International Conference on Computer Science and Computational Intelligence (ICCSCI 2015)

Implementing DEWA Framework for Early Diagnosis of Melanoma

Setiawan Hadi\textsuperscript{a}, Bernard Y Tumbelaka\textsuperscript{b}, Budi Irawan\textsuperscript{c}, Rudi Rosadi\textsuperscript{d}

\textsuperscript{a}Mathematics Department Universitas Padjadjaran, Jalan Raya Bandung Sumedang KM 21 Jatinangor, Sumedang 45363, Indonesia
\textsuperscript{b}Physics Department Universitas Padjadjaran, Jalan Raya Bandung Sumedang KM 21 Jatinangor, Sumedang 45363, Indonesia
\textsuperscript{c}Biology Department Universitas Padjadjaran, Jalan Raya Bandung Sumedang KM 21 Jatinangor, Sumedang 45363, Indonesia
\textsuperscript{d}Informatics Department Universitas Padjadjaran, Jalan Raya Bandung Sumedang KM 21 Jatinangor, Sumedang 45363, Indonesia

Abstract

This paper proposes a simple yet effective and integrated computer vision algorithm used for detecting and diagnosing the earlier stage of melanoma. The framework is built based on three steps of integrated multi aspect approach: segmentation, filtering and localization steps. In the first step, user can select several color spaces and apply leaning and non-learning methods to segment the object. In the filtering step, morphological filter has been applied for image noise removal. In the localization step, connected component labelling and K-means technique are used for objects classification. Type of cancer malignancy is determined based on a score calculated from ABCD characteristics. Experiment has been conducted successfully using skin cancer images taken from internet. This result proved that the developed framework can be used for supporting the early diagnosing of cancer. In general, this research can contribute to the computer science knowledge especially in field of computer vision.

Keywords: melanoma diagnostic; image processing; lightweight algorithm; computer vision

1. Introduction

Melanoma is one of type of skin cancer which is known as the deadliest one. It is reported that the number of people infected are increasing year by year and are geographically distributed world wide\textsuperscript{1}. If the malignant of melanoma can be detected in its early stage, there will be possibilities to reduce the dangerous effect of the disease which can decrease the number of deaths in a society.

However, there are some problems related with early stage of skin cancer diagnosis\textsuperscript{2}. The diagnosis process mainly depends on the experience, knowledge, and skill of a physician during observation. It is reported that by using simple visual analysis of a cutaneous lesion, there is 1 of 3 incorrect diagnosis. To obtain an accurate diagnosis, other exams are necessary, for example biopsies, which are expensive and painful. Another matter is the minimum periodic

\[\text{Corresponding author. Tel.: +62-22-7794696-0000; fax: +62-22-7794545.}
\]
\[E\text{-mail address: setiawan.hadi@unpad.ac.id}\]
visitation of the sickness to dermatologists. This usually happened in people of under developed countries due to financial and availability issues.

Nowadays, a lot of researchers have been working in computer vision-based diagnosis systems by means of hardware and software development by using various techniques. This work is targeting mainly at the initial diagnosing of skin cancer, and for more specific is the finding the malignant melanoma tumor. There are limited publications about lightweight algorithm for skin cancer detection that can be implemented in a portable device such as smart phone. One advantage of this implementation is, the application can be used by people to check their skin as initial diagnosis without supporting of formal trained and skillful people.

In this paper, we introduced a framework that consist of optimized but lightweight computer vision algorithm, implemented for detecting and diagnosing the initial stage of melanoma. The framework building block is based on three steps of integrated multi aspect approach: segmentation, filtering and localization steps, respectively. In the first step, we utilized both learning and non-learning methods. These methods are applied in several popular color spaces. In the filtering step, objects removal algorithm based on morphological image processing has been applied for image cleaning. In the localization step, connected component labeling and K-means are combined for grouping and localizing objects.

The framework, named DEWA (Digital Evaluation of World Around), has been successfully implemented in several application such as face detection and recognition, vision-based Tsunami early warning system, breast and prostate cancer detection and classification, and deployed in portable smart device system. It is integration of several basic primitive yet simple and powerful image processing and computer vision algorithms. It also utilized efficient storage and fast access.

2. Methods

2.1. DEWA Framework

Block diagram of the framework is presented in Figure 1. Generally, there are three main steps included in the framework: (i) image acquisition, (ii) image preparation, and (iii) image refinement and post processing steps. Image acquisition is performed using common image capturing device such as camera in the smart phone. The next step is image preprocessing which basically is image segmentation and image filtering. Image segmentation is performed, first by transforming the captured image into three color transformation, RGB, HSL and YCbCr, and second performing image thresholding based on Gaussian method. The last step is to refine the previous result and perform post processing task to localize the objects. Refinement process is performed by applying K-Means and Connected Component Labeling methods. Post processing task is measuring distance between detected object and ground truth information. This is performed using Ground Truth Bounding Box (GTBB) and Detected Bounding Box (DBB) technique. Finally is the selection process based on the results on each step.

2.1.1. Image Segmentation

This step is focused on separation of object from background. Due to the input image is RGB color image we have to convert it into color-based representation model. In this case we converted the image into HSI, YCbCr, and normalized RGB color space respectively. This conversion process step is needed to get accurate color information of the object. This chrominance space make the process only focus not on the luminance but on the hue of the image. There are many explanation can be found in the literature related to this conversion process including these three color representations.

The result obtained from previous process is color image which has Gaussian distribution following this formula:

\[ g(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp \left( -\frac{(x - \mu)^2}{2\sigma^2} \right) \]  (1)

where \( \mu \) is mean, and \( \sigma \) is variance. The next step is the thresholding process based on two methods, empirical method and Chebyshev theorem.

Image is processed by using formulation:

\[ \sigma_w^2(t) = \omega_1(t)\sigma_1^2(t) + \omega_2(t)\sigma_2^2(t) \]  (2)
\( \omega_1 \) and \( \omega_2 \) are two class probability. These classes are separated by \( t \). The \( \sigma^2_w \) is the variance of the classes. The minimization of the variance is same with the maximization of the class, as presented below:

\[
\sigma^2_b(t) = \sigma^2 - \sigma^2_w(t) = \omega_1(t)\omega_2(t)(\mu_1(t) - \mu_2(t))^2
\]

The algorithm is given as follows:

1. Compute histogram and probability of each intensity level
2. Put initial value on \( \omega_1(0) \) and \( \mu_1(0) \)
3. Do these steps for all threshold value \( t = 1 \ldots maxints \)
   (a) Update the value of \( \omega_i \) and \( \mu_i \)
   (b) Compute the value of \( \sigma^2_b(t) \)
4. The maximum value of \( \sigma^2_b(t) \) is the optimal value of thresholding.

2.1.2. Image Filtering

Image filtering is another essential step which is built based on the distribution map of ground truth skin images. Statistical calculation are implemented and stored for using in the next detection step. Complete description about generating the map has been reported in the literature. Two filtering algorithms are applied in this case, those are spatial filtering and morphological filtering. Spatial filtering is used for simple noise removal, the second technique
is for condensing and washing detected image from orphaned hole errors respectively. This is applied by using combination of erosion and dilation filters.

The next step is to remove non-cancerous objects in image. During first step of detection, it might be a possibilities to obtain group of pixels which can be considered as not cancerous objects, such as hair and other moles. due to pixel color similarities. Fortunately, the uniqueness of melanoma property allows us to remove the non-melanoma objects based on these characteristics: minimum size of detected blob, maximum size of detected blob, blob eccentricity, and blob geometric property.

In the filtering step, objects that are considered as not part of region of interest (ROI) are excluded. Filtering process is applied as follows:

If \( f[x] \) is a binary image with binary value 1 for the object and the binary value 0 for the background, then the transformation of the image of a sub window \( W = \{ y_1, y_2, y_3, \ldots, y_n \} \) of the image can be described as follows:

\[
\psi_b f[x] = b(f[x - y_1], \ldots, f[x - y_n])
\]

where \( b(y_1, \ldots, y_n) \) is \( n \) variables boolean function. Mapping \( f \mapsto \psi_b(f) \) is known as a boolean filter. By choosing various values of \( b \), we will obtain the desired transformation of the image. For example, if \( b \) is the AND operator, the result will be shrinkage of the image. Oppositely, if \( b \) is the OR, the result will be expansion of the image.

### 2.1.3. Object Clustering

Most of clustering techniques are applied in data exploration applications. Literature 4 showed that clustering techniques can be grouped into four, those are hierarchical clustering, partition clustering, density clustering, and grid clustering.

K-means has been known as one of favorite density-based clustering technique. It is implemented widely in a lot of data exploration research. However, original K-means is limited by its characteristics such as cluster number initialization. This limitation will become a constraint when automatic object grouping has to be implemented.

K-means is applied data classification based on their attributes or features into \( K \) number of group (\( K \) is positive integer number). The classification is performed by minimization the distances of data and the related centroid on the cluster. The algorithm steps of K-means algorithm can be described below.

Given this equation:

\[
M = f(x) = \sum_{i=1}^{K} \sum_{j \in C_i} \delta(x_j, \mu_i)^2
\]

K-means will minimize this equation, where \( \delta(x_j, \mu_i) \) is the distance of \( x_j \) and its nearest center \( \mu_i \). The easiest distance, Euclidean formula, is applied for calculating distance of two points, which can be described as follows:

\[
d_{ij} = \sqrt{\sum_{k=1}^{n} (x_{ik} - x_{jk})^2}
\]

### 2.1.4. Improving Clustering Algorithm

As mentioned before, original K-means required number of cluster definition before continuing the process which will be a limitation for automatic object detection. A technique to automatically determine the number of cluster has been proposed. This technique is applied according to the optimization of K’s which are selected in order to discover the best minimum value by applying validity function \( V \), described as follows:

\[
V = \frac{J'}{N \cdot \delta_{\min}^2} + \sum_{i=1}^{k'} \frac{1}{1 + |C_i|}
\]

where \( k' \) is good cluster with enough data, \( J' \) can be referred to the previous equation, but only for selected clusters, \( \delta_{\min} = \min_{\mu_i, \mu_j} \delta(\mu_i, \mu_j) \) is minimum distance among centroids, and \( |C_i| \) is cluster cardinality. The first part of the equation is the target function, whereas the other part is the small cluster penalty factor. The result of target function is the clusters that are optimally separated which will give better solutions.
2.1.5. Connected Component Labeling Algorithm

Connected component labeling algorithm is applied to label the data with same connectivity. Connectivity is relation between two or more pixels intensity values and pixel spatial adjacency in images.

Adjacency criteria is defined based on neighborhood properties. There are two types of pixel neighborhood, 4-neighborhood and 8-neighborhood. Criteria of 4-neighborhood connectivity can be formed as follows:

\[ N_4(p) = \{(x + 1, y), (x - 1, y), (x, y + 1), (x, y - 1)\} \] (8)

Criteria of 8-neighborhood can be defined as

\[ N_8(p) = N_4(p) \cup \{(x + 1, y + 1), (x + 1, y - 1), (x - 1, y + 1), (x - 1, y - 1)\} \] (9)

Generally, connectivity can be described as:

if we have \( p \) and \( q \) values in a set, then connectivity is said as 4-connected and 8-connected if \( q \) is coming from set \( N_4(p) \) and \( N_8(p) \), respectively.

The global algorithm for labeling, known as two-pass algorithm, can be described as follows:

'Pass One

for all pixels in region calculate

if pixel is not background
assign neighbors as connected elements
if neighbors are empty
linked of next label is assigned from set of next label
assign label of pixel into next label
increment next label counter
else
find the smallest label

'Pass Two

for all pixels in region
if pixel is not background
assign label of pixel by finding label on the row and column of the pixel

2.2. ABCD Rule of Dermoscopy

ABCD rule of dermoscopy was introduced by New York University in 1985. The simplicity of the method made this method become the most common technique for skin cancer diagnosis based on dermatoscopic images. The parameters that can be used for determination of a malignant cancer can be described as follows:

(A) Asymmetry - Cancerous lesions are examined for symmetry characteristic. If the lesion is symmetric (0 value) then it can be determined as non-cancerous (benign). Meanwhile for cancerous cases (malignant), 1 value or 2 value (for two orthogonal axes) are considered. In this paper, we utilized asymmetry index (AI) to get the degree of asymmetric. The formula for calculating AI can be described as

\[ AI = \frac{1}{2} \sum_{k=1}^{2} \frac{\Delta A_k}{A_L} \]

where \( k \) is major and minor axis, \( \Delta A_k \) is non overlapping area of lesion. The algorithm is as simple as follows:

find the center of lesion
divide the image into two region, vertically or horizontally
for each region,
calculate the ratio between the lesion and the region size
if the ratio between two regions conforms a threshold value then
the lesion is symmetric
otherwise
the lesion is asymmetric.
(B) **Border irregularity**  Edges of cancerous lesions are in irregular and unclear forms, such as blurred, ragged, or notched. Its value ranges are determined as 0 to 8. Determining border irregularity can be performed using several methods such as fractal dimension, compactness index, pigmentation transition and edge abruptness. In our paper we used fractal dimension for determining border irregularity. In simple term, a fractal dimension can be described as the space-filling capacity measurement of a pattern that shows differentiation between fractal scales and its space. Fractal dimension can be calculated by a method called box-counting. To find the fractal dimension of an image, the Hausdorff dimension calculation method is simpler and effective one. Detail explanation of fractal dimension can be read from⁹.

(C) **Color** Skin pigmentation of cancerous lesion is not uniform. Six known RGB colors must be detected from white to black. Those are white, red, brown (light and dark), blue, and black, which are given value, ranges from 1 to 6. Color distribution of skin cancer image is done by computing RGB values of each pixel fill it into each group. The distribution value can be grouped into 6 values based on colors that have to be detected.

(D) **Diameter** Measure of cancerous lesions are determined larger than 6mm wide. Combination structures with at least 5 patterns are considered as special types of lesions. Growing of a mole should be observed and becoming our concern. Its value ranges 1 to 5. In implementation to calculate the diameter of the lesion, zeroth order moment is applied to get the size of the object and then calibrated based on the resolution of the display system.

Table 1 summarize the description of the ABCD characteristics, and Table 2 described how the value TDS is computer based on ABCD values.

| Table 1. Summarizing The ABCD Values and Weight Factors |
|---------------------------------|-----------------|----------------|
| Criterion         | Score | Weight |
| Asymmetry         | 0-2   | ×1.3   |
| Border            | 0-8   | ×0.1   |
| Color             | 1-6   | ×0.5   |
| Diameter          | 1-5   | ×0.5   |

| Table 2. Criteria of Skin Cancer based on TDS Index |
|-----------------|-----------------|
| TDS Index       | Decision        |
| less than 4.75  | benign skin lesion |
| greater than 4.75 and not more than 5.45 | suspicious cancerous skin lesion |
| greater than 5.45 | malignant melanoma skin lesion |

3. **Implementation**

The integration of DEWA framework and ABCD Rule for detecting skin cancer are performed by developing applications running in multi platforms. The general steps that are implemented in the applications can be described as follows:

1. Load skin cancer image
2. Apply DEWA algorithm to obtain noise-free image
3. Image analysis based on ABCD Rule
   
   (a) Determine the asymmetric value (A)
   (b) Calculate the border irregularity (B)
   (c) Get the color distribution (C) of skin image
(d) Compute the diameter (D)
(e) Apply TDS Index formula to obtain TDS Score

4. Based on TDS score, determine the category of cancer in the skin image

Figure 2 shows an application named SISIK that has been developed and implemented in a smart phone running Microsoft Windows Phone operating system.

3.1. Data for Experiment

Data that is used for experiment are obtained for multiple sources found in the internet. The list of skin cancer image data, resource, and number of images, are shown in Table 3. Not all data are used for experiment. In this case, 29 data are selected, which will be used for test and experiment. Sample data are presented in Figure 3

| Name                                | Number of Data (t) |
|-------------------------------------|--------------------|
| eMedicineHealth                     | 56                 |
| Skin Cancer Pictures by Lisa Fayed  | 13                 |
| SdermNet NZ                         | 50                 |

4. Result

Experimental result is presented in the Table 4. Decision about cancer type is taken based on TDS calculated from A, B, C, D value. From the table we can infer that among 29 controlled-experimental images which are tested, 24
Table 4. Experimental Result.

| ID-DATA | Ground Truth | Asymmetric | Border | Color | Diameter | TDS     | RESULT |
|---------|--------------|------------|--------|-------|----------|---------|--------|
| SC01-Assymetrical | Benign | 2 | 3 | 1 | 3 | 4.4 | Benign |
| SC02-Melanoma | Malignant | 2 | 1 | 4 | 4 | 6.7 | Malignant |
| SC03-Unev. Dist. | Benign | 1 | 5 | 4 | 2 | 4.8 | Suspicious |
| SC04-LargeSize | Benign | 2 | 1 | 3 | 1 | 4.7 | Benign |
| SC05-Melanoma | Malignant | 2 | 8 | 3 | 4 | 6.9 | Malignant |
| SC06-Melanoma | Malignant | 2 | 1 | 6 | 2 | 6.7 | Malignant |
| SC07-Melanoma | Malignant | 2 | 7 | 2 | 4 | 6.3 | Malignant |
| SC08-Melanoma | Malignant | 2 | 4 | 4 | 2 | 6.0 | Malignant |
| SC09-Melanoma | Malignant | 2 | 2 | 6 | 4 | 7.8 | Malignant |
| SC10-Melanoma | Malignant | 2 | 4 | 5 | 2 | 6.5 | Malignant |
| SC11-Melanoma | Malignant | 1 | 6 | 2 | 1 | 3.4 | Benign |
| SC12-Melanoma | Malignant | 2 | 4 | 3 | 2 | 5.5 | Malignant |
| SC13-Melanoma | Malignant | 2 | 4 | 3 | 5 | 7.0 | Malignant |
| SC14-Melanoma | Malignant | 2 | 1 | 2 | 1 | 4.2 | Benign |
| SC15-Melanoma | Malignant | 0 | 4 | 6 | 5 | 5.9 | Malignant |
| SC16-Melanoma | Malignant | 2 | 1 | 4 | 2 | 5.7 | Malignant |
| SC17-Melanoma | Malignant | 2 | 5 | 6 | 2 | 7.1 | Malignant |
| SC18-Melanoma | Malignant | 1 | 8 | 5 | 2 | 5.6 | Malignant |
| SC19-Basal | Benign | 0 | 7 | 6 | 5 | 6.2 | Malignant |
| SC20-Melanoma | Malignant | 1 | 7 | 4 | 3 | 5.5 | Malignant |
| SC21-Squamous | Benign | 0 | 4 | 3 | 5 | 4.4 | Benign |
| SC22-Squamous | Benign | 1 | 1 | 3 | 2 | 3.9 | Benign |
| SC23-Squamous | Benign | 1 | 6 | 2 | 4 | 4.9 | Suspicious |
| SC24-Basal | Benign | 1 | 2 | 4 | 1 | 4.0 | Benign |
| SC25-Squamous | Benign | 0 | 2 | 2 | 5 | 3.7 | Benign |
| SC26-Melanoma | Malignant | 2 | 4 | 4 | 2 | 6.0 | Malignant |
| SC27-Squamous | Benign | 2 | 5 | 1 | 1 | 4.1 | Benign |
| SC28-Squamous | Benign | 0 | 1 | 2 | 4 | 3.1 | Benign |
| SC29-Melanoma | Malignant | 2 | 1 | 6 | 5 | 8.2 | Malignant |

or 83% images are correctly detected compared with Ground Truth data. 5 images or 17% are incorrectly diagnosed. The TDS score will determine the type of cancer based on Table 2. However, the Ground Truth value is not match to the result, so it is considered as incorrect detection.

Image SC03 has high values in Border irregularity and Color parameters, so the TDS score is 4.8 (which is greater a little greater than 4.75) and considered as Suspicious. Image SC11 is small lesion but the Ground Truth is Melanoma. Due the weight of Diameter is 0.5 so the TDS score became small and considered by system as non melanoma. It is similar with SC14 with additional problem in Border irregularity. Image SC14 shape is not too irregular however the ground truth is melanoma, so it is incorrectly detected.

In opposite scheme, image SC19 is detected as benign (same as ground truth) at asymmetric aspect, but the other 3 aspects gave high result, so SC19 considered as melanoma. Meanwhile the image SC23 has high value in two aspects, those are border irregularity and Diameter.

5. Conclusion and Future Work

Early diagnosis of melanoma is important step that can be useful to decrease the danger of skin cancer. By using public device such as smart phone, the diagnosis mechanism can be done by common people. The early and general information about skin cancer can be discovered. It is a fact that the process is not as accurate as done by medical expert or doctor, but in certain condition and situation, this technique minimal can assist the person to take further action better.

This research implements DEWA as initial processing method and ABCD for decision of cancer type. DEWA as framework, utilized three multi aspects and can be considered as light framework because it has been deployed on small portable device successfully. The implementation algorithm to get ABCD characteristics can be extended not
only to Asymmetric Index, Fractal Dimension, RGB color, and Moment techniques, but also developed using other
techniques such as neural network, Hidden Markov, and many others.

The integrated methods that have been applied, even though it has been tested, and give promising result (83%),
need to be verified using more data coming from many areas. In addition, some problems still found in the methods,
such as inaccuracy in detection especially in color aspect need to be corrected. The speed of detection need to be im-
proved especially by optimizing the algorithm. The future technology including smart phone technology will improve
the performance of smart phone in the area of computation, which will be faster and more accurate. Consequently,
the detection process of skin cancer will be easier and can be done in real time fashion.

References

1. I. Maglogiannis, "Overview of Advanced Computer Vision Systems for Skin Lesions Characterization", IEEE Transact. on Information Tech-
nology in Biomedicine, vol. 13, no. 5, September 2009
2. A. Parolin, E. Herzer, and Cludio Rosito Jung. "Semi-Automated Diagnosis of Melanoma Through the Analysis of Dermatological Images", SIBGRAPI, page 71-78. IEEE Computer Society, 2010
3. S. Hadi, A.S. Ahmad, I.S. Suwardi, and F. Wazdi. “Small Device Implementation of DeWa Algorithm”, Int. Conf. on EE and Informatics, 2007.
4. A.K. Jain, M.N. Murty and P.J. Flynn. "Data Clustering: A Review", ACM Computing Surveys, vol. 31 no. 3, 1999
5. Hadi S., Nursantika, D., Purwanti, I. "Integrating Computer Vision Technique to Support Tsunami Early Warning System", Int. Conf. on System
and Informatics, 2012, Yantai China.
6. Aryadi G., Hadi S., Hidayat A., Nursantika D. "Klasifikasi Keganasan Kanker Prostat Berdasarkan Analisis Citra Biomedis", Seminar Nasional
Matematika Universitas Indonesia-Universitas Padjadjaran, 2014.
7. Sholahuddin M.R., Hadi S. “Sistem Deteksi Dan Klasifikasi Kanker Payudara Berbasis Pengolahan Citra Digital Pada Citra Mammogram”,
Undergraduate Thesis, 2013.
8. Stolz, W., Riemann, A., Cognetta, A.B., Pillet, L., Abmayr. W., Holzel, D., "ABCD rule of Dermatoscopy: a new practical method for early
recognition of malignant melanoma", European Journal of Dermatology, no.4, 1994, pp.521-527.
9. Vincent T. Y. Ng ; Tim K. Lee, "Measuring border irregularities of skin lesions using fractal dimensions", Proc. SPIE 2898, 1996.