Pre-Trained Convolutional Neural Network for Classification of Tanning Leather Image

Sri Winiarti, Adhi Prahara, Murinto, Dewi Pramudi Ismi

Informatics Department
Universitas Ahmad Dahlan
Yogyakarta, Indonesia

Abstract—**Leather craft products, such as belt, gloves, shoes, bag, and wallet are mainly originated from cow, crocodile, lizard, goat, sheep, buffalo, and stingray skin. Before the skins are used as leather craft materials, they go through a tanning process. With the rapid development of leather craft industry, an automation system for leather tanning factories is important to achieve large scale production in order to meet the demand of leather craft materials. The challenges in automatic leather grading system based on type and quality of leather are the skin color and texture after tanning process will have a large variety within the same skin category and have high similarity with the other skin categories. Furthermore, skin from different part of animal body may have different color and texture. Therefore, a leather classification method on tanning leather image is proposed. The method uses pre-trained deep convolution neural network (CNN) to extract rich features from tanning leather image and Support Vector Machine (SVM) to classify the features into several types of leather. Performance evaluation shows that the proposed method can classify various types of leather with good accuracy and superior to other state-of-the-art leather classification method in terms of accuracy and computational time.**

Keywords—Leather classification; tanning leather; convolution neural network (CNN); deep learning; support vector machine (SVM)

I. INTRODUCTION

Small and medium sized industries lately have experienced a rapid growth. One of them is leather craft industry which produces various kinds of leather craft item, such as gloves, wallet, belt, sandals, shoes, jacket, and bag. The growth of leather craft industry also affects leather tanning factories to increase their production in order to meet the demand of leather craft materials. To achieve large scale production, an automation system should be implemented in the leather tanning factories.

The automation system involves automatic grading based on type and quality of leather. Type of leather usually distinguished based on color and texture using global and/or local statistical geometrical features with machine learning approach [1], [2]. Quality of leather is mainly determined by the size and location of leather defect. Leather defect can be categorized into five types: lines, holes, stains, wears, and knots [3]. To locate the defects, researchers use morphological operation [4], [5], clustering [6], [7], or machine learning approach [3], [8]-[10].

Leather craft materials usually come from cow, crocodile, lizard, buffalo, goat, sheep, and stingray skin. The animal skin will go through tanning process before it can be used as crafting materials. In every level of tanning process, tanning agent will alter the physical properties and chemical compositions of the skin. The skin will become durable, pliant, and may have different color and texture from the original. Therefore, after tanning process, animal skins will have a large variety of color and texture within the same skin category and have high similarity with other skin categories. This make them difficult to be distinguished. Furthermore, skin from different part of animal body may have different color and texture.

The result of classification determines the grade of tanned leather. The grade of leather will affect the price of leather. Therefore, classification procedure is the most important in automation system of tanning leather production because it is directly affects the price of final tanning leather products. Furthermore, a high return rate and disputes between customer and manufacturing industry which caused by failure in classification of leather usually cause additional costs [4].

From leather craftsmen point of view, the correct result of leather grading is important because it will be used as consideration to determine the type of leather craft product that will be made. Mistakes in determining the type of leather for leather craft products can make the resulting leather craft products become unfavorable and impact on loss in sales. Leather craftsmen in some area do not yet have the knowledge about feasible standard of leather crafting. The type and quality of leather that will be used as leather craft materials are known based on experience and tradition which inherited from generation to generation.

Therefore, this research proposes a method to classify type of leather on tanning leather image and performs performance comparison to evaluate the method. In summary, the contributions of this work are given as follows:

1) The proposed method uses tanning leather images as input because tanning leathers are the final product of leather tanning factory and will become leather craft materials for leather craft industry. Therefore, the proposed method can be used and will benefit both of tanning leather factory and leather craft industry.

2) The features of tanning leather images are extracted using specific layer from pre-trained deep convolutional neural

network (CNN). As the layer goes deeper, the richer features will be extracted from the tanning leather images.

3) Classification is done using linear Support Vector Machine (SVM). SVM performs well in few training data, easy to configure the parameters, and has potential to perform real time classification.

4) Finally, performance of the proposed method is compared with the other state-of-the-art leather classification method.

The rest of this paper is organized as follow: Section 2 presents the related works, Section 3 presents the proposed leather classification method, Section 4 presents the results and discussion, and the conclusion of this work is described in Section 5.

II. RELATED WORKS

Leather classification method has been proposed and developed by many researchers. Most of them focus on making an automation system to detect lather defects [3]-[10]. Leather defects inspection system that has purpose to detect the size and location of defects on leather surface is one of the characteristic to determine the quality of leather along with other characteristics such as the type of leather and the correlation between usable and unusable areas on leather. In this research, authors focus on classifying type of leather as it is also an important aspect that affect the quality of leather.

In the tanning leather classification, the difficulties come from large variety and similarity of color and texture of tanning leather. Researchers use statistical geometrical features with machine learning approach to classify the type of leather. In [1], an improved statistical geometrical features (ISGF) is proposed to classify chrome tanning and vegetable tanning leather. They use two classifiers based on Fisher criterion and Learning Vector Quantization (LVQ) network to compare the chrome tanning and vegetable tanning leather. The result shows that ISGF outperforms the performance of SGF.

In their research, [2] use both global and local features to classify leather. The global features are extracted from the two dimensional power spectrum of enhanced leather trench image. The local features are based on mathematical morphology operation of segmented leather image. The features then compared with set of training image that already classified by the experts.

In this research, authors use rich features obtained from pre-trained CNN. The rich features obtained from pre-trained CNN also have been used in some classification problems [11]-[14]. The features are classified using SVM into several types of leather. Authors perform the same procedure and use the same training data for the other leather classification and compared its performance. The other leather classification uses hand-crafted features which are color moments and some statistical measurements from Gray Level Co-Occurrence Matrix (GLCM). GLCM is well known texture features extraction method and has been used in many texture based classification in wide range of applications [15]-[18]. The features are also classified using SVM.

III. METHODOLOGY

This research proposes a method to classify the type of leather to be used in the quality measurement of leather craft materials. General procedure of the proposed method is shown in Fig. 1. From Fig. 1, tanning leather image is resized to fit the input of pre-trained CNN. Authors use the seventh layer of AlexNet [19] which originally trained for ImageNet challenge [20] to extract the features. The features will be classified using linear SVM into five types of leather: monitor lizard, crocodile, sheep, goat, and cow.
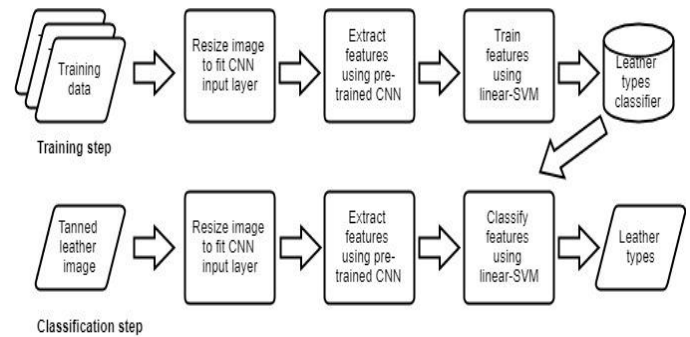

Fig. 1. General procedure of the proposed leather classification method.

A. Tanning Leather Image

Tanning is a process of making leather from the skins of animals. Tanning agents can alter the physical properties and chemical compositions of the leather. The result of tanning is more durable, pliant, and texturally practical material. There are three methods of tanning leather: chroming, vegetable, and combination. Each tanning method produces different materials, aesthetically and texturally e.g. chrome tanned leather is soft and oiled type while vegetable tanned leather is sturdy and hefty. Because tanning agents can change the color and texture of leather surface, the leather images acquired from the industry have large variety even from the same animal. Additionally, the image can be damaged by the external noises.

B. Hand-Crafted Feature Extraction

The hand-crafted leather classification usually based on color and texture. The most common statistical color features are color moments and for statistical texture features are Gray Level Co-Occurrence Matrix (GLCM).

1) Color features extraction: Color features can be extracted using color moments. Color moments are characteristic measurement of color distribution from image. Color moments consist of mean, standard deviation, skewness, kurtosis, and other higher order moments. Color moments are scale and rotation invariant. Color moments that used in this research are explained as follows:

a) Mean is the average value of color on image. If N is the number of pixel on image and p_i is the ith pixel on image then mean (μ) can be calculated using (1).

$$\mu = \frac{1}{N}\sum_{i=1}^{N} p_i \qquad (1)$$

b) Standard deviation is the square root of the variance. The standard deviation (σ) can be calculated using (2).

$$\sigma = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}|p_i - \mu|^2} \qquad (2)$$

c) Skewness is a measure of the asymmetry of the data around the sample mean. If skewness is negative, the data are spread out more to the left of the mean than to the right. If skewness is positive, the data are spread out more to the right. The skewness (s) can be calcualted using (3).

$$s = \sqrt[3]{\frac{1}{N-1}\sum_{i=1}^{N}|p_i - \mu|^3} \qquad (3)$$

d) Kurtosis is a measure of how the outlier-prone a distribution. Kurtosis k can be calculated using (4).

$$k = \sqrt[4]{\frac{1}{N-1}\sum_{i=1}^{N}|p_i - \mu|^4} \qquad (4)$$

2) Texture features extraction: Texture features can be extracted from Gray Level Co-Occurrence Matrix (GLCM). GLCM uses spatial correlation between pixels. Feature extraction using GLCM is done by measuring the occurrence level of paired pixel with specific value on the image then calculate the statistical texture features [21]. The features can be measured from four different orientations or offsets which are 0°, 45°, 90°, and 135° . Some of the statistical measurements of GLCM that used in this research are explained as follows:

a) Contrast is a measure of the intensity between a pixel and its neighbor over the whole image. Contrast is used to measure the local variance level in GLCM matrix. If $p(i,j)$ is the GLCM in coordinate (i,j), then contrast can be calculated using (5).

$$contrast = \sum_{i,j}|i - j|^2 p(i,j) \qquad (5)$$

b) Correlation is a measure of how correlated a pixel is to its neighbor over the whole image. Correlation is used to measure the occurrence of paired pixels in GLCM. If $\mu_i = \sum_{i,j} i \cdot p(i,j)$ and $\mu_j = \sum_{i,j} j \cdot p(i,j)$ is the mean of GLCM i and j , and $\sigma_i = \sqrt{\sum_{i,j} p(i,j)(i - \mu_i)^2}$ and $\sigma_j = \sqrt{\sum_{i,j} p(i,j)(j - \mu_j)^2}$ is the standard deviation of GLCM i and j then correlation can be calculated using (6).

$$correlation = \sum_{i,j}\frac{(i-\mu_i)(j-\mu_j)p(i,j)}{\sigma_i \sigma_j} \qquad (6)$$

c) Energy is the sum of squared elements in the GLCM and can be calculated using (7).

$$energy = \sum_{i,j} p(i,j)^2 \qquad (7)$$

d) Homogeneity is a value that measures the closeness of the distribution of elements in the GLCM to the GLCM diagonal and can be calculated using (8).

$$homogeneity = \sum_{i,j}\frac{p(i,j)}{1+|i-j|} \qquad (8)$$

e) Entropy is a measure of non-uniformity and texture complexity of image. Entropy can be calculated using (9).

$$entropy = -\sum_{i,j} p(i,j)\log p(i,j) \qquad (9)$$

C. Pre-Trained CNN for Feature Extractor

Convolutional Neural Network (CNN) is a powerful machine learning technique. In the deep learning field, CNNs are trained using large collections of diverse images. From these large collections, CNNs can learn rich feature representations for a wide range of images. CNNs have many layers namely input layer, convolutional layers, Rectified Linear Unit (ReLU) layers, cross channels normalization layers, average pooling layers, max pooling layers, fully connected layers, dropout layers, softmax layers, and output classification layers. Each layer in the networks takes in data from the previous layer, transforms the data, and passes the data on the next layer. The network will learn directly from the data and increases the complexity and detail of what it is learning from layer to layer. The function of some layers are explained as follows:

a) Convolutional layer puts the input images through a set of convolutional filters, each of which activates certain features from the images.

b) Pooling layer simplifies the output by performing nonlinear downsampling, reducing the number of parameters that the network needs to learn.

c) ReLU layer allows faster and more effective training by mapping negative values to zero and maintaining positive values.

d) Fully connected layer has outputs a vector of K dimensions where K is the number of classes that the network will be able to predict. This vector contains the probabilities for each class of any image being classified.

e) The final layer of the CNN architecture uses a softmax function to provide the classification output.

In order to leverage the power of CNNs without investing time and effort into training is to use a pre-trained CNN as a feature extractor. Layer that will be used as feature extractor is the fully connected layer that extract richer features compared to the lower layer. Most of the models have been trained on the ImageNet dataset [20], which has 1000 object categories and 1.2 million training images.

One of the model is AlexNet [19] which published in 2012. It can classify images into 1000 different categories, including keyboards, computer mice, pencils, and other office equipment, as well as various breeds of dogs, cats, horses, and other animals. The architecture of AlexNet is shown in Fig. 2 and the sequence of layer that used in the model is listed below:

- Input image layer: 227x227x3
- Convolutional layer 1: 96 11x11 filters at stride 4
- ReLU layer 1
- Cross channel normalization layer 1
- Max pooling layer 1: 3x3 filters at stride 2
- Convolutional layer 2: 256 5x5 filters at stride 1, pad 2
- ReLU layer 2
- Cross channel normalization layer 2

- Max pooling 2: 3x3 filters at stride 2
- Convolutional layer 3: 384 3x3 filters at stride 1, pad 1
- ReLU layer 3
- Convolutional layer 4: 384 3x3 filters at stride 1, pad 1
- ReLU layer 4
- Convolutional layer 5: 256 3x3 filters at stride 1, pad 1
- ReLU layer 5
- Max pooling 3: 3x3 filters at stride 2
- Fully connected layer 6: 4096 neurons
- ReLU layer 6
- Fully connected layer 7: 4096 neurons
- Fully connected layer 8 (softmax layer): 1000 neurons
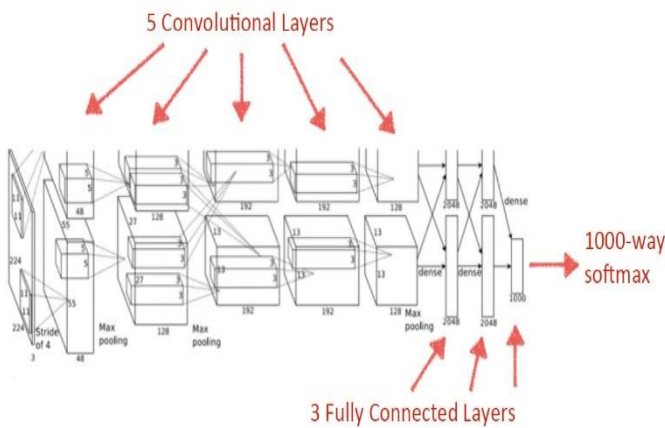- Classification output layer

Fig. 2. The architecture of AlexNet [19].

D. Tanning Leather Image Classification

Support vector machine (SVM) is used to classify tanning leather image from set of features extracted using pre-trained CNN. SVM is a classifier that determined by separation which called hyperplane. Hyperplane can be calculated by maximizing margin or distance from two set of object from two different class. Classification with SVM consist of training and classification. In the training step [22], if there are given training data $x_i \in R^n, i = 1, \dots l$ in two classes and label $y \in R^l$ such that $y_i \in \{1, -1\}$, it can be solved with (10).

$$\min_{w,b,\xi} \frac{1}{2} w^T w + C \sum_{i=1}^{l} \xi_i$$

$$subject\ to\ y_i(w^T \phi(x_i) + b) \geq 1 - \xi_i, \qquad (10)$$

$$\xi_i \geq 0\ i = 1, \dots, l$$

where w is weight vector, b is bias, ξ_i is slack variables, $\phi(x_i)$ maps x_i into higher dimensional space and $C > 0$ is the regularization parameter and the decision function is shown in (11).

$$sgn(w^T \phi(x) + b) = sgn\left(\sum_{i=1}^{l} y_i \alpha_i K(x_i, x) + b\right) \quad (11)$$

where $K(x_i, x_j)$ is the kernel function. After training process, parameter $y_i \alpha_i \forall i$, b, label names, support vectors, and kernel parameter saved as output model from training SVM.

In classification, voting strategy is performed for each data *x* which will be designated to be in a class with the maximum votes [22]. Optimal parameter is selected using *k*-fold cross validation which is a method to do cross validation by dividing training data into *k* set which has (*k*-1) as training data and the rest will be the test data. After training process we will get variable *w, x*, and *b* for each class, then the classification process can be done with these steps:

1) Calculate kernel.
2) Calculate decision function using (11).
3) Repeat step 1 and 2 for other classes.
4) Determine the class by function which gives the most maximum result.

E. Performance Evaluation

The evaluation procedure uses confusion matrix then calculates the accuracy, specificity, sensitivity, and precision using (12), (13), (14), and (15) respectively where TP is true positive, FP is false positive, TN is true negative, and FN is false negative. In order to handle imbalance dataset between positive and negative test data, authors use ratio as weight in the calculation of statistical performance measurement to balance the influence of the dataset.

$$Accuracy = \frac{TP+TN}{TP+FP+TN+FN} \qquad (12)$$

$$Specificity = \frac{TN}{(TN+FP)} \qquad (13)$$

$$Sensitivity = \frac{TP}{(TP+FN)} \qquad (14)$$

$$Precision = \frac{TP}{(TP+FP)} \qquad (15)$$

IV. RESULT AND DISCUSSION

The proposed method runs on Laptop with Intel i7 processor, 16GB of RAM, and graphics card NVIDIA GTX 1050 4GB. Performance evaluation is done using MATLAB. The proposed method will be tested to classify leather into five types: monitor lizard, crocodile, sheep, goat, and cow skin.

Training data consist of 1000 leather images i.e. 200 leather images for each category. Authors conduct performance evaluation test on 3157 leather images. Input tanning leather images are taken using camera with 15-50 cm distances from leather objects and saved into images with size 512x512 pixels. Fig. 3 shows the samples of leather image used in this research. From Fig. 3, tanning leathers may have different texture within the same category which caused by tanning process or the skins are taken from different part of the animal body e.g. reptiles have different color and texture of skin in their belly and back.
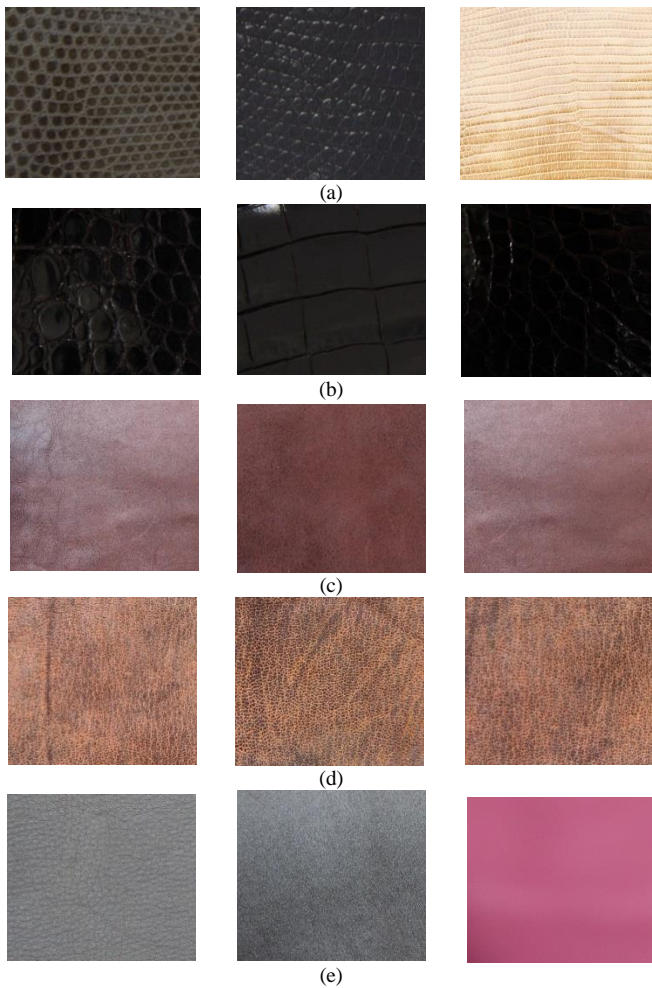
Fig. 3. Sample of leather images (a) monitor lizard (b) crocodile (c) sheep (d) goat (e) cow.

A. Result of Hand-Crafted Feature Extraction

For the hand-crafted feature extraction, input images are converted into grayscale. Authors extract 24 global features from input image, consist of 4 features from color moments namely mean, standard deviation, skewness, and kurtosis and the next 20 features come from statistical texture measurement of GLCM in 4 offsets, namely, contrast, energy, correlation, homogeneity, and entropy. Features are standardized then classified using SVM. Confusion matrix and statistical measurement per category for this hand-crafted leather classification scheme is shown in Table I.

From Table I, all samples of crocodile, sheep, and goat tanning leather images can be recognized perfectly except in the monitor lizard and cow categories only 924 of 1000 and 987 of 1000 samples are recognized. In the case of monitor lizard category, 74 samples falsely classified as crocodile, 1 samples falsely classified as sheep, 1 samples falsely classified as cow, and 12 of cow samples are falsely classified as monitor lizard skins. Because the false positive of monitor lizard samples which classified as crocodile skins, the precision of crocodile category becomes 0.9668 and the sensitivity of monitor lizard category becomes 0.9240. The accuracy of monitor lizard category also drops to 0.9592.

TABLE I.   CONFUSION MATRIX AND STATISTICAL MEASUREMENT OF LEATHER CLASSIFICATION USING HAND-CRAFTED FEATURE EXTRACTION

Confusion Matrix					
Actual	**Predicted**				
	Monitor Lizard	**Crocodile**	**Sheep**	**Goat**	**Cow**
Monitor Lizard	924	74	1	0	1
Crocodile	0	1000	0	0	0
Sheep	0	0	150	0	0
Goat	0	0	0	7	0
Cow	12	0	1	0	987
Statistical Measurement					
Accuracy	0.9592	0.9828	0.9998	1.0000	0.9933
Specificity	0.9944	0.9657	0.9997	1.0000	0.9995
Sensitivity	0.9240	1.0000	1.0000	1.0000	0.9870
Precision	0.9940	0.9668	0.9993	1.0000	0.9995

B. Result of Feature Extraction using Pre-Trained CNN

For the feature extraction using pre-trained CNN, input image is resized to 227x227 to fit CNN input layer. Authors use the eighth layer of AlexNet which is the last fully connected layers before output as activation function to extract features in order to gain rich features for classification. Another reason is to make the features compact because the eighth layer produces 1000 features and the other fully connected layers produce 4096 features. Features are classified using SVM. Confusion matrix and statistical measurement per category for leather classification scheme using pre-trained CNN is shown in Table II.

TABLE II.   CONFUSION MATRIX AND STATISTICAL MEASUREMENT OF LEATHER CLASSIFICATION USING PRE-TRAINED CNN

Confusion Matrix					
Actual	**Predicted**				
	Monitor Lizard	**Crocodile**	**Sheep**	**Goat**	**Cow**
Monitor Lizard	998	0	0	0	2
Crocodile	0	1000	0	0	0
Sheep	0	0	150	0	0
Goat	0	0	0	7	0
Cow	0	0	0	0	1000
Statistical Measurement					
Accuracy	0.9990	1.0000	1.0000	1.0000	0.9995
Specificity	1.0000	1.0000	1.0000	1.0000	0.9991
Sensitivity	0.9980	1.0000	1.0000	1.0000	1.0000
Precision	1.0000	1.0000	1.0000	1.0000	0.9991

TABLE III.   LEATHER CLASSIFICATION PERFORMANCE COMPARISON

Datasets	Leather Classification Method	
	Hand-crafted features (color moments + GLCM)	Features from pre-trained CNN (8th fully connected layer)
Average Accuracy	0.9870	0.9997
Average Specificity	0.9919	0.9998
Average Sensitivity	0.9822	0.9996
Average Precision	0.9919	0.9998
Average Computational Time	0.3793 seconds	0.2015 seconds

From Table II, all test samples are perfectly classified except in the monitor lizard category. In the case of monitor lizard category, 998 of 1000 samples are recognized and only two samples are falsely classified as cow skins. This small mistakes means, the features extracted from pre-trained CNN can exploit the characteristic of each leather category. Furthermore, the feature selection procedure already handled by pre-trained CNN which learned from large categories of image.

The statistical measurements per category in Tables I and II are averaged to calculate the whole performance of the proposed method. Authors also measure the computational time and compare the result from both classification schemes. From Table III, the hand-crafted features extraction scheme produces lower performance than pre-trained CNN. The computational time for hand-crafted features is also longer than pre-trained CNN despite only use small number of features. In the hand-crafted feature extraction, input image needs to be transformed into GLCM matrix before calculating the statistical texture features while in the feature extraction using pre-trained CNN, the features are directly extracted from input image with activation function which are the weights from specific fully connected layer.

The outstanding classification performance of the proposed method mainly caused by utilizing pre-trained CNN in this case AlexNet that designed for harder classification problem with 1000 categories of objects and trained using 1.2 millions of images. Therefore, for smaller and simpler five categories classification task can produce good performance when using pre-trained CNN as feature extractor.

V. CONCLUSION

In this research, a method to classify tanning leather image is proposed. The classification process is used to ensure the quality of leather craft materials produced by leather tanning factories and used by leather craft industries by recognizing the type of animal skin. The method uses tanning leather image as input then perform feature extraction using pre-trained CNN. The classification is done using SVM into five leather categories. The performance evaluation shows that the proposed method can classify leather with good accuracy and precision. The proposed method also superior to hand-crafted classification scheme. For future works, authors aim to add more types of leather to be classified, improve the classification performance, speed-up the computational time, and implement the method in mobile application.

REFERENCES

[1]   L. Wang and C. Liu, "Tanning Leather Classification using an Improved Statistical Geometrical Feature Method," in *2007 International Conference on Machine Learning and Cybernetics*, 2007, pp. 1765–1768.

[2]   Q. Wang, H. Liu, J. Liu, and T. Wu, "A new method for leather texture image classification," in *[1992] Proceedings of the IEEE International Symposium on Industrial Electronics*, pp. 304–307.

[3]   C. Kwak, J. A. Ventura, and K. Tofang-Sazi, "A neural network approach for defect identification and classification on leather fabric," *J. Intell. Manuf.*, vol. 11, no. 5, pp. 485–499, 2000.

[4]   C. Yeh and D.-B. Perng, "Establishing a Demerit Count Reference Standard for the Classification and Grading of Leather Hides," *Int. J. Adv. Manuf. Technol.*, vol. 18, no. 10, pp. 731–738, Nov. 2001.

[5]   F. P. Lovergine, A. Branca, G. Attolico, and A. Distante, "Leather inspection by oriented texture analysis with a morphological approach," in *Proceedings of International Conference on Image Processing*, pp. 669–671.

[6]   K. Hoang, W. Wen, A. Nachimuthu, and X. L. Jiang, "Achieving automation in leather surface inspection," *Comput. Ind.*, vol. 34, no. 1, pp. 43–54, Oct. 1997.

[7]   H. Chen, "The Research of Leather Image Segmentation Using Texture Analysis Techniques," *Adv. Mater. Res.*, vol. 1030–1032, pp. 1846–1850, Sep. 2014.

[8]   R. Viana, R. B. Rodrigues, M. A. Alvarez, and H. Pistori, "SVM with Stochastic Parameter Selection for Bovine Leather Defect Classification," in *Advances in Image and Video Technology*, Berlin, Heidelberg: Springer Berlin Heidelberg, 2007, pp. 600–612.

[9]   M. Jawahar, N. K. C. Babu, and K. Vani, "Leather texture classification using wavelet feature extraction technique," in *2014 IEEE International Conference on Computational Intelligence and Computing Research*, 2014, pp. 1–4.

[10] A. Branca, M. Tafuri, G. Attolico, and A. Distante, "Automated system for detection and classification of leather defects," *Opt. Eng.*, vol. 35, no. 12, p. 3485, Dec. 1996.

[11] Jia Deng, Wei Dong, R. Socher, Li-Jia Li, Kai Li, and Li Fei-Fei, "ImageNet: A large-scale hierarchical image database," in *2009 IEEE Conference on Computer Vision and Pattern Recognition*, 2009, pp. 248–255.

[12] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation," *Proc. 2014 IEEE Conf. Comput. Vis. Pattern Recognit.*, pp. 580–587, Jun. 2014.

[13] K. Yanai and Y. Kawano, "Food image recognition using deep convolutional network with pre-training and fine-tuning," in *2015 IEEE International Conference on Multimedia & Expo Workshops (ICMEW)*, 2015, pp. 1–6.

[14] O. Russakovsky *et al.*, "ImageNet Large Scale Visual Recognition Challenge," *Int. J. Comput. Vis.*, vol. 115, no. 3, pp. 211–252, Dec. 2015.

[15] R. Maurya, Surya Kant Singh, A. K. Maurya, and A. Kumar, "GLCM and Multi Class Support vector machine based automated skin cancer classification," in *2014 International Conference on Computing for Sustainable Global Development (INDIACom)*, 2014, pp. 444–447.

[16] G. Preethi and V. Sornagopal, "MRI image classification using GLCM texture features," in *2014 International Conference on Green Computing Communication and Electrical Engineering (ICGCCEE)*, 2014, pp. 1–6.

[17] M. Hall-Beyer, "Practical guidelines for choosing GLCM textures to use in landscape classification tasks over a range of moderate spatial scales," *Int. J. Remote Sens.*, vol. 38, no. 5, pp. 1312–1338, Mar. 2017.

[18] S. Mukherjee and S. Pandey, "Road Surface Classification Using Texture Synthesis Based on Gray-Level Co-occurrence Matrix," Springer, Singapore, 2017, pp. 323–333.

[19] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Proceedings of the 25th International Conference on Neural Information Processing Systems - Volume 1*. Curran Associates Inc., pp. 1097–1105, 2012.

[20] ImageNet. Large-scale hierarchical image database. Available online at http://www.image-net.org. Accessed on November 22nd, 2017.

[21] R. M. Haralick, K. Shanmugam, and I. Dinstein, "Textural Features for Image Classification," *IEEE Trans. Syst. Man. Cybern.*, vol. SMC-3, no. 6, pp. 610–621, Nov. 1973.

[22] C.-C. Chang and C.-J. Lin, "LIBSVM," *ACM Trans. Intell. Syst. Technol.*, vol. 2, no. 3, pp. 1–27, Apr. 2011.