Assistive Clothing Pattern Recognition for Visually Impaired People

Xiaodong Yang, *Student Member, IEEE*, Shuai Yuan, and YingLi Tian, *Senior Member, IEEE*

Abstract—Choosing clothes with complex patterns and colors is a challenging task for visually impaired people. Automatic clothing pattern recognition is also a challenging research problem due to rotation, scaling, illumination, and especially large intraclass pattern variations. We have developed a camera-based prototype system that recognizes clothing patterns in four categories (plaid, striped, patternless, and irregular) and identifies 11 clothing colors. The system integrates a camera, a microphone, a computer, and a Bluetooth earpiece for audio description of clothing patterns and colors. A camera mounted upon a pair of sunglasses is used to capture clothing images. The clothing patterns and colors are described to blind users verbally. This system can be controlled by speech input through microphone. To recognize clothing patterns, we propose a novel Radon Signature descriptor and a schema to extract statistical properties from wavelet subbands to capture global features of clothing patterns. They are combined with local features to recognize complex clothing patterns. To evaluate the effectiveness of the proposed approach, we used the CCNY Clothing Pattern dataset. Our approach achieves 92.55% recognition accuracy which significantly outperforms the state-of-the-art texture analysis methods on clothing pattern recognition. The prototype was also used by ten visually impaired participants. Most thought such a system would support more independence in their daily life but they also made suggestions for improvements.

Index Terms—Assistive system, clothing pattern recognition, global and local image features, texture analysis, visually impaired people.

I. 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]. Choosing clothes with suitable colors and patterns is a challenging task for blind or visually impaired people. They manage this task with the help from family members, using plastic braille labels or different types of stitching pattern tags on the clothes, or by wearing clothes with a uniform color or without any patterns.

Automatically recognizing clothing patterns and colors may improve their life quality. Automatic camera-based clothing pattern recognition is a challenging task due to many clothing pattern and color designs as well as corresponding large intraclass variations [4]. Existing texture analysis methods mainly focus on textures with large changes in viewpoint, orientation, and scaling, but with less intraclass pattern and intensity variations (see Fig. 1). We have observed that traditional texture analysis methods [3], [5], [10], [11], [15], [19], [23], [26], [29], [32] cannot achieve the same level of accuracy in the context of clothing pattern recognition.

Here, we introduce a camera-based system to help visually impaired people to recognize clothing patterns and colors. The system contains three major components (see Fig. 2): 1) sensors including a camera for capturing clothing images, a microphone for speech command input and speakers (or Bluetooth, earphone) for audio output; 2) data capture and analysis to perform command control, clothing pattern recognition, and color identification by using a computer which can be a desktop in a user's bedroom or a wearable computer (e.g., a mini-computer or a smartphone); and 3) audio outputs to provide recognition results of clothing patterns and colors, as well as system status.

In an extension to [30], our system can handle clothes with complex patterns and recognize clothing patterns into four categories (plaid, striped, patternless, and irregular) to meet the basic requirements based on our survey with ten blind participants. Our system is able to identify 11 colors: red, orange, yellow, green, cyan, blue, purple, pink, black, grey, and white. For clothes with multiple colors, the first several dominant colors are spoken to users. In order to handle the large intraclass variations, we propose a novel descriptor, Radon Signature, to capture the global directionality of clothing patterns. The combination of global and local image features significantly outperforms the state-of-the-art texture analysis methods for clothing pattern recognition. We also show that our method achieves comparable results to the state-of-the-art approaches on the traditional texture classification problems.

This paper is organized as follows. In Section II, we summarize the related work on assistive techniques for visually impaired people and the research work on texture analysis. The computations of global and local features for clothing pattern recognition are described in Section III. Section IV introduces the system and interface design. The details of clothing pattern recognition and color identification are demonstrated in Section V. Section VI presents our experimental results on a challenging clothing pattern dataset and a traditional texture dataset. Section VII describes the preliminary evaluations by blind users. Section VIII concludes the paper.

X. Yang and Y. L. Tian are with the City College, City University of New York, New York, NY 10031 USA (e-mail: xyang02@ccny.cuny.edu; ytian@ccny.cuny.edu).

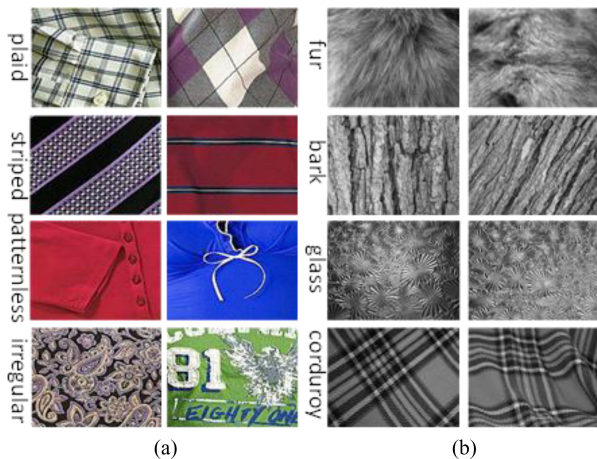S. Yuan is with Teledata Communications Inc., Islandia, NY 11749 USA (e-mail: syuan00@ccny.cuny.edu).

Fig. 1. Intraclass variations in clothing pattern images and traditional texture images. (a) Clothing pattern samples with large intraclass pattern and color variations. (b) Traditional texture samples with less intraclass pattern and intensity variations.
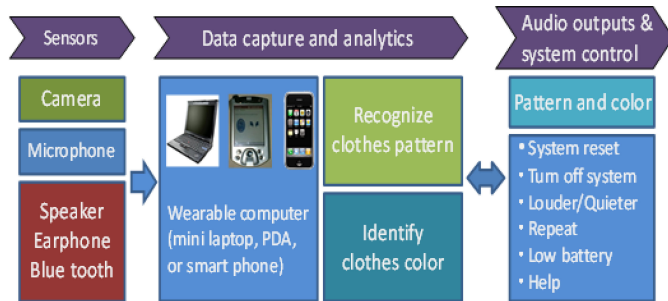

Fig. 2. Overview and architecture design of the camera-based clothing pattern recognition system for blind and visually impaired persons.

II. RELATED WORK

Assistive systems are being developed to improve the life quality and safety for those with special needs [2], [6], [7], [16], [20]–[22], [27], [28], [30], [31], including indoor navigation and way finding, display reading, banknote recognition, rehabilitation, etc. Liu *et al.* [12] built a clothing recommendation system for specific occasions (e.g., wedding or dating). Hidayati *et al.* [14] proposed a method for genre classification of upper-wear clothes. The two systems are both designed without considering key factors for blind users [1]. Yuan *et al.* [31] developed a system to assist blind people to match clothes from a pair of clothing images. This system can provide a user with the information about whether or not the clothing patterns and colors match. However, this system is not able to automatically recognize clothing patterns.

Texture provides essential information for many image classification tasks including clothing pattern recognition. Some early research on texture recognition [3], [5], [10], [15], [19], [23] focused on the analysis of global two-dimensional (2-D) image transformations including in-plane rotation and scaling. Due to the lack of invariance to general geometric transformations, these approaches cannot effectively represent texture images with large 3-D transformations such as viewpoint change and nonrigid surface deformation. Multifractal analysis [25], [26]

has achieved good resilience to 3-D deformations. Texture representations based on this method benefit from the invariance of fractal dimensions to geometric transformations. For example, multi-fractal spectrum (MFS) proposed by Xu *et al.* [26] combined fractal dimensions of pixel sets grouped by density functions and orientation templates. In order to make representations of texture more robust to 3-D image transformations (e.g., viewpoint change and nonrigid surface deformation) and illumination variation, most of the recent methods reply on extracting local image features [11], [29], [32]. A texton dictionary is then generated by clustering the extracted local features. However, multiple features are able to capture properties of an image in different aspects. If different features are highly complementary, their combination will improve the feature representation. For example, Lazebnik *et al.* [11] proposed a texture representation method based on affine-invariant detectors (Harris and Laplacian) and descriptors (RIFT and SPIN). Zhang *et al.* [32] also combined scale invariant feature transform (SIFT) and SPIN for texture classification.

Unlike existing traditional texture images, clothing patterns contain much larger intraclass variations within each pattern category. Although many computer vision and image processing techniques have been developed for texture analysis and classification, traditional texture analysis methods cannot effectively recognize clothing patterns. Here, we develop a camera-based system specifically for visually impaired people to help them recognize clothing patterns and colors.

III. IMAGE FEATURE EXTRACTION FOR CLOTHING PATTERN RECOGNITION

Some clothing patterns present as visual patterns characterized by the repetition of a few basic primitives (e.g., plaids or stripes). Accordingly, local features are effective to extract the structural information of repetitive primitives. However, due to large intraclass variance, local primitives of the same clothing pattern category can vary significantly (see Fig. 1). Global features including directionality and statistical properties of clothing patterns are more stable within the same category. Therefore, they are able to provide complementary information to local structural features. Next, we present extractions of global and local features for clothing pattern recognition, i.e., Radon Signature, statistical descriptor (STA), and scale invariant feature transform (SIFT).

A. Radon Signature

Clothing images present large intraclass variations, which result in the major challenge for clothing pattern recognition. However, in a global perspective, the directionality of clothing patterns is more consistent across different categories and can be used as an important property to distinguish different clothing patterns. As shown in Fig. 8, the clothing patterns of plaid and striped are both anisotropic. In contrast, the clothing patterns in the categories of patternless and irregular are isotropic. To make use of this difference of directionality, we propose a novel descriptor, i.e., the Radon Signature, to characterize the directionality feature of clothing patterns.
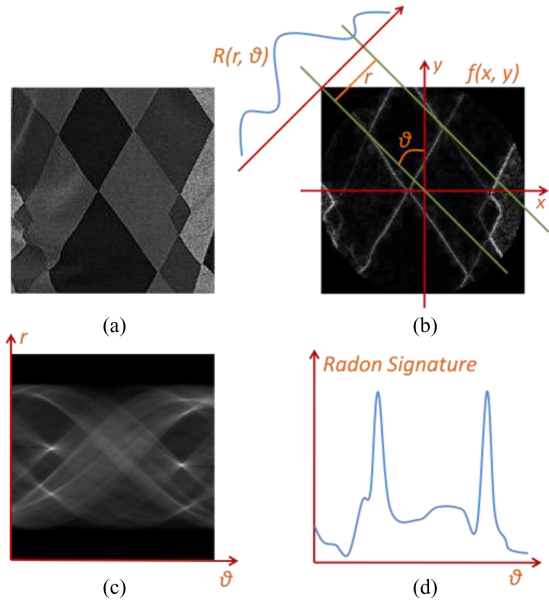
Fig. 3.   Computation of RadonSig. (a) An intensity image of clothing pattern. (b) Radon transform performed on a maximum disk area within the gradient map of (a). (c) Result of Radon Transform. (d) Feature vector of Radon Signature.

Fig. 4.   Clothing patterns samples and associated RadonSig descriptors.

Radon Signature (RadonSig) is based on the Radon transform [8] which is commonly used to detect the principle orientation of an image. The image is then rotated according to this dominant direction to achieve rotation invariance. The Radon transform of a 2-D function $f(x, y)$ is defined as

$$R(r, \theta) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) \, \delta(r - x \cos \theta - y \sin \theta) \, dx dy$$
(1)

where r is the perpendicular distance of a projection line to the origin and θ is the angle of the projection line, as shown in Fig. 3(b). To retain the consistency of Radon transform for different projection orientations, we compute the Radon transform based on the maximum disk area instead of the entire image. The large intraclass variations of clothing patterns also reflect as images in the same category present large changes of color or intensity. To reduce the intensity variations, we use the Sobel operator to compute the gradient map as $f(x, y)$ in (1). Fig. 3(b) illustrates the Radon transform over a disk area of gradient map. $R(r, \theta)$ in (1) is a function with two parameters of r and θ, as shown in Fig. 3(c). The directionality of an image can be represented by $\text{Var}(r, \theta_i)$, the variances of r under a certain projection direction θ_i:

$$\text{Var}(r, \theta_i) = \frac{1}{N} \sum_{j=0}^{N-1} (R(r_j, \theta_i) - \mu(r, \theta_i))^2$$
(2)

$$\mu(r, \theta_i) = \frac{1}{N} \sum_{j=0}^{N-1} R(r_j, \theta_i)$$
(3)

where $R(r_j, \theta_i)$ is the projection value at perpendicular distance of r_j and projection direction of θ_i; $\mu(r, \theta_i)$ is the expected value of $R(r, \theta_i)$; N is the number of sampling bins in each projection line. The RadonSig is formed by the variances of r under all
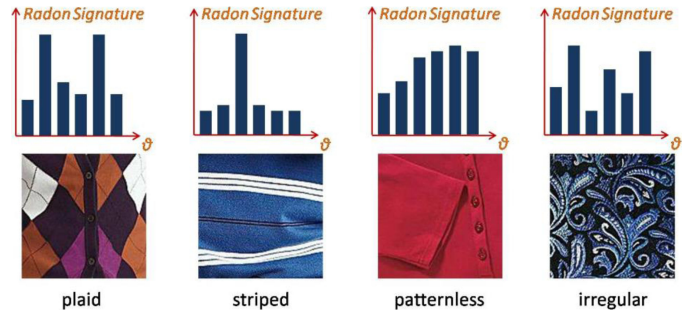
sampling projection directions:

$$[\text{Var}(r, \theta_0), \text{Var}(r, \theta_1), \ldots, \text{Var}(r, \theta_{T-1})]$$

where T is the number of sampling projection directions. It determines the feature dimension of Radon Signature. As the RadonSig is a nonsparse representation, we employ the L2-norm to normalize the feature vector. The principle directions of the image in Fig. 3(a) correspond to the two dominant peaks in the RadonSig in Fig. 3(d).

Fig. 4 illustrates RadonSig descriptors of four sample images from different clothing pattern categories. The plaid patterns have two principle orientations; the striped ones have one principle orientation; as for the patternless and the irregular images, they have no obvious dominant direction, but the directionality of the irregular image presents much larger variations than that of the patternless image. Accordingly, there are two dominant peak values corresponding to two principle orientations in the RadonSig of the plaid image. The RadonSig of the striped image has one peak value associated with the one principle orientation. There is no dominant peak value in the irregular and the patternless cases. But the RadonSig of the patternless image is much smoother than that of the irregular image.

B. Statistics of Wavelet Subbands

The discrete wavelet transform (DWT) decomposes an image I into low-frequency channel $D_j(I)$ under a coarser scale and multiple high-frequency channels under multiple scales $W_{k,j}(I)$; $k = 1, 2, 3$; $j = 1, 2, \ldots, J$, where J is the number of scaling levels. Therefore, in each scaling level j, we have four wavelet subbands including one low-frequency channel $D_j(I)$ and three high-frequency channels $W_{k,j}(I)$. The high-frequency channels $W_{k,j}(I)$; $k = 1, 2, 3$ encode the discontinuities of an image along horizontal, vertical, and diagonal directions, respectively. In this paper, we apply $J = 3$ scaling levels of DWT to decompose each clothing image, as shown in Fig. 5.

Statistical features are well adapted to analyze textures which lack background clutter and have uniform statistical properties. DWT provides a generalization of a multiresolution spectral analysis tool. Therefore, we extract the statistical features from wavelet subbands to capture global statistical information of images at different scales. It is customary to compute the single energy value on each subband [24]. In this paper, we employ
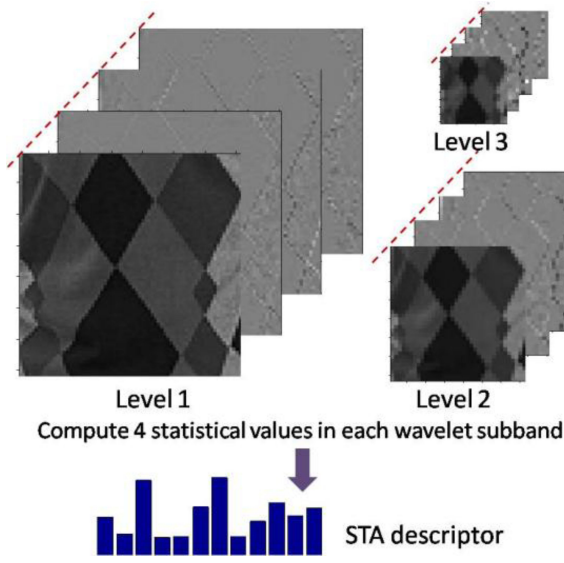
Fig. 5. The computation of STA on wavelet subbands. Three levels of wavelet decomposition are applied to a clothing image. Each decomposition level includes four wavelet subbands of original, horizontal, vertical, and diagonal components arranged from the close to the distant in each level. Four statistical values calculated in each wavelet subband are concatenated to form the final descriptor.

four statistical values including variance, energy, uniformity, and entropy to all wavelet subbands. Thus, the STA is a feature with the dimension of 48 ($3 \times 4 \times 4$). The four normalized statistical values extracted from each wavelet subband can be computed by the following equations:

$$\text{variance} = \sum_{i=0}^{L-1} (z_i - m)^2 \, p(z_i) / (L-1)^2 \quad (4)$$

$$\text{energy} = \sum_{i=0}^{L-1} (z_i - m)^3 \, p(z_i) / (L-1)^2 \quad (5)$$

$$\text{uniformity} = \sum_{i=0}^{L-1} p^2(z_i) \quad (6)$$

$$\text{entropy} = -\sum_{i=0}^{L-1} p(z_i) \log_2 p(z_i) \quad (7)$$

where z_i and $p(z_i)$, $i = 0, 1, 2, \ldots, L-1$ is the intensity level and corresponding histogram; L is the number of intensity levels; $m = \sum_{i=0}^{L-1} z_i p(z_i)$ is the average intensity level.

C. Scale Invariant Feature Transform Bag of Words

The local image features are well adapted to a number of applications, such as image retrieval, and recognition of object, texture, and scene categories [32], as they are robust to partial occlusion, cluttered background, and viewpoint variations. This has motivated the development of several local image feature detectors and descriptors. Generally, detectors are used to detect interest points by searching local extrema in a scale space; descriptors are employed to compute the representations of interest points based on their associated support regions. In this
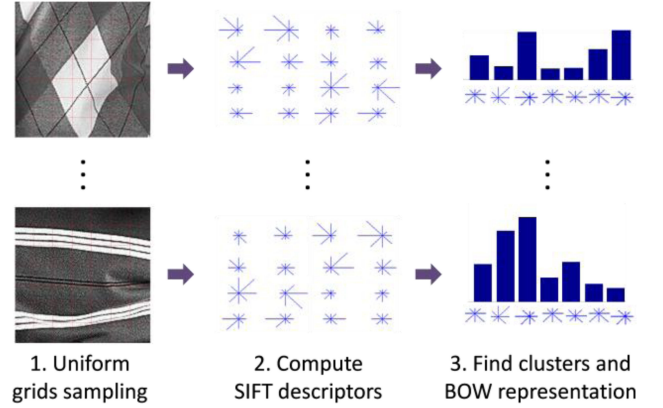

Fig. 6. Process of local image feature extraction.

paper, the uniform grids are used as the interest points sampling strategy, as more sophisticated detectors tend to saturate and fail to provide enough interest points, especially for the textureless images [18]. The evenly sampled interest points are then represented by SIFT descriptors.

We choose the SIFT descriptor as the representation of interest points based on the following reasons: 1) the descriptor with 128 dimensions is compact and fairly distinctive; 2) the representation with careful design is robust to variations in illumination and viewpoints; 3) an extensive comparison against other local image descriptors observed that the SIFT descriptor performed well in the context of image matching [17]. The bag-of-words (BOW) [18] method is further applied to aggregate extracted SIFT descriptors by labeling each SIFT descriptor as a visual word and counting frequencies of each visual world. The local feature representation of an image is therefore represented as the histogram of the quantized SIFT descriptors. We perform L2-norm and inverse document frequency (IDF) normalization of BOW histograms. Fig. 6 demonstrates the process of local features extraction.

IV. SYSTEM AND INTERFACE DESIGN

The camera-based clothing recognition aid prototype for blind people integrates a camera, a microphone, a computer, and a Bluetooth earpiece for audio description of clothing patterns and colors. A camera mounted upon a pair of sunglasses is used to capture clothing images. The clothing patterns and colors are described to blind users by a verbal display with minimal distraction to hearing. The system can be controlled by speech input through a microphone.

In order to facilitate blind users to interact, speech commands input from a microphone are used to provide function selection and system control. As shown in Fig. 7, the interface design includes *basic functions* and *high priority commands*.

Basic functions: A blind user can verbally request the function he/she wants the clothing recognition aid to perform. The recognition results will be presented to the blind user as audio outputs including *recognized*, *not recognized*, and *start a new function*. As for the *recognized* function, the next level functions include *pattern/colors* to announce the recognized clothing
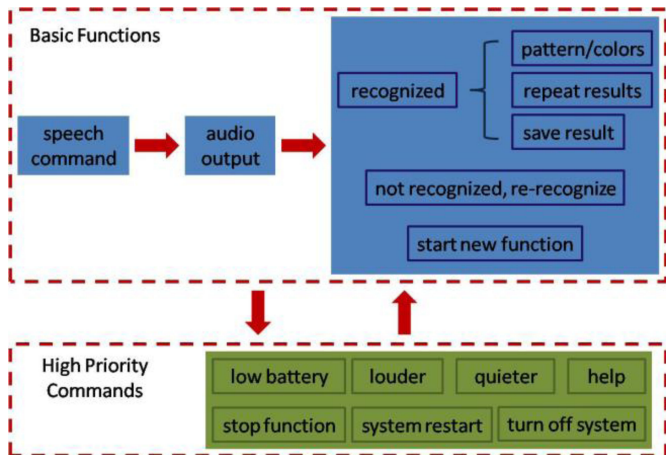
Fig. 7. System interface design for the proposed camera-based clothing pattern recognition system by using speech commands. The high priority commands can be used at any time to overwrite the basic functions.

TABLE I
CLOTHING PATTERNS AND DOMINANT COLORS

Image				
Pattern	plaid	striped	patternless	irregular
Color	yellow(49%) orange(36%) black(9%)	blue(75%) white(19%)	red(98%)	black(41%) red(26%) blue(6%) green(5%)

The complementary information provides more complete descriptions of clothing images.

pattern and dominant colors; *repeat results* to repeat the recognized result; and *save result* to save the clothing image with associated pattern and color information in the computer.

High priority commands: A blind user can set the system configuration by several high priority speech commands such as *system restart*, *turn-off system*, *stop function* (i.e., abort current task), speaker volume and speed control commands (e.g., *louder*, *quieter*, *slower*, and *faster*), and *help*. The high priority commands can be used at any time. A user can speak *help*, and the clothing recognition system will respond with the options associated with the current function. Bone conducted earphones or small wireless Bluetooth speakers can be employed to protect privacy and minimize background sounds. The battery level will also be checked and an audio warning is provided if the battery level is low.

Audio output: As for audio display, we use an operating system speech facility that is standard in modern portable computer systems and smartphones. We currently use Microsoft Speech Software Development Kit which supports scripts. A number of configuration options are also available according to user preference, such as speech rate, volume, and voice gender.

V. RECOGNIZING CLOTHING PATTERNS AND COLORS

The extracted global and local features are combined to recognize clothing patterns by using a support vector machines (SVMs) classifier. The recognition of clothing color is implemented by quantizing clothing color in the HIS (hue, saturation, and intensity) space. In the end, the recognition results of both clothing patterns and colors mutually provide a more precise and meaningful description of clothes to users.

A. Clothing Pattern Recognition

In our system, we empirically set the size of the visual vocabulary to 100. We apply three scaling levels to decompose clothing images. The statistics of wavelet subbands features are therefore formed by a vector with a dimension of 48. In the computation of the Radon Signature, we evenly sample 60 pro-

jection directions from $1°$ to $180°$. The feature vector of the RadonSig has a dimension of 60. We combine all the global and local features by concatenating corresponding feature channels together. The combined feature vector has a dimension of 208.

The combined feature vector is used as the inputs of SVMs classifier with RBF kernel. In our experiments, the optimal parameters of RBF kernel are found by 5-fold cross-validation, and the one-versus-one scheme is used.

B. Clothing Color Identification

Clothing color identification is based on the normalized color histogram of each clothing image in the HSI color space. The key idea is to quantize color space based on the relationships between hue, saturation, and intensity. In particular, for each clothing image, our color identification method quantizes the pixels in the image to the following 11 colors: red, orange, yellow, green, cyan, blue, purple, pink, black, grey, and white.

The detection of colors of white, black, and gray is based on the saturation value S and intensity value I. If the intensity I of a pixel is larger than a upper intensity threshold I^U, and the saturation S is less than a saturation threshold S_T, the color of the pixel is "white." Similarly, the color of a pixel is determined to be black if the intensity I is less than a lower intensity bound I^L and saturation S is less than S_T. For the remaining values of I while S is less than S_T, the color of a pixel is identified as gray. For other colors (i.e., red, orange, yellow, green, cyan, blue, purple, and pink), the hue values are employed. The hue H can be visualized as a $360°$ color wheel. We quantize the color of red in the range of $345° - 360°$ and $0° - 9°$, orange as $10° - 37°$, yellow as $38° - 75°$, green as $76° - 160°$, cyan as $161° - 200°$, blue as $201° - 280°$, purple as $281° - 315°$, and pink as $316° - 344°$. The weight of each color is the percentage of pixels belonging to this color.

If a clothing image contains multiple colors, the dominant colors (i.e., pixels larger than 5% of the whole image) will be output. The clothing patterns and colors mutually provide complementary information. As shown in Table I, the recognized patterns provide additional information about how different colors are arranged, e.g., striped clothes with blue and white color. We test color identification in a clothing color matching experiment. If the dominant colors of a pair of clothing image are the same, the two clothing images are determined as color matched. The proposed color identification achieves 99% matching accuracy in the experiment. More details can be found in [31].

Fig. 8. Four sample images of four clothing patterns categories.

Fig. 9. Four sample images of four sample texture classes in the UIUC Texture dataset.

VI. CLASSIFICATION EXPERIMENTS

In this section, we evaluate the performance of the proposed method on two different datasets: 1) the CCNY Clothing Pattern dataset with large intra-class variations to evaluate our proposed method and the state-of-the-art texture classification methods, and 2) the UIUC Texture dataset to validate the generalization of the proposed approach. Our experiments focus on the evaluation and validation of 1) the complementary relationships between the proposed global and local feature channels; 2) the superiority of our proposed method over the state-of-the-art texture classification approaches in the context of clothing pattern recognition; and 3) the generalization of our approach on the traditional texture classification.

A. Datasets

CCNY clothing pattern dataset: This dataset includes 627 images of four different typical clothing pattern designs: plaid, striped, patternless, and irregular with 156, 157, 156, and 158 images in each category. The resolution of each image is down-sampled to 140 × 140. Fig. 8 illustrates sample images in each category. As shown in this figure, in addition to illumination variances, scaling changes, rotations, and surface deformations presented in the traditional texture dataset, clothing patterns also demonstrate much larger intraclass pattern and color (intensity) variations, which augment the challenges of recognition. The clothing pattern dataset can be downloaded via our research website.[1]

UIUC texture dataset: We also evaluate the proposed method on the UIUC Texture dataset [11] which is a well-established traditional texture dataset. It contains 1000 uncalibrated and unregistered images with the resolution of 640 × 480. There are 25 texture classes with 40 images for each class. The texture images present rotation, scaling and viewpoint change, and non-rigid surface deformation. Fig. 9 shows four sample images of four texture classes in this dataset. As shown in Fig. 8 and Fig. 9, the traditional texture dataset mainly focuses on the geometric
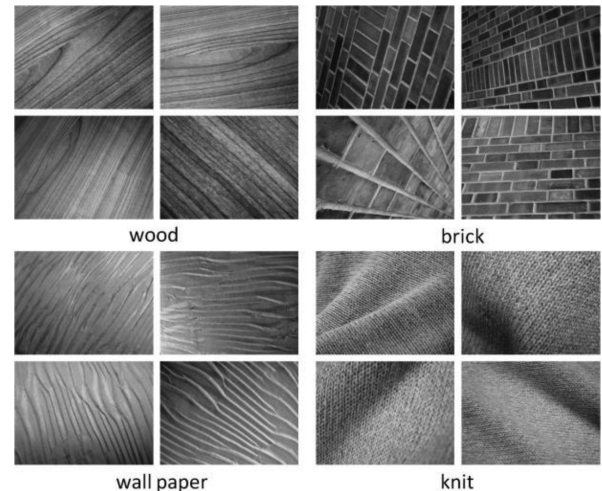
[1]media-lab.engr.ccny.cuny.edu/data

changes of texture surfaces, but with less intraclass pattern and intensity variations compared to the clothing images.

B. Experiments and Discussions on Clothing Pattern Recognition

1) Experimental Setup: In our implementation, the training set is selected as a fixed-size random subset of each class and all remaining images are used as the testing set. To eliminate the dependence of the results on the particular training images used, we report the average of the classification rates obtained for 50 randomly selected training sets. This ensures all the classification rates are statistically comparable. The recognition performance is measured by the average classification accuracy.

2) Effectiveness of Different Features and Combinations: We first evaluate and demonstrate the complementary relationships between global and local features on clothing pattern images. A combination of multiple features may obtain better results than any individual feature channel. However, a combination of features that are noisy, contradictory, or overlapping in terms of class distribution could deteriorate the performance of classification.

To validate the effectiveness of the proposed features, we first evaluate the complementary relationships between different feature channels including global features of the RadonSig and statistics of wavelet subbands (STA), and local features (SIFT). SIFT represents the local structural features; STA is the global statistical characteristics; and RadonSig captures the property of global directionality. Fig. 10 displays the recognition results of different features as a function of training set size. For individual feature channels, SIFT and STA achieve comparable recognition accuracies. While the results based on a single channel of the RadonSig are worse than that of SIFT or STA, the performance of SIFT+RadonSig is better than that of SIFT+STA. Both of them outperform any individual feature channel. Therefore, for clothing patterns recognition, the global and local feature combination of SIFT and RadonSig is more effective than that of SIFT and STA. Furthermore, the
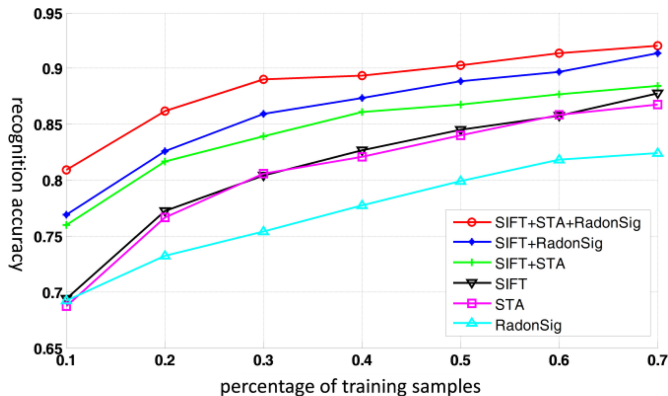
Fig. 10. Comparative evaluation on recognition accuracies of different feature channels and their combinations versus percentage of training samples.

TABLE II
RECOGNITION ACCURACY (%) OF DIFFERENT FEATURE CHANNELS UNDER DIFFERENT VOLUMES OF TRAINING SETS (10%, 30%, 50%, 70% OF THE DATASET USED FOR TRAINING) ON CLOTHING PATTERN DATASET

Feature channel	10%	30%	50%	70%
SIFT	69.42	80.45	84.50	87.73
STA	69.45	80.65	84.28	86.94
RadonSig	69.24	75.41	79.93	82.44
SIFT+STA	75.80	84.03	87.09	88.68
SIFT+RadonSig	76.94	85.91	88.89	91.34
SIFT+STA+RadonSig	**81.06**	**88.09**	**90.59**	**92.55**

combination of all three feature channels further improves the recognition results and dominates in all of different training set sizes.

The comparisons of different feature channels and their combinations validate our intuition that the effectiveness and complementarities of our proposed feature channels. The detailed recognition accuracies of Fig. 10 are listed in Table II. The percentages of training images per class are 10%, 30%, 50%, and 70%, respectively. As shown in Table II and Fig. 10, the recognition accuracy of SIFT+STA+RadonSig using 30% of the images as the training set is comparable or even better than that of other feature channels using 70% of the images as the training set. This observation demonstrates another merit of our proposed approach that it is able to achieve a desirable result by using much less training data.

3) Comparison With the State-of-the-art Texture Analysis Methods for Clothing Pattern Recognition: We further compare the overall performance of our proposed method with that of the state-of-the-art methods including MFS [26], SIFT [13], (H+L)(S+R) [11], and SIFT+SPIN [32], which have achieved the state-of-the-art performances on the traditional texture classification tasks. MFS [26] is an extension of the fractal dimension based on three density functions of image intensity, image gradient, and image Laplacian. It combines both global spatial invariance and local robust measurement. SIFT [13] corresponds to the single-feature channel in our method. It captures the local image structural information. (H+L)(S+R) [11] is based on the extraction of SPIN and RIFT descriptors on affine Harris and Laplacian regions of an image. The elliptic regions detected by two sophisticated detectors are normalized to circles

TABLE III
RECOGNITION ACCURACY (%) OF DIFFERENT METHODS AND VOLUMES OF TRAINING SETS (10%, 30%, 50%, 70% OF THE DATASET USED FOR TRAINING) ON CLOTHING PATTERN DATASET

Method	10%	30%	50%	70%
MFS [26]	56.71	66.68	72.39	75.80
SIFT [13]	69.42	80.45	84.50	87.73
(H+L)(S+R) [11]	52.71	60.34	62.65	64.69
SIFT+SPIN [32]	66.25	78.76	83.20	85.53
Our method	**81.06**	**88.09**	**90.59**	**92.55**

to handle affine transform. SPIN is a rotation-invariant local image descriptor formed by a 2-D histogram of the distribution of pixel intensity values. RIFT is an alternate representation of SIFT. SIFT+SPIN [32] is the combination of two local image features including SIFT and SPIN. In the comprehensive evaluations of texture classification [32], SIFT and SIFT+SPIN organized by the BOW model achieved the state-of-the-art results on traditional texture datasets. In our implementation, the sampling strategies for SIFT and SIFT+SPIN are both uniform grid sampling as mentioned in Section III-C.

Table III shows the recognition accuracies of different methods on the CCNY clothing pattern dataset. The experiments are evaluated by using 10%, 30%, 50%, and 70% of the dataset as training sets, and the rest as testing sets. As shown in this table, our proposed method significantly outperforms other well-established approaches, especially when the training set is small. A closer look at Table III also confirms that our proposed method is able to obtain comparable or even better results by using much less training data. For instance, the accuracy rate of our method using 30% of the images as a training set is better than that of other well-established methods using 70% as a training set. While these methods perform very well on traditional texture datasets, they cannot achieve the same level of accuracy for clothing pattern recognition due to the large intraclass variations. (H+L)(S+R) performs much worse than other methods. This is mainly because of the insufficient interest points detected by two sophisticated detectors, especially for striped and patternless categories. It also confirms the importance of sampling density for clothing pattern classification. It is also interesting to observe that the performance of SIFT+SPIN is worse than that of SIFT alone. This is probably because SPIN is a local image descriptor based on pixel intensities. But the clothing images present large intraclass color variations, which accordingly result in large intraclass intensity variations. Since SPIN is sensitive to intensity changes, it cannot overcome the large intraclass variations. This observation validates that it is important to choose appropriate feature channels for feature combination.

Table IV demonstrates the confusion table results of our method by using 70% of the images as the training set. In the confusion table, each row represents the ground-truth categories of clothing pattern images and each column corresponds to the recognized category. The system recognizes patternless as the best, where the images have the most discriminative local and global properties. On the other hand, plaid tends to confuse with striped.

TABLE IV
CONFUSION TABLE FOR THE CLOTHING PATTERN RECOGNITION EXPERIMENTS
OF OUR PROPOSED METHOD USING 70% OF THE IMAGES AS THE TRAINING SET

	plaid	striped	patternless	irregular
plaid	**0.8855**	0.0494	0.0149	0.0502
striped	0.0383	**0.9234**	0.0123	0.0260
patternless	0.0094	0.0183	**0.9570**	0.0094
irregular	0.0260	0.0451	0.0119	**0.9170**

Each row is the ground truth and each column is the classified results. The average performance is 92.07%.

TABLE V
CLASSIFICATION ACCURACY (%) OF DIFFERENT METHODS AND NUMBER OF
TRAINING IMAGES ON UIUC TEXTURE DATASET

Method	5	10	15	20
MFS [26]	82.24	88.36	91.38	92.74
SIFT [13]	91.13	95.46	96.94	97.56
(H+L)(S+R) [11]	91.12	94.42	96.64	97.02
SIFT+SPIN [32]	93.13	96.87	97.74	98.83
Our method	88.04	94.51	96.31	97.61

C. Experiments With the UIUC Texture Dataset

The proposed method significantly outperforms the traditional texture classification methods in recognizing clothing patterns. In order to validate the generalization of our method, we further compare our method with other state-of-the-art approaches on a traditional texture dataset, i.e., UIUC Texture dataset. The experimental setup is the same as Section VI-B1.

In the performance evaluation on the UIUC Texture dataset, we also compare our proposed method with MFS, SIFT, (H+L)(S+R), and SIFT+SPIN, as mentioned in Section VI-B3. Table V shows the classification rates of different methods on the UIUC Texture dataset. In the experiments, the numbers of training images per class are 5, 10, 15, and 20, respectively. The remaining images in each class are used as the testing set. As shown in this table, our method achieves comparable results to SIFT, (H+L)(S+R), and SIFT+SPIN, and obtain better accuracy rates than MFS. In contrast to the performance of clothing pattern recognition, (H+L)(S+R) performs very well on this dataset as the two sophisticated detectors can localize sufficient interest points on highly textured images. This result again demonstrates the importance of sampling density for robustness of texture classification. SIFT+SPIN performs the best on UIUC Texture dataset while its performance on clothing pattern images is much worse. This is because the intraclass intensities of texture images in this dataset are much more consistent than that in clothing pattern images. SPIN provides complementary appearance feature to SIFT. However, when the number of training images is 5, SIFT+SPIN largely outperforms our method. This is probably because the directionality property in nature texture is not as evident as that in clothing images and the RadonSig needs a relatively large training set to model the directionality of texture images. But in other cases, our method achieves comparable results to SIFT+SPIN. Therefore, it demonstrates the generalization of our method, i.e., it achieves state-of-the-art results on both the clothing pattern images and the traditional texture dataset.

Fig. 11. Example setup of clothing pattern recognition prototype to capture images by a web camera with auto focus on sunglasses.

VII. PROOF-OF-CONCEPT EVALUATION

The proposed clothing pattern and color recognition prototype were tested in a proof-of-concept evaluation with ten blind participants including five men and five women with ages ranging from 36 to 72 years old. The testing was conducted in a laboratory under normal lighting conditions. Because some of the totally blind users are not aware of the environment lighting conditions, the system provides a reminder if the environment is too dark before the testing starts.

As shown in Fig. 11, the clothing images are captured by a camera (Logitech web camera with auto focus) mounted on a pair of sunglasses where the blind user holds the clothes to be recognized. The images captured by people with normal vision are similar to those captured by blind participants after they are trained for several minutes to hold clothes in front of the camera to occupy the image view. The images of the clothes are captured and recognized by the system. The system provides the clothing patterns and verbally presents the three dominant colors to the user.

The number of clothing patterns and colors recognized by the system are less than what humans can provide. In addition, the audio description is not as flexible as what humans can state. However, most blind users expressed wanting such a system to support more independence in their daily life. They also expressed the desire of more detailed colors such as rose red and more clothing patterns. The blind users expressed a desire for faster speech feedback in order to gain more information. Some blind participants expressed a desire to have the camera on a cap instead of on glasses and to have the function available on mobile phones.

VIII. CONCLUSION

Here, we have proposed a system to recognize clothing patterns and colors to help visually impaired people in their daily life. We employ RadonSig to capture the global directionality features; STA to extract the global statistical features on wavelet subbands; and SIFT to represent the local structural features. The combination of multiple feature channels provides complementary information to improve recognition accuracy. Based on a survey and a proof-of-concept evaluation with blind users, we have collected a dataset on clothing pattern recognition

including four-pattern categories of plaid, striped, patternless, and irregular. Experimental results demonstrate that our proposed method significantly outperforms the state-of-the-art methods in clothing pattern recognition. Furthermore, the performance evaluation on traditional texture datasets validates the generalization of our method to traditional texture analysis and classification tasks. This research enriches the study of texture analysis, and leads to improvements over existing methods in handling complex clothing patterns with large intraclass variations. The method also provides new functions to improve the life quality for blind and visually impaired people.

REFERENCES

[1] A. Arditi and Y. Tian, "User interface preferences in the design of a camera-based navigation and wayfinding aid," *J. Visual Impairment Blindness*, vol. 107, no. 2, pp. 18–129, 2013.

[2] D. Dakopoulos and N. G. Bourbakis, "Wearable obstacle avoidance electronic travel aids for the blind: A survey," *IEEE Trans. Syst., Man, Cybern. C*, vol. 40, no. 1, pp. 25–35, Jan. 2010.

[3] L. Davis, S. Johns, and J. Aggarwal, "Texture analysis using generalized co-occurrence matrices," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-1, no. 3, pp. 251–259, Jul. 1979.

[4] D. Gould, "The making of a pattern," *Vogue Patterns*, 1996.

[5] R. Haralick, "Statistical and structural approaches to texture," *Proc. IEEE*, vol. 67, no. 5, pp. 786–804, May 1979.

[6] F. Hasanuzzaman, X. Yang, and Y. Tian, "Robust and effective component-based banknote recognition for the blind," *IEEE Trans. Syst., Man, Cybern. C*, vol. 42, no. 6, pp. 1021–1030, Nov. 2012.

[7] A. Huete, J. Victores, S. Martinez, A. Gimenez, and C. Balaguer, "Personal autonomy rehabilitation in home environment by a protable assistive robot," *IEEE Trans. Syst., Man, Cybern. C*, vol. 42, no. 4, pp. 561–570, Jul. 2012.

[8] K. Khouzani and H. Zaden, "Radon transform orientation estimation for rotation invariant texture analysis," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 6, pp. 1004–1008, Jun. 2005.

[9] I. Kocur, R. Parajasegaram, and G. Pokharel, "Global data on visual impairment in the year 2002," *Bulletin World Health Org.*, 2004.

[10] S. Lam, "Texture feature extraction using gray level gradient based on co-occurrence matrices," in *Proc. Int. Conf. Syst., Man Cybern.*, 1996, pp. 267–271.

[11] S. Lazebnik, C. Schmid, and J. Ponce, "A sparse texture representation using local affine regions," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 8, pp. 1265–1277, Aug. 2005.

[12] S. Liu, J. Feng, Z. Song, T. Zhang, H. Lu, C. Xu, and S. Yuan, "Hi, magic closet, tell me what to wear," in *Proc. ACM Multimedia*, 2012.

[13] D. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Version*, vol. 60, no. 2, pp. 91–110, 2004.

[14] S. Hidayati, W. Cheng, and K. Hua, "Clothing genre classification by exploiting the style elements," in *Proc. ACM Int. Conf. Multimedia*, 2012, pp. 1137–1140.

[15] V. Manian, R. Vasquez, and P. Katiyar, "Texture classification using logical operation," *IEEE Trans. Image Process.*, vol. 9, no. 10, pp. 1693–1703, Oct. 2000.

[16] R. Manduchi and J. Coughlan, "(Computer) vision without sight," *Commun. ACM*, vol. 55, no. 1, pp. 96–104, 2012.

[17] K. Mikolajczyk and C. Schmid, "A performance evaluation of local descriptors," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 10, pp. 1615–1630, Oct. 2005.

[18] E. Nowak, F. Jurie, and B. Triggs, "Sampling strategies for bag-of-features image classification," in *Proc. Eur. Conf. Comput. Vis.*, 2006, pp. 490–503.

[19] T. Randen and J. Husoy, "Filtering for texture classification: A comparative study," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 21, no. 4, pp. 291–310, Apr. 1999.

[20] S. Shoval, J. Borenstein, and Y. Koren, "Auditory guidance with the navbelt—A computerized travel aid for the blind," *IEEE Trans. Syst., Man, Cybern. C*, vol. 28, no. 3, pp. 459–467, Aug. 1998.

[21] E. Tekin, J. Coughlan, and H. Shen, "Real-time detection and reading of LED/LCD displays for visually impaired persons," in *Proc. IEEE Workshop Appl. Comput. Vision*, 2011, pp. 491–496.

[22] Y. Tian, X. Yang, C. Yi, and A. Arditi, "Toward a computer vision based wayfinding aid for blind persons to access unfamiliar indoor environments," *Mach. Vis. Appl.*, vol. 24, no. 3, pp. 521–535, 2012.

[23] M. Varma and A. Ziisserman, "Texture classification: Are filter banks necessary," in *Proc. Comput. Vis. Pattern Recog.*, 2003, pp. 691–698.

[24] Z. Wang and J. Yong, "Texture analysis and classification with linear regression model based on wavelet transform," *IEEE Trans. Image Process.*, vol. 17, no. 8, pp. 1421–1430, Aug. 2008.

[25] H. Wendt, P. Abry, S. Jaffard, H. Ji, and Z. Shen, "Wavelet leader multifractal analysis for texture classification," in *Proc. Int. Conf. Image Process.*, 2009, pp. 3829–3832.

[26] Y. Xu, H. Ji, and C. Fermuller, "Viewpoint invariant texture description using fractal analysis," *Int. J. Comput. Vis.*, vol. 83, no. 1, pp. 85–100, 2009.

[27] X. Yang and Y. Tian, "Robust door detection in unfamiliar environment by combining edge and corner features," in *Proc. IEEE CVPR Workshop Comput. Vis. Appl. Visually Impaired*, 2010, pp. 57–64.

[28] X. Yang, Y. Tian, C. Yi, and A. Arditi, "Context-based indoor object detection as an aid to blind persons accessing unfamiliar environments," in *Proc. Int. ACM Conf. Multimedia*, 2010, pp. 1087–1090.

[29] X. Yang and Y. Tian, "Texture representations using subspace embeddings," *Pattern Recog. Lett.*, vol. 34, no. 10, pp. 1130–1137, 2013.

[30] X. Yang, S. Yuan, and Y. Tian, "Recognizing clothes patterns for blind people by confidence margin based feature combination," in *Proc. Int. ACM Conf. Multimedia*, 2011, pp. 1097–1100.

[31] S. Yuan, Y. Tian, and A. Arditi, "Clothes matching for visually impaired persons," *J. Technol. Disability*, vol. 23, no. 2, pp. 75–85, 2011.

[32] J. Zhang, M. Marszalek, S. Lazebnik, and C. Schmid, "Local features and kernels for classification of texture and object categories: A comprehensive study," *Int. J. Comput. Vis.*, vol. 73, no. 2, pp. 213–238, 2007.

Xiaodong Yang (S'12) received the B.S. degree from the Huazhong University of Science and Technology, Wuhan, China, in 2009. He is currently working toward the Ph.D. degree with the Department of Electrical Engineering at City College, City University of New York, New York, NY, USA.

His research interests include computer vision, machine learning, video surveillance, and multimedia. His current research focuses on large-scale image and video recognition, surveillance event detection, human activity analysis, and computer vision-based assistive systems for visually impaired people.

Shuai Yuan received the B.S. degree from Southeast University, Nanjing, China, in 2008, and the M.S. degree in electrical engineering from the City College, City University of New York, New York, NY, USA, in 2011.

In 2009, he joined Media Lab of the City College of New York and was a Research Assistant while pursuing the M.S. degree. He is currently a Software Developer in Teledata Communications Inc. His research interest is on pattern recognition and classification. His research interests include machine learning, image processing, and object recognition.

YingLi Tian (M'99–SM'01) received the B.S. and M.S. from TianJin University, TianJin, China, in 1987 and 1990, respectively, and the Ph.D. degree from the Chinese University of Hong Kong, Hong Kong, in 1996.

After holding a faculty position at National Laboratory of Pattern Recognition, Chinese Academy of Sciences, Beijing, she joined Carnegie Mellon University in 1998, where she was a Postdoctoral Fellow with the Robotics Institute. Then, she was a Research Staff Member with IBM T. J. Watson Research Center from 2001 to 2008. She is one of the inventors of the IBM Smart Surveillance Solutions. She is currently a Professor with the Department of Electrical Engineering, City College of New York, New York, NY, USA. Her current research interests include a wide range of computer vision problems from motion detection and analysis, assistive technology, to human identification, facial expression analysis, and video surveillance.