Scientific Research

Contour-Based Image Segmentation Using Selective Visual Attention

Engin Mendi[1], Mariofanna Milanova[2]

[1]Department of Applied Science, University of Arkansas at Little Rock, Little Rock, United States; [2]Department of Computer Science, University of Arkansas at Little Rock, Little Rock, United States.
Email: esmendi@ualr.edu

ABSTRACT

In many medical image segmentation applications identifying and extracting the region of interest (ROI) accurately is an important step. The usual approach to extract ROI is to apply image segmentation methods. In this paper, we focus on extracting ROI by segmentation based on visual attended locations. Chan-Vese active contour model is used for image segmentation and attended locations are determined by SaliencyToolbox. The implementation of the toolbox is extension of the saliency map-based model of bottom-up attention, by a process of inferring the extent of a proto-object at the attended location from the maps that are used to compute the saliency map. When the set of regions of interest is selected, these regions need to be represented with the highest quality while the remaining parts of the processed image could be represented with a lower quality. The method has been successfully tested on medical images and ROIs are extracted.

Keywords: *Active Contours, Selective Visual Attention, Image Segmentation, Telemedicine*

1. Introduction

Identifying and extracting the region of interest (ROI) accurately is an important step before coding and compressing the image data for efficient transmission or storage. The main requirement for multimedia encoding techniques is achieving high level ratio of compression for effective use of bandwidth and energy consumption. There is an increased demand for faster transmitting diagnostic medical images in telemedicine applications. ROI must be compressed by lossless or near lossless algorithm while on the other hand, the background region must be compressed with some loss of information that is still recognizable using JP2K standard or Inverse Difference Pyramidal (IDP) decomposition (**Figure 1**).

There are a wide variety of approaches for the segmentation problem. One of the popular approaches is active contour models, also called snakes. The basic idea is to start with a curve around the object to be detected, the curve moves towards an "optimal" position and shape by minimizing its own energy. Based on the Mumford-Shah functional [1-3] for segmentation, Chan and Vese [4] proposed a new level set model for active contours to detect objects whose boundaries are not necessarily defined by a gradient.

Visual attention is the process of selecting and getting visual information based on saliency in the image itself (bottom-up), and on prior knowledge about scenes, objects and their interrelations (top-down) [5,6]. Visual attention addresses both problems by selectively enhancing perception at the attended location, and by successively shifting the focus of attention to multiple locations. It is also important for selecting the object of interest from the input information and [7] provides the brain with a mechanism of focusing computational resources on one object at a time, either driven by low-level image properties (bottom-up attention) or based on a specific task (top-down attention). Moving the focus of attention to locations one by one enables sequential recognition of objects at these locations. The more one knows about an image, the higher the top-down influence part will be. On the other side, for an unknown image, the bottom-up attention mechanism is very important. This is the case when no medical doctor is sending remotely the image.

Hu *et al.* [8] used visual attention algorithm to define a method leading to the automatic choice of the best features for a given medical application. Mancas presents application of computational attention in medical images
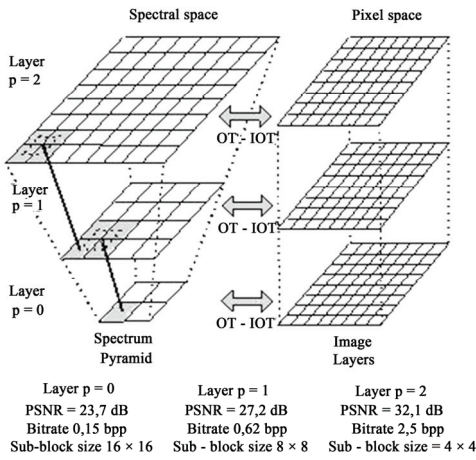
Figure 1. Illustration of the image decomposition called Inverse Difference Pyramid (IDP) [9]

[10]. Attention may be due to: 1) local properties (a feature saliency depends on its neighborhood); 2) global properties (a feature saliency depends on the whole visual field). Attention model can be applied directly on the medical images in order to find rare grey level: for instance liver images, where only the grey level variations should be enough to detect pathologies.

Here ROI was extracted with active contours based on selective visual attention. Chan-Vese active contour model is used for image segmentation and attended locations are determined by SaliencyToolbox [11] which is extension of the saliency map-based model of bottom-up attention [12], by a process of inferring the extent of a proto-object at the attended location from the maps that are used to compute the saliency map. In this paper we extend our previous study of markless segmentation of medical images [13]. Here we compare results using different local and global features for a coarse localization of possibly pathological areas. We also show the results extracting multiple ROIs in a single image. The paper is organized as follows: Section 2 provides an overview of the Chan-Vese model. Section 3 presents the bottom-up salient region selection model. Section 4 describes the application of our approach. Section 5 presents the conclusions of this paper in a summary.

2. Chan-Vese Model

The Mumford-Shah model [1-3] is a variational problem for approximating a given image by a piecewise smooth image of minimal complexity. Let u be differentiable on R and allowed to be discontinuous across C, Mumford-Shah energy functional is as follows:

$$F(\phi, C) = \frac{1}{\sigma^2} \int_R (\phi - f)^2 \, dx + \int_{R \setminus C} \left\| \nabla \phi \right\|^2 dx + \lambda \left| C \right| \quad (1)$$

where R is the image domain, f is the feature intensity,

C is the curve, ϕ is the smoothed image, $\left| C \right|$ is the arc length of C and σ, λ are positive parameters. Segmentation problem is restated as finding optimal approximations of g by piece-wise smooth functions u, whose restrictions to the regions are differentiable.

The Chan-Vese model [4] is a special case of the Mumford Shah model by restricting (1) to piece-wise constant functions ϕ and looking for the best approximation ϕ of f taking only two values. Then the energy functional in (1) is expressed in terms of the level set function by replacing the C by Lipschitz function ϕ:

$$F(\phi, c_1, c_2) = \int_R \left\{ \left| \nabla H_\varepsilon(\phi) \right| + \lambda \left[H_\varepsilon(\phi)(c_1 - f)^2 \right. \right.$$
$$\left. \left. + (1 - H_\varepsilon(\phi))(c_2 - f)^2 \right] \right\} dx \quad (2)$$

where H is the Heaviside function, defined by:

$$H(z) = \begin{cases} 1, & if \quad z \geq 0 \\ 0, & if \quad z < 0 \end{cases}$$

and H_ε is the regularization of H.

Constant functions c_1 and c_2 of level sets can be expressed by minimizing the energy functional with respect to the constants and keeping the level sets fixed:

$$c_1(\phi) = \frac{\int_D f H_\varepsilon(\phi) dx}{\int_D H_\varepsilon(\phi) dx} \quad (3)$$

$$c_2(\phi) = \frac{\int_D f(1 - H_\varepsilon(\phi)) dx}{\int_D (1 - H_\varepsilon(\phi)) dx} \quad (4)$$

Combining the energy terms and replacing the singular term $H_\varepsilon'(\phi)$ by $\left| \nabla \phi \right|$, the corresponding Euler-Lagrange equation for ϕ, using gradient descent in artificial time leads to:

$$\frac{\partial \phi}{\partial t} = \left| \nabla \phi \right| \left\{ \kappa(\phi) - \lambda [(c_1 - f)^2 - (c_2 - f)^2] \right\} \quad (5)$$

where $\kappa(\phi)$ is the curvature of the level sets and $\kappa(\phi) = -div(\frac{\nabla \phi}{\left| \nabla \phi \right|})$. A multigrid scheme on the discretized Euler-Lagrange Equation (5) is used for the minimization of Chan-Vese energy functional.

$$\inf_{\phi, c_1, c_2} F(\phi, c_1, c_2) \quad (6)$$

which is

$$\min_\phi F(\phi, c_1, c_2) = \int_D \left\{ \left| \nabla \phi \right| + \lambda [(c_1 - f)^2 \right.$$
$$\left. + (c_2 - f)^2] \phi \right\} dx \quad (7)$$

The explicit formula provided by (5) is solved by us-

ing gradient descent procedure as described in [14].

3. Bottom-Up Salient Region Selection Model

The model of bottom-up salient region selection presented by [7,11] based on the model of saliency-based bottom-up attention by Itti-Koch [15,16] is implemented as part of the SaliencyToolbox [11]. This model introduces a process of inferring the extent of a proto-object at the attended location from the maps that are used to compute the saliency map.

Itti-Koch model [15,16] is a bottom-up selective visual attention based on serially scanning a saliency map that is computed from local feature contrasts, for salient locations in the order of decreasing saliency (**Figure 2**). Presented with a manually preprocessed input image, their model replicates human viewing behavior for artificial and natural scenes.

Visual input [14] is first decomposed into a set of topographic feature maps. Different spatial locations then compete for saliency within each map, such that only locations which locally stand out from their surround can persist. All feature maps feed, in a purely bottom-up manner, into a master saliency map. The purpose of the saliency map is to represent the saliency at every location in the visual field by a scalar quantity and to guide the selection of attended locations, based on the spatial distribution of saliency. However this model's usefulness [17] as a front-end for object recognition is limited by the fact that its output is merely a pair of coordinates in the image corresponding to the most salient location.

This model is extended [7,11] by a process of inferring the extent of a proto-object, contiguous region of high activity in feature map, at the attended location from the maps t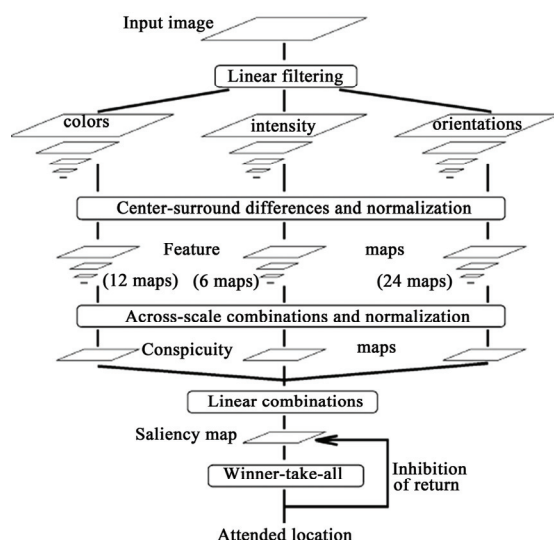hat are used to compute the saliency map. This is achieved by introducing feedback connections in the saliency computation hierarchy in order to estimate the proto-object region based on the maps and salient locations computed in Itti-Koch model [15,16]. Different visual features that contribute to attentive selection are combined into one single topographically oriented saliency map which integrates the normalized information from the individual feature maps into one global measure of conspicuity.

The locations [7] in the saliency map compete for the highest saliency value by means of a winner take-all (WTA) networks of integrate-and-fire neurons. The winning of this process is attended to, and the saliency map is inhibited. Continuing WTA competition produces the second most salient location, which is attended to subsequently and then inhibited, thus allowing the model to simulate a scan path over the image in the order of decreasing saliency of the attended locations.

4. Experimental Results

Image segmentations of attended locations of four medical images were used in the application of the new approach to image segmentation. All conspicuity maps, saliency maps, WTAs and attended locations are operated by SaliencyToolbox [11]. The saliency map is summed by conspicuity maps that provide information of color, intensity and orientation. The attended locations are set as initial contours to be segmented by using Chan-Vese Model [4].

For example, in **Figure 3**, seborrheic keratosis is segmented from a skin image. **Figure 4** shows multiple basal cell carcinoma segmentation. **Figure 5** and **Figure 6** show segmentation of cherry angiomas of the trunk and basal cell carcioma of the cheek, respectively. **Table 1** shows the stimulated time (ms) that attended locations (AL) took. Global low level attention is applied directly on the medical images. Low level features bring some top down information about grey levels. A final attention map, for example **Figure 3(g)**, can help the contour segmentation algorithm by focusing only at separated regions with the greatest chance of being pathological. This approach works on the images where pathological pixel grey level is different from normal tissues grey-level.

For telemedicine applications, we have integrated image segmentation with adaptive compression technique. The proposed compression technique is based on the hypothesis that image resolution exponentially decreases from the fovea to the retina periphery. This hypothesis

Figure 2. General architecture of Itti-Koch model [14]

Table 1. Simulated time (ms) of attended locations

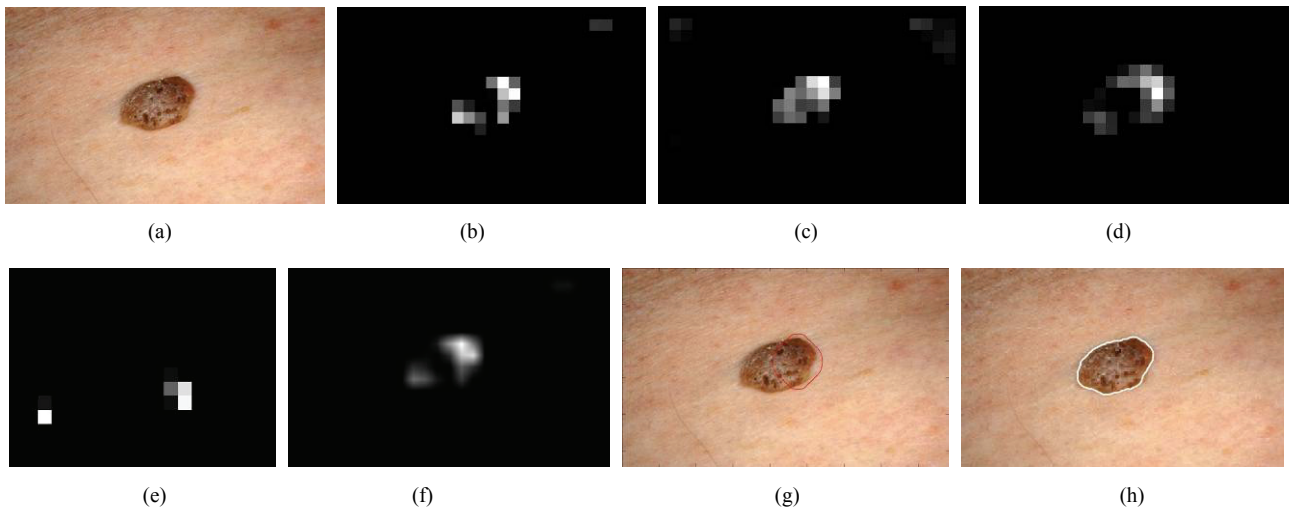|	1st AL	2nd AL	3rd AL	4th AL
Figure 3	239,98			
Figure 4	39,171	200,151		
Figure 5	174,193	72,18		
Figure 6	98,189	157,132	167,242	185,115

Figure 3. (a) Input image; (b) conspicuity map for color contrast; (c) conspicuity map for intensity contrast; (d) conspicuity map for orientation contrast; (e) saliency map combined by conspicuity maps; (f) WTA map for the attended location; (g) attended location; (h) active contours based on the attended location
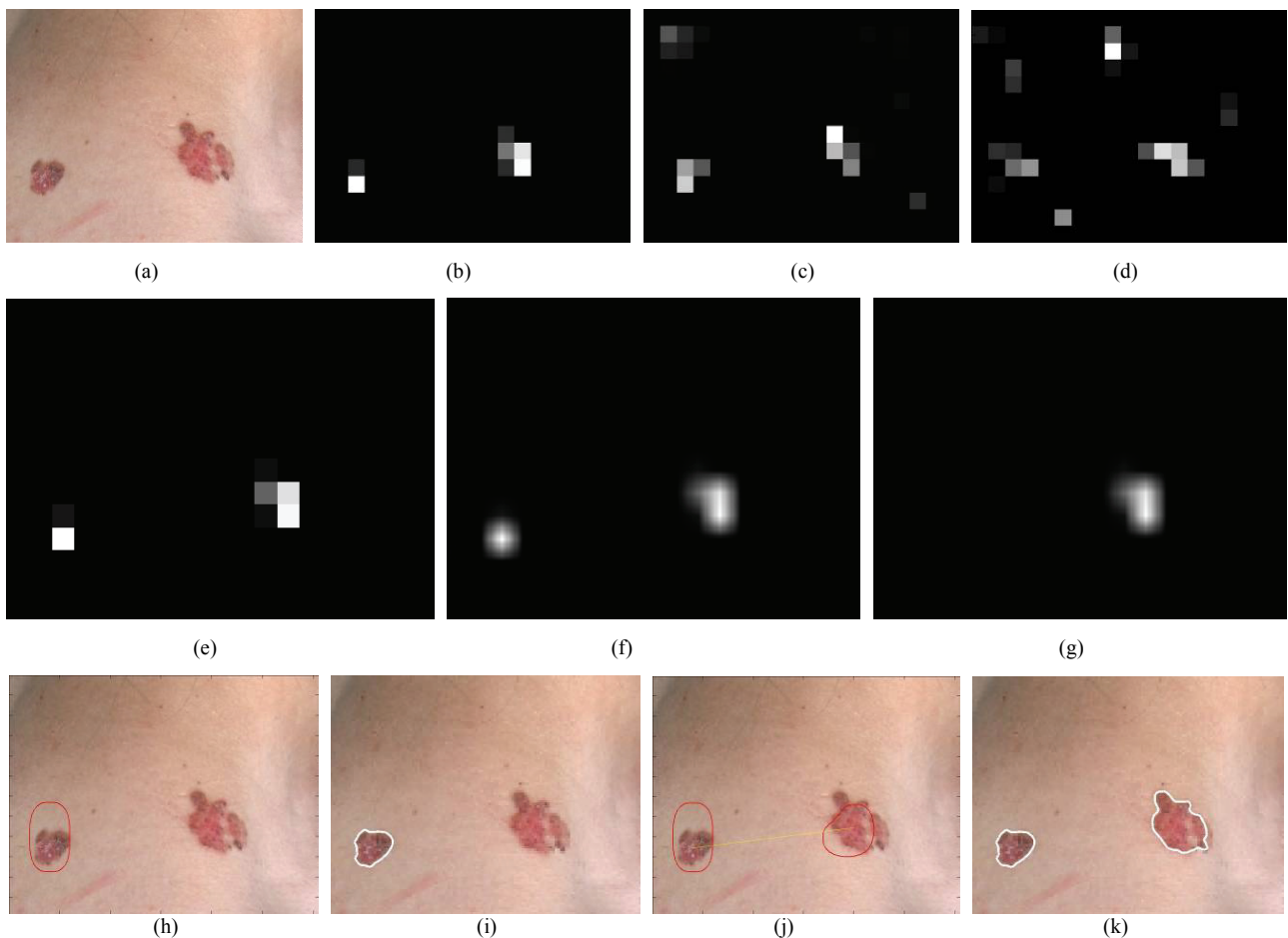

Figure 4. (a) Input image; (b) conspicuity map for color contrast; (c) conspicuity map for intensity contrast; (d) conspicuity map for skin contrast ; (e) saliency map combined by conspicuity maps; (f) WTA map for the first attended location; (g) WTA map for the second attended location; (h) first attended location; (i) active contours based on first attended location; (j) second attended location; (k) active contours based first two attended locations

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

(e)   (f)   (g)

(h)   (i)   (j)   (k)
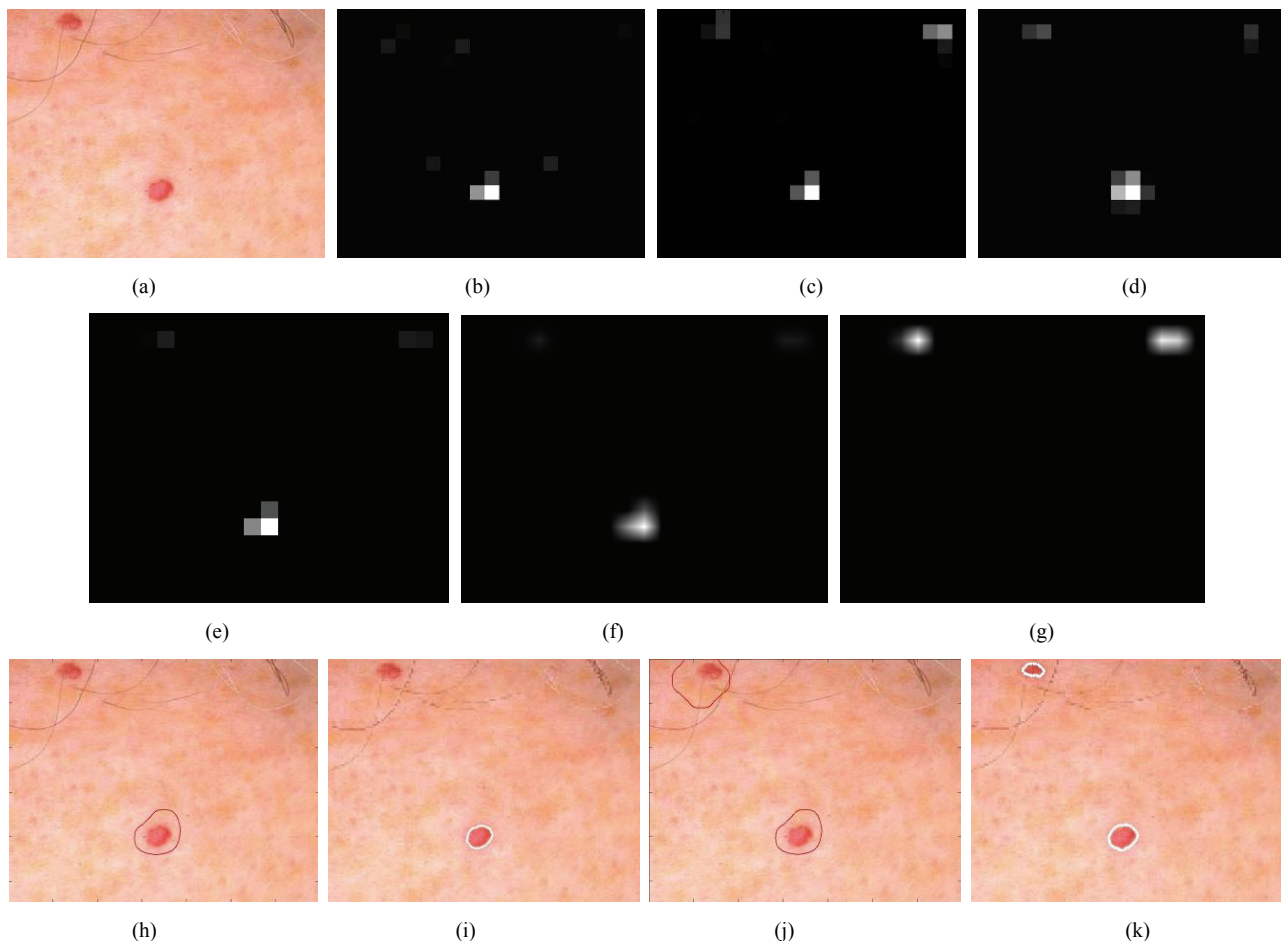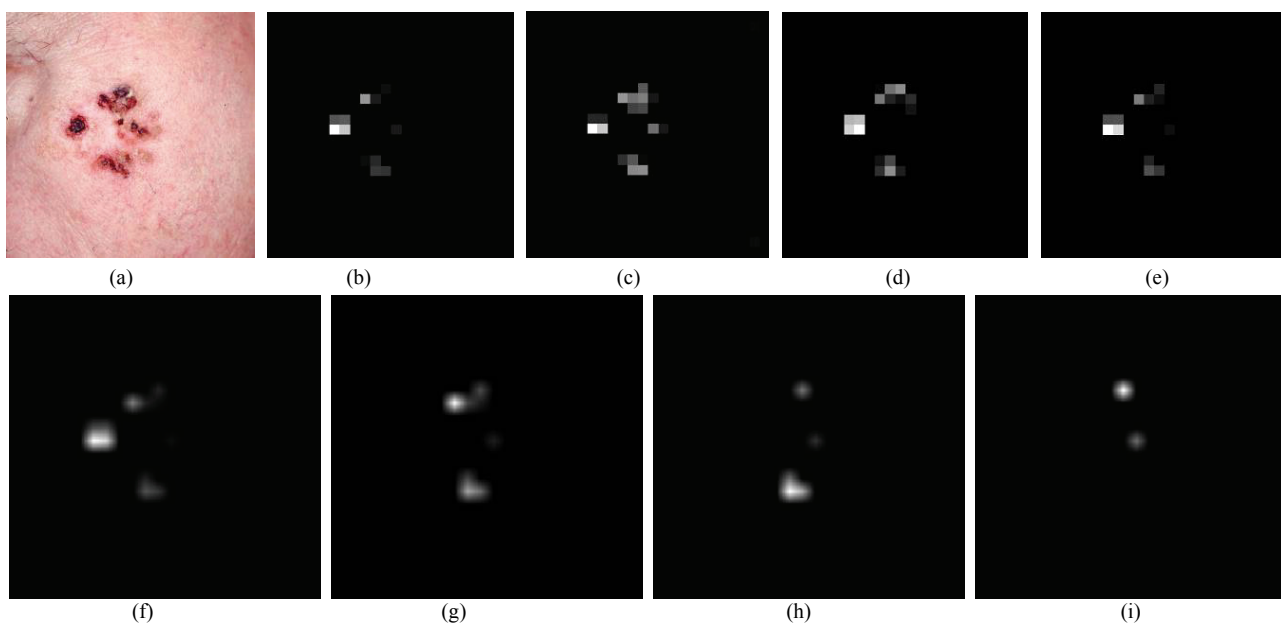
Figure 5. (a) Input image; (b) conspicuity map for color contrast; (c) conspicuity map for intensity contrast; (d) conspicuity map for orientation contrast; (e) saliency map combined by conspicuity maps; (f) WTA map for the first attended location; (g) WTA map for the second attended location; (h) first attended location; (i) active contours based on first attended location; (j) second attended location; (k) active contours based first two attended locations

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

(f)   (g)   (h)   (i)
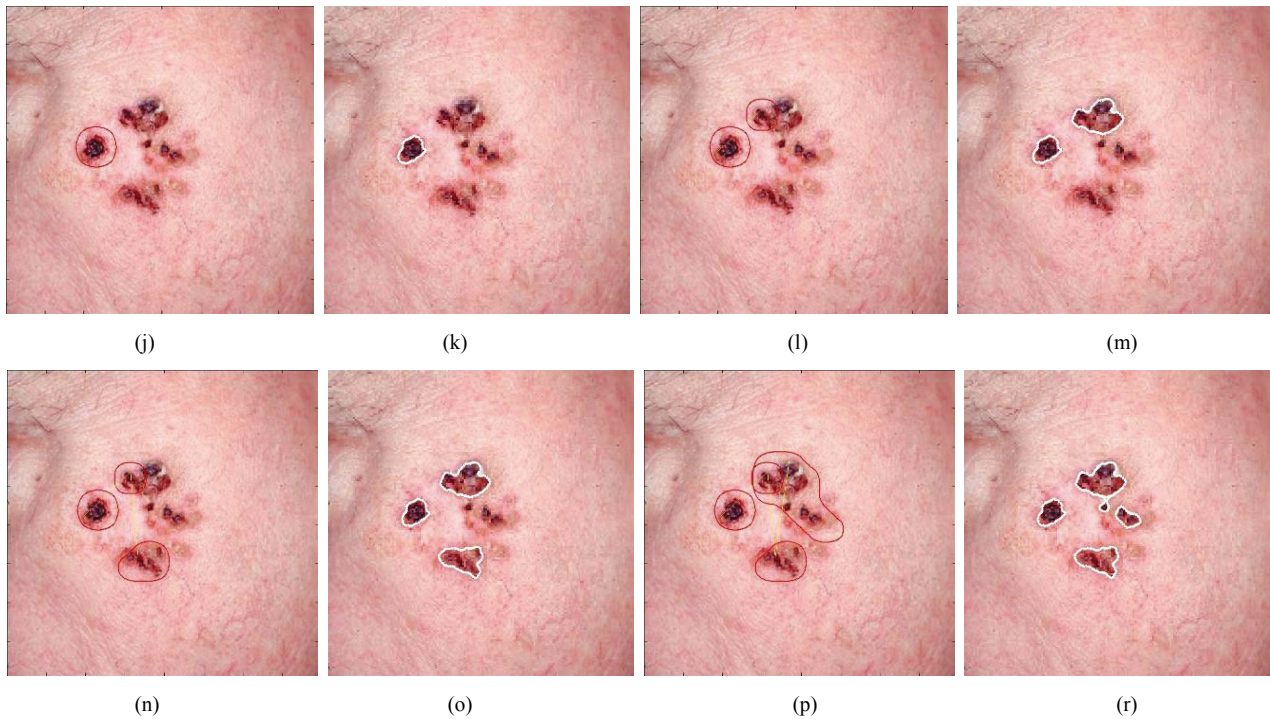
JSEA

Figure 6. (a) Input image; (b) conspicuity map for color contrast; (c) conspicuity map for intensity contrast; (d) conspicuity map for orientation contrast; (e) saliency map combined by conspicuity maps; (f-i) WTA maps for the first, second, third and fourth attended locations, respectively; (j, l, n, p) WTA maps for the first, second, third and fourth attended locations, respectively; (k, m, o, r) active contours based first fourth attended locations, respectively

can be represented computationally with different resolutions. The visual attention points may be considered as the most highlighted areas of the visual attention model. These points are the most salient regions in the image. When going further from these points of attention, the resolution of the other areas dramatically decrease. Different authors work with different filters and different kernel size to mimic this perceptual behavior [18]. These models ignore contextual information representation. When the set of regions of interest is selected, these regions need to be represented with the highest quality while the remaining parts of the processed image could be represented with a lower quality. In result, higher compression is obtained. The adaptive compression technique proposed is based on new image decomposition called Inverse Difference Pyramid (IDP) [9]. This approach is developed by analogy with the hypothesis for the way humans do image recognition using consecutive approximations with increasing similarity. A hierarchical decomposition is used for the image representation. The approximations in the consecutive decomposition layers are represented by the neurons in the hidden layers of the neural networks (NN) [19]. The most specific features of IDP method are that the images are processed in consecutive layers with higher quality. This approach offers the ability to transfer the image via Internet layer by layer, without sending the same information twice.

5. Conclusions

The paper presents a new markerless approach for medical image segmentation by combining saliency attention maps with active contours. The Chan-Vese active contour model [4] has been implemented by setting attended locations as initial contours. Attended locations are extracted with SaliencyToolbox [11]. It is anticipated that this process will be useful for identifying and extracting the ROI accurately. The combination of the two techniques minimizes user interaction and speeds up the entire segmentation process. The method has been successfully tested on medical images and the ROI is extracted. The proposed approach works for allocating tumors in medical images.

REFERENCES

[1] D. Mumford and J. Shah, "Optimal Approximation by Piecewise Smooth Functions and Associated Variationol Problems," *Communications on Pure and Applied Mathematics*, Vol. 42, 1989, pp. 577-685.

[2] D. Mumford and J. Shah, "Boundary Detection by Minimizing Functionals, I," *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, San Francisco, 1985, pp. 22-26.

[3] J. Shah, "Piecewise Smooth Approximations of Functions," *Calculus of Variations and Partial Differential Equations*,

Vol. 2, 1994, pp. 315-328.

[4]   T. F. Chan and L. A. Vese, "Active Contours without Edges," *IEEE Transactions on Image Processing*, Vol. 10, No. 2, 2001, pp. 266-277.

[5]   D. Walters, U. Rutishauser, C. Koch and P. Perona, "Selective Visual Attention Enables Learning and Recognition of Multiple Objects in Cluttered Scenes," *Computer Vision and Image Understanding*, Vol. 100, No. 1-2, 2005, pp. 41-63.

[6]   M. Milanova, S. Rubin, R. Kountchev, V. Todorov and R. Kountcheva, "Combined Visual Attention Model for Video Sequences," *IEEE ICPR'2008 International Conference on Pattern Recognition*, Tampa, 2008, pp. 1-4.

[7]   D. Walter, "Interactions of Visual Attention and Object Recognition: Computational Modeling, Algorithms, and Psychophysics," PhD Thesis, California Institute of Technology, Pasadena, 2006.

[8]   X.-P. Hu, "Hot Spot Detection based on Feature Space Representation of Visual Search," *IEEE Transactions on Medical Images*, Vol. 22, No. 9, 2003, pp. 1152-1162.

[9]   A. Kountchev and A. Mironov, "Algorithms for Pyramid Image Decomposition," In: N. Mastorakis Ed., *Intelligent Systems and Computer Science*, WSEAS Press, Danvers, 1999, pp. 196-200.

[10] M. Mancas, "Image Perception: Relative Influence of Bottom-Up and Top-Down Attention," *Attention in Cognitive Systems*, *Lecture Notes in Computer Science*, Vol. 5395, Greece, 2008, pp. 212-226.

[11] D. Walther and C. Koch, "Modeling Attention to Salient Proto-Objects," *Neural Networks*, Vol. 19, No. 9, 2006, pp. 1395-1407.

[12] C. Koch and S. Ullman, "Shifts in Selective Visual-Attention—Towards the Underlying Neural Circuitry," *Human Neurobiology*, Vol. 4, No. 4, 1985, pp.219-227.

[13] E. Mendi and M. Milanova, "Image Segmentation with Active Contours Based on Selective Visual Attention," 8*th WSEAS International Conference on Signal Processing* (*SIP*'09), Istanbul, 2009, pp. 79-84.

[14] K. Fundana, N. C. Overgaard and A. Heyden, "Deformable Shape Priors in Chan-Vese Segmentation of Image Sequences," *Proceedings International Conference on Image Processing*, San Antonio, Vol. 1, 2007, pp. 285-288.

[15] L. Itti, C. Koch and E. Niebur, "A Model of Saliency-Based Visual Attention for Rapid Scene Analysis," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 20, No. 11, 1998, pp. 1254-1259.

[16] C. Koch and L. Itti, "Computational Modeling of Visual Attention," *Nature Reviews Neuroscience*, Vol. 2, 2001, pp. 194-203.

[17] D. Walther, U. Rutishauser, C. Koch and P. Perona, "On the Usefulness of Attention for Object Recognition," *Workshop on Attention and Performance in Computational Vision at ECCV*, Prague, 2004, pp. 96- 103.

[18] M. Mancas, B. Gosselin and B Macq, "Perceptual Image Representation," *EURASIP Journal of Image and Video Processing*, Vol. 2007, 2007, pp. 1-9.

[19] R. Kountchev, S. Rubin, M. Milanova, V. Todorov and R. Kountcheva, "Non-Linear Image Representation Based on IDP with NN," *WSEAS Transactions on Signal Processing*, Istanbul, Vol. 5, No. 9, 2009, pp. 315-325.