AUTOMATIC LEATHER SPECIES IDENTIFICATION USING MACHINE LEARNING TECHNIQUES

M. Jawahar[1a], S. V. Kanth[1], V. Rajangam[1] and T. Selvi[1]

1 Central Leather Research Institute, Leather Process Technology, Chennai, India

a)Corresponding author: malathy.jawahar@gmail.com

Abstract. Identification and classification of leather species becomes valuable and necessary due to concerns regarding consumer protection, product counterfeiting, and dispute settlement in the leather industry. Identification and classification of leather into species is carried out by histological examination or molecular analysis based on DNA. Manual method requires expertise, training and experience, and due to involvement of human judgment disputes are inevitable thus a need to automate the leather species identification. In the present investigation, an attempt has been made to automate leather species identification using machine learning techniques. A novel non-destructive leather species identification algorithm is proposed for the identification of cow, buffalo, goat and sheep leathers. Hair pore pattern was segmented efficiently using k-means clustering algorithm Significant features representing the unique characteristics of each species such as no.of hair pores, pore density, percent porosity, shape of the pores etc., were extracted. The generated features were used for training the Random forest classifier.Experimental results on the leather species image library database achieved an accuracy of 87 % using random forest as classifier, confirming the potentials of using the proposed system for automatic leather species classification.

1 Introduction

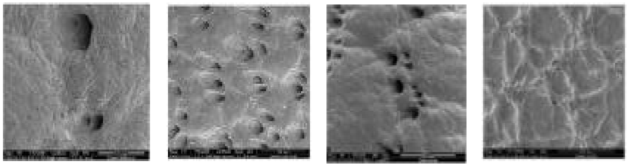
Identifying the species of leather is of paramount importance due to concerns regarding authenticity issues, protecting endangered species, product counterfeiting, consumer protection, etc., Leather industry is fragmented and lacks information exchange systems, neither does it have a provenance system thus information regarding origin of leather is often lost. Labelling of leather is often not done clearly and many a time's disputes arise based on doubts over the origin of the leather. Often disputes between two parties regarding species of leather is brought to expert leather authority to be settled so standardizing the system and removing the human judgment and bias is required to promote free and fair trade thereby enabling the growth of the leather industry.

Each species of the animal has a unique hair pore pattern and this information can be used to detect the species of an unknown leather by examining under a microscope. This method is most efficient while using full grain leathers. It is fast and cost-effective when it comes to identifying leather compared to other methods such as DNA finger printing techniques [1,2] and other histological studies of species identification[3,4].

In this current investigation, image analysis technique has been used to segment hair pores from the leather background and measurement of hair pore properties by extracting image features.

2 MATERIALS AND METHODS

Representative goat, sheep, buffalo and cow leather samples of each species (20 each) as shown in Fig. 1 were selected from the official butt portion for microscopic imaging. Image analysis and machine learning algorithms were implemented in Scilab. The computational workflow for the identification of leather species is shown in Figure 2.



a. Buff b. Cow c. Goat d. Sheep

Fig. 1. Microscopic leather species image.

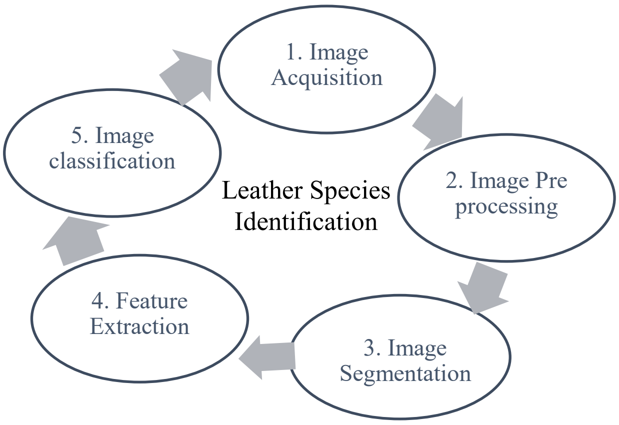


Fig. 2. Work flow diagram of leather species identification.

2.1 Image Preprocessing

Microscopic leather images were preprocessed to highlight the difference between the pores and the surrounding background using adaptive histogram equalization algorithm (Fig.3).

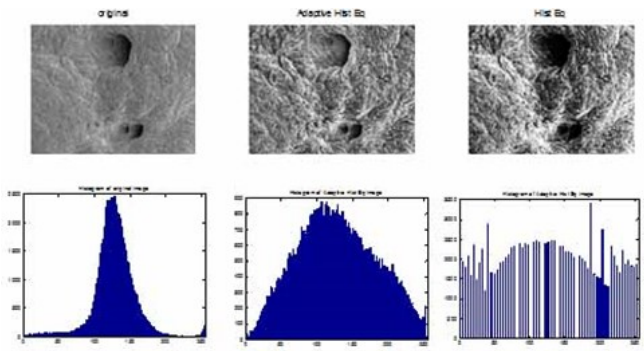


Fig. 3. Preprocessing using Adaptive histogram equalization.

2.2 Image Segmentation

In this study, k-means clustering algorithm is used for segmenting the hair pores from the background leather image. This is an unsupervised learning mechanism where in pixels are clustered or grouped into k clusters based on similarity in intensity and grey values. This algorithm requires k seed points or starting points which dictates the way in which the region will grow along with the membership function which defines the criteria according to which pixels are put into clusters.

Assume the data set is given by $(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n)$, where each observation is a d-dimensional real vector, k-means clustering aims to partition the n observations into k sets where $(k \leq n)$, $\mathbf{S} = \{S_1, S_2, \ldots, S_k\}$ in such a way that it minimizes the within-cluster sum of squares.

$$\underset{\mathbf{S}}{\arg\min} \sum_{i=1}^{k} \sum_{\mathbf{x}_j \in S_i} \|\mathbf{x}_j - \boldsymbol{\mu}_i\|^2 \quad [1]$$

Where μ_i is the mean of points in S_i.

Assignment :

$$S_i^{(t)} = \{x_p : \|x_p - m_i^{(t)}\|^2 \le \|x_p - m_j^{(t)}\|^2 \ \forall\ 1 \le j \le k\}, \quad [2]$$

Further updating:

$$m_i^{(t+1)} = \frac{1}{|S_i^{(t)}|} \sum_{x_j \in S_i^{(t)}} x_j \quad [3]$$

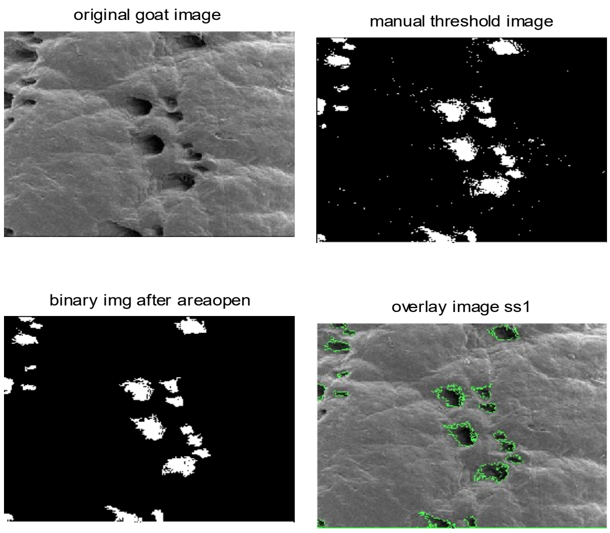


Fig. 4. Segmentation using K-means clustering.

The segmented image (Fig. 4) had many spots which were undesirable and lead to false data. These unwanted data points were effectively removed using erosion of the image followed by dilation. These morphological operations were performed using various structuring elements or masks of which a circular or disk element was found to be most effective.

2.3 Feature Extraction

Segmented image was described using features. In this work, features such as average pore-size distribution, least inter-pore distance, porosity, hair pore density and shape attributes like circularity, aspect ratio, solidity of the pores were computed. The area of the hair pores was estimated by computing the total number of pixels occupied by each pore i.e. region of interest (ROI). Mean, standard deviation, minimum and the maximum intensity value within the ROI were also computed. The Feret's diameter was computed as the longest distance between two parallel lines tangent to the pore. Inter-pore distance was obtained by measuring the distance between the adjacent centroids that ideally represents the centre of the pores. Hair pore density represents the number of pores in a unit area of the image. Porous area fraction or porosity was calculated as the total area occupied by the pores divided by the total area of the investigated image region. Shape of the hair pore was examined in terms of circularity, aspect ratio (AR), solidity.

2.4 Random Forest Classifier

Random Forest (RF) classifier proposed by Breiman[5] is a multiple combination of decision tree. Each tree casts a unit vote for the final classification of the input feature set. Random Forest splits each node using randomly chosen best subset of features. Feature variable importance measure, prediction error on the out-of-bag portion for each tree was computed. The same computation was carried out after permuting each feature variable. Mean and normalized standard deviation was computed. Splitting of tree stops when standard deviation of the difference in feature variable equal to 0.

Random forest uses the gini index to measure the node impurity. It is the measure most commonly chosen for classification-type problems. Gini(T) is defined as

$$Gini(T) = 1 - \sum_{j=1}^{n} P_j^{\,2} \qquad [4]$$

Where P_j represents relative frequency of dataset T with n classes.

Validation was done with unseen dataset (test samples) to evaluate the ability of the classifier to discriminate the four leather species. The performance of the classification in leather images using random forest was measured in terms of Accuracy, Sensitivity, Specificity, and Precision/Positive Predictive Value (PPV), Recall/Negative Predictive Value (NPV) and F score as given in equations 5 to 10.

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FP} + \text{FN} + \text{TN}} \qquad \textbf{[5]}$$

$$\text{Precision or PPV} = \frac{\text{TP}}{\text{TP} + \text{FP}} \qquad [6]$$

$$\text{Recall or NPV} = \frac{\text{TN}}{\text{TN} + \text{FN}} \qquad [7]$$

$$\text{F1 Score} = 2*(\text{Recall} * \text{Precision}) / (\text{Recall} + \text{Precision}) \qquad [8]$$

3 Results and Discussion

In this study, leather grain surface images were obtained using optical microscope. Leather hair pore arrangement and distinctive features for the buffalo, cow, goat and sheep leather species were investigated. Representative sample images of buffalo, cow, goat and sheep leathers are shown in Figure1. 80 leather samples (each species 20) were chosen for image analysis. K-means clustering algorithm was used to segment the hair pore pattern from the leather images which were preprocessed using adaptive histogram equalization algorithm. Features describing the hair pore pattern were extracted. Classification was carried out using random forest to identify the leather species. The work flow diagram is described in Fig 2.

Qualitative analysis of the preprocessed buffalo image using adaptive histogram equalization technique (Fig.3) was found to have a low MSE and a high PSNR value. Segmentation using K-means clustering technique for goat leather shown in Fig. 4. It can be observed that the intensity distributions of the pores and the background often overlapped, thus complicating the separation of information relating to pores alone. Since the hair pores were basically of circular shape, disk structuring element was selected for morphological operation. The basic morphological operators such as dilation and erosion were used to eliminate the unwanted spots thus segmenting the hair pores alone. Performance analysis on segmented image was validated using quality measures Area Error Rate (AER), Overlap Error (OE) and Zijdenbos Similarity index(ZSI) and the results are set out in Table 1.

Table 1. Performance analysis of the segmented image.

Sample	ZSI	AER	OE
Cow	0.8323	0.1967	0.2201
Goat	0.8034	0.2212	0.1898
Sheep	0.7945	0.2215	0.2101
Buffalo	0.8143	0.2123	0.2154

The error between the K-means segmentation algorithm and manually segmented region was re-presented in terms of AER and OE. AER and OE values were closer to 0. This shows that segmentation error is negligible. Zijdenbos Similarity Index (ZSI) quantifies the accuracy of the segmentation [6]. From Fig.1 it can be observed that the cow and buffalo species can be easily segmented and there were no overlapping regions. Hence, ZSI score nearing to 1 was observed for cow (0.83) and buffalo (0.81). Goat and sheep had overlapping regions, therefore ZSI values were 0.80 and 0.79 respectively. The results of the segmentation performance measures show that the K-means clustering algorithm followed by morphological operations could be used for segmentation of leather species. After segmentation of the hair pores, centroid of the hair pores were identified by repeated application of morphological erosion operation and labelled so that feature extraction can be carried out for each label.

Table 2. Hair Pores Distribution based on pore density.

Species	Pore Density (per sq cm)				Percent Porosity
	Overall	Small	Medium	Large	
Goat	2185	1248	936	0	0.027
Cow	2263	0	2263	0	0.046
Sheep	1099	1099	0	0	0.005
Buffalo	304	152	76	76	0.021

All the features extracted from the microscopic image were multiplied with the calibration factor calculated by measuring the length of the known distance in pixels. Hair pores were counted and classified as small (<=1000 sq μm), medium (1000 – 4000 sq μm) and large (> 4000 sq μm). In this study, hair pore density was calculated from the number of pores present per sq cm (Table 2). Buffalo hair pore density was found to be the least among the four species. Overall hair pore density for cow and goat leathers were found to be almost similar but there are differences in the size and distribution of the pores. Cow leathers have uniform pores which are in medium size range whereas. goat hair pores are in both small and medium ranges The pore density of sheep leathers was found to be half the pore density for goat leathers. Porosity defines the ratio of the cumulative surface area occupied by pores to the total surface area of the leather. Goat and cow differ quite considerably in terms of percent porosity. Due to the presence of larger pores buffalo leathers was found to have higher percent porosity than that for sheep leathers. Thus the above extracted features were found to effectively describe the unique characteristics of each of the leather types investigated in this study.

Random forest classifier was trained with extracted features. Performance of the classifier was validated using 5-fold cross validation. The samples were divided into 5 subsets and for every fold one subset was randomly selected as validation set and the remaining 4 subsets were combined together to form as a training set. Average validation scores obtained from the five folds (Table 3) were calculated.

Table 3. Average Classification report for leather species classification.

S.no	Accuracy	Error	Recall	Precision	F-score
Fold 1	83.33	16.67	1.00	0.79	0.88
Fold 2	86.11	13.89	0.92	0.88	0.90
Fold 3	88.89	11.11	0.93	0.93	0.93
Fold 4	80.56	19.44	0.92	0.82	0.87
Fold 5	91.67	8.33	0.96	0.93	0.94
Average	**86.11**	13.88	0.95	0.87	0.90

It can also be observed from Table 3 that the average recall and precision were found to be 95% and 87% respectively. The F-score was 90%. The overall classification accuracy for the proposed method was 86.11%.

4 Conclusion

This study develops automatic leather species recognition system. In the system, a novel non-destructive leather species identification algorithm is proposed for the identification of cow, buffalo, goat and sheep leathers. Hair pore pattern was segmented efficiently using k-means clustering algorithm Significant features representing the unique characteristics of each species such as no.of hair pores, pore density, percent porosity, shape of the pores etc., were extracted. The generated features were used for training the Random forest classifier. Experimental results show that the proposed system can recognize all the four types of leathers with high efficiency and accuracy. In conclusion, it is possible to apply computer vision system to the automatic leather species identification and potential to replace the leather experts.

References

1. *Long, A.L. Addy, V. L., Booth, S., Dwyer, R.O., Loftus, R.; Extraction of DNA from Leather and applications to the*
1. *supply chain. JALCA, 102(1), 22-26, 2007*
2. *Stenzel, S., Bohrisch, J., Pach, M., Meyer, M.; A New Marking System for Leather Based on Encapsulated DNA. JALCA, 110(9), 277-287, 2015.*
3. *Saheb Baha, S.S., Satya Prasad, K., "Automatic detection of Hard Exudates in Diabetic Retinopathy using Morphological segmentation and Fuzzy Logic" International Journal of Computer Science and Network Security, 8 (12), 2008.*
4. *Ziel, A., Haus, R., Tulke, A., Quantification of the pore size distribution (porosity profiles) in microfiltration membranes by SEM, TEM and computer image analysis. J.Mem.Sci, 323(2); 241-246, 10/2008.*
5. *Breiman. L. (2001). Random Forests. Machine Learning, 45, 5–32.*
6. *Alkan, A., Tuncer, S.A., Gunay, A. (2014). Comparative MR image analysis for thyroid nodule detection and quantification. Measurement, 47, 861–868.*