International Workshop on Deep Structure, Singularities and Computer Vision (DSSCV 2005): 72-83

Histogram Statistics of Local Model-Relative Image Regions

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 to statistically characterize histograms of model-relative image regions. A multiscale model is used as an aperture to define image regions at multiple scales. We use this image description to define an appearance model for 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 model-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 regional histograms at a fixed local scale over a fixed global scale.

1 Introduction

Multiscale image descriptors are important for understanding and segmenting deep structures in images. Deformable geometric models have also been shown to be a powerful tool for segmentation. Geometric models generate model-relative image descriptors, which are often used in the human visual system and whose importance is argued in the companion paper by Pizer et al [11]. In this paper, we use a multiscale model-relative image description for the segmentation of 3D deformable objects in medical images. Automatic segmentation methods that statistically learn the likelihood of an object given an image have several desirable qualities. We define an image likelihood measure using non-parametric histograms as our basic image measurement and describe a new method to statistically learn their variation. These histograms are measured in model-relative regions defined at a particular scale using the geometric model as an aperture.

Appearance models at extremely local scale levels are based on the correlation of pixel intensities. Intensities are acquired along profiles normal to the object boundary [4, 15] or from entire model-relative image regions [3, 6]. These methods can be used in conjunction with image filters to summarize information at a larger spatial scale and to 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, which are at larger spatial scales, 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. Histograms provide a rich estimate of a region's intensity distribution but previous work only specifies a reference histogram and not its expected variation [5].

In this paper, we use a region based method that defines several modelrelative regions. This allows a multiscale image description that can be used at a large scale level with one or two global regions defined per object, or at more local scale levels with many smaller regions per object. We segment images using this image description at three fixed scale levels. First, we use global image regions as in previous methods. Then, we describe two approaches to define increasingly local regions. These novel local region approaches have the advantage of histogram measurements with increased locality and tighter distributions, which help 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) [9, 10]. To form a statistical description of each region, we map non-parametric histograms to points in Euclidean space using the Earth Mover's distance (EMD) [1, 7, 13]. Then, we apply standard statistical tools to model histogram variation. 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 local regions to reduce distribution variability. Therefore, we have an image descriptor that can model any intensity distribution while maintaining tightness using regions at an appropriate scale.

Appearance models allow two simplifying assumptions when defining the probability of an image 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 that model-relative images have intensities with no further probabilistic dependence on object shape. The probability of a model-relative image is determined using several image measurements, which are also often assumed to be independent. However, local measurements are highly interrelated due to their small scale 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 [3]. On the other hand, as argued in the companion paper [11] we can reasonably assume that larger scale regional measurements of a model-relative image are independent, if the image is divided into anatomically based local regions and geometric variation is entirely captured by the shape prior.

Thus, we assume regional image measurements relative to object shape are 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 fully train 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 and consider properties of the resulting space. In section 2.2 we define the likelihood of a histogram. In section 2.3 we provide an example.

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 Mallows distance [8]. The EMD representation we use is described for texture classification in [7] and used to build statistical models in [1].

The EMD, and the Mallows 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 Mallows distance between continuous one-dimensional distributions q and r, with cumulative distribution functions Q and R, respectively, is defined as

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

For example, consider the Mallows distance between two Gaussian distributions $N(\mu_1, \sigma_1^2)$ and $N(\mu_2, \sigma_2^2)$. For p = 2, this distance can be shown to be $\sqrt{(\mu_1 - \mu_2)^2 + (\sigma_1 - \sigma_2)^2}$.

For discrete one-dimensional distributions, 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 represented as vectors $\overline{x} = n^{-1/p} * (x_1, \ldots, x_n) = (x'_1, \ldots, x'_n)$ and $\overline{y} = n^{-1/p} * (y_1, \ldots, y_n) = (y'_1, \ldots, y'_n)$ with $x_1 \leq \ldots \leq x_n$ and $y_1 \leq \ldots \leq y_n$. The Mallows distance between x and y is then defined as the L_p vector norm between \overline{x} and \overline{y}

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

Therefore, this representation maps histograms to points in *n*-dimensional Euclidean space in which distances are understood as M_2 histogram distances. In this space, there is a particular straight line path of interest. The mean of any histogram can be changed by an arbitrary amount by adding this amount to every bin in the histogram. Since the mean of a histogram represents its position, changes in histogram position are orthogonal to changes in shape.

Another property of this space is that Gaussian distributions exist in a linear two-dimensional subspace. As for general distributions, one axis of this space represents the Gaussian's mean. As shown above, the remaining orthogonal direction is linear in the Gaussian's standard deviation.

Points in a convex portion of this space represent valid histograms. That is, a point \overline{x} is a valid histogram if and only if $x_1 \leq \ldots \leq x_n$. Therefore, the mean of a set of histograms, or any interpolated histogram, will always be valid. In the next section, the likelihood of a histogram is computed assuming that the mean of a set of histograms and straight-line paths from the mean are representative of the input set. In section 2.3 we demonstrate this with an example.

2.2 Histogram Likelihood

In this section, we statistically define a histogram's likelihood. We can use standard statistical tools for this task since we have sensibly mapped histograms to Euclidean space. For each region, we construct a multi-variate Gaussian model as a parametric estimate of a histogram's likelihood. Gaussian models stretch space, modifying the M_2 metric, to account for the variability in the training data. Thus, Gaussian models naturally enhance the M_2 metric even though they are not proper in the sense that points representing invalid histograms are assigned a non-zero probability.

When constructing a multi-variate Gaussian model, we cannot estimate a full covariance matrix since we are in a high dimension low sample size situation. This is a standard problem in medical imaging since large training sets are often unavailable, which are required to accurately estimate the covariance of a model containing a desirable number of histogram bins. Therefore, we estimate a non-singular covariance of the form

$$k = \sum_{i=1}^{m} U_i U_i^T + \sigma I$$

where each U_i is a vector and I is the identity matrix. We compute the maximum likelihood estimate of k for a fixed m given the training histograms in each region. This estimate can be computing using PCA. The U_i vectors correspond to the principal directions with the m largest eigenvalues, λ_i . These vectors are scaled by $\lambda_i - \sigma$. σ corresponds to the average squared projection error normalized by the number of remaining dimensions.

As discussed in the companion paper [11], regions contain incorrectly labeled voxels as a consequence of the object model having its own scale. When collecting training histograms for each region, we remove such voxels. This allows us to model the true variability in each region and to define a more accurate optimum for segmentation. This approach does not, however, take into account the expected variation of the actual training segmentations. This can result in a covariance estimate that biases segmentations towards either the object interior or exterior. Therefore, we create an unbiased covariance estimate by normalizing each covariance matrix such that the average Mahalanobis distance of the training histograms is the same in each region.

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 a single patient. The interior and exterior of the bladder, prostate, and rectum, within 1 cm of each boundary, define six global regions. For each region, figure 1 shows the 17 25 bin histograms. In general, the interior of the bladder, which consists of bladder wall and urine, has higher CT values than its exterior. The bladder exterior consists of fatty and prostate tissue, with the heavy tail representing the latter. We only model a portion of the rectum and hence its exterior contains interior rectum intensities, making the exterior rectum histogram bimodal.

For each region, we compute the mean of the 17 histograms, m = 2 principal directions of variation, and σ . Figure 2 shows each region's mean and ± 1.5 standard deviations along each principal direction from the mean. The mean and each mode appear 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. In section 3.2 we then present segmentation results using these global image regions.



Fig. 1. Histograms from interior (red) and exterior (blue) bladder, prostate, and rectum regions in 17 images of the same patient.

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. As detailed in the companion paper [12], 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 mrep deformation over a training set, described in the companion paper [12]. We use the method described in section 2 for the image match.

In this paper, our primary concern is to determine the quality of the image likelihood optimum defined by our appearance model. We evaluate this 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 the patient's treatment for prostate cancer. As an initial test of our framework, we segment each image using a leave-one-out strategy, which supplies sufficient training data to estimate adequate and stable statistics. We estimate the model prior and likelihood using m-reps fit to manual segmentations of the training images. We gather shape statistics for the combined bladder, prostate, and rectum object ensemble and define a shape space using six principal geodesics, which captures approximately



Fig. 2. Histograms representing the mean of 17 interior and exterior regions. Shown along with each mean is ± 1.5 standard deviations along the first (left) or second (right) principal direction from the mean (slightly smoothed). The first mode often contains more tail and less peak movement than the second mode. Some of these tail movements have been cropped out of the graphs.

94% of the shape variance. We ignore the model prior and perform a maximum likelihood segmentation within the shape space.

We compare our segmentation 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 space 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 of our appearance model.

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 region using a 25 bin equi-count histogram.

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 the sum of M_2 distances 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 other images. In this case, the image match is the sum of M_2 distances to each average histogram. The last model, Gaussian Global, uses the fully trained likelihood measure introduced in section 2.2. The image match for this model is the sum of Mahalanobis distances in each Gaussian model. Each model independently learns two principal directions of variation and σ .

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 with increased statistical training. 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 all three objects. In the next section we further improve these results using local image regions.

4 Defining Local Image Regions

Next, we use the appearance model described in section 2 with local modelrelative 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

	Volume Overlap			Ave. Surface Dist. (mm)			
Appearance Model	Bladder	Prostate	Rectum	Bladder	Prostate	Rectum	
Training	88.6%	87.8%	82.8%	1.11	1.05	1.15	
Profile	79.8%	76.0%	64.8%	2.07	2.20	2.72	
Simple Global	80.7%	78.4%	67.1%	1.97	1.94	2.47	
Mean Global	81.8%	79.4%	68.0%	1.84	1.86	2.42	
Gaussian Global	84.8%	79.6%	72.1%	1.53	1.86	2.00	

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.

estimates. They also require a shape model that defines a voxel correspondence near the object boundary.

Our dataset contains at least 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. A 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. In section 4.2 we define overlapping regions centered around many boundary points. In section 4.3 we give results using both methods.

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. We used several heuristics to create our manual partitions, which are shown in figure 3. 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.



Fig. 3. 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.

4.2 Local Image Regions

An alternative method to define local regions is to consider a set of boundary points that each describe the center of a region. Define an interior and exterior region for each point by first finding the portion of the surface within a radius of each point. Then, each region consists of all the voxels within a certain distance to the boundary that have model-relative coordinates associated with the region's corresponding surface patch. This approach can define overlapping image regions at any scale and locality, and learning boundaries between local regions is unnecessary.

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 ± 1 cm, as in previous results.

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 described in section 4.1, and the *Local* method refers to the approach described in section 4.2. Both methods use 25 histogram bins and Gaussian models restricted to 2 principal directions of variation. These results show that both the *Local* and *Partition* methods are roughly equivalent to the *Global* method. However, there is a consistent improvement by the *Local* method in the segmentation of the rectum.

5 Conclusions

In this paper we defined a novel multiscale appearance model for deformable objects. We have shown that our histogram based appearance model outperforms a

Table 2	2. Segm	entation	$\operatorname{results}$	using	local	image	regions	. The	Gaussian	appearance	Э
model us	sing the	two loca	al region	meth	ods is	compa	ared to	the glo	bal regio	n method.	

	Vo	lume Overl	ар	Ave. Surface Dist. (mm)			
Appearance Model	Bladder	Prostate	Rectum	Bladder	Prostate	Rectum	
Training	88.6%	87.8%	82.8%	1.11	1.05	1.15	
Gaussian Global	84.8%	79.6%	72.1%	1.53	1.86	2.00	
Gaussian Partition	83.0%	80.5%	72.1%	1.74	1.77	2.01	
Gaussian Local	83.2%	80.5%	73.0%	1.67	1.78	1.95	

profile based appearance model for a segmentation task when only one training image is available. We also described a method to statistically train histogram variation when multiple training images are available and demonstrated its improved segmentation accuracy. Finally, we considered regions at different scales and showed that local image regions 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. Then, we plan to consider other anatomical objects including the kidneys.

In the pelvic region, gas and bone produce outlying CT values. When there is a significant amount of these extreme values our mapping can produce unnatural interpolations. Therefore, we will investigate a technique to identify these intensities in advance and compute a separate estimate of their variation.

As described in [11], we plan to do a multiscale optimization. Such an approach could use the three region scales described in this paper. Furthermore, we will use geometric models to describe soft instead of hard apertures. For example, a voxel's contribution to a measurement could be weighted by a Gaussian, based on its distance to the object's boundary. Using multiscale regions and soft apertures should smooth the segmentation objective function, resulting in a more robust optimization.

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 model-relative position of these regions, to help capture inter-object geometric statistics.

We only considered histograms of pixel intensities in this paper. We will extend this framework 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

We thank J. Stephen Marron for discussions on histogram statistics, Sarang Joshi for discussions on Gaussian parameter estimation, and the rest of the MIDAG team for the development of the m-rep segmentation framework. The work reported here was done under the partial support of NIH grant P01 EB02779.

References

- 1. R. E. Broadhurst. Simplifying texture classification. Technical report, University of North Carolina, http://midag.cs.unc.edu, 2004.
- T. Chan and L. Vese. Active contours without edges. In *IEEE Trans. Image* Processing, volume 10, pages 266–277, Feb. 2001.
- T. F. Cootes, G. J. Edwards, and C. J. Taylor. Active appearance models. In ECCV, 1998.
- T. F. Cootes, C. J. Taylor, D. H. Cooper, and J. Graham. Active shape models their training and application. In *Computer Vision and Image Understanding*, volume 61, pages 38–59, 1995.
- 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. In *IEEE Trans. on Medical Imaging*, volume 24, pages 281–292, Mar. 2005.
- S. Joshi. Large Deformation Diffeomorphisms and Gaussian Random Fields for Statistical Characterization of Brain Submanifolds. PhD thesis, 1997.
- 7. E. Levina. Statistical Issues in Texture Analysis. PhD thesis, 2002.
- 8. E. Levina and P. Bickel. The earth movers distance is the mallows distance: Some insights from statistics. In *ICCV*, pages 251–256, 2001.
- 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. *Medical Physics*, To appear.
- 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*, 55(2):85–106, 2003.
- 11. S. M. Pizer, J. Y. Jeong, R. E. Broadhurst, S. Ho, and J. Stough. Deep structure of images in populations via geometric models in populations. In *DSSCV*, 2005.
- S. M. Pizer, J. Y. Jeong, C. Lu, K. Muller, and S. Joshi. Estimating the statistics of multi-object anatomic geometry using inter-object relationships. In DSSCV, 2005.
- Y. Rubner, C. Tomasi, and L. J. Guibas. A metric for distributions with applications to image databases. In *ICCV*, pages 59–66, 1998.
- 14. I. M. Scott, T. F. Cootes, and C. J. Taylor. Improving appearance model matching using local image structure. In *IPMI*, 2003.
- J. Stough, S. M. Pizer, E. L. Chaney, and M. Rao. Clustering on image boundary regions for deformable model segmentation. In *ISBI*, pages 436–439, Apr. 2004.
- 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. In *IEEE Trans. Medical Imaging*, volume 22, Feb. 2003.