Facial segmentation using the Active Shape Model for the contactless monitoring of vital signals

Raúl Díaz Morales

Tutors: Dipl.-Ing. Tom Wirthgen
         Dipl.-Ing. Sebastian Zaunseder

Responsible Professor: Prof. Dr.-Ing. habil. H. Malberg

Date: 22/10/2012
Abstract

Active Shape Model are used for over 15 years for pattern recognition. Many evolutions based on the original methods have been proposed particularly in the range of facial recognition and medical imaging. This work exposes a new adaptation by using ASM for the application in remote PhotoPlethysmography (rPPG), a method used to detect a person’s heartbeat without exerting any physical contact. Segmentation of a certain areas of the face can provide robustness against motion and changes in brightness for the detection in rPPG.
Contents

1 Introduction 9

2 State-of-the-Art 11
  2.1 Basics of heartbeat detection by rPPG . . . . . . . . . . . . . 11
    2.1.1 rPPG: Principles and drawbacks . . . . . . . . . . . . . 11
    2.1.2 rPPG state-of-the-art . . . . . . . . . . . . . . . . . . . 13
  2.2 Facial recognition . . . . . . . . . . . . . . . . . . . . . . . 16
    2.2.1 Methods for facial segmentation . . . . . . . . . . . . . 16
    2.2.2 Facial segmentation for rPPG . . . . . . . . . . . . . . . 17
    2.2.3 Contents of this work . . . . . . . . . . . . . . . . . . . 18
  2.3 ASM Background . . . . . . . . . . . . . . . . . . . . . . . . 20
    2.3.1 Point Distribution Model and Image Search . . . . . . . 20
    2.3.2 ASM characteristics . . . . . . . . . . . . . . . . . . . . 24
    2.3.3 Kernel Principal Components Analysis in ASM . . . . . . 26

3 Active Shape Model 29
List of Figures

2.1 Hemoglobin and oxyhemoglobin absorption .......................... 12
2.2 Pulse amplitude mapping principle ................................. 14
2.3 Heartbeat amplitude matching images .............................. 14
2.4 ASM featuring and possible segmentation ......................... 19
2.5 Model modes and generated shapes ............................... 22
2.6 ASM gray value search ............................................ 23
3.1 Training set images acquisition set-up ............................. 30
3.2 Example of facial self-occlusion .................................. 32
3.3 Facial landmarks distributions .................................... 34
3.4 Facial features matching quality test ............................... 36
3.5 Examples of matching quality ...................................... 37
3.6 Simplified landmarks distribution .................................. 42
3.7 Training set results for frontal views ............................... 44
3.8 Training set results: best and worst matching ................. 45
3.9 Training set results: best and worst matching ................. 45
List of Tables

2.1 Facial segmentation approaches ........................................ 17

3.1 Model turning poses restrictions. ....................................... 49

1 Description of the participants. ......................................... 69
List of Algorithms

1. ASM image search .............................................. 39
2. Shape alignment in ASM image search ..................... 39
3. Parameters updating in ASM ................................ 39
4. Combination of ASM image search ......................... 43
5. Eyes detection .................................................. 43
Abbreviations

**AAM**  Active Appearance Model

**ASM**  Active Shape Model

**BSM**  Bayesian Shape Model

**HSV**  Hue, Saturation and Value

**kPCA**  kernel Principal Components Analysis

**PDM**  Point Distribution Model

**PCA**  Principal Components Analysis

**rPPG**  remote Photoplethysmography

**RGB**  Red, Green and Blue

**ROI**  Region Of Interest
Chapter 1

Introduction

Nowadays, the importance of non-invasive methods in medical applications is fundamental. Such methods present numerous advantages and technology is increasingly capable to provide innovative solutions on the matter. Several vital signs can be monitored remotely (e.g. without exerting any physical contact). However, cardiovascular parameters are of particular interest since their disorders may immediately and drastically affect the health of a person. Contactless monitoring of the cardiovascular parameters offers solutions that are more comfortable physically and psychologically for the patients [25]. Moreover, it contributes to an easier utilization of the devices, even for non-specialist users.

Remote photoplethysmography (rPPG) can lead to all kind of new applications as much in the medical field as in everyday life [39, 23, 22, 21]. For instance, it could allow improvements on the medical remote assistance and facilitate easy monitoring of health conditions of patients (i.e. old people, patients with chronic diseases) without leaving their domestic environment. Another example is the possible application in non-medical environments. This could be the case for the monitoring of drivers, which can provide more safety in coordination with the technology of the car.

The remote detection methods of cardiovascular parameters are still in development. The biggest difficulty that researches have to deal with is the effectiveness in changeable environments. As the effects which are introduced by secondary factors (i.e. those which are not related to the changes in the blood flow) may exceed the desired signal by magnitudes. Interferences due
to changes of the illumination and motion are difficult to compensate. For this reason, there is still a long road ahead in this field of research, both to enhance the reliability of the detection and to develop possible applications.

The main purpose of this thesis is to improve rPPG in adverse conditions mainly due to the facial movements. To this end, segmentation of measuring areas is performed in order to reduce remarkably the effect of possible interferences. Furthermore, these facial regions are distinguished by some physiological properties in order to improve the quality of the desired signal. The method used to segment the face is a linear model called Active Shape Model (ASM). ASM provides the possibility to detect and track facial features. Once the face and its features are located, the system traces the areas of interest.
Chapter 2

State-of-the-Art

2.1 Basics of heartbeat detection by rPPG

It has been shown that it is possible to extract the heartbeat of a person without physical contact using a standard RGB camera and daylight as the illumination source \[25, 39\]. The associated procedures consist of measuring the variations of brightness in the face of the subject. These variations have a direct relation to the variations of the facial blood flow and therefore with the heartbeat.

2.1.1 rPPG: Principles and drawbacks

Since rPPG measures variations of the reflected light on the human face, a proper detection can be really difficult to perform. Several factors can interfere significantly with the measurement as noise or interfering signals. For instance, the subject under consideration most often is going to remain in motion. Facial movements can present large interfering amplitudes in the same frequency range as the signal of interest. Besides, motion entails changes in the background area and hence undesirable changes in the brightness of the image. In both cases the signal of interest might become untraceable.

Several solutions can be implemented in order to reduce the effect of these interferences. One approach is based on improving the robustness of
Figure 2.1: Hemoglobin and oxyhemoglobin absorption depending on the light wavelength [21].

A second approach is based on segmentation of facial regions where the signal of interest is comparable, or even better, to the interfering signals. In fact the combination of different light colors is not completely necessary, since it has been shown [21] that using only green light it is possible to detect the blood flow variations.

It is known that hemoglobin and oxyhemoglobin –hemoglobin in the state of transporting oxygen– have special absorption characteristics depending on the wavelength of the incident light (see figure 2.1). Specifically, the oxyhemoglobin presents high absorption of light in the spectral range 520-580 nm, that is to say, the range corresponding to green light. Hemoglobin is the second major chromophore for visible light in the skin [14] and can be considered representative in terms of blood light absorption, because the other elements are practically colorless (about 83% of the total is water).

Moreover, it is also important to mention that green light detection presents another advantage over the other light colors. RGB cameras present less noise in the detection of green light and, consequently, better behavior.
in this range of frequencies. This advantage is due the Bayer filters used as color filter array by this type of cameras, which distributes the pixels in such way that there is the same number of green pixels than both red and blue ones.

2.1.2 rPPG state-of-the-art

In the majority of the situations, the rPPG experiments have been carried out taking the entire face as an area of interest. Recently, some papers seem to show that it is possible to enhance the accuracy of the measurements by performing a segmentation of concrete facial regions. Lewandowska et al. [22] present a paper about remote pulse detection with a standard RGB camera. Furthermore, [22] also use thermal images to suggest that part of the forehead between the eyes is a suitable area. The purpose of the thermal images is to show that forehead temperature between different subjects remains relatively constant, which means that this is a region free of both motion and occlusion, and hence suited to extract the results from. Based on this assumption, a little area of forehead is selected. Even so, the segmentation is performed just in the first frame, assuming no facial motion in the whole detection process.

More elaborated studies demonstrate that the forehead is not the only appropriate part from which to take measures. The procedure applied in these cases is called ”pulse amplitude mapping” [39, 21]. It consists of obtaining the heartbeat frequency spectrum of a subject with an ear photoplethysmograph. At the same time, a video of the subject’s face is taken in order to calculate the frequency response in every pixel (or even region of pixels). Then the amplitudes corresponding to the pulse frequency are compared for both graphs (see figure 2.2).

The relation between amplitudes at the same frequency can be represented using color scaling. Figure 2.3 indicates that the forehead is a suitable region to take the measures, but also the cheeks and the part of the neck corresponding to carotid artery.

If in first place, the main aim consisted of avoiding interferences due to the motion or brightness, now it is also relevant to take specific areas of the face, suitable because of their physiological properties. In this case, these ROIs coincide with the areas where it is known that there is a higher blood volume near the surface of the skin (e.g. subpapillary plexus) and the light
Figure 2.2: Pulse amplitude mapping. a) Frequency response measured with ear photoplethysmograph. b) Pixel Frequency response and comparative between amplitudes at the heartbeat frequency [21].

Figure 2.3: Heartbeat amplitude matching images. The pixels in red and orange indicate that there is a higher amplitude in heartbeat frequency range and hence define suitable areas to detect the pulse [21].
absorption is then more noticeable. Based on this facts, this paper is oriented to perform a proper segmentation of the mentioned facial regions.

Therefore, a facial segmentation technique is required. In this field, multiple approaches have been presented based on one or several facial properties as, for example, color or shape. These different approaches present a wide range of levels and qualities of facial segmentation that have to be considered in order to choose properly the one capable to satisfy the rPPG requirements.
2.2 Facial recognition

2.2.1 Methods for facial segmentation

In the middle of the nineties, facial recognition and segmentation underwent a revival [5]. Several approaches were developed in order to provide models capable of locating and extracting facial features. Radeva and Mart [26] presented a free-form model, called rubber snake, capable to segment eyes, eyebrows and mouth with some precision problems. Saber and Tekalp [29] proposed a combination of color and shape algorithms (HSV) to extract the face, while the eyes, nose and mouth segmentation is performed geometrically over the space created by the eigenvectors of the skin spatial covariance matrix. Sobottka and Pitas [32] used color and shape algorithms for tracking the contour of the face as well, but apply a gray-level algorithm for the facial feature extraction. Mu et al. [24] performed eyes and mouth segmentation considering motion and texture differences between these areas and the rest of face over several video image restrictions.

During the last decade, the number of papers about facial segmentation has kept growing. For instance, Xue et al. [44] described feature extraction and facial recognition using a Bayesian Shape Model (BSM). Hammal et al. [18] proposed an algorithm with a combination of 5 independent curves for modeling the mouth and another less geometrical demanding combination of curves for modeling the eyes and eyebrows. Zuo et al. [45] presented a facial feature extraction method composed of a cascade of three different algorithms using a point-based model and a deformable parameter. Other papers formulate algorithms using approaches based on skin-color [6] or combination of different techniques [30, 3], as edge detection, k-means or labeling applied in 3D images.

Active Shape and Appearance Models

Cootes et al. [13] developed reference methods mainly aimed at medical image analysis, as the own Active Shape Model (ASM), that became useful not only for facial feature extraction, but also for other kinds of medical applications. This is the case of Van Ginneken et al. [38], who applied ASM for segmentation of chest radiography, Thodberg and Rosholm [35] used the
Table 2.1: Facial segmentation approaches.

<table>
<thead>
<tr>
<th>Skin Color</th>
<th>Shape (points and curves)</th>
<th>Texture</th>
</tr>
</thead>
<tbody>
<tr>
<td>Color Mapping</td>
<td>ASM</td>
<td>AAM</td>
</tr>
<tr>
<td>HSV</td>
<td>Bayesian</td>
<td>HSV</td>
</tr>
</tbody>
</table>

same method in radiographies of the hand for an accurate osteoporosis diagnosis and Smyth et al. [31] presented an automatic measurement of vertebral shapes.

With regard to ASM utilization in facial recognition, some papers have been presented in order to model the shape of the face and locate the facial features as mouth, nose and eyes [13, 37, 34]. In facial recognition commonly a more complex version of ASM is used that introduces texture parameters. This method is called Active Appereance Model (AAM) [11, 33, 17] and was formulated also by Cootes et al. Several papers use AAM for facial recognition [16, 28, 43].

3D facial recognition


2.2.2 Facial segmentation for rPPG

In this thesis, as already mentioned, segmentation has to be applied over forehead, cheeks and part of the neck. The regions of interest are located inside these parts and consequently are not salient anatomical features. This means that there are no special features regarding to color, texture or brightness inside their limits.
For this reason, both color and texture techniques cannot face the segmentation of the ROIs directly. Even though, the shape-based techniques are not suitable in this case, since they need to process the image information (gray-value, texture, etc) where the shape is going to be matched.

However, as seen in the previous points, these techniques are suited to track the face. Therefore, it is possible to define the ROIs once the face is recognized by knowing the position of its contour and features (i.e. from previous knowledge about the location of the head contour, eyebrows and the fringe, a precise forehead segmentation becomes feasible).

For the rPPG application, it is important to obtain a robust segmentation despite facial motion. Otherwise, the subject has to remain in a specific position during the whole measurement. Such restrictions would limit the applications of rPPG significantly. Consequently, a method capable to fix with precision the ROIs segmentation for any kind of facial pose is highly desirable.

Shape-based approaches present a particular advantage over other kind of methods. They are suited to define a geometrical facial frame into which locate the position of the ROIs, specifically, the point-based shape models where the shape of the modeled object is defined by points. As it can be deduced from literature and as shown in next sections, point-based shape models provide a precise and manageable segmentation.

The Active Shape Model is a point-based shape model capable to satisfy these segmentation requirements. Particularly, it is available an ASM implementation that has been already tested successfully for similar segmentation uses in [42]. In addition, ASM presents low computational costs, compared to other methods, like Bayesian Shape Model or Active Appearance Model, which don’t represent, on the other hand, meaningful advantages as far as the rPPG objectives are concerned.

2.2.3 Contents of this work

It has been shown that it is possible to detect contactless a person’s heartbeat with a standard camera and daylight as illumination source. As the way to do it, is measuring the variations of light absorption, several problems come up, since the signal of interest may become nearly untraceable
compared to the possible artifacts—motion, brightness changes, noise. It seems that segmentation can be a suitable solution to reduce the measuring areas and hence to improve the conditions of the light variations detection despite interferences. From previous studies, it can be concluded that Active Shape Models might be able to perform a robust segmentation of the areas of interest.

In the following, a brief background of Active Shape Models is given. Then the necessary materials to perform the segmentation, as image training set, are discussed and, afterwards, the model is built taking the algorithm and MVTec HALCON code used by Wirthgen et al. [12] as a base of the new implementation. The model is tested and re-evaluated in order to obtain a proper elasticity to perspective variations. Finally, several comments and conclusions are presented.
2.3 ASM Background

2.3.1 Point Distribution Model and Image Search

ASM was formulated for the first time by Cootes and Taylor in the early 90s and consist of two different techniques: Point Distribution Model (PDM) and image searching [12, 9].

**Point Distribution Model** PDM uses a collection of \( n \) points that form a shape. Each shape can be represented as a \( 2n \) dimension vector:

\[
x^T = (x_1, x_2, ..., x_i, y_1, y_2, ..., y_i)
\]

where \((x_i, y_i)\) is the \( i^{th} \) point of the shape, also known as \( i^{th} \) landmark. Landmarks are significant points annotated manually by a specialist in order to get the shape of the object that is to be modeled. Hence a set of images from the same object is needed in order to obtain a representative sample of the object’s shape variations. For each image a collection of landmarks defines a shape. It is important to notice that every landmark has to have the same relative position in each image, otherwise, it is impossible to establish any relation between different shapes. This set of \( s \) known shapes that define the object of study is commonly called a training set.

Procrustes Analysis is performed in order to align the shapes of the training set and obtain uniformity in their position, rotation and scale [10]. Once all the shapes in the training set are conveniently aligned, the main method of PDM, Principal Components Analysis (PCA), is applied. PCA provides a new orthogonal coordinate system based on the variance of the data cloud, fitting the main axis to the highest variance obtained, the second axis to the second highest variance and so on.

Firstly, the mean shape of the training set is calculated:

\[
\bar{x} = \frac{1}{s} \sum_{i=1}^{s} x_i
\]
Then the variations of the data set on the mean shape are found by calculating the covariance matrix:

\[ S = \frac{1}{s-1} \sum_{i=1}^{s} (x_i - \bar{x})(x_i - \bar{x})^T \]  

(2.3)

The covariance matrix presents a high dimensionality. PCA reduces the dimension of the data by finding the eigenvectors \( p_i \) and the corresponding eigenvalues \( \lambda_i \) of \( S \) such that:

\[ Sp_i = \lambda_i p_i \quad (\lambda_i \geq \lambda_{i+1}) \]  

(2.4)

In this new orthogonal space each shape can be represented as:

\[ x = \bar{x} + Pb \]  

(2.5)

where \( P=(p_1 \mid p_2 \mid \ldots \mid p_{2n}) \) is the matrix that contains the eigenvectors of the covariance matrix and \( b \) is the vector of parameters of a deformable model or, in other point of view, the points in the rotated coordinate system defined by \( P \).

The first parameters of \( b \) correspond to the largest eigenvalues, that is why usually the first modes –first \( b \) parameters– represent the most of the total variance. Figure 2.5 shows the generated shapes on the mean shape for the first 5 modes by varying each element of the model vector in a certain range. Indeed first images –first modes– present a wider range in its generated shapes than the lower modes.

In order to simplify the handling of the equations it is possible to approximate the shape \( x \) as:

\[ x \approx \bar{x} + Pb \]  

(2.6)

where now \( P=(p_1 \mid p_2 \mid \ldots \mid p_t) \) consists of the eigenvectors corresponding to the \( t \) largest eigenvalues. In turn, \( b \) is a \( t \)-dimensional vector such that:
Figure 2.5: Every image represents the variations –gray traces– in the range of $\pm 3\sqrt{\lambda_i}$ on the mean shape –black trace– for each one of the first five modes.

$$
b = P^T(x - \bar{x})
$$

(2.7)

The number of modes $t$ can be chosen so as to take a representative proportion of the training set variance. The total variance in the training set can be defined as:

$$
V_T = \sum \lambda_i
$$

(2.8)

Then a constraint is imposed:

$$
\sum_{i=1}^{t} \lambda_i \geq f_v V_T
$$

(2.9)

If, for instance, $f_v = 0.98$, it means that a value of $t$ that represents the 98% of the total training set variance is selected. Although this is a common way to find a suitable number of modes, it is not the only one, as is explained briefly in [13, pag. 17].

Normalizing $b$ can scale the different modes and facilitate its handling:

$$
\hat{b}_i = \frac{b_i}{\sqrt{\lambda_i}}
$$

(2.10)
Variations in the parameters of the vector $b$ in the Eq. 2.5 provide new shapes that were not previously in the training set. Assuming that $b_i$ are independent and gaussian, each of their variances are given by $\lambda_i$ and variations within $\pm 3\sqrt{\lambda_i}$ allow to describe 99% of all shapes in the training set. Hence, by applying these limits to parameter $b_i$, it is guaranteed that the model can generate new shapes similar to those in the original training set.

**Image search** The main aim is matching the shapes the obtained model can generate with the new images (not included initially in the training set). In the image search the gray-level appearance model is commonly used [9]. It describes that the local texture feature around each landmark is the normalized derivative of the profiles sampled perpendicular to the landmark contour and centered at the landmark (Fig. 2.6). This gray-level information is used to estimate the best position of the landmarks in the surroundings of an initial shape during the searching process.

On either side of the contour at which the landmark is located, $k$ pixels are sampled using a fixed step size of 1 pixel. It is possible to put this profile of length $2k + 1$ in a vector $g_i$. Then $g_i$ is normalized dividing it by the sum of absolute element values.

This procedure is applied to each image of the training set, obtaining a set of normalized samples $g_i$ for a given model point. Their mean $\bar{g}$ and

---

Figure 2.6: Search along normal profiles to the model boundary.
covariance $S_g$ are calculated assuming the samples are distributed as a multivariate gaussian. During the search $m$ pixels profiles are sampled around each point. Then Mahalanobis distance is taken as:

$$f(g_s) = (g_s - \bar{g})^T S_g^{-1} (g_s - \bar{g})$$

(2.11)

where $g_s$ is the representation of a new sample and $f(g_s)$ provides the quality of fit. The quality of the fit is tested at each of the $2(m-k)+1$ possible positions along the sample and the one which gives the lowest value of $f(g_s)$ is chosen.

This is repeated for every model point, finding a new position that satisfies the lowest Mahalanobis distance. Then an iterative process is performed in order to match the model shape to the new points given from image search. Each iteration is based on the update of the parameter’s positions, scale, rotation and modes of vector $b$. Subsequently, constraints are applying to $b$ ensuring a plausible shape. Finally, the convergence between the points of the image search and the new shape generated by the model is tested. If it doesn’t converge, another updating of parameters is performed and so on until the model generates a plausible shape that better matches the points.

Usually, a multi-resolution framework ([10, pag. 14]) is used to enhance the efficiency, speed and robustness of the algorithm. This entails an image search of the objects as the one explained before, but for different levels of resolution. Firstly, the search image is applied in a coarse image and then the initial location is refined in a series of finer resolution images. The levels of resolution correspond to the original image (level 0) and different levels where the original image has been smoothed and subsampled to obtain an image with half the number of pixels in each dimension.

### 2.3.2 ASM characteristics

**Brightness** Brightness changes can interfere significantly in the image search process. The effects of brightness due to motion and the change of background surface in the image can involve the largest difficulties. However, this is precisely what the chosen approach, segmentation of specific areas inside the face, is intended to avoid.
Occlusion  Occlusion is one of the biggest obstacles that shape models as ASM can face. The state-of-the-art deals with this problem directly or indirectly. This is the case of Akyol and Zieren [2], who presented a paper where ASM robustness against occlusion is tested to track the head in a hands occlusion situation for sign language speakers. Van Ginneken et al. [38], in the already mentioned paper about the application of ASM in chest radiographs, model also the occlusion of collarbones introducing its shape in the training set images due to the fact that its presence is constant and invariable.

In this work, the occlusion appears in the areas of the interest. The forehead can be occluded by some kinds of fringes depending on the haircut. Glasses and some kind of clothes in the neck region can be a problem as well. To face such situations, some mechanisms are introduced in the implementation of the facial segmentation. Also the effect of a subject wearing glasses on the model matching quality is studied.

Multi-Perspective  ASM has been applied successfully in frontal face images. However, multiple perspective images can entail some troubles, since a linear model (PCA) is used and there is a big proportion of possible shapes that remain unreachable. In order to solve this, some works applied two conceptual different solutions: combination of several linear models for variations between -45 and 45 degrees [40] or the utilization of non-linear models.

Kernel PCA, a non-linear approach, is explained in detail in the next section. Although its implementation is not carried out in the present work, a review of its characteristics provides some clues about its advantages for the ROIs segmentation and it serves as a valid element to enrich the discussion about the methods implemented here.

As far as this thesis is concerned, the procedure consists of building a proper training set with frontal and non-frontal shapes and test the model in order to see how different shape perspectives can be recognized. Some papers describe possible limits for the basic ASM. For example, in [27], it is stated that the model can work in variations of ±20 degrees displacement from the frontal view (turning and nodding).
2.3.3 Kernel Principal Components Analysis in ASM

Kernel Principal Components Analysis (kPCA) has been suggested for various application as, for example, multiple processes monitoring [19, 20, 8]. kPCA also has been applied to Active Shape Models in order to solve their pose matching limitations [27, 36, 41]. Regarding the literature about ASM based on non-linear PCA, several proposals are studied to evaluate its validity as a solution for facial segmentation.

**kPCA**

A review about the mathematical procedure of kPCA is given in this point. The main objective of this explanation is to provide enough details about the method to discuss afterwards its advantages and disadvantages in the context of the application that the current work is dealing with. For an “in-depth” description consult the cited bibliography.

In the case of using a gaussian kernel, which is the most applied in literature, a kernel matrix $K$ is built from the data set $\{x_i : i = 1 \text{ to } M\}$ as it follows:

$$K_{ij} = \Phi(x_i) \cdot \Phi(x_j) = k(x_i, x_j) = \exp\left(-\frac{1}{2\sigma^2}\|x_i - x_j\|^2\right) \quad (2.12)$$

where the training set shapes are computed in pairs with the dot product and $\sigma$ is the width of the kernel. It is recommended in [36] to normalize the kernel matrix:

$$\tilde{K} = K - \frac{1}{M}K - \frac{1}{M}1M + \frac{1}{M}1M$$

where $1_M$ is the $M \times M$ matrix with all elements equal to $\frac{1}{N}$. PCA is now applied to the normalized kernel matrix:

$$M\lambda\alpha = K\alpha \quad (2.14)$$
where $\mathbf{\alpha} = [\alpha_1, \ldots, \alpha_M]^T$. This yields eigenvectors $\mathbf{\alpha}^1, \ldots, \mathbf{\alpha}^M$ with eigenvalues $\lambda^1 \geq \lambda^2 \geq \ldots \geq \lambda^M$. As in linear PCA, it is possible to reduce the dimensionality of the model parameters by taking the number of modes that represent a significant percentage of the total variance of the system (exactly the same procedure as applied in [2.3]).

At this juncture, the principal components $\mathbf{b}$ of a shape vector $\mathbf{x}$ are extracted by projecting them onto the chosen eigenvectors $\mathbf{\alpha}^i$ using the dataset:

$$b_i = \sum_{n=1}^{M} \alpha_{ni}^i k(\mathbf{x}, \mathbf{x}_n)$$

At this point, it is possible to generate new shapes by varying the elements of $\mathbf{b}$, the model vector. However, in ASM this is not enough. As in the case of linear ASM, the generation of shapes has to be at the service of the image matching.

**Image search using kPCA**

The gray value search is implemented the same way as in the linear ASM. Nevertheless, for the non-linear implementation the mean shape is not taken as the initial shape. As [26, 27] note, in kPCA the average shape doesn’t represent the center of the model space anymore.

This is the reason why [27] suggest to start the iterative process with a frontal view instead of the mean shape. In every iteration the constraints in the model parameters have to be applied for the new shape that the gray value model finds. After this, the new parameters have to again become a shape in the original space of the training set. However, this reconstruction to the original space is more complex than it was for the linear ASM.

As it has been shown before, to recover the shape from the vector parameters in linear ASM:

$$\mathbf{x} = \bar{\mathbf{x}} + \mathbf{Pb}$$
In the case of kPCA it is necessary to minimize:

\[ \| \Phi(\hat{x}) - P\Phi(x) \|^2 \]  

(2.17)

where \( \hat{x} \) is the approximation of the reconstruction and \( P \) is a truncation operator. [41] proposes an expression to solve the previous equation for a gaussian kernel.

In the case of [36] the reconstruction of the shape is not applied. The authors disagree with [27] and maintain that is not generally valid to apply the constraints in the vector parameters the same way than in linear PCA. Now the displacements of variance are not on the mean shape. In other words, the mean shape doesn’t represent the null vector model anymore. On the contrary, in kPCA when all the model parameters are zero, they correspond to a shape that is far away from the data of the training set.

For this reason, in [36] a “proximity to data” criteria is proposed in order to search which are the model parameters that are close to the shapes of the training set. The approach is then based on a distance:

\[ f(x) = \sum_{i,j,\alpha} b_i^x b_j^x k(x, x_i) k(x, x_j) \]  

(2.18)

This distance presents a local maximum when the given shape vector \( x \) it is close to the training set data in the kernel space. For the “proximity to data” approach the reconstruction of the vector shape is not necessary in the original space. In every iteration the new shape is taken and used to compute the distance. Then a lower bound is established. If the new data is not close enough to the training set, the shape is brought closer to the local maximum by applying gradient ascent on the equation [2.18]
Chapter 3

Active Shape Model

3.1 Implementation

3.1.1 Materials

From the set-up shown in the figure 3.1, 1024x1280 resolution videos of 20 different subjects are taken. The specifications of the type of camera used can be found at the end of this document. The camera framing is selected in order to center the subject’s head –both face and neck– in the image. Then 600x900 pixels resolution pictures from the videos are extracted and finally different images of 18 subjects are selected to build the training set. Every subject executed the same movements and expressions that previously were written down as a protocol (see appendix A).

On one hand, the utilization of a protocol can limit the possibilities to train the model, since some poses are not described by any of the subjects –lack of randomness. The most variety of subjects and facial features is desirable in order to built a model as robust as possible.

On the other hand, the protocol offers the possibility to obtain similar poses of all the subjects, which can be useful to test the model in a certain way. The systematic movements included in the protocol, like the head turning from -90 to 90 degrees, allowed a precise study of the perspective limits of the model. For example, a training set is built with the same three
Figure 3.1: Set-up with camera recording the face of lay down subject.

images of each subject: frontal view and more or less the same right and left profile. Then, from this specific training set, the limit of the model to reach facial horizontal displacement can be evaluated precisely.

Also, it is important to notice that the subject is stretched out, which can facilitate the control of the movements to ensure a good execution of the protocol. Nevertheless, this circumstance can have a significant impact on the way the training set is built. For instance, if the subject is lying down, likely the default position of the neck can be different than in a situation where the subject is standing.

### 3.1.2 Training set

As seen in section 2.3, a training set of images showing the object to be modeled is necessary. Every image of the training set presents the same object but in different moments, poses and/or perspectives. Then a series of points, called landmarks, are used to describe the shape of the object on the image. The choice of these images determines a specific way to train the model. For instance, if the training set images include just frontal view faces,
it is unlikely that the model can generate any kind of profile position. Likewise, as it will be explained further below, training the model with a certain pose doesn’t necessarily mean that afterwards the model will be able to generate a similar shape. This limitation is due to the statistical nature of the model which is built from variance related displacements around the mean shape.

Accordingly, the best way to built the training set should consist of a "uniform" distribution of the different poses and expressions that are desired to reach in the generation of new shapes. Precise uniformity in the choice is important, due to the variance-based displacements from the mean shape that the model computes. That is the reason why a training set with 90% of frontal views and only 10% of profiles won’t be capable to reach any profile pose, practically the same that in the mentioned case without any profile image.

These considerations, and others that will be shown in the next sections, must be considered to obtain a suited Active Shape Model.

Training set images

Images that only show part of the object are discarded from the training set, either because only a part of the object is inside the framing or because they present some kind of occlusion. ASM bases its matching algorithm on a gray value structure search. Therefore, if it is not possible to have access to this gray value structures, the search becomes completely in vain. The occlusion can be due to:

Other objects interferences Which could be hands, long fringes or a high shirt neck covering part of the face and/or neck. In case of a subject wearing glasses, as far as the eyes are visible through the lens and the nose is not occluded by the frame, a correct shape matching is possible. This particular case is tested in the appendix B.

Self-occlusion Since the conditions of the acquisition of the images are quite controlled, self-occlusion is the most common occlusion that has to be
faced in the training set construction. It consists of the object covering some of its parts during the motion. An example is shown in the figure 3.2, where motion of the face results in an occlusion of the eyebrows corner. Another usual example of the facial self-occlusion is the jaw covering part of the neck’s contour during the turning motion of the head.

Also, it can be convenient to use grayscale images from the start, since they are taken to perform the gray-value appearance model (see 2.3). This way, it is not necessary a posteriori to implement this color conversion in the algorithm.

In the current work, all the test images that are used present the face centered in the frame. Since the gray-value search starts with a reference shape, the meanshape, its alignment is crucial to obtain satisfactory matching results. At the end of this thesis, where future works are commented, some considerations are given about the convenience of implementing an optimized and automatic alignment of the reference shape at the beginning of every image search process.
Landmarking

First of all, a complete landmarking distribution is taken (figure 3.3). In this version, all the landmarks that perfectly define the shape of the face are regarded. This first configuration provides the possibility of a suitable study of the desired model. Thus, it is possible to choose exactly the landmarks that provide a suitable model performance. In this aspect, three points are considered:

Matching Quality  An excessive reduction of the landmarks can lead to a bad precision of the model when it comes to describe the shape of the face. In contrast an excessive concentration can limit the variance of each landmark and the model becomes too rigid to reach a certain range of shapes.

Time of execution  The more landmarks are used, the more time it takes to match a new shape. This is also related with other parameters of the model, such as the size of the gray value profile, which is discussed in the next section.

Application  The main objective in this work is the facial segmentation of several anatomical areas. Hence, some parts of the facial features are not useful to this purpose. Some of these parts can be discarded (i.e. upper profile of the eyes and whole the contour of the mouth), others can be just simplified (i.e. face contour).

However, other features, as the upper part of the eyebrows or the lower part of the eyes are not simplified. The reason behind this design decision is that an oversimplification can easily lead to a malfunction of the gray value search. In these cases, the gray value structures are complex and not well defined most of the time. By using more landmarks the algorithm manage more gray value profiles for an specific shape and the search becomes more robust (i.e. upper eyebrows contour).

Taking into account these considerations, a reduced configuration is performed (see figure 3.3). This distribution reduces the number of landmarks in approximately 56% without representing a significant loss of quality. From
Figure 3.3: Left: full configuration with 120 landmarks defining face and neck shape. Right: the speed-up configuration with 67 landmarks designed by applying the optimization of quality, time of execution and application criteria.

hence on, unless otherwise stated, this configuration is the one used to built and test the model.

3.1.3 Training set evaluation

In [10], a training set test is proposed. The method is based on taking the images of the training set as unseen data. To this end, every time one of the images is discarded and the model is built with the rest of the data (i.e. leave one out cross validation). Thereafter, this new model is used to match the previously discarded image and the quality of the matching is evaluated. The process is done for every image in the training set.

The idea behind the procedure is to test the training set’s capability to train the model in order to handle different poses and different facial shapes–different subjects. In [10] a matching quality test is proposed to measure the distance between the points of the matched shape and the points of the reference shape–training set shape.
In the current work a simplified approach is used to evaluate the quality of the matching. It consists of calculating the intersection between the reference shape and the matched shape. Such an approach is taken having in mind that the final objective here is a proper segmentation, therefore an area based quality criterion is more appropriate. This method is explained below.

### 3.1.4 Matching quality test

Considering two shapes, one matched by the model, $V$, and the other reference shape, $R$, and the background of the image, $H$, the quality of the matching is defined by:

$$ r = \frac{a(V \cap R) - a(V \cap H)}{a(V \cup R)} $$

(3.1)

where the operator area $a(.)$ measures the intersections and unions between the both shapes and the background.

Applying this formula straight to the facial shapes results in evaluation of the quality matching for the contour of the head, but not of the facial features. Since usually the contour presents better matching than the facial features, this way to test can not be considered generally valid.

It is possible to solve this taking the region of the eyes and nose, both facial features used for the segmentation of ROIs. Repeating the test only for these regions it is possible to obtain a precise quality test of the facial features (see figure 3.4).

Then it is possible to merge both tests in just one quality parameter as it follows:

$$ r = \alpha r_1 + (1 - \alpha) r_2 $$

(3.2)

where $r_1$ is the matching quality for the whole face, $r_2$ the matching quality for the region of eyes and nose and $\alpha$ the weight parameter.
For the current work $\alpha$ equals 0.3 since, as already mentioned, the matching of the facial features presents more difficulties than the matching of the contour of the head. For these values, a quality parameter $r$ lower than 0.75 turns out in an poor matching, when $r$ is greater than 0.8 the matching can be considered suitable (see figure 3.5). As it is shown in the chapter dedicated to ROI segmentation, poor matching doesn’t mean necessary a deficient segmentation.

3.1.5 Chosen training set

A training set is built using images of 18 different subjects (two subjects are discarded because their images do not accomplish the previously explained training set requirements). Each subject provides three images corresponding to approximately the same three poses: frontal view of the face, right profile and left profile. Both profiles are picked at the moment right before the corner of the eyebrows is occluded by the contour of the face (as the central image of the figure 3.2).

Hence, the model is trained uniformly with poses of frontal view and the same profile of both sides. This way the model is not restricted a priori as far as turning movements is concerned and the innate limitations of the model can be properly studied. Nevertheless, note that it is indeed restricted when
Figure 3.5: Examples of matching quality. The corresponding quality rates are, from the left: 0.85, 0.76, 0.67

it comes to match different facial expressions since the training set’s frontal views do not include such kind of variations.

3.1.6 Code: characteristics and parameters

Several ASM parameters are evaluated in order to optimize the model and the image search process algorithms. To this end, a compromise between both quality and time of matching is applied for every one of this parameters. This ensures not only a good quality of matching, but also a quick and effective performance of the algorithm.

The parameters under study are related with the gray-value appearance model and the iteration search process. First the gray value profile size is studied. This establishes the dimensions of the gray value area that is going to be modeled. Afterwards, in the image search, the size of the search profile, with a default value of 14 pixels, has to be optimized. Finally, related with the iteration process, two parameters are consider: maximum number of iterations per level of the multiresolution ASM and a break condition with default values of 15 iterations and 0.8 pixels respectively.
Gray value profile size

The image search procedure in ASM is based on a gray-value appearance model. This method models a gray-value profile around every landmark of every training set image. Larger are these profiles, more information about the gray-value is available in the subsequent profile search in the test image.

Therefore, the size of the given profiles has a direct impact on the matching quality. To simplify the optimization, square profiles are considered. This way just the length of the side is under study.

Multiple ASMs are built by setting different gray value profiles in its gray-value appearance model. Thereafter, the same test images are matched by the models obtaining the relation between the profile side size and both quality and time of matching.

Search area size

Already in the method of the image search, the size of the search area is tested. Once the meanshape is centered in the test image, the gray-value search starts around every landmark. The size of the profile that takes to perform this search has an important influence on the matching quality.

In this case, only one model is built using the optimized gray-value profile size found in the previous point. Every test image is matched for different sizes of search area that go from 2 up to 24 pixels by steps of 2 pixels.

Number of iteration and Multiresolution ASM levels

Multiresolution Active Shape Model involves, as already explained, a leveled access to the gray value structures of the images in both gray value model and in the subsequent image search. This means that there is an iterative process for each one of the levels –being the level 0 the original resolution image–, starting always with the coarse image resolution. For this implementation, 3
levels, including the original resolution, are considered (see algorithm 1).

\[
\text{OldShape}=\text{meanshape}; \\
\text{NumLevels}=3; \\
\text{for level} \leftarrow 1 \text{ to } \text{NumLevels} \text{ do} \\
\quad \text{for } \text{it} \leftarrow 1 \text{ to } \text{MaxNumIterations} \text{ do} \\
\quad\quad \text{NewShape} \leftarrow \text{SearchGrayValue}(\text{OldShape}); \\
\quad\quad \text{aNewShape} \leftarrow \text{Align}(\text{NewShape}, \text{MeanShape}); \\
\quad\quad \text{modelShape} \leftarrow \text{UpdateASMParam}(\text{aNewShape}); \\
\quad\quad \text{diff} \leftarrow \text{compare}(\text{NewShape}, \text{OldShape}); \\
\quad\quad \text{OldShape} \leftarrow \text{modelShape}; \\
\quad\quad \text{if } \text{diff} < \text{breakcondition} \text{ then} \\
\quad\quad\quad \text{break; } \\
\quad\quad \text{end} \\
\quad \text{end} \\
\text{end}
\]

Algorithm 1: Image search iterative process in ASM. Functions \text{Align}() and \text{UpdateASMParam}() are given in algorithms 2 and 3 respectively.

\[
\text{Result: } \text{x}_{\text{aligned}} \leftarrow \text{Align}(x_1, x_2) \\
\text{initialization; } \\
\text{t} \leftarrow (t_x, t_y) \leftarrow \text{GravityCenter}(x_1); \\
\text{x}_1 \leftarrow \text{GravityCenterOrigin}(x_1); \\
a \leftarrow (x_1x_2)/|x_1|^2; \\
b \leftarrow (\text{sum}(x_1y_2i - x_1y_2i))/|x_1|^2; \\
\text{scale} \leftarrow \sqrt{a^2 + b^2}; \\
\text{angle} \leftarrow \text{arctan}(b/a); \\
\text{R} \leftarrow \text{RotationalMatrix}(\text{angle}); \\
\text{x}_{\text{aligned}} \leftarrow \text{scale} \times \text{R} \times x_1 + t
\]

Algorithm 2: Shape alignment in ASM image search using Procrustes analysis

\[
\text{Result: } \text{x}_{\text{updated}} \leftarrow \text{UpdateASMParam}(x) \\
\text{initialization; } \\
b \leftarrow P^T(x - \bar{x}); \\
\text{for } i \leftarrow 1 \text{ to } |b| \text{ do} \\
\quad \text{if } b[i] > 3\sqrt{\lambda_i} \text{ then} \\
\quad\quad b[i] = 3\sqrt{\lambda_i}; \\
\quad \text{end} \\
\quad \text{if } b[i] < -3\sqrt{\lambda_i} \text{ then} \\
\quad\quad b[i] = -3\sqrt{\lambda_i}; \\
\quad \text{end} \\
\text{x}_{\text{updated}} \leftarrow \bar{x} + Pb;
\]

Algorithm 3: Parameters updating in ASM
Break condition

The iterative processes have a break condition which decides if the convergence has been reached. In every iteration the new found shape and the previous shape are compared. If the difference between these two shapes is small, it is assumed that a definitive matching has been found. At this moment the iterative process is stopped (see algorithm 1).

This difference is expressed by the next equation, which presents the average point displacement between the old and the current shape for a given iteration:

\[ \text{diff} = \text{mean}(|x_{\text{old}} - x_{\text{current}}| + |y_{\text{old}} - y_{\text{current}}|) \]  (3.3)

where \( x = (x_1, ..., x_n) \) and \( y = (y_1, ..., y_n) \), being \( (x_i, y_i) \) the coordinates of the \( ith \) landmark.

3.1.7 ASM perspective restrictions

As already mentioned, standard Active Shape Model presents some pose perspective limitations. The formation of the training set and, above all, the linearity of the PCA used in ASM have direct impact on this restrictions. However, although to a lesser extent, the kind of the landmark distribution can also have an impact on this matter.

The chosen training set is built with right and left profile images in the time before eyebrow occlusion. The range of head turning in this case is wider than the limitations of the model. This means that the standard ASM never faces this kind of occlusions because it can never reach such poses.

The origin of this limitation in perspective turning and nodding, but also extreme facial expressions, lies in the statistical nature of the PCA applied to build the model. With three images per subject, two of them both profiles and the other of frontal view, the mean shape of the whole training set is always centered around the positions of the frontal face.

Since for the linear PCA the displacements are always from the mean
shape, depending on the variance of the system, there is always a limit in the locations that every landmark can reach. Besides, these limits are established by means of constraints, since a completely freedom of displacement from the mean shape wouldn’t result in profile face matching, but in a full set of non-plausible facial shapes. This is the reason why the existence of a certain shape in the training set doesn’t guarantee that the model is going to be able to match the same or similar shape in the unseen data.

Taking into account the previous considerations, the model is tested in order to match shapes for a different turning poses. The turning limitations are measured taking a point located in a corner of one of the eyes from a frontal view. Then different images of all the way from the central to the profile pose of the head are used as a test data. In the middle of the turning movement the quality of the matching becomes unsatisfactory and, at a certain point, the model is not capable to match the shape. The corresponding image is taken and the point of the eye corner is located again to measure the horizontal displacement (along the abscissa axis). Such displacement is considered as the turning perspective limit of the model.

This evaluation is repeated it for three different landmarks distribution: the already seen complete and speed-up landmark configurations, and a new distribution defined just with the contour of the face and neck, without including the facial features (see figure 3.15).

In presence of these restrictions, the standard ASM can be limited when it comes to accomplish the rPPG performance requirements. Moreover, if it is considered a probable situations where the camera is not recording the frontal view of the subject’s face (i.e. remote detection of heart rate in a hospital bed makes the profile poses rather likely.)

To this end, a modification of the current Active Shape Model implementation is developed in order to built a system capable to track faces for different poses (ideally for all the range from 90 degrees to -90 degrees turning movements).

3.1.8 Combination of Actives Shape Models

The method proposed consists of using different Active Shape models to cover all poses. For this purpose, several ASMs are built, each one with its own
training set and model parameters. Then the algorithm is capable to apply one of the models to its disposal by knowing which pose is shown in the test image. This decision is made based on the position of the eyes.

Models  In a first implementation two models are defined. The first model’s training set consists of several frontal views (approximately between -20 and +20 degrees), including different facial expressions that can result in significant facial deformations (i.e. the elongation of the jaw when the subject pronounces a noticeable sound o).

The training set of the second model is built with images of a face profile in a rough range between +20 and +60 degrees. Its shapes are built taking into account that it is not necessary to describe the contour that is going to be occluded.

Detection Criteria  The algorithm knows when to apply one or the other model by means of locating the position of the eyes. The method has to its disposal a gray value template corresponding to the area around the eye. Then the unseen data is tracked in order to match the template with the gray values in the image.
The test image is divided into regions, each region linked to a specific range of poses. When both eyes are located in the central region the first model is applied. When one of the eyes enters the right region the second model is used (see Fig. 4).

For an implementation capable to manage the turning range until 90 degrees, the extreme profile poses can be detected by the disappearance of one of the eyes. For turning poses to the other side, or nodding movements, the procedure should be the same. This establishes that at least 5 models should be available to track all the turning poses, besides of those needed for the nodding movements (up and down).

\begin{algorithm}
Model1 $\leftarrow$ FrontalFaces;
Model2 $\leftarrow$ RightprofileFaces;
EyesPosition $\leftarrow$ EyesDetection(TestImage);
if $(EyesPosition \cap ProfileRegion) == 0$ then 
\hspace{1cm} ImageSearch(ParamatersModel1);
else 
\hspace{1cm} ImageSearch(ParamatersModel2);
end
\end{algorithm}

Algorithm 4: Test image matching with combination of two ASMs corresponding to frontal and right profile. Function EyesDetection() is given in algorithm 5. ImageSearch pseudocode is shown in 1.

\begin{algorithm}
MedianImage $\leftarrow$ MedianFilter(Image);
MatchedRegions $\leftarrow$ GrayValueMatching(Image, EyesTemplate);
EyesPosition $\leftarrow$ Area&PositionRestrictions(MatchedRegions);
\end{algorithm}

Algorithm 5: Eyes detection
3.2 Results

3.2.1 Training set test results

The results of the training set test (leave one out cross validation) for frontal faces can be found in the figure 3.7. The best and the worst frontal view matching obtained are shown in the figure 3.8.

Separately, the result of the training set test for both sets of profile faces can be consulted in the figure 3.9.

3.2.2 Model parameters optimization

Gray value profile size

Multiple models for different square side lengths are built and used to match a test image. The values of the lengths are selected from 3 to 49 pixels by steps of 2 pixels. During the matching process of the image two parameters are measured: matching quality and time of execution. The charts in the
Figure 3.8: Left: best matching in the training set test. Right: worst matching in the training set test.

Figure 3.9: Matching quality for different subjects: left profiles (left) and right profiles (right).
Figure 3.10: Left: relation between the length of the profile side and matching quality. Right: relation between the length of the profile side and time of execution.

Search area size

The search image process is performed for different search profile sizes: from 2 pixels side length up to 24 by steps of 2 pixels. Two charts represent the relation of the search profile size with the matching quality and with the matching time respectively (figure 3.11).

Number of iteration and Multiresolution ASM levels

Once again, the matching quality and time of execution are tested for several maximum numbers of iterations per iterative process (figure 3.12).

In the same figure it is possible to find the results for the same test, but leaving out the resolution level 0 corresponding to the original image resolution.
Figure 3.11: Left: relation between the length of the search profile and matching quality. Right: relation between the length of search profile and time of execution.

Figure 3.12: Left: relation between the maximum number of iterations and matching quality. Right: relation between the maximum number of iterations and time of execution. In green the response discarding the highest resolution level.
Figure 3.13: Left: relation between the break condition values and matching quality. Right: relation between the break condition values and time of execution.

Figure 3.14: ASM is limited to reach profile poses away from the frontal views.

**Break condition**

Several break condition values from 0 to 1.6 pixels by steps of 0.1 are tested obtaining the results in terms of both quality and time matching in figure 3.13.

### 3.2.3 ASM perspective limitations test

Using the optimized parameters, the model is tested now for a complete sequence of head turning poses (see figure 3.14).

This limitations are quantified in the table 3.1. These numbers correspond to the displacement in pixels of the position of the left eye’s corner from the frontal view (central image in 3.14) and the image where the model is not
Table 3.1: Model turning poses approximated restrictions for the different facial landmarks distributions given in pixels of displacement from the frontal view to the moment when the matching quality becomes unsatisfactory.

<table>
<thead>
<tr>
<th>Landmark distributions</th>
<th>Turning limitation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Complete Configuration</td>
<td>50 pixels</td>
</tr>
<tr>
<td>Speed-up Configuration</td>
<td>50 pixels</td>
</tr>
<tr>
<td>Contour Configuration</td>
<td>70 pixels</td>
</tr>
</tbody>
</table>

Figure 3.15: Example of combination of two different ASMs in a turning movement. In the first two frames the shape is matched by the frontal model. For the third frame, situation in which the frontal model would fail, the shape is matched by the lateral model.

capable to match the shape.

**Combination of Actives Shape Models Test**

The implementation described in the previous section of multiple ASMs combination is tested. Since it consists of two models capable to track frontal and right profile views, a sequence of images where the head is turning to the right are used as a test data. Figure 3.15 shows how both models match different facial poses shapes.
3.3 Discussion

3.3.1 Training set

The training set test reveals some considerations:

1. Different subjects recognition  Despite the low quality in the profile images, in the case of the frontal views the model is capable to offer a correct matching. The capacity of the model to track frontal faces establishes directly the capacity of the model to work in the future with unseen data (i.e. unknown faces). Specifically, as figures 3.7 shows, the model matches over the 80% of the frontal faces with a quality superior to 0.80, approximate threshold that ensures a good quality (see quality matching threshold in 3.1.4).

Two of the images that present a quality below the threshold were included as frontal views when in fact they present some nodding and torsion variations (see in figure 3.7 the best and the worst matching). The restrictions of the model to reach nodding poses are similar than those that affect turning poses for exactly the same reasons. However, nodding poses show less variance from the frontal ones and they should be reachable easily by the model with a proper alignment of the initial shape.

2. Limitations of perspective  As already mentioned, the ASM has strict perspective limitations. This is the reason why the model is not capable to match properly the shape when most of the profile faces of the training set are under study (3.9). The quality in these cases is low, although some features, most of the time the face and neck contour, are pretty well matched. The specific test of ASM perspective limitations can be consulted further below.

Regarding the charts, a few profiles present a satisfactory quality matching. This results correspond to the images of the training set that are not a strictly turning movement, but with a strong face rotation component. Rotation is usually well defined by means of shape alignment process in Active Shape Model.
3. Clues to a possible improvements The first point leads to a possibility to improve the training set with more diverse subject images in order to make it more robust in the future when confronted with unknown faces. Moreover, the previous consideration about the limits in the perspective mark the chance to built a specific training set, by enclosing the facial poses in its images (e.g. it can be useful for the multiple ASMs implementation).

3.3.2 Model parameters optimization

Gray value profile size

The comparatives size/matching quality and size/matching time for the evaluation of this parameter are shown in the figure 3.10.

On one hand, for very small profiles the gray-value appearance model has only little information available and the algorithm is not capable to find a good match. Increasing the length the matching quality improves quickly until it reaches some point where it gets stable. This point should be the better value in terms of a good matching quality.

On the other hand, the time of execution behaves erratically for small lengths. This is mainly due to how the search iterative process responds to having not enough data about the gray value structure that is looking for. If the matching is unsolved because the gray value structure is not found, the break condition is never accomplished since there is no convergence. Then the maximum number of iterations is executed which affects the matching time. Although, around 20 pixels and higher lengths the behavior is the expected: after reaching a minimum, the time of execution increases with the size.

Therefore, the optimum gray-value profile size should represent a compromise between both charts. This value, 23 pixels, involves a minimum execution time, less than 2.2 seconds to match the shape, and a maximum of matching quality.
Search area size

The optimum value, 12 pixels, is found following the same kind of compromise applied in the previous point (see 3.11). In this case, the matching time response is more predictable, since the size of the modeled profile around each landmark is already optimized.

Number of iteration and Multiresolution ASM levels

As it is shown in figure 3.12, more than 10 iterations ensure a good matching quality. Specifically, 12 iterations seem to be a suitable choice. In this case, the time is not increasing linearly, since it gets stable from a number of iterations equal to 17. This is due to, for a long iterative process, the break condition is always accomplished.

Also it is interesting to repeat the test excluding the last level in the image search. The last level corresponds to the original resolution image. Nevertheless the search in this high resolution is not necessary, since the previous levels with lower resolutions are enough to get a good quality decreasing significantly the time of execution.

Break condition

A break condition of 0.8 pixels seems to satisfy the compromise between matching quality and time of execution. Note that if the break condition is omitted, the matching time corresponds to the execution of the maximum number of iterations established in the previous subsection.

3.3.3 Limitations in perspective

From the results, it can be conclude that a simplified landmark structure doesn’t necessarily ensure a wider perspective range. For example, comparing the results for the full and the speed-up landmark configurations (figure 3.3), there is no an improvement in the limitation of the perspective.
To put these results into perspective, the occlusion of the eyebrow appears between 70-80 pixels of displacement. This means that the model for the speed-up landmark configuration reaches shapes around 60-70% of the distance between the frontal view and the occlusion of the eyebrow. However, these displacement values are valid only for a very well defined turning movement. The results for poses with not only turning movement, but also facial rotation can present other matching properties as seen in the training set test for the profile faces.

Although it is difficult to convert this pixels of displacement in degrees of head turning, the obtained values do not seem disagree with, for example, the results in [27], where it is stated that standard ASM can reach poses between -20 and 20 degrees.

Some landmark distribution simplifications do present some improvements in the perspective restriction. This is the case of the contour configuration, where without facial features inside the landmark structure the model seems to become more flexible to reach profile poses. Furthermore, taking just the contour, the gray value search becomes simpler and it is easier to obtain a correct image search.

Nevertheless, the segmentation of the ROIs is difficult to perform without having some knowledge about the location of facial features. For example, to properly segment the forehead the upper limits of the eyebrows are required, as well as the lateral limits of the nose to segment the cheeks. Especially considering a head in motion, since the face undergoes a different perspective deformations according to the pose. In consequence, contour configuration can provide a poor segmentation quality, despite allowing for considerable improvement in the turning and nodding displacements.

**Combination of Actives Shape Models**

The main advantage of this method it is the possibility to solve the ASM perspective problems (see figure 3.15). However, other comments can be extracted from the implementation characteristics and its evaluation results:

**Specific training set** After a study of the ASM perspective restrictions, like the one done in the previous point, it is possible to know the very defined
range of poses that one has to include in every training set of the different models (aspect already envisaged in 3.3.1). The possibility to build specific training sets provides robustness and high precision in the way every model matches its corresponding shapes.

**Matching time** Although here it is not a priority, it is interesting to note that this method does not involve a significant increase in the matching time, since the decision process is, in terms of matching time per image, not significant. Nevertheless, it notably increases the computational cost when the models are built.

**Drawbacks** The implementation is still in embryo. The eyes detection is difficult to implement in order to obtain a stable performance for different subjects, poses and brightness conditions. Several problems appear commonly (e.g. sometimes the gray values of the template are detected in parts of the image that are not corresponding to the eyes).

**Standard ASM comparative** Combining the different ASMs enhances the perspective limitations of standard ASM. Moreover, it allows to reach the whole range of turning poses by selecting an appropriate set of models. Another advantage is that this is possible without big changes over the standard implementation: the criterion of decision lies on an external technique to the model based entirely on image processing.

**Other perspective methods** It is possible to find another approach in literature that uses combinations of different Active Shape Models. In 40 three different models are built –central, and both profiles at -45 degrees and +45 degrees– separating also the contour face model from the facial features model. An energy function is built to link both models and then a genetic algorithm matches the suitable pose using a ”chromosome”. A ”chromosome” is a data package that encodes the sets of parameters in an specific order (7 shape parameters, 4 pose parameters, 3 shape parameters and in the bottom 4 more pose model parameters).

The whole range of poses is still not reached, but only until approximately between -45 and +45 degrees poses. Besides, since they are adding more
computational cost during the image search process, the matching time is also increasing considerably.

Also, the non-linear PCA application can be considered in ASM, kPCA, which theoretically allows the matching of poses along the full 180 degrees turning [27]. Nonetheless, the matching quality presents some drawbacks, mainly related with the capacity of the gray value appearance model to operate with kPCA. For different facial poses, the model has to deal a very wide range of gray value structure for the each landmark. For instance, a point in the eyebrow for a facial frontal view is surrounded by skin and hair, whereas for some turning poses can be located in the contour of the face, between the skin and the background of the image.

It is also important to keep in mind that the reconstruction of a generate shape is an approximation depending on a proper computation of a minimization problem (see [2.17]). Furthermore, kPCA entails an increase of the computational resources comparing with both standard and combination of ASM. This has a remarkable impact on the execution time for both studied approaches: “proximity to data”, based on gradient ascent in every iteration of the search process, or the direct reconstruction of the shape from the model parameters.
Chapter 4

Segmentation of ROI

4.1 Characteristics

This chapter tackles the main objective of this work: segmentation of several anatomical areas in a face in motion. It has been seen that in order to perform this ROIs segmentation, previously it is necessary to acquire a the location of some facial features together with the head contour.

This way, for every test image, a Active Shape Model is used to match the shape of the face. As ASM is a point-based shape model, after the matching process, the points that define the facial shape are well-known. Thereafter, it is possible to locate the ROIs by referring these facial points as follows:

**Forehead** The contour of the face provides the lateral limits of the forehead. The upper and lower margins are established by the position of the fringe a eyebrows respectively.

**Cheeks** The cheeks are defined between the contour of the face and the lateral points of the nose. Besides, the upper limit is established according the position of the eyes.
Figure 4.1: Segmentation performed over the ASM matched shape. In blue, the landmarks used to refer the position of the ROI points.

**Carotid Artery** Defined in relation of the both neck and jaw contours.

A fundamental point in the performance of the segmentation is the motion of the faces. This is not a problem as long as the Active Shape Model matches the face and the features in the test image properly. If the matching is not good enough, naturally the associate segmentation of the ROIs won’t be successful.

### 4.2 Implementation

The ROIs points are located by means of applying displacements in both dimensions from the already matched facial shape points. Not all the facial points are taken as a reference, just those that are significant in the shape of the facial features (see figure 4.1). This way, the segmentation performance is simplified without affecting its capacity to obtain suitable results for a
different poses and subjects.

Once the facial reference points are taken, it is time to decide what kind of displacements have to be applied. To this end, two different kind of ROIs segmentation are performed. The first one, is based on small displacements, keeping the contours of the ROIs close to the contours of the facial features and, generally, maximizing the surfaces of the ROIs. The second kind of segmentation uses larger displacements for all the points, resulting in smaller ROIs.

Besides the implementation based on displacements, some mechanisms are introduced:

**Fringe edge**  The algorithm that segments the forehead tracks the position of the fringe, which can be useful as occlusion handling strategy. The ASM built in the previous chapter with the called speed-up landmark configuration establishes with three landmarks the limits of the fringe.

By default the upper limit of the forehead segmentation is established by the position of the highest lateral landmarks on each side (case shown in figure 4.1). Another available mode of forehead segmentation establishes this upper limit directly with the position of the fringe. Also, it is implemented that if the fringe edge is lower than the lateral landmarks then it becomes the upper limit of the forehead segmentation.

**Discarding non useful segmentations**  After performing the ROIs segmentation, the algorithm computes the area of the segmented regions. This is mostly useful for the forehead and carotid artery segmentation. Either because fringe covers a big part of the forehead or because there is not enough visible neck, the segmentations can be useless to extract measurements from. By establishing a minimum ROI area, the algorithm discards these cases.

### 4.3 Results

Figure 4.2 shows different examples of ROIs segmentation by applying minimum displacements (e.g. taking as much as possible area for each ROI).
Figure 4.2: Examples of the ROI segmentation. For the third and the fourth image, an area restriction by discarding the ROIs that are no big enough due to either occlusion or a certain facial shape matching is applied.

Figure 4.2, on the other hand, presents ROIs segmentation for bigger displacements and, therefore, smaller areas. Finally, in figure 4.4 occlusion handling for the forehead area is shown.

4.4 Discussion

A first implementation of the segmentation is based on taking small displacements from the facial features (see figure 4.2). This design criteria provide large ROIs for the segmentations of the cheeks and the forehead, which seems to be an appropriate first criterion to apply segmentation. However, such an implementation might have as a result a faulty segmentation if the ASM matching of the face is not accurate enough.

In order to enhance this performance, a second implementation is considered. It is possible to make the segmentation less depended by simplifying or reducing the ROIs. For example, if the ROIs are reduced taking larger displacements from the facial matched shape, the segmentation performance becomes more robust against a incorrect facial matchings. This is an interesting property when it comes to design how the ROIs points are going to be located, since it is possible to obtain a proper segmentation even for those images that present a poor facial features matching (see figure 4.3).

Even so, the implementation based on larger displacements provides sometimes practical solutions for the segmentation of both cheeks, but not in the case of the forehead or neck. Also, large displacements make the performance
Figure 4.3: Images on top show a non accurate matching which leads in the faulty segmentation. In both images below, the area of the segmented cheeks is reduced, giving now a valid segmentation for the previous subject (left), as well as for one of the images that already have been segmented correctly with the other area criteria (right).

Figure 4.4: In this case, although the neck is uncover, its segmentation is not well performed (left). This ROI is finally discarded by a minimum area criteria (right).
more unstable for a general performance.

As a last aspect it is important to emphasize why the quality of the segmentation of the ROIs is not evaluated directly, as in the case of the current implementation of ASM in 3.1.3. Firstly, a correct ASM matching practically always leads to a correct segmentation of the ROIs. Therefore, the facial matching evaluation seems to be valid enough to test the subsequent segmentation. Secondly, the regions of interest are rather independent of each other, presenting different performance qualities, and a global quality test can be difficult to quantify and evaluate. Furthermore, considering the test of each ROI separately, no generally valid results have been found. Lastly, as a last resort, the validation of the segmentation results should be based on the ability to detect rPPG signal in the ROIs. Hence, not only the geometrical segmentation is important, but also other parameters as, for example, the temporal stability.

The results also point that the neck segmentation is specially problematic. Even in the cases where the neck is completely uncovered a proper segmentation is not insure. In fact, here the availability of the surface of the neck in the image is not relevant, since it is the structure of the landmark distribution what, for all practical purposes, produces the segmentation performance. The matching shape leaves a quite reduced area to segment between the neck and contour of the the jaw, making the quality of the segmentation very dependent on a correct facial matching (see figure 4.4).
Chapter 5

Conclusion

5.1 ROI segmentation by Active Shape Models

It has been seen that facial segmentation is required to obtain robustness in the heartbeat’s detection by rPPG. To this end, the present work has studied the application of Active Shape Models. This tool should be appropriate to, firstly, extract the shape of the head’s contour and the facial features. Secondly, to locate the anatomical areas of interest forehead, cheeks and part of the neck – inside the facial shape that the model has previously matched.

The system has to be specially prepared to handle faces in motion no matter its pose, expression or features. This makes the matching quality test crucial for the evaluation of Active Shape Model as a suitable solution for the rPPG application. A low matching quality results in an impossible location of the areas of interest, since their points are directly related with the position of the matched facial features.

Undoubtedly, ASM presents a precise way to perform the segmentation of the regions of interest. Whenever the structure of the facial landmarks distribution is complete enough, the limits of the ROIs can be perfectly defined in the face. This factor maximizes an accuracy of the segmentation difficult to obtain with other approaches. Even more so, if these approaches are not based on modeling shape points (i.e. skin color based methods).
The forehead area location is a representative example of this precision in the segmentation. Its lower limit can be established following the contour of the eyebrows, taking also the space in the middle (point with high blood perfusion). Moreover, the upper limit depends of the position of the fringe which is an important advantage to avoid its possible occlusion. These considerations are more important in case of faces in motion, where tracing the areas of interest with that precision along the different poses and facial deformations can become problematic.

As already mentioned, the facial shape matching quality has to be studied in order to find a suitable performance of the model. Considering this kind of evaluation, optimized model parameters have been found during the current research.

The optimization of the model has provided correct and highly reliable matching quality for frontal faces in terms of how the system deals with unknown faces. This point is very important, since at the end the model has to match face shapes in subjects that are not included in its training set. Such capacity has been proven not only by means of testing the training set, discarding at the time one of the images, but also in other tests, as the one shown in the appendix B, where the effect of the glasses in the matching working is studied with a subject completely unknown for the system.

On the other hand, the optimization of some model parameters, such the gray-value profile size or the break condition for the searching process, has reduced the runtime significantly (more than 50% in regards to the implementation with default values). This reduction has been achieved ensuring at all times a satisfactory matching quality.

Although, it has been seen as well that the quality of the matching is circumscribed by a specific range of poses—turning and nodding movements—and facial expressions. The problem of matching facial expressions can be mainly solved improving the training set with more images. Nevertheless the limitations in the profile poses are inherent in the nature of the linear method used to build the model: Principal Components Analysis. This leads to a correct matching quality, and consequently a correct ROI segmentation, only for poses closed from the frontal view position.

A solution has been considered in order to obtain a better performance of the model for the profile views. The utilization of several ASMs, offers a significant enhancement as far as quality of match is concerned. Also it doesn’t
involve higher computational costs in the image search process. Future implementations of this method could lead to a facial matching for the whole range of head turning and nodding, besides to allow the construction of very specific and robust independent models. Although, the location of the eyes to decide which model has to be applied presents a difficult implementation for different subjects.

At the same time, kPCA, a non-linear implementation, has been studied, since it is known as a solution for ASM perspective handling limitations. In terms of matching quality the most effective solution for the whole range of facial poses, although it has been shown that its performance is still presenting some challenges and it is far away to be ideal. Moreover, kPCA comes with high computational costs and, for this reason, is only suitable for very specific applications.

5.2 Future works

The research of segmentation in rPPG can follow two different paths. On one hand, there is the option to keep improving the ASM to enhance its performance since, as it has been stated, it is difficult to find other methods capable to provide a similar precision on the segmentation. Regarding the ASM implementation and its evaluation in the current work, at least three possible improvements are surmised:

**Initial shape alignment**  It has been seen that a good matching quality depends greatly on the mean shape alignment at the beginning of the image search. This alignment can ensure the proximity of each landmark to the corresponding modeled gray-value structure and avoid possible failures of the search iterative process.

During the model test in this work, the mean shape has been centered in the face either by changing the alignment manually or by assuming the same center for every test image. Since ASM is going to face unseen data where the faces have not got the same suitable center, it would represent a significant improvement to find a way to optimize an automatic performance of the initial shape alignment. This way, besides to improve the time of execution by optimizing the iterative process, it is possible to obtain a better
matching quality, not only for diversity in the subjects faces, but also in the response of the model in front of nodding and turning poses.

**Gray-value handling** The gray-value appearance model used in ASM is sensitive to changes in brightness and, above all, when it comes to face a wide diversity of gray-value structures. This method models gray-values around every landmarks of every image in the training set.

It has been tested that if all the images are from the same subject in the same time of acquisition –minimum changes in brightness and low variance in gray-values– the image search takes practically always the minimum number of iterations, tracing a straight transition from the mean shape to the correct match.

Nevertheless, for the training set with three images per subject, the model has to deal with different subjects in different times of acquisition, which means rather wide range of gray-values for the surroundings of the same landmark position, besides to some changes in brightness (although this is quite controlled). This leads to a utilization of a larger number of iterations and even sometimes the search loses the position and tracks new shapes completely out of the regions that is looking for. Note that again the alignment of the initial shape is here indispensable, but not the unique cause of unexpected results. For some cases, a good matching is reached in the middle of the search. Problems occur when the algorithm doesn’t recognize that matching as a good matching and keeps going losing the track beyond the correct position completely.

Therefore it seems that making the gray-value search more robust can result in a significant improvement of the ASM performance. In order to implement this, some skin-color algorithms can be really useful. The images under study are in color, so the application at this moment of a previous processing should not represent a problem. Skin-color algorithms can pave the way for the subsequent gray-value search by discerning the skin pixels from any other kind of pixels. Thereafter the gray-value appearance model could work more guided, reducing the possibility to get lost and provide an invalid matching.
Perspective methods  Finally, to obtain a performance capable to work in the application for rPPG, the methods to solve the perspective limitations of ASM have to be enhanced. To this end, different approaches should be optimized, as the detection of the eyes in the proposed solution in the current document using multiple ASMs.

Also, it would be important to consider that the application is going to work for a sequence of images and the facial poses between consecutive images usually don’t present large variations. On the contrary, there is always a continuity and it is possible to take advantage of it by implementing a system with certain memory, matching every new image in accordance with the previous matched shapes.

Finally, the performance of the ROI segmentation for facial ASM matching might also be enhanced in future works. For example, it could be possible to establish more relation not only between the ROIs and the facial features, but also between the own segmented regions points. This could be useful to introduce, for instance, a ROI discarding based on non plausible segmented regions, besides to offer all kind of possibilities to make the segmentation performance more reliable and robust.

Some other options in the same line of point distribution shape based model, could be the already mentioned Bayesian Shape Models (BSM) or the variety of ASM, Active Appearance Model (AAM), introducing texture parameters. Anyhow, consulting the existing bibliography, there is no sound guarantees that any of this methods could improve the performance of ASM as far as time of execution or perspective limitations are concerned.

Another option could consists of a merging of several techniques. Skin-color algorithms, edge detection, etc. Nevertheless, it is practically sure that the choice of this kind of solutions would lead to a substantial loss of precision.
Appendices
A. Protocol

Table[1] shows the list of the subjects that have collaborated in the acquisition of the data, together with the permission of publishing photos.

There is only one source of illumination, completely centered with the position of the subject. The brightness conditions are controlled by means of reducing the shadows of the people in the room, dropping down the blinds and turning around the computer monitors.

The subject lies down, taking as a good position the one that puts his/her head in the center of the camera framing. Then a brief explanation of the required movements is given. In any case, one of the persons in charge is going to point and say out loud the positions during the recording time. During this period, the movements are executed following this order:

1. Slowly but constantly, the subject turns his/her head right and once he/she reaches the 90 profile, returns to the central position.

2. Same than (1), now turning left.

3. Nodding up until the neck is completely stretched and returning to the front view.

4. Nodding down until the subject can see his/her feed and returning to the front view.

5. In frontal view position, the subject pronounces the vowels "a, e, i, o, u", taking time to vocalize every letter.

6. The subject closes the eyes during a period of approximately 2 seconds.

7. Again with the eyes opened, the subject executes a big smile.

8. Finally, the subject yawns emphasizing the facial expression.

At this point, the acquisition of images from the camera is stopped and the subject is informed that his/her participation is done.
Table 1: Description of the participants.

<table>
<thead>
<tr>
<th>Number</th>
<th>Date</th>
<th>Sex</th>
<th>Age</th>
<th>Publication</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Anonymous</td>
<td>Free</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>07.08.2012</td>
<td>f</td>
<td>23</td>
<td>×</td>
</tr>
<tr>
<td>2</td>
<td>08.08.2012</td>
<td>m</td>
<td>24</td>
<td>×</td>
</tr>
<tr>
<td>3</td>
<td>08.08.2012</td>
<td>f</td>
<td>31</td>
<td>×</td>
</tr>
<tr>
<td>4</td>
<td>07.08.2012</td>
<td>f</td>
<td>27</td>
<td>×</td>
</tr>
<tr>
<td>5</td>
<td>08.08.2012</td>
<td>m</td>
<td>26</td>
<td>×</td>
</tr>
<tr>
<td>6</td>
<td>08.08.2012</td>
<td>m</td>
<td>28</td>
<td>×</td>
</tr>
<tr>
<td>7</td>
<td>07.08.2012</td>
<td>m</td>
<td>24</td>
<td>×</td>
</tr>
<tr>
<td>8</td>
<td>08.08.2012</td>
<td>m</td>
<td>24</td>
<td>×</td>
</tr>
<tr>
<td>9</td>
<td>08.08.2012</td>
<td>m</td>
<td>27</td>
<td>×</td>
</tr>
<tr>
<td>10</td>
<td>08.08.2012</td>
<td>m</td>
<td>28</td>
<td>×</td>
</tr>
<tr>
<td>11</td>
<td>07.08.2012</td>
<td>m</td>
<td>24</td>
<td>×</td>
</tr>
<tr>
<td>12</td>
<td>07.08.2012</td>
<td>m</td>
<td>31</td>
<td>×</td>
</tr>
<tr>
<td>13</td>
<td>07.08.2012</td>
<td>f</td>
<td>25</td>
<td>×</td>
</tr>
<tr>
<td>14</td>
<td>07.08.2012</td>
<td>f</td>
<td>27</td>
<td>×</td>
</tr>
<tr>
<td>15</td>
<td>15.08.2012</td>
<td>m</td>
<td>30</td>
<td>×</td>
</tr>
<tr>
<td>16</td>
<td>08.08.2012</td>
<td>m</td>
<td>28</td>
<td>×</td>
</tr>
<tr>
<td>17</td>
<td>07.08.2012</td>
<td>m</td>
<td>29</td>
<td>×</td>
</tr>
<tr>
<td>18</td>
<td>08.08.2012</td>
<td>f</td>
<td>25</td>
<td>×</td>
</tr>
<tr>
<td>19</td>
<td>08.08.2012</td>
<td>f</td>
<td>23</td>
<td>×</td>
</tr>
<tr>
<td>20</td>
<td>08.08.2012</td>
<td>f</td>
<td>25</td>
<td>×</td>
</tr>
</tbody>
</table>
B. ASM facial matching for a subject wearing glasses

For this test two images from the same subject, with and without glasses, are taken. The subject, as far as the model concerns, is unknown since any of his images it is being included in the training set. Both images show approximately a frontal view, although they present a little variations, either in the expression or in the pose. Figure 1 shows how the model matches both facial shapes.

Both matchings in the figure can be considered acceptable. In any case the little divergences that the matched shapes present are due to the presence of the glasses, but to the limitations of the model and its training set to reach some facial expressions or poses (as it is indicated in 3.1.5 in the matter of the chosen training set restrictions). The lenses allow the access to the whole area of the eyes and do not represent significant changes in the brightness.

Then it is possible to deduce that whenever the image search process can face the gray value structures is looking for, a proper shape matching is obtained. This consideration involves all the cases that do not present any
kind of occlusion and poor or excessive levels of brightness.
The UI-5240CP is an ultra-compact camera with e2v CMOS sensor in 1.3 Megapixel resolution (1280x1024 pixels) and a GigE interface. In addition, the UI-5240CP can also be powered using Power-over-GigE (PoE) or an external mains adaptor. Besides the screw-on RJ45 plug for GigE, the UI-5240CP features a 6-pin Hirose connector for optional power supply and digital in-/outputs (optically decoupled up to 30 V). These camera models incorporate the monochrome/color sensors EV76C560BB/EV76C560BC from e2v.

### Specification

#### The characteristics at a glance

<table>
<thead>
<tr>
<th>Feature</th>
<th>Specification</th>
</tr>
</thead>
<tbody>
<tr>
<td>Interface</td>
<td>GigE</td>
</tr>
<tr>
<td>Lens Mount</td>
<td>C-Mount</td>
</tr>
<tr>
<td>Sensor Technology</td>
<td>CMOS (e2v)</td>
</tr>
<tr>
<td>Model Description (color)</td>
<td>UI-5240CP-C</td>
</tr>
<tr>
<td>Model Description (mono)</td>
<td>UI-5240CP-M</td>
</tr>
<tr>
<td>Resolution (h x v)</td>
<td>1280 x 1024</td>
</tr>
<tr>
<td>Resolution Depth</td>
<td>10bit (10bit ADC)</td>
</tr>
<tr>
<td>Resolution Category / Pixel Class</td>
<td>1.3 Megapixel</td>
</tr>
<tr>
<td>Sensor Size</td>
<td>1/1.8&quot;</td>
</tr>
<tr>
<td>Shutter</td>
<td>Global</td>
</tr>
<tr>
<td>max. fps in Free Run Mode</td>
<td>50 fps</td>
</tr>
<tr>
<td>Exposure Time in Free Run Mode</td>
<td>9µs - 2000ms</td>
</tr>
<tr>
<td>Exposure Time in Trigger Mode</td>
<td>9µs - 2000ms</td>
</tr>
<tr>
<td>AOI Modes</td>
<td>horizontal + vertical</td>
</tr>
<tr>
<td>Binning Modes</td>
<td>horizontal + vertical</td>
</tr>
<tr>
<td>Subsampling Modes</td>
<td>---</td>
</tr>
<tr>
<td>I/O - Strobe</td>
<td>1</td>
</tr>
<tr>
<td>I/O - Trigger</td>
<td>1</td>
</tr>
<tr>
<td>I/O - RS-232</td>
<td>-</td>
</tr>
<tr>
<td>I/O - GPIO</td>
<td>-</td>
</tr>
<tr>
<td>I/O - I2C</td>
<td>-</td>
</tr>
<tr>
<td>HDR (High Dynamic Range)</td>
<td>-</td>
</tr>
<tr>
<td>Sensor Model (m/c)</td>
<td>EV76C560BB / EV76C560BC</td>
</tr>
<tr>
<td>Pixel Pitch in µm</td>
<td>5.30</td>
</tr>
<tr>
<td>Optical Size</td>
<td>6.784 x 5.427 mm</td>
</tr>
<tr>
<td>Protection Classes</td>
<td>IP30</td>
</tr>
<tr>
<td>Dimensions H/W/L</td>
<td>29.00 mm, 29.00 mm, 41.00 mm</td>
</tr>
<tr>
<td>Mass</td>
<td>55.00 g</td>
</tr>
<tr>
<td>Power Supply</td>
<td>PoE + ext. 12 - 24 V</td>
</tr>
</tbody>
</table>
Bibliography


