A Topographic Representation for Mammogram Segmentation

Byung-Woo Hong* and Michael Brady

Medical Vision Laboratory, University of Oxford, Oxford, U.K.
{hong,jmb}@robots.ox.ac.uk

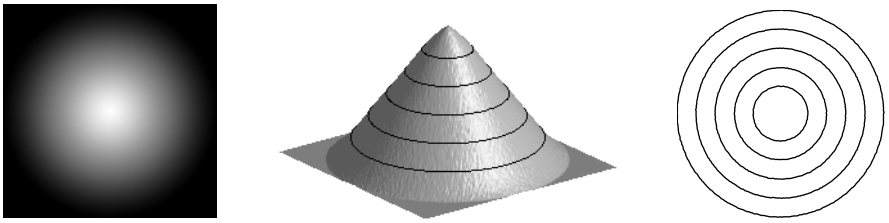
Abstract. This paper presents a novel segmentation method for delineating regions of interest (ROI's) in mammograms. The algorithm concurrently detects the breast boundary, the pectoral muscle and dense regions that include candidate masses. The resulting segmentation constitutes an analysis of the global structure of the object in the mammogram. We propose a topographic representation called the iso-level contour map, in which a salient region forms a dense quasi-concentric pattern of contours. The topological and geometrical structure of the image is analysed using an inclusion tree that is a hierarchical representation of the enclosure relationships between contours. The "saliency" of the region is measured topologically as the minimum nesting depth. Experimental results demonstrate that the proposed method achieves a satisfactory performance as a prompt system in the mass detection.

1 Introduction

Image segmentation aims to delineate regions, each of which is, to a certain extent, homogeneous. However, universally accepted measures of uniformity especially of textured regions do not currently exist. This leads to segmentation being domain and problem specific. This paper aims to develop a segmentation algorithm for (X-ray) mammograms. Breast cancer is the leading cause of death from cancer among women in many countries. Mammography is the most cost effective method to detect early signs of breast cancer. However, mammograms are highly complex images and the signs of disease are often subtle. It has been found that a large number of cancers are missed at screening perhaps as many as 20%. For this reason, there has been considerable effort aimed at developing computer-aided diagnosis (CAD) systems that might provide a consistent and reproducible second opinion to a radiologist. Currently, most CAD systems are designed to prompt suspicious regions. There has been substantial progress in the automatic detection of microcalcifications, but progress has been considerably slower in the reliable detection of malignant masses by computer. We introduce a novel segmentation method for mammograms and investigate its application to the detection of regions of interest (ROI's) including the breast boundary, the pectoral muscle (for MLO views) and dense regions including candidate masses if any. While most previous work has been based on image processing approaches such as texture analysis, edge detection, or statistical

* The author gratefully acknowledges the financial support of the Clarendon Fund Bursary and the ORS Award

R.E. Ellis and T.M. Peters (Eds.): MICCAI 2003, LNCS 2879, pp. 730–737, 2003.



(a) (b) (c)

Fig. 1. (a) A synthetic image of a circular object with a gradual intensity fall from the centre toward the boundary. (b) The image surface of (a) in a three-dimensional space and heights to be quantised are marked on the surface. (c) Contour map generated from quantisation in (b).

learning, we aim to detect significant regions by analysing the topological structure of an image, deliberately, at this stage, using no prior information about the objects to be segmented. In our segmentation method, global structural information is derived from a topographic representation that provides both geometrical and topological properties of the objects in the image.

2 Topographic Representation

A mammogram depicts various types of tissue with different densities in the breast. Because the structure of the breast is complex, a mammogram contains a vast amount of heterogeneous information. One obvious difficulty in mammographic image analysis stems from the high dimensionality of the data to be analysed. We seek a reliable representation that is capable of reducing the amount of irrelevant information and providing robust mammographic descriptions. In this work, a topographic representation is built based on *iso-level contours* or *isophotes*, curves of constant intensities.

2.1 Iso-level Contour Map

An image is considered as a surface in which the intensity at each pixel is regarded as height, as shown in Figure 1. An ordered set of connected points at the same height forms an *iso-level contour* for a corresponding intensity. An iso-contour $C(l)$ for a given intensity level l from an image function $I(x, y)$ is given by:

$$C(l) = \{(x, y)|I(x, y) = l\}, \qquad \forall (x, y) \in \Omega$$

where Ω is a domain of the image I. An iso-level contour is constrained to be a simply connected closed curve that is equivalent to a Jordan curve. A simply connected closed curve is homeomorphic to the unit circle. A digital image $I(i, j)$ is generally modelled as a function in a discrete domain $\Omega \subset \mathbb{N}^2$. However, it is not possible to draw a continuous curve of a constant intensity in a discrete domain such as a noisy image. Thus it is necessary to transform the discrete image domain into a continuous domain; we use bilinear interpolation to obtain a continuous image model $I(x, y)$, where $(x, y) \subset$

$\mathbb{R}^2$ so that the intensity level for the pixel at a non-integer coordinate position can be approximated and iso-level contours are well defined. A topographic representation is obtained by a set of iso-level contours at distinct multiple partition values over the intensity range of an image. This representation is referred to as an *iso-level contour map*. An iso-level contour map $CM(I)$ for an image I is given by:

$$CM(I) = \{C(l_i)|l_i \in L, i = 1, 2, \cdots, n\}, \qquad L = [I_{min}, I_{max}]$$

where L is the intensity range of a given image I between the minimum intensity I_{min} and the maximum intensity I_{max}, and n denotes the number of quantisation levels and is related to the specificity of features to be detected. It is observed that a quasi-concentric pattern of contours appears within an object, and a dense contour pattern is formed in the area of an abrupt intensity gradient. The shapes of contours that are enclosed within an object depend on the shape of that object. Generally, iso-level contours generated within an object at multiple intensity levels have shapes that are similar to the shape of the object boundary. Furthermore, they often form a quasi-concentric pattern that is based on the shape of the object. Thus, iso-level contours from a single object are often nested, especially near the object boundary. A significant shape change in a nested set of iso-level contours typically implies a transition from within an object into the local background. A contrast invariant representation can be achieved by an iso-level contour map with a fixed number of uniformly spaced iso-levels ranging from the minimum to the maximum intensity. This is important because mammograms are taken under varying imaging conditions, across patients and image acquisition systems. This results in widely varying intensity ranges for mammograms.

2.2 Anisotropic Diffusion

Noise is inevitably introduced by the image formation process. Since contour extraction is disturbed by noise, causing contours to become jagged, a noise reduction scheme is necessary to eliminate noise and insignificant details as a preprocessing step. Conventional low-pass filtering generally blurs not only noise, but also edges that are semantically important features in mammograms. Therefore, edges need to be preserved in the denoising process. An anisotropic diffusion filter by Weickert [5] is used to remove noisy fluctuations while enhancing edges so that smooth contours can be extracted. Figure 3 demonstrates how the anisotropic diffusion filtering affects the entire structure of the iso-contour map. Noise and insignificant small features in the raw image lead to an uninformative iso-level contour map that does not convey useful mammographic features. On the other hand, the topological and geometrical structure of the mammogram is well described by the iso-level contours extracted from the denoised image.

3 Segmentation

The proposed segmentation algorithm delineates ROI's by analysing the topological and geometrical structure of the image. The ROI's are referred to as *salient regions*, which appear distinctive against the surrounding background. The image is described by a

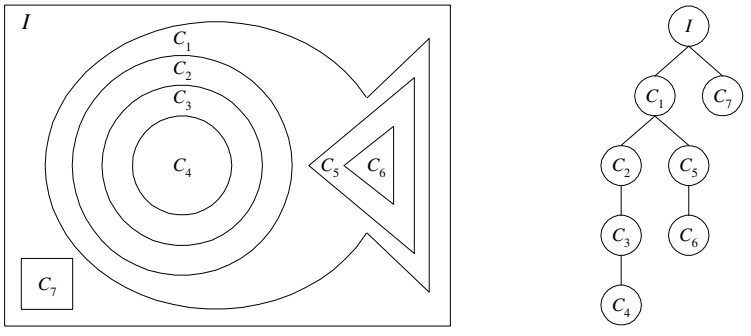


Fig. 2. (a) A example of contour map. (b) Inclusion tree for the contour map (a).

compact representation using the iso-level contour map, which is interpreted utilising the tree structure of enclosure relationships between iso-level contours. It is assumed that a salient region forms a dense quasi-concentric pattern of iso-level contours. A conspicuous dense contour pattern is generally observed near the boundary of a salient object due to an abrupt intensity gradient. The transition from the salient region to the local background often leads to the break up of the nested pattern of the iso-level contours. A saliency measure for iso-level contours is introduced to identify contours that circumscribe salient regions. The global property of contours in terms of topology and geometry is analysed in order to select salient contours. The saliency of each contour is measured using the notion of nesting depth.

3.1 Inclusion Tree

The entire pattern of contours in an iso-level contour map can be described as a relationship of enclosure. The enclosure relationship of contours can be efficiently represented in the form of a tree structure, called an *inclusion tree*, where each node represents a contour in the contour map. In the inclusion tree, a contour encloses the contours at its descendent nodes. In particular, the root node of the inclusion tree represents the boundary of the image. Figure 2 illustrates an example of an iso-level contour map and a corresponding inclusion tree. The enclosure relationship between contours can be established due to the Jordan curve theorem stating that the complement of a Jordan curve C consists of exactly two disjoint regions, an interior $Int(C)$ and an exterior $Ext(C)$. The operator $\sqsubset$ for the enclosure relationship between two iso-level contours, C_i and C_j, is defined as:

$$C_i \sqsubset C_j \qquad \text{if } C_i \subset Int(C_j)$$

A path $P_{i,j}$ from contour C_i to contour C_j in the inclusion tree consists of an ordered sequence of the iso-level contours that begins from contour C_i and ends at contour C_j. The length $L(P_{i,j})$ of a path $P_{i,j}$ is defined to be the number of edges that are directly

traversed from C_i to C_j along the path $P_{i,j}$. The degree D of a node is the number of children it has. In Figure 2, path $P_{1,4}$ from contour C_1 to contour C_4 is given as $P_{1,4} = (C_1, C_2, C_3, C_4)$ such that $C_1 \supset C_2 \supset C_3 \supset C_4$. The length $L(P_{1,4})$ of the path $P_{1,4}$ is 3 and the degree $D(C_1)$ of C_1 is 2. Nodes are called *branching* if their degree is greater than 1 (e.g. I, C_1) and their immediate children are called *base nodes* (e.g. C_1, C_2, C_5, C_7). Nodes with 0 degree are called *terminal nodes* (e.g. C_4, C_6, C_7). The hierarchical representation of the inclusion tree provides an efficient way to examine both the topological and the geometrical structure of an image.

3.2 Minimum Nesting Depth

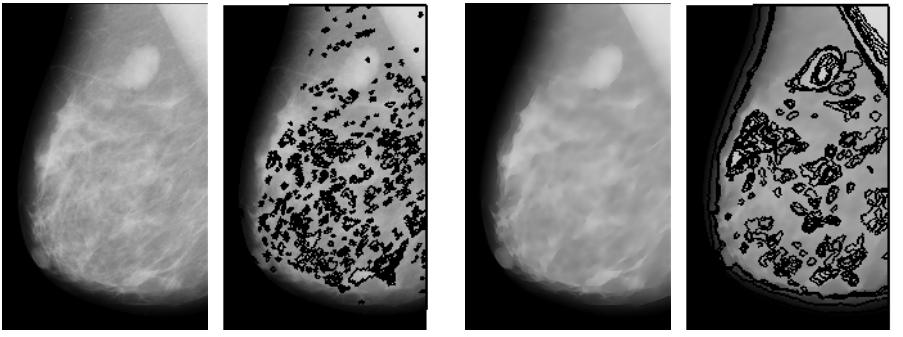
The saliency of each iso-level contour is measured by analysing the inclusion tree that is built based on the iso-level contour map. Topological changes of contour structure and the depth of nested contour structure are related mainly to measuring the saliency of contours. Bifurcation of the inclusion tree is indicative of a significant change in the iso-level contour structure and therefore in the corresponding image contents. It may indicate the separation into different objects or into constituent parts of the same object. The contours at branching nodes in the inclusion tree imply a topological change of contour structure and they are initially selected. Then, the contours at base nodes are chosen as candidate salient contours since they may support the nesting structure of the contours. The saliency of those selected contours is measured by the nesting depth. The nesting depth for a contour is given by the number of contours from the innermost contour to the contour within the nesting structure. In the inclusion tree, the nesting depth for contour C_b is equivalent to the length of the path from the contour C_b to the innermost contour C_t that is represented by the terminal node of the subtree T_b whose root node represents the contour C_b. However, there may exist more than one terminal node in the subtree T_b due to subsequent bifurcation. In this case, the minimum among the values of the nesting depth from the contour C_b to all terminal nodes in the subtree T_b, called the *minimum nesting depth*, is taken as a saliency measure. Then, the minimum nesting depth $MND(C_b)$ for contour C_b is given by:

$$MND(C_b) = \min_t L(P_{b,t}), \qquad \forall t, C_t \in T_b, D(C_t) = 0$$

where $P_{b,t}$ denotes a path from contour C_b to contour C_t's that are all terminal nodes in the subtree T_b whose root node represents C_b. In summary, the segmentation of salient regions in an image is performed by selecting the base contours in the iso-level contour map of the image and the saliency of the base contours is measure by the MND. The base contours with higher MND correspond to the boundaries of distinctive regions with abrupt intensity changes.

4 Application to Mammography

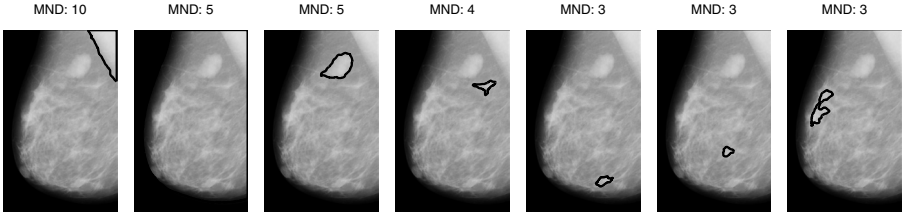
The segmentation algorithm has been applied to mammograms, to detect ROI's that include breast boundaries, pectoral muscles and candidate masses. The breast boundary can be specified as the contour where the intensity level begins to increase from the



(a) (b) (c) (d)

Fig. 3. (a) An original mammogram with a pectoral muscle and a mass. (b) The contour map generated from (a). (c) A denoised image of (a) using anisotropic diffusion filter. (d) The contour map generated from (c).

background intensity level with an additional shape constraint. Pectoral muscles appear to be uniformly bright regions at predictable locations in MLO views. Similarly, candidate masses are typically distinctive regions that are bright relative to the surrounding background, but their positions cannot be predicted, nor can the local contrast although statistically cancers may be found more often in certain parts in the breast (e.g. VOQ). In the mammogram shown in Figure 3 (a), the dense region with a triangular shape on the top right corner represents the pectoral muscle, while the distinctive bright oval shaped region to the left to the pectoral muscle is a mass. The contour map generated from the original mammogram, shown in Figure 3 (b), does not describe well the geometrical structure of the image. This is because of noise and small unimportant features. An anisotropic diffusion filter was applied to remove noise and insignificant features. The diffused image where noise is suppressed and important structures are kept with their edges enhanced is shown in Figure 3 (c) and its contour map is presented in Figure 3 (d). In the contour map generated from the denoised image, salient regions appear to form prominent nested contours. The inclusion tree was built based on the contour map from



(a) (b) (c) (d) (e) (f) (g)

Fig. 4. The extracted salient contours from the contour map in Figure 3 (d) taking a threshold about the minimum nesting depth. The contours are superimposed on the original image.

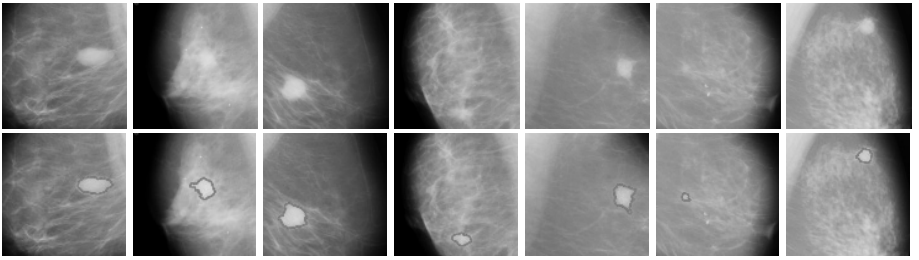


Fig. 5. Examples of correctly segmented masses. Top row: Original mammograms that include masses. Bottom row: Segmentation results superimposed on the original images.

the denoised image and the contours at base nodes in the inclusion tree were searched for candidates of the boundary of salient regions. Then, the MND was measured for the selected contours at the base nodes and a threshold was set with respect to the MND to remove less significant regions. The segmented salient regions are presented with their MND in Figure 4 where the the pectoral muscle (a), breast boundary (b), the mass (c) and some other dense tissue regions (d)-(g) are detected. The breast boundary and the pectoral muscle can be identified among extracted salient regions using topological and geometrical constraint. The breast boundary encloses the whole internal structure of the breast and has a low intensity level. The pectoral muscle is located at top right or top left corner in the breast region depending on the side of the breast and has a high intensity level. The breast boundary roughly has the shape of a half oval and the pectoral muscle generally appears as a roughly triangular shape. However, it is more difficult to discriminate masses due to the variability of their appearance. Relevant features, including density, size, shape, margin, and texture can be employed to select masses from the set of candidates. Additionally, these features can be used to classify a mass as benign or malignant. In this paper, it is aimed to analyse the structure of the breast identifying breast boundary and pectoral muscle and detecting dense regions, one of which may represent a mass. To assess the performance of the algorithm, a set of 48 mammograms (MLO views), including masses varying in size and subtlety, were selected from the MIAS database. The algorithm detected 46 masses correctly and 2 masses were missed,

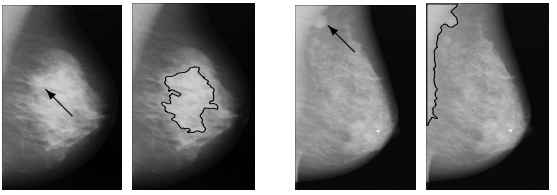


Fig. 6. Two cases of poor results. (a) A mammogram with a subtle mass indicated by an arrow. (b) The poor segmentation result for the mass in (a). (c) A mammogram with a mass that appears to be attached to the pectoral muscle. (d) The poor segmentation result for the mass in (c).

concurrently identifying every breast boundary and pectoral muscle. Some examples of correctly detected masses varying from subtle to obvious are presented with their original images in Figure 5. The mammograms that include 2 missed masses and their poor segmentation results are shown in Figure 6. One of the masses missed has low contrast since it is contained in a dense tissue area. The other missed mass appears to be attached to the pectoral muscle. The intensity of the mass may not be distinctive from the intensity of the connected region between the mass and the pectoral muscle. The experimental results show that the proposed segmentation algorithm is efficient and successful in analysing mammographic features achieving very high rate of detection of mass in mammograms.

5 Discussion and Conclusion

We have developed a segmentation method to detect salient regions in mammograms. Salient regions correspond to distinctive areas that may include the breast boundary, the pectoral muscle, candidate masses and some other dense tissue regions. A topographic representation has been developed using iso-level contours and the topological and geometrical relationship between contours is efficiently analysed utilising the inclusion tree. The saliency of contours is measured by the minimum nesting depth. The breast boundary and the pectoral muscle can be easily identified from the extracted salient regions using anatomical information. A breast coordinate system can be established after segmentation of the breast boundary and the pectoral muscle. It may provide useful information for the identification of masses and for the registration of two mammograms. In addition, the breast boundary, together with extracted dense regions, can be used for measuring the breast density, which is correlated to the risk of cancer. Masses included in extracted dense regions can be identified employing features such as shape, location, density, margin and texture even though the identification and classification of masses are not dealt with in this paper. Experimental results show that a topographic representation is largely invariant to brightness and contrast and it provides a robust and efficient representation for the characterisation of mammographic features. Segmentation based on analysing the inclusion tree enables detection of salient regions. It appears that a global structural approach of the sort inverstigated is a useful counterpart to a local statistical approach for the segmentation of mammograms. The rate of detection of masses indicates that this method may be used as the basis for an effective prompting tool to assist radiologist in the diagnosis of breast cancer.

References

1. Caselles, V., Coll, B. and Morel, J.: Topographic maps and local contrast changes in natural images. International Journal of Computer Vision. **33** (1999)
2. Highnam, R. and Brady, M.: Mammographic Image Analysis. Kluwer Academic (1999)
3. Kok-Wiles, S.L., Brady, J.M. and Highnam, R.P.: Comparing mammogram pairs for the detection of lesions. 4th IWDM (1998)
4. Shiffman, S., Rubin, G. and Napel, A.: Medical Image Segmentation Using Analysis of Isolable-Contour Maps. IEEE Trans. on Medical Imaging. **19** (2000)
5. Weickert, J.: A Review of Nonlinear Diffusion Filtering. Scale-Space Theory in Computer Vision, Lecture Notes in Computer Science, Vol. 1252. Springer Verlag (1997)