SPARSE REPRESENTATION BASED VISUAL ELEMENT ANALYSIS

Xue Li, Hongxun Yao*, Xiaoshuai Sun, Rongrong Ji, Xianming Liu, Pengfei Xu
School of Computer Science and Technology, Harbin Institute of Technology, Email: h.yao@hit.edu.cn

ABSTRACT

Modern clothes are designed based on various visual elements of different fashion styles. Traditional vision-based clothes recommendation methods focused on searching clothes which are similar with user preferred samples in the aspects of colors and partial shape elements. In this paper, we propose a method of recommending clothes by mining visual elements of different fashion styles. Independent Component Analysis (ICA) is employed to extract sparse features, and then Term-Frequency (TF) analysis is applied to discover visual elements from these independent components. Finally, we test three ranking metrics for clothes recommendation including Euclidean distance of TFs, Cosine distance of TFs and Minimum TF. Experimental results based on web commercial images demonstrate the effectiveness of the proposed method.

Index Terms—Clothes recommendation, independent component analysis, term frequency, style mining

1. INTRODUCTION

We are always confused about what to wear every day, or how to choose from the sea of clothes in the online shops. Unlike other online media rankings [1][2], clothes images have unique characters. Several clothes recommendation systems have been reported. Previous works focused on increasing the harmony between different dressing items, especially for color information. Harmonic rules are introduced by Tokumaru et al. [3] in “Color design support system”, and then these rules are further applied to online clothes shopping in the Virtual Stylist Project [4] composing a virtual fitting system by synthesizing the dressing image. The related works in recent years tended to recommend clothes appropriate for the occasions or environments. The “What am I gonna wear?” project at MIT [5] was a system that recommended what to wear according to the real-life scenarios indicated by users. “Mirror Appliance”[6] referred to the current weather information and the user’s current schedule to help find the right clothes.

However, traditional clothes recommendation systems ignore a fact that clothing has become a constantly changing trend. To search for an up-to-date dress seems more necessary and challenging compared with traditional clothes recommendations. When purchasing new clothes, we would like to ask ourselves, “What’s new?”, “What will make me look different from last year?”, that’s the fashion trends that consumers are seeking for, and from which we start our recommendation system.

Without professionalism, appreciation of the arts, or keen sense of fashion, few people can predict the latest ideas or seize the most fashionable pulse. For this reason, we intentionally or unintentionally pay attention to the latest clothing tips from the trend forecasting and reporting made by those fashion experts and pioneers, and knows about the refreshing fashion elements nowadays from the suddenly new collections. For instance, one of the hottest trends for Fall & Winter 2010 is the leopard print, which is widely used in overcoats, dresses, and footwear. Back to 2009, punk rock fashion made a strong revival, and designers used a number of rivets which is a main element of this trend. Thus, what are in vogue should be these fashion elements. Furthermore, to present some kind of trend (punk, countryside, royal, etc.), it can come down to the fashion element.

In view of the above, our work 1) summarizes the fashion trend by analyzing the visual fashion elements from the pictures of world-wide fashion shows and collections, and 2) recommends fashionable clothes conforming to the trend of a specific brand and time. However, the proposed method is related to but also quite different from traditional content based image retrieval techniques. ICA is applied to extract visual elements, which reflect certain fashion styles or user preference, from a group of pictures instead of one single query image. As shown in Figure 1, group pictures can be used to learn fashion styles or user intentions (user preferred visual styles), which a single image is nearly impossible to tell.

![Figure 1](image-url)
2. THE METHOD

The framework of our method is shown in Figure 2. To capture the design style of a certain fashion trend, we use a group of clothes images obtained from related fashion shows or collections or the users as query images. First, ICA is employed to model the images in short-term sparse representation [7] and a group of independent components are obtained from these query images. Then, visual elements are mined from those independent components using statistical methods. Finally, images in the database are ranked based on the similarity between the visual elements and input clothes.

2.1. Mining Visual Elements

2.1.1 Adaptive Sparse Representation of Clothes

Clothing image is usually affected by its outline shape, print types, dominant color, and fabric material [8]. Color and Edge [3][4][8] are two most popular features for clothes representation in previous clothes recommendation systems. However, a large portion of information may be lost during the traditional feature extracting stage, where style elements are decomposed into an incomplete feature space. It is difficult to learn an existing visual element if it is incompletely represented. In order to project the image into an effective feature space while maintaining as much information as possible, the short-term sparse features [7] learned by independent component analysis (ICA) are employed instead of traditional color and edge features.

ICA [9] was first proposed for the blind signal segmentation problem, and later proved to be an effective model for sparse representation. Based on ICA framework, Sun et al. [7] proposed a short-term sparse representation method utilizing contextual constrained training scheme. Sun’s method is able to provide a better sparse representation for natural visual contents with zero information lost and improved sparsity. In this work, the ICA based sparse representation method is selected for the following two reasons:

1) The basis functions learned by ICA are quite similar, both in form and functionality, with the receptive fields of simple cells in primary visual context, which means they provide a similar feature space used by human vision system for visual sensing and perception.

2) The coefficients of each independent component are statistically independent which allows us to analysis each feature channel separately. Besides, from visual observation, we find that some of the independent components appear to be exactly the visual elements of the training clothes.

In order to bring contextual constraints to our clothes representation learning strategy, we process the clothes images in semantic groups which are defined by analyzers or common users. Image patches are sampled from the given group of images by scanning every picture in a block by block manner. Each image patch is resized from 15×15×3 (Height×Width×Color) to a 1-D vector. All patch vectors are then combined into one single data matrix, which is further processed by FastICA [10] algorithm. Basis functions learned from a picture group composed of clothes sharing similar visual elements are shown in Figure 3.

![Figure 3. Independent components learned from a semantic group. The left is typical clothes image selected from a semantic group and the right is the Independent Components learned from the image group.](image)

2.1.2 Visual Elements Analysis

Features that commonly appeared in a group of clothes are probably the core visual elements for the corresponding design style. To discover this kind of features, we introduce a classic concept called TF/IDF (term frequency–inverse document frequency), which is densely used in textual information retrieval research. TF is defined as the frequency of a given term appearing in a specific document and IDF measures the general importance of a term in all documents, by counting the documents the term appears in. TF can be used to measure the importance of an element in a data group. Similar with text, TF of a given independent component can be defined as:
By quantitatively analyzing the elements, we can draw a conclusion that independent components are with much larger ones. Based on this observation, we can draw a conclusion that independent components with small non-zero function, IG is the image group. Visual elements can be divided into main elements and affiliate elements. The visual elements obtained in Section 2.1, we can further divided into main elements and affiliate elements. The visual elements tend to own small independent components for different images, visual elements mining in different clothes collections.

\[ TF(I, IC) = \frac{N(I, IC)}{\sum_{i} N(I, IC)} \]  

(1)

where \( I \) is a given image, \( IC_i \) is the \( i \)th independent component, \( N(I, IC) \) is the number of patches with activated \( IC_i \) in \( I \). Given \( P \) as the \( m \)th vectored patch sampled from image \( I \), \( N(I, IC) \) can be calculated by Eq. 2-3.

\[ N(I, IC) = \sum_{m} \delta(P_m * IC_i) \]  

(2)

\[ \delta(x) = \begin{cases} 1 & x > \alpha \\ 0 & \text{else} \end{cases} \]  

(3)

By quantitatively analyzing the TF values of different independent components for different images, visual elements tend to own small TFs (non-zero) and background elements are with much larger ones. Based on this observation, we can draw a conclusion that independent components with small non-zero TFs tend to be the target visual elements. Besides, visual elements can be further divided into main elements and affiliate elements. The TF values of main visual elements among different images tend to be stable which means they can be easily distinguished from affiliate elements according to TF variance. Based on the analysis above, the probability for an independent component to be a main visual element of a group of clothes images is defined as:

\[ P(IC_i) = \frac{1 - E(TF(I, IC_i))}{\text{var}(TF(I, IC))}, I_i \in IG \]  

(4)

where \( E(.) \) is expectation function, \( \text{var}(.) \) is variance function, IG is the image group. Visual elements can be directly obtained by selecting the top \( N \) elements from the independent components according to the probability in Eq.4. Figure 4 shows some visual results of visual elements mining in different clothes collections.

3. EXPERIMENTAL RESULTS

3.1. Data Collection

We build a clothes image dataset, which consists of 600 garment images from popular B2C e-commerce websites (www.amazon.com, www.bluefly.com, etc.) and 600 images from fashion collections in different brands and seasons in recent years, which are divided into 20 tags such as D&G Spring & Summer 2011 and Christian Dior Spring & Summer 2011. All images are resized to 320×250 to reduce computation cost. 10 subjects (including 6 female and 4 male, aged between 21 and 29) were asked to grade the images in the database with different tags. 0-10 represents different confidential levels in which 10 means the image definitely belongs to the category and 0 means definitely not.

3.2. Experimental Settings

After ranking the images in the database for each query image group by adjusting the numbers of filtering the visual elements and adopting the above three ranking rules, we calculate the average scores of top 20 images for every query, which is the final score for the testing method.

3.3. Results

There are 7 query groups altogether, each of which includes 5 typical images selected from the corresponding fashion shows. For each query, 9 groups of ranking results are obtained corresponding to different parameter settings and ranking metrics. We use average confidential score to evaluate each ranking results. Figure 5 shows quantitative comparisons and Figure 6 shows visual results. As shown in these results, the parameter for minimum TFs has large influence on the results. When the parameter is too small
(e.g. 5) or too large (e.g. 50), the performance becomes worse. Balmain 2010 is a special case, since the dress collocation in this collection is relatively much simpler, and the actual number of visual elements is much smaller than 25. As shown in the comparison when the parameter is fixed as 25, the minimum $\text{TF}_s$ performs best among three different ranking methods.

Figure 5. Quantitative results. Top box: performance of Minimum $\text{TF}_s$ using 5, 25 and 50 visual elements; bottom: performance of 25 elements with three ranking metrics.

4. CONCLUSIONS

In this paper, we propose a novel clothes recommendation method by mining visual elements of different fashion styles. We adopt ICA to extract sparse features and apply Term-Frequency analysis to mine visual elements which are further used for clothes ranking. The proposed method is able to capture the design style of a certain fashion trend by mining a group of clothes images obtained from related fashion shows or collections, and is demonstrated to be effective on web commercial data.

5. ACKNOWLEDGMENTS

This work is supported by the National Natural Science Foundation of China (No. 61071180) and NEC cooperative project (No. LC04-20101201-03).

6. REFERENCES