

# Clothing Style Recognition using Fashion Attribute Detection

Guang-Lu Sun  
sunguanglu66@126.com

Hong-Han Chen  
8hhchen8@gmail.com

Xiao Wu  
wuxiaohk@swjtu.edu.cn

Qiang Peng  
qpeng@swjtu.edu.cn

School of Information Science and Technology  
Southwest Jiaotong University  
Chengdu, China

## ABSTRACT

In this paper, a new framework is proposed for clothing style recognition in natural scenes. Clothing region is first detected through the fusion of super-pixel segmentation, saliency detection and Gaussian Mixture Model (GMM). Next, a group of fashion attribute detectors are trained to get the likelihood of each attribute in the clothing image. Finally, the correlation matrix between clothing styles and fashion attributes is adopted to predict the clothing style. For evaluation, we collect a dataset for clothing style recognition which contains 5 styles and 14 fashion attributes. Extensive experiments demonstrate that the proposed framework has a promising ability to recognize the clothing style.

## Categories and Subject Descriptors

H.3.3 [Information Storage and Retrieval]: Information Search and Retrieval

## Keywords

Clothing style, Fashion, Attribute

## 1. INTRODUCTION

With the rapid development of shopping websites, online clothing shopping is very popular nowadays. Facing massive clothing products, it is essential for users to search a desired clothing effectively. Generally, the search results returned by search engines are based on the text relevance. However, the tags of clothing images are usually full of noises. Aiming to solve this problem, some works were proposed for automated clothing attribute annotation by training a few attribute classifiers based on visual features of clothing images.

Unfortunately, clothing retrieval based on attributes cannot meet the needs of users completely. Different from cloth-



Figure 1: Some examples of clothing styles, such as British, Sexy and Punk.

ing attributes, style is an important cue for describing clothing from the human point of view, which meet well with the search habits of users. Therefore, we consider to describe the clothing based on styles. However, it poses some challenges as follows: 1) The background of each clothing image is very cluttered in most natural scenes. 2) There are large visual variances in the clothes with the same style (as shown in Figure 1). 3) Clothing style belongs to high-level concept of human, so there is a semantic gap between low-level features and styles.

To solve the above problems, a framework is proposed (as shown in Figure 2). First, a new clothing region localization method is introduced to reduce the interference of the background, which combines super-pixel segmentation, saliency detection and GMM model. Second, in the step of fashion design, each style of a clothing has its representative fashion attributes. Inspired by this intuition, a new method is proposed for clothing style recognition, which integrates the fashion attribute detection model. Finally, a correlation matrix between styles and fashion attributes is introduced to boost the performance of style prediction. The main contributions of our work can be summarized as follows:

1. The proposed framework explores a new perspective to tackle the problem of clothing style classification in natural scenarios.

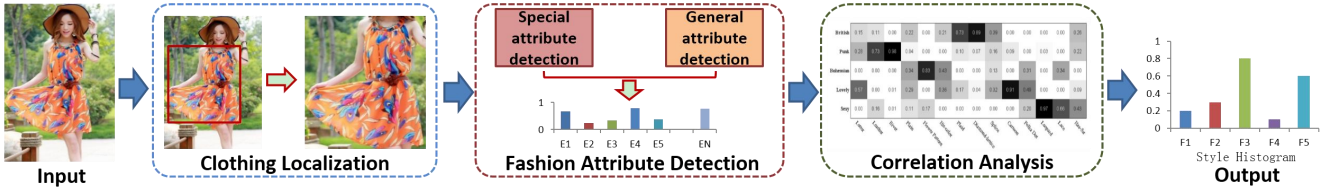


Figure 2: The whole framework for clothing style recognition

2. A clothing region localization method is proposed, which is more robust than the existing works.
3. A series of fashion attribute detectors are trained, and then the clothing style is predicted by utilizing the correlation matrix between fashion attributes and styles.

## 2. RELATED WORK

With the explosive growth of clothing images, the study of clothing images has attracted more attention of researchers, mainly including clothing extraction, clothing attribute learning and applications based on clothing models.

Most clothing extraction methods [1, 2] extracted the clothing objects by distinguishing the foreground and background regions. To more accurately the clothing region, the foreground region was predicted [2] by using face detection [3] and skin detection [4].

There existed a few interesting applications based on clothing images. Occupation can be predicted by exploiting clothing contextual information [5]. The clothing can be classified based on color and texture features [6]. A practical problem of cross-scenario clothing retrieval was addressed by combining pose estimation and transfer learning technology [7]. An occasion-oriented clothing recommendation system [8] was constructed by considering two key criteria: wear properly and wear aesthetically.

Clothing attribute learning was normally used to solve the classification problem. A classification model of clothing attributes was introduced in [9], which consists of a multi-class learner based on a Random Forest. The prediction of independent attribute classifiers was improved by exploring mutual dependencies between the attributes based on Conditional Random Field in [10]. However, these two approaches depended on upper body detection and pose estimation, respectively. Style Finder system [11] created a more detailed attribute list for women fashion coat, and trained a set of attribute classifiers by extracting multi-features on the whole image. But it was only designed within one scenario. Due to the strong association between fashion attributes and styles, we tried to boost the semantic classification performance of clothing by utilizing fashion attributes.

## 3. FRAMEWORK

### 3.1 Clothing Region Localization

Currently, most works on clothing localization were based on face detection or pose estimation. However, there are plenty of clothing images without face or person. Motivated by the following intuitions, we explore to design a robust clothing region localization method: (1) The clothing is often put in the center of an image. The spatial location of clothing is submitted to the Gaussian distribution. (2) The clothing is usually photographed in a relative high



Figure 3: Examples of clothing region localization

contrast background for highlighting the subjective. Hence, the clothing region is salient from the human visual perception. (3) In the super-pixel method [12], the clothing and the background are divided into small blocks, and the visual features of those blocks belonging to the clothing have the high similarity. Therefore, the clothing regions in images can be obtained by combining spatial information, salient region detection [13] and super-pixel method [12].

Given a clothing image  $I(x, y)$ ,  $x$  and  $y$  are the coordinates of each pixel. The details of the clothing localization process are expressed as follows:

Firstly, the Gaussian weight map  $P_G(x, y)$  of an input image can be generated adaptively according to Eqn. (1).

$$f(x_1, x_2 \dots x_N) = \frac{1}{(2\pi)^{N/2} |\Sigma|^{1/2}} \exp\left(-\frac{1}{2} (x - \mu)^T \Sigma^{-1} (x - \mu)\right) \quad (1)$$

where  $N$  is the number of dimensions.  $\Sigma$  and  $\mu$  are Gaussian parameters.

To determine the clothing regions, saliency detection [13] is implemented for input image to get the saliency value of each pixel, denoted as  $P_s(x, y)$ .

Meanwhile, the image is segmented with the super-pixel detection, and then the probability of each super-pixel block is calculated by combining the Gaussian map and the saliency map, which is given in Eqn. (2).

$$P_{sp}(i) = \frac{1}{N} \sum_{x, y \in sp_i} P_s(x, y) * P_g(x, y) \quad (2)$$

where  $i$  is the index of super-pixel,  $N$  is the total number of pixels in a super-pixel block. If the number of pixels in a super-pixel block is greater than a certain threshold, it will be treated as the clothing region. Then the clothing model  $F_m$  is constructed by extracting visual features of the clothing region  $R_F$ . The background model is built by utilizing visual features of super pixel blocks which located in the border of the image. The unassigned super-pixel blocks are classified into the clothing region or the background region according to the similarity distance, as Eqn. (3).

$$d = \|sp_j - M\|^2 \quad (3)$$

where  $sp(j)$  indicates one of the unassigned super-pixel blocks,  $M$  represents the clothing model or the background model.

Finally, the clothing region is effectively extracted. The extracted clothing regions are shown in Figure 3.



Figure 4: The illustration of clothing styles and fashion attributes

### 3.2 Fashion Attribute Detection

Each style of a clothing contains unique fashion attributes. Our defined fashion attributes of each style are shown in Figure 4. Some fashion attributes detection have been done by previous works and have achieved good performance. We follow the similar way as previous works for the fashion attribute detection. Unfortunately, for some fashion attributes, the performance of previous methods is unacceptable. Due to this reason, a group of detectors for specific fashion attributes is constructed.

#### 3.2.1 Rivet and Leopard Detectors

Generally, the intensity of a rivet or a leopard spot is brighter or darker than their neighbor areas, and there are more local features in these regions compared to other regions. Inspired by [14], the density of local points can be used to distinguish a rivet or leopard region from other attributes. The density of local points is calculated as Eqn. (4).

$$\hat{f}_k(s) = \sum_{d_i \leq h} \frac{3}{4h^2} \left(1 - \frac{d_i^2}{h^2}\right)^2 \quad (4)$$

where  $h$  is the size of the sliding window,  $\hat{f}_k(s)$  denotes the density value of the detected area,  $d_i$  indicates the distance between the local feature and the center of the sliding window.

#### 3.2.2 Plaid and Letter Detectors

The plaid attribute is easy to be recognized because it contains many cross lines. Firstly, we convert the color image to grayscale and perform adaptive local binarization process. Next, the line detection is carried out with the probabilistic Hough transform model. Finally, the angle histogram of the line is exploited for plaid detection. Two bins in angle histogram are usually high, while other bins have low values.

For the letter detection, a method [15] is adopted, which is designed for text detection in natural scene. In order to refine the performance, an Optical Character Recognition (OCR) technology [16] is integrated for letter detection.

### 3.3 Style Recognition by Correlation Matrix

To map the distribution of fashion attributes of clothing to the style space, the relationship matrix between fashion attributes and styles is constructed, which is shown in Figure 5. The value of the matrix is calculated by the statistic

British	0.15	0.11	0.00	0.22	0.00	0.21	0.73	0.89	0.39	0.00	0.00	0.00	0.00	0.26
Punk	0.28	0.73	0.98	0.04	0.00	0.00	0.10	0.07	0.16	0.09	0.00	0.03	0.00	0.22
Bohemian	0.00	0.00	0.00	0.34	0.83	0.43	0.00	0.00	0.13	0.00	0.31	0.00	0.34	0.00
Lovely	0.57	0.00	0.01	0.29	0.00	0.36	0.17	0.04	0.32	0.91	0.49	0.00	0.00	0.09
Sexy	0.00	0.16	0.01	0.11	0.17	0.00	0.00	0.00	0.00	0.00	0.20	0.97	0.66	0.43
	Letter	Leather	Rivet	Plain	Flower	Hit-color	Diamond	Plaid	Splice	Cartoon	Polka-dot	Leopard	Lace	Has-fur

Figure 5: The correlation matrix between styles and fashion attributes

analysis. The correlation matrix proves the assumption that clothing style recognition by fashion attributes detection is reasonable. The style of clothing can be determined by the correlation matrix, which can be calculated as Eqn. (5).

$$F_i = \operatorname{argmax} \left\{ S_{ij} = W_j * P_i^T \right\} \quad (5)$$

where  $W_j$  is a vector representing the relationship between the  $j_{th}$  style and the fashion attributes,  $P_i$  is the distribution of the  $i_{th}$  clothing in the fashion attribute space,  $S_{ij}$  is the score of the  $i_{th}$  clothing on the  $j_{th}$  style.

## 4. EXPERIMENTS AND DISCUSSION

### 4.1 Dataset

Since there is no suitable clothing dataset to evaluate the clothing style recognition task, we collect the clothing images from Taobao, the biggest commercial website in China. The total number of clothing images is 7275. We define 5 styles and 14 fashion attributes. The information on this dataset is shown in Table 1 and Table 2.

### 4.2 Experiments on Fashion Attribute Detection and Style Recognition

The fashion attribute dataset is composed of general attributes and specific attributes. For general attributes, feature selection is undergone from multiple features like LBP, HOG, PHOW and Wavelet. For specific attribute detection, the specific attribute detectors described in Section 3.2 are used. The accuracy of fashion attribute detection is listed in the Table 1.

To evaluate the effectiveness of the proposed framework for clothing style recognition, style finder [11] is treated as a baseline method, which achieved good performance on clothing attribute detection in clean background images.

The performance comparison of style recognition is shown in Figure 6. Our method achieves the best performance, which outperforms the style finder. There are mainly two reasons: on one hand, style finder trained each style classifier by using visual features directly, which ignored the fact that there exist huge visual differences in the clothing of the same style. It is unreasonable to recognize the clothing style only by style classifiers based on visual features. But fashion attribute is visually more stable than the whole clothing. So our method can handle the problem of visual difference. On the other hand, there exists significant semantic gap for style recognition, the fashion attributes act as tie bridge to reduce the semantic gap. Due to the visual stability of fashion attributes, a more accurate result for fashion attributes detection can be obtained. Meanwhile,

Table 1: The information on fashion elections and their detection performance

Attribute	POS	NEG	Accuracy(%)
Letter	403	6,872	83.2
Leather	432	6,839	75.1
Rivet	628	6,647	87.3
Plain	397	6,878	71.5
Flower pattern	571	6,704	86.2
Hit color	432	6,843	69.7
Plaid	633	6,642	79.3
Splice	386	6,889	65.5
Cartoon	627	6,648	67.8
Polka dot	379	6,896	78.7
Leopard	550	6,725	81.9
Lace	675	6,600	70.5
Has-fur	267	7,008	80.7

Table 2: The data information of styles and detection accuracy

Style	POS	NEG	Accuracy(%)
British	1,510	5,765	70
Punk	1,367	5,908	67.5
Bohemian	1,250	6,025	83
Lovely	1,556	5,719	71
Sexy	1,592	5,683	68.3

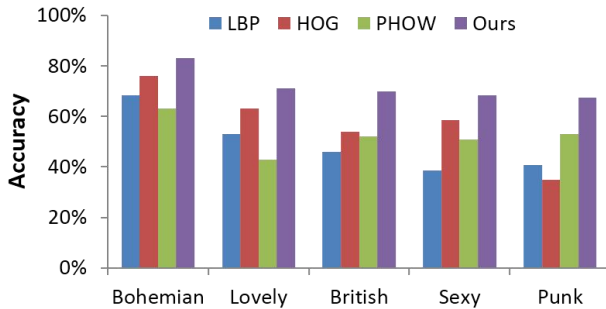


Figure 6: Performance comparison on style recognition

there is a closer relationship between a style and its representative fashion attributes. Therefore, fashion attributes are of great support for the problem of semantic gap. The promising results indicate that it is feasible to recognize the clothing style with fashion attribute detection.

## 5. CONCLUSION

In this work, we propose a solution from a new perspective to recognize the clothing style in natural scenario by detecting fashion attributes. A group of fashion attribute detectors are constructed and then the clothing style is determined using the correlation matrix between fashion attributes and styles. Promising results are achieved on clothing style image dataset. In our future work, we plan to build a large scale clothing recommendation system based on clothing styles.

## 6. ACKNOWLEDGEMENT

This paper was supported by National Natural Science Foundation of China (No. 61373121, No. 61036008), Program for Sichuan Provincial Science Fund for Distinguished

Young Scholars (No. 2012JQ0029, No. 13QNJJ0149) and Sichuan Provincial Science and Technology Innovation Seeding Fund (20131012, 2014-062).

## 7. REFERENCES

- [1] Z. Hu, H. Yan, and X. Lin. Clothing segmentation using foreground and background estimation based on the constrained delaunay triangulation. *Pattern Recognition*, 41(5):1581–1592, 2008.
- [2] X. Wu, B. Zhao, L. Liang, and Q. Peng. Clothing extraction by coarse region localization and fine foreground/background estimation. In *MMM*, pages 316–326. Springer, 2013.
- [3] M. Jones and P. Viola. Fast multi-view face detection. *Mitsubishi Electric Research Lab TR-20003-96*, 3:14, 2003.
- [4] P. Kakumanu, S. Makrogiannis, and N. Bourbakis. A survey of skin-color modeling and detection methods. *Pattern Recognition*, 40(3):1106–1122, 2007.
- [5] Z. Song, M. Wang, X. Hua, and S. Yan. Predicting occupation via human clothing and contexts. In *ICCV*, pages 1084–1091. IEEE, 2011.
- [6] K. Yamaguchi, M. H. Kiapour, L. E. Ortiz, and T. L. Berg. Parsing clothing in fashion photographs. In *CVPR*, pages 3570–3577. IEEE, 2012.
- [7] S. Liu, Z. Song, G. Liu, C. Xu, H. Lu, and S. Yan. Street-to-shop: Cross-scenario clothing retrieval via parts alignment and auxiliary set. In *CVPR*, pages 3330–3337. IEEE, 2012.
- [8] S. Liu, J. Feng, Z. Song, T. Zhang, H. Lu, C. Xu, and S. Yan. Hi, magic closet, tell me what to wear! In *ACM MM*, pages 619–628. ACM, 2012.
- [9] L. Bossard, M. Dantone, C. Leistner, C. Wengert, T. Quack, and L. Van Gool. Apparel classification with style. In *ACCV*, pages 321–335. Springer, 2013.
- [10] H. Chen, A. Gallagher, and B. Girod. Describing clothing by semantic attributes. In *ECCV*, pages 609–623. Springer, 2012.
- [11] W. Di, C. Wah, A. Bhardwaj, R. Piramuthu, and N. Sundaresan. Style finder: Fine-grained clothing style detection and retrieval. In *CVPR*, pages 8–13. IEEE, 2013.
- [12] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Susstrunk. Slic superpixels compared to state-of-the-art superpixel methods. *IEEE Trans. on PAMI*, 34(11):2274–2282, 2012.
- [13] T. Liu, Z. Yuan, J. Sun, J. Wang, N. Zheng, X. Tang, and H. Y. Shum. Learning to detect a salient object. *IEEE Trans. on PAMI*, 33(2):353–367, 2011.
- [14] S. Segui, M. Drozdal, F. Vilarino, C. Malagelada, F. Azpiroz, P. Radeva, and J. Vitria. Categorization and segmentation of intestinal content frames for wireless capsule endoscopy. *IEEE Trans. on ITB*, 16(6):1341–1352, 2012.
- [15] H. Chen, S. S. Tsai, G. Schroth, D. M. Chen, R. Grzeszczuk, and B. Girod. Robust text detection in natural images with edge-enhanced maximally stable extremal regions. In *ICIP*, pages 2609–2612. IEEE, 2011.
- [16] S. Mori, C. Y. Suen, and K. Yamamoto. Historical review of ocr research and development. *Proceedings of the IEEE*, 80(7):1029–1058, 1992.