Histogram Statistics of Local Image Regions for Object Segmentation

Robert E. Broadhurst, Joshua Stough, Stephen M. Pizer, Edward L. Chaney

Medical Image Display & Analysis Group (MIDAG)
University of North Carolina, Chapel Hill NC 27599, USA
`reb@cs.unc.edu`

Abstract. We present a novel approach, based on local image histograms, for statistically characterizing the appearance of deformable models. In deformable model segmentation, appearance models measure the likelihood of an object given a target image. To determine this likelihood we compute pixel intensity histograms of local object-relative image regions from a 3D image volume near the object boundary. We use a Gaussian model to statistically characterize the variation of non-parametric histograms mapped to Euclidean space using the Earth Mover's Distance.

The new method is illustrated and evaluated in a deformable model segmentation study on CT images of the human bladder, prostate, and rectum. Results show improvement over a previous profile based appearance model, out-performance of statistically modeled histograms over simple histogram measurements, and advantages of local image regions over global regions.

1 Introduction

The segmentation of 3D deformable objects is an important and challenging task, especially in medical imaging. Automatic segmentation methods that statistically learn a prior on object shape and the likelihood of an object given an image have several desirable qualities. In this paper, we define an image likelihood measure using non-parametric histograms as our basic image measurement and describe a new method to statistically learn their likelihood. We acquire tighter distributions and locality by defining several object-relative image regions.

One category of appearance models is based on the correlation of pixel intensities. Intensities are acquired along profiles normal to the object boundary [3, 15] or from entire object-relative image regions [4, 7]. These methods can be used in conjunction with image filters to summarize information at a larger spatial scale and measure image structure such as texture, gradients, or corner strength [14]. Local methods, however, have difficulty capturing the inter-relations among pixel intensities in a region.

Region based methods are better than local methods at capturing pixel inter-relations. This is accomplished by aggregating pixel intensities over global image regions such as object interior or exterior, in one of two ways. In the first, region

statistics, such as mean and variance, are computed. These statistics are either learned during training or functions of them are defined to be minimized [2, 16]. Although the variation of region statistics can be learned during training, the statistics themselves capture limited information. In the second, each region is represented by a histogram, and a distance to a learned reference histogram is defined [6]. Histograms provide a rich estimate of a region's intensity distribution but previous work only specifies a reference histogram, and not its expected variation.

In this paper, we apply standard statistical tools to model histogram variation. To do this we map non-parametric histograms to points in Euclidean space using the Earth Mover's Distance (EMD) [1, 9, 13]. Straight-line paths between histograms in the resulting space provide interpolated histograms representing plausible distributions. The lack of distribution assumptions allow inhomogeneous regions to be modeled, though this typically results in loose distributions. In this case, we define several object-relative regions to obtain tighter distributions and add locality to our histogram measurements. This helps drive our segmentation algorithm to a more clearly defined optimum. In order to define these local regions we need a shape model that specifies a voxel to voxel correspondence near the object boundary; for this we use m-reps (see section 3.1) [10, 11].

Another property of appearance models, is two simplifying assumptions often used in defining an image's probability given a model. Image dependence on a model can be decomposed into describing the image relative to the model, and further correlations between the image and object shape. Appearance models can reasonably assume, when modeling anatomic objects, that object-relative images have intensities with no further probabilistic dependance on object shape. The probability of an object-relative image is determined using several image measurements, which are also often assumed to be independent. Local measurements are highly interrelated due to their small scale, however, so it is inaccurate to consider them as independent. It is also difficult to model local measurement inter-relations, since this requires a global high dimensional appearance representation with a complicated and hard to train covariance [4]. We can reasonably assume, on the other hand, that regional measurements of an object-relative image are independent, if the image is divided into anatomically based local regions and geometry variation is entirely captured by the shape prior.

In this paper, we assume regional image measurements are only relative to object shape, and are then conditionally independent. This defines image likelihood as the product of the probability densities derived from each region.

In section 2 we introduce our histogram methodology and construct a statistically learned histogram likelihood measure. In section 3 we overview our segmentation framework and give segmentation results using global image regions. In section 4 we extend this work to local image regions.

2 Statistical Modeling of Non-Parametric Histograms

We are interested in fully training a non-parametric histogram based appearance model. To do this we map histograms to points in Euclidean space in such a way that straight-line paths between two points produce a natural interpolation between the corresponding histograms. This mapping allows us to use standard statistical tools, such as Principal Component Analysis (PCA) and Gaussian modeling.

In section 2.1 we construct this mapping. We consider two properties of this space and construct a histogram's likelihood in section 2.2, and provide an example in section 2.3.

2.1 Mapping Histograms to Euclidean Space

Our mapping can be understood by considering the similarity measure defined between two histograms, that will correspond to Euclidean distance. We use the EMD, which was introduced by Rubner et al. for image retrieval [13] and has since been shown to be equivalent to the Mallow's distance [8]. The EMD representation we use is described for texture classification in [9] and used to build statistical models in [1].

The EMD, and the Mallow's distance for discrete distributions, can be thought of as measuring the work required to change one distribution into another, by moving probability mass. The position, as well as frequency, of probability mass is therefore taken into account yielding two major benefits. First, over-binning a histogram, or even using its empirical distribution, has no additional consequences other than measuring any noise present in the distribution estimate. Second, this distance measure to some extent mimics human understanding [13].

The Mallow's distance between continuous one-dimensional distributions q and r, with cumulative distribution functions Q and R, respectively, is defined as

$$M_p(q, r) = \left(\int_0^1 |Q^{-1}(t) - R^{-1}(t)|^p dt \right)^{1/p}.$$

Consider the Mallow's distance, for example, between two Gaussian distributions with the same variance and means m_1 and m_2, respectively. This distance can can be shown to be $|m_1 - m_2|$ for any p, which reflects the expected difference between samples of the two distributions.

In the case of discrete one-dimensional distributions, we consider two distributions x and y represented by empirical distributions with n observations, or equi-count histograms with n bins and the average value of each bin stored. Considering these values in sorted order, x and y can be written as vectors $\overline{x} = (x_1, x_2, \ldots, x_n)$ and $\overline{y} = (y_1, y_2, \ldots, y_n)$ with $x_1 \leq x_2 \leq \ldots \leq x_n$ and $y_1 \leq y_2 \leq \ldots \leq y_n$. The Mallow's distance between x and y is then defined as their L_p vector norm (to within a scale factor)

$$M_p(x, y) = \left(\frac{1}{n} \sum_{i=1}^{n} \|x_i - y_i\|^p \right)^{1/p}.$$

Using this representation, therefore, maps histograms to points in an n dimensional Euclidean space in which distances are understood as M_2 histogram distances.

2.2 Histogram Likelihood

In this section, we first examine properties of the constructed Euclidean space and then statistically define a histogram's likelihood.

To better understand this space, we consider two histograms obtained from different 3D regions around a bladder in a CT image, and examine example histograms along the line between the two mapped points. Figures 1 and 2 show uni-modal and roughly bi-model histograms, respectively, and support the use of linear interpolation in this space.
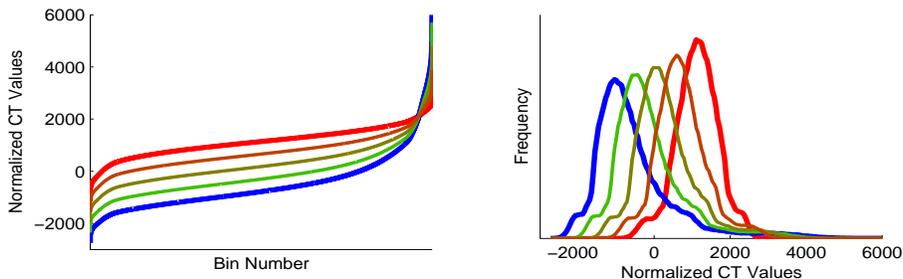

Fig. 1. Linear interpolation of two uni-modal histograms from global interior (blue) and exterior (red) regions of the bladder.

Only points in a convex portion of the Euclidean space represent valid histograms. That is, a point $\overline{x}$ is a valid histogram if and only if $x_1 \leq x_2 \leq \ldots \leq x_n$. Therefore, the mean of a set of histograms, or any interpolated histogram, will always be valid.

We face the usual high dimension low sample size situation with this space, since large training sets are often unavailable and a fair number of bins are desired to represent a histogram. So, to train the likelihood of a histogram given a set of training histograms we first apply PCA for dimension reduction. Next, we build a standard multi-variate Gaussian model for each region. Gaussian models can be thought of as stretching the space, changing the underlying metric, to account for the variability in the training data. In this sense a Gaussian model
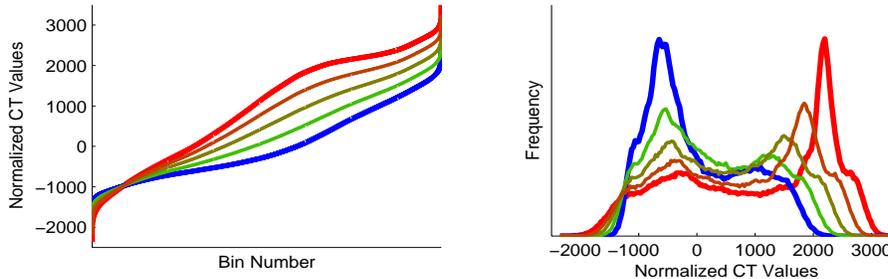
Fig. 2. Linear interpolation of two roughly bi-modal histograms from a local exterior bladder region near the prostate in two images.

should construct a histogram likelihood reflecting the training histograms. In this space a Gaussian model is not proper because it can assign a non-zero probability to points which do not represent valid histograms. For segmentation, however, we only need to determine the likelihood of valid histograms.

2.3 Global Regions Example

We present the following example to demonstrate the construction of a histogram's likelihood. We use 17 CT images of the pelvic region from one patient. The interior and exterior of the bladder, within 1 cm of its boundary, define two global regions. The histograms measured in these two regions, using 25 bins and normalized to have a combined zero mean, are shown in figures 3(a) and 3(d). In general, the interior, including bladder wall and urine, has higher CT values than the exterior as shown in figure 3(d). The exterior region consists mostly of fatty and prostate tissue, with the heavy tail representing the latter.

For dimension reduction, we apply PCA to the pooled 34 points in 25 dimensional Euclidean space. In this case, we choose 5 principal directions. Figure 4 shows scatter plots of points projected onto pairs of the first 3 principal directions. In particular, the first plot in figure 4 shows that the first principal direction separates the two regions.

Finally, to complete the training process we apply PCA separately to each region to build Gaussian models. Figures 3(b) and 3(e) show the mean for each region and ±0.5 standard deviations along the first principal direction from the mean. Figures 3(c) and 3(f) show the same for the second principal direction. These modes seem representative of the training data.

3 Segmentation using Global Regions

In this section, we use global regions, as defined in section 2.3, for segmentation. To do this we first discuss, in section 3.1, our shape model and segmentation framework. We then present segmentation results using these global image regions in section 3.2.
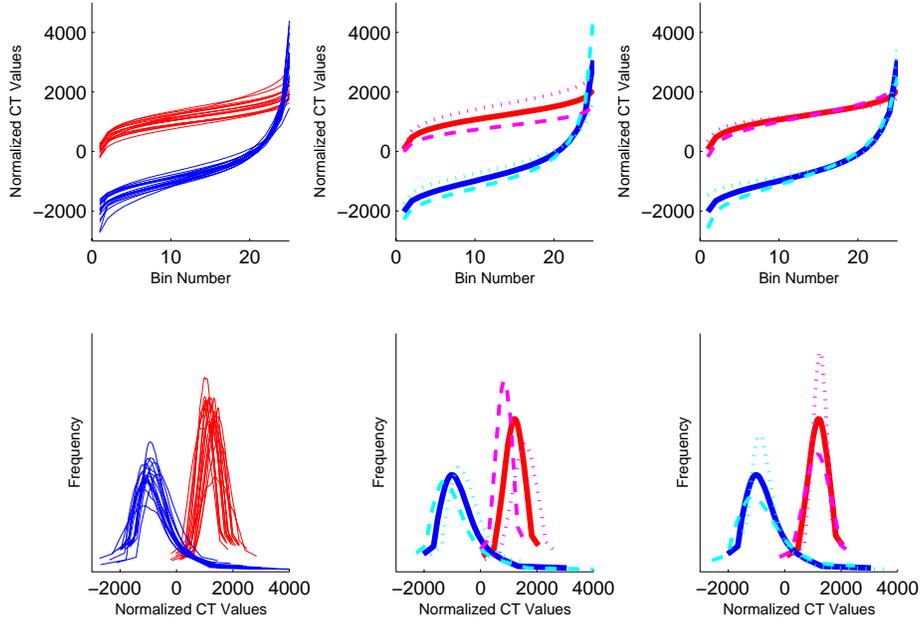
Fig. 3. (a)-(f) left to right. (a) and (d) are histograms representing 17 interior (red) and exterior (blue) bladder regions. (b) and (e) show the mean histogram of each region and ± 0.5 along the first principal direction from the mean. (c) and (f) show the same for the second principal direction.
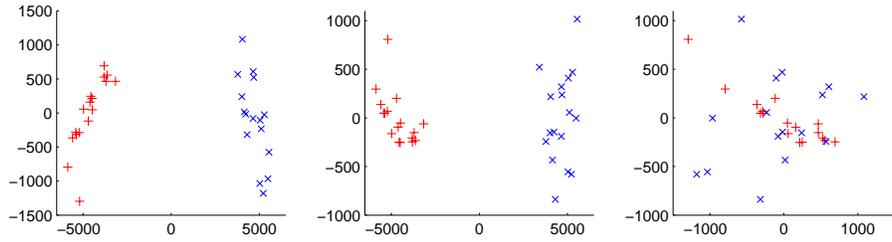

Fig. 4. Points in 25 dimensional Euclidean space representing histograms from 17 interior (red +) and exterior (blue x) bladder regions. The axes are the first 3 principal directions in pairs. Left: 1 vs 2, center: 1 vs. 3, right: 2 vs. 3.

3.1 The Segmentation Framework

Our goal is to automatically segment the bladder, prostate, and rectum in CT images. We use the m-rep model of single 3D figures, as in [10], to describe the shape of these deformable objects. The object representation is a sheet of medial atoms, where each atom consists of a hub and two equal-length spokes. The representation implies a boundary that passes orthogonally through the spoke ends. Medial atoms are sampled in a discrete grid and their properties, like spoke length and orientation, are interpolated between grid vertices. The model defines a coordinate system which dictates surface normals and an explicit correspondence between deformations of the same m-rep model and the 3D volume in the object boundary region. This allows us to capture image information from corresponding regions.

M-reps are used for segmentation by optimizing the posterior of the geometric parameters given the image data. This is equivalent to optimizing the sum of the log prior and the log likelihood, which measure geometric typicality and image match, respectively. Geometric typicality is based on the statistics of m-rep deformation over a training set [5]. We use the method described in section 2 for the image match. Furthermore, this optimization is multi-scale, with object ensemble, object, and atom stages.

In this paper, we evaluate our appearance model by segmenting the bladder, prostate, and rectum from an intra-patient dataset consisting of 17 images. Each image is from the same CT scanner and has a resolution of $512 \times 512 \times 81$ with voxel dimensions of $0.977 \times 0.977 \times 3.0$ millimeters. These images are acquired sequentially during the course of a patient's treatment for prostate cancer. Thus we segment each image using only the previous images for training. We estimate the model prior and likelihood by using m-reps fit to manual segmentations of the appropriate training set. Shape statistics are only gathered for the object ensemble and object stages, restricting the optimization during segmentation to these two stages.

Our primary concern in this paper is to determine the quality of the image likelihood optimum. For this initial test, we restrict our segmentations to the last 8 images so that we have sufficient training data to estimate adequate and stable statistics. We present segmentation results for the bladder, prostate, and rectum, and compare our results to a profile based method. This profile method uses normalized correlation with profiles from the first image and is described in [15]. All other aspects of these segmentation algorithms are identical, including the shape prior and automatic rigid body initialization. Comparisons are made relative to manual segmentations and put into context by showing our shape model's ability to represent the manual segmentations during training. Training performance serves as a baseline for the best expected performance.

3.2 Segmentation Results using Global Regions

We now evaluate the performance of three versions of our appearance model. For all three, we use two global regions for each object, defined as the object interior

and exterior within a fixed 1 cm collar region of the boundary. We represent each histogram with 25 equi-count bins and use 5 principal modes of variation for the Gaussian models.

The three versions of our appearance model learn increasingly more information during training. The *Simple Global* model creates a reference histogram for each region from the first image. The image match is a function of the EMD to each reference histogram. This model can be directly compared with the profile approach, since only the first image is supplied to both. The *Mean Global* model calculates the average histogram for each region using all the previous images. In this case, the image match is a function of the EMD to each average histogram. The last model, *Gaussian Global*, fully trains the likelihood of each region. The image match for this model is a function of the Mahalanobis distance of each target histogram projected into the PCA-defined 5-dimensional subspace.

Table 1 reports volume overlap, defined as intersection over union, and average surface distance, defined as the average shortest distance of a boundary point on one object to the boundary of the other object. Results show segmentation accuracy improves as more statistical training is done. Table 1 also shows a significant improvement of the global histogram based appearance models over the previous profile based model. Directly comparing the profile and histogram based methods, *Simple Global* achieves better results for the bladder and prostate, and worse results for the rectum. In the next section we improve these results using local image regions.

Table 1. Segmentation results of our appearance model using global image regions. Results are measured against manual contours, and compared against a previous profile based method and the ideal of our shape model attained during training.

	Volume Overlap			Ave. Surface Dist. (mm)		
Appearance Model	Bladder	Prostate	Rectum	Bladder	Prostate	Rectum
Training	87.5%	87.2%	87.9%	1.25	1.10	0.76
Profile	81.5%	73.3%	63.7%	1.93	2.58	2.74
Simple Global	82.8%	78.5%	56.6%	1.77	2.05	3.30
Mean Global	83.5%	77.6%	66.6%	1.65	2.12	2.43
Gaussian Global	84.7%	80.0%	67.8%	1.49	1.90	2.31

4 Defining Local Image Regions

We now use the appearance model described in section 2 with local object-relative image regions. Local regions have tighter intensity distributions than global regions since intensities are more locally correlated. This results in an image likelihood measure with a more clearly defined optimum, especially when

global regions consist of multiple homogeneous tissue regions. Since smaller regions are summarized, however, local regions provide less accurate distribution estimates. They also require a shape model that defines a voxel correspondence near the object boundary.

Our dataset contains two examples of global region inhomogeneity. First, the exterior bladder region consists of both prostate and fatty tissue. The bowel can also be present, though this is not the case in this dataset. The second example is the exterior rectum region. We only model the portion of the rectum near the prostate, so there are two arbitrary cutoff regions with exterior distributions matching those of the rectum's interior.

We describe two approaches to define local regions. In section 4.1, we manually partition the global interior and exterior regions. We define overlapping regions centered around many boundary points in section 4.2, and give results using both methods in section 4.3.

4.1   Partitioning Global Image Regions

Local regions can be defined by partitioning an object's surface, and hence the 3D image volume near the surface, into local homogeneous tissue regions. Such a partitioning can either be specified automatically, based on distribution estimates from a training set (see future directions), or manually delineated using anatomic knowledge.

In this section, we manually define several interior and exterior local regions for the bladder, prostate, and rectum, using limited anatomic knowledge. To create our manual partitions, shown in figure 5, we used several heuristics. First, more exterior regions are defined since there is more localized variability in the object exterior. For the bladder model a local exterior region is defined near the prostate. A local region is also defined for the portion of the bladder opposite the prostate since this region experiences the most shape variability between images. Lastly, for the rectum model a local exterior region is defined in each arbitrary cutoff region.

4.2   Local Image Regions

An alternative way to define local regions, is to consider a sampling of boundary points and set each point as a region's center. We define an interior and exterior region for each point by first finding the portion of the surface within a radius of each point. Then, we find all the voxels with object-relative coordinates associated with each surface patch, and within a certain distance to the boundary. This approach can define overlapping image regions at any scale and locality, and boundaries between local regions do not need to be learned.

For the bladder, prostate, and rectum we use 64, 34, and 58 boundary points, respectively. Each region is set to a radius of 1.25 cm and the collar region is kept at $\pm$ 1 cm, as in previous results.

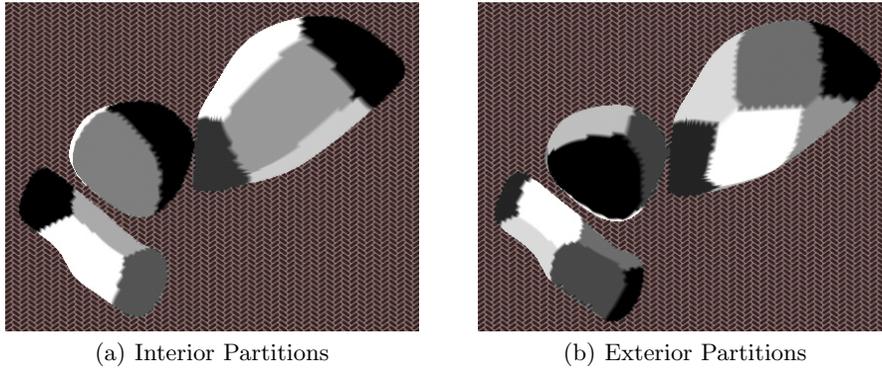| (a) Interior Partitions	(b) Exterior Partitions

Fig. 5. Manual surface partitions of the bladder, prostate, and rectum defining local interior (a) and exterior (b) regions. For the bladder, prostate, and rectum we define 6, 3, and 4 interior regions, and 8, 5, and 8 exterior regions, respectively.

4.3 Results

Table 2 gives segmentation results using the Gaussian appearance model from section 2 for both local region approaches. The *Partition* method refers to the approach in section 4.1, and the *Local* method refers to the approach in section 4.2. Both methods use 25 histogram bins and Gaussian models restricted to 3 principal directions of variation. These results show a slight advantage of the *Local* method both over the *Global* and *Partition* methods. Although the actual comparison numbers are sensitive to the number of principal directions, for different settings the *Local* method consistently performs the best. The greatest improvement is in the segmentation of the rectum.

Table 2. Segmentation results using local image regions. The Gaussian appearance model using the two local region methods is compared to the global region method.

Appearance Model	Volume Overlap			Ave. Surface Dist. (mm)		
	Bladder	Prostate	Rectum	Bladder	Prostate	Rectum
Training	87.5%	87.2%	87.9%	1.25	1.10	0.76
Gaussian Global	84.7%	80.0%	67.8%	1.49	1.90	2.31
Gaussian Partition	84.8%	80.5%	69.0%	1.49	1.84	2.24
Gaussian Local	85.4%	80.7%	70.8%	1.45	1.81	2.05

5   Conclusions

In this paper we defined a novel appearance model for deformable objects. We have shown that an appearance model based on histograms outperforms one based on profiles given one training image. We also described a method to statistically train histogram variation when multiple training images are available, and demonstrated its improved segmentation accuracy. Finally, local image regions were considered and shown to have some benefits over global regions, especially for rectum segmentation.

6   Future Directions

We only present initial segmentation results in this paper. Our next step is to validate these findings in a more comprehensive intra-patient study of the pelvic region. We then plan to consider other anatomical objects.

Experiments have shown our Gaussian appearance model to be sensitive to the number of training samples. We can use the *Mean* model when insufficient intra-patient training data exists to build reliable statistics. Alternatively, we plan to use a Gaussian model with inter-patient statistics for the early images and then transition to intra-patient statistics.

Statistics of the *Local* Gaussian model can also benefit from global training. Global training makes the assumption that image measurements can be clustered into approximately identically distributed groups [15]. Measurements in each group are then pooled for training, allowing stable statistics to be computed using a small training set.

We desire a more principled approach considering tissue composition for defining regions in the *Partition* method. We hope to characterize the intensity distributions of particular tissue types, to estimate the tissue mixtures over image regions using mixture modeling, and finally to optimize the regions for maximum homogeneity. In addition, we may train on the object-relative position of these regions, to help capture inter-object geometric statistics.

Currently, we classify voxels as being inside or outside of each region, and assign equal weights to those considered inside. In the future, the contribution of each voxel will be assigned a weight for each regional measurement. For example, this could be a Gaussian weighting, based on a voxel's distance to the object's boundary. Assigning these weights should smooth the segmentation objective function, resulting in a more robust optimization. Multi-scale regions can also be used to provide a smoother objective function. Such an approach could start with global interior and exterior regions for each object and then progressively define more local regions.

We only consider histograms of pixel intensities in this paper. An extension is to estimate the distribution of additional features, such as texture filter responses or Markov Random Field estimates. Although the EMD defines a distance measure between multi-dimensional distributions, we plan to assume the independence of these features and then apply the same techniques described in this paper.

Acknowledgements

References

1. R.E. Broadhurst. Simplifying Texture Classification. UNC MIDAG Technical Report, http://midag.cs.unc.edu, 2004.
2. T. Chan and L Vese. Active contours without edges. IEEE Trans. Image Processing, vol. 10(2), pp. 266-277, Feb. 2001.
3. T.F. Cootes, C.J. Taylor, D.H. Cooper, and J. Graham. Active Shape Models Their Training and Application. Computer Vision and Image Understanding, Elsevier, Vol. 61(1), pp.38-59, 1995.
4. T.F. Cootes, G.J. Edwards, and C.J. Taylor. Active Appearance Models. In proc. ECCV, 1998.
5. P.T. Fletcher, C. Lu, S.M. Pizer, and S. Joshi. Principal Geodesic Analysis for the Study of Nonlinear Statistics of Shape. IEEE Trans. on Medical Imaging, vol. 23(8), pp. 995-1005, Aug. 2004.
6. D. Freedman, R.J. Radke, T. Zhang, Y. Jeong, D.M. Lovelock, and G.T.Y. Chen, Model-Based Segmentation of Medical Imagery by Matching Distributions. Accepted to IEEE Trans. on Medical Imaging, August 2004 (revised version to appear).
7. S. Joshi. Large Deformation Diffeomorphisms and Gaussian Random Fields for Statistical Characterization of Brain Submanifolds. PhD thesis, Washington University, 1997.
8. E. Levina, P. Bickel. The Earth Movers Distance is the Mallows Distance: Some insights from statistics. In Proc. ICCV, pp. 251-256, 2001.
9. E. Levina. Statistical Issues in Texture Analysis. Ph.D. Thesis, UC Berkley, 2002.
10. S.M. Pizer, T. Fletcher, Y. Fridman, D.S. Fritsch, A.G. Gash, J.M. Glotzer, S. Joshi, A. Thall, G Tracton, P. Yushkevich, and E.L. Chaney. Deformable M-Reps for 3D Medical Image Segmentation. IJCV, vol. 55(2), pp. 85-106, Kluwer Academic, Nov.-Dec. 2003.
11. S.M. Pizer, P.T. Fletcher, S. Joshi, A.G. Gash, J. Stough, A. Thall, G. Tracton, and E.L. Chaney. A Method & Software for Segmentation of Anatomic Object Ensembles by Deformable M-Reps. To appear, Medical Physics, 2004.
12. J. Puzicha, Y. Rubner, C. Tomasi, and J. M. Buhmann. Empirical evaluation of dissimilarity measures for color and texture. In Proc. CVPR, 1999.
13. Y. Rubner, C. Tomasi, and L. J. Guibas. A metric for distributions with applications to image databases. In Proc. ICCV, pp. 59-66, 1998.
14. I.M. Scott, T.F. Cootes, C.J. Taylor. Improving Appearance Model Matching Using Local Image Structure. In proc. IPMI, 2003.
15. J. Stough, S.M. Pizer, E.L. Chaney, and M. Rao. Clustering on Image Boundary Regions for Deformable Model Segmentation. In proc. ISBI, pp. 436-439, Apr. 2004.
16. A. Tsai, A. Yezzi, W. Wells, C. Tempany, D. Tucker, A. Fan, W.E. Grimson, and A. Willsky. A Shape-based approach to the segmentation of medical imagery using level sets. IEEE Trans. Medical Imaging, 22(2), Feb. 2003.