Level Set Methods in an EM Framework for Shape Classification and Estimation

Andy Tsai[1,2], William Wells[3,4], Simon K. Warfield[3,4,5], and Alan Willsky[2]

[1] Department of Medicine, Massachusetts General Hospital, Harvard Medical School, Boston, MA, USA
[2] Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge, MA, USA
[3] Department of Radiology, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA
[4] Computer Science and Artificial Intelligence Laboratory, Massachusetts Institute of Technology, Cambridge, MA, USA
[5] Department of Radiology, Children's Hospital, Harvard Medical School, Boston, MA, USA

Abstract. In this paper, we propose an Expectation-Maximization (EM) approach to separate a shape database into different shape classes, while simultaneously estimating the shape contours that best exemplify each of the different shape classes. We begin our formulation by employing the level set function as the shape descriptor. Next, for each shape class we assume that there exists an unknown underlying level set function whose zero level set describes the contour that best represents the shapes within that shape class. The level set function for each example shape is modeled as a noisy measurement of the appropriate shape class's unknown underlying level set function. Based on this measurement model and the judicious introduction of the class labels as hidden data, our EM formulation calculates the labels for shape classification and estimates the shape contours that best typify the different shape classes. This resulting iterative algorithm is computationally efficient, simple, and accurate. We demonstrate the utility and performance of this algorithm by applying it to two medical applications.

1 Introduction

Shape classification can be defined as the systematic arrangement of shapes within a database, based on some similarity criteria. It has received considerable attention in recent years with important applications to problems such as computer aided diagnosis, handwriting recognition, and industrial inspection. The various classification techniques in the literature can be broadly categorized into those based on feature matching and those based on dense matching. Dense matching algorithms are computationally expensive as they try to transform or warp one shape into another based on some energy optimization scheme. For example, Del Bimbo and Pala [3] derived a similarity measure between two shapes based on the amount of elastic deformation energy involved in matching

C. Barillot, D.R. Haynor, and P. Hellier (Eds.): MICCAI 2004, LNCS 3216, pp. 1–9, 2004.

the shapes. Cohen *et. al.,* [2] developed an explicit mapping between two shape contours based on finite element analysis. Basri *et. al.,* [1] used the sum of local deformations needed to change one shape into another as the similarity metric in comparing two shapes.

Feature matching algorithms are more popular and utilize low-dimensional feature vectors extracted from the shapes for classification. For example, Dionisio and Kim [5] classified objects based on features computed from polygonal approximations of the object. Kawata *et al.,* [8] extracted surface curvatures and ridge lines of pulmonary nodules from 3D lung CT images to discriminate between malignant and benign nodules. In Golland *et al.,* [7], skeletons are used to extract features which are then used within different linear classification methods (Fisher linear discriminant and linear Support Vectors method). Gdalyahu and Weinshall [6] constructed syntactic representation of shapes (with primitives consisting of line segments and attributes consisting of length and orientation), and used a variant of the edit matching procedure to classify silhouettes.

We consider our algorithm as a feature matching algorithm. The individual pixels associated with each shape's level set representation are the features associated with that particular shape. One might argue that the dimensionality of this feature space is too high, and is not really a feature space as it does not capture only the salient information pertinent to a shape. However, we believe that it is the over representation or redundancy within this feature space that lends simplicity to our formulation and affords us the ability to capture very subtle differences among shapes for classification. We then incorporated this high-dimensional feature vector within an EM framework to provide us with a principled approach of comparing shapes for classification.

The rest of this paper is organized as follows. Section 2 illustrates how we incorporated the level set methods into the EM framework for shape classification and estimation. In Section 3, we present experimental evaluations of our algorithm by applying our technique to two medical problems. We conclude in Section 4 with a summary of the paper and a discussion on future research.

2 Shape Classification and Estimation

Given a database of example shapes, the goal of our algorithm is two-fold: (1) to separate the example shapes into different groups of approximately the same shapes (based on some similarity measure), and (2) to estimate the shape contour for each group that best represents or typifies the shapes contained within that group. Accomplishing these two tasks of shape classification and estimation is difficult, and is the problem which we focus on in this paper.

In this shape classification and estimation problem, if the underlying shape contour of each shape class is known *a priori*, then various pattern recognition techniques in the literature can be employed to separate the shapes within the database into different groups. Similarly, if the class labels of every example in the database is known *a priori*, then the underlying contour for each shape class can be estimated by calculating the "average" shape contour within each

shape class. Needless to say, it is difficult to calculate or determine either the representative shape contour of each class or the class labels of the example shapes in the database without knowledge of the other. However, we will show in this paper that it is possible to estimate both using the EM algorithm.

2.1 Shape Representation

We begin by describing our choice of the shape descriptor. Let the shape database $\mathcal{T}$ consist of a set of L *aligned* contours $\{\mathcal{C}_1, \mathcal{C}_2, ..., \mathcal{C}_L\}$.[1] We employ the signed distance function as the shape descriptors in representing each of these contours [9]. In particular, each contour is embedded as the zero level set of a signed distance function with negative distances assigned to the inside and positive distances assigned to the outside. This technique yields L level sets functions $\{Y_1, Y_2, ..., Y_L\}$, with each level set function consisting of N samples (using identical sample locations for each function).

2.2 Measurement and Probabilistic Models

For simplicity of derivation and clarity of presentation, we assume that there are only two shape classes within the database which we would like to group. It is important to realize, however, that it is straightforward to generalize our algorithm to classify more than two classes. By limiting ourselves to classify only two shape classes, we can employ the binary class label $C = \{C_1, C_2, ..., C_L\}$ to indicate which of the two shape classes each of the example shapes belong to. Specifically, each C_l $\forall l = 1, ..., L$ takes on the values of 0 or 1.

In our problem formulation, we postulate that there are two unknown level set functions $X = \{X_1, X_2\}$, one associated with each of the two shape classes with the property that the zero level sets of X_1 and X_2 represent the underlying shape contours of the two shape classes A and B. Importantly, there are no restrictions placed on whether X is a signed distance function. Next, we view each example shape's level set function Y_l as a noisy measurement of either X_1 or X_2. Based on this formulation, the explicit dependence of Y_l on X and C_l is given by the following measurement model:

$$Y_{l_i} = \begin{bmatrix} C_l & (1 - C_l) \end{bmatrix} \begin{bmatrix} X_{1_i} \\ X_{2_i} \end{bmatrix} + v_i \qquad \forall i = 1, ..., N \tag{1}$$

where $v \sim \mathcal{N}(0, \sigma^2 I)$ represents the measurement noise with σ as the standard deviation of the noise process.[2] This measurement model gives us the following

[1] Any alignment strategy that will result in the shapes having the same orientation and size can be employed.

[2] The notation I represents the identity matrix and the notation $\mathcal{N}(\mu, \Lambda)$ represents a Gaussian random vector with mean μ and variance Λ.

conditional probability of Y given C and X:

$$p(Y|X,C) = \prod_{l=1}^{L} p(Y_l\big|X,C_l) = \prod_{l=1}^{L} \Big[C_l \quad (1-C_l)\Big] \begin{bmatrix} \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(Y_{l_i}-X_{1_i})^2}{2\sigma^2}} \\ \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(Y_{l_i}-X_{2_i})^2}{2\sigma^2}} \end{bmatrix} . \quad (2)$$

Of note, this probability model bears resemblance to the stochastic framework introduced in [10] for the construction of prior shape models.

We assume that the class labels C and the level set representations of the shape contours X are statistically independent and hence

$$p(C\big|X) = p(C) \; . \quad (3)$$

Without any prior knowledge regarding the classifications of the various example shapes in the database, we set

$$p(C_l) = \begin{cases} 0.5 & \text{if } C_l = 0 \\ 0.5 & \text{if } C_l = 1 \end{cases} \qquad \forall l = 1, ..., L \; . \quad (4)$$

2.3 The EM Framework

The EM procedure, first introduced by Dempster *et. al.* [4] in 1977, is a powerful iterative technique suited for calculating the maximum-likelihood (ML) estimates in problems where parts of the data are missing. The missing data in our EM formulation is the class labels C. That is, if the class labels for the different shapes within the database are known, then estimating the underlying shape contour which best represents each shape class would be straightforward. The observed data in our EM formulation is Y, the collection of level set representations of the example shapes. Finally, X is the quantity to be estimated in our formulation.

The E-step. The E-step computes the following auxilliary function Q:

$$Q(X\Big|X^{[k]}) = \Big\langle \log p(Y,C|X)\Big|Y, X^{[k]}\Big\rangle \quad (5)$$

where $X^{[k]}$ is the estimate of X from the kth iteration, and $\langle \cdot \rangle$ represents the conditional expectation over C given Y and the current estimate $X^{[k]}$. Using Bayes' rule and our earlier simplified assumption that C and X are statistically independent, Q can be rewritten as

$$Q(X\Big|X^{[k]}) = \Big\langle \log p(Y|X,C)\Big|Y, X^{[k]}\Big\rangle + \Big\langle \log p(C)\Big|Y, X^{[k]}\Big\rangle . \quad (6)$$

Since the M-step will be seen below to be a maximization of $Q(X\big|X^{[k]})$ over X, we can discard the second term in Eq. (6) since it does not depend on X.

Expanding the remaining term in Eq. (6), we have that[3]

$$Q(X|X^{[k]}) = -\sum_{l=1}^{L} \left[\langle C_l | Y_l, X^{[k]} \rangle \quad \left(1 - \langle C_l | Y_l, X^{[k]} \rangle\right) \right] \begin{bmatrix} \sum_{i=1}^{N} \left(Y_{l_i} - X_{1_i}\right)^2 \\ \sum_{i=1}^{N} \left(Y_{l_i} - X_{2_i}\right)^2 \end{bmatrix}. \quad (7)$$

As evident from above, the core of the E-step is the computation of $\langle C_l | Y_l, X^{[k]} \rangle$. Using the formula for expectations, Bayes's rule, and Eqs. (2), (3), and (4), we find that

$$\langle C_l | Y_l, X^{[k]} \rangle = \frac{\prod_{i=1}^{N} e^{-\frac{\left(Y_{l_i} - X_{1_i}^{[k]}\right)^2}{2\sigma^2}}}{\prod_{i=1}^{N} e^{-\frac{\left(Y_{l_i} - X_{1_i}^{[k]}\right)^2}{2\sigma^2}} + \prod_{i=1}^{N} e^{-\frac{\left(Y_{l_i} - X_{2_i}^{[k]}\right)^2}{2\sigma^2}}}. \quad (8)$$

This equation is equivalent to calculating the posterior shape class probabilities assuming that the underlying level set functions X_1 and X_2 are known.

The M-step. Estimates of X_1 and X_2 are obtained in the M-step of our formulation by maximizing the auxiliary function Q. In other words, the M-step calculates the $X^{[k+1]}$ such that

$$X^{[k+1]} = \arg\max_{X} Q(X|X^{[k]}). \quad (9)$$

To solve for $X^{[k+1]}$, we imposed the zero gradient condition to Eq. (7). In particular, by differentiating $Q(X|X^{[k]})$ with respect to X_{1_i} and X_{2_i} for each pixel i, and setting each resulting equation to 0, we obtain the following two expressions:

$$\begin{aligned} X_{1_i}^{[k+1]} &= \frac{\sum_{l=1}^{L} \left\langle C_l \middle| Y_{l_i}, X_i^{[k]} \right\rangle Y_{l_i}}{\sum_{l=1}^{L} \left\langle C_l \middle| Y_{l_i}, X_i^{[k]} \right\rangle} \qquad \forall i = 1, ..., N \\ X_{2_i}^{[k+1]} &= \frac{\sum_{l=1}^{L} \left(1 - \left\langle C_l \middle| Y_{l_i}, X_i^{[k]} \right\rangle\right) Y_{l_i}}{\sum_{l=1}^{L} \left(1 - \left\langle C_l \middle| Y_{l_i}, X_i^{[k]} \right\rangle\right)} \qquad \forall i = 1, ..., N \end{aligned}. \quad (10)$$

Eq. (10) is equivalent to an ML estimator of the level set functions X_1 and X_2 when the shape labels are known. Of interest, note that both X_1 and X_2 are weighted averages of distance maps from different shape examples. As a result, neither X_1 and X_2 are signed distance functions because distance functions are not closed under linear operations.

[3] Here, in order to take the log of $p(Y|X, C)$ in Eq. (2), we have used the fact that each C_l is a binary indicator which selects the appropriate probability distribution function.

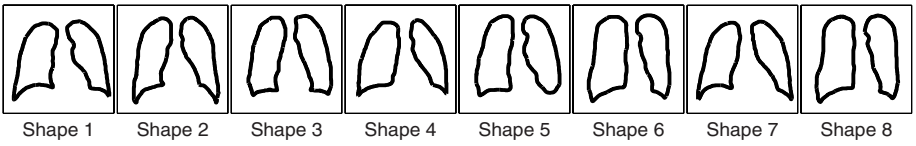
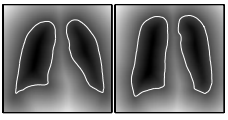


Fig. 1. Database of eight contours outlining the right and left lung fields from a collection of chest radiographs.



Shape	1	2	3	4	5	6	7	8
Radiologist	Nl	Nl	Dz	Nl	Dz	Dz	Nl	Dz
EM	A	A	B	A	B	B	A	B

Fig. 2. Left: Level set estimates of the two shape classes with the zero level set marked in white. Right: Comparion of chest radiograph labelings between a radiologist and our EM algorithm. The radiologist classified the chest radiographs into normal (Nl) or one with emphysema (Dz). The EM algorithm classified them into class A or class B.

3 Experimental Results

In this section, we present two medical applications to illustrate the performance of our algorithm. In both examples, we employ the alignment strategy proposed in [11] to guarantee that the example shapes in the database are all aligned to one another in terms of size and orientation. Furthermore, in these experiments, we start the iteration on the E-step, and initialize X_1 and X_2 to be the average of signed distance maps from two mutually complementary subsets of the shape database. Specifically, X_1 is set to be the average of the first four signed distance maps, and X_2 is set to be the average of the latter four signed distance maps. After convergence of our algorithm, we threshold the class labels to obtain the classification results we show in this section.

3.1 Chest Radiographs of Normal and Emphysematous Patients

Emphysema is a lung disease which involves the destruction of alveoli and its surrounding tissue. Typical findings on chest xrays of emphysema patients include hyperinflation of the lung fields and flattened diaphram. Figure 1 shows a database consisting of eight sets of contours with each set representing the outlines of the right and left lung fields from a different patient's chest radiograph. The eight patients' chest radiographs have been classified *a priori* by a radiologist as having either normal or emphysematous lung. The experiment here is two-fold: (1) to classify the eight sets of contours shown in Figure 1 into two groups based on our EM-based classifier, and (2) to compare the grouping scheme generated by our algorithm with the one from the radiologist's.

The experimental results are shown in Figure 2. The two images shown in Figure 2 represent the level set functions of the two shape classes with the

zero level set of each level set function outlined in white. The white contours can be thought of as the representative shape of each shape class. The table in the figure shows that the grouping scheme generated by our EM algorithm exactly matched the one generated by the radiologist. In particular, notice that Class A corresponds to normal and Class B corresponds to diseased patients. Not surprisingly, Class B's representative shape shows the hyperinflated lung as well as the flatten diagphram typical of emphysematous patients. For this particular experiment with each shape having 300×300 pixels, it took 5 iterations to converge requiring approximiately 1.67 seconds on an Intel Xeon 4.4GHz dual processor computer.

3.2 Cerebellum of Neonates with Dandy-Walker Syndrome

Dandy-Walker Syndrome is a congenital brain malformation associated with agenesis of the cerebellum. Our task is to separate the cerebellum database shown in Figure 3 into normal cerebellums and those afflicted with Dandy-Walker Syndrome. The eight cerebellums in the database are known *a priori* to either have the disease or not. The experiment here is: (1) to classify the eight cerebellums into two different groups based on our EM-based shape classifier, and (2) to compare the results of the grouping scheme generated by our EM algorithm with the one known *a priori*.

The experimental results are shown in Figure 4. The two shapes shown in Figure 4 are the representative shapes of the two shape classes. The table in the figure shows that the grouping scheme generated by our EM algorithm matched the correct answer. In particular, notice that Class A corresponds to normal and Class B corresponds to diseased patients. Class B's representative shape shows partial agenesis of the superior aspect of the cerebellum. For this 3D experiment, each example shape's level set function is represented by $256 \times 256 \times 50$ pixels. In terms of processing time, 10 iterations were required for convergence taking approximiately 39 seconds on an Intel Xeon 4.4GHz dual processor computer.

4 Conclusions and Future Research Directions

We have outlined a novel approach for statistical shape classification and estimation based on the EM algorithm and the level set representation of shapes. The approach we have outlined is flexible as it can handle the classification of complex shapes (including those that have dimensionality greater than two and those with complex topologies) into multiple shape classes (and not just limited to two classes). The experimental results we show here are encouraging as it demonstrates low classification errors with fast processing times. We are currently exploring the use of other implicit shape representations (other than distance transforms) that will not cause any inconsistencies in the shape representation during the calculation of the M-step. We are also interested in extending this formulation to enable it to provide users with information regarding the specific differences among the different shape classes.

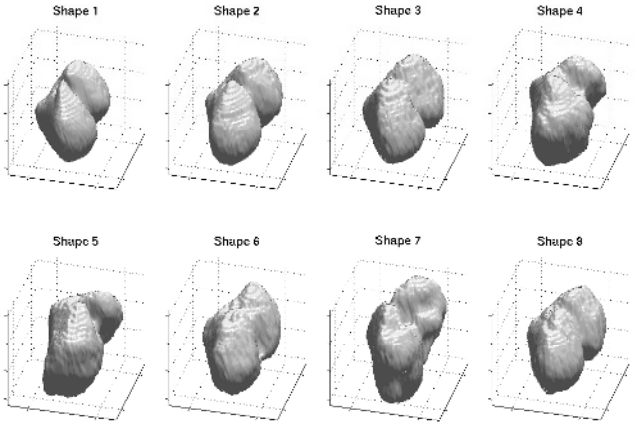
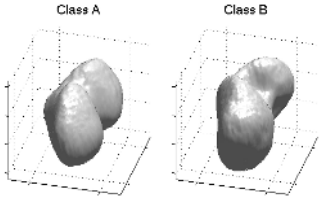


Fig. 3. Database of normal and diseased cerebellums.



Shape	1	2	3	4	5	6	7	8
Dx	Nl	Nl	Nl	Dz	Dz	Nl	Dz	Nl
EM	A	A	A	B	B	A	B	A

Fig. 4. Left: Shape estimates of the two shape classes. Right: Comparion of labelings between the truth and our EM classifier. The diagnosis (Dx) of the cerebellums is either normal (Nl) or one with Dandy-Walker syndrome (Dz). The EM algorithm classified the cerebellums into class A or class B.

Acknowledgements. This work was supported by MGH Internal Medicine Residency Program, AFOSR grant F49620-00-1-0362, Whitaker Foundation, NIH grants R21 MH67054, R01 LM007861, P41 RR13218, NSF ERC9731748, NIH 5 P41 RR13218, and JHU EEC9731748. The authors would like to thank Drs. A. Mewes and C. Limperopoulos for their help in the segmentation of the cerebellum data set.

References

1. R. Basri, L. Costa, D. Geiger, and D. Jacobs, "Determining the similarity of deformable shapes," *IEEE Workshop: Phys Based Modeling in Comput Vis*, pp. 135-143, 1995.
2. I. Cohen, N. Ayache, and P. Sulger, "Tracking points on deformable objects using curvature information," *ECCV*, pp. 458-466, 1992.
3. A. Del Bimbo and P. Pala, "Visual image retrieval by elastic matching of user sketches," *IEEE Trans PAMI*, vol 19, pp. 121-132, 1997.

4. A. Dempster, N. Laird, and D. Rubin, "Maximum-likelihood from incomplete data via the EM algorithm," *J of Royal Statist Soc Ser B,* vol. 39, pp. 1-38, 1977.
5. C. Dionisio and H. Kim, "A supervised shape classification technique invariant under rotation and scaling," *Int'l Telecommunications Symposium*, 2002.
6. Y. Gdalyahu and D. Weinshall, "Flexible syntactic matching of curves and its application to automatic hierarchical classification of silhouettes," *IEEE Trans. on PAMI*, vol 21, pp. 1312-1328, 1999.
7. P. Golland, E. Grimson, and R. Kikinis, "Statistical shape analysis using fixed topology skeletons: corpus callosum study," *IPMI*, pp. 382-387, 1999.
8. Y. Kawata, N. Niki, H. Ohmatsu, R. Kakinuma, K. Eguchi, M. Kaneko, and N. Moriyama, "Classification of pulmonary nodules in thin-section CT images based on shape characteristics," *IEEE Trans on Nucl Sci*, vol. 45, pp. 3075-3082, 1998.
9. S. Osher and J. Sethian, "Fronts propagation with curvature dependent speed: Algorithms based on Hamilton-Jacobi formulations," *J. of Comput. Phys.*, vol. 79, pp. 12–49, 1988.
10. N. Paragios and M. Rousson, "Shape priors for level set representations," *ECCV*, June 02, Copenhagen, Denmark.
11. A. Tsai, A. Yezzi, W. Wells, C. Tempany, D. Tucker, A. Fan, E. Grimson, and A. Willsky, "A shaped-based approach to segmentation of medical imagery using level sets.," *IEEE Trans on Medical Imaging*, vol. 22, pp. 137-154, 2003.