ISSN 2278-3091
Volume 9, No.1.1, 2020
International Journal of Advanced Trends in Computer Science and Engineering
Available Online at http://www.warse.org/IJATCSE/static/pdf/file/ijatcse7691.12020.pdf
https://doi.org/10.30534/ijatcse/2020/7691.12020

A Palm Oil Fresh Fruit Bunch Feature Extraction for CBIR Ripeness System

Danial Md Nor[1]*, Nor Afifa Md Noor[2], Mohd Helmy Abd Wahab[3], Ibrahim Masood[4], Ashari Kasmin[5]

[1]Faculty of Electrical and Electronics Engineering, Universiti Tun Hussein Onn Malaysia Parit Raja, Johor, Malaysia
[2]Faculty of Electrical and Electronics Engineering, Universiti Tun Hussein Onn Malaysia Parit Raja, Johor, Malaysia
[3]Faculty of Electrical and Electronics Engineering, Universiti Tun Hussein Onn Malaysia Parit Raja, Johor, Malaysia
culty of Mechanical and Manufacturing Engineering, Universiti Tun Hussein Onn Malaysia Parit Raja, Johor, Malaysia
[5]Faculty of Engineering Technology, Universiti Tun Hussein Onn Malaysia Parit Raja, Johor, Malaysia
* danial@uthm.edu.my

ABSTRACT

In this new development era, image processing application has become high demanding whether in factory sector or agricultural sector as everyone has understand the advantages of the image processing techniques. Content Based Image Retrieval (CBIR) is an image pro-cessing system that returns the retrieved images with the same characteristics as the query image. By applying this CBIR for the palm oil fresh fruit bunch (FFB) ripeness grading system, the manual grader can ease their work as there will be less physical works needed to clas-sify the FFB. Moreover, there also will be no issues about the unsynchronized of the FFB ripeness grading as human grader are too subjec-tive. There are several techniques used to extract the colour of the palm oil FFB image features. Each of the techniques used will be different in terms of how the extraction has been made and it can be seen by retrieved images after the similarity matching. For this project, the feature extraction of colour descriptor is used and two techniques from the feature extraction are applied for the palm oil FFB ripeness grading sys-tem which are Equivalent Ellipse (EE) for Kernel Density Estimation (KDE) Colour and Colour Moment. Colour Moment used the standard deviation and skewness to obtain the retrieved image. Whereas the EE for KDE Colour, will analyze the colour of an image based on the density of the mean and variance. KDE often associated with EE. Result from this system assessed retrieved images which had been calcu-lated by the similarity of feature vector from retrieved query image by using Euclidean Distance (ED) measurement. Results of data had been shown through the standard evaluation technique made by precision and recall to compare both of the techniques results. The obtained graph has shown that EE for KDE Colour was better than Colour Moment as the accuracy for each technique are 98% and 59% respectively.

Key words : Image processing, CBIR, colour descriptor, ripeness, grading, palm oil fresh fruit bunch (FFB).

1. INTRODUCTION

Nowadays, the development is going up in all around the world. At the same time, the demand of electricity has a major change that has occurred in this world is the modern technology, where many software and hardware development exist to improve the image processing especially for animation or graphics either in Internet or television[1]. The understanding of the image processing technique purpose has been increased in order to get the validation or how the image is built. Agricultural modern sector also are not missed towards the modern technologies of this image processing technique as the demand of good quality food products produced in a bit has mark up and palm oil is one of the major products[2]. Currently, the local palm oil plantations still rely on manual grading of palm oil fresh fruit bunch (FFB) through human vision[3].

Based on Malaysian Palm Oil Board 2003, colour of palm oil fruit is one of the influential factors which can classify the grade and quality of the palm oil fresh fruit bunch (FFB)[4]. Variations of image processing techniques engaged in classifying the grade of fruits ripeness automatically using Neuro-fuzzy technique, machine vision, inductive attribute and artificial neural network[5]. In this research project, the palm oil fresh fruit bunch (FFB) is utilized as sample of the testing improvement of the FFB ripeness grading system by using image processing techniques. This is because the image processing technique is invented especially to avail the typical palm oil ripeness grading system. Overall, there are sixteen (16) variations of palm oil FFB as shown in Figure 1 but in this case the image processing techniques will cover three main ripeness category of palm oil FFB which are ripe, unripe and over ripe[3].

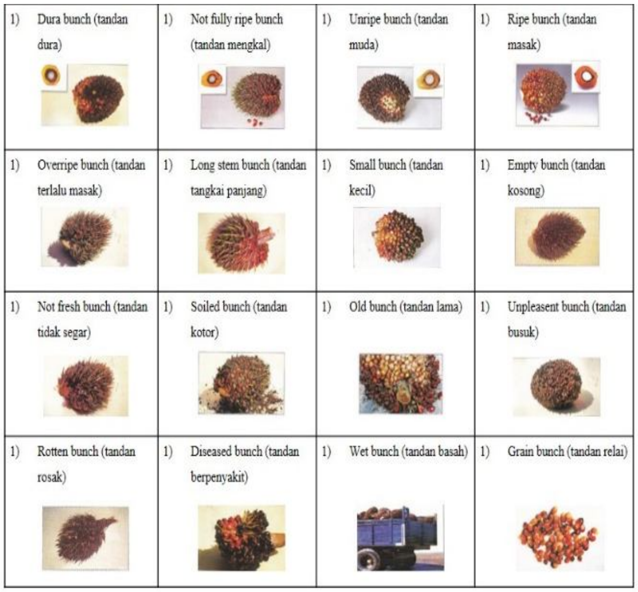

Figure 1: The variations of palm oil fresh fruit bunch (FFB)

Low-level feature extraction is the most basic step of image processing procedures which aim at symbolic of the image content [6]. Besides, it also aims to improve the quality of the input image by means of enhancement or noise removal processes. In this project, the input image is the image of palm oil FFB fruit captured. There are dozens of procedures has been proposed from the researches for the extraction of this low-level feature extraction descriptor but only few of them are adopted which can be distinguished as colours, shapes, textures, points, edges and regions[6]. Somehow, there is no exact theory found to be available for the descriptor used for the low-level feature extraction. Meanwhile, many techniques in image processing still referred to low level feature extraction as it is the most basic procedure to extract the huge data into simpler and essential result needed. As compared to high-level feature extraction, the procedure is more complicated because involving the human brain. The feature vector used in this project is colour descriptor as it can encode the colour feature vector of the palm oil fresh fruit bunch (FFB) image.

In agricultural sector, the relationship between colour and the level of ripeness has widely used. Factually, colour descriptor of images consist of a number of histogram descriptor as it describes a whole or within the interest region of image. The histogram descriptor usually represented by histogram colour vector, Equivalent Ellipse (EE) for Kernel Density Estimation (KDE) colour, colour correlogram and colour moment [3][4]. Some of the advantages of this colour descriptor are insensitive towards any changes in image and histogram resolution, the high percentage of effectiveness, simplicity in implementation and computational and it requires low storage[7].

2. CONTENT-BASED IMAGE RETRIEVAL (CBIR)

The volume of image database is expanding with the fast evolution of image retrieval invention and the tool that suits for this purpose is content-based image retrieval(CBIR)[8]. CBIR act as the tools that utilizes the visual content of an image, to find the identical images in dominant areas of image databases, according to a user's choice[9]. Researchers has proved the CBIR is one of the successful technology to improve the interface between user and computer[10]. Based on low level feature extraction, colour, texture and shape can be the best descriptor for this CBIR system but in this project only colour descriptor is highlighted. The basic algorithm used in CBIR are extraction, selection and classification.

2.1 Colour Moment

Color moment is used to recognize images based on their color feature by using a probability distribution of the whole image. Usually, the normal distribution are calculated by the mean and variance. The greater probability distribution of color in an image can be used to identify the image based on color[11]. The color distribution of an image is calculated as shown in Table 2 [12]:

Table 1: Equation of colour moment

Moments	Equation	Description
Mean	$\mu = \frac{\sum_{i,j} x_{i,j}}{N}$	Average intensity of an image that have N pixels.
Standard Deviation	$\sigma = \left(\sqrt{\frac{\sum_{i,j} x_{i,j}}{N}}\right)^2$	Variance in the standard deviation is the intensity variation around the mean
Skewness	$\frac{\sum_{i,j}\left(X_{i,j} - \mu\right)^3}{N\sigma^3}$	Shows degree asymmetry in the distribution

The index entry for one image consist of:

Index size = Number of regions x Number of colour
= Channels x 3 floating point numbers

If the image is divided horizontally into three equal regions then 27 floating point numbers per image are stored.

2.2 Equivalent Ellipse (EE) for Kernel Density Estimation (KDE) Colour

An equivalent Ellipse (EE) is used to represent each of the mentioned density surface, which reduce the size of the features from 2^{2m} to six features. The calculation of this technique is same as the color moment but difference in the analysis of color images. As the Kernel Density Estimation (KDE) technique need to associate with the Equivalent Ellipse (EE), the surface density should be considered through its center, semi-major axis, minor axis, the mean value and the angle of orientation, the matching process is taken from the estimation of surface density. Position of peaks and valleys on the surface density are different in the original image and the captured image, because of the presence of the dominant color coating. This condition usually occurs due to changes in lighting conditions or image capture equipment. Based on both EE of captured there are still original image that are preserved and location EE that has shifted. In some cases only center close but EE and axis orientation is different, that help to distinguish the image has a different color[13]. The equation involved are as followed:

$$\text{Mean, } \mu = \frac{\sum_{i,j} x_{i,j}}{N} \text{ ....[1]}$$

$$\text{Variance, } \sigma = \sqrt{\frac{\sum_{i,j} x_{i,j}}{N}} \text{ ....[2]}$$

In most cases in analysing the result of EE, there are difficulties in the identification methods, which anchor on the matching of density surfaces. The primary goal involves the identification of images that have similar colour content and the reduction of elements to speed up matching. Thus, this EE is preferred for the density surface.

3. PALM OIL FFB GRADING

Currently, palm oil FFB is graded manually by human visual inspections based on their observation on the percentage of detached fruitlets from FFB. It can be observed the ripe bunch is in orange reddish colour, the outside mesocarp is in orange colour and less than ten (10) fresh fruitlets are separated from the FFB. While, for the overripe bunch it is in darkish reddish colour and more than fifty percent (50%) fresh fruitlets are separated from the FFB but at least ten percent (10%) fresh fruitlets attached to the FFB. The last one which is unripe bunch. This unripe bunch is in purple blackish, the outside mesocarp is in yellowish colour and none of the fresh fruitlets is separated from the FFB even if there are separated, it might not due to the normal ripening process. Somehow, it is challenging to classify the FFB based on the visual inspection of the colour especially the darkish red (i.e. overripe), orange reddish (i.e. ripe) and purple blackish (i.e. under ripe). The frequently issues of dissatisfaction and dispute among estate owners and factories supervisors is due to the improper grading. Hence, to overcome this problem, a palm oil fresh fruit bunch (FFB) ripeness grading system by using image processing is proposed.

4. MATERIALS AND METHODS

The proposed system for this project is implemented with the aided of algorithm, technique and tools. The result which are relevant to a query images is known as the retrieved image and considered it as the output. The proposed system structure are shown in Figure 2 as followed:

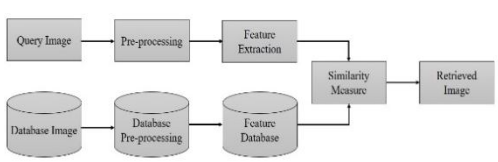

Figure 2: The block diagram of content based image retrieval system (CBIR)

The query and database had the similar process which are pre-processing, feature extraction and similarity measure where the image is retrieved based on the similar feature vector with the query image. Starting from the query image, the user provide 30 images of palm oil fresh fruit bunch (FFB) as the sample of query for the system. While for the database, 96 images of palm oil fresh fruit bunch (FFB) is provided by the user. Then, an image is selected as the query image to go through the pre-processing where the image from the document will be format to the setting resolution. The setting resolution is 540x720 pixel. The database image also will be format to the setting resolution.

Continue with the most important part in this system which is feature extraction. In this part, the color feature is extract from the query in the form of vector so do the images in the database. All images stored in feature database are generated by feature vector. Then the query image and the database image are calculated to find the similarities. The result that has been retrieved is not only one image but there is a few more images and it all depends on the similarity with the query images. This system using Euclidean Distance (ED) to simplified the image. The sorted Euclidean Distance (ED) between the query image and database is arranged in ascending order. The lower the value of the ED, the greater the similarities between the query and retrieve image. The feature vector are calculated for the precision and call up to measure the retrieval performance of algorithm. Lastly, the system retrieves and display a sequence of images by ranked. The ranked of images are depend on the calculation of ED techniques.

4.1 Calculation for the Similarity

In this project, the result obtained are based from two color feature vector which are EE of KDE Color and Color Moment. Through this method, the user can store images into database and able to compared with the query images. The feature vector from query image will be extracted and the similarity with the images in database is computed by using Euclidean Distance (ED). The distance metric of Euclidean Distance (ED) is commonly used due to the efficiency and effectiveness[14]. The gap between two vectors is measured by calculating the square root of the sum of the squared absolute differences and it can be calculated as[14], [15] :

$$\Delta d = \sqrt{\sum_{i=0}^{n} (|Q_i - D_i|)^2} \text{ .... [3.1]}$$

The method to evaluate the image in this project using precision and recall. The precision in image retrieval can be defined as the measurement of the retrieved relevant images to the query of the total retrieval images[16]. The measurement of the precision can be calculated as:

$$\text{Precision} = \frac{A}{B} \text{ .... [3.2]}$$

Where ‘A’ is the relevant retrieved images and ‘B’ is the total retrieved images.

While, the recall in images retrieval can be define as the measurement of the retrieval relevant image to the total data base images[14]. The measurement of the recall can be calculated as:

Recall = $\frac{A}{C}$ ………. [3.3]

Where ‘A’ is the relevant retrieved image and ‘C’ is the total number of relevant images in the database.

5. RESULTS AND DISCUSSION

In this section, two results based on two different techniques for color feature extraction will be discussed. Based on the results obtained, the retrieved images for both color feature technique are different and it can be summarize by Precision and Recall method.

5.1 Colour Moment Technique

Table 2, presents the summary of the analysis of the precision and recall of the images by using Color Moment. For precision, the number of relevant retrieved images are same for each query and the total number of the retrieved images is 100. For a table of recall, the number of relevant images is also same for each query image and the total of relevant images in the database defined as 100.

Table 2: Precision and recall for the query image by EE for KDE Color

Overripe FFB			Ripe FFB			Under ripe FFB		
Query Image	Precision	Recall	Query Image	Precision	Recall	Query Image	Precision	Recall
1	0.1	0.1	11	0.1	0.1	21	0.1	0.1
2	0.4	0.4	12	0.2	0.2	22	0	0
3	0.2	0.2	13	0.2	0.2	23	0.1	0.1
4	0.1	0.1	14	0.4	0.4	24	0.2	0.2
5	0.4	0.4	15	0	0	25	0.4	0.4
6	0	0	16	0.1	0.1	26	0	0
7	0	0	17	0.2	0.2	27	0.2	0.2
8	0.4	0.4	18	0.4	0.4	28	0.2	0.2

Table 3, there are different types of ripeness which are overripe, ripe and under ripe. Based on these three types, the parameter of precision and recall are calculated. Precision is given by number of relevant images retrieved to the total number of image retrieved, whereas Recall is the number of relevant images retrieved to the number of relevant images in the database. The accuracy rows proved that this technique can retrieved many similar image.

Table 3: Results of precision, recall and accuracy of EE for KDE Color

Images	No. of Images in Database	No. of Retrieve Images	No. of Relevant Images	Precision	Recall	Accuracy
Overripe FFB	100	100	24	0.24	0.24	0.62
Ripe FFB	100	100	19	0.19	0.19	0.595
Under ripe FFB	100	100	12	0.12	0.12	0.56

Based on Figure 4, it shows the occurrence of the point of intersection of precision and recall. This is important because there will be the enactment of CBIR techniques. The greater the crossing point area, the better the retrieved image. Based on the graph, it can be say that the retrieved image is not giving the best result because no intersection occurred.

Figure 4: Graph of precision/recall against the query image of colour moment

5.2 EE for KDE Colour Technique

Table 4, presents the summary of the analysis of the precision and recall of the images by using EE of KDE Color. For precision, the number of relevant retrieved images are same for each query and the total number of the retrieved images is 100. For a table of recall, the number of relevant images is also same for each query image and the total of relevant images in the database defined as 100.

Table 4: Precision and Recall for the Query Image by EE of KDE Colour

Overripe FFB			Ripe FFB			Under ripe FFB		
Query Image	Precision	Recall	Query Image	Precision	Recall	Query Image	Precision	Recall
1	1	1.00	11	1	1.00	21	0.6	0.60
2	1	1.00	12	1	1.00	22	1	1.00
3	1	1.00	13	1	1.00	23	1	1.00
4	1	1.00	14	1	1.00	24	1	1.00
5	1	1.00	15	1	1.00	25	1	1.00
6	1	1.00	16	1	1.00	26	1	1.00
7	1	1.00	17	1	1.00	27	1	1.00
8	1	1.00	18	1	1.00	28	0.7	0.7

Table 5, there are different types of ripeness which are overripe, ripe and under ripe. Based on these three types, the parameter of precision and recall are calculated. Precision is given by number of relevant images retrieved to the total number of image retrieved, whereas Recall is the number of relevant images retrieved to the number of relevant images in the database. The accuracy rows proved that this technique can retrieved many similar image.

Table 5: Results of precision, recall and accuracy for EE for KDE Colour

Images	No. of Images in Database	No. of Retrieve Images	No. of Relevant Images	Precision	Recall	Accuracy
Overripe FFB	100	100	100	1	1	1
Ripe FFB	100	100	100	1	1	1
Under ripe FFB	100	100	88	0.88	0.88	0.94

Based on Figure 5, it shows the occurrence of the point of intersection of precision and recall. This is important because there will be the enactment of CBIR techniques. The greater the crossing point area, the better the retrieved image. Based on the graph, the intersection cannot be seen because both of them are overlapping which means the image obtained is the most perfect.

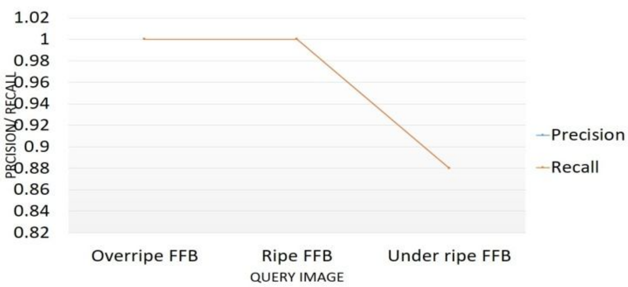

Figure 5: Graph of precision/recall against the query image of EE for KDE Colour

5.3 The Comparison between EE for KDE Colour and Colour Moment

Refer to Table 8, it shows the percentage accuracy for each technique that are being analyzed. While in Figure 6, it shows clearer differences accuracy between EE of KDE Color and Color Moment. Thus, it can be conclude that EE of KDE Color technique is better to recover the image as compared with the technique of Color Moment. Based on the observation, overripe FFB and ripe FFB give the accuracy at 100% while using the EE of KDE Color technique. This can be the indicator to determine which technique have a better performances in this CBIR system.

Table 6: % Accuracy of each technique.

Ripeness of FFB	% Accuracy for EE of KDE Color	% Accuracy for Color Moment
Overripe FFB	100	62
Ripe FFB	100	60
Under ripe FFB	94	56
Average %	98	59

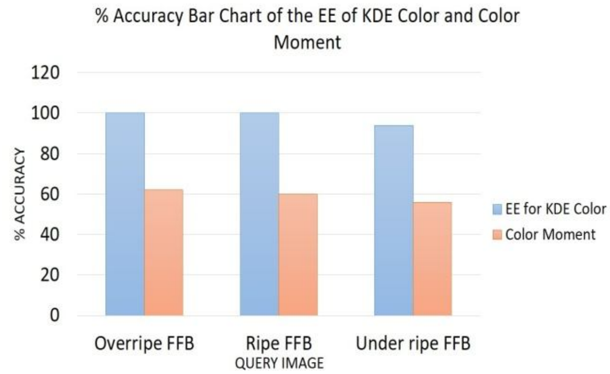

Figure 6: Comparison between EE for KDE colour and colour moment

6. CONCLUSION

The purpose of this system is to overcome the manual grading of the palm oil fresh fruit bunch (FFB) in agricultural industry by using image processing technique. As a conclusion, the development of this CBIR on the palm oil FFB is able to analyze, identify and retrieve the similar image from the database if query image is provided. This system has given two choices to the user to choose appropriate technique for the palm oil FFB ripeness grading. The selection of the good performances of the techniques is based on the precision and recall method. The development of CBIR on the palm oil FFB ripeness grading system showed that the EE of KDE Color technique has a good performances compared to Color Moment as the percentage accuracy are 98% and 59% respectively. Besides, the time to retrieve the similar image by using EE of KDE Color is faster compared to Color Moment. Overall, the project has been successfully implemented. The result obtained are satisfied and this project objectives are achieved.

RECOMMENDATIONS

There are a few recommendations for this project to enhance the system for more convenient in future. First, use the other technique which is the combination with other features such as shape and texture. This way is to improve the performances of the CBIR to detect the ripeness of palm oil FFB. Next, add more images of the palm oil FFB with a white background. This addition will able to give a better results especially when plotting the graphs. The system can also be integrate with the cell phone or any other imaging device to make the grader more convenient to classify the ripe, unripe and overripe of palm oil FFB.

REFERENCES

[1] M. S. M. Alfatni *et al.*, “Oil palm fresh fruit bunch ripeness classification based on rule- based expert system of ROI image processing technique results,” *IOP Conf. Ser. Earth Environ. Sci.*, vol. 20, pp. 012–018, 2014.

[2] N. Jamil, A. Mohamed, and S. Abdullah, “Automated grading of palm oil Fresh Fruit Bunches (FFB) using neuro-fuzzy technique,” *SoCPaR 2009 - Soft Comput.*

Pattern Recognit., no. 1, pp. 245–249, 2009.

[3] P. Fruit, "Novel Method of Grading Fresh Fruit Bunches (FFB) of Oil Palm Fruit Novel Method of Grading Fresh Fruit Bunches (FFB) of Oil," *Conf. Sci. Soc. Res. 14*, no. November 2016, 2009.

[4] J. Roseleena, J. Nursuriati, J. Ahmed, and C. Y. Low, "Assessment of palm oil fresh fruit bunches using photogrammetric grading system," *Int. Food Res. J.*, vol. 18, no. 3, pp. 999–1005, 2011.

[5] N. Fadilah, J. M. Saleh, H. Ibrahim, and Z. A. Halim, "Oil palm fresh fruit bunch ripeness classification using artificial neural network," *2012 4th Int. Conf. Intell. Adv. Syst.*, vol. 1, pp. 18–21, 2012.

[6] W. Förstner, "A Framework for Low Level Feature Extraction," *3rd Eur. Conf. Comput. Vis.*, pp. 383–394, 1994.

[7] R. S. Chora, "Image Feature Extraction Techniques and Their Applications for CBIR and Biometrics Systems," vol. 1, no. 1, 2007.

[8] K. Juneja, "A Survey on Recent Image Indexing and Retrieval Techniques for Low-level Feature Extraction in CBIR systems," 2015.

[9] M. Saad, "Low-Level Color and Texture Feature Extraction for Content-Based Image Retrieval," *Final Proj. Report, EE K*, 2008.

[10] Z. Lei, L. Fuzong, and Z. Bo, "A cbir method based on color-spatial feature," pp. 166–169, 1999.

[11] A. M. Vyas, B. Talati, and S. Naik, "Colour Feature Extraction Techniques of Fruits: A Survey," *Int. J. Comput. Appl.*, vol. 83, no. 15, pp. 15–22, 2013.

[12] V. P. Singh, "Improved Image Retrieval Using Color-Invariant Moments," pp. 1–6, 2017.

[13] M. Jones and J. Rehg, "Statistical Color Models with Application to Skin Detection 2 Histogram Color Models," *Comput. Vis. Pattern Recognit.*, vol. 46, no. 1, pp. 1–23, 1999.

[14] F. Malik and B. Baharudin, "Analysis of distance metrics in content-based image retrieval using statistical quantized histogram texture features in the DCT domain," *J. King Saud Univ. - Comput. Inf. Sci.*, vol. 25, no. 2, pp. 207–218, 2013.

[15] S. Sergyan, "Color histogram features based image classification in content-based image retrieval systems," *2008 6th Int. Symp. Appl. Mach. Intell. Informatics*, no. February, pp. 221–224, 2008.

[16] F. Mokhtarian and A. Mackworth, "Scale-Based Description and Recognition of Planar Curves and Two-Dimensional Shapes," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-8, no. 1, pp. 34–43, 1986. https://doi.org/10.1109/TPAMI.1986.4767750