Active Shape Models with Invariant Optimal Features (IOF-ASM) Application to Cardiac MRI Segmentation

S Ordas[1], L Boisrobert[1], M Huguet[2], AF Frangi[1]

[1]Computer Vision Laboratory, Aragon Institute of Engineering Research, University of Zaragoza, Spain
[2]Centre Cardiovascular, CETIR Sant Jordi, Barcelona, Spain

Abstract

This paper addresses the problem of accurate segmentation of the endocardial contours of both ventricles and the epicardium of the entire heart. It reports on an extension of the Active Shape Model framework [1] making use of a non-linear appearance model as in the Optimal Features ASMs (OF-ASM) approach by van Ginneken et al. *[2] and incorporating a reduced set of differential Cartesian invariant features as local image descriptors. The new method is coined Invariant Optimal Features ASM (IOF-ASM). Validation results on a vast cardiac MR data-set have shown that the new approach outperforms the original ASM* $(p < 0,001)$ *and that both optimal features algorithms have barely the same performance, yielding reliable segmentations.*

1. Introduction

Following a routinely acquired MRI scan an enormous amount of meaningful clinical data is potentially available. Accurate segmentation of the cardiac chambers constitutes a prerequisite for almost any quantitative diagnostic procedure. Nevertheless, manual delineation is still the most usual approach for quantification. Since the usual geometry of the cardiac ventricles is known, it seems natural to incorporate prior shape and/or texture knowledge into the segmentation process. ASM is a flexible methodology that has been used for the segmentation of a wide range of applications, including medical imagery. In the original approach of Cootes *et al.* [1] the statistical model of the shape is derived from the covariance matrix of a Point Distribution Model (PDM) while a set of local Grey-Level Profiles (GLPs) are used to capture the local grey-level variations observed at each landmark position of the PDM. The shape model is projected onto a subspace following a Principal Component Analysis (PCA) of the PDM, and retaining only a number of desired modes of variation. The fitting procedure is an alternation of landmark displacement and model fitting in a multi-resolution fashion. With respect to the appearance model, ASMs models are based on a crucial assumption: the local texture around the landmarks has to be consistent for all the instances of the object. Nevertheless, in practice, local grey-levels around the landmarks can vary widely and pixel profiles around a textured object boundary are not very different from those in other parts of the image. In key:foo2, an extension to the original linear ASM approach is introduced, using a non-linear grey-level appearance model. The method was coined ASM with Optimal Features (OF-ASM). A reduced set of optimal local image features obtained by feature selection are fed into a k-Nearest-Neighbor classifier (kNN) that decides whether a given location is more likely to lie inside or outside the object. This metric replaces the *Mahalanobis* distance of the GLPs used in the linear ASM fitting stage. In this paper we present an extension to the non-linear appearance approach of the OF-ASMs, incorporating a reduced set of differential invariant features as local image descriptors. The new method is coined, Active Shape Models with Invariant Optimal Features (IOF-ASM). In this way, the method is invariant to Euclidean transformations given that only cartesian differential Euclidean invariants (up to a given order) are chosen as local image descriptors.

2. Theory

2.1. Linear ASM

Overview - In an ASM, the statistical model of the shape is derived from the PDM of a set of landmarks points, used to model the shape of an object and its variations. The appearance model is built from a set of local GLPs (normalized first order derivatives), used to describe the local intensity variations at each landmark position. The fitting procedure is an alternation of landmark displacement and model fitting in a multi-resolution manner.

Shape Model - A set of n landmark points, is used to describe the object geometry. These landmark points are (manually) determined in a set of s training images. From these collections of landmark points, a PDM is constructed as follows. The coordinates of the landmark points are concatenated to form a shape vector $\{x_1, y_1, ..., x_n, y_n\}$. The shape vectors of all the training set are translated,

0276–6547/03 $17.00 © 2003 IEEE

Computers in Cardiology 2003;30:633–636.

rotated and scaled in order to align them and remove the pose. The mean shape $\bar{x} = \frac{1}{s}\sum_{i=1}^{s} x_i$ and the covariance matrix $S = \frac{1}{s-1}\sum_{i=1}^{s}(x_i - \bar{x})(x_i - \bar{x})^T$ are computed. The eigenvectors corresponding to the t largest eigenvalues λ_i of the covariance matrix are retained in the matrix $\Phi = \{|\phi_1|...|\phi_n|\}$. A shape approximation is obtained $x \approx \bar{x} + \Phi b$ where b is a vector of t elements given by $b = \Phi^T(x - \bar{x})$. When fitting the model to a set of points, the values of b are constrained to lie within the range, $\pm\beta\sqrt{\lambda_i}$ where β controls the flexibility of the model and is usually between 2 and 3. A set of t eigenvalues is retained so as to explain a certain proportion f_v of the variance in the training shapes, usually ranging from 90% to 99.5%. The desired number of modes is given by the smallest t for which $\sum_{i=1}^{t}\lambda_i \leq f_v \sum_{i=1}^{n}\lambda_i$.

Appearance Model - The typical image structure that describes the texture around each landmark is obtained from the GLPs : k pixels are sampled (using linear interpolation) around each landmark, perpendicular to the contour. The first derivative (finite differences approximation) of the $2k + 1$ length vector is calculated and normalized. The normalization is such that the sum of absolute values of the elements in the derivative profile is one. Denoting these normalized derivatives profiles as $g_1, ..., g_s$, the mean profile $\bar{g}$ and the covariance matrix S_g are computed for each landmark. This allows for the computation of the *Mahalanobis* distance $f(g_f) = (g_f - \bar{g})^T S_g^{-1}(g_f - \bar{g})$ between a new GLP g_f and the GLPs of the appearance model, given by their mean and covariance matrices. The GLPs are constructed for a number N_σ of resolutions. The finest resolution uses the original image and a step size of 1 pixel when sampling the profiles. Subsequent levels are constructed by doubling the image scale and the step size.

Fitting - Each landmark is displaced along the direction perpendicular to the contour to n_s positions on either side, evaluating a total of $2n_s + 1$ positions. The step size is $2^{(i-1)}$ pixels for the ith resolution level. After moving all landmarks, the shape model retrieves a plausible shape from the displaced points, yielding an updated segmentation. This is repeated $N_{\max}$ times at each resolution, in a coarse-to-fine fashion.

2.2. OF-ASM

Non-linear appearance model - The *Mahalanobis* distance assumes a normal multivariate unimodal distribution of GLPs but, in practice, they will exhibit any arbitrary statistical distribution. In this approach, the landmark points are moved to better locations along a profile perpendicular to the object contour at each landmark location; but this time, the best displacement will be the one for which everything on one side of the profile is believed to be outside the object, and viceversa. OF-ASMs base this decision on optimal local image features obtained by feature selection and a texture classifier (see below). Images can be locally described by their Taylor series expansion, provided the derivatives at the point of expansion can be computed up to a given order. In order to assure a well-posed derivative calculation of the discontinuous intensity function, derivatives of images can be computed by convolution with derivatives of Gaussians at a particular scale σ. This set of derivatives were named the "local jet" [3]. Given a set of filtered images, features for each location are extracted by taking the first two moments (i.e. mean and std. dev.) of the histogram around each location, using a Gaussian window function. The size of this window function is characterized by a second scale parameter α. The construction of local histograms, extracted from a Gaussian aperture function, is called a Locally Orderless Image [4]. In the OF-ASM approach, for most of the applications in which was tested, the first and second moments: m_1, m_2 (N_m = 2), all derivatives up to second order: L, L_x, L_y, L_{xx}, L_{yy}, L_{xy} (N_L = 6), with $L_m = \frac{\partial L}{\partial m}$ and $L_{mn} = \frac{\partial^2 L}{\partial m \partial n}$, usually two to five inner scales: ($N_\sigma = 2-5$), and a fixed relation between the inner scale σ and the histogram extent α, ($\alpha = 2\sigma$), are used. Thus, the dimension of the feature vector is $N_m \times N_L \times N_\sigma$. For each landmark, a squared grid with an odd number of points $N_{grid} \times N_{grid}$ around it, is defined. The spacing is $2^{(i-1)}$ pixels for the ith resolution level. Therefore, for each landmark and for each resolution level, a feature vector with N_m x N_L x N_σ elements is sampled at N_{grid} x N_{grid} points with decreasing inter-point spacing, from coarse to fine.

Sequential Feature Selection (SFS) - SFS algorithms attempt to select the best reduced-set of features that optimize a criterion function from a large set of available features. In general, the optimal SFS algorithm depends on the application. A way of doing the SFS is as follows: after a Sequential Feature Forward Selection (SFFS) stage, at most f_m features (for each landmark and resolution) are retained. The selected group of features can be trimmed by Sequential Feature Backward Selection (SFBS) if that improves the performance.

Fitting - The feature vector at each position along the profile is fed into the kNN classifier. Its output is either 1 (inside the object) or 0 (outside the object). A weighted voting of N_w neighbors is used. The weight of each vote is e^{-d^2}, where d is the Euclidean distance to each neighbor in the feature space. The objective function $f(g)$ to be minimized is the sum of absolute differences between the expected probability and the predicted probability, for each point along the profile $f(g) = \sum_{i=-k}^{-1} g_i + \sum_{0}^{k}(1 - g_i)$ where

the index along the profile g, oriented from the outside to the inside of the object, runs from $-k$ to $+k$. This non-linear metric replaces the *Mahalanobis* distance of the linear ASM approach.

2.3. IOF-ASM

The limitation of using the derivatives directly in the feature vector of the OF-ASM approach is the lack of invariance with respect to translation and rotation (Euclidean transformations). Consequently, these operators can only cope with textured boundaries with the same orientations of those seen in the training set. Cartesian Differential Invariants (CDIs) describe the differential structure of an image independently of the chosen cartesian coordinate system. Florack *et al.*[3] described their construction from Gaussian differential operators and their local image geometry independence to Euclidean transformations. The method continues evaluating at each location and for each resolution the first two moments ($N_m = 2$) of the locally orderless histograms of extent α ($\alpha = 2\sigma$) for the feature images L, L_{ii} ($L_{xx} + L_{yy}$), $L_i L_i$ ($L_x^2 + L_y^2$), $L_i L_{ij} L_j$ ($L_x^2 L_{xx} + 2L_{xy} L_x L_y + L_y^2 L_{yy}$), $L_{ij} L_{ji}$ ($L_{xx}^2 + 2L_{xy}^2 + L_{yy}^2$) ($N_L = 5$), at N_σ inner scales. The dimension of the feature vector is, N_m x N_L x N_σ. Note that the CDIs still depend on the choice of the scale parameter σ.

3. Experiments

Data Set and Gold Standard - The performance of the algorithms was tested on a data set comprising a total of 1356 routinely acquired short-axis MR images. The gold standard segmentations were manually drawn by an expert. The studies correspond to 95 subjects, 21 healthy and 74 patients suffering various common cardiac pathologies (myocardium infarction, hypertrophic obstructive cardiomyopathy, LV aneurysm, etc.). From each patient, 5 phases and 3 slices were manually segmented. The latter were selected from the base to the mid-ventricle portion, assuring that the right ventricle was present in all the slices. The models were built from a training set including 21 subjects (8 healthy and 13 patients). The rest of the data set was used for validation. The shape model PCA included 34 modes of variation explaining 99.5 % of the total variance. The acquisition parameters were: TR: 3.75 4ms, TE: 1.5 1.58 ms, FA: 45, slice thickness: 8-10 mm, slice size: 256×256 pixels, resolution: 1.56×1.56 mm and FOV: 400×300 mm^2, on a General Electric MRI facility.

Model Parameters - The shape model includes three closed contours, namely the endocardial left ventricle (LV ENDO) and right ventricle (RV ENDO) borders, and the epicardial border of the whole heart (HEART EPI). Following common practice in clinical quantitative cardiac MR analysis, endocardial contours were traced behind the papillary muscles and trabeculae and the epicardial contour along the inner border of the epicardial fat layer. Starting from a reference point at the posterior junction of the LV and RV, the three contours were equi-spatially sampled defining a total of 74 landmarks (n). Table 1 shows the values of each parameter for the three methods.

Parameter	Explanation	ASM	OF	IOF
n	# of landmarks	74	74	74
s	# of training images	315	315	315
f_v	explained variation	99%	99%	99%
$2n_s$+1	sampled positions (model)	5	NA	NA
N_{grid}xN_{grid}	squared grid size	NA	5 x 5	5 x 5
N_m	# of moments	NA	2	2
N_σ	# of resolutions	3 (1;2;4)	3(1;2;4)	3(1;2;4)
N_L	# of differential features	NA	6	5
N_m x N_L x N_σ	feature vector dimension	NA	48	40
f_m	max. # of features after SFS	NA	5	5
$N_{\max}$	iterations per resolution	2;4;8	2;4;8	2;4;8
2k+1	sampled positions (fitting)	11	11	11
N_w	voting nearest neighbors	NA	5	5

Table 1. Parameters used in the experiments.

Quantitative assessment indices - The utilized quantitative assessment indices were (1) the areas enclosed by the three contours expressed in cm^2 and (2) the point-to-curve (P2C) border positioning error (average signed, unsigned and maximum), in mm. In all experiments the performance when fitting the shape directly to the true landmarks was also computed because it indicates an upper bound for the tests.

4. Results

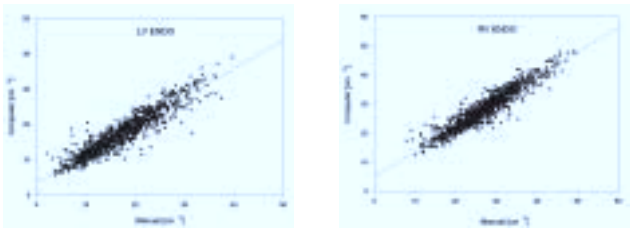
Table 2 summarizes the results of the experiments. Each of the two optimal features methods was significantly better ($p < 0.001$, paired t–test) than the linear ASM approach, but there were no significant differences between them. A good correlation was obtained between computerized and gold standard measurements of LV ENDO, RV ENDO and HEART EPI areas measures. See Fig. 1 (only the IOF-ASM results are shown.). HEART EPI results are not shown but are similar to RV ENDO.

5. Discussion and conclusion

Computing a local texture operator sampled around a local area, can deal with the textured boundary segmentation problem. Instead of assuming a Gaussian intensity profile distribution, a non-parametrically estimation is utilized which is important when having large variations in the background and/or in the object. In this way the inclusion of a non-linear appearance model in the ASM structure allows to model any number of arbitrary texture configurations for each landmark or shape sub-part. The presented method enables the automatic segmentation of routinely acquired 2D cardiac MR images taking advantage of prior shape knowledge and new local descriptors. It

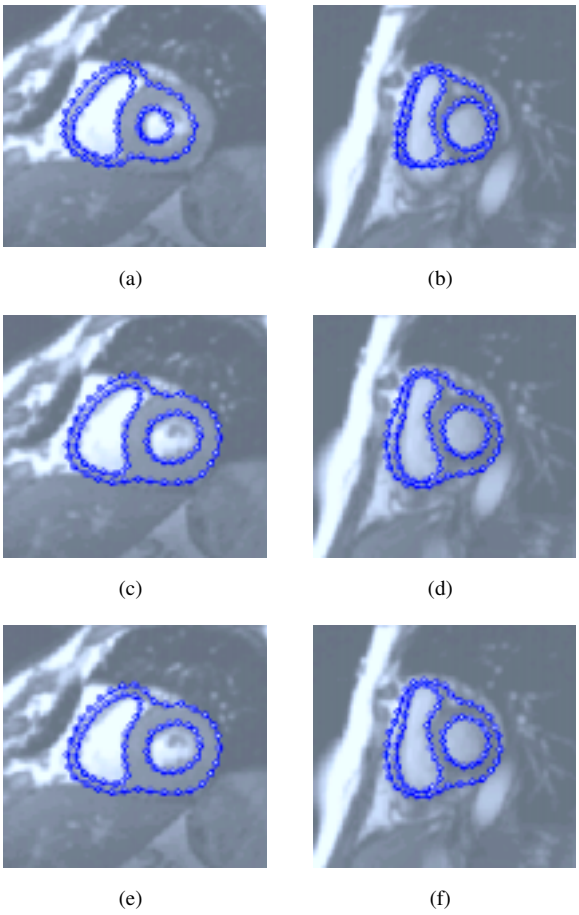
LV ENDO	Signed P2C	Unsigned P2C	Maximum
ASM	-2.89±2.36	3.07±2.13	8.54±4.43
OF-ASM	-0.79±2.61	2.01±1.85	6.06±3.43
IOF-ASM	-0.32±2.48	1.80±1.74	5.89±3.12
Fit of shape model	-1.21±0.39	1.22±0.35	3.08±0.65
RV ENDO	unsigned P2C	signed P2C	Maximum
ASM	0.57±2.78	1.98±2.04	9.66±5.11
OF-ASM	-0.01±1.66	1.14±1.21	6.12±3.11
IOF-ASM	0.31±1.87	1.20±1.47	6.26±3.56
Fit of shape model	-1.15±0.27	1.15±0.25	3.42±0.67
HEART EPI	unsigned P2C	signed P2C	Maximum
ASM	-1.88±3.76	3.50±2.32	13.83±3.99
OF-ASM	-0.01±2.34	1.49±1.80	7.76±3.37
IOF-ASM	0.48±2.48	1.52±2.01	7.73±3.52
Fit of shape model	-1.21±0.29	1.21±0.28	3.35±0.60
Areas Error	LV ENDO	RV ENDO P2C	HEART EPI
ASM	3.52±3.52	4.57±4.25	3.65±3.24
OF-ASM	2.05±2.29	2.24±2.76	2.53±3.30
IOF-ASM	2.10±2.16	2.55±2.81	2.97±3.81

Table 2. Mean ± SD of the signed, unsigned and maximum P2C errors in mm and of the area calculation error in cm^2.



(a) (b)

Figure 1. Regression graph between computer and gold standard measures. (a) LV ENDO. (b) RV ENDO.



(a) (b) (c) (d) (e) (f)

Figure 2. Two representative examples of the segmentation obtained by the three approaches. (a)-(b) ASM. (c)-(d) OF-ASM. (e)-(f) IOF-ASM.

consistently obtains high quality segmentations in spite of the poor quality of some images and the amount of shape variation induced by some diseases (specially for RV ENDO and HEART EPI). The results indicate that the two OF- approaches are clearly better than the original ASM method, but in this image modality and type of slice orientation, there was no significant difference between them. Some typical segmentation results are shown in Figure 2. The computational cost on a regular PC hardware for both OF- methods (2-3 minutes) per slice constitutes an issue to improve as ASMs are very fast (less than a second). As opposed to previous studies, the performance of the different methods was tested on a data set with an enormous amount of images, from an ordinary cardiac MR facility. Such a large validation demonstrates the value of optimal features ASMs cardiac segmentation.

Acknowledgements

This work is supported by MCyT grants TIC2002-04495-C02 and TIC2000-1635-C04-01. SO and LB are supported by MEC-FPU grants AP2002-3955 and AP2002-3946 respectively. AFF holds a MCyT Ramn y Cajal Research Fellowship.

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*, 61(1):38-59, 1995.

[2] B. van Ginneken, A.F. Frangi, J.J. Staal, B.M. ter Haar Romeny, M.A. Viergever, "Active shape model segmentation with optimal features" *IEEE Transactions on Medical Imaging*, 21(8), 924-33, 2002.

[3] Luc Florack, *The Syntactical Structure of Scalar Images*, Ph.D. thesis, Universiteit Utrecht, 1993.

[4] J.J. Koenderink and A.J. van Doorn. "The structure of locally orderless images". *International Journal of Computer Vision*, 31(2/3):159-168, 1999.

Address for correspondence:

Sebastian Ordas
Computer Vision Laboratory
Aragon Institute of Engineering Research
University of Zaragoza
Maria de Luna,1 "Ada Byron" Building
E-50018, ZARAGOZA, Spain
Tel.: +34-976-762705 Fax.: +34-976-762111
sebordas@unizar.es