공학석사 학위논문

Feature extraction using graph Laplacian for LCD panel defect classification

LCD 패널 상의 불량 분류를 위한

Laplacian 그래프를 이용한 특성 추출 방법

2012년 8월

서울대학교 대학원

컴퓨터공학부

김 규 동

Feature extraction

using graph Laplacian

for LCD panel defect classification

LCD 패널 상의 불량 분류를 위한

Laplacian 그래프를 이용한 특성 추출 방법

지도 교수 유 석 인

이 논문을 공학석사 학위논문으로 제출함

2012년 8월

서울대학교 대학원

컴퓨터공학부

김 규 동

김 규 동의 석사 학위논문을 인준함

2012년 8월

위 원 장	**신 영 길**	(인)
부위원장	**유 석 인**	(인)
위 원	**이 제 희**	(인)

Abstract

Feature extraction

using graph Laplacian

for LCD panel defect classification

Gyudong Kim

School of Computer Science

The Graduate School

Seoul National University

There are four types of defects on LCD panel. For exact classification for the defects, good feature selection and classifier are necessary. In this paper, various features such as brightness, shape and statistical features are stated and Bayes classifier using Gaussian mixture model is used as classifier. But noisy or irrelevant features can harass the classification result. Feature extraction method can reduce the influence of irrelevant features and dimensionality to analyze complicated data well. Principal Component Analysis was one of the most famous feature extraction method had appropriate performance. However PCA would produce poor result if many noisy features exist. To solve that problem of PCA, feature extraction method based on spectral graph theory is proposed. Unlike PCA, proposed method using graph Laplacian matrix based on similarity instead of covariance matrix for analyzing spectral system. Experimental result shows that feature extraction method using graph Laplacian produces better performance than the result using PCA. And also proposed method is very robust to randomly added noisy features.

Keywords: Defect classification, Gaussian mixture model, Spectral graph theory, Graph Laplacian, Principal component analysis

Student Number: 2010-23244

Contents

Chapter 1. Introduction 1

Chapter 2. Defects and features 4

2.1. Definition 4

2.2. Brightness features 6

2.3. Shape features 6

2.4. Statistical features 9

Chapter 3. Feature extraction 11

3.1. Principal component analysis 11

3.2. Graph Laplacian 12

Chapter 4. Gaussian mixture models 14

4.1. Training 14

4.2. Classification 17

Chapter 5. Experiment 18

Chapter 6. Conclusion 21

Reference

List of Tables

Table 1. Result of experiment 1 19

Table 2. Result of experiment 2 19

List of Figures

Figure 1. Examples of the defects on LCD panels 5

Figure 2. Shape of defect and convex hull 8

Figure 3. How to get the co-occurrence matrix from the image 10

Figure 4. Comparison of feature extraction result using PCA and graph Laplacian 13

Figure 5. Probability distribution estimation with single Gaussian model and Gaussian mixture model 14

Chapter 1. Introduction

Defect classification is the one case of object classification, and classification is the one of main research topic of computer vision area. Object classification methods have researched for a long time. In recent, BoF(Bag-of-Feature) methods for representing image is widely used for object classification problem. The defects also can be translated to many features and classified as objects.

Nowadays demands for LCD panel is increasing, but production process of those panels is delicate and hard. Even small dust or a scar can influence viewing performance of LCD largely. Therefore, considerate and fine detection and treatment for those substances, for example dust, scar or defect, is needed. But even if the defects were founded, all of those things should not be treated equally. Some defects must be repaired because it can cause serious problem, but other defects can be removed by just washing with water or something. Thus it is necessary to classify the defects exactly for reducing production cost and boosting process efficiency.

Bayesian classification is one of the most famous classification methods. Bayesian theory gives a mathematical calculus of degrees of belief. It was reviewed and described in [1], [9] and [10]. Also [1] applied the theory to varying unsupervised classification and made the theory general. Many classification methods using mixture models were also proposed and researched for a long time. [2] stated finite mixture analysis is definitively a powerful framework for model-based cluster analysis. There are some mixture models using to classify object, and Gaussian mixture model is most frequently used.

In classification problem with BoF methods, feature selection and extraction is important issue. If there are irrelevant or noisy feature, classification result can be poor. Principal component analysis(PCA) is the approach that can dealing with both feature extraction and dimensionality reduction. PCA is one of the spectral algorithms that analyze the spectrum like eigenvector and eigenvalue. It is known that PCA has property which makes correlated original data to uncorrelated data. But PCA was not robust to many noisy or irrelevant features in data. Therefore feature extraction method based on spectral graph theory is proposed.

Spectral methods could be used for different purposes like clustering and feature selection. Spectral clustering is frequently researched work in recent. [3] stated similarity of spectral clustering and PCA. Spectral clustering used graph Laplacian to analyze the data structure. Recently spectral method using graph Laplacian for feature selection was also proposed in [4], [11]. Despite PCA analyzed eigen-system of data covariance matrix, proposed method use graph Laplacian matrix based on similarity. As calculating with similarity measure, eigen-system of graph Laplacian preserves data structure well. By those good conserving properties, proposed method can be more robust to noisy or irrelevant features than feature extraction using PCA.

This paper is organized as follows. Section 2 denotes defect and feature definition. Section 3 describes feature extraction using graph Laplacian and Section4 explains Gaussian mixture models for classification. Section 5 presents experiment method and result. This paper ends up with a conclusion in Section 6.

Chapter 2. Defects and features

2.1. Definition

There are four categories of defects on LCD. Each defect is called peel, open, substance and remain. As seen in Figure 1, defects have some properties. But some defects in other category can be too similar to classify with human eyes. Therefore, BoF method is needed to represent those defects.

For describing defects with features, definition of feature is needed. Classifying defects with BoF was researched in many area, and [5], [6] denote some useful features. In this paper, many features are used including brightness features, shape features and statistical features.

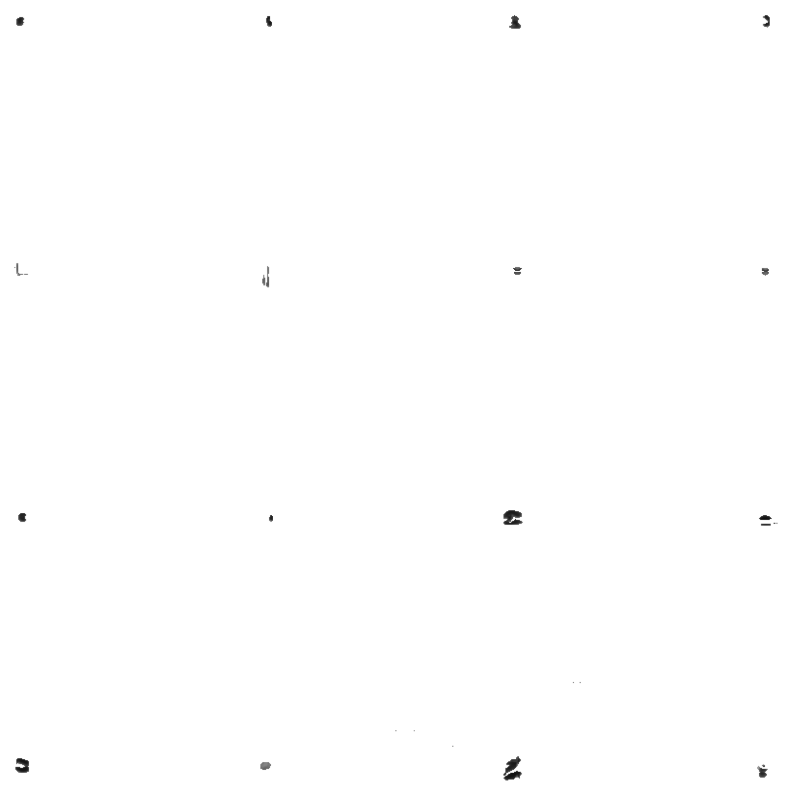

Figure 1. Examples of the defects on LCD panels. Each row represents peel, open, substance, remain defects respectively in descending order.

2.2. Brightness features

Each defect can have different characteristics. For example, some open defects could be brighter averagely than some peel defects. And some substance defects would have larger variance of intensity compared with some remain defects. Therefore, brightness features can be important for classifying the defects. Also those brightness features are easily observed. In this paper, five brightness features are used for classification. Those are like this:

Intensities average of defect area,

Intensities standard deviation of defect area,

Intensities median value of defect area,

Intensities minimum value of defect area,

Intensities maximum value of defect area.

2.3. Shape features

Appearance of defects can give some clues for discriminating the types of defects. For instance, peel defects are more similar to circular shape than open defects generally. Various shape features can be acquired when using contour and convex hull of defect area. Contour means the exterior or the surface of the object. In Figure 2, C is the contour of the defect. Convex hull is a convex shaped surface or contour perfectly covering the defect which has minimum length. Convex hull of the defect of Figure 2 is C_{con}. When a contour is more and more convex, a length of convex hull will be more similar to a length of contour itself.

Having a contour C and convex hull C_{con}, following properties can be calculated:

P : Length of contour
S : Size of contour
P_{con} : Length of convex hull
S_{con} : Size of convex hull

And these properties are used for extracting more valuable features to classify the defects:

$$\begin{aligned}
&\text{Length: } L = \frac{P}{2} \\
&\text{Breadth: } B = S \\
&\text{Compactness: } Com = \frac{P^2}{4\pi S} \\
&\text{Roughness: } R = \frac{P}{P_{con}} \\
&\text{Area Ratio: } S_r = \frac{S}{S_{con}}
\end{aligned} \tag{1}$$

If shape of the defect is more irregular or bumpy, original contour and convex hull of the defect will be more different. Then length of convex hull become shorter and breadth of convex hull will be larger with respect to original contour. Therefore roughness shall be increased and area ratio will be decreased. And compactness is also good feature to distinguish defects. If the defect has perfectly circular shape, then compactness of the defect will be 1. But in case of the defect have rectangular shape which is not circular shape, the compactness of the defect will be larger than 1. Therefore, we can find more circular-shaped defects which can be said more compact defects. Through these analyses of shape features, the defects can be classified appropriately.

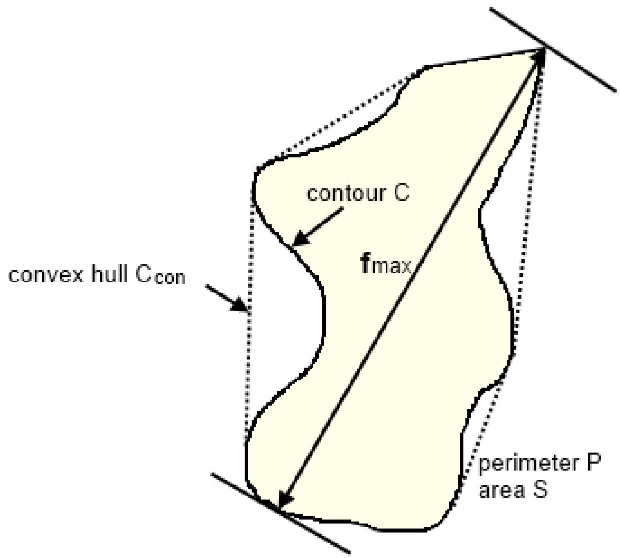

Figure 2. Shape of defect and convex hull. f_{max} is the largest diameter of defect.

2.4. Statistical features

In [5], it was stated that the grey level co-occurrence matrix was a well-known statistical tool for extracting second-order texture information from images. The co-occurrence matrix P_d counts changes of intensities and records those as matrix's elements. Let I is the grayscale original image and P_d is square matrix defined on a given displacement vector $\boldsymbol{d}$ = (dx,dy). And I_p is the intensity of the pixel p=(i,j) in image I. The entry (i,j) of the matrix P_d is the number of occurrence of the pair of gray level i and j which are apart from each other $\boldsymbol{d}$ away. Then the elements of co-occurrence matrix P_d with displacement vector $\boldsymbol{d}$ are calculated as follows.

$$P_d(i,j) = \left\| \{ p : (i,j) = (I_p, I_{p+d}),\ for\ all\ \ p\ in\ the\ image\ I \} \right\| \quad (2)$$

An example is given in Figure 3 to demonstrate how to get the co-occurrence matrix of right from gray level image of left with respect to the displacement vector $\boldsymbol{d}$ = (1,1). A symmetric co-occurrence matrix P can be calculated by using the co-occurrence matrix P_d and its transpose matrix P_{-d}.

$$P = P_d + P_{-d} \quad (3)$$

Though many features can be computed from the co-occurrence matrix P, only five features are used in this paper.

$$\text{Energy: } \sum_i \sum_j P^2(i,j)$$

$$\text{Entropy: } \sum_i \sum_j P(i,j) \log_2[P(i,j)]$$

$$\text{Contrast: } \sum_i \sum_j (i, j)^2 P(i, j) \tag{4}$$

$$\text{Homogeneity: } \sum_i \sum_j \frac{P(i, j)}{1 + |i - j|}$$

$$\text{Correlation: } \sum_i \sum_j \frac{(i - \mu_x)(j - \mu_y) P(i, j)}{\sigma_x \sigma_y}$$

where μ_x, μ_y, σ_x and σ_y are the means and the standard deviations corresponding to vectors p_x, p_y that are calculated as

$$\begin{aligned} p_x &= \sum_j P(i, j) \\ p_y &= \sum_i P(i, j) \end{aligned} \tag{5}$$

Since symmetric co-occurrence matrix P_d is used in this paper, μ_x equals to μ_y and also σ_x is same with σ_y.

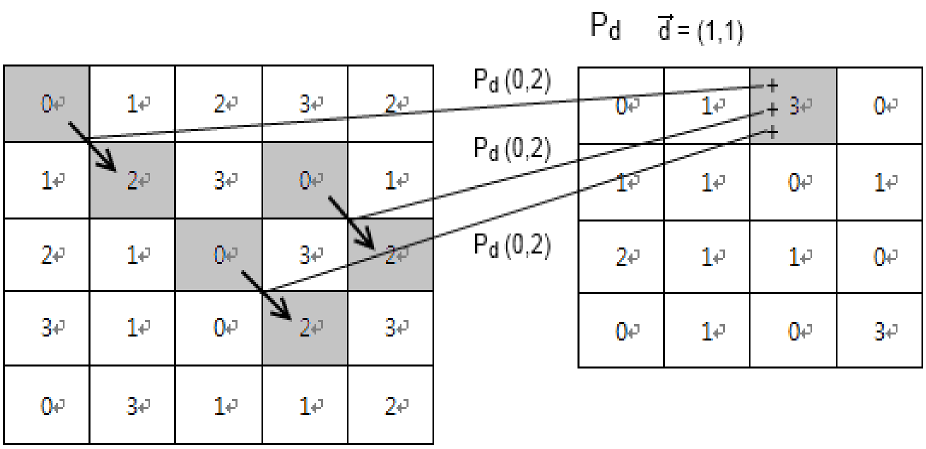

Figure 3. How to get the co-occurrence matrix from the image.

Chapter 3. Feature extraction

To achieve good classification result, feature must be selected considerately. Noisy feature can harass the learning and classification process badly. And if there are too many features, computational time for learning can be too exhaustive. Principal Component Analysis was the one of those feature extraction approaches which made correlated original data to uncorrelated one and reduce dimensionality. But sometimes PCA were not robust to noisy features. Therefore, spectral feature extraction method using graph Laplacian is proposed in this section.

3.1. Principal component analysis

PCA was spectral analysis method that analyze spectrum of data. General PCA was conventional one which uses eigenvector and eigenvalue of data's covariance matrix. PCA extracts a subset of new features from the data set by means of functional mapping. It has a property that maximizes the variance of extracted features. That characteristic was known as effective for many cases. But that property was not good in some cases such as a situation that noisy or relevant features existed. In this paper, spectral graph theory is used for dealing with that problem.

3.2. Graph Laplacian

Using graph Laplacian's eigensystem, the structure of original data can be preserved. Figure 4 describe difference of covariance matrix's eigenvector and graph Laplacian's eigenvector. Despite the projection using PCA's largest eigenvector do not divide two class well in Figure 4 (b), graph Laplacian's two largest eigenvectors are doing well relatively as seen as Figure 4 (c) and (d). Therefore, proposed method use graph Laplacian matrix's eigenvector as feature instead of original feature's projection to covariance matrix's eigenvector in PCA.

To make graph Laplacian matrix, similarity or adjacency matrix is needed. Similarities can be measured by various function, but in this paper RBF function is used:

$$S_{ij} = e^{-\alpha \cdot \left\| x^i - x^j \right\|^2} \tag{6}$$

where x_i means ith sample in data set, and α is parameter for adjusting magnitude and distance of x_i and x_j is calculated as Euclidean distance. When similarity matrix is acquired, symmetric graph Laplacian matrix can be calculated:

$$D = diag(S\mathbf{1}) \tag{7}$$

$$\mathbf{1} = (1,...,1)^T$$

$$L_{sym} = D^{-\frac{1}{2}} S D^{-\frac{1}{2}} \tag{8}$$

Finally, normalized graph Laplacian matrix can be acquired:

$$L_{rw} = D^{-1} L_{sym} \tag{9}$$

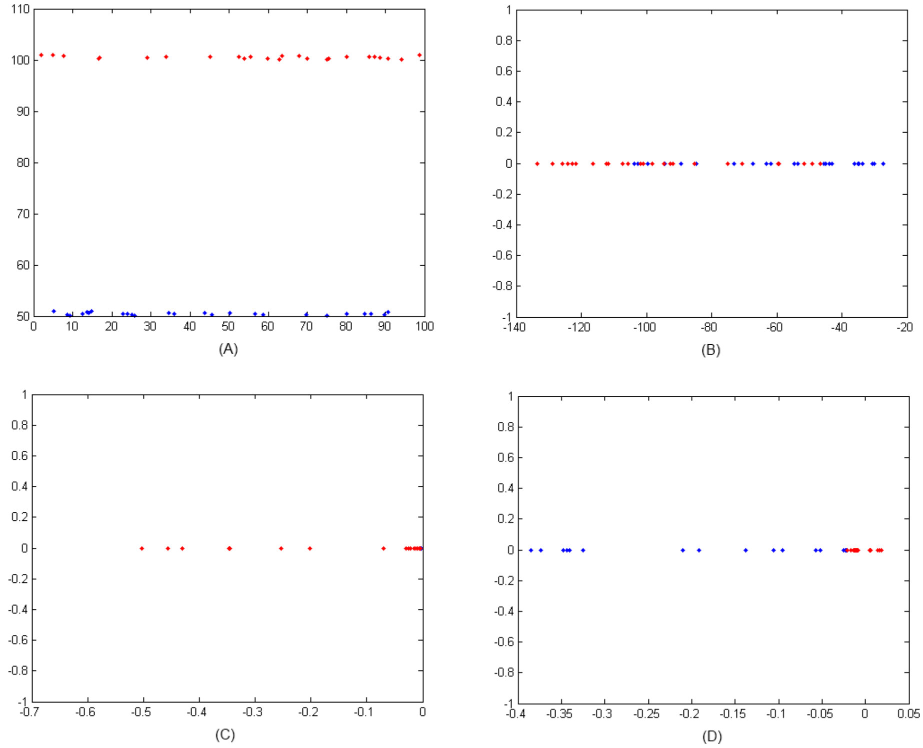

Figure 4. Comparison of feature extraction result using PCA and graph Laplacian

Chapter 4. Gaussian mixture models

3.2. Training

Bayes classifier using Gaussian mixture model(GMM) is composed of multiple Gaussian component to approximate probability model of data. If predicting probability distribution of data with single Gaussian model, it can be poor for the complicate data existed in real case. But using multiple component, an estimate of data will be more precise. These examples are presented at Figure 5. As seen in Figure 5, right estimation result using GMM is better than left one using single Gaussian model.

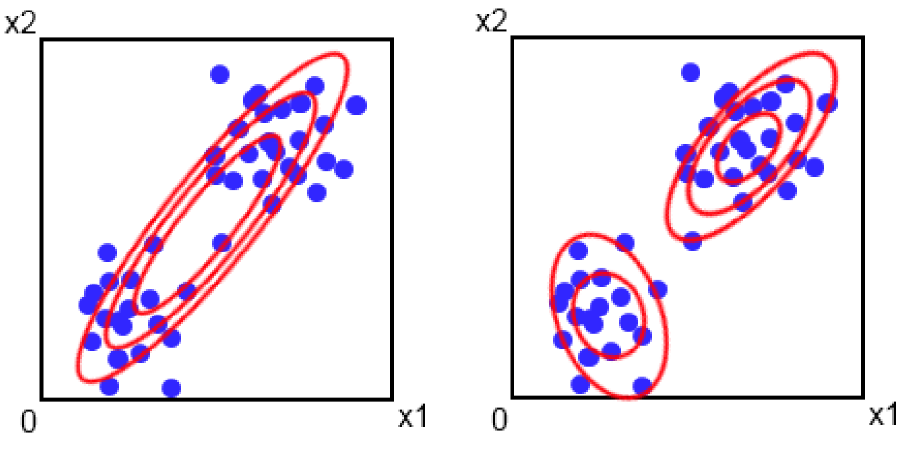

Figure 5. Probability distribution estimation with single Gaussian model and

Gaussian mixture model.

GMM is expressed as linear superposition of K Gaussian distribution.

$$p(\mathbf{x}) = \sum_{k=1}^{K} \pi_k N(\mathbf{x} \mid \mu_k, \Sigma_k) \tag{10}$$

where π_k is the mixing coefficient, and sum of π_k is 1.

In GMM, latent variable z exists and it has properties as follows.

$$\begin{aligned} p(z_k = 1) &= \pi_k \\ p(\mathbf{z}) &= \prod_{k=1}^{K} \pi_k^{z_k} \end{aligned} \tag{11}$$

And z_k correspond with x_k. Using z, responsibility can be acquired.

$$\begin{aligned} \gamma(z_k) &\equiv p(z_k = 1 \mid \mathbf{x}) \\ &= \frac{p(z_k = 1) p(\mathbf{x} \mid z_k = 1)}{\sum_{j=1}^{K} p(z_j = 1) p(\mathbf{x} \mid z_j = 1)} \\ &= \frac{\pi_k N(\mathbf{x} \mid \mu_k, \Sigma_k)}{\sum_{j=1}^{K} \pi_j N(\mathbf{x} \mid \mu_j, \Sigma_j)} \end{aligned} \tag{12}$$

If regarding π_k as a prior probability of $z_k = 1$, then $\gamma(z_k)$ is a posterior probability of observing corresponding **x**. In other word, responsibility can be thought as a degree of probability that a data can be included at kth component.

Estimation of GMM is performed by EM algorithm which maximizes likelihood function and updates parameters repeatedly. Process is executed as follows.

1) Initialize average vector μ, covariance vector Σ, mixing coefficient vector π, and calculate initial value of log likelihood function.

2) Expectation stage

   Using present parameters, calculate responsibility $\gamma(z_k)$ by equation (12).

3) Maximization stage

   Using present responsibility values, re-compute parameters.

$$\mu_k^{new} = \frac{1}{N_k}\sum_{n=1}^{N}\gamma(z_{nk})\mathbf{x}_n$$

$$\Sigma_k^{new} = \frac{1}{N_k}\sum_{n=1}^{N}\gamma(z_{nk})(\mathbf{x}_n - \mu_k^{new})(\mathbf{x}_n - \mu_k^{new})^{\mathrm{T}}$$

$$\pi_k^{new} = \frac{N_k}{N} \tag{13}$$

$$\text{where} \quad N_k = \sum_{n=1}^{N}\gamma(z_{nk})$$

4) Calculate log likelihood function.

$$\ln p(\mathbf{X}\,|\,\pi,\mu,\Sigma) = \sum_{n=1}^{N}\ln\{\sum_{k=1}^{K}\pi_k N(\mathbf{x}_n\,|\,\mu_k,\Sigma_k)\} \tag{14}$$

5) If convergence conditions of log likelihood function or parameters are not satisfied, then go back to step 2.

It is noted that log likelihood function's value always increased when executing E stage and M stage successively. Therefore, log likelihood function can always reach to local maxima.

3.2. Classification

After training of Gaussian mixture models for each four defect categories, classification can be performed. Process of classification is simple that the defect category which has more similar properties with input defect will be chosen. Measure of similarity is calculated by log likelihood function of each category's Gaussian mixture model same as (14). Now the category which produces the highest log likelihood value with input is selected as input's category.

Chapter 5. Experiment

Data set using for experiment have 483 defect samples including 98 peel samples, 107 open samples, 179 substance samples and 99 remain samples. Each sample is composed of features described in above sections. 5-fold cross validation is used for training and test of experiment. Full process of experiment is like this: a) Extract features from the samples, and make BoF sets. Extraction of features means measuring length or breadth and calculating average or standard deviation of intensities from the defects. b) Acquire new features using conventional PCA or graph Laplacian. Covariance matrix for PCA or graph Laplacian matrix L is calculated using entire data set including training and test data and will be divided into training and test data again. c) Classify the data using Gaussian mixture model(GMM) with cross validation method. The number of components in GMM is fixed to 8.

All experiment is performed by MATLAB and top 6 eigenvectors are used as new projection axis of PCA and as new features in proposed method. Experiment 1 is executed to compare the feature extraction method. Classification result of experiment 1 is described in table 1.

Accuracy(%)

Feature extraction / Defect type	None	PCA	Graph Laplacian
Peel	94.90	96.94	96.94
Open	80.19	98.11	99.06
Substance	96.05	99.44	99.44
Remain	87.76	88.78	100.00
Total	90.61	96.45	98.96

Table 1. Results of experiment 1.

Accuracy(%)

Feature extraction / Defect type	None	PCA	Graph Laplacian
Peel	86.74	88.78	96.94
Open	66.98	91.51	99.06
Substance	98.31	98.87	99.44
Remain	79.59	85.71	100.00
Total	85.18	92.48	98.96

Table 2. Results of experiment 2.

Table 1 shows that the positive classification rate using feature extraction method is higher than the one using nothing. And performance of the classification using graph Laplacian as feature extraction method is better than the performance of the one using PCA as feature extraction method. In detail, open and remain defects are enhanced very largely with proposed method.

To know about consistency of each feature extraction method to noisy features, Experiment 2 is performed. In this case, 5 random generated noisy features are added to original BoF sets. Except features, same condition with Experiment 1 is used for Experiment 2. The result of Experiment 2 is presented at Table 2.

As expected, classification rate of no feature extraction method and PCA-based feature extraction method decreased comparing with Experiment 1. But it is noted that the result of proposed method in Experiment 1 and in Experiment 2 are equal. It means that proposed feature extraction method is consistent with noisy features with respect to another method.

Chapter 6. Conclusion

In this paper, various features and two feature extraction methods are introduced for LCD panel defect classification. As defects have diverse shape, size and intensity, it is necessary to consider from many aspects. Therefore many features including brightness features, shape features and statistical features are used for constructing bag-of-feature set of defects. Also for dealing with irrelevant features and dimensionality reduction, feature extraction method like principal component analysis is presented. But PCA feature extraction method is weak to some kind of noisy features, feature extraction using graph Laplacian is proposed in this paper.

Experimental results show that classification performance with feature extraction method is superior to the one without feature extraction method, and also proposed method has better result than PCA method. Moreover, it is notable that proposed method is very robust to noisy features comparing with PCA feature extraction method.

Classification experiment is executed with Bayes classifier using Gaussian mixture models in this paper. But features extracted by proposed method also easily can be used by another classifiers like support vector machine or artificial neural network.

Reference

[1] R. Hanson, J. Stutz, and P. Cheeseman. "Bayesian classification theory". Technical Report FIA-90-12-7-01, NASA Ames Research Center, 1990

[2] G. Celeus, "Mixture models for classification". In: DeckerR., et al., editors. Advances in Data Analysis. Berlin: Springer; 2007

[3] Xiang T. S. Gong, "Spectral clustering with eigenvector selection", Pattern Recognition, vol. 41, no 3, pp. 1012-1029, 2008

[4] Y. Jiang, J. Ren, "Eigenvalue sensitive feature selection", ICML, 2011

[5] X. Zhang, C. Krewet, B. Kuhlenkötter, "Automatic classification of defects on the product surface in grinding and polishing", Int J Mach Tools Manuf, 46 (1), pp. 59–69, 2006.

[6] P. Cavalin, L. S. Oliveira, A. L. Koerich and A. S. Britto, Jr. "Wood defect detection using grayscale images and an optimized feature set", Proc. IEEE Ind. Electron, pp.3408, IECON, 2006.

[7] R. M. Haralick, K. Shanmugam, I. Dinstein. "Textural Features for Image Classification". IEEE Transactions on Systems, Man, and Cybernetics SMC-3 (6): 610–621, 1973.

[8] C. M. Bishop, *PATTERN RECOGNITION AND MACHINE LEARNING*, Springer, 2006

[9] H. Permuter, J. Francos, I. Jermyn, "A study of Gaussian mixture models of colour and texture features for image classification and sementation", Pattern

Recognition 39(4) 695-706, 2006

[10] O. Boiman, E. Shechtman, M. Irani, "In defense of nearest-neighbor based image classification", CVPR, 2008

[11] L. Zelnik-Manor, P. Perona, "Self-tuning spectral clustering", Advances in Neural Information Processing Systems 17, 1601-1608, MIT Press, Cambridge, MA, 2005

국문 초록

LCD 패널 상의 불량 분류를 위한 Laplacian 그래프를 이용한 특성 추출 방법

서울대학교 대학원

컴퓨터공학부

김 규 동

LCD 패널 위에 존재하는 불량은 크게 4가지 종류로 나눌 수 있다. 각 불량은 서로 다른 방식으로 다뤄야 하기 때문에 정확한 분류방법이 필요하다. 정확한 분류를 위해서는 불량은 잘 나타내는 특성들과 좋은 분류기를 사용해야 한다. 이 논문에서는 불량을 표현하기 위해 불량 영역의 밝기, 모양, 통계적인 특성들을 사용하고, 가우시안 혼합 모델을 이용한 베이즈 분류방법을 사용하여 불량을 분류하게 된다. 하지만 불량을 나타내는 특성 중에 노이즈가 존재하거나 분류에 관련이 없는 특성들이 많이 존재하는 경우에는 분류 결과가 안 좋게 나올 수 있

다. 따라서 여기서는 특성 추출방법을 사용하게 된다. 특성 추출 방법은 연관성 이 적은 특성들이 분류에 미치는 영향을 줄여줄 뿐만 아니라 데이터의 차원을 줄여주어 분석을 용이하게 해주면서 계산 속도를 향상시키는 효과를 낼 수 있다. 주요 성분 분석 방법은 이러한 특성 추출 방법의 중 가장 유명한 방법중의 하나로 좋은 성능을 낸다고 알려져 있다. 하지만 주요 성분 분석 방법 역시 많은 수의 의미 없는 특성들이 존재하는 경우에는 나쁜 성능을 보여준다. 이러한 문제를 해결하기 위해 이 논문에서는 스펙트럴 그래프 이론을 이용한 특성 추출 방법을 제안하였다. 주요 성분 분석 방법이 데이터의 공분산 행렬을 사용하는 것과는 달리, 제안된 방법은 샘플 간의 유사도를 기반으로 한 그래프 라플라시안 매트릭스를 생성하여 그 고유 값과 고유 벡터를 사용하는 방법이다. 실험 결과를 보면 알 수 있듯이 그래프 라플라시안을 이용한 특성 추출 방법은 주요 성분 분석을 사용한 경우보다 더 좋은 분류 성공률을 보여준다. 또한 제안한 방법이 임의로 의미 없는 특성이 추가된 경우의 실험에 대해서도 매우 꾸준한 성능을 보여줌을 알 수 있다.

Keywords: 불량 분류, 가우시안 혼합 모델, 스펙트럴 그래프 이론, 그래프 라플라시안, 주요 성분 분석

Student Number: 2010-23244