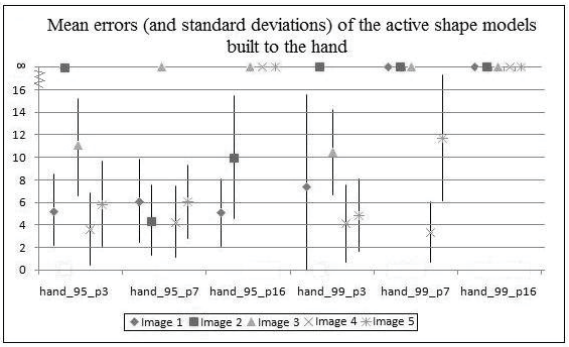

Figure 11: Obtained mean errors (± standard deviations) of the active shape models built to the hand in each 5 test images.

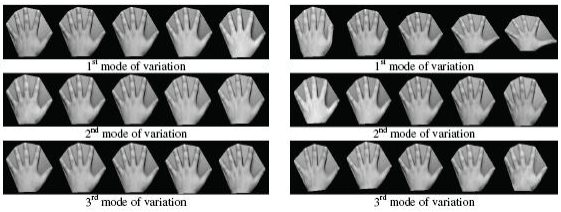

Figure 12: Variation effect (±2sd) of the first three modes of the texture model (left) and of the appearance model (right) built to the hand.

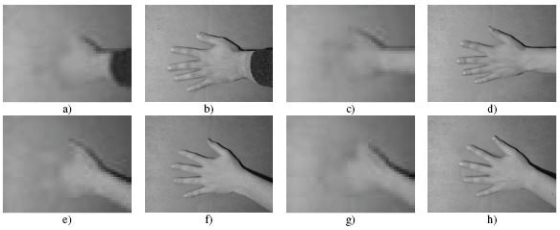

Figure 13: Test images with initial position of the mean shape model built overlapped (a, c, e, g) and the final results of the segmentation process (b, d, f, h) using the active appearance model.

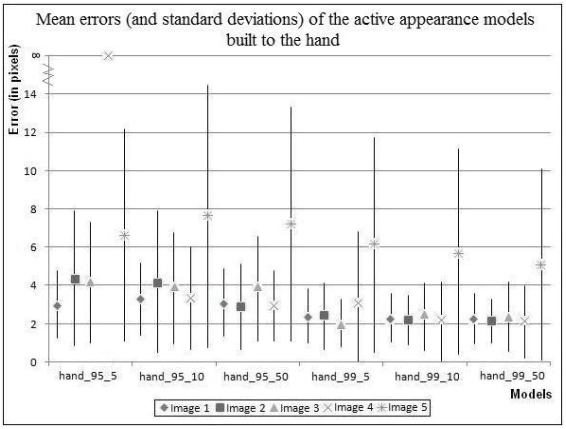

Figure 14: Obtained mean errors (±2sd) of the active appearance models built to the hand in each 5 test images.

In Fig.12, we present an example of varying the first three modes of the texture model and the appearance model. Through the analysis of the referred figure it is possible to see that the texture model considers the information about the individuals of the training set and the appearance model combines the shape with texture information. Fig.13 shows the segmentation results obtained for the four test images and Tab. 3 shows the errors of the segmentations using these models. Once again, blank cells referred to cases that the segmentation was not successful. Fig.14 translates graphically the values of the previous table, like in Fig.11, we represent the mean errors (and standard deviations) obtained for each test image for all active appearance models built; and in the cases where the segmentation fails the mean error was set to infinite.

Table 3: Obtained errors of the active appearance models built to the hand in the test images (mean and sd).

Models	Image 1	Image 2	Image 3	Image 4	Image 5
Hand_95_5	3.00±1.76	4.37±3.54	4.15±3.17	-	6.64±5.54
Hand_95_10	3.28±1.88	4.21±3.70	3.87±2.91	3.35±2.69	7.63±6.85
Hand_95_50	3.12±1.78	2.91±2.24	3.85±2.73	2.93±1.83	7.23±6.12
Hand_99_5	2.43±1.43	2.41±1.74	2.04±1.23	3.08±3.06	6.13±5.60
Hand_99_10	2.33±1.27	2.18±1.29	2.37±1.75	2.12±2.07	5.77±5.37
Hand_99_50	2.26±1.33	2.15±1.16	2.38±1.82	2.11±1.89	5.11±5.01

From the obtained errors, we can see that except of the failing of the first model on the segmentation of the fourth image, all others obtain good results. We can also observe that the models that use 99% of the variance generally obtain better results than the ones with 95% of the variance, as expected. The highest the variance retained by the model, the lowest is the error of segmentation in new images. But, it is important to refer that the amount of information used in these models is fairly superior than in the active shape models, justifying the differences obtained between the two models; from this we can conclude that active appearance models are more stable and robust than active shape models.

It is also possible to notice that the longer the used profile is, the better the segmentation results are, which is also an expected result.

In this work, we also generated hand models with different number of landmarks. According to our experiments, using 127 landmark points, 5 on high curvature zones and 8 between these zones, active shape models achieved better results in the segmentation in new images, with the same initial conditions. The mean error obtained decreased when we use the active shape models built with 127 landmarks around 2 pixels comparing with the shape models built using 79 landmarks. On the other side, the mean error obtained using the active appearance models with more landmark points did not change, which further confirms the stability and robustness

of these models.

4.2 Face models analysis

For modelling faces represented in images, we used a training set of 22 images and other 4 images for testing purpose [Cootes, (2004)]. The referred images were randomly selected from sequence of frames taken from a video of a person engaged in conversation. The subject was in a room talking with another person over an extended period of time. A static camera was trained on each individual, framed so that the movement of the face during the sequence was almost entirely within the image. The size of images is 720×576 pixels.

The active shape model was build using the first method presented in section 2.2 to automatically extract landmark points of faces represented in images and other two methods for the same purpose were used to build active appearance models. The goal where was to check if the use of landmarks that represent significant boundary zones is important to construction of the models and thus means better segmentation results.

We will present results for active models using the three approaches described in section 2: we extracted 44 landmark points with the face contour method; 49 landmark points with the regular mesh approach; with the third method, we extracted 54 and 75 landmark points, respectively with type 1 and type 2.

For the automatic active shape model built, using 44 landmark points, the first 10 modes of variation could explain 90% of all the shape variance of the object modelled, Fig.15 shows the first 3 modes of variation.

The parameters of the models generated for the face were identical to the ones built for the hand. We built models considering 95% and 99% of all shape variance and with reduced, medium and long profiles.

The size of the face images is 720×576 pixels and the active search runs at most 10 iterations on each resolution level, starting from the image of level 3. Again, the maximum number of iterations chosen was the number who presented better results, an inferior number was not sufficient and a superior one was not necessary.

In order to measure the quality of segmentation of the modelled object in new images we calculate the minimum and maximum values, mean and standard deviation of the Euclidean distances of both the landmarks of the object and the obtained result.

Fig.16 presents the segmentation results using the active shape model built in three test images. The grey levels profile used for the construction of these active shape models were 7 pixels long and the explained percent was 95%. In Tab. 4 and Fig.17 we show the resultant errors of the segmentations using the built models.

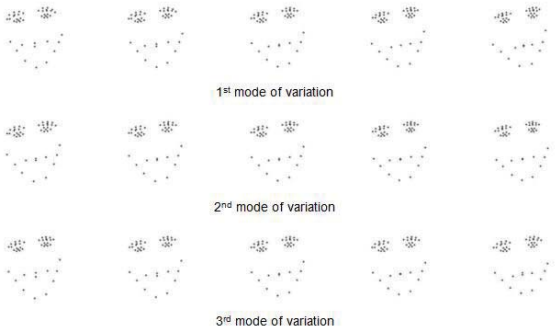

Figure 15: Effects of varying each of the first three parameters of the face shape model ($\pm$2sd).

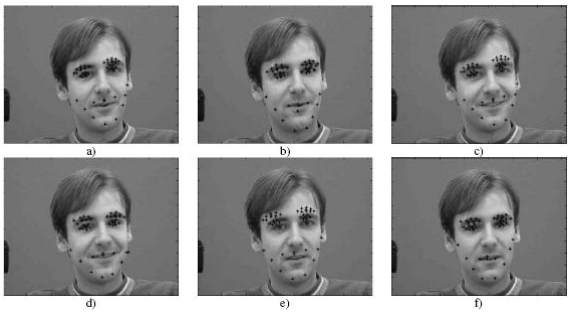

Figure 16: Test images with initial position of the mean shape model built overlapped (a, c, e) and the final results of the segmentation process (b, d, f) using the active shape model.

Table 4: Obtained errors of the active shape models built to the face in the test images (mean and standard deviation).

Models	Image 1	Image 2	Image 3	Image 4
Face_95_p3	7.76±5.02	7.48±7.71	8.63±8.21	12.10±9.35
Face_95_p7	6.20±5.26	8.19±7.48	9.05±8.59	7.33±6.16
Face_95_p16	8.85±6.56	10.23±7.60	10.49±6.87	8.22±6.20
Face_99_p3	12.45±8.68	10.17±9.30	11.38±10.16	7.40±5.61
Face_99_p7	7.85±6.65	11.86±8.18	11.27±7.90	9.73±7.32
Face_99_p16	9.25±6.43	10.49±10.13	9.45±6.71	11.50±8.37

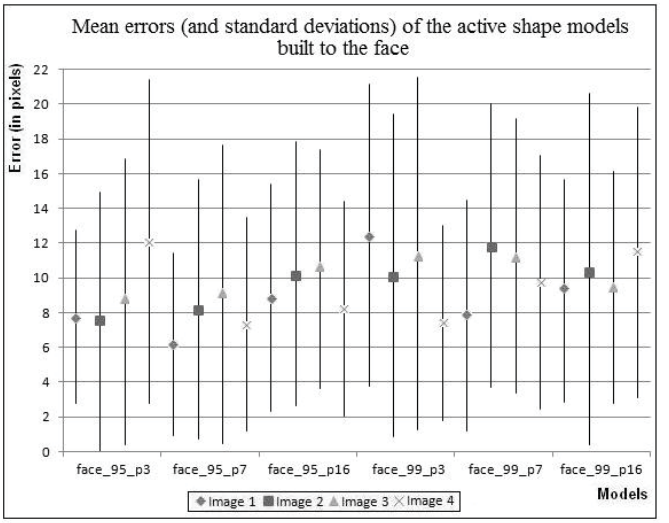

Figure 17: Obtained mean errors ($\pm$ standard deviations) of the active shape models built to the face in each 5 test images.

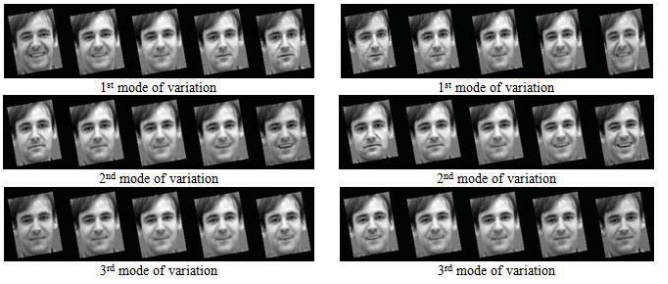

Figure 18: First three modes of texture (left) and appearance (right) variation of the adaptive multiresolution face mesh model ($\pm$2sd).

Table 5: Obtained errors of the active appearance models built to the face in the test images (mean and standard deviation).

Models	Image 1	Image 2	Image 3	Image 4
Face_44_95_5	6.14±4.11	4.81±4.50	5.05±3.63	-
Face_44_95_10	6.11±4.15	4.74±4.52	5.06±3.70	5.68±4.16
Face_44_95_50	5.17±3.59	4.08±4.46	4.83±3.69	5.37±3.97
Face_49_95_5	3.17±1.09	2.79±0.49	4.92±1.25	-
Face_49_95_10	2.93±0.22	1.30±0.10	3.60±1.80	2.54±1.77
Face_49_95_50	2.97±0.54	1.30±0.17	3.51±1.69	2.56±1.76
Face_54_95_5	1.52±0.83	3.63±3.65	2.86±1.52	2.08±1.27
Face_54_95_10	1.58±0.43	3.39±3.31	2.60±1.90	1.70±0.85
Face_54_95_50	1.56±0.64	3.55±3.24	2.70±0.94	1.71±0.91
Face_75_95_5	1.47±0.78	3.59±3.59	2.77±1.61	2.05±1.14
Face_75_95_10	1.65±0.54	3.69±3.57	2.70±1.04	2.08±1.19
Face_75_95_50	1.56±060	3.39±3.28	2.53±0.81	1.63±0.85
Face_44_99_5	5.18±3.20	4.68±4.91	5.24±3.57	5.59±4.17
Face_44_99_10	5.23±3.18	4.56±4.71	5.31±3.50	-
Face_44_99_50	5.20±3.35	4.31±4.55	4.90±3.48	6.04±5.02
Face_49_99_5	1.25±0.23	1.41±0.06	3.48±1.62	2.46±1.42
Face_49_99_10	1.40±0.22	1.34±0.13	3.54±1.53	2.39±1.56
Face_49_99_50	2.12±0.12	1.26±0.03	3.78±1.49	2.41±1.60
Face_54_99_5	2.17±0.64	-	-	-
Face_54_99_10	1.88±0.71	3.61±3.53	-	3.34±1.71
Face_54_99_50	1.67±0.52	3.37±3.54	2.61±1.01	3.05±1.61
Face_75_99_5	2.16±0.60	-	-	3.29±1.51
Face_75_99_10	1.80±0.67	3.49±3.51	2.77±1.11	3.29±1.58
Face_75_99_50	1.66±0.53	3.27±3.39	2.60±0.98	3.03±1.58

Through the observation of Tab. 4 and Fig.17 we can see that all the models were capable of identifying the object in a new image and also that they achieved good segmentation results. As for the case of the hand, the models that retained 95% of all the shape variance had better results than the ones with 99%. Following the former justification, using more variation modes will make the model to over deform in the first iterations of the active search. Looking to the dimension of the used profiles, it is possible to see that the medium profile achieves the best results.

In order to analyse the results of using automatic methods to build active appearance models, we constructed several models using the same 22 images of the training set with 44, 49, 54 and 75 landmark points generated by the three methods presented in section 2. So, we built models using 95% and 99% of all the shape variance and using different number of pixels in the texture model (5000, 10000 and 50000 pixels).

In Fig.18, we show the first three modes of variation of the texture and appearance using the model built with 54 landmarks and the adaptive multiresolution mesh of type 1. Through analysing the figure we can see that the texture modes combine the expressions of the individual and the appearance model combine the face shape

with the texture.

In order to analyse the efficiency of the use of automatic extraction methods in active appearance models in recognizing objects in new images, we used 4 test images. In the search stage, the maximum number of iterations by level was 10 and segmentation began at resolution level 5. In Fig.19, we present some iterations of the search process of the active appearance model built using the face contour extraction landmark method with 95% of all shape variance and 10000 pixels used on one test image. In Tab. 5, we represent the mean and standard deviation errors of the segmentation using different models.

Through analysing the errors in Tab. 5, we can see that the models are capable of identifying the object in new images. In Fig.20, we can see that the difference between using 95% or 99% of all shape variance is not so significant; however, the models who retained 99% of all shape variance fail more often. Looking at Fig.21, we can confirm that the models who obtain better results are the ones who use longer texture profiles, since it uses more information of the object texture.

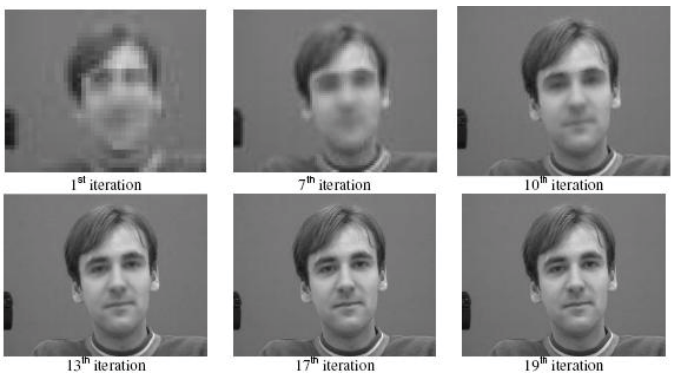

Figure 19: 1^{st}, 7^{th}, 10^{th}, 13^{th}, 17^{th} and 19^{th} iteration of the segmentation process using the active appearance model built for the face contour model.

The face contour extraction method (44 landmark points) is the one with the worst result. The face adaptive multiresolution mesh type 2 using 95% of all shapes variance obtains the best results on segmentation. On the other side, the face regular mesh using 99% of the variance could gain better segmentation results. The face regular mesh and face adaptive multiresolution mesh always obtain better results than the face contour extraction method, leading us to speculate that it is possible to built shape and appearance models not considering as landmarks significant points of the contour and also obtain good segmentation results.

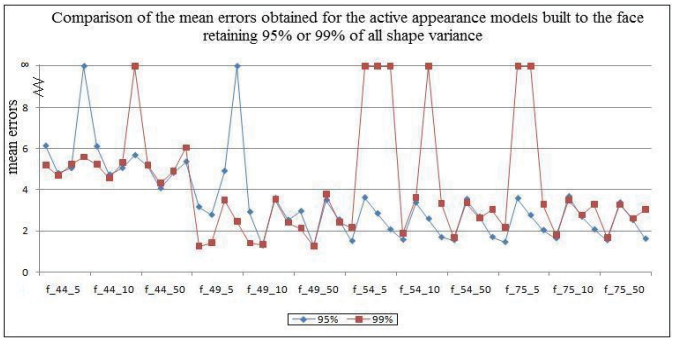

Figure 20: Comparison of the mean errors obtained for the active appearance models built to the face retaining 95% or 99% of all shape variance.

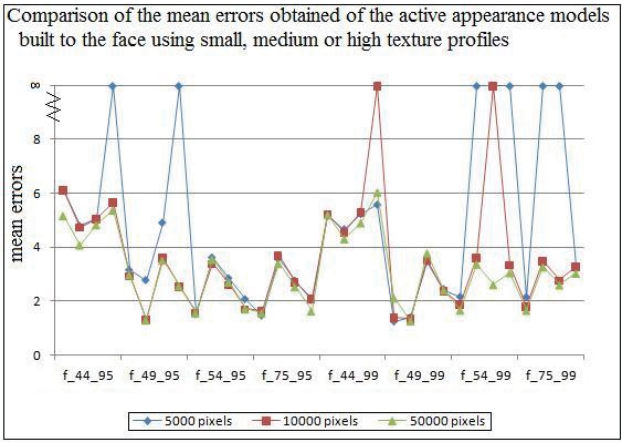

Figure 21: Comparison of the mean errors obtained of the active appearance models built to the face using small, medium or high texture profiles.

At last, considering the segmentation results obtained for both hand and face with the same initial conditions, active appearance models always have better performances than active shape models. The mean and standard deviation of the segmentation errors are lower in appearance models as can be observed by comparing from Tab. 2 to 5. This occurs because active appearance models use much more information of the grey levels than active shape models.

5 Conclusions and Future Work

In this work we improve point distribution model by presenting new methods capable of automatically extract landmarks from objects like hand palms and faces represented in images. Additionally, we described the construction of point distribution models, its variants and some applications of these models.

Point distribution models are built through the analysis of the statistics of the coordinates of the landmarks that represent the deformable object in analysis. After aligning the object shapes, a Principal Component Analysis is made and the mean shape of the object and the main modes of its variation are obtained. PDMs can be used to identify (segment) the studied objects in new images through active shape models and active appearance models. These models are built through combining the shape of the objects with its grey levels.

As the step of extracting the landmarks normally needs manual operations, it is usually the most time consuming step of the construction of the point distribution model. So we decided to develop four methods capable of automatically extracting landmarks from objects like hand palms and faces. The first two methods were constructed in order to replace user intervention for an automatic method, then we built another method to verify if it was really important to use as landmarks significant points of the object to be modelled and finally, and due to the results of the previous method, we joined the principles of the former two methods and analysed the result of it.

So, the first method, applied on objects like hand palms, starts by using an algorithm to identify skin regions and then to obtain the landmarks of hand palms by calculating the zones with higher curvature. The other methods also use an algorithm to detect skin regions in images; the face contour method places landmarks in regions like eyes, eyebrows, mouth and chin; the face regular mesh method considers the landmark points as the nodes of a regular rectangular mesh adjusted over the face region; and the face adaptive multiresolution mesh uses the previous two methods to build a multiresolution mesh considering the positions of face, eyes and mouth.

The methods developed to automatically extract landmark points from objects like hand palms and faces represented in images showed to be reliable, considering the images used, and allow the building of active shape models and active appearance models in almost fully automatic way (the user only has to specify the number of landmarks to use).

Both active shape and active appearance models obtained very interesting segmentation results of the modelled objects in new images. Active shape models only use the information around each landmark point of the modelled object, while the

active appearance models use also the grey level information of the object. The first one tends to be less robust than the latter. Because of this, it is advisable to use landmarks with significant locations to build active shape models. On the other side, active appearance models allow the construction of a robust model using relatively few landmark points. Therefore, the first one is preferred in problems in which the extraction of landmark points is not an easy process.

The segmentation results obtained in this work showed that the active appearance models built with the regular face mesh model and the adaptive face mesh model give better results than the face contour model, leading us to believe that, at least for active appearance models, the use of landmarks as the nodes of a regular mesh is easier, less demanding computationally and presents better results than choosing significant points as landmarks.

The next step in this work could be, for instance, related with the test of these models considering images with different illumination or with different subjects, regarding the improvement of the robustness of the models built.

As for future work, making the automatic extraction of landmark points of other type of objects, like bones or organs will facilitate the use of point distribution models in medical imaging. The study of point distribution models on 3D objects is also an interesting area to be focused in future works. Other possibility is building of statistical models that use the physical proprieties of the modelled objects.

Acknowledgement: The first author would like to thank the support of the PhD grant SFRH/28817/2006 from FCT- Fundação para a Ciência e Tecnologia from Portugal. The work presented was partially done in the scope of the project "Segmentation, Tracking and Motion Analysis of Deformable (2D/3D) Objects using Physical Principles", with reference POSC/EEA-SRI/55386/2004, financially supported by FCT.

References

Aixut, T.; Meneses, Y. L., et al. (2003): Constraining deformable templates for shape recognition, 6th Conference on Quality Control by Artificial Vision - QCAV, Tennessee, USA.

Angelopoulou, A. N.; Psarrou, A. (2004): Evaluating Statistical Shape Models for Automatic Landmark Generation on a Class of Human Hands. In *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Istanbul.

Bailleul, J.; Ruan, S., et al. (2003): Automatic Atlas-based Building of Point Distribution Model for Segmentation of Anatomical Structures from Brain MRI.

In *7th International Symposium on Signal Processing and Its Applications*, Paris, France.

Baker, S.; Matthews, I. (2004): Automatic Construction of Active Appearance Models as an Image Coding Problem. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 26, No. 10, pp. 1380-1384.

Campadelli, P.; Cusmai, F., et al. (2003): A color based method for face detection. In *International Symposium on Telecomunications*, Isfahan, Iran.

Carvalho, F. J. S.; Tavares, J. M. R. S. (2007): Eye detection using a deformable template in static images. In *VIPimage - I ECCOMAS Thematic Conference on Computational Vision and Medical Image Processing*, Porto, Portugal.

Caunce, A.; Taylor, C. J. (1998): 3D Point Distribution Models of the Cortical Sulci. In *6th International Conference on Computer Vision*, Bombay, India.

Cootes, T. F. (2004): Build_aam, from `http://www.wiau.man.ac.uk/~bim/software/am_tools_doc/download_win.html`.

Cootes, T. F. (2004): Talking Face, from `http://www.isbe.man.ac.uk/~bim/data/talking_face/talking_face.html`.

Cootes, T. F.; Edwards, G. J., et al. (1998): Active Appearance Models. In *Proceedings of European Conference on Computer Vision*, Freiburg, Germany, Springer.

Cootes, T. F. and Taylor, C. J. (1992): Active Shape Models - 'Smart Snakes'. In *Proceedings of the British Machine Vision Conference*, Leeds.

Cootes, T. F.; Taylor, C. J. (1993): Active Shape Model Search using Local Grey-Level Models: A Quantitative Evaluation. In *British Machine Vision Conference*, Guildford, BMVA Press.

Cootes, T. F.; Taylor, C. J., et al. (1992): Training Models of Shape from Sets of Examples. In *Proceedings of the British Machine Vision Conference*, Leeds.

Cootes, T. F.; Taylor, C. J., et al. (1994): Active Shape Models: Evaluation of a Multi-Resolution Method for Improving Image Search. In *British Machine Vision Conference*, York, England, BMVA.

Gonçalves, P. C. T.; Tavares, J. M. R. S., et al. (2008): Segmentation and Simulation of Objects Represented in Images using Physical Principles. *CMES: Computer Modeling in Engineering & Sciences*, Vol. 32, No. 1, pp. 45-55.

Hamarneh, G. (1999): ASM (MATLAB), from `http://www.cs.sfu.ca/~hamarneh/software/code/asm.zip`.

Hicks, Y.; Marshall, D., et al. (2002): Automatic Landmarking for Building Biological Shape Models. *International Conference of Image Processing, Rochester,*

USA, Vol. 2, pp. 801-804.

Hill, A.; Taylor, C. J. (1994): Automatic Landmark Generation for Point Distribution Models. In *Fifth British Machine Vision Conference*, York, England, BMVA Press.

Hsu, R.; Abdek-Mottaleb, M., et al. (2002): Face detection in color images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 24, No. 5, pp. 696-706.

Jaume, S.; Macq, B., et al. (2002): Labeling the Brain Surface Using a Deformable Multiresolution Mesh. In *5th International Conference on Medical Image Computing and Computer-Assisted Intervention*, Tokyo, Japan.

Kass, M.; Witkin, A., et al. (1987): Snakes: Active Contour Models. *International Journal of Computer Vision*, Vol. 1, pp. 321-331.

Lanitis, A.; Taylor, C. J., et al. (1995): An Automatic Face Identification System Using Flexible Appearance Models. *Image and Vision Computing*, Vol. 13, No. 5, pp. 392-401.

Lim, J. S. (1990): *Two-Dimensional Signal and Image Processing*, PTR Prentice Hall.

Lohmann, G.; Cramon, D. Y. (2000): Automatic labelling of the human cortical surface using sulcal basins. *Medical Image Analysis*, Vol. 4, pp. 179-188.

Oliveira, F. P. M.; Tavares, J. M. R. S. (2008): Algorithm of Dynamic Programming for Optimization of the Global Matching between Two Contours Defined by Ordered Points. *CMES: Computer Modeling in Engineering & Sciences*, Vol. 31, No. 1, pp. 1-11.

Rijsdam, J. (1999): *An automatic left-ventricular search and fitting method using a three-dimensional point distribution model*, MSc thesis, Leiden University, Leiden.

Schaap, J. (1999): *3D Point Distribution Models and their application in medical image segmentation using Active Shape Models: A pilot study*, MSc thesis, Delft University of Technology, Delft.

Stegmann, M. B.; Gomez, D. D. (2002): Hand images, from `http://www2.imm.dtu.dk/pubdb/views/publication_details.php?id=403`.

Sung, J.; Kanade, T., et al. (2007): A Unified Gradient-Based Approach for Combining ASM into AAM. *International Journal of Computer Vision*, Vol. 75, No. 2, pp. 297-309.

Tavares, J. M. R. S.; Carvalho, F. J. S., et al. (2008): Computer Analysis of Objects' Movement in Image Sequences: Methods and Applications. *International Journal for Computational Vision and Biomechanics*, (in press).

Tian, Y.; Kanade, T., et al. (2000): Dual-state parametric eye tracking. In *Conference on Automatic Face and Gesture Recognition*, Grenoble, France.

Vasconcelos, M. J.; Tavares, J. M. R. S. (2006): Methodologies to Build Automatic Point Distribution Models for Faces Represented in Images. In *CompIMAGE - Computational Modelling of Objects Represented in Images: Fundamentals, Methods and Applications*, Coimbra, Portugal.

Wang, S.; Lim, K., et al. (2007): A Geometric Deformation Constrained Level Set Method for Strutural Shape and Topology Optimization. *CMES: Computer Modeling in Engineering & Sciences*, Vol. 18, No. 3, pp. 155-181.

Yuille, A. L.; Cohen, D., et al. (1992): Feature extraction from faces using deformable templates. *International Journal of Computer Vision*, Vol. 8, pp. 104-109.

Zhang, Y.; Cheng, B., et al. (2008): Kinematic analysis of lumbar spine undergoing extension and dynamic neural foramina cross section measurement. *CMES: Computer Modeling in Engineering & Sciences*, Vol. 29, No. 2, pp. 55-62.

Zheng, H.; Daoudiy, M., et al. (2004): Blocking Adult Images Based on Statistical Skin Detection. *Electronic Letters on Computer Vision and Image Analysis*, Vol. 4, pp.1-14.