

“© 2010 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works.”

Image Segmentation for Surface Material-type Classification using 3D Geometry Information

Andrew Wing Keung To, Gavin Paul and Dikai Liu

ARC Centre of Excellence for Autonomous Systems

University of Technology

Sydney NSW, Australia

{awto, gpaul, dkliu}@eng.uts.edu.au

Abstract—This paper describes a novel approach for the segmentation of complex images to determine candidates for accurate material-type classification. The proposed approach identifies classification candidates based on image quality calculated from viewing distance and angle information. The required viewing distance and angle information is extracted from 3D fused images constructed from laser range data and image data. This approach sees application in material-type classification of images captured with varying degrees of image quality attributed to geometric uncertainty of the environment typical for autonomous robotic exploration. The proposed segmentation approach is demonstrated on an autonomous bridge maintenance system and validated using gray level co-occurrence matrix (GLCM) features combined with a naive Bayes classifier. Experimental results demonstrate the effects of viewing distance and angle on classification accuracy and the benefits of segmenting images using 3D geometry information to identify candidates for accurate material-type classification.

Index Terms—perspective projection, material-type classification, image classification, gray level co-occurrence matrix, naive Bayes classifier

I. INTRODUCTION

Steel bridges are subjected to deterioration due to rusting, which reduces the effective lifespan of bridges. To prolong the lifespan of a steel bridge, maintenance by means of rust removal via grit-blasting and repainting must be performed regularly. However, the grit-blasting is hazardous towards humans due to issues including possible injuries caused by high pressurized mixture of air and grit and prolong exposure to lead and asbestos contaminated environments. An autonomous robotic system provides a solution to these issues and can eliminate human exposure to hazardous environments, and improve safety and efficiency of bridge maintenance.

In order for an autonomous robotic system to conduct grit-blasting in complex bridge environments, this system requires real-time environment knowledge. This robotic system should have the capacity of building environmental awareness, including the geometry of the environment, the surface types and conditions of steel structural members, through exploration and mapping, surfaces identification and material-type classification. Camera images and laser

range sensor data describing the environment can be gathered during robot manipulator-based exploration [1][2]. In exploration sensors mounted to the end-effector of the robot arm are manoeuvred within the complex environment to gather data from multiple view-points.

A laser range finder is used to build an environment's geometric map and for resolution limited classification of surface material-types [3]. Therefore, material-type classification approaches typically use cameras to provide higher resolution vision-based data. Vision-based classification has been investigated extensively through different visual signal processing and statistical modeling techniques including fractal analysis [4], wavelet transform [5], gray level co-occurrence matrix [6] and Markov Random Fields [7] for the extraction of material type characteristics. Feature learning algorithms applied towards feature-based classification includes Support Vector Machine, K-nearest neighbor and naive Bayes classifier. However, experiments from the discussed literatures assumed the collected images are of certain quality standard; this is impractical for an autonomous robotic system collecting image data in a complex environment. Therefore image quality inconsistencies due to data collection variations including viewing distance and angle needs to be addressed during classification in order to improve accuracy.

This paper presents an approach to evaluating images of a complex environment to identify candidates for material-type classification. Images of an environment are fused with the 3D geometric information of the environment and used in image quality calculation for candidate segment selection. Candidate selection is integrated with a texture-based material-type classifier for surface type classification. The breakdown of this paper is as follows: Section II discusses 3D geometry fused image generation, candidate selection and material type classifier. Section III presents experimental results. Section IV presents the conclusion and future work.

II. METHODOLOGIES

A. Overview

The process for classification candidates selection is shown in Fig.1. Camera image data and laser range data are collected and fused together through perspective projection to generate a 3D geometry fused image (i.e. additional geometric information for each image pixel corresponding to the surface represented by the pixel relative to the camera origin.). The fused image is then used to calculate image qualities required for the selection of candidate segments for classification. Finally, the identified candidate segments are classified utilising GLCM features combined in a naive Bayes classifier.

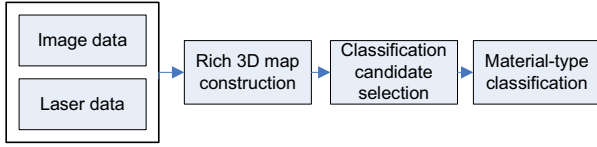


Fig. 1. Overview of the proposed segmentation process

B. 3D Geometry Fused Image Generation

The perspective projection technique presented in [8][9] is applied to match a 2D images as shown in Fig.2a with 3D mesh surfaces generated by the laser range finder readings [10] to construct a 3D geometry fused image. The process of perspective projection involves locating the image plane in global 3D space. The size and orientation of the image plane is determined by the homogenous transform matrix of the camera's origin, the camera lens field of view and the distance of the furthest point in the image to the camera's origin. Ray casting from the camera's origin to image plane is performed to correlate relevant mesh points in the 3D map to pixel points on the 2D image as illustrated in Fig.3b. A 3D geometry fused image is used for the calculation of viewing distance and angle of surface to determine image quality.

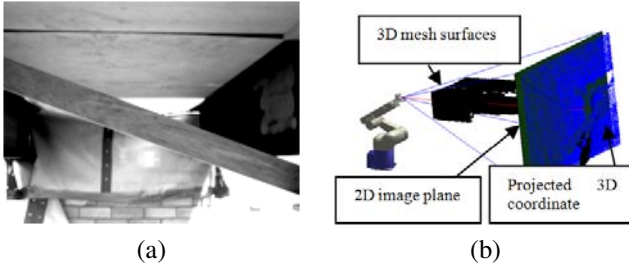


Fig. 2. a) 2D image of an environment; b) 3D perspective projection of laser scan data onto the image plane

C. Accurate Classification Candidate Selection

A decision-making model based upon image quality is utilised to select candidate segment for accurate classification. Image quality in this model is defined as the aggregated consideration of image quality factors including; focal quality, spatial resolution and distortion. The satisfaction of these image quality factors will determine the likelihood of correct classification for an image segment. These image quality factors are further discussed in detail below.

Focal quality describes the measurement of texture acutance level relative to the plane of focus and depth of field d . As shown in the Fig.3, the plane of focus describes the plane parallel to the image plane with ideal image focus quality (i.e. maximum acutance level permitted by the lens quality and camera resolution). Depth of field describes the region beyond and prior to the plane of focus where the acutance level loss is considered negligible. To calculate the depth of field region range D_n to D_f using [11], the following parameters are defined; focal length F , lens f-number f , circle of confusion c and the image plane to lens distance s ,

$$D_n = \frac{Fs^2}{F^2 + fc(s - F)} \quad (1)$$

$$D_f = \frac{Fs^2}{F^2 - fc(s - F)} \quad (2)$$

The focal quality component of the decision model is satisfied when the surface to camera origin distance d_o is within the calculated depth of field region D_n to D_f ,

$$D_f \leq d_o \leq D_n \quad (3)$$

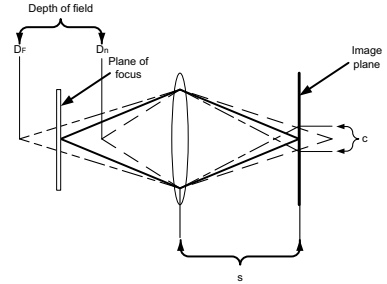


Fig. 3. Plane of focus and depth of field diagram

Spatial resolution considers the pixel density of an image segment i.e. pixels per inch. The calculation of spatial resolution for surface segments is determined by the image resolution and the viewing distance of the surface. To satisfy the spatial resolution quality consideration in the decision model, the viewing distance for an object must be less than or equal to the viewing distance threshold T ,

$$d_o \leq T \quad (4)$$

Spatial resolution conformance between the training data and the test data is implemented in the decision model to address scale intolerant material-type classification methodologies. Viewing distance range R_{lower} and R_{upper} are defined to restrict the spatial resolution variation between the training and the testing data sets. To satisfy the spatial resolution conformance in the decision model, the viewing distance for an object must be within the defined viewing range.

$$R_{lower} \leq d_o \leq R_{upper} \quad (5)$$

Surface distortion considers the variation in texture characteristics of a surface in an image due to viewing angle change. The range of incidence angle from θ_{min} to θ_{max} shown in Fig. 4a and Fig. 4b, specifies the viewing angle range for acceptable surface distortion. To satisfy the distortion, the camera's centre ray angle θ_s is required to be within the defined viewing angle range,

$$\theta_{min} \leq \theta_s \leq \theta_{max} \quad (6)$$

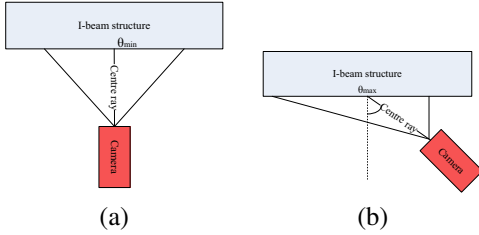


Fig. 4. Viewing angle a) Allowable minimum angle of incidence; b) Allowable maximum angle of incidence

D. Material-Type Classifier

Gray Level Co-occurrence Matrix (GLCM) is utilised for feature extraction in the material-type classifier. GLCM, as discussed in [12] is a texture analysis technique capable of analysing gray level intensity of overall or individual colour channels of an image, and thus is suitable for material feature extraction [7][13]. The following GLCM feature set is selected to model the environment materials: contrast, correlation, energy and homogeneity.

Contrast is defined as the sum of square variances where i and j are the respective row and column value of the matrix and p is the normalised co-occurrence value in each GLCM element. A contrast value of 0 indicates a uniform image,

$$\sum_{i,j} |i - j|^2 p(i, j) \quad (7)$$

Correlation measures the linear dependency of gray levels on those of neighbouring pixels where μ_i and μ_j are the GLCM column and row mean values and σ_i and σ_j are the GLCM column and row standard deviations. The correlation

value reflects the quality of the linear relationship, a value of 1 indicating a perfectly correlated image,

$$\sum_{i,j} \frac{(i - \mu_i)(j - \mu_j)p(i, j)}{\sigma_i \sigma_j} \quad (8)$$

Energy is calculated as the sum of squared co-occurrence values p . The energy value reflects the textural uniformity observed from the repetition of pairs of pixel in the image. An energy value of 1 indicates a constant image.

$$\sum_{i,j} p(i, j)^2 \quad (9)$$

Homogeneity is the inverse difference moment measuring the distribution closeness of elements in the GLCM to the matrix diagonal. The weight values decreases exponentially away from the diagonal of the matrix,

$$\sum_{i,j} \frac{p(i, j)}{1 + |i - j|^2} \quad (10)$$

Probability distribution functions of GLCM features are constructed and a naive Bayes classifier [14][15][16] is then employed to perform the material-type classification. Naive Bayes is a simple and popular approach to classification where it is reasonable to assume feature independence, complete material-type awareness and uniform material probability for all surface materials. Given a particular material model $M_k \in M$ where M represents the material models for the environment, the probability of a material is,

$$P(M_k|f) = \frac{P(f|M_k)P(M_k)}{P(f)} \quad (11)$$

where $f=(f_1, \dots, f_n)$ and n denotes the number of features in a material model. The denominator in (11) is invariant across all material models, therefore it can be consider as a normalisation parameter and hence can be rewritten as,

$$P(M_k|f) \propto P(f|M_k)P(M_k) \quad (12)$$

The outcome of each individual model feature value f_t is defined to be independent for all other feature outcome values. Thus the classifier can be expressed,

$$P(M_k|f) \propto \prod_{t=1}^n P(f_t|M_k)P(M_k) \quad (13)$$

where n denotes the number of features in a material model and $P(f_t|M_k)$ is the conditional probability describing the likelihood of the material feature f_t given the material model M_k . The class prior $P(M_k)$ is defined to be uniformly distribution across all material that exists in the environment.

The maximum likelihood rule is applied for the selection of the most probable hypothesis m ,

$$m = \operatorname{argmax}_m (P(m = M|f)) \quad (14)$$

As discussed, the candidate segments determined by the use of 3D geometry fused images and a image quality decision model are used in the proposed material-type classifier for accurate material type classification.

III. EXPERIMENTS

In order to verify the presented approach, the approach is integrated into an autonomous 3D map building system [2]. Three experiments are conducted in a real-world environment.

A. Environment Setup

The environment used in these experiments Fig.5b is modeled upon a real-world environment Fig.5a. A 6-DOF robotic manipulator is placed in a replica bridge environment Fig.6 a which consists of an I-beam channel (i.e. a bridge's under deck) constructed with materials including rusted I-beams, plywood, timber and steel plates. Data collection is performed by a multi-sensor attachment consisting of a 1.3MP c-mount camera with a 6mm fixed focal length lens and a Hokuyo laser range finder.



Fig. 5. a) A real bridge structure environment; b) Experiment setup

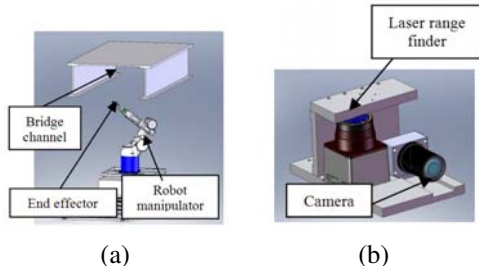


Fig. 6. a) A Robot manipulator within a bridge channel; b) Multi-sensor package

B. Experiment 1: Focal Quality and Spatial Resolution

Experiment 1 demonstrates the effects of viewing distance on material-type classification for each material type. For this experiment, the lens is calibrated with a plane of focus at viewing distance of 100mm and F-stop number of 4. The gathered image samples are shown in Fig. 7 containing different material-types with viewing distance ranging from 50mm to 300mm at 50mm increments. The 400×400 pixel image samples are segmented into 40×40 pixel sub-sample. Material-type classification is performed on the sub samples to determine the overall classification accuracy of the image sample. The training data set used in this experiment consists of 40×40 pixel material sub-samples acquired at a fixed viewing distance of 100mm and viewing angle of 90° .

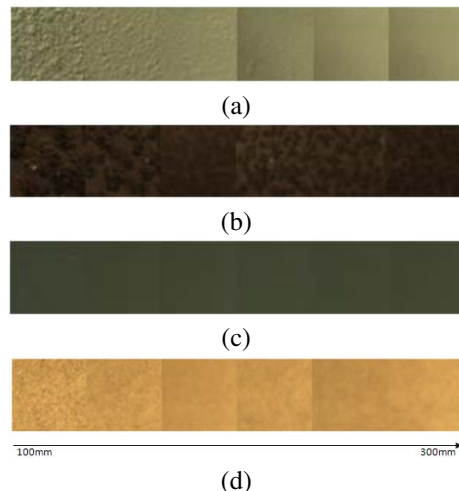


Fig. 7. Experimental training set at varying viewing distances a) blasted metal; b) rusted metal; c) steel; d) wood;

Results from this experiment are presented in Table I. The results indicate a trend of decreasing accuracy of classification as the viewing distance of the sample deviates from the training data set (from 100mm viewing distance). The highest accuracy result for blasted metal, rusted metal and wood was achieved at 100mm and 300mm for steel. Overall steel remained highly accurate in all viewing distance samples due to the low variance in texture characteristics throughout the tested distance range. Omitting the classification results of steel, rusted metal demonstrated the best performance in maintaining correct material classification throughout the range of viewing distance, while blasted metal and wood both displayed a rapid decrease in classification accuracy beyond the 100mm viewing distance.

TABLE I
RESULTS OF CLASSIFICATION WITH VARYING DISTANCES

Viewing Distance (mm)	Material classification accuracy (%)			
	Blasted metal	Rusted metal	Steel	Wood
50	3%	5%	83%	2%
100	96%	91%	79%	94%
150	1%	77%	95%	0%
200	0%	64%	96%	0%
250	0%	69%	90%	1%
300	0%	57%	98%	0%

C. Experiment 2: Viewing Angle

Experiment 2 demonstrates the effect of viewing angle variation on material-type classification. For this experiment, image collection is performed by manoeuvring the robot manipulator at 15° increments from 90° to 45° with the centre pixel of each image acquisition positioned to maintain

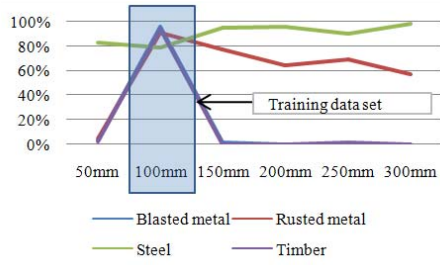


Fig. 8. Classification accuracy graph

a viewing distance to surface of 100mm. Lens configuration remains the same as experiment 1 with plane of focus at a viewing distance of 100mm and an iris F-stop number of 4. The image at each viewing distance is standardised into a 1280×960 pixel sample and segmented into 40×40 sub-sample. Training data set remains the same as experiment 1.

Experimental results are presented in Table II and Fig.10. Similar to Experiment 1, the classification accuracy of steel has not deteriorated relative to changing viewing angle due to the low contrast and sparse texture characteristics of steel. False classification relative to viewing angles is shown in Fig.9. The figure presents the sequence of viewing angle image captures for blasted metal and the associated visual classification results. Fig.9a shows a data sample captured in conformance with the training data set at viewing distance and angle of 100mm and 90° , respectively. The corresponding classification result illustrated in Fig.9b shows a scattered distribution of false classification with teal representing steel classification and other colours representing other material-types. Fig.9d, 9f and 9h indicates a progressive clustering pattern of false classification towards the left. The clustering pattern of false classification can be inferred by viewing the original images presented in Fig. 9c, 9e and 9g. At each progressive viewing angle, the deterioration of sharpness and image quality towards the left of the image becomes increasing evident.

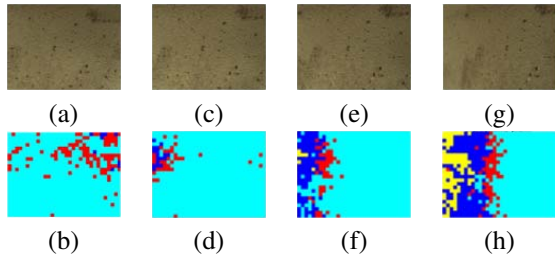


Fig. 9. Blasted metal a) 90° sample; b) 90° result; c) 75° sample; d) 75° result; e) 60° sample; f) 60° result;g) 45° sample; h) 45° result;

D. Experiment 3: Real World Environment

Experiment 3 is conducted in a real-world complex bridge environment. The camera is positioned at multiple viewing

TABLE II
CLASSIFICATION RESULTS WITH DIFFERENT VIEWING ANGLES

Viewing angle (degrees)	Material classification accuracy (%)			
	Blasted metal	Rusted metal	Steel	Wood
90	85%	91%	90%	41%
75	93%	85%	90%	68%
60	79%	80%	94%	29%
45	55%	55%	92%	32%

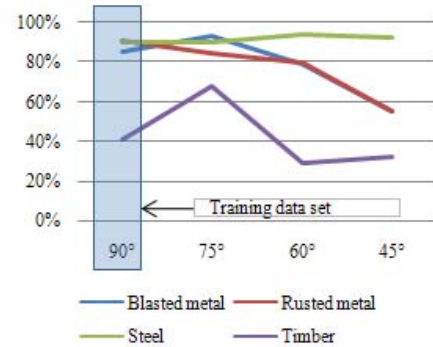


Fig. 10. Classification accuracy with different viewing angle

points for image acquisition of the cross member surface with a viewing distance range of 90-110mm and viewing angle range of $85 - 95^\circ$. Shown in Fig.11a and 11b are experiment sample images of the cross member surface. The lens configuration and training data set remains unchanged from previous experiments.

Table III shows the comparison of accuracy between the classification result of the overall image Fig.11a and Fig.11b, and the classification result of the candidate segments Fig.11c and Fig.11d. As shown in the table, the classification accuracy of the overall image is significantly lower than the classification accuracy of the candidate segmented images. Fig.11e and Fig.11f are the classification results showing the material-type labeled for each sub-sample.

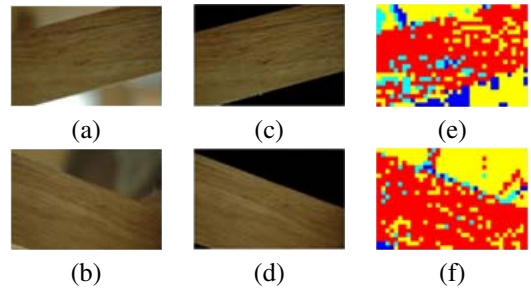


Fig. 11. a) sample 1; b) sample 2; c) selected segment in sample 1; d) selected segment in sample 2; e) sample 1 results; f) sample 2 results

TABLE III
MULTI-SURFACE CLASSIFICATION RESULTS

Test sample	Overall accuracy (%)	Selected surface accuracy (%)
Sample 1	57%	81%
Sample 2	63%	85%

E. Discussion

Results presented from the three experiments have demonstrated the effects of viewing distance and viewing angle on classification accuracy of material types by using the training data obtained from one viewing distance and angle data set. Experiments 1 and 2 justified the considerations of viewing distance and angle as the parameters in the image quality decision model, and the their effects on the classification accuracy. Experiment 3 demonstrated the real-world application of the approach in a complex bridge environment.

The experimental results also highlighted the limitation of the approach towards classification of texture poor material-types, which has been shown from the results that the viewing distance and the viewing angle do not have significant effect on the classification accuracy of timber and steel.

IV. CONCLUSION AND FUTURE WORK

The proposed approach to segmentation on a 3D geometry fused image for material type classification has fulfilled the objective of identifying classification candidates segments based on image quality. It has been shown that it is viable to evaluate image quality of 2D images using viewing distance and angle of surface segments from the 3D geometry information. Future work includes investigation into comparison of different classification methods, and the optimal way to incorporate the proposed approach in exploration of complex 3D environments to improve material awareness.

ACKNOWLEDGMENT

This work is supported by the Australian Research Council (ARC) Linkage Grant (ARC-LP0776312), the ARC Centre of Excellence for Autonomous Systems (CAS) (funded by funded by the ARC and the New South Wales State Government), the Roads and Traffic Authority (RTA) and the University of Technology, Sydney.

REFERENCES

[1] J. Xu, D. K. Liu, and G. Fang, "An efficient method for collision detection and distance queries in a robotic bridge maintenance system," vol. 362, pp. 71–82, 2007.

[2] G. Paul, D. Liu, N. Kirchner, and G. Dissanayake, "An effective exploration approach to simultaneous mapping and surface material-type identification of complex three-dimensional environments," *J. Field Robot.*, vol. 26, no. 11-12, pp. 915–933, 2009.

[3] N. Kirchner, T. Taha, D. K. Liu, and G. Paul, "Simultaneous material type classification and mapping data acquisition using a laser range finder," *Int. Conf. on Intelligence Technologies*, pp. 124–129, 2007.

[4] E. Momma, Y. Kimura, H. Ishii, T. Ono, M. Harada, T. Aoyama, and T. Higuchi, "Rust classification using image analysis of steel structures," *3rd ICST*, pp. 409–413, 2008.

[5] W. Gaohong, Z. Yujin, and L. Xinggang, "Wavelet transform-based texture classification with feature weighting," *ICIP Proceedings.*, vol. 4, pp. 435–439, 1999.

[6] A. P. Bradley, B. C. Lovell, and P. T. Jackway, "Scale-space texture analysis," 1995.

[7] C. Unsalan and A. Ercil, "Automated inspection of steel structures," *Recent Advances in Mechatronics, Springer-Verlag*, pp. 468–480, 1999.

[8] D. Dion Jr, "3d triangular mesh matching through a sequence of registered 2d and 3d images," p. 1977, 2000.

[9] L. Liu and I. Stamos, "Multiview geometry for texture mapping 2d images onto 3d range data," pp. 2293–2300, 2006.

[10] G. Paul, D. K. Liu, and N. Kirchner, "An algorithm for surface growing from laser scan generated point clouds," *Robotic Welding, Intelligence and Automation, Springer-Verlag*, vol. 362, pp. 481–491, 2007.

[11] A. P. Pentland, "A new sense for depth of field," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 9, no. 4, pp. 523–531, 1987.

[12] R. M. Haralick, Dinstein, and K. Shanmugam, "Textural features for image classification," *IEEE Transactions on Systems, Man, and Cybernetics*, pp. 610–621, 1973.

[13] J. Y. Tou, Y. H. Tay, and P. Y. Lau, "A comparative study for texture classification techniques on wood species recognition problem," *ICNC*, vol. 5, pp. 8–12, 2009.

[14] B. Ren and L. Cheng, "Research of classification system based on naive bayes and metaclass," pp. 154–156, 2009.

[15] Z. Qin, "Naive bayes classification given probability estimation trees," pp. 34–42, 2006.

[16] H. Huang and C. Hsu, "Bayesian classification for data from the same unknown class," vol. 32, pp. 137–145, 2002.