2D Segmentation Using a Robust Active Shape Model With the EM Algorithm

Carlos Santiago, Jacinto C. Nascimento, *Member, IEEE*, and Jorge S. Marques

***Abstract*— Statistical shape models have been extensively used in a wide range of applications due to their effectiveness in providing prior shape information for object segmentation problems. The most popular method is the active shape model (ASM). However, accurately fitting the shape model to an object boundary under a cluttered environment is a challenging task. Under such assumptions, the model is often attracted toward invalid observations (outliers), leading to meaningless estimates of the object boundary. In this paper, we propose a novel algorithm that improves the robustness of ASM in the presence of outliers. The proposed framework assumes that both type of observations (valid observations and outliers) are detected in the image. A new strategy is devised for treating the data in different ways, depending on the observations being considered as valid or invalid. The proposed algorithm assigns a different weight to each observation. The shape parameters are recursively updated using the expectation-maximization method, allowing a correct and robust fit of the shape model to the object boundary in the image. Two estimation criteria are considered: 1) the maximum likelihood criterion and 2) the maximum *a posteriori* criterion that use priors for the unknown parameters. The methods are tested with synthetic and real images, comprising medical images of the heart and image sequences of the lips. The results are promising and show that this approach is robust in the presence of outliers, leading to a significant improvement over the standard ASM and other state-of-the-art methods.**

***Index Terms*— Image segmentation, active shape model, expectation-maximization algorithms.**

I. Introduction

OBJECT segmentation remains a complex task that is required in a large spectrum of applications. Statistical shape models (SSM) have had an important part in the development of sophisticated segmentation methods, due to their ability to capture prior information about the shape of the object. However, the estimation of the model parameters is often hampered by the presence of noisy observations [1]. This paper proposes a novel Bayesian framework approach for the estimation of the model parameters that allows for a reliable and robust segmentation of the object in the presence of outliers.

The most popular SSM is the Active Shape Model (ASM) method, proposed by Cootes et al. [2]. It consists of characterizing the shape of an object by learning the mean shape and its most significant modes of deformation from training data. In this method, the contour of an object is characterized by the parameters of a pose transformation and by the deformation parameters. This provides a simple and effective way of modeling the shape space of the object, and preventing unpredictable or unexpected segmentations.

Fitting the learned model to the boundary of an object in an image comprises two steps: 1) searching for observation points located on the object's boundary in the image, and 2) estimating the model parameters (pose and deformation parameters) based on those observations. The ASM method assumes the distribution of residuals between the model and the observations is Gaussian. However, in most real applications this is not true (e.g., ultrasound images of the heart have a significant level of multiplicative noise) and the least squares method used to estimate the parameters may lead to poor segmentations in the presence of outliers [3]. One approach that has been used to overcome this limitation is to improve the boundary detection method [4]–[9]. By reducing the number of outliers detected, these works are able to improve the accuracy of the segmentations. An alternative approach is to use an estimation method that is able to deal with the presence of outliers.

This paper focuses on the second type of approach and proposes a new method for estimating the shape model parameters that is robust in the presence of outliers. The algorithm assumes that some of the observation points are outliers and takes this into account when estimating the parameters.

The paper is organized as follows. Section II presents an overview of the state of the art. Section III formalizes the problem, and the proposed framework is described in Sections IV-V. The experimental setup and statistical results are shown in Section VI-VII, as well as a comparison with state of the art works. Finally, Section VIII concludes the paper.

II. State of the Art

The standard ASM estimates the shape model parameters under the assumption that all the observation points belong to the object boundary. In most applications, this is not true and, consequently, the segmentations obtained in the presence of noisy observations are often poor. This drawback of the ASM method caused subsequent works to propose alternative ways of estimating the shape model parameters [1], [9]–[12].

Most works consider multiple candidates for the location of each model point. This makes the algorithm less dependent on the accuracy of the boundary detection method and increases

Manuscript received April 11, 2014; revised September 29, 2014 and March 20, 2015; accepted April 1, 2015. Date of publication April 20, 2015; date of current version May 20, 2015. This work was supported in part by the Fundação para a Ciência e a Tecnologia under Grant UID/EEA/50009/2013 and in part by the Ph.D. Program under Grant SFRH/BD/87347/2012. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Scott T. Acton.

The authors are with the Instituto de Sistemas e Robótica, Instituto Superior Técnico, Lisbon 1049-001, Portugal (e-mail: carlos.santiago@ist.utl.pt; jan@isr.ist.utl.pt; jsm@isr.ist.utl.pt).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TIP.2015.2424311

the probability that the true object boundary is among the detected observation points. Then, the estimation of the model parameters is accomplished by simultaneously selecting the subset of observation points that maximizes a specific objective function, which typically promotes i) segmentations along points with specific image features, e.g., edge points (located along an image edge), and ii) segmentations with shapes similar to those observed in the training set.

Different approaches have been used to accomplish this. Wang and Staib [10], for instance, use the gradient descent method to obtain the model parameters that maximize the posterior probability, which includes a prior about the parameters learned from a training set and the likelihood of each pixel location. Similarly, Cootes et al. [9] proposes an algorithm for maximizing the overall quality of fit between the shape model and the image, based on a random forest regressor that evaluates the quality of fit of the image pixels around each model point. However, these two methods do not explicitly account for the presence of outliers. Instead, they rely on the objective function (posterior probability/overall quality of fit) to guide the model toward the true object boundary.

Few works have proposed to deal with outliers explicitly. For instance, the approach used in [11] is based on the Robust Point Matching (RPM) algorithm [13]. The RPM algorithm pairs observation points with model points, and the points that are left unpaired are considered outliers and disregarded in the estimation of the parameters. Rogers and Graham [1], on the other hand, tests two types of approaches: 1) M-estimators, which weight the observation points based on some criteria, and 2) random sampling strategies (e.g., the Random Sample Consensus (RANSAC) algorithm [14]) to determine the best subset of observation points to be used in the estimation of the shape model parameters. Lekadir et al. [12] proposes to determine which observation points are outliers using a local shape dissimilarity metric. This metric is based on the distance between pairs of model points. Observation points that do not respect the local shape of the object are treated as outliers. A disadvantage of this approach is that only one observation point is allowed for each model point.

We propose a new method for estimating the model parameters using a Bayesian approach that assigns each observation a different weight based on the probability of that observation belonging to the object boundary. The main difference between [1] and our work is the following: in [1], the weights are obtained using an heuristic strategy (see [1, eq. (10)]) designed by Huber [15]; we propose a principled and well founded way to automatically determine these weights in a probabilistic framework. Since outliers often receive low probabilities, their influence in the parameter estimation is reduced, making the method robust in the presence of outliers. A similar approach has been previously used in [16] for the estimation of a Snake model [17]. However, a drawback in [16] is that the Snake model can assume unexpected shapes, since is does not use prior information about the object shape. The following sections describe the problem and the proposed methodology.

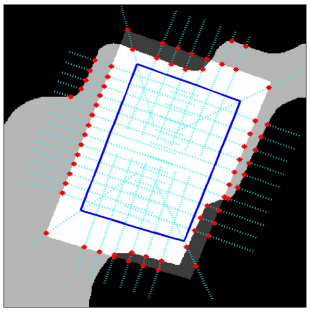


Fig. 1. Example of the detection of observation points. The blue line corresponds to the model; the dashed cyan lines correspond to the search lines; and the red dots correspond to the detected observation points.

III. Problem Formulation

Given an image I, we wish to approximate the boundary of an object of interest by a sequence of 2D points, $\mathbf{x} = ({\mathbf{x}^1}^\top, ..., {\mathbf{x}^N}^\top)^\top$, where $\mathbf{x} \in \mathbb{R}^{2N\times 1}$, $\mathbf{x}^i = (x_1^i, x_2^i)^\top$ are the coordinates of i-th point and N is the number of points in the sequence. The contour model adopted in this paper is shape model, which uses prior information about the shape of the object.

The estimation of the shape model is based on a training set $\mathcal{D} = \{(I, \mathbf{x})_j\}_{j=1}^{|\mathcal{D}|}$, containing the training images $I_j : \Omega \rightarrow [0, 255]$ and the respective annotations $\mathbf{x}_j$ (training shapes), where Ω stands for the image lattice. The shape model is learned in a two step procedure [2]: 1) align all the training shapes $\mathbf{x}_j$, with $j = 1, .., |\mathcal{D}|$ and 2) perform a Principal Component Analysis (PCA) to find the modes of variation. Every contour in the shape space can be approximately defined by the average contour $\bar{\mathbf{x}} = ({\bar{\mathbf{x}}^1}^\top, ..., {\bar{\mathbf{x}}^N}^\top)^\top \in \mathbb{R}^{2N\times 1}$, with $\bar{\mathbf{x}}^i \in \mathbb{R}^2$, deformed by a linear combination of the K main modes of variation $\mathbf{D} \in \mathbb{R}^{2N\times K}$, leading to

$$\mathbf{x} \simeq \bar{\mathbf{x}} + \mathbf{D}\mathbf{b}, \tag{1}$$

where the vector $\mathbf{b} \in \mathbb{R}^{K\times 1}$ contains the deformation coefficients. Accordingly, the i-th model point is described by

$$\mathbf{x}^i \simeq \bar{\mathbf{x}}^i + \mathbf{D}^i\mathbf{b}, \tag{2}$$

where $\mathbf{D}^i \in \mathbb{R}^{2\times K}$ is a matrix containing the lines of $\mathbf{D}$ associated to the i-th model point.

It will be assumed that the shape model undergoes a global (pose) transformation $\mathbf{T}$ and each transformed point, $\widetilde{\mathbf{x}}^i$, is given by

$$\begin{aligned}\widetilde{\mathbf{x}}^i &= \mathbf{T}(\mathbf{x}^i)\\ &= \begin{bmatrix} a_1 & -a_2 \\ a_2 & a_1 \end{bmatrix}\begin{bmatrix} x_1^i \\ x_2^i \end{bmatrix} + \begin{bmatrix} t_1 \\ t_2 \end{bmatrix}\\ &= \mathbf{A}\mathbf{x}^i + \mathbf{t}\\ &= \mathbf{A}(\bar{\mathbf{x}}^i + \mathbf{D}^i\mathbf{b}) + \mathbf{t}.\end{aligned} \tag{3}$$

To segment a new image, we have to determine the best model parameters. For this we need an initial estimate of the contour configuration, $\widetilde{\mathbf{x}}$. A set of observation points is then extracted from the image in the vicinity of each model point $\widetilde{\mathbf{x}}^i$. These observations should ideally be located at the boundary of the object of interest. In this work, the object boundary is searched along lines orthogonal to the contour at each model

point, as shown in Fig. 1. In each search line, the observation points were detected using the match filter for edges (see [18] Section 5.2), which was originally proposed in the context of object tracking. Other search methods can be used instead (e.g., searching within a region around each model point, as in [9]). This allows us to obtain set of observation points $\mathbf{Y}^i = \{\mathbf{y}^{ij}, j = 1, ..., M^i\}$ located in the vicinity of the model point $\widetilde{\mathbf{x}}^i$. In general, $M^i \neq M^j$ for $i \neq j$. It is important to remark that many of the observations contained in the set $\mathbf{Y}^i$ do not belong to the object boundary and should be considered as outliers. Unfortunately, we do not know, beforehand, which observation points are valid (truly belonging to the object boundary) or invalid (outliers). To tackle this issue, we consider two possible sensor-observation models and assign a binary label $k^{ij} \in \{0, 1\}$ to the observation point, $\mathbf{y}^{ij}$, where $k^{ij} = 1$, if $\mathbf{y}^{ij}$ is considered valid, and $k^{ij} = 0$ otherwise. The label probabilities, $p_0 = p(k^{ij} = 0)$ and $p_1 = p(k^{ij} = 1)$ are also unknown, i.e., they also need to be estimated.

The model described so far comprises three sets of parameters: 1) global transformation parameters $(\mathbf{a}, \mathbf{t})$, 2) local deformation parameters $\mathbf{b}$, and 3) the label probabilities $\mathbf{p} = (p_0, p_1)$. In the following, we will denote the set of all parameters by $\boldsymbol{\Theta} = (\boldsymbol{\theta}, \mathbf{p})$, with $\boldsymbol{\theta} = (\mathbf{a}, \mathbf{t}, \mathbf{b})$ where $\mathbf{a} = (a_1, a_2)$ and $\mathbf{t} = (t_1, t_2)$.

Let us consider the j-th observation point detected in the vicinity of the i-th model point, $\mathbf{y}^{ij} = \left(y_1^{ij}, y_2^{ij}\right)^\top$, and let k^{ij} be the corresponding model label assigned to $\mathbf{y}^{ij}$. Assuming that the observation $\mathbf{y}^{ij}$ is valid ($k^{ij} = 1$) the following sensor model is considered

$$\mathbf{y}^{ij} = \widetilde{\mathbf{x}}^i + \mathbf{v}^i, \tag{4}$$

where $\mathbf{v}^i \sim \mathcal{N}(\mathbf{0}, \Sigma^i)$ is a zero mean white Gaussian noise with diagonal covariance matrix Σ^i, estimated from the training set. Combining (3) and (4) yields

$$\begin{aligned}\mathbf{y}^{ij} &= \mathbf{T}(\mathbf{x}^i) + \mathbf{v}^i \\ &= \mathbf{A}(\bar{\mathbf{x}}^i + \mathbf{D}^i\mathbf{b}) + \mathbf{t} + \mathbf{v}^i.\end{aligned} \tag{5}$$

It follows that the probability of $\mathbf{y}^{ij}$ being a valid observation is given by

$$p\left(\mathbf{y}^{ij}\middle|k^{ij} = 1\right) = \mathcal{N}\left(\mathbf{y}^{ij}; \mathbf{A}(\bar{\mathbf{x}}^i + \mathbf{D}^i\mathbf{b}) + \mathbf{t}, \Sigma^i\right), \tag{6}$$

where $\mathcal{N}(\cdot; \mu, \Sigma)$ denotes a normal distribution with mean μ and covariance matrix Σ. On the other hand, it will be assumed that the the observation point $\mathbf{y}^{ij}$ with label $k^{ij} = 0$ (*i.e.,* outlier) follows a uniform distribution within a validation gate $V_{\widetilde{\mathbf{x}}^i}$ in the vicinity of $\widetilde{\mathbf{x}}^i$, and

$$p\left(\mathbf{y}^{ij}\middle|k^{ij} = 0\right) = \mathcal{U}(V_{\widetilde{\mathbf{x}}^i}). \tag{7}$$

Given the probabilistic model used to represent the data, we wish to estimate the model parameters (global transformation $(\mathbf{a}, \mathbf{t})$ and deformation coefficients $\mathbf{b}$) needed to fit the shape model to the observation points, $\mathbf{Y} = \{\mathbf{y}^{ij}\}$ extracted from the image. Ideally, we would like to know the binary labels $\mathbf{K} = \{k^{ij}\}$ associated to each observation points $\mathbf{y}^{ij}$, but this information is not available. Therefore, $\mathbf{K}$ is an unobserved (hidden) set of labels. The class-conditional generative model is thus obtained by marginalizing with respect to the missing label sequence, as follows

$$p(\mathbf{Y}|\boldsymbol{\Theta}) = \sum_{\mathbf{K}\in\{0,1\}^M} p\left(\mathbf{Y}, \mathbf{K}|\boldsymbol{\Theta}\right) \tag{8}$$

where $M = \sum_{i=1}^{N} M^i$ is the total number (often large) of detected observations.

IV. Expectation Maximization Framework

The estimation of the model parameters by maximization of (8) would require taking into account all possible combinations of $\mathbf{K}$, which is infeasible. In fact, it is not possible to obtain a closed form expression nor to analytically optimize it. This difficulty is circumvented by searching for a sub-optimal solution using the Expectation-Maximization (EM) method [19]. The EM method iteratively updates the transformation and deformation parameters by maximizing an auxiliary function, $Q\left(\boldsymbol{\Theta}; \widehat{\boldsymbol{\Theta}}_{(t)}\right)$ where $\widehat{\boldsymbol{\Theta}}_{(t)}$ is the most recent estimate of the unknown parameters. In this work, two approaches will be described: *A*. maximizing the expectation of the likelihood function, which leads to the maximum likelihood (ML) estimate; and *B*. maximizing the expectation of the joint probability function, which leads to the maximum *a posteriori* (MAP) estimate. These two approaches are detailed next.

A. Maximum Likelihood Estimation

In the ML framework, the complete log-likelihood of a set of observation points $\mathbf{Y}$ and labels $\mathbf{K}$ is given by

$$\begin{aligned}\mathcal{L}(\mathbf{Y}, \mathbf{K}, \boldsymbol{\Theta}) &= \log p\left(\mathbf{Y}, \mathbf{K}|\boldsymbol{\Theta}\right) \\ &= \log\left(p\left(\mathbf{Y}|\mathbf{K}, \boldsymbol{\Theta}\right) p\left(\mathbf{K}\right)\right).\end{aligned} \tag{9}$$

Assuming conditional independence between observations, the previous equation can be factorized as follows

$$\begin{aligned}\mathcal{L}(\mathbf{Y}, \mathbf{K}, \boldsymbol{\Theta}) &= \log\left(\prod_{i=1}^{N}\prod_{j=1}^{M^i} p\left(\mathbf{y}^{ij}\middle|k^{ij}, \boldsymbol{\Theta}\right) p\left(k^{ij}\right)\right) \\ &= \sum_{i=1}^{N}\sum_{j=1}^{M^i} \log p\left(\mathbf{y}^{ij}\middle|k^{ij}, \boldsymbol{\Theta}\right) + \log p\left(k^{ij}\right),\end{aligned} \tag{10}$$

where $p\left(k^{ij}\right)$ is the probability of the label k^{ij} that depends on whether the observation $\mathbf{y}^{ij}$ belongs to the object boundary or is an outlier, i.e., p_1 or p_0, respectively.

The EM algorithm comprises the two following steps: 1) the *E-step*, in which we update the expectation of the log-likelihood by computing the probability of each observation, based on a previous estimate of the parameters; and 2) the *M-step*, in which we update the parameters by maximizing the expectation obtained in the E-step. These steps are detailed in the following subsections.

E-Step: Let $\widehat{\boldsymbol{\Theta}}_{(t)} = (\widehat{\mathbf{a}}, \widehat{\mathbf{t}}, \widehat{\mathbf{b}}, \widehat{\mathbf{p}})_{(t)}$ be the estimates of the model parameters at iteration t. Given a set of observations, $\mathbf{Y}$, and the most recent estimates, we can define the

auxiliary function $Q_{\mathrm{ML}}\left(\boldsymbol{\Theta};\widehat{\boldsymbol{\Theta}}_{(t)}\right)$ as the expected value of the log-likelihood function

$$\begin{aligned}
&Q_{\mathrm{ML}}\left(\boldsymbol{\Theta};\widehat{\boldsymbol{\Theta}}_{(t)}\right)\\
&\quad=\mathbb{E}_{\mathbf{K}}\left[\mathcal{L}(\mathbf{Y},\mathbf{K},\boldsymbol{\Theta})\middle|\mathbf{Y},\widehat{\boldsymbol{\Theta}}_{(t)}\right]\\
&\quad=\sum_{i=1}^{N}\sum_{j=1}^{M^i}\mathbb{E}_{\mathbf{K}}\left[\log p(\mathbf{y}^{ij}\mid k^{ij},\boldsymbol{\Theta})+\log p(k^{ij})\middle|\mathbf{Y},\widehat{\boldsymbol{\Theta}}_{(t)}\right]\\
&\quad=\sum_{i=1}^{N}\sum_{j=1}^{M^i}\sum_{l=0}^{1}p(k^{ij}=l|\mathbf{y}^{ij},\widehat{\boldsymbol{\Theta}}_{(t)})\\
&\qquad\times\left[\log p(\mathbf{y}^{ij}\mid k^{ij}=l,\boldsymbol{\theta})+\log p(k^{ij}=l)\right]\\
&\quad=\sum_{i=1}^{N}\sum_{j=1}^{M^i}w_0^{ij}\left(\log p\left(\mathbf{y}^{ij}\middle|k^{ij}=0,\boldsymbol{\theta}\right)+\log p_0\right)\\
&\qquad+w_1^{ij}\left(\log p\left(\mathbf{y}^{ij}\middle|k^{ij}=1,\boldsymbol{\theta}\right)+\log p_1\right),
\end{aligned}\tag{11}$$

where $w_l^{ij}, l\in\{0,1\}$ denotes the confidence degree of observation $\mathbf{y}^{ij}$, given by

$$\begin{aligned}
w_1^{ij}&=p\left(k^{ij}=1\middle|\mathbf{y}^{ij},\widehat{\boldsymbol{\Theta}}_{(t)}\right)\\
&\propto\widehat{p}_{1_{(t)}}\;p\left(\mathbf{y}^{ij}\middle|k^{ij}=1,\widehat{\boldsymbol{\theta}}_{(t)}\right)\\
&\propto\widehat{p}_{1_{(t)}}\;\mathcal{N}\left(\mathbf{y}^{ij};\widehat{\mathbf{A}}_{(t)}(\bar{\mathbf{x}}^i+\mathbf{D}^i\widehat{\mathbf{b}}_{(t)})+\widehat{\mathbf{t}}_{(t)},\Sigma^i\right)
\end{aligned}\tag{12}$$

$$\begin{aligned}
w_0^{ij}&=p\left(k^{ij}=0\middle|\mathbf{y}^{ij},\widehat{\boldsymbol{\Theta}}_{(t)}\right)\\
&\propto\widehat{p}_{0_{(t)}}\;\mathcal{U}\left(V_{\bar{\mathbf{x}}^i}\right)
\end{aligned}\tag{13}$$

such that $w_0^{ij}+w_1^{ij}=1$. These weights correspond to the probability of the labels assigned to the observation $\mathbf{y}^{ij}$ being $k^{ij}=1$ and $k^{ij}=0$, respectively, given the current model estimate.

M-Step: Given the most recent estimates of the unknown parameters at iteration t, $\widehat{\boldsymbol{\Theta}}_{(t)}=(\widehat{\mathbf{a}},\widehat{\mathbf{t}},\widehat{\mathbf{b}},\widehat{\mathbf{p}})_{(t)}$, the M-step aims to solve the following optimization problem

$$\widehat{\boldsymbol{\Theta}}_{(t+1)}=\arg\max_{\boldsymbol{\Theta}}Q_{\mathrm{ML}}\left(\boldsymbol{\Theta};\widehat{\boldsymbol{\Theta}}_{(t)}\right).\tag{14}$$

This is done by taking the derivative of $Q_{\mathrm{ML}}(\boldsymbol{\Theta};\widehat{\boldsymbol{\Theta}}_{(t)})$ with respect to each parameter and equating to zero. We simplify this step by sequentially updating the transformation and deformation parameters, as follows: 1) compute $\widehat{\mathbf{a}}_{(t+1)}$ and $\widehat{\mathbf{t}}_{(t+1)}$, assuming $\mathbf{b}=\widehat{\mathbf{b}}_{(t)}$ fixed; then 2) compute $\widehat{\mathbf{b}}_{(t+1)}$ assuming both $\mathbf{a}=\widehat{\mathbf{a}}_{(t+1)}$ and $\mathbf{t}=\widehat{\mathbf{t}}_{(t+1)}$ fixed; and finally, 3) update the probabilities of each model $\widehat{\mathbf{p}}_{(t+1)}$. These three steps can be solved using standard matrix calculus as will be described next.

1) Update of the Transformation Parameters: Let $\mathbf{x}^i=(x_1^i,x_2^i)^\top=\bar{\mathbf{x}}^i+\mathbf{D}^i\widehat{\mathbf{b}}_{(t)}$, $i=1,\ldots,N$, be the deformed shape according to the current estimate of the deformation parameters, $\widehat{\mathbf{b}}_{(t)}$. The optimization of equation (14) with respect to the transformation parameters, $\mathbf{t}=(t_1,t_2)^\top$ and $\mathbf{a}=(a_1,a_2)^\top$, yields the following equations

$$\begin{aligned}
&\frac{\partial}{\partial\mathbf{a}}Q_{\mathrm{ML}}(\boldsymbol{\Theta};\widehat{\boldsymbol{\Theta}}_{(t)})=0\\
&\iff\sum_{i=1}^{N}\sum_{j=1}^{M^i}w_1^{ij}\mathbf{X}^{i\top}\Sigma^{i-1}\left(-\mathbf{y}^{ij}+\mathbf{X}^i\mathbf{a}_{(t+1)}+\mathbf{t}_{(t+1)}\right)=0
\end{aligned}\tag{15}$$

$$\begin{aligned}
&\frac{\partial}{\partial\mathbf{t}}Q_{\mathrm{ML}}(\boldsymbol{\Theta};\widehat{\boldsymbol{\Theta}}_{(t)})=0\\
&\iff\sum_{i=1}^{N}\sum_{j=1}^{M^i}w_1^{ij}\Sigma^{i-1}\left(-\mathbf{y}^{ij}+\mathbf{X}^i\mathbf{a}_{(t+1)}+\mathbf{t}_{(t+1)}\right)=0,
\end{aligned}\tag{16}$$

where

$$\mathbf{X}^i=\begin{bmatrix}x_1^i & -x_2^i\\ x_2^i & x_1^i\end{bmatrix}$$

Combining (15) and (16) leads to the following linear system of equations

$$\begin{aligned}
&\sum_{i=1}^{N}\sum_{j=1}^{M^i}w_1^{ij}\begin{bmatrix}\mathbf{X}^{i\top}\Sigma^{i-1}\mathbf{X}^i & \mathbf{X}^{i\top}\Sigma^{i-1}\\ \Sigma^{i-1}\mathbf{X}^i & \Sigma^{i-1}\end{bmatrix}\begin{bmatrix}\mathbf{a}_{(t+1)}\\ \mathbf{t}_{(t+1)}\end{bmatrix}\\
&\qquad=\sum_{i=1}^{N}\sum_{j=1}^{M^i}w_1^{ij}\begin{bmatrix}\mathbf{X}^{i\top}\Sigma^{i-1}\mathbf{y}^{ij}\\ \Sigma^{i-1}\mathbf{y}^{ij}\end{bmatrix}.
\end{aligned}\tag{17}$$

The transformation parameters are updated by solving (17). This can be interpreted as a weighted least squares solution to the alignment of two sets of points (the observation points $\{\mathbf{y}^{ij}\}$ and the corresponding model points $\{\mathbf{x}^i\}$). This estimation of the transformation parameters differs from the standard ASM [2] in the following:

1) There is no restriction on the number of observation points detected in the vicinity of each model point, M^i, whereas in the standard ASM, $M^i=1$; and
2) Each observation point has a specific weight, w_1^{ij} in the estimation, as opposed to a constant weight of 1 in the standard ASM.

The first difference means that more observation points can be detected, thus reducing the importance of the boundary detection method and increasing the chance of detecting the true object boundary. As for the latter, since outliers often receive lower values of w_1^{ij} than the valid observations, the update of the transformation parameters will be less influenced by the presence of outliers.

2) Update of the Deformation Parameters: Once the transformation parameters

$$\widehat{\mathbf{A}}_{(t+1)}=\begin{bmatrix}\widehat{a_1} & -\widehat{a_2}\\ \widehat{a_2} & \widehat{a_1}\end{bmatrix}_{(t+1)},\quad\widehat{\mathbf{t}}_{(t+1)}=\begin{bmatrix}\widehat{t_1}\\ \widehat{t_2}\end{bmatrix}_{(t+1)},$$

have been updated using (17), we maximize $Q_{\mathrm{ML}}(\boldsymbol{\Theta};\widehat{\boldsymbol{\Theta}}_{(t)})$ in (11) with respect to $\mathbf{b}$, leading to

$$\begin{aligned}
&\left[\sum_{i=1}^{N}\sum_{j=1}^{M^i}w_1^{ij}\mathbf{D}^{i\top}\widehat{\mathbf{A}}_{(t+1)}^{\top}\Sigma^{i-1}\widehat{\mathbf{A}}_{(t+1)}\mathbf{D}^i\right]\widehat{\mathbf{b}}_{(t+1)}\\
&\quad=\left[\sum_{i=1}^{N}\sum_{j=1}^{M^i}w_1^{ij}\mathbf{D}^{i\top}\widehat{\mathbf{A}}_{(t+1)}^{\top}\Sigma^{i-1}\left(\mathbf{y}^{ij}-\widehat{\mathbf{A}}_{(t+1)}\bar{\mathbf{x}}^i-\widehat{\mathbf{t}}_{(t+1)}\right)\right].
\end{aligned}\tag{18}$$

Note that the deformation parameters obtained by solving (18) may correspond to an unexpected shape. Therefore we adopt the additional step proposed in [2]. We want the Mahalanobis distance, d, to be lower than a specific threshold, $d_{\max}$,

$$d^2 = \sum_{l=1}^{K} \frac{\widehat{b}_l^2}{\lambda_l} \leq d_{\max}^2. \tag{19}$$

where $\widehat{b}_l$ denotes the l-th component of $\widehat{\mathbf{b}}_{(t+1)}$, and λ_l is the eigenvalue associated to the l-th deformation mode. The threshold is chosen so that most of the shapes in the training set satisfy (19) (a typical value is $d_{\max} = 3$). If $\widehat{\mathbf{b}}_{(t+1)}$ does not satisfy (19), we rescale it to the closest acceptable shape as follows

$$\widehat{\mathbf{b}}_{(t+1)} \leftarrow \widehat{\mathbf{b}}_{(t+1)} \frac{d_{\max}}{d} \quad \text{if } d > d_{\max}. \tag{20}$$

3) Update of the Models Probabilities: Finally, the estimates of the probabilities of each model, $\widehat{p}_{0(t)}, \widehat{p}_{1(t)}$, are updated. Maximizing $Q_{\mathrm{ML}}(\boldsymbol{\Theta}; \widehat{\boldsymbol{\Theta}}_{(t)})$ with respect to p_0, p_1 yields

$$\widehat{p}_{1_{(t+1)}} = \frac{\sum_{i=1}^{N}\sum_{j=1}^{M^i} w_1^{ij}}{\sum_{i=1}^{N}\sum_{j=1}^{M^i} w_1^{ij} + w_0^{ij}} = \frac{\sum_{i=1}^{N}\sum_{j=1}^{M^i} w_1^{ij}}{\sum_{i=1}^{N} M^i}$$
$$\widehat{p}_{0_{(t+1)}} = 1 - \widehat{p}_{1_{(t+1)}}. \tag{21}$$

where $\sum_{i=1}^{N} M^i$ is the total number of detected observation points.

B. Maximum a Posteriori Estimation

The estimation of the model parameters by the MAP method is done in a similar way. However, instead of using the complete log-likelihood, this approach uses the complete log-joint probability, which can be expressed as

$$\begin{aligned}\mathcal{P}(\mathbf{Y}, \mathbf{K}, \boldsymbol{\Theta}) &= \log p(\mathbf{Y}, \mathbf{K}, \boldsymbol{\Theta}) \\ &= \log\Big(p(\mathbf{Y}, \mathbf{K}|\boldsymbol{\Theta})\, p(\boldsymbol{\Theta})\Big) \\ &= \log p\,(\mathbf{Y}|\mathbf{K}, \boldsymbol{\Theta}) + \log p\,(\mathbf{K}) + \log p(\mathbf{a}, \mathbf{t}) \\ &\quad + \log p(\mathbf{b}) + \log p(\mathbf{p})\end{aligned} \tag{22}$$

The first two terms are the same as in (10); the remaining terms are the prior probabilities for the transformation parameters, $(\mathbf{a}, \mathbf{t})$, deformation parameters, $\mathbf{b}$, and sensor probabilities, $\mathbf{p}$. As previously, the EM algorithm is used to estimate the model parameters by maximizing the expectation of (22).

The prior for the deformation parameters $\mathbf{b}$ is obtained from the PCA method (used to learn the shape model), which assumes they are normally distributed with zero mean

$$p(\mathbf{b}) = \mathcal{N}\left(\mathbf{b}; \mathbf{0}, \Sigma^{\mathbf{b}}\right) = c_{\mathbf{b}} \exp\left(-\frac{1}{2}\mathbf{b}^\top {\Sigma^{\mathbf{b}}}^{-1}\mathbf{b}\right), \tag{23}$$

where $c_{\mathbf{b}}$ is a normalization constant, and $\Sigma^{\mathbf{b}}$ is a $K \times K$ diagonal matrix whose entries are the eigenvalues $\Sigma^{\mathbf{b}}_{ll} = \lambda_l$, obtained by the PCA method. This prior assigns higher probability to shapes that are similar to the average shape in the training set, i.e., not too deformed.

Regarding the transformation parameters, in this work we will use non-informative priors to define the prior term $p(\mathbf{a}, \mathbf{t})$. Although we assume that we have no prior information about the transformation parameters, these priors can be useful in many applications. For instance, in tracking problems where we want to segment a particular frame f, priors can be included using the results obtained in the previous frame $f-1$, to ensure a smooth pose update throughout the frames and also to guide the segmentation when the observations do not provide enough evidence of the location of the object.

Non-informative priors, pioneered by Jeffreys [20], [21], are a class of probability density functions capable of expressing ignorance about the model parameters. Rather than describing *a priori* beliefs, they should be viewed as a way of letting the data "dominate", while staying inside a Bayesian framework [22].

For our particular problem, and since the transformation parameters, $\mathbf{a} = (a_1, a_2)^\top$ and $\mathbf{t} = (t_1, t_2)^\top$, can be assumed to be independent, the prior can be factorized as $p(\mathbf{a})p(\mathbf{t})$. The Jeffreys' priors for parameters $\mathbf{a}$ and $\mathbf{t}$ are given by (see Appendix A for details)

$$p(\mathbf{a}) \propto \sqrt{|\mathcal{I}(\mathbf{a})|} = c_{\mathbf{a}} \tag{24}$$
$$p(\mathbf{t}) \propto \sqrt{|\mathcal{I}(\mathbf{t})|} = c_{\mathbf{t}}. \tag{25}$$

where $\mathcal{I}(\cdot)$ is the Fisher information matrix, and $c_{\mathbf{a}}$ and $c_{\mathbf{t}}$ are constants that do not depend on the parameters $\mathbf{a}$ and $\mathbf{t}$, which means $p(\mathbf{a})$ and $p(\mathbf{t})$ are both uniformly distributed in $\mathbb{R}^2$.

For the sensor probabilities, we also adopt the standard non-informative Jeffreys' prior,

$$p(\mathbf{p}) \propto \sqrt{|\mathcal{I}(\mathbf{p})|} = c_{\mathbf{p}} \tag{26}$$

where $c_{\mathbf{p}}$ is a constant that does not depend on $\mathbf{p}$.

Replacing the priors (23)-(26) in (22) yields

$$\begin{aligned}\mathcal{P}(\mathbf{Y}, \mathbf{K}, \boldsymbol{\Theta}) &= \log p\,(\mathbf{Y}|\mathbf{K}, \boldsymbol{\Theta}) + \log p(\mathbf{K}) \\ &\quad + \log c_{\mathbf{a}} + \log c_{\mathbf{t}} + \log c_{\mathbf{p}} + \log c_{\mathbf{b}} \\ &\quad - \frac{1}{2}\mathbf{b}^\top {\Sigma^{\mathbf{b}}}^{-1}\mathbf{b},\end{aligned} \tag{27}$$

Assuming independence between the observations, we can write

$$\begin{aligned}\mathcal{P}(\mathbf{Y}, \mathbf{K}, \boldsymbol{\Theta}) &= \sum_{i=1}^{N}\sum_{j=1}^{M^i} \log p\left(\mathbf{y}^{ij}\middle|k^{ij}, \boldsymbol{\Theta}\right) + \log p\left(k^{ij}\right) \\ &\quad + C - \frac{1}{2}\mathbf{b}^\top {\Sigma^{\mathbf{b}}}^{-1}\mathbf{b},\end{aligned} \tag{28}$$

where $C = \log c_{\mathbf{a}} + \log c_{\mathbf{t}} + \log c_{\mathbf{p}} + \log c_{\mathbf{b}}$ is a constant.

E-Step: Given an estimate of the model parameters, $\widehat{\boldsymbol{\Theta}}_{(t)}$, and a set of observation points, $\mathbf{Y}$, the expectation of (28) is

$$\begin{aligned}Q_{\mathrm{MAP}}\left(\boldsymbol{\Theta}; \widehat{\boldsymbol{\Theta}}_{(t)}\right) &= \mathbb{E}_{\mathbf{K}}\left[\mathcal{P}(\mathbf{Y}, \mathbf{K}, \boldsymbol{\Theta})\middle|\mathbf{Y}, \widehat{\boldsymbol{\Theta}}_{(t)}\right] \\ &= \mathbb{E}_{\mathbf{K}}\left[\sum_{i=1}^{N}\sum_{j=1}^{M^i} \log p\left(\mathbf{y}^{ij}\middle|k^{ij}, \boldsymbol{\Theta}\right) + \log p\left(k^{ij}\right)\right] \\ &\quad + C - \frac{1}{2}\mathbf{b}^\top {\Sigma^{\mathbf{b}}}^{-1}\mathbf{b},\end{aligned} \tag{29}$$

where the first term in the expectation is given in (11). Therefore,

$$Q_{\text{MAP}}\left(\boldsymbol{\Theta};\widehat{\boldsymbol{\Theta}}_{(t)}\right)=Q_{\text{ML}}\left(\boldsymbol{\Theta};\widehat{\boldsymbol{\Theta}}_{(t)}\right)+C-\frac{1}{2}\mathbf{b}^{\top}\Sigma^{\mathbf{b}^{-1}}\mathbf{b}. \quad (30)$$

The constant C can be discarded since it does not change the outcome of the maximization step. On the other hand, the last term will affect the output of the deformation parameters estimation and it appears "naturally" in this formulation due to the prior for $\mathbf{b}$.

M-Step: The maximization of $Q_{\text{MAP}}\left(\boldsymbol{\Theta};\widehat{\boldsymbol{\Theta}}_{(t)}\right)$ in (30) with respect to the model parameters is similar to the ML estimation. As previously mentioned, because we use non-informative priors for the transformation parameters and sensor probabilities, the only difference to the ML estimation is in the case of the deformation parameters $\mathbf{b}$. The updates will be detailed in the following steps.

1) Update of the Transformation Parameters: Since the prior terms in (30) do not depend on the transformation parameters, $(\mathbf{a},\mathbf{t})$, these parameters are updated by solving (17), as in the previous section.

2) Update of the Deformation Parameters: Maximizing (30) with respect to the deformation parameters, $\mathbf{b}$, yields an additional term, due to the Gaussian prior in (23). The update is achieved by solving

$$\left[\Sigma^{\mathbf{b}^{-1}}+\sum_{i=1}^{N}\sum_{j=1}^{M^i}w_1^{ij}{\mathbf{D}^i}^{\top}\widehat{\mathbf{A}}_{(t+1)}^{\top}{\Sigma^i}^{-1}\widehat{\mathbf{A}}_{(t+1)}\mathbf{D}^i\right]\widehat{\mathbf{b}}_{(t+1)}$$
$$=\left[\sum_{i=1}^{N}\sum_{j=1}^{M^i}w_1^{ij}{\mathbf{D}^i}^{\top}\widehat{\mathbf{A}}^{\top}{}_{(t+1)}{\Sigma^i}^{-1}\left(\mathbf{y}^{ij}-\widehat{\mathbf{A}}_{(t+1)}\bar{\mathbf{x}}^i-\widehat{\mathbf{t}}_{(t+1)}\right)\right], \quad (31)$$

where the term $\Sigma^{\mathbf{b}^{-1}}$ "pulls" the coefficients of $\widehat{\mathbf{b}}_{(t+1)}$ closer to zero. This means that the algorithm will try to keep the deformation coefficients low, which guarantees that shape of the segmentation is similar to the shapes in the training set.

3) Update of the Model Probabilities: The derivative of (30) with respect to the model probabilities, p_0 and p_1, yields the same results as in the ML estimation case, which means the model probabilities are updated using (21).

C. Relation With Karush-Kuhn-Tucker (KKT) Conditions

The KKT conditions can be used to add inequality constraints to an unconstrained optimization problem. This is a generalization of the Lagrange multipliers, which allow adding equality constraints to the solution.

Consider the update of the deformation parameters in the ML approach. The additional step in (20) aims at finding the estimate $\widehat{\mathbf{b}}$ that satisfies the condition in (19). This can be viewed as an inequality constraint on the solution of the optimization problem in (14). Formally, we wish to find

$$\widehat{\mathbf{b}}_{(t+1)}=\begin{array}{ll}\arg\max\limits_{\mathbf{b}} & Q_{\text{ML}}\left(\boldsymbol{\Theta};\widehat{\boldsymbol{\Theta}}_{(t)}\right)\\ \text{subject to} & \mathbf{b}^{\top}\Sigma^{\mathbf{b}^{-1}}\mathbf{b}\le d_{\max}^2.\end{array} \quad (32)$$

The KKT conditions allow us to obtain the solution to this problem by finding the value of $\mathbf{b}$ that satisfies the following equation

$$\nabla Q_{\text{ML}}\left(\boldsymbol{\Theta};\widehat{\boldsymbol{\Theta}}_{(t)}\right)-\mu\nabla_{\mathbf{b}}g(\mathbf{b})=0 \quad (33)$$

where $\mu\ge 0$ is a constant - a KKT multiplier, and $g(\mathbf{b})=\mathbf{b}^{\top}\Sigma^{\mathbf{b}^{-1}}\mathbf{b}-d_{\max}^2$ is the inequality constraint in (19). This leads to the solution

$$\left[\mu\Sigma^{\mathbf{b}^{-1}}+\sum_{i=1}^{N}\sum_{j=1}^{M^i}w_1^{ij}\mu{\mathbf{D}^i}^{\top}\widehat{\mathbf{A}}_{(t+1)}^{\top}{\Sigma^i}^{-1}\widehat{\mathbf{A}}_{(t+1)}\mathbf{D}^i\right]\widehat{\mathbf{b}}_{(t+1)}$$
$$=\left[\sum_{i=1}^{N}\sum_{j=1}^{M^i}w_1^{ij}{\mathbf{D}^i}^{\top}\widehat{\mathbf{A}}_{(t+1)}^{\top}{\Sigma^i}^{-1}\left(\mathbf{y}^{ij}-\widehat{\mathbf{A}}_{(t+1)}\bar{\mathbf{x}}^i-\widehat{\mathbf{t}}_{(t+1)}\right)\right]. \quad (34)$$

This means that the MAP estimation of the deformation parameters in (31) is a particular case of (34), for $\mu=1$.

D. Comparison Between the ML and the MAP Approaches

In this work, the difference between the ML and the MAP approaches lies in the estimation of the deformation parameters. The ML formulation uses an additional step (20) that shrinks the estimate of the deformation parameters in order to guarantee that the obtained shape is acceptable [2] (i.e., not over-deformed). This is required because the objective function $Q_{\text{ML}}\left(\boldsymbol{\Theta};\widehat{\boldsymbol{\Theta}}_{(t)}\right)$ defined in (11), does not contain any (prior) information about the expected estimates of the deformation parameters. Without that constraint the resulting estimates could lead to shapes significantly different from those found in the training set. On the other hand, in the MAP formulation, the objective function $Q_{\text{MAP}}\left(\boldsymbol{\Theta};\widehat{\boldsymbol{\Theta}}_{(t)}\right)$, defined in (30), includes a constraint on the deformation parameters due to the prior $p(\mathbf{b})$. This constraint is obtained in a more natural and principled way and "pushes" the coefficients of $\mathbf{b}$ to zero, thus inducing a shrinkage effect similar to the one in (20). Furthermore, in the ML formulation, the shrinkage step in (20) scales all the components of $\mathbf{b}$ uniformly, disregarding the importance of each mode of variation, whereas in the MAP estimation the shrinkage effect is weighted by the covariance matrix $\Sigma^{\mathbf{b}}$ (recall (31)), which means it depends on the eigenvalue of each deformation mode.

V. Algorithm Overview

The algorithm described in the previous sections can be summarized as follows. Given initial estimates of $\widehat{\boldsymbol{\Theta}}_{(t=0)}$, the shape model parameters are iteratively updated by:

1) *Searching for observation points* in the vicinity of the model;
2) *Computing the observation probabilities* w_1^{ij} and w_0^{ij} using (12) and (13) (E-step);
3) *Updating the transformation parameters* a_1, a_2, t_1, t_2 using (17) (M-step);
4) *Updating the deformation parameters* $\mathbf{b}$ using (18) and (20) in the ML case, or (31) in the MAP case (M-step); and
5) *Updating the model probabilities* p_0 and p_1 using (21) (M-step).

This process is repeated until no significant changes are observed in the contour. This algorithm will be denoted as Expectation-Maximization Robust Active Shape Model (EM-RASM).

VI. Experimental Setup

The evaluation of the EM-RASM method is divided in two types of experiments: 1) using synthetic data, which illustrates the different behavior of the proposed method compared to the standard ASM used, see [2], [5], [23]–[26]; and 2) using real images for two different applications.

The experiments using synthetic data provide evidence of the superiority of the EM-RASM over the standard ASM using two example images.

Regarding the experiments using real images, the first application is the segmentation of the left ventricle (LV) in ultrasound images. This is a complex problem due to the presence of multiplicative noise in the images, which lead to the detection of many outliers. The second application is the segmentation of the lip in face images from the Cohn-Kanade (CK+) database [27]. In the latter, the accuracy of the proposed method is compared to two other state of the art methods: the Adaptive Snakes (AS), proposed in [16]; and the combination of ASM with RANSAC, proposed in [1]. The first method (AS) also uses the EM algorithm to deal with outliers, but the contour of the object is free, meaning that neither *a priori* shape information nor training is provided to the Snake model. This means that the method is a purely bottom-up approach, and that the segmentation may have an unexpected shape. The second method uses an ASM to describe the contour and the RANSAC method to determine which observation points are outliers.

The proposed method requires an initial guess of the model parameters. In all the tests performed, the model was initialized with the average shape $\bar{\mathbf{x}}$ (i.e., $\mathbf{b} = \mathbf{0}$). The initial guess for the transformation parameters was obtained by aligning the average shape $\bar{\mathbf{x}}$ with a contour obtained by human input using standard least squares method. In all the tests showed in the next section, the initial guess for the models probabilities was $p_0 = p_1 = 0.5$. We found no evidence suggesting that the initial values for these probabilities significantly changed the output of the algorithm.

The segmentations were evaluated by comparing the obtained contours with the true object boundary (ground truth). The accuracy of the segmentations were quantitatively determined using the Dice coefficient [28].

VII. Results

A. Synthetic Images

The performance of the EM-RASM method was evaluated in two synthetic images. These images consist of two different corrupted versions of a binary image of a rectangle, with intensity value 1 inside the rectangle and 0 outside. In the first example, the binary image was corrupted by white Gaussian noise with zero mean and variance $\sigma^2_{\text{noise}} = 0.5$ (see Fig. 2 on the top left). In the second example, the rectangle image was corrupted by black regions (see Fig. 2 on the top right).

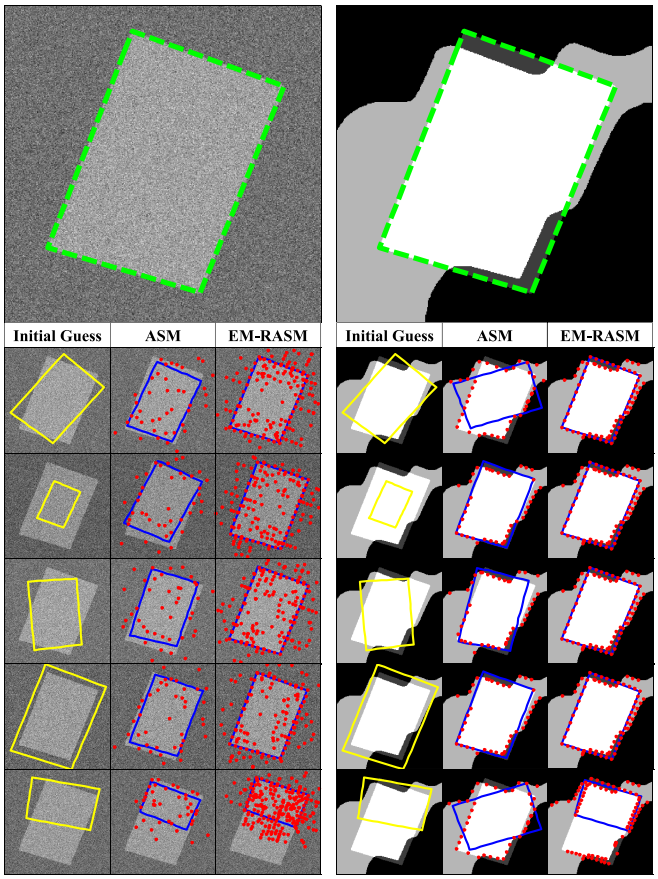


Fig. 2. Segmentation of a rectangle image: (top left) corrupted by white Gaussian noise with variance $\sigma^2_{\text{noise}} = 0.5$; and (top right) corrupted by black regions. The ground truth segmentation is shown in dashed green. Each row shows an example using a different initial guess (left column) and the final segmentation using the standard ASM (middle column) and the EM-RASM (right column). The red dots correspond to the detected observation points in the last iteration.

We wish to retrieve the correct location of the rectangle given an initial guess. However, in the corrupted versions of the image, additional observation points will be detected on edges that do not belong to the rectangle boundary, which means the set of observation points will be noisy. In these examples, the results using the ML and MAP estimation methods were very similar. For the sake of clarity in the presentation, we show only the results using the ML method.

The shape model was learned using synthetic data generated by adding random Gaussian perturbations to the true object boundary $\bar{\mathbf{x}}$. More specifically, each training example $\mathbf{x} \in \mathbb{R}^{2N\times 1}$ is a realization of

$$\mathbf{x} = \bar{\mathbf{x}} + e, \quad e \sim \mathcal{N}(\mathbf{0}, \sigma^2_{\text{train}}\mathbf{I}) \tag{35}$$

where $\sigma_{\text{train}} = 2$ is the standard deviation imposed on the synthetic model points. Since the Gaussian noise is isotropic and identical for all model points, we do not expect the shape model to identify any significant modes of variation.

Fig. 2 shows the output of the proposed algorithm and the output of the standard ASM for different initial guesses. In the first example, the standard ASM was unable to cope with the outliers and, consequently, it did not fit the rectangle boundary accurately. The EM-RASM, on the other hand, was able to accurately segment the rectangle despite the detection

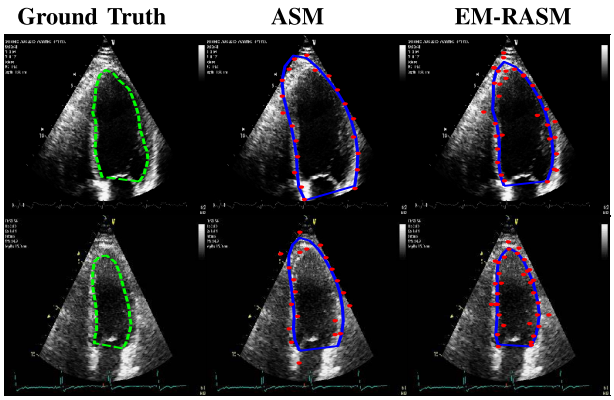


Fig. 3. Segmentation of the left ventricle in ultrasound images. Each row shows one example, where the green dashed line shows the ground truth (left column); the blue lines correspond to the estimated segmentation using the standard ASM (mid column) and the proposed algorithm (ML) (right column); and the red dots represent the detected observations in the last iteration.

of many outliers (note the large number of red dots in the images that do not lie along the rectangle edges). In the second synthetic example, the gradient associated to the edges of the black regions is stronger than the one associated to the edges of the rectangle. Consequently, some of the detected observation points are outliers. The figure shows that the presence of these outliers causes the standard ASM to be unable to correctly segment the image, since it tries to fit all the observation points simultaneously (including outliers). The EM-RASM, on the other hand, is able to accurately segment the rectangle and disregard the outliers. The last row in the figure shows an example in which neither the standard ASM nor the EM-RASM were able to segment the rectangle. This was due to a poor initialization of the model parameters, as shown by the initial guess on the left column.

B. Real Images - Left Ventricle Segmentation

We applied the EM-RASM method to the segmentation of the left ventricle in 2D ultrasound image sequences. The dataset is composed of five 2D sequences (five different patients), each with 16-20 frames, for a total of 87 images.

The shape model was trained using manual medical annotations of the left ventricle contours (ground truth). Each training example was obtained by resampling, in arc-length, the medical contours with a fixed number of points from the bottom left to the apex (top) and from the bottom right to the apex. We tested the proposed algorithm and the standard ASM using a leave-one-sequence-out cross validation strategy.

In each test sequence, the initial guess for the transformation parameters was obtained by aligning the average contour $\bar{\mathbf{x}}$ with a contour obtained by human input using the standard least squares method. A different human input contour was used in each test sequence, and the resulting initial guess was used in all the frames of the sequence (i.e., we did not propagate the contours from one frame to the next).

Fig. 3 shows two examples with the segmentation obtained with EM-RASM (using the ML approach) and with the standard ASM. Note that a large number of the detected observations (red dots) are outliers. The figure shows that

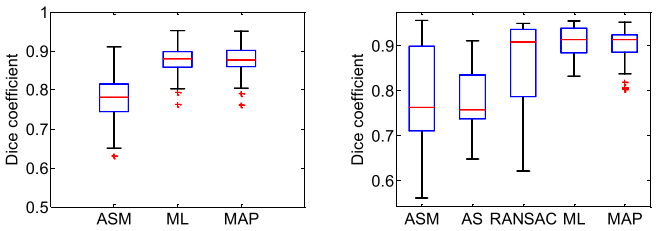


Fig. 4. Statistical results for the segmentation of: (left) the LV in ultrasound images; and (right) the lip in face images and comparison with state of the art methods.

the EM-RASM performed better than the standard ASM and was able to fit the LV boundary, whereas the accuracy of the segmentation obtained using the standard ASM was hampered by the outliers.

Statistical results are presented in Fig. 4 (left). The results show that the ML and the MAP approach perform similarly and that they lead to a significant improvement in accuracy over the standard ASM.

C. Comparison With State of the Art - Lip Segmentation

We will use the example of the lip segmentation problem to compare the performance of the proposed method with two other state of the art methods: the Adaptive Snakes, proposed in [16], and the combination of ASM and RANSAC, proposed in [1]. The first method uses the Snake model to describe the contours and estimates the model parameters using an outlier model and the EM algorithm (similarly to the approach proposed in this work), whereas the second method uses a shape model and estimates the model parameters using the RANSAC method.

The dataset consists of four sequences of face images. These sequences were acquired from the neutral expression samples of the Cohn-Kanade expression database [27], each with 10-20 frames, for a total of 58 images. The training shapes were obtained by resampling, in arc-length, the lip contours provided in the database so that the number of points in the lower lip and the number of points in the upper lip were the same in all the contours. As in the previous case, the shape model was trained using a leave-one-sequence-out cross validation strategy. The initial guess for the transformation parameters for each test sequence was obtained by aligning the average contour $\bar{\mathbf{x}}$ with a contour obtained by human input. The same initial guess was used in all the frames of the test sequence.

Fig. 5 shows two examples of the segmentations obtained with EM-RASM (using the ML approach) and with the standard ASM. It is possible to see that the ability to detect more than one observation point for each search line increases the probability of detecting the true object boundary, particularly if that boundary is not associated with a strong edge (e.g., in Fig. 5 second and third rows). On the other hand, despite the consequent increase in the number of outliers, the proposed approach was able to fit the valid observations. This does not happen in the standard ASM, in which the obtained contour was misguided by the outliers. However, if the number of outliers is small, both methods perform similarly.

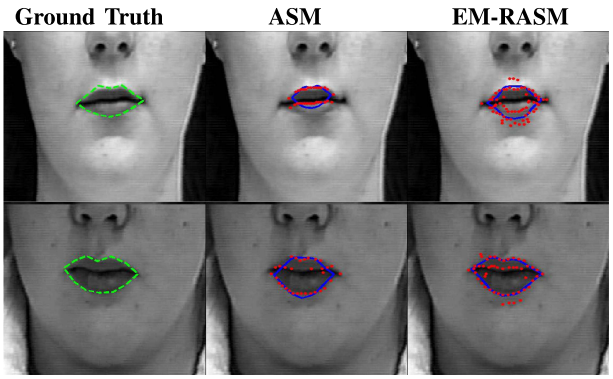


Fig. 5. Segmentation of the lip in face images. Each row shows a different example, where the green dashed line shows the ground truth (left column); the blue lines correspond to the estimated segmentation using the standard ASM (mid column) and the proposed algorithm (ML) (right column); and the red dots represent the detected observations in the last iteration.

Fig. 4 (right) shows the statistical evaluation of the segmentation accuracy. As in the previous segmentation problem, the proposed method outperforms the standard ASM. These results also show that the proposed method achieves a better accuracy than the two other state of the art methods. The RANSAC approach proposed in [1] also performs significantly better than the standard ASM and has a performance similar to the proposed method in some sequences. On the other hand, the method proposed in [16] (AS) has a performance similar to the ASM, but note that the AS does not use shape information obtained in a training stage. Finally, it is also possible to see that the ML and the MAP approaches achieve a very similar performance.

VIII. Conclusion

This paper combines active shape models (ASM) with robust estimation of the model pose and deformation using an outlier model. The estimation of the model parameters is achieved using the EM method, which weights each observation point by the probability of that point belonging to the object boundary. We show that this approach is robust in the presence of outliers and is able to overcome this limitation of the standard ASM, both in synthetic and real images.

Future work should focus on extending the proposed framework to more reliable observations. For instance, using other (application-specific) feature detection methods may lead to better results by decreasing the number of outliers detected. Furthermore, edge points along the same image edge often belong to the same object in the image. This can be used to improve the computation of the weights associated to the observations.

Appendix

The Jeffreys' prior [29] is a non-informative prior that can be used to express *ignorance* about some model parameter $\alpha \in \mathbb{R}^n$. The Jeffreys' prior is defined by

$$p(\alpha) \propto \sqrt{\det\left(\mathcal{I}(\alpha)\right)}, \tag{36}$$

where $\mathcal{I}(\alpha)$ is the Fisher information matrix introduced by Fisher [30]. Its best known use is in the Cramer-Rao (lower) bound (see [31], [32], or [33]).

The coefficients of $\mathcal{I}(\alpha)$ are given by

$$\mathcal{I}(\alpha)_{ij} = -\mathbb{E}_{\mathbf{Y}}\left[\frac{\partial^2 \log p\left(\mathbf{Y}|\alpha\right)}{\partial\alpha_i \partial\alpha_j}\right], \tag{37}$$

where $p\left(\mathbf{Y}|\alpha\right)$ is the likelihood of a set of observations $\mathbf{Y}$, given the model parameter α.

Now consider the problem described in Section III. The log-likelihood of a set of observation points $\mathbf{Y}$ is given by[1]

$$\log p\left(\mathbf{Y}|\mathbf{a},\mathbf{t}\right) = \sum_{i=1}^{N}\sum_{j=1}^{M^i} \log \mathcal{N}\left(\mathbf{y}^{ij}; \mathbf{A}\mathbf{x}^i + \mathbf{t}, \Sigma^i\right).$$

After straightforward manipulations, the Fisher information matrix for each of the transformation parameters, $\mathbf{a}$ and $\mathbf{t}$, is given by

$$\mathcal{I}(\mathbf{a}) = -\sum_{i=1}^{N}\sum_{j=1}^{M^i} {\mathbf{X}^i}^\top {\Sigma^i}^{-1}\mathbf{X}^i \tag{38}$$

$$\mathcal{I}(\mathbf{t}) = -\sum_{i=1}^{N}\sum_{j=1}^{M^i} {\Sigma^i}^{-1}. \tag{39}$$

This leads to the following Jeffreys' priors

$$p(\mathbf{a}) \propto \sqrt{|\mathcal{I}(\mathbf{a})|} = \sqrt{\left|-\sum_{i=1}^{N}\sum_{j=1}^{M^i} {\mathbf{X}^i}^\top {\Sigma^i}^{-1}\mathbf{X}^i\right|} = c_a \tag{40}$$

$$p(\mathbf{t}) \propto \sqrt{|\mathcal{I}(\mathbf{t})|} = \sqrt{\left|-\sum_{i=1}^{N}\sum_{j=1}^{M^i} {\Sigma^i}^{-1}\right|} = c_t, \tag{41}$$

which do not depend on the transformation parameters $\mathbf{a}$ and $\mathbf{t}$.

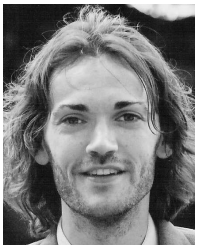
References

[1] M. Rogers and J. Graham, "Robust active shape model search," in *Computer Vision—ECCV*. Berlin, Germany: Springer-Verlag, 2006, pp. 517–530.
[2] T. F. Cootes, C. J. Taylor, D. H. Cooper, and J. Graham, "Active shape models—Their training and application," *Comput. Vis. Image Understand.*, vol. 61, no. 1, pp. 38–59, Jan. 1995.
[3] F. R. Hampel, E. M. Ronchetti, P. J. Rousseeuw, and W. A. Stahel, *Robust Statistics: The Approach Based on Influence Functions*, vol. 114. New York, NY, USA: Wiley, 2011.
[4] T. F. Cootes, G. J. Edwards, and C. J. Taylor, "Active appearance models," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 23, no. 6, pp. 681–685, Jun. 2001.
[5] B. van Ginneken, A. F. Frangi, J. J. Staal, B. M. ter Haar Romeny, and M. A. Viergever, "Active shape model segmentation with optimal features," *IEEE Trans. Med. Imag.*, vol. 21, no. 8, pp. 924–933, Aug. 2002.
[6] M. Wimmer, K. Stulp, S. Pietzsch, and B. Radig, "Learning local objective functions for robust face model fitting," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 30, no. 8, pp. 1357–1370, Aug. 2008.
[7] D. Cristinacce and T. Cootes, "Automatic feature localisation with constrained local models," *Pattern Recognit.*, vol. 41, no. 10, pp. 3054–3067, Oct. 2008.
[8] P. Arbeláez, M. Maire, C. Fowlkes, and J. Malik, "Contour detection and hierarchical image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 5, pp. 898–916, May 2011.
[9] T. F. Cootes, M. C. Ionita, C. Lindner, and P. Sauer, "Robust and accurate shape model fitting using random forest regression voting," in *Computer Vision—ECCV*. Berlin, Germany: Springer-Verlag, 2012, pp. 278–291.

[1]We are only considering the model $k = 1$ from equation (6), since the outlier model does not depend on the parameters $\mathbf{a}$ and $\mathbf{t}$.

[10] Y. Wang and L. H. Staib, "Boundary finding with correspondence using statistical shape models," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, Jun. 1998, pp. 338–345.
[11] J. Abi-Nahed, M.-P. Jolly, and G.-Z. Yang, "Robust active shape models: A robust, generic and simple automatic segmentation tool," in *Medical Image Computing and Computer-Assisted Intervention—MICCAI*. Berlin, Germany: Springer-Verlag, 2006, pp. 1–8.
[12] K. Lekadir, R. Merrifield, and G.-Z. Yang, "Outlier detection and handling for robust 3D active shape models search," *IEEE Trans. Med. Imag.*, vol. 26, no. 2, pp. 212–222, Feb. 2007.
[13] H. Chui and A. Rangarajan, "A new point matching algorithm for non-rigid registration," *Comput. Vis. Image Understand.*, vol. 89, nos. 2–3, pp. 114–141, Feb. 2003.
[14] M. A. Fischler and R. C. Bolles, "Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography," *Commun. ACM*, vol. 24, no. 6, pp. 381–395, Jun. 1981.
[15] P. J. Huber, *Robust Statistics*. New York, NY, USA: Wiley, 1981.
[16] J. C. Nascimento and J. S. Marques, "Adaptive snakes using the EM algorithm," *IEEE Trans. Image Process.*, vol. 14, no. 11, pp. 1678–1686, Nov. 2005.
[17] M. Kass, A. Witkin, and D. Terzopoulos, "Snakes: Active contour models," *Int. J. Comput. Vis.*, vol. 1, no. 4, pp. 321–331, Jan. 1988.
[18] A. Blake and M. Isard, "Image processing techniques for feature location," in *Active Contours*. New York, NY, USA: Springer-Verlag, 1998, pp. 97–113.
[19] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *J. Roy. Statist. Soc. B, Methodol.*, vol. 39, no. 1, pp. 1–38, 1977.
[20] H. Jeffreys, "An invariant form for the prior probability in estimation problems," *Proc. Roy. Soc. London A, Math. Phys. Sci.*, vol. 186, no. 1007, pp. 453–461, Sep. 1946.
[21] H. Jeffreys, *The Theory of Probability*. Oxford, U.K.: Oxford Univ. Press, 1961.
[22] T. Z. Irony and N. D. Singpurwalla, "Non-informative priors do not exist a dialogue with José M. Bernardo," *J. Statist. Planning Inference*, vol. 65, no. 1, pp. 159–177, Dec. 1997.
[23] T. F. Cootes and C. J. Taylor, "A mixture model for representing shape variation," *Image Vis. Comput.*, vol. 17, no. 8, pp. 567–573, Jun. 1999.
[24] G. Behiels, D. Vandermeulen, F. Maes, P. Suetens, and P. Dewaele, "Active shape model-based segmentation of digital X-ray images," in *Medical Image Computing and Computer-Assisted Intervention—MICCAI*. Berlin, Germany: Springer-Verlag, 1999, pp. 128–137.
[25] G. Hamarneh and T. Gustavsson, "Combining snakes and active shape models for segmenting the human left ventricle in echocardiographic images," in *Proc. Comput. Cardiol.*, Sep. 2000, pp. 115–118.
[26] S. Milborrow and F. Nicolls, "Locating facial features with an extended active shape model," in *Computer Vision—ECCV*. Berlin, Germany: Springer-Verlag, 2008, pp. 504–513.
[27] P. Lucey, J. F. Cohn, T. Kanade, J. Saragih, Z. Ambadar, and I. Matthews, "The extended Cohn–Kanade dataset (CK+): A complete dataset for action unit and emotion-specified expression," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2010, pp. 94–101.
[28] L. R. Dice, "Measures of the amount of ecologic association between species," *Ecology*, vol. 26, no. 3, pp. 297–302, Jul. 1945.
[29] H. Jeffreys, *The Theory of Probability*. New York, NY, USA: Oxford Univ. Press, 1998.
[30] R. A. Fisher, "On the mathematical foundations of theoretical statistics," *Philosoph. Trans. Roy. Soc. London A*, vol. 222, nos. 594–604, pp. 309–368, 1922.
[31] S. Kay, *Fundamentals of Statistical Signal Processing*, vol. 1. Englewood Cliffs, NJ, USA: Prentice-Hall, 1993.
[32] L. L. Scharf, *Statistical Signal Processing*, vol. 98. Reading, MA, USA: Addison-Wesley, 1991.
[33] H. L. Van Trees, *Estimation and Modulation Theory*, vol. 1. 1968.

Carlos Santiago received the Biomedical Engineering and M.Sc. degrees from the Technical University of Lisbon, Portugal, in 2009 and 2011, respectively. He is currently pursuing the Ph.D. degree with the Instituto Superior Técnico, Lisbon, Portugal, where he is also with the Institute for Systems and Robotics. His research interests include statistical image processing, shape analysis, and medical image analysis.



Jacinto C. Nascimento (S'00–M'06) received the Electrical Engineering degree from the Instituto Superior de Engenharia de Lisboa, in 1995, and the M.Sc. and Ph.D. degrees from the Instituto Superior Técnico (IST), Technical University of Lisbon, in 1998 and 2003, respectively. He is currently an Assistant Professor with the Informatics and Computer Engineering Department, IST, Lisbon, and a Researcher with the Institute for Systems and Robotics. He has published 30 publications in international journals and 80 in conference proceedings (many of which of the IEEE), has served on the Program Committee of many international conferences, and has been a reviewer for several international journals. His research interests include statistical image processing, pattern recognition, machine learning, medical imaging analysis, video surveillance, and general visual object classification.

Jorge S. Marques received the Electrical Engineering and Ph.D. degrees, and the Aggregation Title from the Technical University of Lisbon, Portugal, in 1981, 1990, and 2002, respectively. He is currently an Associate Professor with the Electrical and Computer Engineering Department, Instituto Superior Técnico, Lisbon, and a Researcher with the Institute for Systems and Robotics. He has authored over 150 papers in international journals and conferences and also authored the book *Pattern Recognition: Statistical and Neural Methods* (IST Press, 2005, 2nd edition, Portuguese). He was the Co-Chairman of the IAPR Conference IbPRIA 2005, the President of the Portuguese Association for Pattern Recognition from 2001 to 2003, and an Associate Editor of the *Statistics and Computing Journal* (Springer). His research interests are in the areas of statistical image processing, shape analysis, and pattern recognition.