A BINARY LEVEL SET METHOD BASED ON K-MEANS FOR CONTOUR TRACKING ON SKIN CANCER IMAGES

Azadeh Noori Hoshyar, Adel Ali Al-Jumaily, Yee Mon Aung
University of Technology Sydney, Broadway 2007, Australia
Azadeh.NooriHoshyar@student.uts.edu.au, Adel.Al-Jumaily@uts.edu.au, Yee.M.Aung@student.uts.edu.au

ABSTRACT
A great challenge of research and development activities have recently highlighted in segmenting of the skin cancer images. This paper presents a novel algorithm to improve the segmentation results of level set algorithm with skin cancer images. The major contribution of presented algorithm is to simplify skin cancer images for the computer aided object analysis without loss of significant information and to decrease the required computational cost. The presented algorithm uses k-means clustering technique and explores primitive segmentation to get initial label estimation for level set algorithm. The proposed segmentation method provides better segmentation results as compared to standard level set segmentation technique and modified fuzzy c-means clustering technique.

KEY WORDS
Segmentation, Clustering, Level set, K-means, Skin cancer.

1. Introduction
Segmentation is an essential issue in digital image processing and is used both for image description and as an important step before classification. The segmentation algorithms are generally based on similarity, which can be categorized as; thresholding [1], template matching [2], [3], region growing [4], edge detection [5], and clustering [6].

Clustering is a process of classifying a set of objects into classes with similar characteristics. It has been widely applied in many areas such as image processing [7], [8], machine learning [9],[10], pattern recognition [11]-[13], data mining [14]-[16], statistics [17], [18]. Recently clustering algorithms found crucial applications in medical imaging field [19].

Among clustering algorithms, k-means is the most popular one due to its simplicity and fast running speed. K-means is known as a simple and fast numerical, non-deterministic, unsupervised and iterative method which has proved to provide good clustering results [20]-[23].

Level set method is another powerful and robust segmentation technique which is flexible under challenging conditions. It depends on both extrinsic and intrinsic factors such as intensity and curvature, respectively [24][25]. Different researchers [26][27]indicated that level set method may decrease the mutability of complex segmentation tasks in medical applications. They mentioned the flexibility of level set techniques cause to lengthen computation time and consequently will limit its application in medical area. In some other works [28][29],the level set algorithm has been proposed to not be used merely for classification purposes.

In traditional level set methods, initialization extensively plays an essential role in curve evolution process. Re-initialization of level set function is significantly used as a numerical remedy to keep the curve evolution more stable and achieve effective results. Though, as pointed out by Gomes, many presented re-initialization schemes provide undesirable results along with increased computational cost [30]. On the other hand, it is becoming significantly obvious that none of the methods alone are adequate and that the application of different approaches will clear various aspects of data to be explored [31]-[34]. Thus, this work attempts to incorporate the strength of two techniques and overcome the disadvantages. Here, an implicit joint k-means-level set algorithm for classifier decision boundaries is presented. It applies k-means algorithm to get initial label estimation for level-set algorithm. In short, the contribution of this work is to propose a new and creative segmentation algorithm for development of the medical expert systems for achieving increased precision in diagnosis. The algorithm increases the accuracy of segmentation while decrease the computational cost.

The performance of proposed algorithm of level set classification is tested and compared using a data set comprising of forty skin cancer images. The parameters for performance evaluation include False positive error, Hammoude distance, True detection rate and Similarity. The results showed that the proposed approach is quite competitive with many classifiers that are being employed in practice.

The rest of this paper is structured as follows: We shortly go over the k-means framework in Section 2. In Section 3, the proposed algorithm is explained in detail. Experiment results are demonstrated in Section 4, and Section 5 is assigned to conclusion.

2. The Basic K-means
K-means algorithm is a clustering algorithm which firstly presented by MacQueen in 1967[35]. This algorithm divides the pixels into k clusters and heavily relies on
selecting the number of clusters k and initial cluster centroids \( v_i \), \( i = 1, \ldots, 2, k \). The centroids of clusters are calculated based on the average of pixel intensities in each cluster. These initial centers effectively influence on number of iterations in k-means algorithm. After calculating the centroids \( v_i \), the distance of pixels and centres are estimated and each pixel \( x_j \) is iteratively assigned to the closest cluster as in equation 1.

\[
de_i = \| x_j - v_i \| 
\]

The matrix of \( U \) with the membership values are determined by

\[
U = [u_{ij}]
\]

Where \( u_{ij} \in [0, 1] \) for all \( i \) and \( j \). \( \sum_t^k u_{ij} = 1 \) for all \( j \) and \( 0 < \sum_{j=1}^n u_{ij} < n \) for all \( i \). (k=number of clusters, t=0, \( n= \)number of pixels)

The cluster centres are updated by computing the mean of each cluster as in equation 3.

\[
V_i = \frac{\sum_{j=1}^n u_{ij} x_j}{\sum_{j=1}^n u_{ij}} \text{ for all } i
\]

This process is repeated till it reaches the center values same as the prior values since it indicates the current values are the optimal results.

This algorithm optimize the objective function \( J_w(U, v) \) in a manner that \( kn \)

\[
J_w(U, V) = \sum_{t=1}^k \sum_{j=1}^n \| X_j - V_i \|^2_2,
\]

It is spotted that the algorithm is sensitive to cluster initialization and distance measure [19][20][36].

3. The Level Set Framework

Level set algorithms, firstly presented by Osher and Sethian [37, 38], offer an impressive implementation of curve evolution. This approach is based on enchasing contour \( C \) as the zero level set of the graph of a higher dimensional function \( \Phi(x, y, k) \), as in equation 5

\[
C_k = \{ (x, y) \mid \Phi(x, y, k) = 0 \}
\]

Where \( k \) represents an artificial time-marching parameter, and then evolves the graph as moves pursuant to the prescribed flow. Therefore, the level set can change topology and expand singularities as remains smooth and preserve the form of a graph.

In this manner, the curve evolution is defined as equation 6.

\[
\frac{\partial c}{\partial k} = VN
\]

Where \( V \) is the speed of curve evolution, \( N \) is the normal vector of inward unit. From \( \partial \Phi(C_k, k) \), the following equation is achieved:

\[
(\partial \Phi/(\partial k)) + \nabla \Phi \cdot (\partial \Phi/(\partial k)) \cdot 0
\]

According to the level set function definition described above, the vector \( N \) may be written as \( N = -\nabla \Phi/||\nabla \Phi|| \). Then it can be implemented corresponding to curve evolution equation (8):

\[
\frac{\partial \Phi}{\partial k} = V ||\nabla \Phi||
\]

The initial level set function is generated using the initial given curve. In addition, the function requires to be re-initialized consecutively during the update process which generally takes a lot of computing time.

4. The Proposed Active Contour Tracking Method

The presented method seeks a new target location \( t1 \) in the current frame exploiting the k-means procedure and initializing the level set segmentation algorithm from the location \( t0 \) of the target in prior frame. The weights are computed based on the scale of bandwidth \( h \) centered at \( t0 \). In other word, the initial curve of each sample is set up and evolved based on the target location \( t0 \) obtained by the k-means algorithm. The energy function is refined using the previous knowledge of target model achieved by k-means. The presented active contour tracking algorithm is explained in detail as follow.

4.1 The Proposed Algorithm

As described above, k-means clustering algorithm needs \( K \) clusters which should be initialized manually. The number of clusters is 3 in this paper because the skin cancer images are expected to be clustered into less or equal to 3 parts consisting the background (skin), tumor, and possible extra parts. That is, the first cluster is named cluster 1 and the last one is cluster 3. Since the skin cancer image is gray-scale image in which the minimum intensity value is equal to 0 and the maximum intensity value is equal to 255, the centroid of cluster 1 is 0 and the centroid of cluster 3 is 255. Equation 9 indicates the calculation of centroid in cluster \( k \), \( C_k \)

\[
C_k = \text{rand} \times 255 \quad (k = 1 \text{ and } k < = 3)
\]

Input: The original skin cancer image

Output: Initial Segmented image

1. Initialize centroid of clusters randomly
2. loop while \( CTr = CTr-1 \)
   - If \( CTr = CTr-1 \), then disk, \( i = \{ CTr - I \} \) for all pixels
   - Assign pixels to the closest distance
   - Estimate the new centroid for 3 clusters \( ctk = \text{Avg}(I_k) \)
3. The segmented skin lesion is achieved at location \( t1 \) with the scale of \( h \)
4. End loop

Input:
- The original skin cancer image

Output:
- Initial Segmented image

1. Initialize centroid of clusters randomly
2. loop while \( CTr = CTr-1 \)
   - If \( CTr = CTr-1 \), then disk, \( i = \{ CTr - I \} \) for all pixels
   - Assign pixels to the closest distance
   - Estimate the new centroid for 3 clusters \( ctk = \text{Avg}(I_k) \)
3. The segmented skin lesion is achieved at location \( t1 \) with the scale of \( h \)

Algorithm 1. K-means clustering with 3 clusters
Algorithm 1 shows the total procedure of k-means clustering of this paper. After determining the initial centroid values of the clusters, the pixels are distributed to one of 3 clusters which the centroid value is the most closest to that pixel intensity. The centroids of clusters are updated by the mean intensity values of its pixels. This process is repeated till reaching the constant value. The target segmented location t1 achieved by this algorithm is used to initialize the level set algorithm. This segmented result is converted to binary for feeding the level set algorithm.

For each sample \( t_i^k \) in level set framework, we initialize a curve by a target segmented location \( t_1 \) with the scale of \( h \) achieved by k-means segmentation algorithm. Hence we evolve the curve using level set algorithm upon the time \( k, I_k \), and the target model \( t \):

\[
C_k^i = \text{evo} \left( s_k^i, I_k, t \right) = S_k^i(M) \tag{10}
\]

Where \( S_k \) indicates the contour at time \( k \) and iterates \( M \) times to the direction in which the energy function \( E_{\text{img}} \) is reduced

\[
E_{\text{img}} = E_R(c_1, c_2, \partial) \tag{11}
\]

Where \( c_1, c_2 \) indicate positive constant, \( \partial \) is the indicator of regions (\( \partial = 1 \) lesion, \( \partial = -1 \) background)

By the end of this process, the true target segmentation is achieved which includes the energy smaller than other samples in evolution process. Therefore, the algorithm of this process is illustrated in Algorithm 2.

| Input: target segmented location \( t_1 \) with a scale of \( h \) achieved by k-means segmentation algorithm |
| Output: True target segmentation \( t_2 \) |

1. Initialize a curve by a target segmented location \( t_1 \)
2. Run curve evolution in \( M \) iterations toward the direction of energy reduction
3. Calculate the weights
4. If the \( (t_2 - t_1) < \varepsilon \) and mark \( t_2 \) as the result

Algorithm 2. Level set segmentation algorithm initializing by k-means clustering algorithm

5. Experiment Result

In this section, the proposed method is tested on 40 images taken from digital cameras. In order to demonstrate our improved method, multiple algorithms have been run upon the same condition. The first algorithm is the traditional level set (TLS), proposed in [38], the second is modified fuzzy c-means clustering (FCM) [39] in two cut off (\( sw = 0 \), cut between the small and middle class, \( sw = 1 \), cut between the middle and large class), and the last one is the proposed method of this paper.

The original images are initially segmented into 3 clusters for accurate evaluation of lesion area in the image. As this image is used as the curve evolution of level set algorithm, it should be converted to binary image and used to feed the level set algorithm. Figure 1 shows the segmented results achieved by our proposed method (k-means-level set (KLS)).

To perform the comparison of our segmented results with the ground truth images which have been segmented manually by the experts, let \( SR \) and \( GT \) indicate the result of automatic segmentation method and the ground truth segmentation, respectively. Figure 2 indicates this comparison.

Figure 2. Comparison of ground truth image with the segmented image by proposed algorithm a) Ground truth image b) Segmented result c) Pixels in segmented lesion as well as ground truth (Subscription of “a”, “b” images)

Figure 3 demonstrates the results achieved by “modified fuzzy c-means clustering” method which has been applied to compare with our proposed method.
Figure 3. Modified fuzzy c-means clustering in two cut-off position a) Original image b) Otsu thresholding c) FCM (sw=0) d) FCM (sw=1)

Figure 4 shows the comparison which has been done between our proposed method (KLS), TLS, FCM (sw=0) and FCM (sw=1).

Four different metrics of Border error, Similarity, Hammoude distance and Rms error are used to quantify the boundary differences. Each metric is calculated between the segmented results achieved by TLS, KLS, FCT (sw=0) and FCT (sw=1) with ground truth images (manually segmented images by experts). Table 1 shows the statistical comparison of algorithms by these metrics. As it is obvious from table 1, the average border error, Hammoude distance and Rms error between the ground truth and segmented images in KLS is the lowest and similarity is the highest among others. It shows the better performance of KLS than others.

Table 1

<table>
<thead>
<tr>
<th></th>
<th>TLS</th>
<th>KLS</th>
<th>FCT (sw=0)</th>
<th>FCT (sw=1)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Border error</td>
<td>0.176912</td>
<td>0.075871</td>
<td>0.124453</td>
<td>0.550871</td>
</tr>
<tr>
<td>Similarity</td>
<td>0.810429</td>
<td>0.953229</td>
<td>0.892076</td>
<td>0.458453</td>
</tr>
<tr>
<td>Hammoude distance</td>
<td>0.447771</td>
<td>0.301471</td>
<td>0.391818</td>
<td>0.676335</td>
</tr>
<tr>
<td>Rms Error</td>
<td>0.335165</td>
<td>0.198329</td>
<td>0.278412</td>
<td>0.656047</td>
</tr>
</tbody>
</table>

For detecting whether our proposed method has statistically significant difference from others, one-way analysis of variance (ANOVA) as a statistical inference is applied with a .05 significance level. It is known as a popular and powerful tool which is robust to non-homogeneity of the data [40]. The anova test will compare all forty border errors between TLS segmentation results and ground truth with all forty border errors between KLS segmentation results and ground truth with all forty border errors between FCT (sw=0) segmentation results and ground truth and finally with all forty border errors between FCT (sw=1) segmentation results and ground truth. The purpose is to find out how significant is the results of our method as compared to other methods. In other word, determine if our method has significantly decreased the border error as compared with other methods. The same process is performed for other three metrics as well. Table 2, which demonstrate One-way ANOVA results of the skin cancer image data, give evidence to the results achieved above in table 1. It represents mean values of Border error, Similarity, Hammoude distance and Rms error metrics on forty experiments by KLS and TLS, KLS and FCT (sw=0), and KLS and FCT (sw=1) achieved by 1-way anova test. As it can be seen in table 2 on the first column, all the mean values are less than 0.05, which shows the significant difference between KLS than TLS. In second column mean values of Rms error shows the significant difference between KLS and FCT (sw=0), while the other metrics shows although there are improvement (according to table 1), it is not significant. The mean values of third column show the significant difference between KLS and FCT (sw=1).
Table 2
Anova test on Border error, Similarity, Hammoude distance and Rms error between the “proposed method and TLS”, “KLS and FCT (w=0)”, “KLS and FCT (w=1)”

<table>
<thead>
<tr>
<th></th>
<th>KLS, TLS</th>
<th>KLS, FCT (sw=0)</th>
<th>KLS, FCT (sw=1)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Border error</td>
<td>0.0073</td>
<td>0.1332</td>
<td>4.53E-10</td>
</tr>
<tr>
<td>Similarity</td>
<td>0.0068</td>
<td>0.0638</td>
<td>1.92E-10</td>
</tr>
<tr>
<td>Hammoude distance</td>
<td>0.0236</td>
<td>0.1352</td>
<td>5.82E-06</td>
</tr>
<tr>
<td>Rms Error</td>
<td>0.0211</td>
<td>0.0376</td>
<td>2.77E-10</td>
</tr>
</tbody>
</table>

Table 3
Comparison of Elapsed time between proposed method and TLS

<table>
<thead>
<tr>
<th></th>
<th>TLS</th>
<th>KLS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Computational time</td>
<td>0.731007</td>
<td>0.716403</td>
</tr>
</tbody>
</table>

The figure 5 shows the comparative view and difference of Border error, Similarity, Hammoude distance and Rms error by KLS and TLS, KLS and FCM (sw=0), and KLS and FCM (sw=1) have been achieved by Anova test.

As noticed before, another contribution of this paper achieved is computational time which significantly reduced in compare with TLS method. Table 3 indicates the comparison results.

Figure 5. Difference of Border error, Similarity, Hammoude distance and Rms error of KLS and TLS, KLS and FCM (sw=0), and KLS and FCM (sw=1)
The figure 6 shows the difference of computational cost between KLS and TLS.

Table 3  
| KLS, TLS | 0.0384 |

Figure 6. Comparison of computational cost of KLS and TLS by Anova test

Using a joint algorithm, we indicated that the combination of k-means and level set for skin cancer segmentation outperforms the commonly used classification methods such as level set and modified fuzzy c-means clustering, in two cut off (sw=0, sw=1), in terms of accuracy.

6. Conclusion

The structuring of an effective segmentation for medical expert systems to assist medical doctors is the purpose of this research. The joint k-means-level set algorithm is proposed. The algorithm employs the k-means segmentation result to initialize level set clustering technique to improve the segmentation results. The precursory along with the conclusive results are much stimulating. As the reader may note, this segmentation architecture is being proposed in such a way that assures the accuracy and effectiveness on its results. In this work, the images data and pathologists’ interaction helped a lot to achieve a good system performance. The experiments have been performed on forty images to evaluate the efficiency of proposed algorithm. Four metrics of Border error, Similarity, Hammoude distance and Rms error were used for this purpose. The promoted results show the successful performance of our proposed method when compared to traditional level set segmentation method and modified fuzzy c-means clustering, in two cut off (sw=0, cut between the small and middle class, sw=1, cut between the middle and large class). To approve our improved results, ANOVA test has been applied for assurance. Additionally, the proposed algorithm can easily be retargeted to apply in other domains of interest.

References


