Detection and Classification of Apple Fruit Diseases using K-means clustering and Learning Vector Quantization Neural Network

[1]Mr. Abhijeet V. Jamdar, [2]Prof. A. P. Patil

Department of E&C
J.J.M.C.O.E
Jaysingpur, India

***Abstract*: In development of India there is major contribution of agricultural field. There are so many fruits that are exported from India and give more profit to the farmers. Diseases in fruit cause devastating problem in economic losses and production in agricultural industry. Manual identification of defected fruit is very time consuming. Web-based system used to help non experts in identifying fruit diseases, based on the picture representing the symptoms of the fruit.**

In this paper, we suggest a solution for the detection and classification of three common diseases on apple fruit (apple rot, apple scab, apple blotch). The image processing based proposed approach is composed of the three main steps; in the first step K-Means clustering segmentation algorithm is used for the image segmentation, in the second step some state of the art features are extracted from the segmented image, and finally images are classified into one of the classes by using Learning Vector Quantization Neural Network. Experimental results showed that the proposed system can significantly support accurate detection and automatic classification of apple fruit diseases. For the proposed solution we get more than 95% classification accuracy.

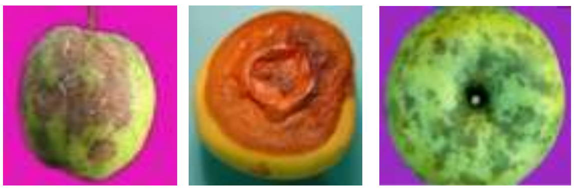
***Keywords*: Keyword: K-means clustering segmentation algorithm, Learning Vector Quantization Neural Network, apple fruit diseases**

I. INTRODUCTION

Agriculture is the mother of all cultures. It has played a key role in the development of human civilization. Agricultural production system is an outcome of a complex interaction of soil, seed and agro chemicals (including fertilizers). Therefore, there is need to take good judgment for management of all the inputs is essential for the sustainability of a complex system. The focus on increasing the productivity, without considering the ecological impacts has resulted into environmental degradation. Diseases and insect pests are the major problems in agriculture. These require careful diagnosis and timely handling to protect the crops from heavy losses. In plant, diseases can be found in various parts such as fruit, stem and leaves. In development of India there is major contribution of agricultural field. There are so many fruits like apple, oranges, mango and grapes etc. that are exported from India and give more profit to the farmers.

Diseases in fruit cause devastating problem in economic losses and production in agricultural industry worldwide. Generally the naked eye method is used to identify the diseases. In this method experts are involved who have the ability to detect the changes in leaf color. This method involves lots of efforts, takes long time and also not practical for the large fields. Many times different experts identify the same disease as the different disease. This method is expensive as it requires continuous monitoring of experts.

Apple fruit diseases can cause major losses in yield and quality appeared in harvesting. To know what control factors to take next year to avoid losses, it is crucial to recognize what is being observed. Some common diseases of apple fruits are apple scab, apple rot, and apple blotch [5], as shown in Fig. 1. Apple scabs are gray or brown corky spots. Apple rot infections produce slightly sunken, circular brown or black spots that may be covered by a red halo. Apple blotch is a fungal disease and appears on the surface of the fruit as dark, irregular or lobed edges.



(a) (b) (c)

Fig.1 Three common diseases on Apple fruit:
(a) Apple Scab, (b) Apple Rot, (c) Apple Blotch

A typical image recognition process includes image preprocessing, segmentation, feature extraction and pattern recognition. Image segmentation is one of the key steps, and the precision of the segmentation directly influences the reliability of feature extraction and the accuracy of recognition.

This paper presents an efficient solution for detection and classification of diseases on apple fruit. The proposed approach is composed of the three main steps; in the first step K-Means clustering segmentation algorithm is used for the image segmentation, in the second step some state of the art features are extracted from the segmented image, and finally images are classified into one of the classes by using Learning Vector Quantization Neural Network. using K-means clustering technique based on color features from the images. In order to validate the proposed approach, we have considered three types of the diseases in apple; apple scab, apple rot and apple blotch.

The rest of the paper is structured as follows: Section 2 gives a brief overview of previous work; Section 3 explains the proposed approach for the apple fruits detection and classification problem; Section 4 reports the experimental results; and Finally, Section 5 draws the conclusion.

II. LITERATURE REVIEW

Shiv Ram Dubey and Pushkar Dixit (2013) presented a novel defect segmentation of fruits based on color features with K-means clustering unsupervised algorithm. They had taken apple as a case study and evaluated the proposed approach using defected apples For defect segmentation they used color images of fruits. Defect segmentation is carried out into two stages. At first, the pixels are clustered based on their color and spatial features, where the clustering process is accomplished. Then the clustered blocks are merged to a specific number of regions [1].

In [2] Shiv Ram Dubey (2012) proposed an image processing approach has been used for fruit disease identification. The research has conducted for apple disease namely apple scab, apple rot and apple blotch. For image segmentation K-means clustering technique is used. Feature extraction is done from segmented images. Features considered for feature extraction are color histogram, color coherence vector, local binary patterns and complete local binary patterns. For fruit disease identification Multiclass support vector machine is used.

Feiyun Zhang (2013) proposed new method for recognition of wheat leaf disease images based on learning vector quantization neural network (LVQNN) and multiple feature parameters of their images. The diseased color image region was segmented by K-means hard clustering and then color feature parameters and texture parameters were extracted by using lifting wavelet transform and pulse coupled neural network (PCNN) from the color segmentation image. At last, combination feature parameters were used to index LVQNN to recognize the wheat leaf disease [3].

In [4] Rajeshwar Dass, Priyanka and Swapna Devi (2012) described the different segmentation techniques used in the field of ultrasound and SAR Image Processing. Image segmentation is the process of partitioning an image into multiple segments, so as to change the representation of an image into something that is more meaningful and easier to analyze. Several general-purpose algorithms and techniques have been developed for image segmentation.

In [5] "Automatic fruit and vegetable classification from images by Anderson Rocha & etc. proposed a system introduces a unified approach that can combine many features and classifiers that requires less training and is more adequate, where all features are simply concatenated and fed independently to each classification algorithm. To deal with Contemporary Vision and Pattern Recognition problems using just one feature descriptor is a difficult task and feature fusion may become mandatory. Although normal feature fusion is quite effective for some problems, it can yield unexpected classification results when the different features are not properly normalized and preprocessed. The introduced fusion approach is validated using a multi-class fruit-and-vegetable categorization task in a semi-controlled environment, such as a distribution center or the supermarket cashier.

Digital image recognition of plant diseases could reduce the dependence of agricultural production on the professional and technical personnel in plant protection field and is conducive to the development of plant protection informatization [6]. In order to find out a method to realize image recognition of plant diseases, four kinds of neural networks including backpropagation (BP) networks, radial basis function (RBF) neural networks, generalized regression networks (GRNNs) and probabilistic neural networks (PNNs) were used to distinguish wheat stripe rust from wheat leaf rust and to distinguish grape downy mildew from grape powdery mildew based on color features, shape features and texture features extracted from the disease images.

In [10] Dheeb Al Bashish, Malik Braik, and Sulieman Bani-Ahmad (2010) proposed and evaluated a framework for detection of plant leaf/stem diseases. In developing countries the approach for detection of such diseases is based on the naked eye observation by the experts can be expensive and time consuming. The proposed framework is image-processing based and is composed of the following main steps; in the first step the images are segmented using the K-Means technique, in the second step the segmented images are passed through a pre-trained neural network. As a testbed, they used a set of leaf images taken from Al-Ghor area in Jordan.

III. PROPOSED APPROACH

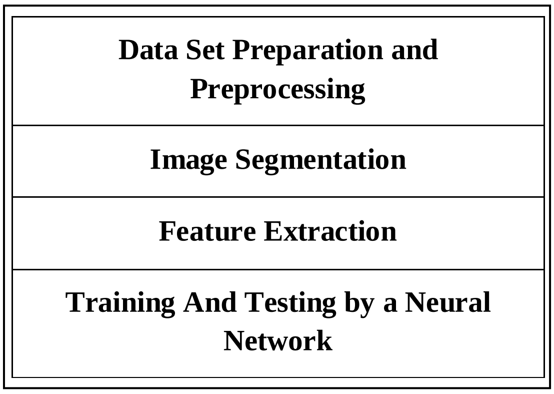


Fig.2 Basic Procedure of proposed Approach

The steps of the proposed approach are shown in the Fig.2. The images of diseased apple are collected from internet. Preprocessing is the technique for enhancing data images prior to computational processing. In this approach the diseased color image region is segmented by K-means clustering algorithm. After segmentation, features are extracted from the segmented image of the fruit. Finally, training and classification are performed on a Learning Vector Quantization Neural Network.

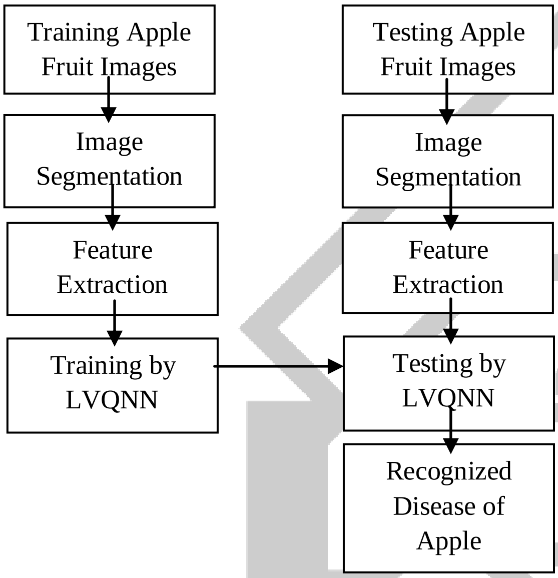


Fig.3 Framework of proposed approach

The framework of the proposed solution is shown in the Fig. 3. Each phase of the proposed method is described in the rest of this section.

A. Image Segmentation

K-Means clustering technique is used for the image segmentation. K-Means clustering algorithm was developed by J. MacQueen (1967). K-means is generally used to determine the natural groupings of pixels present in an image. It is attractive in practice, because it is straightforward and it is generally very fast [1].

It partitions the input dataset into k clusters in which one cluster contains the majority of the diseased part of the image (in this approach we partitions the image into three clusters). Each cluster is represented by an adaptively changing center, starting from some initial values named seed-points. K-means clustering computes the distances between the inputs (also called input data points) and centers, and assigns inputs to the nearest center.

The introduced framework of defect segmentation operates in six steps as follows,

Step 1. Read the input image of defected fruit.

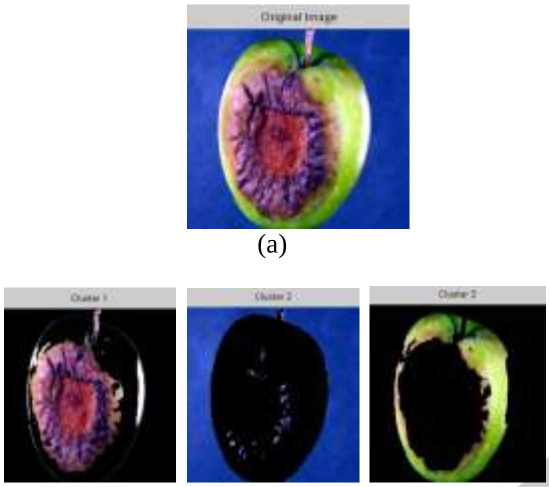
Step 2. Transform Image from RGB to L*a*b* Color Space. We have used L*a*b* color space because it consists of a luminosity layer in 'L*'channel and two chromaticity layer in 'a*' and 'b*' channels. Using L*a*b* color space is computationally efficient because all of the color information is present in the 'a*' and 'b*' layers only.

Step3. Classify Colors using K-Means Clustering in 'a*b*' Space. To measure the difference between two colors, Euclidean distance metric is used.

Step 4. Label Each Pixel in the Image from the Results of K-Means. For every pixel in our input, K-means computes an index corresponding to a cluster. Every pixel of the image will be labeled with its cluster index.

Step 5. Generate Images that Segment the Input Image by Color. We have to separate the pixels in image by color using pixel labels, which will result different images based on the number of clusters.
Step 6. Select the segment containing disease.

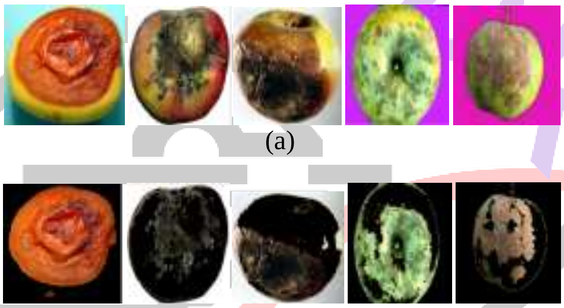
In this experiment input image are partitioned into three segments. Fig. 4 demonstrates the output of K-Means clustering for an apple fruit infected with apple rot disease.



(a) (b) (c) (d)

Fig. 4. K-Means clustering for an apple fruit that is infected with apple rot disease with three clusters (a) The infected apple fruit image, (b) first cluster, (c) second cluster, (d) third cluster

Fig. 5 also depicts some more image segmentation results using the K-Means clustering technique.



(a) (b)

Fig. 5 Segmentation results (a) Images before segmentation, (b) Images after segmentation.

B. Feature Extraction

In the proposed approach, we have used some state of the art color and texture features to validate the accuracy and efficiency. The features used for the apple fruit disease classification problem are Global Color Histogram, Color Coherence Vector, Local Binary Pattern, and GLCM features.

1) Global Colour Histogram (GCH)

The Global Color Histogram (GCH) is the simples approach to encode the information present in an image [16]. The goal is to calculate the HSV global histograms for all the images [17].

2) Color Coherence Vector (CCV)

An approach to compare images based on color coherence vectors are presented in [18]. Color histograms are used to compare images in many applications. Their advantages are efficiency, and sensitivity to small changes in camera viewpoint. Color coherence vector (CCV) stores the number of coherent versus incoherent pixels with each color. By separating coherent pixels from incoherent pixels, CCV's provide finer distinctions than color histograms. CCV's can be computed at over 5 images per second on a standard workstation.

3) Local Binary Pattern (LBP)

Given a pixel in the input image, LBP [19] is computed by comparing it with its neighbors:

$$LBP_{N,R} = \sum_{n=0}^{n-1} s(v_n - v_c)2^n, s(x) = \begin{cases} 1, x \geq 0 \\ 0, x < 0 \end{cases}$$

Where, vc is the value of the central pixel, vn is the value of its neighbors, R is the radius of the neighborhood and N is the total number of neighbors. Suppose the coordinate of vc is (0,0), then the coordinates of vn are (Rcos(2π n / N), Rsin(2π n /N)) . The values of neighbors those are not present in the image grids may be estimated by interpolation. Let the size of image is I*J. After the LBP code of each pixel is computed, a histogram is created to represent the texture image:

$$H(k)=\sum_{i=1}^{I}\sum_{j=1}^{J} f(LBP_{N,R}(i,j),k), k\in[0,K],$$

$$f(x,y)=\begin{cases}1, x=y\\ 0, otherwise\end{cases}$$

Where, K is the maximal LBP code value. In this experiment the value of 'N' and 'R' are set to '8' and '1' respectively to compute the LBP feature.

4). GLCM Features

We use some global measures: mean (μ), contrast (Cn), homogeneity (Hg), energy (En), variance (σ2), correlation (Cr), and entropy (Hn) [5].

We use following 7 GLCM features to train the neural network.

- Mean :

$$\mu = \frac{1}{2}\sum_i i h_s[i]$$

- Contrast :

$$C_n = \sum_j j^2 h_d[j]$$

- Homogeneity :

$$H_g = \frac{1}{1+j^2} h_d[j]$$

- Energy :

$$E_n = \sum_i h_s[i]^2 \sum_j h_d[j]^2$$

- Variance :

$$\sigma^2 = \frac{1}{2}\left(\sum_i (i-2\mu)^2 h_s[i] + \sum_j j^2 h_d[j]\right)$$

- Correlation :

$$C_r = \frac{1}{2}\left(\sum_i (i-2\mu)^2 h_s[i] - \sum_j j^2 h_d[j]\right)$$

- Entropy :

$$H_n = -\sum_i h_s[i]\log(h_s[i]) - \sum_j h_d[j]\log(h_d[j])$$

We calculate values of these features for all the images in dataset. We create Excel Sheet of all these values and use this Excel Sheet is as input to train the Neural Network.

C) Training and testing by LVQNN –

The LVQ neural network uses competitive learning method under the supervision state. In the period of learning, the desired output value determines the assigned category of the input samples. Accordingly, the weakness of unsupervised learning algorithm is eliminated [3].

The LVQ neural network structure is composed of three layers: the first layer is input, the second layer is a competitive layer and the third layer is output. The network structure is shown in figure.

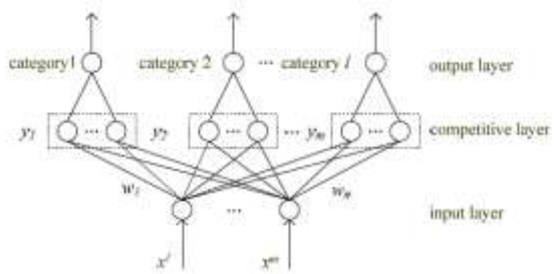


Fig.6 Learning Vector Quantization Neural Network

Learning algorithm of LVQ neural network is as follows:

Step 1: Initialized weight vector (Cluster centers), W={w1,w2,…….,wn} , giving initial value randomly to the weight vector, initialized learning rate **ß** (0) , determined number of training tn .

Step 2: Enter P training mode sample vector x^p

Step 3: By the formula to find a cluster center, that is victory node neurons.

$$\left\| w_k - x^p \right\| < \left\| w_j - x^p \right\| \quad j = 1, 2, \ldots, n$$

Step 4: According to the classification correct or not, winning neuron weights should be adjusted. If the classification is correct, the weight is moved towards the input sample vector. That is:

$$w_k^{new} = w_k^{old} + \beta(x^p - w_k^{old})$$

In turn, Right value is forced to move in the opposite direction of the input vector, that is:

$$w_k^{new} = w_k^{old} - \beta(x^p - w_k^{old})$$

Step 5: Adjust the learning rate **ß** , **ß** Gradually decreases as iterated function.

$$\beta[t] = \beta[0]\left[1 - \frac{t}{t_n}\right]$$

Step 6: Check the training termination condition, if t < tn , k = k +1, enter a new sample vector and back to the step (2) continue to perform until t = tn.

IV. CLASSIFICATION EXPERIMENT RESULTS

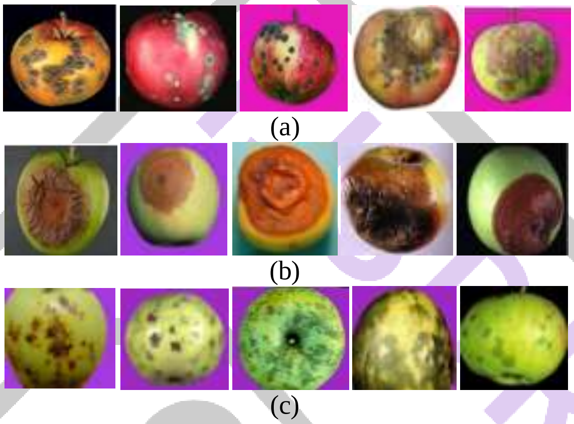


Fig.3. Sample images of data set collected from Internet infected with (a) apple scab, (b) apple rot and (c) apple blotch diseases.

We select 24 samples composed of 7 GLCM features values for each disease as training samples. So, the network has 7 input neurons. Table1 shows the LVQ identification results of apple diseases,

Disease category	**No. of samples**	**Correct identific-ation no.**	**Recognit-ion rate**
Apple Scab	10	09	90%
Apple Rot	08	08	100%
Apple Blotch	06	06	100%

Fig.4 Apple Rot Camera Image

Also we take apple infected with apple rot from market and we take photo of that apple shown in fig.4. Then we use that as input image. Finally we get correct recognition result. So our system is useful for the farmers for fast and real time recognition of apple diseases.

V. CONCLUSION

An image processing based solution is proposed and evaluated in this paper for the detection and classification of apple fruit diseases. The proposed approach is composed of mainly three steps. In the first step image segmentation is performed using K-Means clustering technique. In the second step features are extracted. In the third step training and classification are performed on LVQNN. We have used three types of apple diseases namely: Apple Blotch, Apple Rot, and Apple Scab as a case study and evaluated our program.

Experimental results showed the apple fruit disease recognition rate of the algorithm can reach more than 95%. It proved that the LVQ neural network could effectively recognize the apple disease of apple scab, apple rot and apple blotch.

References:

[1] Shiv Ram Dubey, Pushkar Dixit, Nishant Singh, Jay Prakash Gupta, " Infected Fruit Part Detection using K-Means Clustering Segmentation Technique", International Journal of Artificial Intelligence and Interactive Multimedia, Vol.2, 2013, Page(s): 65-72 .

[2] Shiv Ram Dubey, Anand singh Jalal, **"**Detection and Classification of Apple Fruit Diseases using Complete Local Binary Patterns" IEEE, Third international conference on Computer and communication Technology, 2012.

[3] Feiyun Zhang, "Image Recognition of Wheat Leaf Disease based on Learning Vector Quantization Neural Network and Combination Feature Parameters" , Journal of Convergence Information Technology, April 2013 Vol.8 No.8, Page(s):106-112.

[4] Rajeshwar Dass, Priyanka, Swapna Devi, "Image Segmentation Techniques", International Journal of Electronics and Communication Technology, Vol. 3, Issue 1, Jan.-March 2012.

[5] A. Rocha, C. Hauagge, J. Wainer, and D. Siome, "Automatic fruit and vegetable classification from images," Computers and Electronics in Agriculture, Elsevier; vol. 70, 2010. Page(s): 96-104.

[6] Haiguang Wang, Guanlin Li, Zhanhong Ma, Xiaolong Li, " Application of Neural Networks to Image Recognition of Plant Diseases", 2012 International Conference on Systems and Information, 2012 IEEE, Page(s) : 2159-2164.

[7] Manisha A. Bhange, Prof. H. A. Hingoliwala, "A Review of Image Processing for Pomegranate Disease Detection", International Journal of Computer Science and Information Technologies, Vol. 6 (1), 2015, Page(s): 92-94.

[8] Dheeb Al Bashish, Malik Braik, and Sulieman Bani-Ahmad, "Detection and classification Leaf Diseases using K-means based Segmentation and Neural-network based Classification", Information Technology Journal 10(2),2011,Page(s): 267-275.

[9] Xiaojing Niu, Shihui Guo, Meili Wang, Hongming Zhang, Xianqiang Chen, Dongjian He, "Image Segmentation Algorithm for Disease Detection of Wheat Leaves", International Conference in Advanced Mechatronics Systems, August 10-12 2014, Page(s): 270-273.

[10] Dheeb Al Bashish, Malik Braik, and Sulieman Bani-Ahmad, "A Framework for Detection and classification of Plant Leaf and Stem Diseases", 2010 International Conference on Signal and Image Processing, 2010 IEEE, Page(s): 113-118.

[11] Savita N. Ghaiwat, Parul Arora, "Detection and Classification of Plant Leaf Diseases Using Image processing Techniques: A Review," International Journal of Recent Advances in Engineering & Technology, Volume-2, Issue - 3, 2014.

[12] Monika Jhuria, Ashwani Kumar,Rushikesh Borse,"Image Processing For Smart Farming: Detection of Disease and Fruit Grading", IEEE, Proceedings of the 2013 IEEE SecondInternational Conference on Image Information Processing, 2013

[13] Yuan Tian, Chunjiang Zhao, Shenglian Lu and Xinyu Guo, "SVM-based Multiple Classifier Syatem for Recognition of Wheat Leaf Diseases," Proceedings of 2010 Conference on Dependable Computing, November 20-22, 2010, page(s): 189–193.

[14] O. Kleynen, V. Leemans, and M. F. Destain, "Development of a multi-spectral vision system for the detection of defects on apples," Journal of Food Engineering, vol. 69, 2005, Page(s) :41–49.

[15] S. R. Dubey and A. S. Jalal, "Species and Variety Detection of Fruits and Vegetables from Images", International Journal of Applied Pattern Recognition (IJAPR), vol. 1, no. 1, 2013, Page(s): 108 – 126.

[16] R. Gonzalez, R. Woods, Digital Image Processing, 3rd ed., Prentice-Hall, 2007.

[17] Bongani Malinga, Daniela Raicu, Jacob Furst, " Local vs. Global Histogram-Based Color Image Clustering"

[18] Greg Pass, Ramin Zabih and Justin Miller, "Comparing Images Using Color Coherence Vectors"

[19] T. Ojala, M. Pietikäinen, and T. T. Mäenpää, "Multiresolution gray-scale and rotation invariant texture classification with Local Binary Pattern," IEEE Trans. on Pattern Analysis and Machine Intelligence, vol. 24, no. 7, pp. 971-987, 2002.