DISCRIMINATIVE OPTIMIZATION OF 3D SHAPE MODELS FOR THE GENERALIZED HOUGH TRANSFORM

ANA BELÉN MARTÍN RECUERO[1], PETER BEYERLEIN[2], HAUKE SCHRAMM[3]

Abstract. To achieve a high level of automation in medical image processing, techniques for automatic detection of anatomical objects are required. Recently, it has been shown that the Generalized Hough Transform (GHT)[1], a technique widely used for 2D object detection, can also be successfully applied to 3D images. The central knowledge source of the GHT is a usually manually generated shape model, describing the shape of the considered object as a set of points. In this work we outline an automatic procedure for generating efficient and discriminative shape models for usage in GHT-based object detection. The technique (1) splits the N shape model points into N individual knowledge sources, (2) recombines them into a Maximum Entropy Distribution and (3) optimizes their individual weights in this distribution using Minimum Classification Error Training (MCE). By this, an individual weighting of model points with respect to their importance for the object detection task is achieved. Since the technique estimates positive and negative weights, the resulting shape model captures both the shape of the considered object (points with positive weights) as well as the shape of confusable structures or anti-shapes (points with negative weights). Since unimportant points can be identified by their low absolute weight and removed from the shape model, it is possible to learn shapes from scratch, e.g. an initial random point cloud. First results with this technique will be presented, showing that an efficient 100 point shape model for femur detection can be learned from scratch by using only three training images.

1. Introduction

An automatic procedure for detecting and segmenting anatomical objects in 3-D images is necessary for achieving a high level of automation in many medical applications. Recently it has been shown [2] that the GHT can be used for a coarse 3-D delineation of anatomical objects with well defined shape in medical images. The computational and memory requirements of the GHT are in general very large, especially in case of considering 3-D images and various free transformation parameters. Although it is possible to substantially limit its algorithmic complexity by restricting the number of transformation parameters and their quantization, further speed-up is required to allow for application in interactive user environment. The number of considered image points (voxels) and the number of model points directly influence the processing time and memory requirements. Besides, shape models might partially match other concurrent shapes in the image appearing the possibility of a wrong object detection in favor of the concurrent shapes. To this end, techniques are required which allow for a "smart" selection of the

Key words and phrases. Generalized Hough Transform, Medical Image Analysis, Object Detection, Shape Model, Minimum Classification Error Training, Maximum Entropy Distribution.

points representing the target object to improve the trade-off between computational complexity of the GHT and the object detection accuracy. Another aspect of the GHT is the cumbersome user interaction to manually delineate the target object to create a specific model of each object to detect. This paper outlines an automatic procedure to create 3-D models from scratch, reducing the user interaction to a minimum and which reduces the computational complexity of the GHT and provides discrimination capabilities to the model, thus improving the detection accuracy.

2. The Generalized Hough Transform

The GHT employs the shape of an object to transform a feature (i.e. edges of an image) into a multi-dimensional function of a set of unknown object transformation parameters. The maximum of this function over the parameter space determines the optimal transformation for matching the model to the image, that is, for detecting the object. In our framework, the GHT relies on two main knowledge sources:

- Shape knowledge, usually stored as so-called "R-table"
- Statistical knowledge about the grey value and gradient distribution at the edges remaining in an image after applying a preprocessing.

The GHT aims at finding optimal transformation parameters for matching a given shapel model, located for example in the origin of the target image, to its counterpart. To this end, a geometric transformation of the target object $\mathbf{M} = \{\mathbf{p_1^m}, \mathbf{p_2^m}, ..., \mathbf{p_{N_m}^m}\}$ is introduced, defined by $\mathbf{A} \cdot \mathbf{p_i^m} + \mathbf{t}$, with $\mathbf{A}$ being a linear transformation matrix and $\mathbf{t}$ being a translation vector. Each edge point $\mathbf{p_i^e}$, denoted as $\mathbf{E} = \{\mathbf{p_1^e}, \mathbf{p_2^e}, ..., \mathbf{p_{N_e}^e}\}$ corresponds to a given model point $\mathbf{p_j^m}$. The correspondence between an edge and a model point can for example be determined by a comparison of the edge point gradient direction with the target object's surface direction at the position of the model point. Without any knowledge about corresponding points, it must be assumed that *any* edge point corresponds to a given model point. The correspondence is assumed to result from a transformation of the model point according to

$$\mathbf{p_i^e} = \mathbf{A} \cdot \mathbf{p_j^m} + \mathbf{t} \tag{2.1}$$

If, the other way around, we aim at determining the translation parameters $\mathbf{t}$ which may have led to a specific edge point $\mathbf{p_i^e}$, given a corresponding model point $\mathbf{p_j^m}$ and a transformation matrix $\mathbf{A}$, we arrive at

$$\mathbf{t}(\mathbf{p_j^m}, \mathbf{p_i^e}, \mathbf{A}) = \mathbf{p_i^e} - \mathbf{A} \cdot \mathbf{p_j^m} \tag{2.2}$$

This equation can be utilized to perform a brute-force *voting* procedure in order to identify the set of transformation parameters, $\mathbf{t}$ and $\mathbf{A}$, which optimally matches the given model to the image. Therefore, the *detection procedure* works as follows: For each possible corresponding pair $(\mathbf{p_j^m}, \mathbf{p_i^e})$ and each possible setting of the quantized parameters in $\mathbf{A}$ a *voting* for the corresponding translation and linear transformation parameters in the discretized parameter space is done. In case of a good match between the transformed model and the edge image, a large number of pairs $(\mathbf{p_j^m}, \mathbf{p_i^e})$ will vote for the same cell in the multi-dimensional accumulator array. Thus, after processing all model and edge points, the optimal transformation parameters $\mathbf{t}$ and $\mathbf{A}$ are given by the cell with the highest count.

3. Minimum Classification Error Training for Model Optimization

A crucial aspect of the generalized Hough transform, which has rarely been addressed so far, is how to optimally combine the information from different model regions or even points into a single decision function. Thus, it is proposed to log-linearly combine a set of base models, representing (groups of) model points, into a probability distribution of the maximum-entropy family. A minimum classification error training can be applied to optimize the base model weights with respect to a predefined error function. The classification of unknown data might be performed by using an extended Hough model that contains additional information about model point grouping and base model weights. Apart from an increased classification performance, the computational complexity of the Hough transform can be reduced with this shape model.

3.1. **Log-linear model combination.** We consider the classification of image feature observations[1] x_n into a class $k \epsilon \{1, ..., K\}$ using the generalized Hough transform. The class k may represent an object location, or arbitrary transformation parameters. To solve this classification task, a set of M posterior probability base models $p_j(k|x_n)$, $j = 1, ..., M$ is applied. These base model distributions represent single Hough model points or groups of points and may be derived from the Hough space voting result on some training volume data by the relative voting frequencies:

$$p_j(k|x_n) = \frac{N(j,k,x_n)}{\sum_{\forall k'} N(j,k',x_n)} \tag{3.1}$$

Here, $N(j, k, x_n)$ represents the number of votes by model point (or region) j for hypothesis k if the features x_n have been observed. Alternatively, the probability distribution could be estimated by a multi-modal Gaussian mixture.

In the next step, the base models are log-linearly combined into a probability distribution of the maximum-entropy family [3]. This class of distributions ensures maximal objectivity and has been successfully applied in various areas.

$$p_\Lambda(k|x_n) \quad = \quad e^{-\log Z(\Lambda,x_n) + \sum_{j=1}^{M} \lambda_j \log p_j(k|x_n)} \tag{3.2}$$

The value $Z(\Lambda, x_n)$ is a normalization constant with

$$Z(\Lambda, x_n) = \sum_{k'} \exp\left[\sum_{j=1}^{M} \lambda_j \log p_j(k'|x_n)\right] \tag{3.3}$$

The coefficients $\Lambda = (\lambda_1, ..., \lambda_M)^T$ can be interpreted as weights of the models j within the model combination.

As opposed to the well-known maximum entropy approach, which leads to a distribution of the same functional form, our approach optimizes the coefficients Λ with respect to a classification error rate of the following discriminant function:

$$\log \frac{p_\Lambda(k|x_n)}{p_\Lambda(k_n|x_n)} \quad = \quad \sum_{j=1}^{M} \lambda_j \log \frac{p_j(k|x_n)}{p_j(k_n|x_n)} \tag{3.4}$$

In this equation, k_n denotes the correct hypothesis. Since the weight λ_j of the base model j within the combination depends on its ability to provide information for

[1]Note that x_n may represent the features of a complete image or even a set of images.

correct classification, this technique allows for the optimal integration of any set of base models.

3.2. **Minimum classification error training.** Assume, we are given a set of training volumes $n = 1, \dots, H$ with correct class assignment k_n and generate a feature sequence x_n for each volume. By performing a preliminary classification with equal weights (i.e. $\lambda_j = \mathit{const}\ \forall j$), a set of rival classes $k \neq k_n$ can be determined. In order to quantify the classification error for each rival class k, an appropriate distance measure $\Gamma(k_n, k)$ must be selected. Of course, this choice strongly depends on the class definition. In case of a translation classification problem for example, where the solution is a simple 2D or 3D position vector, the Euclidean distance between the correct point and its rival could be used. An even simpler idea is to use a binary distance measure, which is '1' for the correct class and '0' for all others.

The model combination parameters should then minimize the classification error count $E(\Lambda)$

$$E(\Lambda) = \sum_{n=1}^{H} \Gamma\left(k_n, \arg\max_k \left(\log \frac{p_\Lambda(k|x_n)}{p_\Lambda(k_n|x_n)}\right)\right) \tag{3.1}$$

on representative training data to assure optimality on an independent test set. As this optimization criterion is not differentiable, we approximate it by a smoothed classification error count:

$$E_S(\Lambda) = \sum_{n=1}^{H} \sum_{k \neq k_n} \Gamma(k, k_n) S(k, n, \Lambda), \tag{3.2}$$

where $S(k, n, \Lambda)$ is a smoothed indicator function. If the classifier (3.4) selects hypothesis k, $S(k, n, \Lambda)$ should be close to one, and if the classifier rejects hypothesis k, it should be close to zero. A possible indicator function with these properties is

$$S(k, n, \Lambda) = \frac{p_\Lambda(k|x_n)^\eta}{\sum_{k'} p_\Lambda(k'|x_n)^\eta}, \tag{3.3}$$

where η is a suitable constant. An iterative gradient descent scheme is obtained from the optimization of $E_S(\Lambda)$ with respect to Λ [3]:

$$\begin{aligned}
\lambda_j^{(0)} &= 1 \quad \text{(Uniform Distribution)} \\
\lambda_j^{(I+1)} &= \lambda_j^{(I)} - \varepsilon \cdot \eta \sum_{n=1}^{H} \sum_{k \neq k_n} S(k, n, \Lambda^{(I)}) \cdot \\
&\quad \cdot \tilde{\Gamma}(k, n, \Lambda^{(I)}) \cdot \log \frac{p_j(k|x_n)}{p_j(k_n|x_n)} \\
\Lambda^{(I)} &= (\lambda_1^{(I)}, \dots, \lambda_M^{(I)})^T \qquad (3.4) \\
&\quad j = 1, \dots, M \\
\tilde{\Gamma}(k, n, \Lambda) &= \Gamma(k, k_n) - \sum_{k' \neq k_n} S(k', n, \Lambda) \Gamma(k', k_n).
\end{aligned}$$

This iteration scheme reduces the weight of model points or groups which favor weak[2] hypotheses while increasing the weight of base models which favor good hypotheses.

3.3. **Classification with extended Hough model.** The classification of an unknown volume data set is performed with an extended Hough model, that incorporates information about the capability of each model point to detect the object in the correct position. This is represented as model point weights,(as obtained from minimum classification error training). The classification algorithm proceeds as follows:

1. Apply GHT using input features x to fill the Hough space accumulator.
2. Determine $p_j(k|x)$ for all base models j and classes k using the accumulator information (e.g. with equation (3.1)).
3. Compute the discriminant function (3.4) for each class k with the λ_j obtained from minimum classification error training.
4. Decide for the class with highest discriminant function.

3.4. **Model Point Selection.** Through a model point selection, we aim at reducing the computation complexity of the GHT, and at the same time improving or at least maintaining the classification performance. We decide to eliminate model points that are not decisive for the classification, that is model points that are not discriminant. Now, the issue is how to define the discrimination potential of a model point. The model points with higher discrimination potential are those that only vote for the cell k_n or in turn for a cell k with a low loss $\Gamma(k, k_n)$ (1). Through our investigations, we realized that model points with high number of votes in cells k but low number of votes in the cell k_n also obtain large absolute λ values (2).

This fact can directly been observed in the iteration scheme for the calculation of the Λ (equation 3.4). Assume that a solution cell k is considered a good[3] hypothesis of the GHT and therefore is selected by the smoothing function (equation 3.3) during the training. In this case, the model point dependent contributions will be re-estimated according to the difference in number of votes between the selected cell k and k_n, as well as the distance measure $\Gamma(k, k_n, \Lambda)$.

For the first kind of model points (1), the gradient descendent iteration scheme (equation 3.4) increments the value of their λ. For the second kind of model points (2), a decrement of the λ value of the model point takes place.

On the other hand, the model points with lower discrimination capability are those with more uniform contributions to the Hough Space. The smaller the difference in votes between the cell k and the cell k_n, the more softly the λ value will vary.

Taking into account these reasonings, we define a discriminative model point for the GHT as

(1) a model point matching the object or its surroundings, but not matching other remaining shapes/edges in the image. The λ of these model points are positive.

[2]Weak means that the distance to the correct hypothesis is large.

[3]The smoothing function selects a cell k according to their number of votes with respect to the whole Hough Space. The number of considered cells k can be controlled through the parameter η (equation 3.3).

(2) a model point exclusively matching concurrent shapes. The λ of these points are negative.

Both of these model points obtain large absolute λ values.

It is remarkable, that a very negative weight applied to our model point dependent contributions to the Hough Space would also mean an improvement in the discrimination of our object detection task. Explained in a different way, this model point is providing information for the classification in two different ways:

(1) related to the concurrent objects in the image. We have to consider that these model points cause high number of votes in concurrent solution cells. The probability of deciding for this cell as the best solution of the object detection task decreases along with the negative contribution of this model point to the solution k. In this way, the probability of a misclassification decreases.
(2) related to the target shape. That is, the cell k_n or in turn cells with low cost $\Gamma(k, k_n, \Lambda)$. The probability of the classification of this cell as the best solution of the object detection will increase when the number of votes coming from this model point in the present cell is low. Note the effects of normalization, a negative value of the normalized feature multiplied by the negative value of the model point weight is equal to a positive increment to be added to the cell. Consequently, the probability to classify this cell as the best hypothesis of the object detection task will increase.

This observation leads to a shape model where the model points included in it, do not only belong to the object to detect, but also to other shapes remaining in the preprocessed image.

4. Automated Generation of Shape-Variant Hough Models

Presently, the generation of shape models for the generalized Hough transform requires substantial user interaction and has to be repeated each time a new shape is introduced. The ideas described in this section aim at improving this situation by introducing an automatic determination process for shape-variant Hough models that requires only minimal user interaction. Technically, it is based on an random model initialization in combination with an appropriate (discriminative) weighting of Hough model points. With this method, the generation of a new shape-variant Hough model is simply done by labeling the location of this shape in a small set of training volumes. The generated model will be well adapted to all training shapes and therefore incorporates the shape variability of the whole training data set.

The algorithm for automatic generation of shape-variant models proceeds as follows:

1. Apply feature detection (e.g. Sobel edge detection) on all training volumes
2. For each training volume: ask user for object location
3. Generate a spherical random scatter plot of model points using two input parameters: (1) number of points, (2) concentration decline in dependence of the distance to the center. Or use as input for the cloud of model points the remaining edge points after applying the pre-processing of the image.
5. Run the iterative discriminative model point weighting procedure, described in Section 3
6. Run an iterative selection of model points based on their absolute weight value as explained in 3.4 followed by a re-estimation of the discriminative

model point weighting procedure (previous point), until a given stop criterion is reached (i.e. failure in the posterior segmentation task).

The generated shape-variant model and its model weights can directly be used in a GHT-based classification.

5. Experiments

We set up an experiment with an initial cloud of 10000 model points. The location of the right femur (object to detect) is given in three Computed Tomography (CT) training images. The training and test images include the pelvic zone and the left femur. The images are 512x512 voxels with a variable number of slices (between 54 and 92). The voxel extension is 0.94, 0.94 and 3 mm in x, y and z direction respectively. We run the iterative discriminative model point weighting procedure and iteratively remove model points based on their smallest absolute value. We evaluated the system on seven independent test images and on the training data. The selection of model points has been performed according to the largest absolute value. We measured the Euclidean distance to the ground truth in Hough Space units for the test data. The table 5 presents the results for different number of model points selected from an initial number of 10000 random model points. The means and standard deviations are presented in the table. We observe a better performance while reducing model points until reaching 100 model points. The larger the distance error measure the worse performance of the classifier and the more difficult to perform a posterior segmentation.

Model points	1000	500	250	100	50
Euclidean distance	1.38±0.30	1.38±0.30	1.28±0.26	0.88±0.6	2.66±3.46

Table 1. Mean and standard deviation of validation metrics for seven patients.

A successful segmentation was possible given the outcome of the Discriminative Generalized Hough Transform (DGHT) using optimized shape models with 100 model points in nine of the ten images tested. In the failed case, only a little misplacement was observed.

We applied the GHT and DGHT with 1000 model points from the random cloud (see figure 1). The GHT hypothesized solution is far away from the ground truth The Hough Space is more uniform for the case of the DGHT and the solution is located in the right position.

6. Conclusions

We introduce a new concept of weighted positive and negative model point contributions for the GHT. This approach supposes the introduction of a new discriminative object detection technique. The learned model contains model points referring to the target object: "shape" with positive contributions to the Hough Space and model points referring to other concurrent objects: "anti-shape" with negative contributions reducing the probability of detecting the object in a concurrent wrong location. The absolute value of the weight refers to the capability of a model point in the shape model to discriminate between different solutions. Based on this assumption, it is possible to iteratively exclude model points with

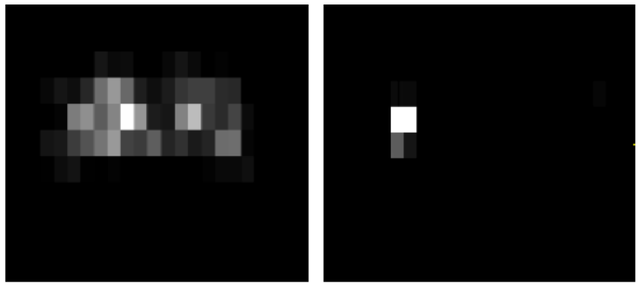

Figure 1. Hough Space after applying the GHT with standard shape model and DGHT with 1000 random model points. The location solution corresponds to the brighter cell.

the smallest absolute weight until a stop criterion is reached. The learn-shape is based on the matches of the shape model to the remaining structures in the image. Therefore, a fixed preprocessing must always be applied. The generation of shape models from scratch incorporates in the model other remaining image structures to support the object detection. This way, more information is used to detect the target. The results presented seem promising. However, more training data and crosslink validation are necessary to support the first evaluation experiments. We demonstrate a considerable improvement compared with a random shape model with uniform model point contributions. Thus, this procedure makes feasible the generation of shape models from scratch and additionally allows for detecting objects discriminatively. A focused Hough Space is achieved compared to the GHT using a standard shape model. The DGHT increases detection accuracy for similar GHT computational complexity.

Acknowledgment. The authors wish to thank Vasanth Philomin for consultancy on implementing the generalized Hough transform and many fruitful discussions, Jochen Peters and Olivier Ecabert for providing their model based segmentation environment and general help and J. von Berg and C. Lorenz (Philips Research, Hamburg) for providing the mesh model of the heart.

References

1. Ballard H. Generalizing the Hough transform to detect arbitrary shapes. Pattern Recognit 1981;13:111-22
2. Schramm H, Ecabert O, Peters J, Philomin V, Weese J. Toward fully automatic object detection and segmentation. In: Reinhardt JM, Pluim JP, Josien PW, editors. Medical Imaging 2006: Image Processing. Proc. SPIE 2006 March;6144:11-20.
3. Beyerlein P. Discriminative Model Combination. Proceedings of the IEEE Workshop on Automatic Speech Recognition and Understanding, Santa Barbara, USA; 1997 Dec. p. 238-45.
4. Beyerlein, P. Discriminative Model Combination. Proceedings of the International Conference on Acoustics, Speech, and Signal Processing (ICASSP); 1998 May 5-15, Seattle, WA, USA. 1998. Vol 1 p. 481-4.
5. Beyerlein B. DMC-based Density Parameter Estimation for Classification in Speech Recognition and Medical Informatics. 2006 to be published.
6. Cootes TF, Taylor CJ, Cooper H, Graham J. Active shape models-their training and application. Comput Vis Image Underst 1995;61(1):38-59.

7. Darroch JN, Ratcliff D. Generalized Iterative Scaling for Log-Linear Models. Annals of Mathematical Statistics 1972;43(5):1470-80.
8. Deselaers T, Keysers D, Ney H. Discriminative Training for Object Recognition using Image Patches. CVPR 2005, International Conference on Computer Vision and Pattern Recognition, San Diego, CA; 2005 Jun. vol 2 p. 157-62.
9. Deselaers T, Keysers D, Ney H. Improving a Discriminative Approach to Object Recognition using Image Patches. DAGM 27th Symposium on Pattern Recognition, Vienna, Austria, 2005. vol 3663 p. 326-33.
10. Jaakkola T, Meila M, Jebara T. Maximum Entropy Discrimination. In: Adavances in Neural Information Processing Systems 12, MIT Press, Cambridge, MA; 2000. p. 470-6.
11. Keysers, D.Modognition. PhD. diss. RWTH Aachen, Computer Science Department, Aachen, Germany, 2006.
12. Keysers D, Ney H. Linear Discriminant Analysis and Discriminative Log-linear Modeling. ICPR 2004, International Conference on Pattern Recognition, Cambridge, UK; 2004 Aug. vol 1 p.156-9.
13. Lazebnik S, Schmid C, Ponce J. A Maximum Entropy Framework for Part-Based Texture and Object Recognition. Proceeding of the IEEE on International Conference on Computer Vision, Beijing, China; 2005 Oct. vol 1 p. 832-8.
14. Lee k, Stree WN. Generalized Hough Transforms with Flexible Templates. Proceeding of the Int. Conf. on Artificial Intelligence (IC -AI), Las Vegas, NV; 2000 Jun. vol 3 p. 1133-9.
15. Macherey W, Keysers D, Dahmen J, and Ney H. Improving Automatic Speech Recognition Using Tangent Distance. Eurospeech 2001, 7th European Conference on Speech Communication and Technology. Aalborg, Denmark; 2001 Sep. Vol 3 p. 1825-8.
16. Ney H. On the Probabilistic Interpretation of Neural Network Classifiers And Discriminative Training Criteria. IEEE Trans Pattern Anal Mach Intell 1995 Feb;17(2):107-119.
17. Nigam K, Lafferty J, McCallum A. Using Maximum Entropy for Text Classification. IJCAI-99 Workshop on Machine Learning for Information Filtering, Stockholm, Sweden; 1999. p. 61-7.
18. Weese J, Kaus R, Lorenz C, Lobregt S, Truyen R, Pekar V. Shape constrained deformable models for 3D medical image segmentation. In: Insana MF, Leahy RM, editors. Image Processing in Medical Imaging. Heidelberg (Berlin):Springer;2001. p. 380-7. (LNCS; vol 2082).
19. Xu, L. Oja, E, Kultanen, P. A new curve detection method:Randomized Hough Transform (RHT). Pattern. Recognit 1990;11(5)631-5.

[1] Philips Research Europe - Aachen, [2] Wildau University of Applied Sciences, Germany, [3] Germany, Kiel University of Applied Sciences, Germany