Robust Feature Descriptors For Object Segmentation Using Active Shape Models

Daniela Medley, Carlos Santiago and Jacinto C. Nascimento

Institute for Systems and Robotics, Instituto Superior Técnico, Lisbon, Portugal[*]

Abstract. Object segmentation is still an active topic that is highly visited in image processing and computer vision communities. This task is challenging due not only to difficult image conditions (e.g., poor resolution or contrast), but also to objects whose appearance vary significantly. This paper visits the Active Shape Model (ASM) that has become a widely used deformable model for object segmentation in images. Since the success of this model depends on its ability to locate the object, many detectors have been proposed. Here, we propose a new methodology in which the ASM search takes the form of local rectangular regions sampled around each landmark point. These regions are then correlated to variable or fixed texture templates learned over a training set. We compare the performance of the proposed approach against other detectors based on: (i) the classical ASM edge detection; (ii) the Histogram of Oriented Gradients (HOG); and (iii) the Scale-Invariant Feature Transform (SIFT). The evaluation is performed in two different applications: facial fitting and segmentation of the left ventricle (LV) in cardiac magnetic resonance (CMR) images, showing that the proposed method leads to a significant increase in accuracy and outperforms the other approaches.

Keywords: Active Shape Models, Image segmentation, Texture regions, Histogram of Oriented Gradients, Scale-Invariant Feature Transform

1 Introduction

Segmenting images containing objects whose appearance vary significantly is a challenging task. Statistical models have become a widely used approach in this context. These models are able to represent large shape and appearance variations of the object of interest. A popular method is the Active Shape Model (ASM). Since its early introduction by Cootes et al. [1], ASM has become a well-recognized powerful tool due to its ability to segment objects with significant shape variability. This method characterizes an object shape by a set of specific points, denoted as landmark points, and models it by a mean shape and its most significant modes of variation learned from a training set. To fit the model to an object, the ASM searches within the image for candidate positions of the object

[*] This work was supported by FCT UID/EEA/50009/2013

landmarks, which we call observations. Traditionally, this search consists of finding the strongest edge along a profile line for each model point. Although this approach may work in some simple applications, in most real world problems it is a naive approach that might fail to cover all the object features, generating noisy observations (outliers). If some of the observations are outliers, the accuracy of the ASM is severely compromised, resulting in a decreased segmentation performance. Therefore, a crucial component for the success of these models lies in their ability to find the correct position of the object landmarks.

Alternative approaches have since been proposed to overcome this ASM drawback. The Active Appearance Model (AAM), also proposed by Cootes *et al.* [2], is not only able to provide shape information about an object, but also takes into account its variation in appearance, i.e., its textural information. Contrary to ASM, the AAM search method consists of using the texture residual between the learned model and the test image in order to find the best model parameters to match the image. In later work, Cristinacce and Cootes [3] presented the Constrained Local Model (CLM). This method is very similar to AAM, but instead of modelling the whole object texture, it models local templates surrounding each landmark point. It uses Principal Component Analysis (PCA) to learn statistical templates of the appearance from a training set, which are adjustable for each test image.

Powerful feature descriptors have also been combined with these deformable models, in order to improve the observation detection, namely, Histograms of Oriented Gradients (HOG) [4] and Scale-Invariant Feature Transform (SIFT) [5, 6]. These models have shown an increased performance in comparison to the standard model approach. Nevertheless, these have an increased complexity and appear to be slower and computational demanding.

This paper proposes two different and efficient observation detection methods that are used within the ASM framework. Both approaches are similar in the sense that they search for observations within a rectangular region around each landmark point and both find the point that maximizes the correlation with a texture template learned from a training data. They differ on how the templates are obtained. In the first method, each template, associated to a landmark point, is computed as the mean texture in the training data, obtaining a *fixed template* that will be the same for the testing stage. The second method uses not only the mean texture, but also the variation modes obtained through PCA. This allows fitting the templates for each test image, i.e, the method uses *variable templates*. This approach resembles the CLM algorithm but uses a different fitting strategy. More specifically, the CLM uses a whole response surface to update the model, whilst we determine the templates around each landmark point and extract the location of best observation point, which is then used to estimate the ASM parameters.

Despite of been easily used for different image interpretations, to show the advantage of the proposed methods we compare them to three other detection approaches in the problem of facial fitting and the segmentation of the left ventricle in cardiac magnetic resonance images: (i) the classical ASM edge detection;

(ii) a detector based on HOG features; and (iii) a detector based on SIFT features.

The remaining of this paper is organized as follows: Sect. 2 describes the ASM framework and each observation detection method used in this work. The datasets and evaluation results are shown in Sect. 3. Finally, in Sect. 4 the overall conclusions are presented.

2 Methodology

This section starts by briefly revising the ASM framework used in this work. In Sect. 2.2, we detail all the different detection methods analysed and how they are combined with the ASM methodology.

2.1 Active Shape Model

The ASM algorithm [1] describes the shape of an object by learning its statistics from annotated images in a training set. More specifically, this model uses the mean shape and its main modes of deformation computed using PCA. Formally, any shape $\mathbf{x}$ in the ASM framework can be analytically described as

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

where, $\bar{\mathbf{x}} \in \mathbb{R}^{2N \times 1}$ is a vector representing the mean shape computed from a training set, $\mathbf{D} \in \mathbb{R}^{2N \times K}$ is a matrix containing the first K modes of deformation, and $\mathbf{b} \in \mathbb{R}^{K \times 1}$ contains the deformation coefficients that weight each of the deformation modes. Then, the position of the shape in an image is governed by a similarity transformation, which accounts for the scale, rotation and translation of the shape $\mathbf{x}$.

To fit the model to the object in a new image, the ASM searches for observation points and estimates the parameters that minimize the distance between the model points and the corresponding observations. In the next section, we describe five ASM search methods: the classical ASM edge detection, the ASM-HOG, the ASM-SIFT and the proposed fixed and variable texture template methods, denoted as ASM-FTT and ASM-VTT, respectively (see Fig. 1).

2.2 ASM Search Methods

Traditionally, in the ASM framework the observations correspond to the strongest edge along profile lines orthogonal to the contour at each model point (see Fig. 1 a)). In most real world problems, this approach generates outliers that misguide the estimation of the ASM parameters. Thus, alternative observation detectors have the potential to improve the ASM performance.

All the four search methods described next share a common framework as they are region-based and use a template descriptor to search for observation points. The first stage of the detectors is to obtain a set of template descriptors and regions in which features are detected. To accomplish this, each training

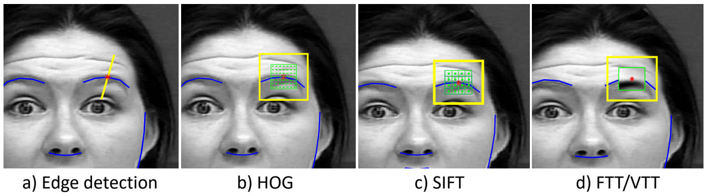

Fig. 1: Representation of the different search methods for one landmark point (red) of the model (blue), where the search region is represented in yellow: a) the original ASM edge detection; b) the HOG-based detection with the descriptor of the landmark in green; c) the SIFT-based detection with the descriptor of the landmark in green; and d) the proposed search methods based on texture (FTT and VTT).

patch is sampled around each landmark point over the training set. A feature extraction function $\mathcal{F}$ is applied to each patch, resulting in a feature vector for each pixel location. Thus, an average descriptor-based template for each landmark can then be built and normalized (e.g. zero mean and unit variance).

Formally, let the training set of images be defined as $\mathcal{D} = \{\mathbf{I}_j\}_{j=1}^{|\mathcal{D}|}$. Assuming the contour is defined as $\mathbf{x} = [\mathbf{x}_i^\top]_{i=1,...,N}$ with N landmark points, i.e., $\mathbf{x} \in \mathbb{R}^{2N\times 1}$, we extract N patches for each j-th image $\mathbf{I}_j$, building the following set of patches

$$\mathcal{P}_j = \left\{\mathbf{P}_j^1, ..., \mathbf{P}_j^N\right\}_{j=1}^{|\mathcal{D}|} \ , \tag{2}$$

where $\mathbf{P}_j^i$ stands for the i-th patch extracted from the j-th image and is given by $\mathbf{P}_j^i = [\mathbf{I}_j]_{T\times T}^i$. The operator $[.]_{T\times T}^i$ crops the image $\mathbf{I}_j$, centered at the landmark point $\mathbf{x}_i$, with size of $T \times T$ pixels. Now, for each pixel location in each patch $\mathbf{P}_j^i$, we apply a feature extraction function $\mathcal{F}$, obtaining a descriptor-based image $\mathcal{F}(\mathbf{P}_j^i)$.

For each i-th landmark point, the feature based template $\mathcal{T}^i$ of size $T \times T$ can be computed as follows

$$\mathcal{T}^i = \frac{1}{|\mathcal{D}|}\sum_{j=1}^{|\mathcal{D}|}\mathcal{F}(\mathbf{P}_j^i), \quad i = 1, ..., N \ . \tag{3}$$

When searching in a test image $\mathbf{I}_j$, a larger search region, with $R \times R$ dimension (with $R > T$), is created around each i-th model point, *i.e.* $\mathbf{R}_j^i = [\mathbf{I}_j]_{R\times R}^i$. As previously, we extract features from this search region, which we denote as $\mathcal{R}_j^i = \mathcal{F}(\mathbf{R}_j^i)$. By defining a sliding window within the region $\mathcal{R}_j^i$, it is possible to compute the similarity between $\mathcal{T}^i$ and each sub-region of $\mathcal{R}_j^i$ (with $T \times T$ dimension). In the following sections we describe the different methodologies under the above framework.

ASM-HOG Histogram of oriented gradients [7] is a feature descriptor that characterizes the local appearance of the object based on the gradient orientations in different bins of small portions of an input image (see Fig. 1 b)).

To combine this feature descriptor with an ASM search, we define $\mathcal{F}$ in (3) as a HOG feature extractor applied to each landmark point and use the following approach. Given a training set, we can compute a HOG feature vector for each landmark point and an average template $\mathcal{T}^i$ as given in (3). When searching for observation points in the test image, we define the search region $\mathcal{R}_j^i$, sampled around the i-th landmark point, and determine the HOG feature vector for each pixel within this region, through the sliding window. Finally, the resulting HOG features for each pixel in the search region are compared with the corresponding template, $\mathcal{T}^i$. The most similar region will correspond to the new likely location of the object landmarks, *i.e.*, the detected observation. Since HOG feature vectors are histograms, several similarity measures can be applied. Histogram Intersection is one of those metrics that has shown a better performance than the standard Euclidean distance in image applications [8]. For grey images it can be defined as follows [9]

$$\vartheta(h_1, h_2) = \sum_{l=1}^{n} \min(h_1(l), h_2(l)) \ , \tag{4}$$

where h_1 and h_2 are the histograms to be compared of the template $\mathcal{T}^i$ and of each sub-region of $\mathcal{R}_j^i$, respectively, with n bins each.

ASM-SIFT Scale-invariant feature transform [10] is also able to extract distinctive features from images based on the local gradients histograms around each landmark and are invariant to scale, illumination and pose. Contrary to HOGs, these histograms are computed with respect to the dominant orientation of the landmark. It starts by scanning an image to identify potential interest points, known as *keypoints*. Next, the dominant orientation of the keypoint is determined, based on local image gradient direction, and the local descriptor is built.

In order to combine the SIFT descriptor with the ASM algorithm (see Fig. 1 c)), we propose a different strategy for its implementation. For a given training image, we compute a SIFT descriptor by applying a SIFT extraction function $\mathcal{F}$ to each patch (see (3)), forcing the keypoints to be the landmark points, with a gradient orientation relative to the normal vector of the contour at each model point. This allows the corresponding feature vector to be used in the ASM search for observations. Thus, an average template patch can also be built and normalized for each landmark point, consisting on SIFT descriptors (see (3)).

Similarly to the ASM-HOG, when analysing a new image, the search process starts by defining a search region $\mathcal{R}_j^i$ around each current landmark point $\mathbf{x}_i$ and determine the SIFT descriptors for each pixel in this region, with a sliding window. Since the SIFT descriptor is also based on histograms, we use the same comparison metric as before, which is defined in (4).

Proposed Search Using Fixed Texture Templates (ASM-FTT) In this method we also follow the procedure addressed in Sect. 2.2. First, for each j-th image in the training set, we extract the set of patches $\mathcal{P}_j$ (see (2)). Then for each landmark point $\mathbf{x}_i$, $i = 1, \ldots, N$, we compute the corresponding $T \times T$ fixed template $\mathcal{T}^i$ following (3), where the feature extraction function $\mathcal{F}$ concerns to the image texture itself (see Fig. 1 d)). Finally, for a given test image, we set the $R \times R$ region $\mathcal{R}_j^i$ in which we compute a (normalized) cross-correlation between the template $\mathcal{T}^i$ and the search region $\mathcal{R}_j^i$ as follows[1] [11]

$$\gamma(u,v) = \frac{\sum_{x,y}[\mathcal{R}(x,y) - \overline{\mathcal{R}}_{u,v}][\mathcal{T}(x-u,y-v) - \overline{\mathcal{T}}]}{\{\sum_{x,y}[\mathcal{R}(x,y) - \overline{\mathcal{R}}_{u,v}]^2 \sum_{x,y}[\mathcal{T}(x-u,y-v) - \overline{\mathcal{T}}]^2\}^{0.5}} \ , \tag{5}$$

where the notation $\mathcal{R}(x,y)$ stands for the region pixels (x,y) in the region $\mathcal{R}_j^i$; $\overline{\mathcal{T}}$ is the mean of the template $\mathcal{T}$ and $\overline{\mathcal{R}}_{u,v}$ is the mean of the region $\mathcal{R}(x,y)$ under the template.

Therefore, by performing a normalized cross-correlation between the individual training patch $\mathcal{T}^i$ and the search region $\mathcal{R}_j^i$ for each current i-th landmark point, it is possible to determine the new likely location of the object landmarks. This can be determined by analysing the resulting image response, in order to find the strongest match.

Proposed Search Using Variable Texture Templates (ASM-VTT) The search method proposed in this section is similar to the previous one. However, instead of using fixed feature templates, which remain unchangeable during the search process, we describe a method that adapts the texture templates to the test image. More specifically, we apply PCA to the training patches $\mathbf{P}_j^i$, for $j = 1, ..., |\mathcal{D}|$. Similarly to the shape analysis described in Sect. 2.1, we approximate each patch $\mathbf{P}_j^i$ of size $T \times T$ by a vectorized linear combination of K main modes of variation as follows

$$\mathbf{g} \simeq \overline{\mathbf{g}} + \mathbf{D}_g \mathbf{b}_g \ , \tag{6}$$

where $\overline{\mathbf{g}} \in \mathbb{R}^{T^2 \times 1}$ is the mean normalized texture vector, $\mathbf{D}_g \in \mathbb{R}^{T^2 \times K}$ contains the main modes of variation of the patch, and $\mathbf{b}_g \in \mathbb{R}^{K \times 1}$ is a vector containing the deformation coefficients that weight each variation mode in $\mathbf{D}_g$.

Given a test image, $\mathbf{I}_t$, and an initial guess of the model position, a set of image-specific texture templates can be generated by computing the $\widehat{\mathbf{b}}_{g_t}$ as follows

$$\widehat{\mathbf{b}}_{g_t} = \mathbf{D}_g^{-1}(\mathbf{g}_t - \overline{\mathbf{g}}) \ , \tag{7}$$

where $\mathbf{g}_t$ is the vectorized texture patch around a specific landmark point of $\mathbf{I}_t$. The vector $\mathbf{b}_{g_t}$ contains the parameters that best match the statistical model to the test image. Without additional constraints, $\mathbf{b}_{g_t}$ may correspond to an unrealistic patch. Therefore, an additional step is required to constraint the solution in (7). To achieve this, the Mahalanobis distance $d_{\mathcal{M}}$ is used to measure the acceptability of the generated patches. More concisely, $d_{\mathcal{M}}$ has to be lower

[1] In this equation we suppressed the subscripts i and j for the simplicity of the notation.

than a specific predefined threshold $d_{\max}$

$$d^2_{\mathcal{M}} = \sum_{l=1}^{L} \frac{\widehat{b}_l^2}{\lambda_l} \leq d^2_{\max} \ , \tag{8}$$

where $\widehat{b}_l$ denotes the l-th component of $\widehat{\mathbf{b}}_{g_t}$ and λ_l is the eigenvalue associated to the l-th deformation mode. If $\widehat{\mathbf{b}}_{g_t}$ does not satisfy (8) the variation mode is rescaled as follows

$$\widehat{\mathbf{b}}_{g_t} = \widehat{\mathbf{b}}_{g_t} \frac{d_{\max}}{d} \ . \tag{9}$$

Finally, the normalized cross-correlation can then be applied, as described in (5), between the template and the predefined search region, sampled around each model point. By finding the strongest match, the new likely location of the object landmarks can thus be determined.

3 Experimental Setup

This section presents the experimental setup used to evaluate the proposed framework. Two different applications were used to test each feature detection method: facial fitting and segmentation of the left ventricle (LV) in cardiac magnetic resonance (CMR) images.

3.1 Dataset

In the first application, we addressed the problem of fitting an ASM to facial image sequences. For that, we used the publicly available Cohn-Kanade (CK+) database [12, 13] of emotion sequences taken from frontal view, where the manual face annotation are available, i.e., the ground truth (GT). Among several emotion sequences, we took the "surprise" sequences, since they contain more challenging lip boundary deformations and large eyebrow displacements. The dataset comprises 56 different sequences, each with 10-20 frames, with a total of 912 images (each with 490 × 640 size). The leave-one-sequence-out cross validation was used for performance evaluation. For initialization purposes and in order to have an initial guess of the model position, we used the Viola-Jones detector [14], for finding the faces in each image, which results in a rectangle containing the face. The learned mean shape is then aligned to the centre of the rectangle, resulting in a rough initialization of the model points.

The second application is the segmentation of the LV in CMR images. For that purpose, we use the publicly available dataset [15], which contains data from 33 different patients. For each patient, the CMR data is a sequence of 20 volumes with 8 to 15 slices, in a total of 7980 images, coupled with the manual LV border annotations, which is used as the ground truth. Each slice is a 256 × 256 image with an average resolution of 1.4 ± 0.2 mm/pixel, nevertheless, the LV is only present in 4 to 10 of the them (the remaining slices are disregarded). Each volume is segmented independently, whereby, as for the previous application, the leave-one-sequence-out cross validation is also used for each patient volume, i.e., for each one of the 20 volumes of the 33 patients, the statistics are learned from that same volume of the remaining 32 patients. The initialization is only performed for the first slice of each volume of the test sequence and, for that, we use the ground truth. The remaining slices are initialized by successively propagating the previous slice segmentation.

3.2 Error Metric

Segmentations are evaluated by comparing the estimated contour with the true object boundary (the ground truth). Their performance is quantitatively measured using two different metrics: the average distance error [16] for the facial segmentation and the Dice coefficient [17] for the LV segmentation, as detailed next.

Average Distance ($\boldsymbol{d}_{\mathbf{AV}}$) Let us assume $\mathbf{x} = [\mathbf{x}_i^\top]_{i=1,\ldots,N}$ and $\mathbf{y} = [\mathbf{y}_i^\top]_{i=1,\ldots,N}$, with $\mathbf{x}_i$, $\mathbf{y}_i \in \mathbb{R}^2$ being two point vectors representing the estimated and the ground truth of the face, respectively. The AV between $\mathbf{x}$ and $\mathbf{y}$ is defined as the average Euclidean distance, as follows

$$d_{\mathrm{AV}}(\mathbf{x}, \mathbf{y}) = \frac{1}{N} \sum_{i=1}^{N} \|\mathbf{y}_i - \mathbf{x}_i\| \ . \tag{10}$$

Dice Coefficient ($\boldsymbol{d}_{\mathbf{Dice}}$) Assuming now $\mathbf{S}_1$ and $\mathbf{S}_2$ as two binary images associated with the estimated contour $\mathbf{s}_1$ and the ground truth $\mathbf{s}_2$, respectively, such that the pixels inside the LV segmentation have value one and the pixels outside have the value zero. We can compute the Dice coefficient as follows

$$d_{\mathrm{Dice}}(\mathbf{S}_1, \mathbf{S}_2) = \frac{2\ C(\mathbf{S}_1 \wedge \mathbf{S}_2)}{C(\mathbf{S}_1) + C(\mathbf{S}_2)} \ , \tag{11}$$

where $C(.)$ is a function that counts the number of pixels within the region and the operator $\wedge$ denotes a pixel-wise AND. Note that d_{Dice} is always between 0 and 1, where a value of 1 reflects a perfect match between the two segmentations.

3.3 Results

Facial Segmentation The performance of the different feature detectors was evaluated and compared first for facial fitting. The same initial guess was used in all the tested methods with $N = 44$ as the number of model landmark points. Figure 2 shows three examples of the segmentation obtained by each method. These images show the improved accuracy of the proposed methods with fixed (ASM-FTT) and variable (ASM-VTT) templates.

To ascertain the robustness of the proposed approach, in the experimental setup, several patch dimensions were tested, both for the search region and the feature templates. From the extensive experimental evaluation, we found the regions which achieved the best results were in the interval of $\{11 \times 11,\ 23 \times 23,$

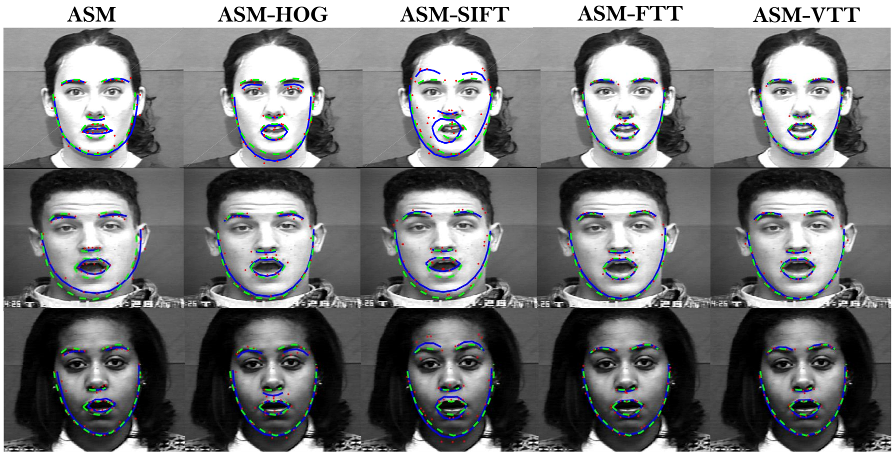

Fig. 2: Examples of the facial segmentation. Each column shows the result of a different method: ASM, ASM-HOG, ASM-SIFT, ASM-FTT and ASM-VTT, respectively The green dashed line shows the ground truth, whereas the blue line shows the estimated segmentation and the red dots represent the observation points in the last iteration.

$35 \times 35\}$ and $\{33 \times 33,\ 55 \times 55,\ 67 \times 67\}$ pixels, both for the template and the search region, respectively. It is important to remark that for the ASM implementation the patch size only determines the length of the profile lines, and for larger regions ASM quickly starts to degrade its performance. Figure 3 presents a statistical evaluation of the segmentation accuracy for each of the studied methods. Note that the results shown only concern the statistical performance for a search region of 55×55 pixels and a template of 23×23 pixels, as the results obtained sre similar for the different tested values. The results show that the proposed methods performed better than both the standard ASM approach and the alternative ones, namely the ASM-HOG and the ASM-SIFT, which is possible to verify through the decrease of the average distance with the use of the proposed models. Furthermore, both ASM-FTT and ASM-VTT have a similar performance and both lead to a significant improvement over the remaining tested methods.

Left Ventricle Segmentation The final application herein presented is the segmentation of the LV. Figure 4 shows three examples of the segmentation obtained with the different methods. These images clearly show the improved accuracy of the proposed methods both with fixed (ASM-FTT) and variable (ASM-VTT) templates.

Once more, the parameters were chosen after an extensive evaluation. Note that the images tested here have a reduced size, thus, in accordance to the first application, we find the region of 33×33 pixels as the most suitable for the search region, 15×15 pixels for the template size and $N = 40$ for the model landmark

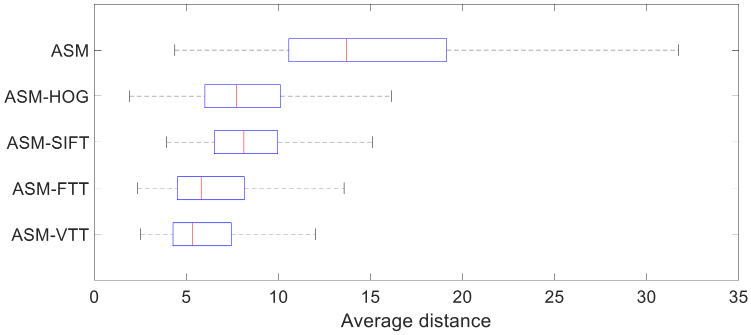

Fig. 3: Comparison of the statistical results of each method for the facial segmentation, using the average distance in pixels, d_{AV}.

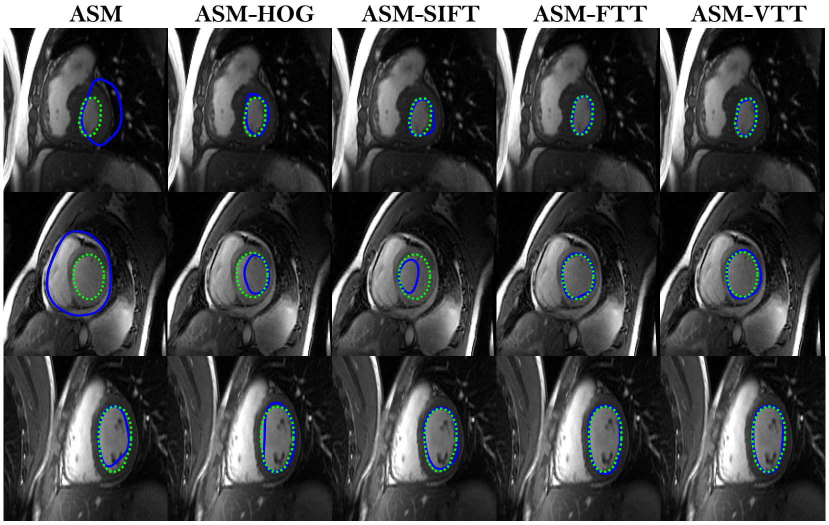

Fig. 4: Examples of LV segmentation with the different methods in three CMR images from different patients. Each column shows the result of a different method: ASM, ASM-HOG, ASM-SIFT, ASM-FTT and ASM-VTT, respectively. The green dashed line shows the ground truth, whereas the blue line represents the estimated segmentation.

points number. The statistical evaluation of the segmentation accuracy for each of the methods are represented in Fig. 5. The results show that ASM-FTT and ASM-VTT methods have a similar performance and both lead to a significant improvement in accuracy, outperforming the other tested approaches.

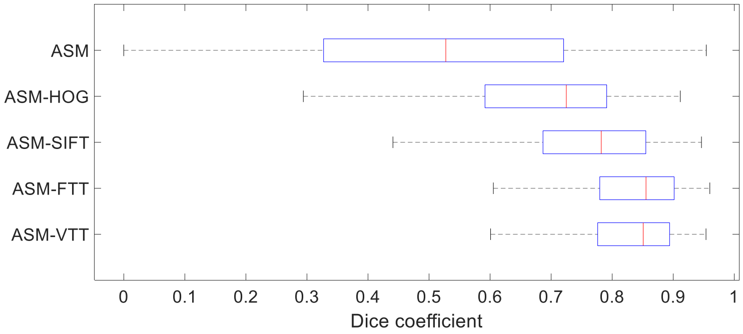

Fig. 5: Comparison of the statistical results of each method for the LV segmentation, using the Dice coefficient, d_{Dice}.

4 Conclusion

This paper proposes two novel ASM-based search methods to detect reliable observation points. The first one is based on a fixed texture template which remains unchangeable during the test phase, the ASM-FTT. The second method is based on statistical texture templates whose variation coefficients change with each test, the ASM-VTT. The two proposed methodologies are compared with: (i) the classical ASM edge detector, (ii) the ASM combined with a HOG detector, and (iii) the ASM combined with a SIFT detector. From the experimental evaluation, we applied these methodologies in two datasets, for the segmentation of faces in image sequences and for the segmentation of the left ventricle in CMR images. The obtained results are relevant and promising, as we have shown that the proposed methodologies lead to a better performance compared to the other three approaches. Further improvements can still be achieved by considering multiple observations points for each model point, instead of just one per patch. If the strongest feature in the patch is invalid (i.e., not belonging to the contour), this can jeopardize the segmentation accuracy. This means that more features should be extracted from each patch in the attempt to get the reliable one. Further work will extend the proposed approach to deal with multiple features for each patch.

References

1. T. F. Cootes, C. J. Taylor, D. H. Cooper, and J. Graham, “Active Shape Models - Their Training and Application,” *Computer vision and image understanding*, vol. 61, no. 1, pp. 38–59, 1995.
2. T.F. Cootes, G.J. Edwards, and C.J. Taylor, “Active Appearance Models,” *Proc. European Conference on Computer Vision (ICCV)*, vol. 2, pp. 484–498, 1998.
3. D. Cristinacce and T. F. Cootes, “Feature Detection and Tracking with Constrained Local Models,” *Procedings of the British Machine Vision Conference 2006*, pp. 95.1–95.10, 2006.
4. Epameinondas Antonakos, Joan Alabort-i Medina, Georgios Tzimiropoulos, and Stefanos Zafeiriou, “HOG Active Appearance Models,” in *Proceedings of IEEE International Conference on Image Processing (ICIP)*, 2014, pp. 224–228.
5. Fuji Ren and Zhong Huang, “Facial expression recognition based on AAM-SIFT and adaptive regional weighting,” *IEEJ Transactions on Electrical and Electronic Engineering*, vol. 10, no. 6, pp. 713–722, 2015.
6. Dianle Zhou, Dijana Petrovska-Delacrétaz, and Bernadette Dorizzi, “Automatic landmark location with a combined active shape model,” *IEEE 3rd International Conference on Biometrics: Theory, Applications and Systems, BTAS 2009*, 2009.
7. Navneet Dalal and Bill Triggs, “Histograms of oriented gradients for human detection,” *Proceedings - 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR 2005*, vol. I, pp. 886–893, 2005.
8. A Vadivel, AK Majumdar, and S Sural, “Performance comparison of distance metrics in content-based image retrieval applications,” *International Conference on*, , no. September, 2003.
9. Michael J. Swain and Dana H. Ballard, “Color indexing,” *International Journal of Computer Vision*, vol. 7, no. 1, pp. 11–32, 1991.
10. David G. Lowe, “Distinctive Image Features from Scale-Invariant Keypoints,” *International Journal of Computer Vision*, vol. 60, no. 2, pp. 91–110, 2004.
11. J .P. Lewis, “Fast Normalized Cross-Correlation,” *Vision Interface*, vol. 1995, no. 1, pp. 1–7, 1995.
12. P. Lucey, Jeffrey F. Cohn, T. Kanade, J. Saragih, Z. Ambadar, and I. Matthews, “The extended Cohn-Kanade dataset (CK+): A complete facial expression dataset for action unit and emotion-specied expression,” in *Proceedings of the Third International Workshop on CVPR for Human Communicative Behavior Analysis*, 2010, pp. 94–101.
13. T. Kanade, J. F. Cohn, and Y. Tian, “Comprehensive Database for Facial Expression Analysis,” in *Proceedings of the 4th IEEE International Conference on Automatic Face and Gesture Recognition*, 2000, number March, pp. 46–53.
14. P. Viola and M. Jones, “Rapid object detection using a boosted cascade of simple features,” in *Computer Vision and Pattern Recognition.*, 2001, vol. 1, pp. 511–518.
15. Alexander Andreopoulos and John K. Tsotsos, “Efficient and generalizable statistical models of shape and appearance for analysis of cardiac MRI,” *Medical Image Analysis*, vol. 12, no. 3, pp. 335–357, 2008.
16. Jacinto C. Nascimento and Jorge S. Marques, “Robust shape tracking with multiple models in ultrasound images,” *IEEE Transactions on Image Processing*, vol. 17, no. 3, pp. 392–406, 2008.
17. Lee R. Dice, “Measures of the Amount of Ecologic Association Between Species,” *Journal of the Neurological Sciences*, vol. 26, no. 3, pp. 297–302, 1945.