Clothing genre classification by exploiting the style elements

Article · October 2012

DOI: 10.1145/2393347.2396402

3 authors:

Shintami Chusnul Hidayati
Academia Sinica
12 PUBLICATIONS **535** CITATIONS
SEE PROFILE

Wen-Huang Cheng
Academia Sinica
176 PUBLICATIONS **2,400** CITATIONS
SEE PROFILE

Kai-Lung Hua
National Taiwan University of Science and Technology
140 PUBLICATIONS **1,412** CITATIONS
SEE PROFILE

Clothing Genre Classification by Exploiting the Style Elements

Shintami C. Hidayati
Dept. of CSIE, Nat'l
Taiwan Univ. of Sci. & Tech.
Taipei, Taiwan, R.O.C.
m9915810@mail.ntust.edu.tw

Wen-Huang Cheng
Research Center for IT
Innovation, Academia Sinica
Taipei, Taiwan, R.O.C.
whcheng@citi.sinica.edu.tw

Kai-Lung Hua
Dept. of CSIE, Nat'l
Taiwan Univ. of Sci. & Tech.
Taipei, Taiwan, R.O.C.
hua@mail.ntust.edu.tw

ABSTRACT

This paper presents a novel approach to automatically classify the upperwear genre from a full-body input image with no restrictions of model poses, image backgrounds, and image resolutions. Five style elements, that are crucial for clothing recognition, are identified based on the clothing design theory. The corresponding features of each of these style elements are also designed. We illustrate the effectiveness of our approach by showing that the proposed algorithm achieved overall precision of 92.04%, recall of 92.45%, and F-score of 92.25% with 1,077 clothing images crawled from popular online stores.

Categories and Subject Descriptors

I.5.4 [**Pattern Recognition**]: Applications—*Computer Vision*

General Terms

Algorithm, Experimentation

Keywords

Clothing genre, classification, style element

1. INTRODUCTION

Dressing is a way of life. People dress to express their personalities, emotions, and self-confidence. Any person with the right advice will be able to learn how to dress properly. However, most people are not natural-born fashion stylist. Choosing a "right" combination of clothes to wear has become a tedious and even annoying routine for everyday life [8]. The same situation occurs when shopping online for clothes. Although most of the existing online stores provide keyword-based or content-based search, they do not well support the people's practical needs to look for clothes with the desired style. To address this problem, the development of effective techniques for automatically classifying the clothing genres is crucial. For example, a recognized clothing genre, say a sleeveless upperwear with string, can be recommended to be combined with a short pant.

According to the theory of clothing design [9], for example, a piece of upperwear can be defined by its style elements such as the presence of collar, print style, and sleeve, to determine the clothing genre. In particular, a collar refers to the part of upperwear that folds over and fastens around the neck. The widths and shapes of collars have varied substantially over time because they rise and fall with fashion trends in eras. Therefore, the style elements like collar type can be important clues used to distinguish the genre of clothes.

With the media acquisition and processing technologies are rapidly growing, multimedia application plays important role in supporting our real daily life. Several clothes images retrieval and clothes recommendation systems have been proposed. Approaches proposed by Cheng *et al.* [2], Wang *et al.* [11], and Chen *et al.* [1] focused on retrieving clothing images by color and partial clothing attributes that are similar with query image given by user, regardless of matching clothing genre. "Mirror appliance" [7] and "What am I gonna wear?" [8] were clothing recommendation projects whereas the clothing genres were assigned manually. However, it is not quite feasible for the user to manually annotate all the image data if the image database is large. Automatic recognition of three upperwear genres for frontal-view images taken with homogeneous background has been proposed by Zhang *et al.* [12, 13].

In this work, we propose a novel framework for automatically recognizing the clothing genre, with an initial focus on upperwear clothes. We adopted the theory of clothing design to define eight kinds of the clothing genres as shown in Figure 1. The determination of the clothing genre for upperwear clothes are defined based on a set of basic style elements, as listed in Table 1. For extracting the style elements, in the proposed framework, we first employ an upperbody detector to get body parts of a person standing arbitrary pose in images. The extracted feature vectors of the style elements are then exploited for learning the genre models and used for predicting the clothing genre.

2. STYLE ELEMENTS OF CLOTHES

In this section, we describe all the style elements and their corresponding features used in the proposed algorithm that classifies a full-body person image as one of the following eight clothing genres: formal shirt (FS), henley shirt (HS), informal shirt (IS), long sleeves T-shirt (LS), polo shirt (PS), spaghetti (SG), tank top (TT), and T-shirt (TS). In the

MM'12, October 29–November 2, 2012, Nara, Japan.

Figure 1: The defined clothing genre.

Table 1: Measurement of five style elements for each clothing genre.

Genre	Collar	Print style	Shoulder skin	Front button	Sleeve
FS	yes	plain	no	full	long
HS	no	loud/plain	no	half	long/short
IS	yes	loud	no	full	long/short
LS	no	loud/plain	no	no	long
PS	yes	loud/plain	no	half	long/short
SG	no	loud/plain	yes, thin strap	no	no
TT	no	loud/plain	yes, thick strap	no	no
TS	no	loud/plain	no	no	short

framework, we use five style elements [15] (collar, print style, shoulder skin, front button, and sleeve) to discriminate different clothing categories. The measurements of these style elements for each clothing category are summarized in Table 1. Note that it might be simple for human eyes to determine the measurements of these style elements, however it is not trivial for automated image analysis. The idea and the design of corresponding features for these style elements will be described in the following subsections.

2.1 Collar Types

Collar is the part of a garment around the neck and shoulders. Clothes with collar are usually considered more formal than clothes without itWe utilize three features to detect the existence of collar. First, we employ body-part detector [3] to obtain the information of torso τ from the input full-body person image. Through [3], we will learn the location of torso τ. Since the collar will only appear in the upper part of the torso, we will calculate three collar features in the collar region $\Re_c$ which is the upper one-third area of τ.

We notice that a collar clothes has more salient corners compared with non-collar clothes. Therefore, we detect the location of all corners in $\Re_c$ through [4]. The output of the corner detection [4] is $c = [x_{C_1}, y_{C_1}; x_{C_2}, y_{C_2}; ...; x_{C_n}, y_{C_n}]$,

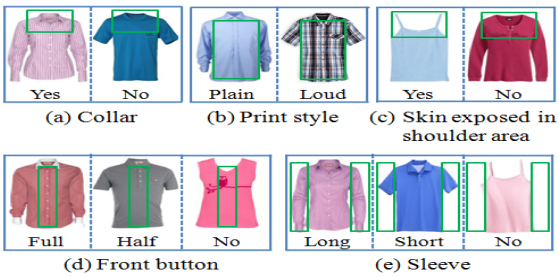

Figure 2: Five style elements employed in the proposed algorithm.

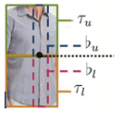

Figure 3: Four areas for front button region.

where (x_{C_i}, y_{C_i}) represents the normalized location of i^{th} detected corner, and n is the number of detected corners. We hence define f_{c_1} as $n/\Re_c$, f_{c_2} as variance$(X = x_{C_1}, ..., x_{C_n})$, and f_{c_3} as variance$(Y = y_{C_1}, ..., y_{C_n})$. Here f_{c_2} and f_{c_3} are used to distinguish collar corners from pattern collars since in general collar corners are distributed more closely while pattern corners are distributed more sparsely.

2.2 Print Style Types

Print style element is used to evaluate the pattern complexity of an input garment. As illustrated in Figure 2(b), we observe that patterned clothes have more corners and more colors in torso area as compared to plain clothes. Hence, we characterize the print style type by two features which reflect the number of detected corners and colors in τ, respectively.

Given $p = [x_{\tau_1}, y_{\tau_1}; x_{\tau_2}, y_{\tau_2}; ...; x_{\tau_m}, y_{\tau_m}]$ is the output of the corner detection [4] on the whole region of torso τ, where m is the number of detected corners. We therefore define $f_{p_1} = \frac{m}{A_\tau}$.

The second feature, f_{p_2}, is obtained as follows: we first convert the RGB color space to YCbCr color space. We then partition the torso region of the test image into a 5×5 grid of blocks. For each block b, we detect edges by canny operator in both Cb and Cr channels. We then calculate the number of pixels that are detected as edges for block b and define this number as n_{p_b}. We define N as the number of blocks that satisfy $n_{p_b} > T_1$. If $N < T_2$, then $f_{p_2} = 0$, otherwise f_{p_2} equals to the average number of colors among all blocks, where T_1 and T_2 are pre-defined thresholds. The number of colors in a block is obtained by Frequency Image method [6].

2.3 Shoulder Skin Types

From Table 1 and Figure 2(c), we know most clothes would cover the skin in shoulder area except SG and TT categories. Besides, the difference between SG and TT can be recognized based on the strap width. Therefore we propose to use three features for the shoulder skin style element. We define f_{e_1} as n_s/A_τ, where n_s is the number of pixels that are detected as skin; $(f_{e_2}, f_{e_3}) = (\frac{w_s}{w_\tau}, \frac{w_{sd}}{w_\tau})$, where w_s, w_{sd}, and w_τ indicate the strap width, the distance between two straps, and the torso width, respectively.

2.4 Front Button Types

Front button is a round fastener sewn to the center-front of a dress shirt, it is generally held by placket which comprises two layers of fabric. Since button detection is difficult, we convert the button detection problem into the placket detection problem.

The person in the input image may not stand straight, therefore once we obtained the torso information, we will first rotate the torso to straight position. Besides, the front buttons (or the corresponding plackets) are not always located in the center of the torso image as illustrated in Figure 3. Therefore we identify the physical center of the torso

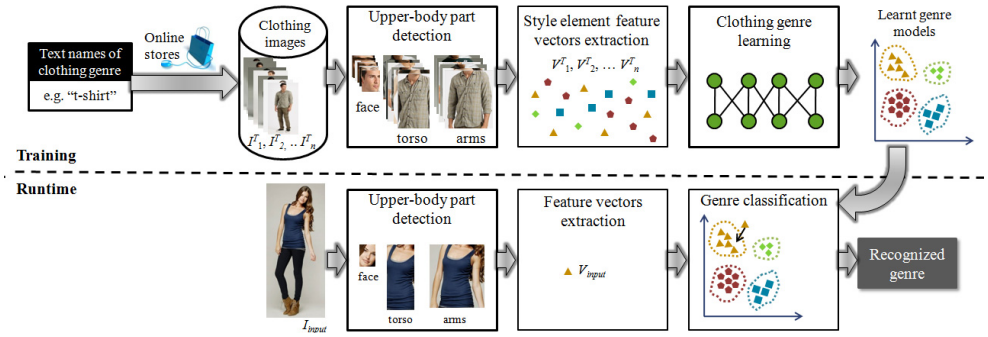

Figure 4: Overview of the proposed method.

by detecting the vertical lines in torso. Given a torso τ, we will first construct a histogram $\mathcal{H}$ with the size of torso width w_τ. $\mathcal{H}(i)$ indicates the number of pixels that are detected as vertical lines at the i^{th} location on the x-coordinate. We will then calculate the center of the x-axis of the front button area as follows:

$$x_c = \frac{\sum_{i=1}^{w_\tau} \mathcal{H}(i) \times i}{\sum_{i=1}^{w_\tau} \mathcal{H}(i)}.$$

We define four regions τ_u, τ_l, $\flat_u$, and $\flat_l$ as the upper-half part of the torso τ, the lower-half part of τ, the upper-half part of the front button area, and the lower-half part of the front button area. Note that the width of $\flat_u$ and $\flat_l$ is $\frac{1}{4}w_\tau$. Therefore four front button features f_{b_1}, f_{b_2}, f_{b_3}, and f_{b_4} are calculated as the number of pixels that are detected as vertical lines in regions τ_u, τ_l, $\flat_u$, and $\flat_l$, respectively.

2.5 Sleeve Types

Sleeve is the part of a garment covering all or part of an arm. As shown in Table 1 and Figure 2(e), sleeve length can be classified into three cases: long-sleeve, short-sleeve, and no-sleeve. Once we knew the torso location, we could easily obtain the arm location. The area of skin exposed in the arm location is inversely proportional to the sleeve length, hence we define feature $f_s = \frac{n_{\texttt{arm}}}{A_\tau}$, where $n_{\texttt{arm}}$ is the number of skin pixels detected in the arm location.

3. CLOTHING GENRE CLASSIFICATION

Our proposed method tackles the problem of recognizing clothing genre for a given full-body person image. This framework is shown in Figure 4. We divide our operations into two phases, i.e. training phase and runtime phase.

In training phase, our goal is to build learnt genre models. We collected images of full-body person, $I_i^T(i = 1, 2, ..., n)$, through online stores. Our queries were text names of the clothing genres, e.g. "spaghetti", "tank top", and "T-shirt". These I_i^T images are used as our training dataset. We execute three operations in training phase. These operations are as follows:

1. Apply upper-body detector [3] for $I_i^T(i = 1, 2, ..., n)$ in order to detect torso region τ and area gained from combination of left arm, torso, and right arm (called arm region $\mathcal{A}$). To do upper-body estimation, firstly we must provide a bounding-box around head and shoulders of person that will be detected in the image. Input parameters for this bounding-box including its minimum x-coordinate as $x_{bb} = \min(x_\mathcal{F}) - 1.5w_\mathcal{F}$, minimum y-coordinate as $y_{bb} = \min(y_\mathcal{F}) - h_\mathcal{F}$, height as $h_{bb} = 4h_\mathcal{F}$, and width as $w_{bb} = 4w_\mathcal{F}$; where $x_\mathcal{F}$, $y_\mathcal{F}$, $h_\mathcal{F}$, and $w_\mathcal{F}$ correspond to x-coordinate, y-coordinate, height, and width of face region $\mathcal{F}$ respectively. We employ Viola-Jones [14] face detector to detect the presence of $\mathcal{F}$ in the image.

   Because person in the image was not captured in exactly frontal viewpoint, we rotate torso area by

   $$\theta^o = \arctan\left(\frac{y_{\min(x_\tau)} - y_{\max(x_\tau)}}{w_{x_\tau}}\right) \times \frac{180^o}{\pi},$$

   where $y_{\min(x_\tau)}$ denotes y-coordinate of minimum x-coordinate of τ and $y_{\max(x_\tau)}$ denotes y-coordinate of maximum x-coordinate of τ. This rotation is in a counterclockwise direction around its center point.

2. Construct feature vectors for each $I_i^T(i = 1, 2, ..., n)$, denoted as $V_i^T(i = 1, 2, ..., n)$. $V_i^T(i = 1, 2, ..., n)$ consists of combination of all extracted features $V_i^T = [f_c, f_p, f_e, f_b, f_s]_i^T$. We use τ to extract f_c, f_p, f_e, and f_b. While $\mathcal{A}$ is used to extract f_s. Feature extractions for f_e and f_s are based on the detected skin in τ and $\mathcal{A}$, respectively.

3. Give label for each $I_i^T(i = 1, 2, ..., n)$ in 8 categories manually. We employ non-linear SVM classifier to learn the normalized feature vectors $V_i^T(i = 1, 2, ..., n)$ and its label, in order to have learnt genre models.

Given a full-body person image I_{input} in the runtime phase, our system performs three operations to recognize the clothing genre for I_{input}. These operations are as follows:

1. Similar to face and upper-body detection in training phase, apply face detector [14] to detect $\mathcal{F}$ and then apply body part detector [3] to detect τ and $\mathcal{A}$. After that, we rotate τ by θ^o in a counterclockwise direction around its center point.

2. Extract the defined style elements for I_{input}. Extracted features for I_{input} is similar to extracted features for $I_i^T(i = 1, 2, ..., n)$, i.e. $V_{input} = [f_c, f_p, f_e, f_b, f_s]_{input}$.

3. Predict classification genre for V_{input} using learnt genre model got from training phase.

4. EXPERIMENTS

In this section, we describe our experiments and evaluation of the performance in order to validate effectiveness of our approach.

Table 3: Confusion matrix

Ground truth	Classification results								
	FS	HS	IS	LS	PS	SG	TS	TT	**Recall**
FS	**18**	0	0	0	0	0	0	0	**100.00%**
HS	0	**18**	0	2	0	0	0	1	**85.71%**
IS	0	0	**20**	0	0	0	0	0	**100.00%**
LS	0	1	0	**17**	0	0	0	1	**89.47%**
PS	0	0	0	0	**25**	0	0	0	**100.00%**
SG	0	0	0	0	0	**10**	2	0	**83.33%**
TS	1	0	1	0	0	0	**36**	0	**94.74%**
TT	0	0	0	0	0	3	0	**19**	**86.36%**
Precision	**94.74%**	**94.74%**	**95.24%**	**89.47%**	**100.00%**	**76.92%**	**94.74%**	**90.48%**	

Table 2: Number of images for each category in our database.

Genre	FS	HS	IS	LS	PS	SG	TT	TS
# of Images	108	134	122	119	153	82	134	225

Table 4: Effective analysis of the proposed features (see Section 4 for details).

	Precision		Recall		F-score	
Overall	92.04		92.45		92.25	
No f_{c_1}	88.63	-3.41	89.66	-2.79	89.15	-3.10
No f_{c_2}	85.96	-6.08	88.69	-3.76	87.30	-4.95
No f_{c_3}	85.65	-6.39	88.75	-3.70	87.17	-5.08
No f_{p_1}	86.36	-5.68	86.87	-5.58	86.61	-5.64
No f_{p_2}	89.02	-3.02	91.42	-1.03	90.20	-2.05
No f_{e_1}	76.44	-15.60	79.23	-13.22	77.81	-14.44
No f_{e_2}	88.84	-3.20	90.81	-1.64	89.81	-2.44
No f_{e_3}	86.96	-5.08	88.18	-4.27	87.57	-4.68
No f_{b_1}	87.54	-4.50	90.28	-2.17	88.89	-3.36
No f_{b_2}	86.01	-6.03	87.17	-5.28	86.58	-5.67
No f_{b_3}	84.52	-7.52	85.97	-6.48	85.24	-7.01
No f_{b_4}	85.10	-6.94	87.47	-4.98	86.27	-5.98
No f_s	79.70	-12.34	77.63	-14.82	78.66	-13.59

We build a garment image dataset, consisting of 1,077 clothing images, from popular e-commerce web sites. Each image is a full body image without any restriction of races, genders, ages, poses, and image resolutions. The number of images for each category is shown in Table 2. The entire image database is randomly divided into two sets: one contains 85% of images from each category and is used for training; the other set is used for testing.

In Table 3, we present the confusion matrix for our proposed framework. Each column of the matrix represents the instances in a predicted class, while each row represents the instances in an actual class. On average, the proposed algorithm achieved overall precision of 92.04%, recall of 92.45%, and F-score of 92.25%.

In Table 4, we investigate the strength of each feature. We remove the features one at a time, repeat the experiment, and record the corresponding precision, recall, F-score values, and the degradation percentage as compared to the original scheme which contains the complete set of features, in each row of Table 4. Experimental results show that the removal of any of the proposed features would cause performance degradation, this confirms the effectiveness of all proposed features.

5. CONCLUSION

We have proposed a novel framework for fully automatic clothing genre recognition. Experimental results show that the proposed algorithm achieved overall precision of 92.04%, recall of 92.45%, and F-score of 92.25% on a dataset of more than 1.000 images and demonstrated the effectiveness of our approach.

Acknowledgments

This work was supported in part by National Science Council of Taiwan via NSC101-2221-E-011-138, NSC101-2221-E-001-016, NSC100-2218-E-011-021, and NSC100-2221-E-001-024.

6. REFERENCES

[1] Z. Chen *et al.* Generating vocabulary for global feature representation towards commerce image retrieval. In *ICIP'11*.

[2] C.-I. Cheng and D. S.-M. Liu. An intelligent clothes search system based on fashion styles. In *ICMLC'08*.

[3] M. Eichner *et al.* 2d articulated human pose estimation and retrieval in (almost) unconstrained still images. Technical Report 272, ETH Zurich, 2010.

[4] X.-C. He and N. H. C. Yung. Corner detector based on global and local curvature properties. *Optical Engineering*, 2008.

[5] P. Kakumanu *et al.* A survey of skin-color modeling and detection methods. *Pattern Recognition*, 2007.

[6] T. Kashiwagi and S. Oe. Introduction of frequency image and applications. In *SICE*, 2007.

[7] S. Nagao *et al.* Mirror appliance: Recommendation of clothes coordination in daily life. In *HFT*, 2008.

[8] E. Shen, H. Lieberman, and F. Lam. What am i gonna wear?: Scenario-oriented recommendation. In *IUI*, 2007.

[9] L. Svendsen. *Fashion: A philosophy*. Reaktion Books, 2006.

[10] H. Tsujita *et al.* Complete fashion coordinator: A support system for capturing and selecting daily clothes with social network. In *AVI*, 2010.

[11] X. Wang and T. Zhang. Clothes search in consumer photos via color matching and attribute learning. In *ACM Multimedia*, 2011.

[12] W. Zhang *et al.* Real-time clothes comparison based on multi-view vision. In *ICDSC* 2008.

[13] W. Zhang *et al.* An intelligent fitting room using multi-camera perception. In *IUI*, 2008.

[14] P. A.Viola and M. J.Jones. Robust real-time face detection. *IJCV*, 2004.

[15] Wikipedia Shirt — Wikipedia, The Free Encyclopedia. 2012. `http://en.wikipedia.org/wiki/Shirt` [Online; accessed 09-November-2011]