Predicting Types of Clothing Using SURF and LDP based on Bag of Features

Wisarut Surakarin
Department of Computer Engineering,
Faculty of Engineering, Chulalongkorn University,
Bangkok, Thailand
ws.sarut@gmail.com

Prabhas Chongstitvatana
Department of Computer Engineering,
Faculty of Engineering, Chulalongkorn University,
Bangkok, Thailand
prabhas@chula.ac.th

***Abstract*— Low-level feature, such as Local Directional Pattern (LDP) was used to describe textures and shapes in the image. The advantage of the LDP feature is its robustness under random noise and illumination/light changes. This paper proposed a new approach to classifying and recognizing types of clothing by using Speeded-Up Robust Features (SURF) and Local Directional Pattern (LDP) based on Bag of Features (BoF) model. The key processes of the proposed system are firstly, the human are located and segmented clothing in the image. Secondly, Speeded-Up Robust Features (SURF) is used for detecting the interesting points and LDP features are used to create a codebook. Finally, a support vector machine (SVM) is used to classify the types of clothing. The dataset consists of seven categories of clothing such as sweaters, suits and shirts. Our dataset consists of total 1131 images out of which the training set is 991 images and the remainder is the testing set. The result of the recognition rate achieves an average F-score of 63.36%.**

Keywords*— *Local Direction Pattern; Bag of Words; Clothing Segmentation; Clothing Classification

I. Introduction

Clothing is a part of fashion and it is important in daily life. There are many people who choose the types of clothing based on the situation on that day. In addition, it can represent the preference, lifestyle, culture and social status. However, classification and recognition of clothing types are a challenge in computer vision. Many researches address the classification and recognition of clothing. Clothing classification has many useful applications, for example, searching and content-based image retrieval. It represents the appearance and the shape of an object in the image. Nodari et al. [1] and Cushen et al. [2] implemented visual search applications for mobile devices. They used content-based image retrieval to search similar image of clothes. In addition, clothing recognition is used to identify the person. Ming and Kai [12] demonstrated real-time clothing recognition system and it is used to identify the people. Because the viewpoints of clothing can be changed, it is a problem to the classification and recognition of the types of clothing in an image.

Speeded Up Robust Features (SURF) [3] is a feature detector. It has succeeded in the situation when small viewpoints change for object recognition and detection. SURF selects an area of distinctive locations in the image such as corners and blobs and it finds the stable feature points. These points are called *keypoints or interest points*. Many content-based image retrieval methods use SURF to detect interest points in clothing and use them to do matching of similar clothing image. There are advantages and disadvantages of using SURF detector. The advantages are that SURF is good at handling the blurring, scaling, translation and rotation invariants. However, it is not good at handling illumination/light changes. In complementing SURF, LDP [4] features are used to describe local textures and surface characteristics because LDP feature is robust under random noise and illumination/light changes.

The main purpose of this work was to classify and recognize the types of clothing by combination of SURF to find *interest points* in a raw image and LDP to create a *codebook* based on bag of features model. The input image is constrained to upper body, full body and frontal face image with a clear background. The dataset is acquired from online e-commerce, social network, imageNet [5] and Google image. Support vector machine (SVM) is used as a baseline classifier for seven categories that included jackets, shirts, t-shirts, polo shirts, suits, sweaters and tank tops in this study.

The remainder of this paper is divided into six sections. Section 2 describes related works. Section 3 proposes the system of clothing segmentation and classification. Section 4 reports the experiment and the results. Section 7 is the conclusion of this study.

II. Related Works

Previous works focused on content-based image retrieval and image searching. For example, Miura et al. [6] proposed image retrieval for supporting customer to query a similar clothing image and recommend a fashion coordinate. Chiao-Meng et al. [7] used multiple features, which included clothing types, appearance, textures and colors for their clothing similar image retrieval algorithm. Grana et al. [8] and Tianfa et al. [9] proposed color space for searching similar color in fashion image retrieval. The feature of SURF was extracted by Bag of Words model. The Bag of Words provides high performance for large dataset to query an image. Many work focus on clothing segmentation from upper body or full body. Nan Wang et al. [10] proposed clothing segmentation of upper body to estimate the shape and grouping. In addition, recognition of

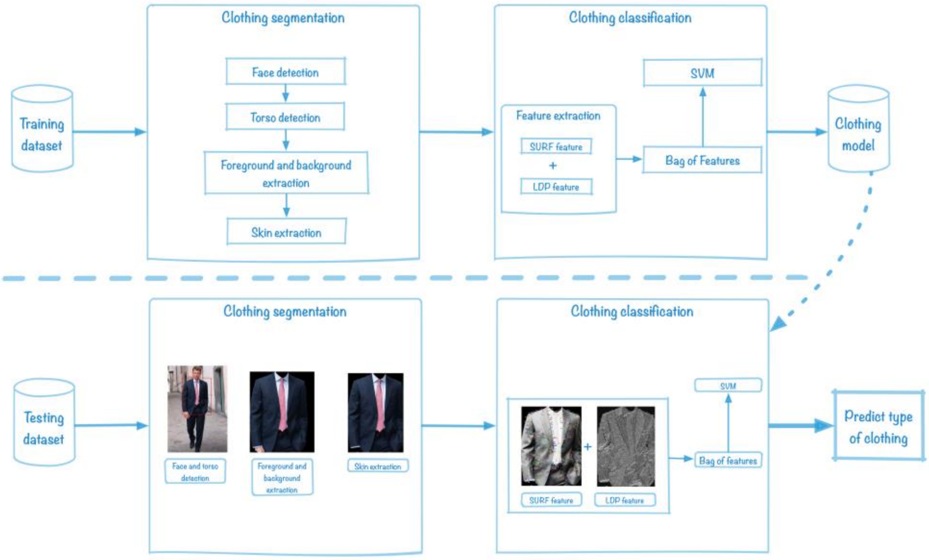

Figure 1. The proposed system of predicting types of clothing

identify has been introduced by Gallagher et al. [11]. They use facial and clothing to recognize people in an image. Ming and Kai [12] achieved recognition in real time. Hidayati et al. [13] proposed clothing genre classification which described dominant features of clothing. Li and Weilong [14] using superpixels and graph cut to segmented clothing. In this work, we segmented clothing by using graph cut method.

In this paper, we do not focus on clothing image retrieval. The main work is in classification and recognition of clothing type. The system is divided into two components. Firstly, detection of clothing in an image and secondly prediction types of clothing in the image.

III. Proposed System

The system is shown in Figure 1. It is divided into two components, which include, clothing segmentation, feature extraction and clothing classification. The output is the type of clothing in an image.

A. Clothing segmentation

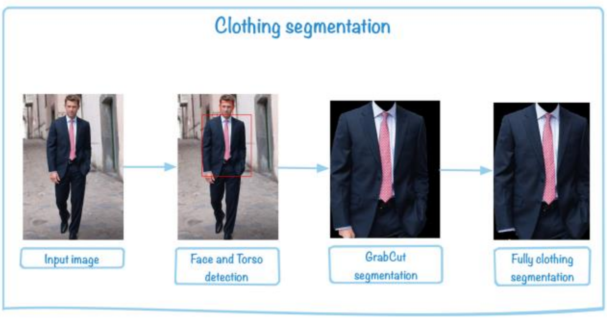

Figure 2. Clothing segmentation

. Figure 2 shows the clothing segmentation process. The input image [15] is retrieved from Google image. There are three main steps to clothing segmentation and locate clothing in the image.

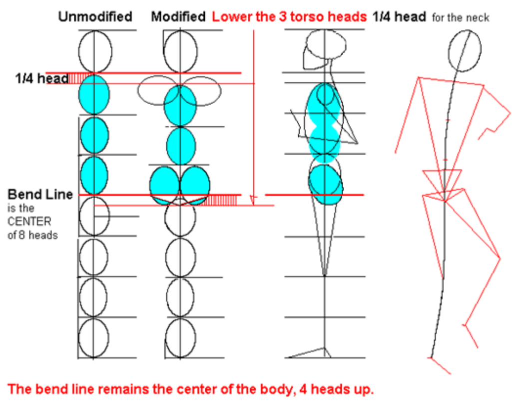

Figure 3. Human drawing proportion

1) Face and torso detection

The most popular method to find an upper body is to detect a face first. Face detection algorithm [16] proposed by Viola-Jones determines the locations and sizes of the human face in the image. We assume upper body must be below the head and crop-bounding box of upper body with regions of interesting (ROI) using human proportion [17] as shown in Figure 3. The head size is calculated from the bounding box of the face.

2) Foreground and background extraction

. We eliminate the non-clothing and background from the bounding box of torso detection by using GrabCut [18] algorithm. GrabCut is an object extraction algorithm based on graph cuts to separate foreground and background in an image.

3) Skin extraction

After segmenting the foreground using GrabCut algorithm with a bounding box, the clothing can be identified. However the skin of person is stay in the bounding box too. The skin is eliminated by using color segmentation. The color space is converted from RGB to HSV before the skin is eliminated. Albiol et al. [19] proposed optimum color space for skin detection. The skin pixels are difficult to be classified according to a set of color model because of the illumination in the real world can change in a different environment. In this work, we set the threshold skin pixel using ranges H = [0, 42], S = [32, 235], V = [60, 255] and use equation (1) to create mask human skin. If a $pixel_{(x,\, y)}$ is a skin color set to one or is not a skin set to zero.

$$skin(x, y) = \begin{cases} 1, & if\ a\ pixels\ are\ not\ in\ rages \\ 0, & otherwise \end{cases} \tag{1}$$

Before feature extraction, we normalized size the bounding box of upper body to $size_{m,n}(128, 256)$ in all of image dataset.

B. Feature extraction using Local Directional Pattern (LDP)

Local Directional Pattern (LDP) [4] is a low-level feature and it is most popular in applications of facial image analysis [20, 21]. LDP considers the edge response values from different directional at each pixel in the image and create a binary bit code from the relative strength magnitude. The LDP is inspired by the Local Binary Pattern (LBP) to texture analysis. LDP is robust to illumination/light change and random noise. We also proposed to combination of SURF detect points of interest and LDP codes to generate codebook.

Firstly, we compute of eight directional response edge values by using Kirsch mask m_i. Kirsch mask in all different orientations $m_i = \{0, 1, 2 \ldots 7\}$ centered. As shown in Figure 3. We take absolute for all of eight directional values.

$$\underset{\text{South East } (m_7)}{\begin{bmatrix} -3 & -3 & -3 \\ -3 & 0 & 5 \\ -3 & 5 & 5 \end{bmatrix}} \quad \underset{\text{South } (m_6)}{\begin{bmatrix} -3 & -3 & -3 \\ -3 & 0 & -3 \\ 5 & 5 & 5 \end{bmatrix}} \quad \underset{\text{South West } (m_5)}{\begin{bmatrix} -3 & -3 & -3 \\ 5 & 0 & -3 \\ 5 & 5 & -3 \end{bmatrix}} \quad \underset{\text{West } (m_4)}{\begin{bmatrix} 5 & -3 & -3 \\ 5 & 0 & -3 \\ 5 & -3 & -3 \end{bmatrix}}$$

$$\underset{\text{North West } (m_3)}{\begin{bmatrix} 5 & 5 & -3 \\ 5 & 0 & -3 \\ -3 & -3 & -3 \end{bmatrix}} \quad \underset{\text{North } (m_2)}{\begin{bmatrix} 5 & 5 & 5 \\ -3 & 0 & -3 \\ -3 & -3 & -3 \end{bmatrix}} \quad \underset{\text{North East } (m_1)}{\begin{bmatrix} -3 & 5 & 5 \\ -3 & 0 & 5 \\ -3 & -3 & -3 \end{bmatrix}} \quad \underset{\text{East } (m_0)}{\begin{bmatrix} -3 & -3 & 5 \\ -3 & 0 & 5 \\ -3 & -3 & 5 \end{bmatrix}}$$

Figure 4. Kirsch edge mask in eight directions

We want to choose the top k to generate binary codes. We define the k most response of directional values to generate the LDP binary codes. All of the top k are set bit to one and other (8-k) are set to zero. After that the LDP codes, which is eight directions, is using equation 2.

$$LDP_k(x, y) = \sum_{i=1}^{8} c_i(m_i - m_k) \times 2^i \tag{2}$$

$$c_i(x) = \begin{cases} 1, & x \geq 0 \\ 0, & x < 0 \end{cases}$$

Where, m_i is a rank in eight directional responses, m_k is the m^{th} most significant directional response and c_i(x) is the response value of direction. As shown in Figure 6 clothing image after using generated LDP codes with k = 4. We calculate LDP codes in Figure 4 and result of LDP image in Figure 5.

201	23	122
74	12	64
91	152	89

Index	m_7	m_6	m_5	m_4	m_3	m_2	m_1	m_0
Mask result	|-8|	|208|	|88|	|480|	|64|	|320|	|-776|	|-248|
Rank	8	5	6	2	7	3	1	4
Binary code k = 4	0	0	0	1	0	1	1	1
Decimal code	23							

Figure 5. Generating LDP code with k = 4

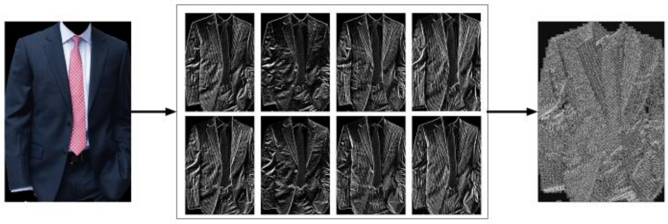

Figure 6. Generating LDP image with k = 4

C. Clothing classification

Clothing classification component has four sub-components to predict types of clothing by using SVM. The first sub-component, we use combination of SURF and LDP features. SURF was detecting interest points in the image. Second, the codebook was created from interest points.

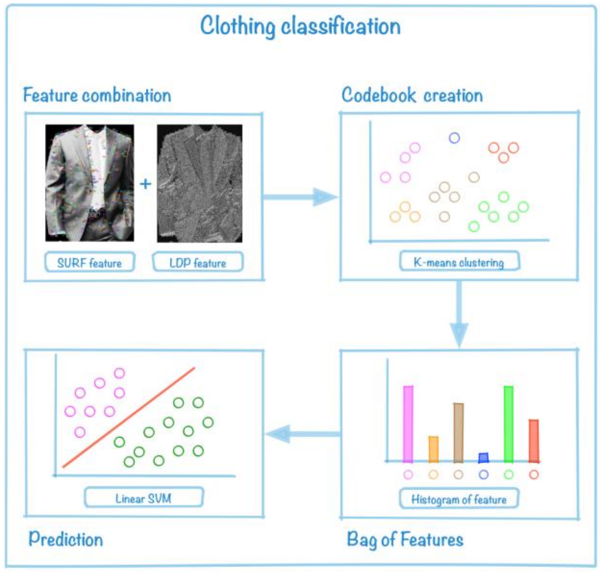

Figure 7. Predicting of types of clothing

The proposed system is motivated by the Bag of Features (BoF) [22] approach. BoF is used in document classification. It works by counting vocabulary of words. BoF have been applied to computer vision. The visual vocabulary is constructed to represent the dictionary by clustering local feature from training dataset.

SURF is used to locate structures and detect the interest points.. After that, the LDP detector is used to encode the local textural structures of the clothing images and create visual vocabulary as called *codebook*. The codebook created by LDP features from each interest points. All of interest points are clustered using *k-means* method. We classify features into 100 clusters. Histogram of feature was generated after clustering by *k-means* method. We use histogram of feature to predict types of clothing by using SVM.

IV. Experiment And Result

In this section, we describe the classification performance by comparing the experiment of SURF, combination of the SURF and LBP, and our proposed combination of the SURF and LDP. As shown in Figure 9, each category shows the accuracy of classification.

A. Dataset

We collected image, which have upper body or full body image, clearly background and frontal face from online e-commerce, social network, imageNet [5] and Google image. Our dataset consists of a total of 1131 images out of which training set are 991 images and other are testing set. Seven categories: jacket, shirt, t-shirt, polo shirt, suit, sweater and tank top are classified. In our dataset, the number of training and testing for each category is shown in Table 1.

TABLE I. Dataset for each category

	Jacket	Shirt	T-shirt	Polo shirt	Suit	Sweater	Tank top
Train	140	120	121	160	170	135	145
Test	20	20	20	20	20	20	20

B. Accuracy Score

We trained SVM, which uses linear kernel with configuration parameter C = 1.1, the other parameters were auto-selected. We used F-score criterion to report. The results of types of clothing are shown in Table 3. Confusion matrix which columns represent predicting class and rows represent actual class. We achieve an average F-score 63.36%, precision of 64.29% and recall of 64.64%. Table 4 shows that precision, recall, and F-score for each category.

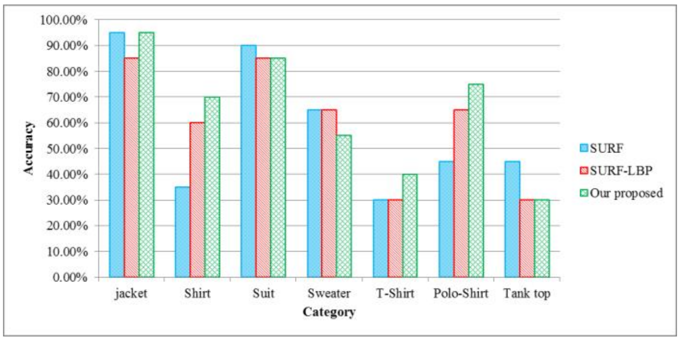

Figure 9. Comparison of accuracy score in different method on the image dataset

TABLE II. Comparison in terms of F-score of different method

	SURF	SURF-LBP	Our proposed
Precision	57.86%	60.00%	64.29%
Recall	57.76%	61.79%	64.64%
F-score	56.46%	59.28%	63.36%

TABLE III. A confusion matrix of our proposed

	jacket	*Shirt*	*Suit*	*Sweater*	*T-Shirt*	*Polo-Shirt*	*Tank top*
jacket	19	0	3	2	1	0	0
Shirt	1	14	0	5	1	0	0
Suit	0	2	17	0	3	0	1
Sweater	0	4	0	11	3	0	8
T-Shirt	0	0	0	0	8	0	3
Polo-Shirt	0	0	0	0	1	15	2
Tank top	0	0	0	2	3	5	6

TABLE IV. F-score of our proposed for each category

	precision	recall	F score
Jacket	95.00%	76.00%	84.44%
Shirt	70.00%	66.67%	68.29%
Suit	85.00%	73.91%	79.07%
Sweater	55.00%	42.31%	47.83%
T-Shirt	40.00%	72.73%	51.61%
Polo-Shirt	75.00%	83.33%	78.95%
Tank top	30.00%	37.50%	33.33%
Overall	64.29%	64.64%	63.36%

V. Conclusions

In this paper, we proposed clothing recognition and classification by using SURF and LDP. Support vector machine was used to predict seven types of clothing. The experimental data showed our proposed method achieves an average F-score of 63.36%, precision of 64.29% and recall of 64.64%. In the future work, we plan to present advertise apparel from type of clothing system.

VI. References

[1] A. Nodari, M. Ghiringhelli, A. Zamberletti, M. Vanetti, S. Albertini, and I. Gallo, "A mobile visual search application for content based image retrieval in the fashion domain," in *Content-Based Multimedia Indexing (CBMI), 2012 10th International Workshop on*, 2012, pp. 1-6.

[2] G. A. Cushen and M. S. Nixon, "Mobile visual clothing search," in *Multimedia and Expo Workshops (ICMEW), 2013 IEEE International Conference on*, 2013, pp. 1-6.

[3] H. Bay, T. Tuytelaars, and L. Van Gool, "SURF: Speeded Up Robust Features," in *Computer Vision – ECCV 2006*. vol. 3951, A. Leonardis, H. Bischof, and A. Pinz, Eds., ed: Springer Berlin Heidelberg, 2006, pp. 404-417.

[4] M. H. Kabir, T. Jabid, and O. Chae, "A Local Directional Pattern Variance (LDPv) Based Face Descriptor for Human Facial Expression Recognition," presented at the Proceedings of the 2010 7th IEEE International Conference on Advanced Video and Signal Based Surveillance, 2010.

[5] D. Jia, D. Wei, R. Socher, L. Li-Jia, L. Kai, and F.-F. Li, "ImageNet: A large-scale hierarchical image database," in *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*, 2009, pp. 248-255.

[6] S. Miura, T. Yamasaki, and K. Aizawa, "SNAPPER: Fashion Coordinate Image Retrieval System," in *Signal-Image Technology & Internet-Based Systems (SITIS), 2013 International Conference on*, 2013, pp. 784-789.

[7] H. Chiao-Meng, S. Chen, M. Cheng, and Y. C. F. Wang, "A sparse-coding based approach to clothing image retrieval," in *Intelligent Signal Processing and Communications Systems (ISPACS), 2012 International Symposium on*, 2012, pp. 314-318.

[8] C. Grana, D. Borghesani, and R. Cucchiara, "Class-Based Color Bag of Words for Fashion Retrieval," in *Multimedia and Expo (ICME), 2012 IEEE International Conference on*, 2012, pp. 444-449.

[9] W. Tianfa, Y. Yule, S. Ling, and Z. Yong, "Clothing image retrieval using color moment," in *Computer Science and Network Technology (ICCSNT), 2013 3rd International Conference on*, 2013, pp. 1016-1020.

[10] W. Nan and A. Haizhou, "Who Blocks Who: Simultaneous clothing segmentation for grouping images," in *Computer Vision (ICCV), 2011 IEEE International Conference on*, 2011, pp. 1535-1542.

[11] A. C. Gallagher and C. Tsuhan, "Clothing cosegmentation for recognizing people," in *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on*, 2008, pp. 1-8.

[12] Y. Ming and Y. Kai, "Real-time clothing recognition in surveillance videos," in *Image Processing (ICIP), 2011 18th IEEE International Conference on*, 2011, pp. 2937-2940.

[13] S. C. Hidayati, W.-H. Cheng, and K.-L. Hua, "Clothing genre classification by exploiting the style elements," presented at the Proceedings of the 20th ACM international conference on Multimedia, Nara, Japan, 2012.

[14] Li and Weilong, "Interactive Clothing Image Segmentation Based on Superpixels and Graph Cuts," in *Computer Sciences and Applications (CSA), 2013 International Conference on*, 2013, pp. 659-662.

[15] "mens' suits," in *http://www.city-suits.co.uk/media/catalog/product/cache/1/image/9df78eab33525d08d6e5fb8d27136e95/s/o/somerton-3159b_2_1.jpg*, ed.

[16] P. Viola and M. J. Jones, "Robust Real-Time Face Detection," *Int. J. Comput. Vision,* vol. 57, pp. 137-154, 2004.

[17] (2014). *Human Figure Drawing Proportions*. Available: http://www.realcolorwheel.com/human.htm

[18] C. Rother, V. Kolmogorov, and A. Blake, ""GrabCut": interactive foreground extraction using iterated graph cuts," presented at the ACM SIGGRAPH 2004 Papers, Los Angeles, California, 2004.

[19] A. Albiol, L. Torres, and E. J. Delp, "Optimum color spaces for skin detection," in *Image Processing, 2001. Proceedings. 2001 International Conference on*, 2001, pp. 122-124 vol.1.

[20] T. Jabid, M. H. Kabir, and C. Oksam, "Gender Classification Using Local Directional Pattern (LDP)," in *Pattern Recognition (ICPR), 2010 20th International Conference on*, 2010, pp. 2162-2165.

[21] D.-J. Kim, S.-H. Lee, and M.-K. Sohn, "Face Recognition via Local Directional Pattern," *International Journal of Security and Its Applications,* vol. 7, March 2013.

[22] S. O'Hara and B. A. Draper, "Introduction to the Bag of Features Paradigm for Image Classification and Retrieval," *Computing Research Repository (CoRR),* 2011.