Recognizing Clothes Patterns for Blind People by Confidence Margin based Feature Combination

Xiaodong Yang, Shuai Yuan, and YingLi Tian
Department of Electrical Engineering
The City College of New York, CUNY
New York, NY, 10031, USA
{xyang02, syuan00, ytian}@ccny.cuny.edu

ABSTRACT

Clothes pattern recognition is a challenging task for blind or visually impaired people. Automatic clothes pattern recognition is also a challenging problem in computer vision due to the large pattern variations. In this paper, we present a new method to classify clothes patterns into 4 categories: stripe, lattice, special, and patternless. While existing texture analysis methods mainly focused on textures varying with distinctive pattern changes, they cannot achieve the same level of accuracy for clothes pattern recognition because of the large intra-class variations in each clothes pattern category. To solve this problem, we extract both structural feature and statistical feature from image wavelet subbands. Furthermore, we develop a new feature combination scheme based on the confidence margin of a classifier to combine the two types of features to form a novel local image descriptor in a compact and discriminative format. The recognition experiment is conducted on a database with 627 clothes images of 4 categories of patterns. Experimental results demonstrate that the proposed method significantly outperforms the state-of-the-art texture analysis methods in the context of clothes pattern recognition.

Categories and Subject Descriptors

I.4.8 [**Scene Analysis**]: Object Recognition

General Terms

Algorithms, Design.

Keywords

Clothes pattern, recognition, computer vision, blind, visually impaired.

1. INTRODUCTION

Based on statistics from the World Health Organization (WHO), there are more than 161 million visually impaired people around the world, and 37 million of them are blind [9]. In everyday's life, choosing suitable clothes becomes a challenging task for blind or visually impaired people. They manage this task either through helping form their family members, or by using plastic Braille labels or different types of stitching patterns tags on the clothes, or just wearing clothes without any patterns. Recent advances in theories, sensors, and embedded computing hold the promise to enable computer vision technique to address their needs. Although some methods have been developed to determine whether clothes color and pattern are matched to help blind people [12], they cannot recognize the categories of clothes patterns.

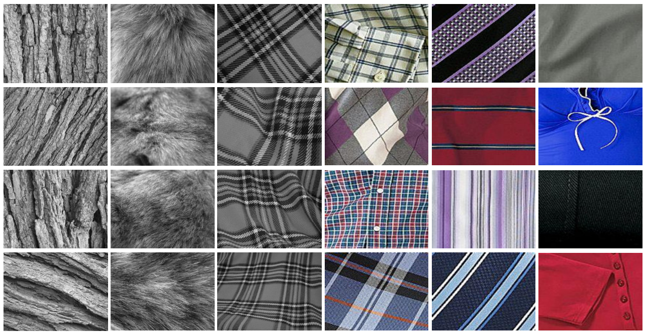

Figure 1: Intra-class variations in traditional texture images and clothes patterns. Column 1-3: texture samples from three categories with less intra-class pattern variations. Column 4-6: clothes pattern samples of three categories with large intra-class variations. Samples of the same column belong to the same category.

Extensive research has been done for texture analysis to make representations of texture robust to viewpoint changes, non-rigid deformation, illumination variance, rotation, scaling, and occlusion [3, 11, 13]. A traditional representation of texture is the statistical features extracted from wavelet subbands of texture images. This method [11] utilizes the spectral information of texture images at different scales to characterize the global energy distribution property. However, such multiresolution approach ignores the local structural information. Most recent state-of-the-art texture recognition approaches [3, 13] represent texture as a histogram of textons by clustering local image features. Textons are the repetitive basic primitives to characterize a specific texture pattern. Because of the robustness to photometric and affine variations, SIFT [5] is commonly used to capture the structural information of texture. On the other hand, it was observed [13] that the combination of multiple complementary features usually achieves better results than the most discriminative individual feature.

As shown in Fig. 1, traditional texture analysis methods mainly focus on recognizing textures with large changes of viewpoint, orientation, and scaling, but with less intra-class pattern variations. However, for application of clothes pattern recognition, in addition to the above variations, there are large intra-class variations due to the huge number of clothes pattern

*Area Chair: Alexander Hauptmann

MM'11, Novermber 28—December 1, 2011, Scottsdale, Arizona, USA.
Copyright 2011 ACM 978-1-4503-0616-4/11/11...$10.00.

designs of each specific clothes pattern. In this paper, we recognize clothes patterns into 4 categories: stripe, lattice, special, and patternless. The 4 categories are able to meet the basic requirements based on our initial survey with potential blind users. In order to handle the large intra-class variation and maintain the discrimination of inter-class variance, both statistical feature and structural feature are employed. While it is customary to simply concatenate multiple feature descriptors as feature combination strategy, we propose a confidence margin based scheme to combine individual features in a more compact and more discriminative way. Recognition experiments on the clothes pattern database validate the superiority of our proposed method over the-state-of-art texture analysis approaches.

2. MULTIPLE FEATURE CHANNELS

2.1 Primitive Feature

Statistical features are generally employed to analyze texture which lacks clutter and has uniform statistical properties. The most common approach is to extract energy values for all wavelet subbands of texture images. Wavelet subbands represent a generalization of multiresolution analysis tool. Several energy functions, such as magnitude, magnitude square, and rectified sigmoid [8], can be used to extract statistical features from each subband. Statistical features from wavelet subbands capture global spectral information of texture images at different scales. In this paper, we employ 5 statistical values including variance, smoothness, homogeneity, entropy, and laws energy [2] to constitute the statistical feature (STA).

Local image descriptors are able to capture the structural information that is ignored in the statistical features. The bag-of-words model [7] is then used to quantize local image descriptors to *visual words* in finite vocabularies. This technique treats texture images as loose collections of independent patches. The distribution of patches in the *visual words* vocabulary is then used to characterize the texture. Patches of texture images are represented by local image descriptors. In the performance evaluation of several local image descriptors [6], it was observed SIFT-based descriptor outperformed others because of its strong invariance to photometric change and affine transformation. SIFT descriptor is created by sampling gradient maps of a support region over a 4×4 grids, with 8 orientation bins for each grid. The magnitude of each point within the support region is weighted by a Gaussian window function to emphasize for the gradients closed to the center of the region and decrease the impact of small changes in the position of the region. The feature vector with 128 elements is then normalized and thresholded to remove elements with small values.

2.2 Multiple Features

In order to deal with the large intra-class variations presented in the clothes patterns, we employ both statistical feature (STA) and local structural feature (SIFT). Inspired by the spectral information and multiresolution provided by Wavelet subbands, we extract SIFT and STA from the original image and its associated wavelet subbands. Thus, for each local patch, STA and SIFT are constructed from original image, and its corresponding horizontal component, vertical component, and diagonal component of wavelet subbands. As shown in Fig. 2, the local patch surrounded by the red square is the support region to build SIFT descriptor; and the local patch surrounded by green square is the support region to compute STA descriptor. Note the support region of STA is larger than the support region of SIFT. This is because STA, the descriptor representing statistical property, needs larger region or more samples to be stable and meaningful. SIFT descriptor is built by sampling gradient map of its support region over 4×4 grids, with 8 orientation bins for each grid. Similarly, STA is calculated by sampling statistical properties of its support region over 4×4 grids. Each grid is further expended to 3 concentric grids with 5 statistical values (Sec. 2.1) representation for each grid. Therefore, SIFT descriptor and STA descriptor from each channel are of the dimensions of 128 and 240, respectively. We consider each descriptor/subband pair as a separate feature channel.

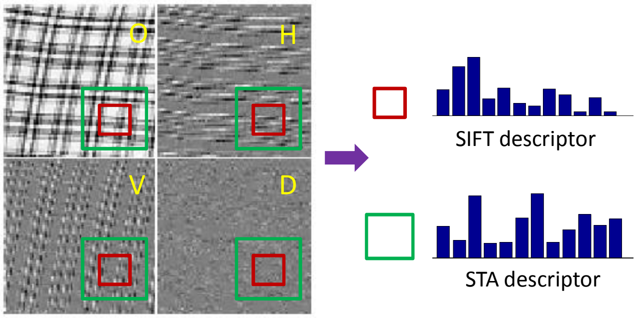

Figure 2: Multiple feature channels. SIFT and STA extracted from patches of wavelet subbands: original component (O), horizontal component (H), vertical component (V), and diagonal component (D).

3. FEATURE COMBINATION

Multiple complementary features capture properties of a phenomenon in different aspects. So, a combination of multiple complementary features is able to obtain better results than any individual feature channel. It is customary to directly concatenate feature vectors of multiple channels. While this method is simple and straightforward, it suffers the following drawbacks: 1) it cannot manifest the complementary relationships between different feature channels; 2) it always results in a very high feature dimension causing the curse of dimensionality [1]; 3) it might submerge some feature channels because of the imbalanced feature dimensions. In this paper, we employ a confidence margin based feature combination scheme to combine multiple feature channels. The final feature combined in this way has a low dimension but more discriminative power.

The confidence margin [4] is the measure of how close an instance is to the classification boundary of a classifier. It represents the reliability of prediction output based on a specific feature. In the context of classification, an instance close to the class boundary is less reliable than the one deep in the class territory. The Support Vector Machines (SVM) [10] is used as the classifier in our clothes pattern recognition system. SVM finds a maximum margin hyperplane in the feature space and can be solved by the Lagrange dual problem:

$$\min_{a} \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} a_i a_j y_i y_j K(x_i, x_j) - \sum_{i=1}^{N} a_i$$

$$s.t. \sum_{i=1}^{N} y_i a_i = 0, \qquad 0 \le a_i \le R$$

Here $K(x_i, x_j)$ is the kernel that generates an inner product of the transformed x_i and x_j in the feature space; y_i is the label of x_i; R

is the regularization parameter. SVM makes predictions of x by the sign of $g(x) = \sum_{i=1}^{N} a_i y_i K(x_i, x) + b$. $|g(x_i)|$ is defined as the magnitude of confidence margin of the instance x_i. SVM is fundamentally a two-class classifier. The one-versus-one [1] approach is used to recognize N $(N > 2)$ categories by training $N(N-1)/2$ different 2-class SVMs on all possible pairs of classes. An instance is recognized as the category with the highest number of votes. Accordingly, an instance x_i has $N(N-1)/2$ confidence margin values, which are used to constitute the confidence margin vector cm_i which is an alternate representation of x_i. There are four categories $(N = 4)$ in the clothes pattern dataset. So no matter what dimension of original features x_i, their confidence margin representations cm_i are all with the same dimension of 6. The proposed feature combination scheme, rather than using original feature vectors, employs their confidence margin representations to form the final combined feature. If F feature channels are used, $[x_i^1, x_i^2, \dots, x_i^F]^T$ is the final feature vector generated by traditional feature combination method, and $[cm_i^1, cm_i^2, \dots, cm_i^F]^T$ is the final feature vector obtained by our proposed method; where cm_i^f is the confidence margin vector of original feature x_i^f in the feature channel f.

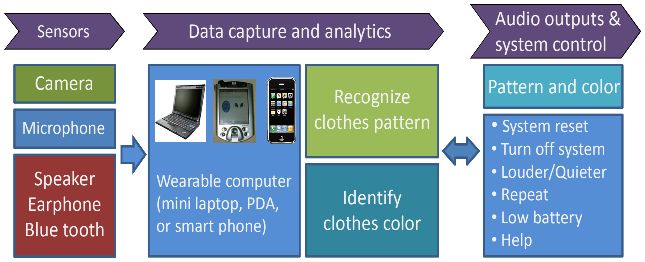

Figure 3: Prototype hardware and architecture design of clothes pattern recognition system for the blind and visually impaired persons.

4. SYSTEM AND INTERFACE

In addition to clothes patterns, color is also an important property of clothes. We apply the clothes color classification method in [12] to recognize clothes color to flowing categories: red, orange, yellow, green, cyan, blue, purple, pink, black, gray, and white. The system integrates both clothes pattern recognition and clothes color recognition.

The clothes pattern and color recognition system consists of three sections: 1) a camera connected to a computing machine to perform clothes pattern and color recognition; 2) speech commands for system control and configuration; 3) audio feedback to provide recognition results for both patterns and colors of clothes. The prototype of the proposed system integrates different sensors including a camera, a microphone, and audio output devices which can be an earphone, Bluetooth, or speakers. A camera is used to capture images of clothes. A wearable computer (can be a PDA or a smart phone) is used to capture and analyze data. The recognition results are described to the blind user by verbal display with minimal distraction of the user's hearing sense. The user can control the system by speech via microphones. The information is processed by the computing machine.

In our prototype, we develop a program based on Microsoft SDK tools for speech recognition and audio feedback to make it easy to use by blind users. As a user gives simple speech commands through a microphone, the system can directly recognize input commands, execute corresponding functions, and provide final audio outputs. In addition, the system can be set by a number of high priority speech commands such as *System reset*, *Turn off system*, *Repeat result*, *Help*, and speaker volume control commands (i.e. *Louder* and *Quieter*). To protect privacy and minimize masking environmental sounds, bone conduction earphones or small wireless blue tooth speakers can be used. The system will also check battery level and send out an audio warning when the battery level is low. Fig. 3 shows the system interface for development.

5. EXPERIMENTS AND DISCUSSIONS

A clothes pattern dataset with large intra-class variation is collected to evaluate our proposed clothes pattern recognition method and other state-of-the-art texture classification approaches. Experimental results demonstrate the complementary relationships of features from multiple channels, i.e. SIFT and STA features from original image and associated wavelet subbands. In addition, the recognition results also validate the superiority of the proposed confidence margin based feature combination scheme over the traditional feature combining method.

5.1 Datasets

To evaluate the recognition performances of our proposed method on clothes patterns, we collected a dataset including 627 images of 4 different typical clothes pattern designs: stripe, lattice, special, and patternless with 157, 156, 158, and 156 images in each category, respectively. This dataset will be released to public. The resolution of each image is 140×140. Fig. 4 illustrates sample images in each category. As shown in this figure, as well as lighting variances, scale changes, rotations, and surface deformations presented in traditional texture dataset, clothes patterns also demonstrate much larger intra-class variations, which augments the challenges for recognition.

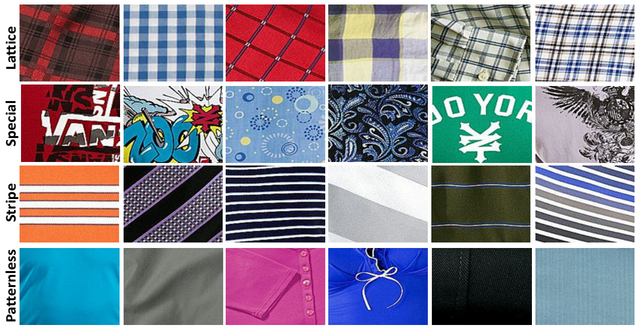

Figure 4: Sample images in the clothes pattern dataset. Rows 1-4 correspond to the categories of lattice, special, stripe, and patternless.

5.2 Results and Discussions

We evaluate the recognition performance of our proposed method and state-of-the-art texture classification approaches on the clothes pattern dataset. In order to detect sufficient keypoints for each category, we evenly select keypoints in every 4 pixels. The sizes of support regions for STA and SIFT are empirically determined as 23×23 and 17×17. The final descriptors compared in recognition experiments include STA, SIFT, WSTA, WSIFT, WSIFT+WSTA, and WSIFT*WSTA. STA and SIFT are statistical descriptors and SIFT descriptors extracted from original images; WSTA and WSIFT correspond to statistical descriptors and SIFT descriptors extracted from original images and associated wavelet subbands; WSIFT+WSTA denotes the combined feature by simply concatenating the feature vectors of WSIFT and WSTA; WSIFT*WSTA refers the final feature combined by confidence margins of SIFT and STA from original images and wavelet subband images. The recognition experiments are evaluated by using 10%, 30%, 50%, and 70% of the dataset as training sets, and the rest as testing sets. The recognition results for different combinations of descriptors and training sets are demonstrated in Table 1. The reported recognition rates are the average over 100 random subsets. The number in bold correspond to the best recognition rate for different training volumes.

As shown in Table 1, the combination of multiple feature channels usually achieves better recognition results than individual feature channel. The recognition accuracy based on WSTA is better than STA, and WSIFT is better than SIFT as well, which verifies the complementary relationships of multiple feature channels. This is because that the horizontal, vertical, and diagonal images of wavelet subbands provide complementary frequency information to the spatial information of original images. On the other hand, the improvement of recognition rate based on WSIFT+WSTA over WSIFT or WSTA is not obvious. When only 10% of images are used as training set, WSIFT+WSTA even deteriorates the recognition performance. This is probably because the training set is too small compared to the large dimension of WSIFT+WSTA. However, if we use WSIFT*WSTA, or the descriptor combined by the confidence margin method, the recognition accuracy is significantly improved. This improvement benefits from confidence margin based combination method is able to manifest the complementary relationships between different feature channels, balance the feature dimensions of different feature channels, and generate feature vectors in a much lower dimension. Therefore, the final feature combined in this way has more discriminative power but far fewer feature dimensions. Furthermore, when the training set is larger than 10%, our proposed method is less sensitive to the volume of training data. For instance, our method can acquire comparable results by using 30% and 50% images as training set. In other words, we can train the proposed method to achieve comparable results by using less data.

Table 1. Recognition accuracy under different combinations of descriptors and volumes of training sets on clothes pattern dataset. The number in parenthesis of each descriptor denotes its dimensionality.

Method	10%	30%	50%	70%
STA (240)	62.50%	74.63%	76.81%	79.3%
SIFT (128)	69.66%	80.70%	84.30%	85.0%
WSTA (960)	70.40%	81.36%	85.78%	85.89%
WSIFT (512)	74.52%	83.17%	86.80%	87.70%
WSIFT+WSTA(1472)	72.30%	83.76%	86.00%	88.89%
WSIFT*WSTA (42)	**76.46%**	**87.50%**	**88.94%**	**91.86%**

6. CONCLUSION

In this paper, we have proposed a confidence margin based feature combination method to combine multiple feature channels to recognize clothes patterns which have large intra-class variations. Multiple feature channels are constructed by STA descriptors and SIFT descriptors from original images and associated wavelet subbands. The combination of multiple feature channels provides complementary information to improve recognition accuracy. Furthermore, the proposed confidence margin based feature combination method is able to represent the combined feature in a more compact and more discriminative format. The future research will focus on collection and recognition of more categories of clothes patterns.

7. ACKNOWLEDGEMENT

This work was supported by NIH 1R21EY020990, NSF grant IIS-0957016 and EFRI-1137172, and ARO grant W911NF-09-1-0565.

8. REFERENCES

[1] Bishop C. Pattern Recognition and Machine Learning. 2007. *Springer*.

[2] Harwood D., Ojala T., Pietikainen M., Kelman S., and Davis L. 1995. Texture Classification by Center-symmetric Auto-Correlation Texture Classification by Center-symmetric Auto-Correlation Using Kullback Discrimination of Distributions. *Patter Recognition*.

[3] Lazebnik S., Schmid C., and Ponce J. 2005. A Sparse Texture Representation Using Local Affine Regions. *IEEE Trans. on Pattern Analysis and Machine Intelligence.*

[4] Li L., Pratap A., Lin H., and Abu-Mostafa Y. 2005. Improving Generalization by Data Categorization. *Knowledge Discovery in Databases*.

[5] Lowe D. 2004. Distinctive Image Features from Scale-invariant Keypoints. *International Journal of Computer Vision*.

[6] Mikolajczyk K. and Schmid C. 2005. A Performance Evaluation of Local Descriptors. *IEEE Trans. on Pattern Analysis and Machine Intelligence*.

[7] Nowak E., Jurie F., and Triggs B. 2006. Sampling Strategies for Bag-of-Features Image Classification. *European Conference on Computer Vision*.

[8] Randen T., Husoy J. 1999. Filtering for Texture Classification: A Comparative Study. *IEEE Trans. on Pattern Analysis and Machine Intelligence.*

[9] Kocur I., Parajasegaram R., and Pokharel G. 2004. Global Data on Visual Impairment in the Year 2002. *Bulletin of the World Health Organization*, 82,844-851.

[10] Vapnik V.N. 1995. The Nature of Statistical Learning Theory. *Springer-Verlag.*

[11] Wang Z. and Yong J. 2008. Texture Analysis and Classification with Linear Regression Model based on Wavelet Transform. *IEEE Trans. on Image Processing*.

[12] Yuan S., Tian, Y., Arditi A. 2011. Clothes Matching for Visually Impaired Persons. *Journal of Technology and Disability.*

[13] Zhang J., Marszalek M., Lazebnik S., and Schmid C. 2007. Local Features and Kernels for Classification of Texture and Object Categories: A Comprehensive Study. *International Journal of Computer Vision*.