Using the Topological Tree for skin lesion structure description

Rita CUCCHIARA, Costantino GRANA
Università degli Studi di Modena e Reggio Emilia, Modena, Italy

Abstract. In this work we describe the Topological Tree (TT) as a knowledge representation method that relates some important visual and spatial features of image regions, namely the *color similarity*, the *inclusion* and the *spatial adjacency*. Starting from color-based region segmentation of an image into disjoint regions, their spatial relationships can be devised and described with graph-based methods. We are interested in the region's propriety "to be included into" (in the sense of "surrounded by") another region. This property could be very useful in biomedical imaging and in particular in the diagnosis of skin melanoma. The TT can be constructed after segmentation, by computing the spatial relationships of regions or can be generated directly during the segmentation: to this aim we present a novel recursive fuzzy c-means (FCM) clustering algorithm based on the PCA of the color space. In the paper, in addition to the TT definition and the construction algorithm description, some results are presented and discussed.

1. Introduction

Medical imaging is a very spread discipline concerning the exploitation of image analysis and computer vision techniques in order to support and improve the physician's work. In dermatology, image analysis is widely adopted in the effort of algorithmically reproducing clinical evaluations. A great work has been done to formalize the dermatological knowledge and their experience in visual analysis, in order to extract significant rules useful for melanoma diagnosis. In the field of image analysis, most works concentrated on the separation of lesion from background, since it is obviously a critical preliminary step in any analytic procedure. Grayscale threshold operations, color clustering, edge-finding and non-linear diffusion have been proposed, while a metric for performance evaluation of the various techniques has been developed in [3]. Even if color is a key feature in dermatoscopic diagnosis (the third of the ABCD criteria presented in [1]), not much effort was provided to adequately exploit color information. Moreover, it has been pointed out that a segmentation stage should consider not only color similarity, but also spatial information [2].

Due to these motivations, in strict collaboration with dermatologists, we defined an approach to knowledge representation specially conceived to outline color features and spatial relationships between meaningful areas of the skin lesion. In particular we focus our attention on the *inclusion* property between regions, as suggested by the dermatologists' analysis approach, and organize the relations description in a structured representation (called *Topological Tree*, TT) based on a tree whose branches represent the inclusion property while nodes contain region information.

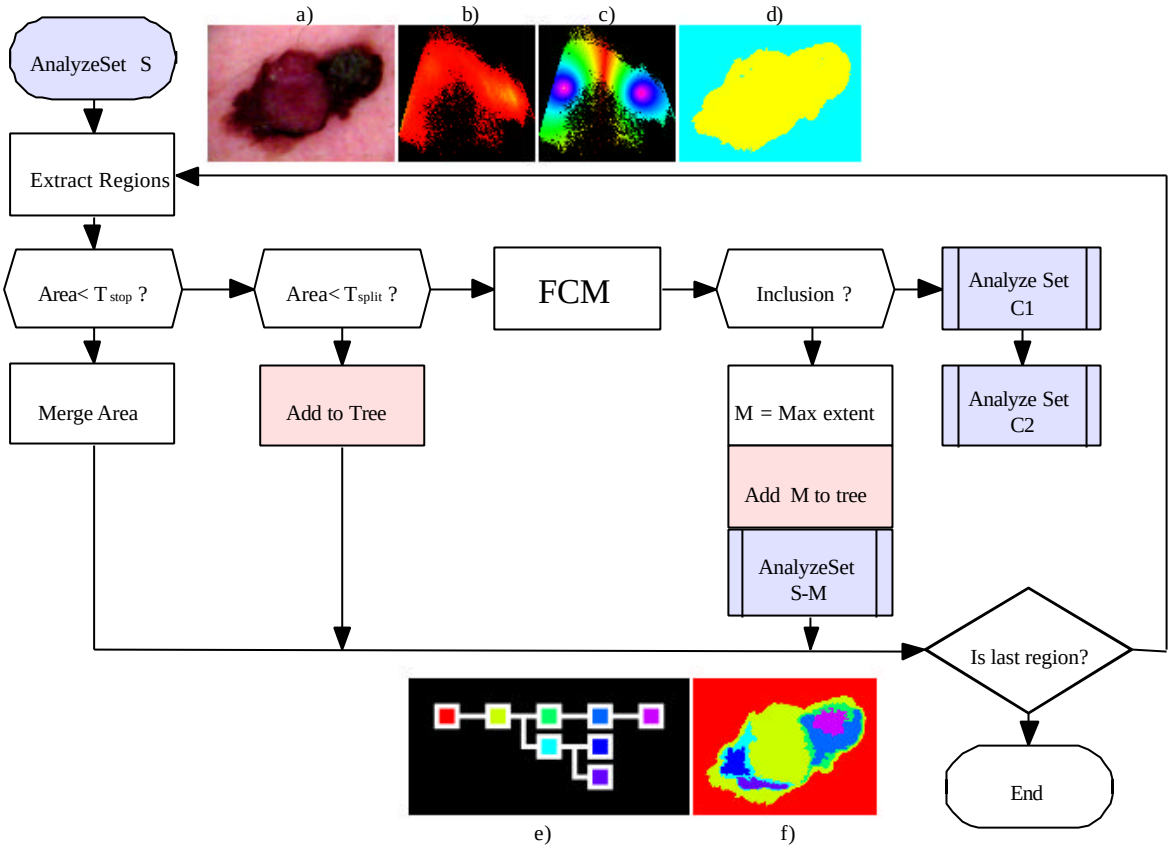

Figure 1. Flow chart of the TT construction algorithm.

The approach we proposed is oriented to describe images and objects that are naturally characterized by areas surrounded each other, such as the pigmented lesions in the figures that follow. Nevertheless the approach is very general and could be applied in many other contexts. The topological tree exhibits some interesting properties that can be exploited to extract knowledge from images for information retrieval, image understanding and diagnosis purposes.

2. The Topological Tree

The main idea behind structuring the region decomposition of the lesion in a tree derives from the observation that the pixels group easily into two clusters, one brighter, corresponding to the healthy skin and the other darker, corresponding to the lesion. Iterating the same concept on the lesion's interior could provide us with an informative view of the internal structure.

Color is evidently the main clinical criterion in this type of analysis; however, a blind segmentation in multiple, differently colored areas could lead to segmentation into areas of poor interest. In Fig. 1 an example of skin lesion divided into two color clusters with our method shows intuitively our idea of representing the inclusion propriety by the topological tree (Fig.1e). Let us give some definition of the nature of regions and their relationships used in the topological tree.

Given an image $I = \{x = (i, j), 1 \le i \le M_1, 1 \le j \le M_2\}$, a segmentation into disjoint regions provides the Region set as a set of connected regions $\mathbf{R} = \{R_1, \ldots, R_k\}$ such that $\bigcup R_i = I$ and $\bigcap R_i = \emptyset$. Segmentation is often attained by clustering approaches, dividing

the image into a number of clusters; points of the same cluster are alike in the sense that they share a common visual property (e.g. the color). In order to obtain a Region set the clustering process should be followed by an analysis of the spatial relationships between points, according with a given neighborhood system $N = \{N_x : x \in I\}$, where $N_s \subset I$ is a set containing the neighbors of x (e.g. the *8-connectivity* property). In a graph-based representation, spatial relationships between regions, extracted by segmentation, are further exploited in order to give a global description of the image. An example is the *adjacency graph* [6], a graph $\mathbf{G}(V,E)$ whose vertexes are the image regions ($V = \mathbf{R}$) and whose arcs show the adjacency property, that is a neighborhood system at region level. We could say that R_i is *adjacent* to $R_j \Leftrightarrow \exists x_i \in R_i, x_j \in R_j : x_j \in N_{x_i}$.

Now we aim to define an *inclusion* propriety that describes the fact that one region is surrounded by (being or not adjacent to) another region: first, we consider an "extended" set of image regions $\overline{\mathbf{R}} = \mathbf{R} \cup R_0$, being R_0 a dummy region representing the external boundary of the image. Let us define the inclusion propriety as follows:

Definition 1 A region $R_i \in \mathbf{R}$ is *included* in $R_j \in \overline{\mathbf{R}}$ if not exists $\{x_0, \ldots, x_M\}$ such that $x_k \in N_{x_{k+1}}, x_0 \in R_i, x_M \in R_0, x_k \in R_j, 0 < k < M$.

This topological feature can be exploited to define the Topological tree:

Definition 2 A *Topological Tree* (TT) is a tree (N,B) whose nodes are regions of an image, and whose branches are oriented arcs and describe the inclusion relationship between adjacent regions (parent and child): given $n_1, n_2 \in N$, $(n_1, n_2) \in B$ means that n_2 is *included* in n_1.

In order to use the TT for describing a segmented images we can assume that $N = \overline{\mathbf{R}}$. The inclusion property is neither reflexive neither symmetric but is transitive. Therefore all the regions of a sub-tree of a region are included in that region and the TT root (i.e. R_0) includes all other regions. An isomorphism between a TT and a sub-graph of the adjacency graph on $\overline{\mathbf{R}}$ could be demonstrated.

3. Using the TT for skin lesion structure description

The TT construction could be obtained after the segmentation, verifying the inclusion property. Conversely, we defined an algorithm, that we called *recursive FCM*, that is specifically oriented to provide both steps: segmenting regions by means of color clustering and insert them in a TT structure. For details about the algorithm see [4,5]. The method is based on a dichotomy of the segmentation set (see Fig.1) into two color clusters obtained after Principal Component Analysis (PCA) by means of a fuzzy C-means (FCM) clustering. In Fig. 1b,c,d you can see the 2D color histogram constructed after PCA, its clustering and the results obtained at the first step. Then inclusion is tested, and the connected regions are added to the tree or are evaluated again by the algorithm, recursively. Since the TT construction is done recursively we have defined the stop conditions related with the region of interest (ROI) that we want to represent in this model. In particular, a skin ROI is a set of pixels of the skin image exhibiting three properties: uniform color, connected pixels and significant area. Thus we stop the tree construction when a region has an indivisible color or its size is so little that is not necessary to divide again. A peculiar characteristics is that the stop condition is controlled by features of the region under examination, without relationship with the rest of the image, while the standard criterion applied in clustering segmentation is the cluster number, that has to be fixed a priori or devised by features of the whole image.

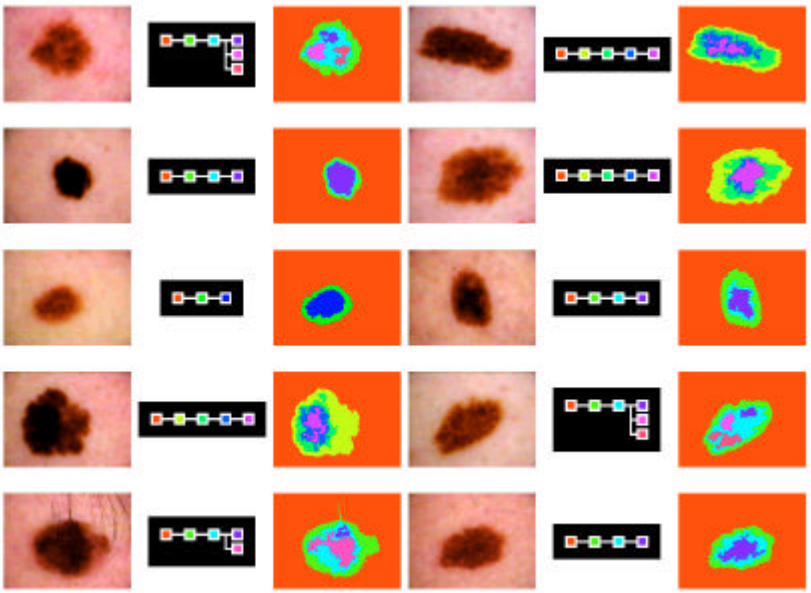

Figure 2. TT results on benign lesions.

Indeed, the main contribution of the work is the knowledge representation power: different aspects of the image, important for melanoma diagnosis, can be examined directly working on the TT and tuning the stop conditions: we can point out the general structure, or we can focus on details such as black dots. In this paper we underline the general structure of a skin lesion therefore we stop the tree construction after few steps, tuning the area threshold high enough to have large skin ROIs. We tested the algorithm over many images, using a large skin lesion database. To evaluate the performance of our system here we present visual results of a sample of 20 lesions (10 lesions are benign and 10 malignant melanoma), as shown in Fig. 2 and 3, which show the lesion, the constructed TT and the image segmentation with pseudo-colors.

We actually do not have a quantitative measure of performance but we have only a qualitative figure, i.e. the feedback of human experts; they find two important aspect very valuable for their diagnosis work: a region segmentation provided in a way that emulates their point of view and a knowledge representation that is able to resume some key-aspects of benignant and malignant melanoma.

For the first point, we actually emulate with the recursive FCM the exploration approach for finding similarity between lesion points: this work is moved from the observation that benign lesions are formed with a growing process from an highly pigmented center towards slightly pigmented surroundings, while many melanomas present a less regular structure because of peripheral growing, regression areas, depigmentation zones or different colored dark areas. In this manner we can represent these areas: it does not matter if they are precisely segmented at pixel level, but is important the possibility of further compute their spatial relationships such as relative symmetry, inclusion ,relative position and so on.

However, not only the set of regions but also the TT itself, can be used for diagnosis hint. The tree structure, and in particular the number of leafs and the number of branches and levels could be significant features for lesion classification. Benignant lesion are generally very regular so that often the tree degenerates in a simple list (with a single region included in its father), instead malignant melanoma present a more complex tree. Obviously this can not be considered the single discriminant feature but the tree structure could be used in classification combined with other visual descriptors. For instance, proceeding with the dichotomy, even in benign lesions more than one disjoint areas included one into another are created: nevertheless they are often symmetrically distributed

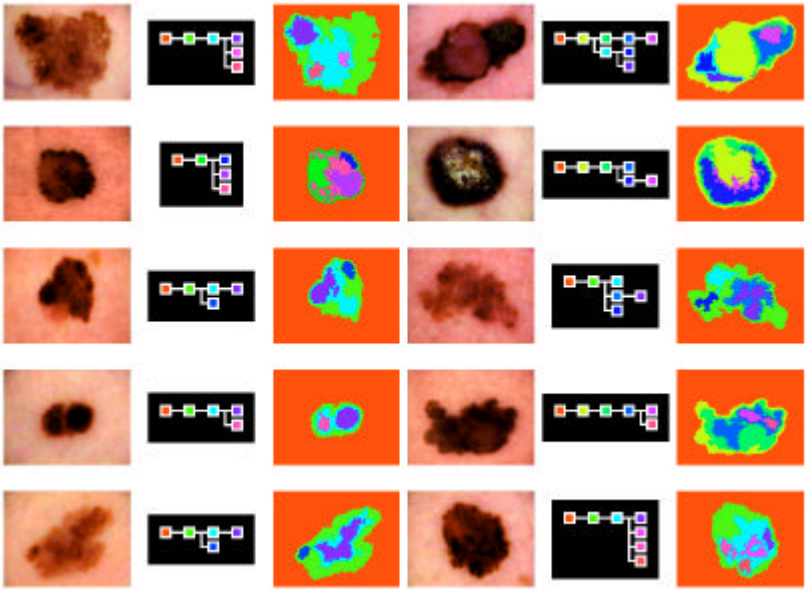

Figure 3. TT results on melanomas.

in the father region, and symmetrically divided again. Therefore , even if not unique, the TT structure can be taken into account as a meaningful signature of the lesion.

4. Conclusions

The main goal of this work is to present the potential capability of the knowledge representation method of the Topological Tree, applied to dermatological images.

Overall performance was considered quite good based on dermatologist's opinion, finding it a useful hint about regions of interest. Cases of wrong segmentation were produced by unsatisfactory evaluation of region inclusion, since regions should be considered included in some cases, even if they're not perfectly included from a strict topological point of view; different topological criteria are under evaluation.

The proposed technique presents some interesting novelties and is able to provide the user with a view of the lesion's structure. An interesting future direction is the study of a reliable metric to evaluate the results, integrating dermatology's suggestions and knowledge to improve automatic classification with more traditional visual features.

References

[1] Z. B. Argenyi, Dermatoscopy (epiluminescence microscopy) of pigmented skin lesions, *Dermatologic Clinics*, 15(1):79-95, 1997

[2] M.G. Fleming, C. Steger, J. Zhang, J. Gao, A.B. Cognetta, I. Pollak, C.R. Dyer, Techniques for a structural analysis of dermatoscopic imagery, *Computerized Medical Imaging and Graphics*, 22(5):375-389, 1998

[3] G.A. Hance, S.E. Umbaugh, R.H. Moss, W.V. Stoecker, Unsupervised Color Image Segmentation with Application to Skin Tumor Borders, *IEEE Engineering in Medicine and Biology*, 15(1):104-111, 1996

[4] R. Cucchiara, C. Grana, M. Piccardi, Iterative fuzzy clustering for detecting regions of interest in skin lesions, *AI*IA Notizie*, 15(1):36-39, 2001

[5] R. Cucchiara, C. Grana, A. Prati, S. Seidenari, G. Pellicani, Building the Topological Tree by Recursive FCM Color Clustering, accepted for pubblication in *IAPR International Conference on Pattern Recognition (ICPR 2002)*

[6] Z. Wu, R. Leahy, An Optimal Graph Theoretic Approach to Data Clustering: Theory and Its Application to Image Segmentation, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 15(11):1101-1113, 1993