MVA2000 IAPR Workshop on Machine Vision Applications, Nov. 28–30, 2000, The University of Tokyo, Japan

8—17

Automatic Segmentation of Features in Medical CT Imagery

A. L. Reno and D.M. Booth
DERA
St. Andrews Road, Malvern
WR14 3PS, UK.
alreno@dera.gov.uk, dmbooth@dera.gov.uk

Abstract

A method is described for fusing the results of different feature detection algorithms in order to segment regions in CT (computed tomography) images. While many feature detectors are available, each has its own characteristics strengths and weaknesses. In this work we aim to anticipate the complementary nature of such detectors, and so exploit strengths and tolerate weaknesses in each. To achieve this, the system is initially trained, so that the expected responses from each feature detector can be determined. A likelihood function then provides a combined measure of the similarity of each image pixel to the target region of interest and to the background region. This provides a means to segment the region of interest from the background using all information provided by the detectors. During segmentation, constraints are applied so that the shape of the extracted region closely resembles the anticipated appearance of the target region of interest.

1. Introduction

A new technique for automatically segmenting features of clinical interest in medical CT (computed tomography) images is described. A difficulty with this task is that there is usually very low contrast between the organs, which are of interest to clinicians, and other areas of such as bone and soft tissue which comprise the background. There is also variation in organ shape between different patients, and in most instances there is a degree of uncertainty in the exact position of the boundary. In fact, tests show that the manual segmentation of such regions by clinicians often varies considerably, showing that there is much subjectivity in the task. The approach described here, instead of relying on a single specialised segmentation algorithm, fuses the output of one or more very simple pixel level classifiers, and combines this with structural information concerning the anticipated shape of the target region. Therefore with this approach we are not relying on highly accurate results from each feature classifier. Rather we aim to exploit the joint results that they provide, even though on their own these results may not be very good. The emphasis is therefore on error anticipation and tolerance, rather than on error free performance.

The classifiers that we consider here provide any pixel level discrimination between the target region and the background. The system must initially be trained to recognise the expected responses of each of the classifiers in the two separate regions. This is achieved with a set of training images in which the regions corresponding to the target object and the background have been manually identified. The mean vector and covariance matrix corresponding to the responses of the classifiers are then computed for the target and background regions. A model of the target shape, in the form of one or more aligned binary images, is then used to provide our prior knowledge. This is an alternative approach to methods based on active contour or point models. It is quite flexible because it is possible to include any type of region based classifier, it is simple to train, and computationally efficient.

2 Pixel Classification

Pixel level classifiers simply discriminate between image pixels in the target (foreground) region and the background region. This discrimination is provided because, statistically, their expected response differs over these two regions. In this sense, we can utilise the output from any algorithm that performs such a function. Here, we will use three texture based classifiers, and a segmentation algorithm.

Changes in texture properties often signify the boundary of a particular region or organ. Therefore,

small texture variations over the CT image can be used to assist in the identification of the region of interest. The texture measures are computed at each 15×15 window in the image. From these sub-regions, a grey–level co-occurrence matrix, $H(i,j)$ is computed. The 256 grey–levels are first downsampled to 16 levels though, to reduce the size of the matrix. Three texture measures which were found to provide reasonable discrimination are the texture entropy, energy and homogeneity. The definitions of these, as given by Jain, Kasturi, and Schunck [3], are as follows,

$$Entropy = -\sum_i \sum_j H(i,j) \log H(i,j), \quad (1)$$

$$Energy = \sum_i \sum_j H^2(i,j), \quad (2)$$

$$Homogeneity = \sum_i \sum_j \frac{H(i,j)}{1+|i-j|}. \quad (3)$$

There is naturally some correlation between these measures (see Figure 1) and this needs to be taken into account. As well as texture, we also make use of a simple segmentation algorithm. With this, a fixed threshold is used to divide the image into two regions, which correspond to the denser organ and bone regions (light), and the less dense softer tissue regions (dark). None of these pixel classifiers give a definitive result. Although in isolation each provides relatively poor discrimination, collectively they are much more useful. The classifiers are combined as follows.

Firstly, we assume that each vector $\mathbf{I} = (\mathbf{I}_1, \mathbf{I}_2, \ldots, \mathbf{I}_n)$, corresponding to the classification at a given image pixel, has a multivariate normal density. The reason for doing this is described more fully in [5]. Then we must determine a mean and covariance matrix, both for the target region of interest, and for the background region. These can be obtained from training images, in which the region of interest has been identified, and therefore background and foreground regions are known. We can, for convenience, refer to these parameters as $\alpha = (\bar{\mathbf{I}}_f, \mathbf{\Sigma}_f)$ and $\bar{\alpha} = (\bar{\mathbf{I}}_b, \mathbf{\Sigma}_b)$, where the first is the foreground mean and covariance, and the second is the background mean and covariance.

Given a new image, in which we wish to identify the particular structure of interest, we first need to compute a likelihood image. With our assumption of normally distributed feature responses, this is given by

$$\log \frac{p(\mathbf{I}|\alpha)}{p(\mathbf{I}|\bar{\alpha})} = \log \sqrt{\frac{|\mathbf{\Sigma}_b|}{|\mathbf{\Sigma}_f|}} + \frac{1}{2}(\mathbf{I}-\bar{\mathbf{I}}_b)^T \mathbf{\Sigma}_b^{-1}(\mathbf{I}-\bar{\mathbf{I}}_b) - \frac{1}{2}(\mathbf{I}-\bar{\mathbf{I}}_f)^T \mathbf{\Sigma}_f^{-1}(\mathbf{I}-\bar{\mathbf{I}}_f). \quad (4)$$

This image specifies, at every pixel, the ratio of the probability of foreground to background (the Bayes' factor). Simple calculation shows that the sum of this term over any group of pixels gives the conditional joint probability that the corresponding region in the image is foreground [4]. Similarly, we may impose information about the anticipated shape of the target region. Suppose we have a set of aligned reference templates $B_1, B_2, \ldots$, which are examples of the shape of the target feature. Each example image is binary, and specifies pixel values of 1, for foreground, and 0 for background. The binary complement of each of these images is also needed, $\overline{B}_1, \overline{B}_2, \ldots$, where each specifies pixel values of 0, for foreground, and 1 for background. Then we compute the image

$$\log \frac{G(x,y)}{\bar{G}(x,y)} = \log \frac{\text{smooth}_\sigma\left(\sum_i B_i\right)}{\text{smooth}_\sigma\left(\sum_i \overline{B}_i\right)} \quad (5)$$

where smooth_σ represents smoothing with a Gaussian of standard deviation σ. (In the example, we use just one template B, and this is sufficient.) With these two components, it is not difficult to see that the posterior probability of any given region Γ (or multiple regions) is

$$\log \pi(\mathbf{x}) = \oint_\Gamma \left(\int_0^{y(u)} \log \frac{G(x,y)}{\bar{G}(x,y)} dy \right) dx + \oint_{\Gamma'} \left(\int_0^{y(u)} \log \frac{p(\mathbf{I}|\alpha)}{p(\mathbf{I}|\bar{\alpha})} dy \right) dx. \quad (6)$$

This is comprised of an image based term and a shape based term. The Γ' represents the fact that the shape of the region as observed in the image, may need to undergo some geometric transformation. This needs to account for factors such as scale, position and rotation for example, which of course are unknown and need to be determined. We can represent the region shape, Γ, as a plane closed contour, comprised of $x(u)$ giving the x–coordinate and $y(u)$ giving the y-coordinate. Then, the coordinates of the corresponding region Γ' in the image are, under an affine transformation,

$$x'(u) = \gamma_1 x(u) + \gamma_3 y(u) + \gamma_5, \text{ and} \quad (7)$$

$$y'(u) = \gamma_2 x(u) + \gamma_4 y(u) + \gamma_6. \quad (8)$$

Similarly, we need to also have the inverse transformation, and this can be represented by λ. Any candidate solution can then be represented $\mathbf{x} = (\Gamma, \gamma)$, or alternatively $\mathbf{x}' = (\Gamma', \lambda)$, where where $\boldsymbol{\gamma}$ and $\boldsymbol{\lambda}$ each holds the six transformation parameters. We first perform the integrations in the brackets in (6), and store these images separately, since they will not change during the

computation of $\log \pi(\mathbf{x})$. So the posterior probability is given then by a sum around the region boundary and its transformation, over these two constant images. The value of each of these two constant images, at each point on the region boundary is taken, and multiplied by dx and dx' respectively, as the boundary is traversed in an clockwise direction. Note, the sign of dx and dx' at each point is important.

The derivative function of (6) is

$$\frac{\partial \log \pi(\mathbf{x})}{\partial \mathbf{x}} = \int_{\Gamma} \frac{\partial(x, y, x', y')}{\partial(x, y, \gamma)} \begin{bmatrix} -y_u(u) \log \frac{\mathcal{G}_{ref}(\mathbf{v})}{\bar{\mathcal{G}}_{ref}(\mathbf{v})} \\ x_u(u) \log \frac{\mathcal{G}_{ref}(\mathbf{v})}{\bar{\mathcal{G}}_{ref}(\mathbf{v})} \\ -y'_u(u) \log \frac{p(\mathbf{I}'|\alpha)}{p(\mathbf{I}'|\bar{\alpha})} \\ x'_u(u) \log \frac{p(\mathbf{I}'|\alpha)}{p(\mathbf{I}'|\bar{\alpha})} \end{bmatrix} du, \tag{9}$$

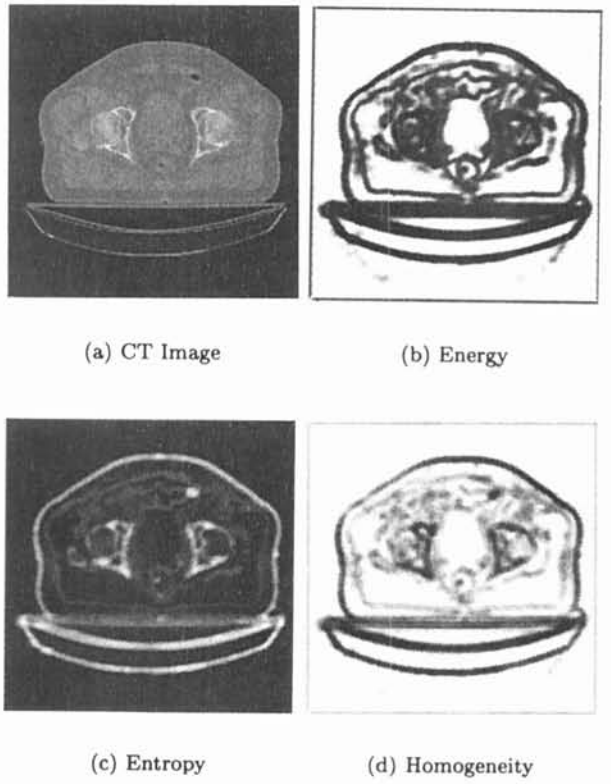
where the Jacobian is straightforward to compute from the definitions already given. A similar term can be found when the inverse transformation is used. So the derivative at any point $x(u)$, $y(u)$, is taken as the contribution of the integral (9) between $\mathbf{\Gamma}(u)$ and $\mathbf{\Gamma}(u + du)$. But the derivative of each of the transformation parameters $\gamma_1, \gamma_2, \ldots$ is given by the total contribution of the integral around the whole region perimeter.

3 Parameter Estimation

We estimate the best region using (6) and (9). Using a Markov Chain Monte Carlo procedure, starting from an initial guess, the candidate region is iteratively modified by proposing adjustments to shape and pose parameters. These proposals are generated from a Gaussian distribution, and are selected at two different scales, one large and the other small. The mean value of the small scale adjustments is given by the gradient function, Equation (9). This enables the estimation to mimic a jump–diffusion process, the type of which was suggested by Grenander and Miller [2] for extracting mitochondria shapes. Proposed adjustments to the parameters are either accepted or rejected, depending on Green's [1] generalised acceptance test,

$$\alpha(\mathbf{x} \rightarrow \mathbf{y}) = \min\left(1, \frac{\pi(\mathbf{y}) q_m(\mathbf{y} \rightarrow d\mathbf{x})}{\pi(\mathbf{x}) q_m(\mathbf{x} \rightarrow d\mathbf{y})}\right), \tag{10}$$

where $\mathbf{x}$ is the current configuration, $\mathbf{y}$ is the proposed new configuration, and $q_m(\cdot \rightarrow \cdot)$ is the probability of the move proposal. The exact form of these terms is described in more detail in [4]. We have found this optimisation method to work well, though we have not yet experimented with other approaches.



(a) CT Image (b) Energy

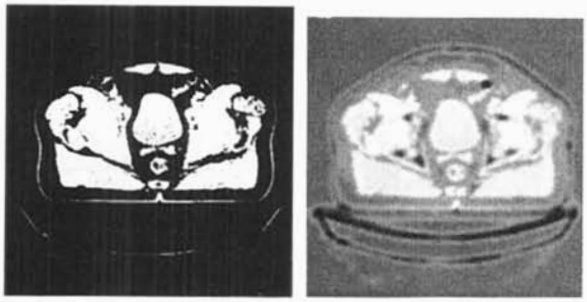
(c) Entropy (d) Homogeneity

Figure 1. Features extracted from a CT image.

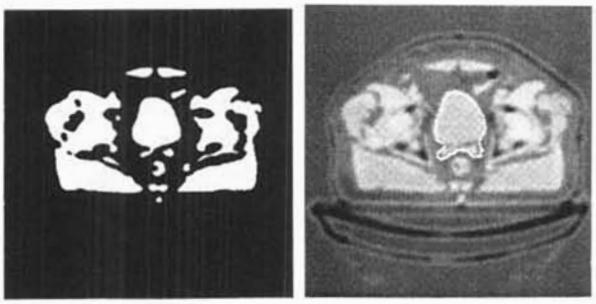
4 Results

A given CT image is segmented by finding the region which optimises, or gives an acceptable solution, to the posterior density above. The optimisation is achieved by gradually adjusting the shape and pose of the region for a fixed number of iterations. The most probable state (one found with highest value for $\log \pi(\mathbf{x})$) is then taken as the best found solution. The illustration in Figure 1 gives an example. The task is to segment the bladder, at the centre of the image. Figure 1(a) shows the CT image. Figure 1(b)-(d) show the results of four separate classifiers. The first three show the texture energy, entropy and homogeneity computed for this image. The mean and covariance matrices used to compute these images were first obtained in a training phase. Using another scan from the same patient, the target region was manually identified. A single reference shape for this example was obtained by manually tracing the same region in another CT scan of the same patient. The binary template, and its complement were convolved with a Gaussian of standard deviation 2 pixels, and combined to produce the shape model described above.

The three features shown in Figures 1(b-d) and the result of the threshold–based segmentation algorithm, shown Figure 2(a) can be combined or fused. These four images provide the feature vector $\mathbf{I} = (\mathbf{I}_1, \mathbf{I}_2, \ldots, \mathbf{I}_4)$ at every pixel over the original CT image. Furthermore, we have $\alpha = (\bar{\mathbf{I}}_f, \mathbf{\Sigma}_f)$ and $\bar{\alpha} = (\bar{\mathbf{I}}_b, \mathbf{\Sigma}_b)$ from our training phase. Now, we can use Equation (4) to fuse the features together into a likelihood image. The likelihood image in this example is shown in Figure 2(b), while Figure 2(c) shows the positive areas (white) in this same image. The positive areas indicate pixels which are more similar to the target region, than to the background, since only when $p(\mathbf{I}|\alpha) > p(\mathbf{I}|\bar{\alpha})$ in the log ratio, will the likelihood be positive. The region was initialised as a square at the centre of the image, and after 50000 iterations of the algorithm, the highest resulting region is shown in Figure 2(d).



(a) Segmentation (b) Log likelihood



(c) Positive likelihood (d) Segmentation result

Figure 2. Using the likelihood image to extract the required region.

5 Conclusions

A new method has been described for extracting features in medical CT images. The approach involves combining the outputs of several different pixel level classification algorithms. Although each of these, in isolation, gives a relatively poor result, when fused together, their joint capability is much better. Hence, failure, or weaknesses in one classifier, may not cause failure of the entire approach. Using this approach has the advantage that it only requires a simple training phase, and once this is done, the computation is efficient, and is linear with respect to the perimeter of the region of interest.

References

[1] P. J. Green. Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrica*, 82(4):711–732, 1995.

[2] U. Grenander and M.I. Miller. Representations of knowledge in complex systems (with discussion). *Journal of the Royal Statistical Society B*, 56(4):459–603, 1994.

[3] R. Jain, R. Kasturi, and B. Schunck. *Machine vision.* McGraw Hill, 1995.

[4] A.L. Reno. Bayesian object segmentation. In H.R. Arabnia, editor, *Proceedings of the International Conference on Imaging Science and Technology*, pages 35–41. CSREA Press, June 1999.

[5] A.L. Reno and D.M. Booth. Statistical region measures for separation of figure from ground. In *Proceedings of the 6th IEEE Conference on Electronics, Circuits and Systems*, volume II, pages 863–866, 1999.