International Journal of Multimedia and Ubiquitous Engineering
Vol.10, No.11 (2015), pp.303-314
http://dx.doi.org/10.14257/ijmue.2015.10.11.29

The Contribution of Feature Selection and Morphological Operation For On-Line Business System's Image Classification

Mokhairi Makhtar[1], Nur Shazwani Kamarudin[2], Syed Abdullah Fadzli[3], Mumtazimah Mohamad[4], Fatma Susilawati Mohamad[5] and Mohd Fadzil Abdul Kadir[6]

[1,3,4,5,6]*Senior Lecturer, Faculty of Informatics and Computing, University Sultan Zainal Abidin, Tembila Campus, Terengganu, Malaysia*
[2]*Research Scholar, Faculty of Informatics and Computing, University Sultan Zainal Abidin, Tembila Campus, Terengganu, Malaysia*
mokhairi@unisza.edu.my

Abstract

Automatic image annotation is one of crucial and attractive field of image retrieval. Classification process is part of the important phase in automatic image annotation (AIA). With the explosive growth of methods in this research area, this paper proposes 5 processing steps before image annotation using Amazon dataset, i.e., image segmentation, object identification, feature extraction, feature selection and image features classification. A lot of research has been done in creating numbers of different approaches and algorithm for image segmentation. Otsu is one of the most well known method in image segmentation region based. The proposed model aims to provide the highest accuracy after undergo those processing steps. This paper conducted several experiments for image classification starting from image segmentation in order to demonstrate usefulness and competiveness among different type of classifiers. It also target to study the effect of morphological operation and feature selection to the accuracy. For the classification experiment, it was tested using four types of classifiers: BayesNet, NaiveBayesUpdateable, RandomTree and IBk.

Keywords: *Image Classification, Feature Selection, Morphological Operation*

1. Introduction

The processes of classifying images involve data collection, image segmentation, image cleaning or filtering, feature extraction and analysis of the results. To produce a result with high accuracy and time consuming, best effort needed from the beginning of process until the end. The kind of dataset itself will influence results starting from segmentation until classification. Classification is the final step proposes in this model. Before classification itself, there are 4 more steps included in the model. Each of them have crucial role in term of providing best classification results.

This paper proposes a global data from Amazon.com towards image classification. The algorithms used in these model are commonly divided into five tasks:

i. image segmentation
ii. morphological operation
iii. feature extraction
iv. feature selection
v. image classification

Image segmentation is partitionaing the images in order identify the objects for further processes. The morphological operation smooth images after segmentation process. Followed with feature extraction that transform rich content of images into various

ISSN: 1975-0080 IJMUE

contents of features [1]. Classification process will classify the features based on categories.

The remainder of this paper is organised as follows. Section 2 describes current work on image analysis starting from image segmentation until classification. In Section 3, the image classification model are provided. Its experimental results are presented in Section 4. Section 5 summarizes and concludes the work.

2. Literature Review

An image segmentation algorithm typically requires many processing steps and a designer must decide which steps to use [2]. Stated in [3], the most famous and reliable segmentation techniques are Edge Detection, Threshold, Histogram, Region based methods, and Watershed Transformation. To narrow down the image segmentation process, here the list of models that being used until now [4], a) Object Background/Threshold Model, b) Neural Model, c) Markov Random Field Model, d) Fuzzy Model, e) Fractal Model, f) Multi-resolution and g) Transformation Model namely Watershed Model and Wavelet Model.

Segmentation is considered the first step in image analysis. This paper uses thresholding techniques and to be exact Otsu algorithm towards Amazon dataset.

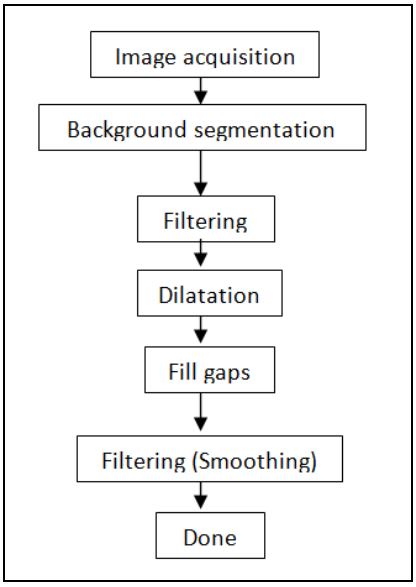

Figure 1. Examples of a Typical Image Segmentation Algorithm [2]

Object background/threshold model itself have many method underlies in this category. One of them is Otsu method. It is originally introduced by Nobuyuki Otsu in [5]. The method working by selecting a threshold automatically from a gray level histogram. Even though Otsu criterion [5] is simple and ease, [6] in their evaluation paper shows that Otsu still can give a better results depends on the nature of the image.

The author of [7] stated that watershed method is the process of labelling different object in an image, while [8] said that it is the process of dividing lines. Watershed method is an effective use for transforming image as a 3D terrain surface [9]. It may also produce an over-segmented image if the data used not suitable with the method.

Canny edge detector currently known as the best detector by ensuring good noise immunity and provide minimum error by detecting true edge point [6]. From experiment done by [6], they compare the segmentation results using Otsu and Canny. For medical image, Canny gives more accurate results by giving almost fully segmented from the edges of the original images.

For the purpose of image analysis, mathematical morphology process is needed in order to clean and smooth the image whether before or after segmentation process. According to [10], there are 4 morphological operators, *i.e.*, dilation, erosion, opening, closing. To make the morphological process easier, image must be converted into binary image.

Dilation alter the object boundaries of an image by adding pixels, while erosion removes pixels on object boundaries. The opening and closing morphology operators are combination of both dilation and erosion. Morphological opening shift small objects from an image while preserving the shape and size of bigger objects in the image. It is started by erosion then followed by dilation. For morphological closing, dilation process take part on the first place then followed by erosion. In [11], they used opening-closing operation to smooth the image with the basic of fuzzy mathematical morphology.

Various content of image features can be extracted through feature extraction task. The extracted features will then be used in further task such as feature selection and classification [1]. Shape-based feature extraction were divided into two categories which is region-based and contour-based. The shape descriptor from contour-based feature extraction calculate features from object contour such as circularity, aspect ratio, discontinuity, complexity, sharpness, angleness, *etc.* According to [13], the author develop a Matlab procedure using *regionprops* function from the Image Processing Toolbox in order to compute the image features based on object contour.

One of the most important steps before detection and classification is feature selection. In order to perform a variety of tasks such as classification and annotation, good features with discriminative, robust, easy to compute and efficient algorithm are needed [14]. In [15], the experiment performed using predictive toxicology data and accuracy was increase after feature selection rather than using original feature set.

In order to determine a suitable classification system for image classification, major steps such as feature extraction, select suitable classification approach, post-classification processing and accuracy assessment needed [16]. The experiment in [13] was conducted image classification based on shape features. Three type of classifier chosen which id decision trees, k-nearest neighbour and support vector machine. The SVM based classifier gives the best accuracy, 86% compare to kNN with 80% and the lowest DT, 69%.

3. Image Classification Model

Figure 2. Propose Classification Model

Given the pre-processing steps proposed that involves five stages:

i. Image segmentation distinguish objects from the background.
ii. Morphological operation is to smooth the image.
iii. Feature extraction process extract valuable features value from images using Matlab built in function, *regionprops*.
iv. Feature selection done the selection process in order to provide the best accuracy results.
v. Image classification is the stage where all training images classified using 8 type of classifiers. Feature selection are then used to compare the accuracy of each classifiers.

3.1.1. Image Segmentation

Segmentation is a fundamental step in image description or classification [17]. The goal of image segmentation is to partition the image plane into meaningful areas [18]. The main objectives of this paper is to separate the background and foreground of each images. Because of that, thresholding method were selected. The basic principle of thresholding is to select an optimal gray-level threshold value for separating objects of interest in an image from the background based on their gray-level distribution [19].

Thresholding is one of well-known technique. Using a large dataset from Amazon.com, image data were segmented using thresholding based method. The most popular methods is Otsu method [20]. Otsu method is one of global thresholding techniques.

A global thresholding technique thresholds the entire image with a single threshold value [20]. This method is based on discriminant analysis.

As stated in [5-6, 21] that threshold operation can partition the image into two classes A_1 and A_2 at gray Q such that A_1 = [22 ... , Q] and A_2 = {Q + 1, Q + 2, ... , k-1}, where k is the total number of the gray levels of the image. Let the number of pixels at **i** gray level be n_i , and $N = \sum_{i=0}^{k-1} n_i$ be the total number of pixels in a given image. The probability of occurence of gray level **i** is defined as $p_i = \frac{n_i}{N}$, $p_i \geq 0$, $\sum_{i=0}^{k-1} p_i = 1$. A_1 and A_2 are normally corresponding to the object of interested and the background, the probabilities of the two classes are $P_{A1} = \sum_{i=0}^{Q} p_i$ and $P_{A2} = \sum p_i = 1 - P_{A1}$.

The means of the classes A_1 and A_2 can be computed as :

$$\mu_{A1} = \sum_{i=0}^{Q} \frac{i^* p_i}{P_{A1}} \quad (1)$$

$$\mu_{A1} = \sum_{i=Q+1}^{k-1} \frac{i^* p_i}{P_{A2}} \quad (2)$$

So we can get the equivalent formula :

$$\sigma^2(Q) = P_{A1} P_{A2} (\mu_{A1} - \mu_{A2})^2 \quad (3)$$

The optimal threshold Q* can be obtained by maximizing the between-class variance.

$$Q^* = Arg \max_{0<Q<k-1} \sigma^2(Q) \quad (4)$$

The RGB colour images are first segmented using Otsu's algorithm to separate between background and foreground which produced binary images as demonstrated in (Figure 3).

3.1.2. Morphological Operation

In order to make the comparison for the classification result whether morphological operation(MO) needed or not, the experiment done by classifying both data that with MO or without MO. From the literature review, most of them stated that, for the sake of better visual effect, images must undergo the morphological operation. For this experiment, the techniques that applied to the images is morphological close filter. The morphological close operation is a dilation followed by an erosion, using the same structuring element for both operations. This filtering process produced much clean result after segmentation with the smooth boundaries, reduce small inward bumps, join narrow breaks and fills small holes caused by noise [13].

As for this paper, after image segmentation process, images then feed into morphological close operation and produces results illustrated in (Figure 4).

Figure 3. The Segmentation Result of Otsu

(a)(b)(c) Original image
(d)(e)(f) Segmented image

3.1.3. Feature Extraction

The process of feature extraction started after segmentation process. It will compute the image features of all segmented images. Features represent the object presented in each image. So, chosen features to be extract must be discriminative and sufficient.

The computation of features from an input image were done by developing a Matlab procedure using *regionprops* function from Image Processing Toolbox. The procedure of obtaining image features from an input image begin with the computation of properties of the image, like area, eccentrcity, extent, solidity,filled area, *etc.* The features were computed using build in formulas in Matlab [12].

Figure 4. The Morphological Operation Result

(a)(b)(c) Image after segmentation
(d)(e)(f) Image after morphological operation

F1 : Area; Compute the actual number of pixels in images.

F2 : Major Axis Length ; Scalar specifying the length (in pixels) of the major axis of the ellipse that has the same normalized second central moments as the region.

F3 : Minor Axis Length ; Scalar; the length (in pixels) of the minor axis of the ellipse that has the same normalized second central moments as the region.

F4 : Eccentricity ; The eccentricity is the ratio of the distance between the foci of the ellipse and its major axis length.

F5 : Orientation ; Scalar; the angle (in degrees ranging from -90 to 90 degrees) between the *x*-axis and the major axis of the ellipse that has the same second-moments as the region.

F6 : Convex Area ; Scalar that specifies the number of pixels in 'ConvexImage'.

F7 : Filled Area ; Scalar specifying the number of on pixels in FilledImage.

F8 : Euler Number ; Scalar that specifies the number of objects in the region minus the number of holes in those objects.

F9 : EquivDiameter ; Scalar that specifies the diameter of a circle with the same area as the region. Computed as sqrt(4*Area/pi).

F10 : Solidity ; Scalar specifying the proportion of the pixels in the convex hull that are also in the region. Computed as Area/ConvexArea.

F11 : Extent ; Scalar that specifies the ratio of pixels in the region to pixels in the total bounding box. Computed as the Area divided by the area of the bounding box.

3.2.4 Feature Selection

A large set of features is a description for the particular data collection. And, feature selection is a process of selecting the combination of features [23]. Indeed, feature selection were divided into three categories, i.e. , filters, wrappers and hybrid method. This paper uses the built-in attribute selector provided in Weka. The Correlation-based Feature Selection Subset Evaluator and Principal Component algorithm were used for the feature selection tasks. While GreedyStepwise and Ranker was the searching method used for feature selection.

4. Experiments and Results

4.1. Data Collection

The total number of data uses for this experiments is 3299 images. There are 16 categories of images from the collection. For training data, 30% images for each categories are randomly selected from the collection to undergo those stages in (Figure 2). All programming starting from image segmentation until feature extraction is written in MATLAB®. (Table 1) listed numbers of images for each categories that selected for the training data.

Table 1. Training Dataset

Categories	Number of images
Handbag	97
Clutch	351
Men-short	583
Men-jacket	475
Men-buttondownshirt	93
Women blouse	681
Women dress	441
Women jeans	126
Boots	51
Heels and pump	351
Sport shoes	50
Total	3299

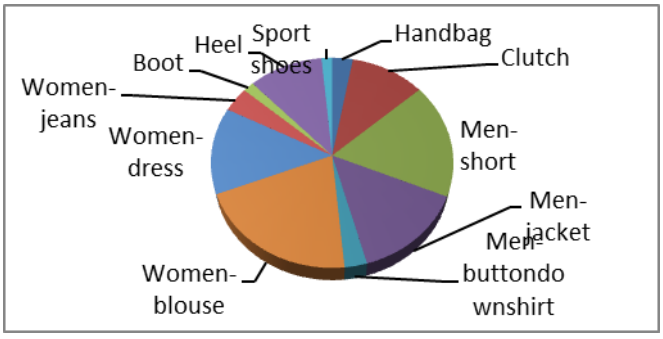

Figure 5. Categorised Images

4.2. Classification Results

The main objective of image classification is to calculate the accuracy of classified images based on the categories stated. The test were performed on the Amazon dataset which consist of 12 features for each single image.This paper uses Weka with 10-fold cross validation to run the classification experiment and those classifier chosen:

weka.classifiers.bayes.BayesNet,
weka.classifiers.bayes.NaiveBayesUpdateable,
weka.classifiers.trees.RandomTree,
weka.classifiers.lazy.IBk.

Table 2 lists the results of classification accuracy for images without feature selection. As stated in the table, accuracy did not gives much differences between data that undergo the morphological operation or not. Without feature selection, RandomTree is the outstanding classifier with result, 93.9507% with the morphological operation.

Table 2. Model Accuracies on Dataset without Feature Selection

	BayesNet	Naive Bayes Updateable	RandomTree	IBk
With Morphological Operation	78.7796	91.3175	93.9507	92.2587
Without Morphological Operation	80.3279	90.1032	93.8069	92.2283

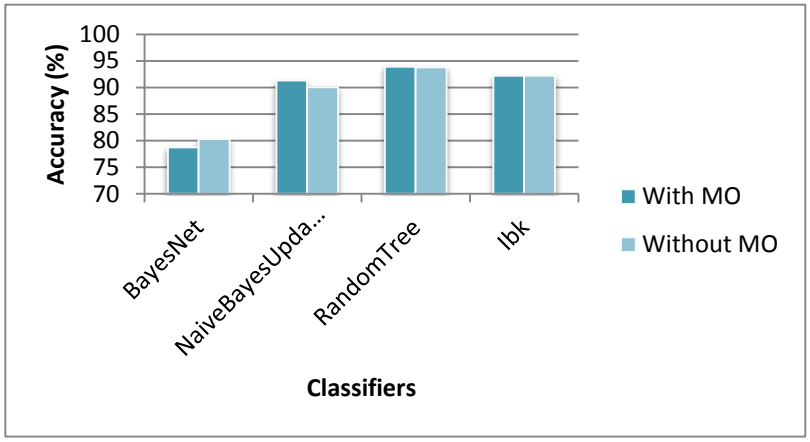

Figure 6. Accuracy Graph on Dataset without Feature Selection

Table 3. Model Accuracies on Dataset with Feature Selection: Ranker + PrincipalComponents Algorithm

	BayesNet	Naive Bayes Updateable	RandomTree	IBk
With Morphological Operation	88.7371	92.9872	97.4803	93.8676
Without Morphological Operation	91.5301	93.0176	98.1785	93.9891

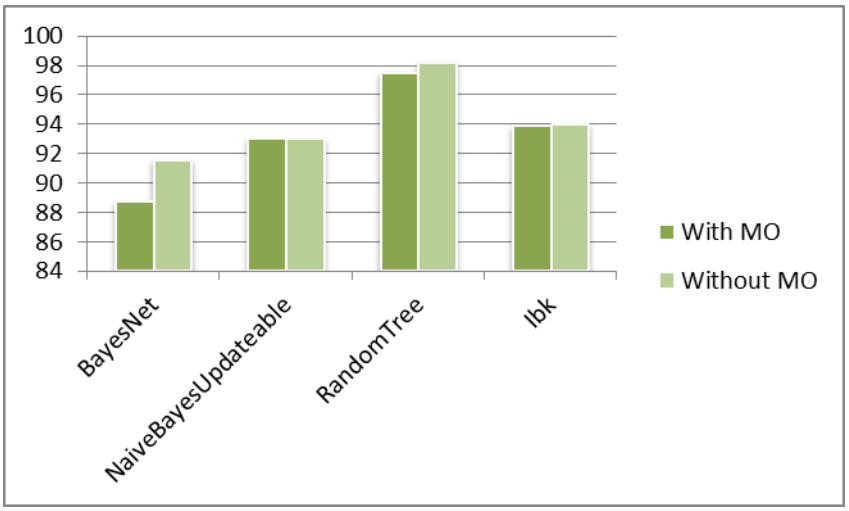

Figure 7. Accuracy Graph on Dataset with Feature Selection Based on Table 3

While Table 3 shows the result for images that undergo the feature selection step with algorithm Ranker + PrincipalComponents. RandomTree classifier gives the highest accuracy compare to others with 98.1785% before morphological operation.

Table 4. Model Accuracies on Dataset with Feature Selection: GreedyStepwise + CFSSubset Evaluator Algorithm

	BayesNet	Naive Bayes Updateable	RandomTree	IBk
With Morphological Operation	97.0856	94.8391	99.3018	95.5981
Without Morphological Operation	98.1178	94.7177	99.4839	95.2034

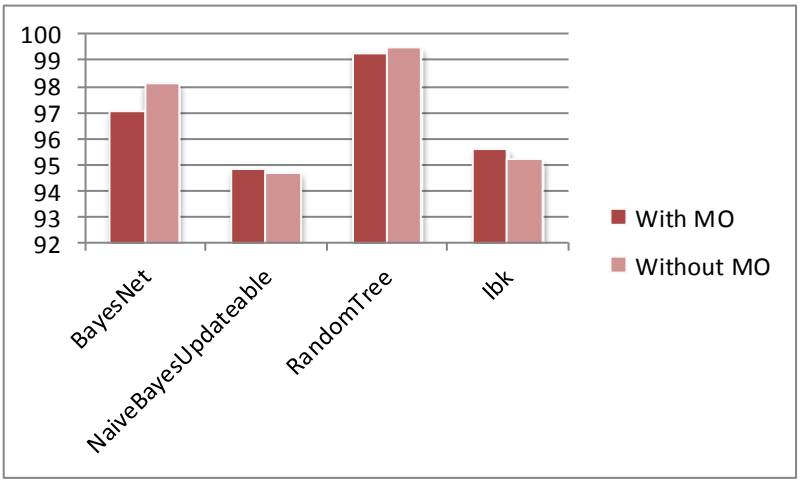

Figure 8. Accuracy Graph on Dataset with Feature Selection Based on Table 4

Furthermore, Table 4 listed accuracies after feature selection using GreedyStepwise + CFSSubsetEvaluator. It shows that RandomTress classifier provide an outstanding result, 99.4839% for data that did not undergo morphological operation.

Table 5. Detail of Classification Results for Each Image Categories Using RandomTree Classifier with Feature Selection (GreedyStepwise + CFSSubsetEvaluator) and without Morphological Operation

Categories	Total Image	Accurately Classified	Percentage of Accuracy (%)
Handbag	97	96	98.969
Clutch	351	351	100
Men-short	583	582	99.828
Men-jacket	475	474	99.789
Men-buttondownshirt	93	92	98.925
Women-blouse	681	681	100
Women-dress	441	440	99.773
Women-jeans	126	126	100
Boot	51	47	92.157
Heel	351	349	99.43
Sport shoe	50	47	94

Overall, the paper reports good results in using Amazon dataset to get highest accuracy with data that skip morphological operation with feature selection. As stated in Table 5, the classification result using RandomTree classifier with feature selection algorithm (GreedyStepwise + CFSSubsetEvaluator) and without morphological operation gives the highest accuracy among all.

5. Conclusion

This paper prove that morphological operation did not give much effect for accuracy on Amazon dataset. Feature selection do give huge effect to the accuracies. It helped in facilitating data visualization and data understanding, reducing the measurement and storage requirements, reducing training and utilization times, and defying the curse of dimensionality to improve the prediction performance. The experiment shows that percentage of accuracy increase after feature selection and Greedy + CFSSubsetEvaluator algorithms provide the best results. Feature selection illustrates that features are valuable in terms of classification process and image analysis for further usage. Although the study only focuses on Amazon dataset for the experiment, this approach is also applicable on other dataset with different kind of images.

Further work in applying different image segmentation method on the same dataset to compare accuracy of classification results.

Acknowledgment

This work is partially supported by UniSZA (Grant No. UNISZA/13/GU(029).

References

[1] Ryszard, "Image Feature Extraction Techniques and Their Applications for CBIR and Biometrics Systems", International Journal of Biology and Biomedical Engineering, vol. 1, no. 1, **(2007)**.
[2] Hedberg H., "A Survey of Various Image Segmentation Techniques".
[3] Khan M. W., "A Survey: Image Segmentation Techniques", International Journal of Future Computer and Communication, vol. 3, **(2014)**.
[4] V. Dey [a], Y. Zhang[a] and M. Zhong[b], "A Review On Image Segmentation Techniques With Remote Sensing Perspective", (ISPRS10), vol. XXXVIII(Part 7A**)**, **(2010)**.
[5] OTSU N., "A Threshold Selection Method from Gray-Level Histograms", 2EEE Transactions on Systrems, Man, and Cybernetics, vol. 9, **(1979)**.
[6] A. A. Mohammed Al-Kubati, J. A. M. S. and Murad A. A. Taher, "Evaluation of Canny and Otsu Image Segmentation", in International Conference on Emerging Trends in Computer and Electronics Engineering (ICETCEE'2012), Dubai, **(2012)**.
[7] Shahzad A., "Enhanced Watershed Image Processing Segmentation. Journal of Information & Communication Technology", vol. 2, **(2008)**.
[8] Soille L. V. A. P., "Watersheds in Digital Spaces: An Efficient Algorithm Based on Immersion Simulations", in IEEE Transactions on Pattern Analysis and Machine Intelligence, **(1991)**.
[9] Qinghua Ji R. S., "A noval method of image segmentation using watershed transformation", in International Conference on Computer Science and Network Technology, IEEE, **(2011)**.
[10] Pandey S., "Study and Implementation of Morphology For Image Segmentation", in Department of Electrical and Instrumentation Engineering, Thapar University, **(2010)**, p. 76.
[11] Yubin L.Y. L., "An Algorithm of Image Segmentation Based on Fuzzy Mathematical Morphology", in International Forum on Information Technology and Applications, **(2009)**.
[12] Available from: http://www.mathworks.com/help/images/ref/bwlabel.html.
[13] J. F. Nunes, P. M. M. and Joao Manuel R. S. Travares, "Shape Based Image Retrieval and Classification", **(2010)**.
[14] O. Tuzel[1], F. Porikli[3] and Peter Meer[1,2], "Region Covariance: A Fast Descriptor for Detection and Classification", in European Conference on Computer Vision (ECCV). 201 Broadway, Cambridge, Massachusetts, **(2006)**.
[15] M. Makhtar, D.C.N. and M. Ridley, "Predictive Model Representation and Comparison: Towards Data and Predictive Models Governance", **(2010)**.
[16] Weng D. L. Q., "A survey of image classification methods and techniques for improving classification performance", International Journal of Remote Sensing, **(2007)**.
[17] L. Cinque[a,] G. Foresti[b] and L. Lombardi[c], "A clustering fuzzy approach for image segmentation", Pattern Recognition, **(2003)**.
[18] M. Singh[1,] A. M., "A Survey Paper on Various Visual Image Segmentation Techniques", International Journal of Computer Science and Management Research, vol. 2, no.1, **(2013)**, p. 7.
[19] Miss Hetal J. Vala and P. A. B., "A Review on Otsu Image Segmentation Algorithm", International Journal of Advanced Research in Computer Engineering & Technology (IJARCET), vol. 2, no. 2, **(2013)**.
[20] P. K. Sahoo, S. S., and A. K. C. Wong, "A Survey Of Thresholding Techniques in Computer Vision, Graphics, and Image Processing", **(1988)**, pp. 233-260.
[21] M. Huang, W. Y. and D. Zhu, "An Improved Image Segmentation Algorithm Based on the Otsu Method, in ACIS International Conference on Software Engineering", Artificial Intelligence, Networking and Parallel/Distributed Computing, **(2012)**.
[22] Information Retrieval. 2 June 2013 [cited 2012; Available from: http://en.wikipedia.org/wiki/Information_retrieval.
[23] Gabbouj E.G.M., "Feature selection for content-based image retrieval", **(2008)**.

