International Journal of Engineering and Technical Research (IJETR)
ISSN: 2321-0869, Volume-2, Issue-11, November 2014

Comparative Study of various Image Segmentation Techniques

Harshil Shah, Rahul Shah

***Abstract*— Image segmentation is a considered to be an integral part of the OCR process. The positives of the recent advancements in the OCR are imputed to the better manipulation of images in terms of segmentation. The proposed paper is a review of the progress made. Segmentation methods are listed under three main techniques. The first technique, Hough Transform, takes into consideration various mathematical equations to expedite the segmentation process. The second technique uses the connected components to segment the images. Lastly it uses clustering algorithm to define several of the data points of the image.**

***Index Terms*—CCL, Expected Maximization, Hough Transform, Image Segmentation.**

I. INTRODUCTION

Image Segmentation [1],[2] is the process of partitioning a digital image into multiple segments. The objective of segmentation is to simplify and change the representation of an image into something that can be dealt better in terms of analysis. Image Segmentation is done based on certain characteristics, such as color, shape and orientation. Image segmentation can be done by using a 2-d dimensional bitmap with each pixel of the image represented by one bit of the bitmap. Formal definition of image segmentation is defined as the function such that it divides the image x into sub images x_j such that every sub image belong to a particular equivalence class defined by the relation. The methods for image segmentation are described as below.

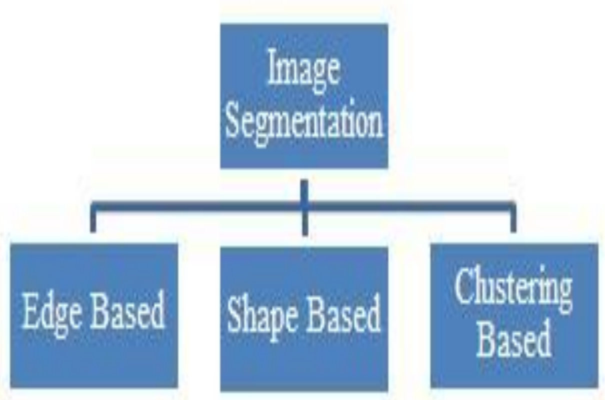

Fig. 1 Classification of various segmentation techniques

Manuscript received November 05, 2014.

Harshil Shah, BE Computer Engineering, Dwarkadas J. Sanghvi College of Engineering, Mumbai, India, +919322065106.

Rahul Shah, BE Computer Engineering, Dwarkadas J. Sanghvi College of Engineering, Mumbai, India, +919769568468.

II. EDGE BASED TECHNIQUE

This method attempts to resolve image segmentation by detecting edges or boundaries between two distinct or contrasting regions. In the paper [3] proposed by Satadal Saha, they have used an edge-based technique to improve the overall efficiency of the system. The approach is used is converting the input image using Hough transform for directional segmentation of lines and words from any type of images The above transform is used iteratively from sentences to words and from words to characters. The Hough image is generated from the binarized edge map of the image. The Hough transform uses various parameters to tune the overall transform for better results.

A. Algorithm

1. Locate all the feature points in the image space.
2. For each feature point in the image space, a set of lines are plotted in the Hough space.
3. The intersections in the Hough space are plotted into a 2-d accumulator.
4. After all the plotting, a local maxima is found in the accumulator
5. If required, plot back each maxima into the image space.

The representation of the Hough plane is done in terms of (p, β) instead of (x, y). The given representation is used so as to avoid the situation when the line is parallel to the y-axis. This would lead to slope of the line tending to infinity. Thus the equation of the line changes to -

$$x \cos \beta + y \sin \beta = p \qquad (1)$$

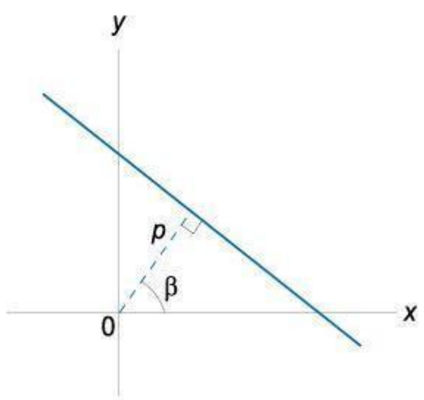

Fig.2 Alternative representation of straight line in (p, β) plane.

B. *Preprocessing and Tuning*

The various parameters used for the tuning deltaRho, deltaTheta, startTheta, endTheta, connectDistance, pixelCount. The image before getting transformed goes through a number of stages. The image are pre-processed, binarized (using Otsu algorithm [4]). The edge detection of the objects is done using various masks. The masks used are as follow:

$$\begin{bmatrix} 1 & 2 & 1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{bmatrix} \quad \begin{bmatrix} 1 & 0 & -1 \\ 2 & 0 & -2 \\ 1 & 0 & -1 \end{bmatrix}$$

$$\begin{bmatrix} 2 & 1 & 0 \\ 1 & 0 & -1 \\ 0 & -1 & -2 \end{bmatrix} \quad \begin{bmatrix} 0 & 1 & 2 \\ -1 & 0 & 1 \\ -2 & -1 & 0 \end{bmatrix}$$

All the marked white lines in the Hough transformed images are segmented through CCL algorithm. In the algorithm [6], any white pixel searches for its white neighbors.8-connected neighbors are searched and non-recursive function call is used to reduce usage of system resource and time complexity. Lastly, a bounding box is created which envelopes the words recognized.

III. SHAPE BASED TECHNIQUE

The technique which is useful in segmentation is shaped based. These techniques take into consideration the homogeneity of a particular area (forming a region). The paper proposed [5] uses discriminative learning connected component based classification. Here they train a self-tunable multilayer perceptron (MLP) classifier for distinguishing between text and non-text connected components using shape and context information as a feature vector.

A. *Shape of the connected component*

In most of the documents, the size of the non-text components is larger than that of the text components. Thus, size information plays a key role in classification. But it alone cannot suffice the need of classification and hence we also use shape of the text and non-text components which can be learned by the MLP classifier.

Hence for generating the feature vector each connected component is rescaled to a 40X40 pixel window. It is only downscaling. If the length or height is greater than 40 then it is downscaled to 40 else if it is less than 40 it is fit to the center of the window. The advantage of doing so is to distinguish the shape of the smaller and larger components.

Together with raw rescaled connected component, the shape based feature vector is also composed of four other size based features:

1. Normalized length - It is the ratio of the length of the component to the length of the input image.

2. Normalized height - It is the ratio of the height of the component to the height of the input image.

3. Aspect ratio of a component - It is ratio of length to height

4. The ratio of the number of foreground pixels to the total rescaled area.

B. *Surrounding context of connected component*

Generally, the text components are aligned horizontally in the document as compared to the non-text components. Hence, we use the surrounding components also to build the feature vector. Each connected component with its surrounding connected area is rescaled to the 40X40 window size for generating the context based feature vector. The surrounding context area is not fixed for all connected components but it is a function of components length (l) and height (h). The function is such that, for each connected component the area of dimensions is 5xl by 2xh. The size of the context based feature vector is 1600.

Hence the total size of the feature vector is 3204 which consists of raw rescaled shape (1600), raw rescaled context (1600), and four size based features.

C. *Classification*

For classification, the paper makes use of Auto-MLP, a self-tuning classifier that can automatically adjust learning parameters. For these classifiers, learning parameters are chosen from parameter space which has been sampled according to probability distribution function. All of these MLPs are trained for few epochs and then half of these classifiers are selected for next generation based on performance. After MLP classifier, it labels each connected component based on the classification probabilities as text and non-text.

IV. CLUSTERING BASED TECHNIQUES

Apart from edge and shaped methods, there are techniques which are derived from data mining to facilitate the process of segmentation. The paper [6] proposed the algorithms used are K-Means, EM which are useful in terms of segmenting images.

A. *K-Means Algorithm*

K-Means algorithm is an example of unsupervised clustering algorithm. It classifies the input data points into different clusters based on their Minkowski distance.

$$(\sum_{i=1}^{n}|x_i - y_i|^p)^{\frac{1}{p}} \qquad (2)$$

The algorithm assumes that the bits of the image form a vector space and tries to clusters them naturally into according to their intensities. The points are clustered around centroids μi ∀ i ranging from 1 to k in pursuit of minimizing the distance of the data points from the centroids of their respective clusters. The algorithm uses an iterative approach to cluster the data points. Here the data points are nothing but the pixel density.

The algorithm is given below

1. Calculate the histogram graph of the intensities of the pixel of a particular image.

2. Randomly select k data points that will act as a centroid for a particular cluster.

3. Follow the given steps again until the cluster a label of the image does not change anymore.

4. Cluster the points based on the metric used for the relative change in the intensities from the centroid intensities.

$$c^{(i)} := \arg min \left\|x^{(i)} - y^{(i)}\right\|^2$$

5. Compute the new centroid for each of the clusters.

$$\mu^{(i)} := \frac{\sum_{i=1}^{m} 1\{c_{(i)} = j\} x^{(i)}}{\sum_{i=1}^{m} 1\{c_{(i)} = j\}}$$

The parameter on which the above algorithm is tuned is k. k denotes the number of clusters to be formed for given set of data points. The characters in a text are clusters into similar cluster due to the fact that most of the characters are of same intensities and thus belong to the same cluster.

B. Expected Maximization

When it comes to unsupervised learning, the most omnipresent algorithm used is Expected Maximization. The data model is dependent on the hidden variables and the method depends on computing the maximum a posterior (MAP) estimate of the parameters. In Expected Maximization, the steps are performed iteratively till all consecutive iterations give the same value. The Expectation Step (E step) computes the probability of hidden variables being observable. The next step i.e. the Maximization Step (M step) maximizes the probability of the expected probability found in the previous step. Now, again the E step and M step are repeated so that the values of the result reach a constant point. The tuning factor or the parameter is calculated in the M step are used in the previous step. The above explanation can be mathematically expressed as:

Given training dataset $\{x_{(i)}, x_{(i2)}, \dots x_{(m)}\}$ and model $p(x, z)$ where z is the latent variable, we have:

$$l(\theta) = \sum_{=1}^{m} \log p(x; \theta)$$
$$= \sum_{=1}^{m} \sum_{z} p(x, z; \theta)$$

It is evident in the above equation that the log probability is described in terms of x, z and θ. But since z, the hidden variable is not known; we approximate its value. The approximations used are derived mathematically using the E & M steps and is given below.

As observed from the above equation, the log probability is described in terms of x, z and θ. Expectation Step, ∀ i:

$$Q_i\left(z^{(i)}\right) := p(z^{(i)} | x^{(i)}; \theta$$

Maximization Step, ∀ z:

$$\theta := arg \ \max_\theta \sum_i \sum_{z^{(i)}} Q_i(z^{(i)}) \log \frac{p(x^{(i)}, z^{(i)}; \theta)}{Q_i(z^{(i)})}$$

Where Qi is the posterior distribution of z (i)’s given the x (i)’s.

Theoretically, the Expected Maximization Algorithm is an alternative to the K Means Algorithm where a point that is a member of a given cluster is incomplete and is not integral.

V. COMPARISON OF VARIOUS APPROACHES

Table 1. Comparison of Various Approaches

Author	Work done on	Concept Used	Data set	Result
Satadal Saha, Subhadip Basu, Mita Nasipuri and Dipak Kr. Basu	Business Card Reader (BCR) system	Hough Transform	45 documents containing 812 lines comprising of 7300+ words	85.70%
Syed Saqib Bukhari, Mayce Ibrahim Ali Al Azawi, Faisal Shafait, Thomas M. Breuel	Subset of UW-III, ICDAR 2009 page segmentation competition test images and circuit diagrams	Machine Learning on Connected Components	95 documents and 18 images	97.25
Suman Tatiraju, Avi Mehta	Gray-scaled images.	K-means, Expected Maximization	-	-

VI. CONCLUSION

The above survey concludes that remarkable work has been done for image segmentation. But there is more scope for improvements. Some of the key improvements could be in terms of segmentation of cursive handwriting in images. In conclusion, we hope that this lucid discussion will clarify the approaches and methodologies involved in it and would aid to the future researchers.

ACKNOWLEDGEMENT

Special thanks to Mrs. Khushali Deulkar for guiding us through the implementation of the algorithm and enriching the quality of the research.

REFERENCES

[1]. Richard G. Casey and Eric Lecolinet. ”A Survey of Methods and Strategies to Image Segmentation”, IEEE Transactions On Pattern Analysis And Machine Intelligence, Vol. 18, No. 7,1999. p. 690-706

[2]. Archana A. Shinde, D.G.Chougule. ”Text Pre-processing and Text Segmentation for OCR”, IJCSET Volume 2, Issue 1; 2012. p. 810-812

[3]. Satadal Saha, et al. “A Hough Transform based Technique for Text Segmentation”, Journal of Computing, Volume 2, Issue 2, February 2010, ISSN 2151-9617. p.134-141

[4]. Nobuyuki Otsu (1979). "A threshold selection method from gray-level histograms". IEEE Trans. Sys., Man., Cyber. 9 (1): 62–66.

[5]. Syed Saqib Bukhari, et al. “Document Image Segmentation using Discriminative Learning over Connected Components”, Proceeding DAS '10 Proceedings of the 9th IAPR International Workshop on Document Analysis Systems p. 183-190

[6]. Suman Tatiraju, Avi Mehta, “Image Segmentation k-means clustering, EM and Normalized Cuts”, University of California – Irvine.

Harshil Shah, B.E. in Computer Engineering, Dwarkadas J. Sanghvi College of Engineering, Mumbai, India.

Rahul Shah, B.E. in Computer Engineering, Dwarkadas J Sanghvi College of Engineering, Mumbai, India.