Customized Hough Transform for Robust Segmentation of Cervical Vertebrae from X-Ray Images

Abraham Tezmol[1], Hamed Sari-Sarraf [1], Sunanda Mitra[1], Rodney Long[2], Arunkumar Gururajan[1]

[1]*Department of Electrical Engineering*
Texas Tech University
hamed.sari-sarraf@coe.ttu.edu

[2]*National Library of medicine*
long@nlm.nih.gov

Abstract

This report addresses the issues involved in developing a robust segmentation technique capable of finding the location and orientation of the cervical vertebrae in x-ray images. This technique should be invariant to rotation, scale, noise, occlusions and shape variability. A customized approach, based on the Generalized Hough transform (GHT), that captures shape variability and exploits shape information embedded in the accumulator structure to overcome noise and occlusions is proposed. This approach effectively finds estimates of the location and orientation of the cervical vertebrae boundaries in digitized x-ray images.

1. Introduction

It is well known that GHT has the ability to locate occurrences of a previously defined template in a target image regardless of noise and occlusions and variations in orientation and scale [1,2], which is the scenario in most x-ray images of the cervical vertebrae. Although GHT is a promising candidate to achieve good segmentation, much of the success of this technique depends on three parameters: (1) gradient information, (2) representativeness of the template, and (3) reckoning of the votes in the accumulator structure.

GHT uses an edge image to correlate points in a previously defined template to those in the target image using local gradient information. If a good edge image is not available, too few votes will be correlated to the template. Typical implementations of GHT assume that a clear edge image, from which gradient information can be obtained, is available. Since this is not the case in x-ray images, an edge detection process is used. A customized process performs Gaussian filtering on a copy of the original image and the result is subtracted from the original. From the resulting edge image, extraction of the gradient information is performed using optimum gradient operators.

Success of GHT greatly depends upon the representativeness of the chosen template. Even if good gradient information is available, the template must adequately represent the target object in order to obtain the necessary votes in the accumulator. Across a large set of images, it is common to find great variability in shape of the cervical vertebrae.

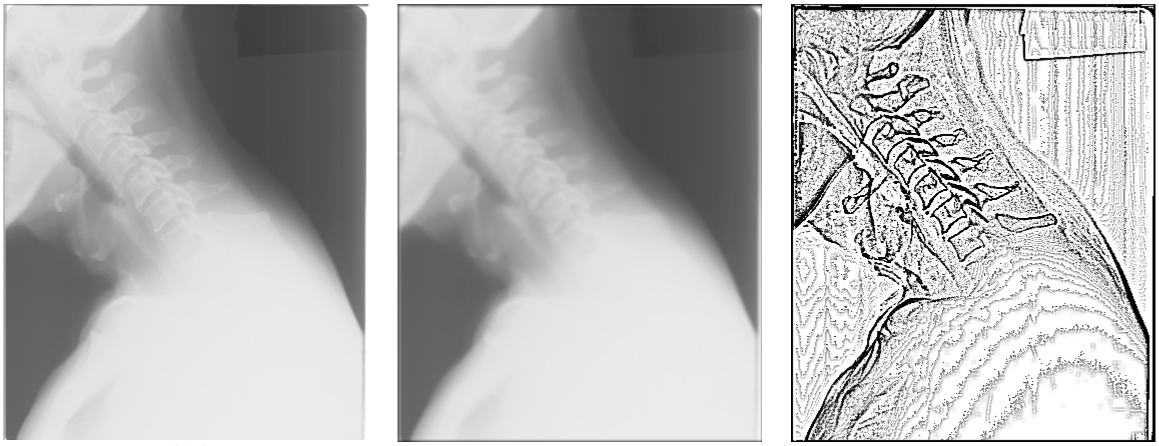
The next and very important step in GHT is the reckoning of votes in the Hough domain. Finding the best estimates of orientation and location of the cervical vertebrae is a direct consequence of the above step. Typical implementations of GHT use a simple criterion: the best estimates correspond to the bin with the largest number of votes. In the proposed approach, post processing of the accumulator structure is carried out and this leads to valuable information about the shape and orientation of the cervical vertebrae. This analysis is described in detail in the later sections of this report.

2. Customized Hough Transform

There are various modifications of the Hough transform which reduce the computational complexity like the fast Hough transform [3] and the adaptive Hough transform [4].

This section describes in detail, the customized approach based on GHT, which captures shape variability and exploits shape information embedded in the Hough domain to overcome noise and occlusions. The following subsections describe the different processing steps involved in the computation of GHT and the subtleties introduced in each processing step.

Fifth IEEE Southwest Symposium on Image Analysis and Interpretation (SSIAI'02)
0-7695-1537-1/02 $17.00 © 2002 IEEE



1(a) 1(b) 1(c)

Figure 1. Unsharp masking (a) Original image (b) Gaussian smoothed image and (c) image after unsharp masking

2.1. Extraction of gradient information

Unsharp masking is the proposed main criterion here to obtain an edge image. Mathematically the resulting image after unsharp masking can be represented as

$$f_u(x,y) = f(x,y) - f_b(x,y) \quad (1)$$

where f(x,y) represents the original image [Fig.1(a)], $f_b(x,y)$ represents the Gaussian blurred image [Fig.1(b)], and $f_u(x,y)$ represents the unsharp masked image [Fig. 1(c)].

The fact that a blurred image is subtracted from the original one will provide only abrupt local variations in contrast (edges) and eliminate a considerable amount of the information that is not of interest (background). Then we use an averaging filter to reduce the amount of unwanted high frequency components and binarize the resulting edge image. Finally the Sobel edge operator is passed through the image and the resulting image has thin double edges and though this is not an optimum solution for reducing the number of edge points, it is certainly efficient in terms of computational cost.

Some of the most commonly used discrete gradient operators are the Sobel and Prewitt operators. Both the operators can be used to estimate the gradient information and as can be seen, the discrete implementation of the gradient is only an approximation to the continuous gradient operation. This inconsistency between the discrete and continuous gradient information was studied by Ando [5] and minimized using the gradient descent method. The optimum 3 x 3 gradient operator is given as:

$$\nabla_x = \frac{\partial}{\partial x} \approx \begin{bmatrix} -0.112737 & -0.274526 & -0.112737 \\ 0 & 0 & 0 \\ 0.112737 & 0.274526 & 0.112737 \end{bmatrix}$$

$$\nabla y = \frac{\partial}{\partial y} \approx \begin{bmatrix} -0.112737 & 0 & 0.112737 \\ -0.274526 & 0 & 0.274526 \\ -0.112737 & 0 & 0.112737 \end{bmatrix} \quad (2)$$

Experiments were conducted on a Hough based scheme using the x-ray images for the sake of comparison between the new operators and the traditional gradient operators like Sobel and Prewitt. It was found that the 3 x 3 optimum gradient operators yield the highest peak in the accumulator and the most compact and uniform region surrounding a possible candidate for the reference point.

2.2 Definition of template

GHT is a template-based technique. It looks for instances of previously defined template in the target image. Even if good gradient information is available, the template must adequately represent the target object in order to obtain the necessary number of votes in the accumulator. The use of multiple templates should capture shape variability and closely represent the different classes of shapes seen across the data set. However, the use of a collection of templates may be computationally expensive. Consequently, the selection of a single template that represents the target object is of primary interest.

As a first approximation, we have used the mean of 50 templates that were obtained from the manual landmarking of 50 images [6] provided by the second National Health and Nutrition Examination survey (NHANES II) database. Each one of the 50 templates is made up of 80 landmark points. The mathematical representation of the templates and that of the mean template is as follows:

Fifth IEEE Southwest Symposium on Image Analysis and Interpretation (SSIAI'02)
0-7695-1537-1/02 $17.00 © 2002 IEEE

$$(\text{Template})_i = [\ x_{i1}, \ldots, x_{i80}, y_{i1}, \ldots, y_{i80}\ ]^T$$

$$\text{mean\_template} = \frac{1}{50} \sum_{i=1}^{i=50} (Template)_i \qquad (3)$$

2.3. Accumulator updating and reckoning of votes

Updating the accumulator, as well as picking the best estimate for the reference point is the remaining part of the algorithm.

A. Hough Domain update process

Here, template matching is done based on the gradient information and for each match the accumulator bin is updated and this search and update process is repeated for every value of the scale 's' and rotation 'Φ'. A look-up table (R-table) for the template edge image is created with different values of (r,α), where r is the distance from the edge to the chosen reference point and α is the angle between the radius vector and the horizontal. The R-table summarizes the shape information of the template and is a part of the parametric representation of the image. Now, with the R-table, template matching is done on the target image and votes for possible matches are collected in the accumulator structure.

B. Reckoning of votes in Hough Domain

The accumulator is a 4-D structure where 2 of the dimensions (m x n) represent the spatial coordinates and the remaining two (j x k) represent scale and rotation variations of the template. As a consequence, the 4-D accumulator can be seen as a (j x k) collection of 2-D images of the spatial location of the reference and each of those images can be treated individually. If the 2-D accumulator is plotted for a fixed value of scale and rotation [Fig. 2(a)], we would expect as per theory, a well defined peak. In practice, the peak is neither unique nor well defined. Votes do not form uniform and compact regions and this makes the task of finding a good candidate for the reference point very difficult. Excessive noise in the original images makes votes sparse in the accumulator, yielding non-uniform structures. To fix this problem, Gaussian filtering is performed and this greatly improves the chances of finding an absolute maximum peak and this leads to a more relevant observation: Several peaks tend to align describing the orientation and location of the cervical vertebrae. The neighborhood peaks represent a more complete figure of merit about the overall performance of the template.

A small window, having the highest peak as the central point is defined as our region of interest. The next step is to try and find the approximate orientation of the spine and the location of the neighboring peaks with the knowledge of the location of the highest peak. Let γ be the angle that approximates the orientation of the template and 'Φ' the current rotation angle of the template. Then μ defined as

$$\mu = \gamma + \Phi \qquad (4)$$

gives a good approximation of the orientation of the cervical spine in the target image.

This angle is then used to rotate the small window as shown in Fig. 2(b). An intensity profile passing through the center of the window is defined as the characteristic profile and this has the property of passing through the center of the highest peak and if the rotation angle is correct, it should also pass through the neighboring peaks. A plot of the characteristic profile can be seen in Fig. 2(c).

Once the highest peak has been found, the votes corresponding to this peak are eliminated, so that the next search yields the second highest peak. The same process is repeated and the third peak is also found. The output of the process will be $\hat{A}$, the cumulative value of the highest peak and its 2 neighbors. This step is a quantitative measure of how well the template matches the whole cervical vertebrae. The next subtlety involved in this analysis is the shape of the peaks. A blurred peak will indicate that in spite of the considerable quantity of votes that form the peak, the template matching is poor, yielding a wide-spread structure. On the other hand, a sharp peak will indicate a more accurate match since most of the votes concentrate in that small region. Hence, we use the gradient operator to measure the decay rate of the highest peak, and refer to this measure as $\hat{G}$.

The new criterion for reckoning of votes can be defined as a weighted sum of two quantities

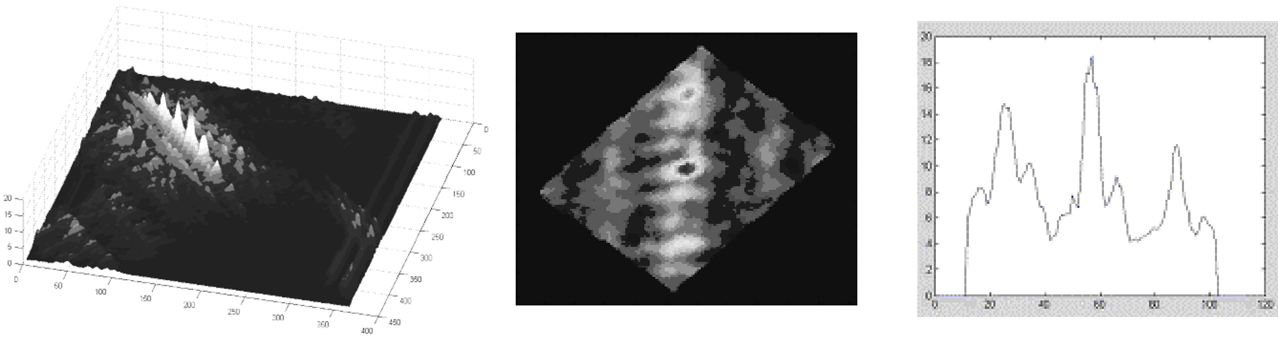
$$\text{Proposed criterion} = \beta\,(\hat{A}) + \varepsilon\,(\hat{G}) \qquad (5)$$

where β and ε are constants less than one.

As stated earlier, for every value (Φ, s), a 2-D image of the accumulator is created and for each of those, we compute the pair of corresponding $(\hat{A}, \hat{G})$ values and store them. The (x,y) location of the highest peak is stored as the best candidate for the reference point. While the traditional criterion chooses the best (Φ, s) pair to be the highest value in the accumulator, the new criterion for reckoning of votes chooses the best (Φ,s) pair using both $\hat{A}$ and $\hat{G}$ values as follows:

1. Find the highest value in the accumulator for each value of scale and orientation.
2. Define the subset of best (Φ, s) pairs:

$$(\Phi, s)_{best} = \{\ (\Phi, s) \mid \hat{A}\ (\Phi, s) \geq \beta\ .\hat{A}(\Phi, s)\}$$



2(a) 2(b) 2(c)

Figure 2. (a) View of the accumulator at a given scale and rotation angle. The highest peak defines the central point. (b) The resulting sub-image of the accumulator after rotation. (c) Plot of the resulting characteristic profile.

3. Within this subset, find the (Φ, s) pair with the highest decay rate:

$$\hat{G}_{max} = \max\{\hat{G}(\Phi, s)\} \ \forall \ (\Phi, s) \in (\Phi, s)_{best}$$

$$(\Phi, s)_{\hat{G}max} = \{(\Phi, s) \in (\Phi, s)_{best} \mid \hat{G}(\Phi, s) = \hat{G}_{max}(\Phi, s)\}$$

4. The corresponding P_r (x_r, y_r) for the $(\Phi, s)_{\hat{G}max}$ gives the best approximation for the reference point.

This restatement of the criterion of reckoning of votes allows us to include the shape information available in the accumulator that was not used by the traditional approach. This leads to a complete, yet simple criterion, which as we will see in the next section, leads to more accuracy in segmentation.

3. Experimental Results

A set of 50 images selected from the NHANES II database was used as our data set. For each image, a template was defined by placing a series of landmark points (LMP) based on morphometric points, which were placed by expert radiologists and represented our ground truth.

Values for β between 0.8 and 0.95 were chosen and the values for ε were chosen to be between 0.05 and 0.2. These values are experimental and work well with this application.

To assess the performance of the proposed approach, two quantitative measurements were used and three main experiments were carried out.

Measurements:

1. Number of original LMP that fall within a bounding box surrounding the template at its final location.
2. Measurement of the difference in orientation between the original shape, described by the LMP, and the final output of the algorithm.

Experiments:

i) GHT was applied to every image in the data set using its corresponding template.
ii) GHT was applied to the data set using the mean template and the reckoning of votes was done by the traditional criterion.
iii) GHT was applied to the data set using the mean template and the reckoning of votes was done by the proposed criterion.

Results from experiment (i) showed an average of 78 out of 80 LMP falling within the bounding box and an average difference in orientation of less than 1 degree. This result [Fig. 3(a)] demonstrates that GHT is indeed a legitimate approach to the segmentation problem if the correct template is used. Results from experiment (ii) showed an average of 62.48 out of 80 LMP falling within the bounding box and an average difference in orientation of 7.21 degrees. Here we can see that due to the use of a single template for all the images, the success rate is decreased [Fig. 3(b)].

More accurate results were obtained for experiment (iii): 72.06 out of 80 LMP, on average, fell inside the bounding box and the orientation error was 4.16 degrees on average. These results demonstrate that the proposed criterion leads to a more accurate segmentation process [Fig. 3(c)].

4. Conclusions and future work

The task of segmenting the cervical vertebrae has been addressed using a customized and robust approach that leads to promising results. It takes into account valuable shape information present in the voting structure. Current work involves the inclusion of more templates to increase accuracy. As mentioned before, in spite of how much gradient information is available and could be used by the proposed criterion, representativeness of the template is

3(a) 3(b) 3(c)

Figure 3. (a), (b) and (c): Results of experiments (i), (ii) and (iii), respectively. Original LMP marked in black, final output marked in gray.

still a very important issue. Shape variability captured by the mean template is limited. If a better representation of the shape variability is desired, more templates should be included. Adding more templates to capture more variability can be seen as a clustering problem. The collection of 50 templates represents the 100% shape variability seen across the data set. The idea is to decrease this set to a minimum number of templates while keeping a good amount of variability. Tradeoffs between accuracy and computational complexity are under study. One of the main drawbacks of GHT is execution time, which increases with the inclusion of more templates. Therefore a good tradeoff between the desired accuracy and the computational complexity must exist for this approach to be feasible.

5. Acknowledgements

This work is sponsored in part by a grant from the Communications Engineering branch of the National library of medicine.

6. References

[1] D.H. Ballard, "Generalizing the Hough Transform to detect arbitrary shapes", *Pattern Recognition*, 13(2), pp. 111-122, 1981.
[2] V.F. Leavers, "Which Hough Transform?", I*mage Understanding*, Vol. 58, No.2, pp. 250-264, Sep. 1993.
[3] H. Li, M. A. Lavin, and R. L. LeMaster, "Fast Hough Transform", *Proceedings 3rd Workshop on Computer Vision*, Bellair, MI, 1985.
[4] J. Illingworth and J. Kittler, "The adaptive Hough Transform", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 9 (5), 1987.
[5] S. Ando, "Consistent gradient operators", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 22, No.3, pp.252-255, March 2000.
[6] G. Zamora, H. Sari-Sarraf, S. Mitra, and R. Long, "Estimation of Orientation and Position of Cervical Vertebrae for Segmentation with Active Shape Models", *Proceedings of SPIE*, San Diego, CA, USA, 2000.

0-7695-1537-1/02 $17.00

IEEE COMPUTER SOCIETY