What are the Fashion Trends in New York?

Shintami C. Hidayati[1,2], Kai-Lung Hua[1], Wen-Huang Cheng[2], and Shih-Wei Sun[3]
[1]Dept. of CSIE, National Taiwan University of Science and Technology, Taipei, Taiwan (R.O.C)
[2]Research Center for Information Technology Innovation, Academia Sinica, Taipei, Taiwan (R.O.C)
[3]Dept. of New Media Art, Taipei National University of the Arts, Taipei, Taiwan (R.O.C)
{d10115802, hua}@mail.ntust.edu.tw, whcheng@citi.sinica.edu.tw, swsun@newmedia.tnua.edu.tw

ABSTRACT

Fashion is a reflection of the society of a period. Given that New York City is one of the world's fashion capitals, understanding its change in fashion becomes a way to know the society and the times. To keep up with fashion trends, it is important to know what's "in" and what's "out" for a season. Though the fashion trends have been analyzed by fashion designers and fashion analysts for a long time, this issue has been ignored in multimedia science. In this paper, we present a novel algorithm that automatically discovers visual style elements representing fashion trends for a certain season. The visual style elements are discovered based on the stylistic coherent and unique characteristics. The experimental results demonstrate the effectiveness of our proposed method through a large number of catwalk show videos.

Categories and Subject Descriptors

H.3.3 [**Information Search and Retrieval**]: Retrieval Models

Keywords

Clothing image analysis; fashion trend; data mining; visual summarization

Figure 1: Change of the women's dress fashion on New York Fashion Week, where models present the fashions during a catwalk fashion show.

1. INTRODUCTION

New York City (NYC) is one of the world's fashion capitals which have major influences on international fashion trends. Understanding the change in fashion becomes an important way to know the society and the times. Therefore, we see there is a need for fashion trends spotting.

Particularly, fashion is a style in clothing, footwear, accessories, or makeup that have been carried by a wide audience during a particular time. Fashion trends reflect to what's "in" and what's "out" for the season. So far, fashion trend analysis is mainly performed by fashion experts who have intensive fashion knowledge. However even for fashion experts, it is time-consuming to collect and analyze fashion trends since there are numerous fashion shows each year. In literature, clothing recognition [1], segmentation [2], and recommendation [3] have received much attention in computer vision and multimedia research. However, to the best of our knowledge, there is no automatic fashion trends analysis framework existed. Hence, this paper is the first work that offers this service.

One major channel to determine fashion trends is fashion show [4], where the latest fashion trends are debuted. In this event, fashion designers showcase their upcoming lines of clothing through models during fashion week, as illustrated in Figure 1. According to [5], color, cut, head decorations, and pattern, as illustrated in Figure 2, are four major elements to investigate fashion style. These fashion elements are considered as fashion trends at a particular fashion week if they satisfy two requirements: 1) coherence, i.e. frequently occurring within a fashion week, and 2) uniqueness, i.e. they occur much more often in a fashion week than in other fashion weeks.

In this work, therefore, we propose a novel framework for automatically discovering visual style elements that well represent fashion trends, with initial focus on New York Fashion Week (NYFW). Given fashion show videos, we first detect

MM'14, November 3–7, 2014, Orlando, Florida, USA.

http://dx.doi.org/10.1145/2647868.2656405 .

Figure 2: Four major elements for fashion style investigation.

the catwalk model. Then, we seek patterns that are both frequently occurring within the given fashion week event and unique for this fashion week, by integrating unsupervised and supervised learning. As research pioneers in this field, we collected a fashion show image dataset that are extracted from videos [6] of several designers that have presented their collections at ten different seasons of NYFW. Our experiments confirmed that the proposed algorithm offers a satisfactory fashion trends identification through subjective and objective evaluations.

2. SYSTEM FRAMEWORK

In this section, we reveal the details of how the proposed framework exploits NYFW videos available on Fashion TV [6] as the input to discover fashion trends at NYFW.

2.1 Catwalk Model Detection

In order to identify fashion styles during fashion week events, we need to identify catwalk models first, so that we will be able to extract photos with showcased garments. Catwalk models are people who demonstrate clothing and accessories on a long and narrow stage, called catwalk, at fashion shows. Let $\vartheta_v^{\omega,x}$ denotes fashion show video of designer x who has an exhibition at NYFW season ω. Assume there are M frames in $\vartheta_v^{\omega,x}$, that is, $\vartheta_v^{\omega,x} = \{\mu_{v1}^{\omega,x}, \mu_{v2}^{\omega,x}, ..., \mu_{vM}^{\omega,x}\}$, we represent our dataset containing fashion show videos as $\mathcal{D}_v = \{\vartheta_v^{1,1}, \vartheta_v^{1,2}, ..., \vartheta_v^{1,G}, \vartheta_v^{2,1}, \vartheta_v^{2,2}, ..., \vartheta_v^{E,G}\}$, where E is the number of NYFW events and G is the number of designers.

Following [7], we track the presence of faces over all video frames in $\vartheta_v^{\omega,x}$ and then identify if the detected faces belong to the same person. Since catwalk models in general show their frontal faces and the audiences show their side faces, we further employ a constraint for frontal faces when using [7]. Next, we make use of the detected faces to obtain model's full-body by employing [8]. Through [8], which estimates the spatial layout of humans via pictorial structure model, we obtain a full-body image with clean background. We denote $\vartheta_i^{\omega,x}$ be a subset of $\vartheta_v^{\omega,x}$, where $\vartheta_i^{\omega,x}$ contains only one frame for distinct models while $\vartheta_v^{\omega,x}$ contains more than just one frame for a model. A set containing full-body image of catwalk models in our dataset can be represented as $\mathcal{D}_i = \{\vartheta_i^{1,1}, \vartheta_i^{1,2}, ..., \vartheta_i^{1,G}, \vartheta_i^{2,1}, \vartheta_i^{2,2}, ..., \vartheta_i^{E,G}\}$.

2.2 Visual Style Element Initialization

Having the collection of full-body image of catwalk models, $\mathcal{D}_i$, we divide it into two sets: (1) one set containing catwalk models from the selected fashion week event whose visual style elements of the fashion trends are to be discovered, namely positive set $\mathcal{D}_i^+$ (e.g. catwalk models at NYFW S/S 2014), and (2) the other set containing catwalk models from other NYFW events, namely negative set $\mathcal{D}_i^-$ (e.g. all catwalk models at NYFW except from NYFW S/S 2014).

Table 1: Number of fashion styles in our dataset.

Designer \ Season	Spring / Summer 2010	Autumn / Winter 2010	Spring / Summer 2011	Autumn / Winter 2011	Spring / Summer 2012	Autumn / Winter 2012	Spring / Summer 2013	Autumn / Winter 2013	Spring / Summer 2014	Autumn / Winter 2014
BCBG Max Azria	31	29	37	34	30	34	34	30	32	32
Calvin Klein	35	34	34	34	32	33	33	32	35	34
DKNY	46	40	37	34	13	40	48	40	47	42
Nicole Miller	39	37	37	34	39	36	39	44	39	39
Ralph Lauren	50	50	57	53	33	56	52	58	52	38
Rebecca Taylor	44	34	35	39	41	40	33	36	40	41
Richard Chai	26	27	29	28	26	28	27	26	31	27
Ruffian	33	34	36	32	30	31	13	32	31	23
Vera Wang	31	39	44	36	40	39	38	48	40	40
Total	**335**	**324**	**346**	**324**	**284**	**337**	**317**	**346**	**347**	**316**
	3,276									

The representations of visual information at a pixel level generally could not provide enough information to support a useful feature. In addition, using the whole image would require huge amounts of training data to represent spatial configuration of objects. Thus, we consider representing visual elements of fashion styles using image patches. We partition every full-body image of catwalk model from both $\mathcal{D}_i^+$ and $\mathcal{D}_i^-$ into non-overlapping square patches.

As aforementioned, the coherence (to occur frequently) and uniqueness (to be sufficiently different from other fashion shows) of visual style elements are the two key factors for discovering fashion trends. If we somehow knew that a set of patches of the dataset had similar visual style element, we could train a discriminative classifier to build a similarity metric of visual features for these patches. Therefore, we first initialize cluster of patches.

We capture the 4 major elements of fashion trends (Figure 2) by exploiting color and texture information. For color information, the extracted features include HSV color histogram, color correlogram [9], and color moments. For texture information, the extracted features include HOG [10], Haar wavelet transform, and 2D Gabor wavelets [11]. Thus, throughout our algorithm, we represent each patch using 3 different type of focus: (1) color features to discover fashion trends in terms of color information; (2) texture features to discover fashion trends in terms of texture information; and (3) the combination of color and texture features to discover fashion trends in terms of both color and texture information.

To determine candidate patches that might represent visual style element of fashion trends, for each patch in positive set, we search the top 20 nearest neighbor patches in the full dataset of $\mathcal{D}_i$ (both positive set, $\mathcal{D}_i^+$, and negative set, $\mathcal{D}_i^-$) and we use normalized correlation distance as the similarity measure. Since patches portraying stylistic coherence and uniqueness of visual style elements in positive set tend to have higher proportion to match similar patches in positive set, we keep the patches that pass the following rule: for any patch, if there are more than 10 patches, from its top 20 nearest neighbors, belong to $\mathcal{D}^+$. Given there are P number of surviving candidate patches from previous step, we then cluster similar visual style elements by running the standard k-means clustering. In our experiments, we set k with a quite high value of $k = \frac{P}{5}$ since k-means clustering tends not to generalize high-dimensional data well. After that, we further remove candidate patches from clusters that have less than 3 members. Finally, we obtain groups of similar visual style element as the initial cluster.

Table 2: The average accuracy of our fashion trend element detector.

		Extracted Features		
		Color	Texture	Color + Texture
NY Fashion Week	Spring/Summer 2010	80.33 %	90.67 %	88.00 %
	Autumn/Winter 2010	64.67 %	91.00 %	88.67 %
	Spring/Summer 2011	87.00 %	92.00 %	92.33 %
	Autumn/Winter 2011	67.00 %	74.00 %	72.33 %
	Spring/Summer 2012	90.67 %	88.33 %	89.00 %
	Autumn/Winter 2012	50.33 %	67.67 %	60.33%
	Spring/Summer 2013	67.67 %	84.67 %	83.67 %
	Autumn/Winter 2013	66.33 %	65.67 %	68.00 %
	Spring/Summer 2014	94.67 %	90.67 %	92.33 %
	Autumn/Winter 2014	81.00 %	76.00 %	71.67 %

2.3 Iterative Clustering

As discussed in [13], the standard distance metric of unsupervised clustering (e.g. k-NN and k-means clustering) does not work well for capturing the important parts within medium-sized patches because it treats all the pixels equally. If the candidate patch has irrelevant visual representation, the retrieved matches from unsupervised clustering will also have that irrelevant element. Thus, to improve stylistic coherence and uniqueness of patches that serve as visual style elements of fashion trends, we learn the distance metric in an iterative manner to keep improving our initial visual element detector obtained from the previous step.

To be more specific, with an initial set of clusters containing the surviving candidate patches, we train a linear SVM classifier [12] for each cluster, treating patches within the cluster as positive examples and all negative-set patches as negative examples. However, when we iterate the SVM learning directly by utilizing the top t detectors from previous round as positive set, it would just improve the top t detectors a little bit due to the overfitting problem [13].

To prevent overfitting, we further apply cross-validation at each round. We divide both positive and negative sets into two equal, non-overlapping subsets. At each iteration, we run the detectors trained on the previous round to select the top t detection for retraining. We form new clusters from the top t firings of each detectors by considering all SVM scores above -1 to be firings. We repeat this procedure for 4 iterations, in which most of top t patches in a cluster converge. In this work, we set $t = 5$ for all experiments.

After the final iteration, we sort the resulting detectors according to the percentage of top 30 firings that are in the positive set. We return the top ten detectors as the visual style elements that represent fashion trends at a certain season of New York Fashion Week.

3. EXPERIMENTAL RESULTS

3.1 Dataset and Experimental Settings

Dataset: In this work, we constructed a fashion show dataset composed of images with the resolution of 1024×1024, these images are extracted from Fashion TV [6] through the proposed method explained in section 2.1. As shown in Table 1, our dataset collected 3,276 different catwalk models that had been presented by nine fashion design companies for the past 10 seasons of NYFW.

Experimental settings: We represented visual style elements of catwalk models by square patches at the size of 128×128. For color features, we quantized HSV color histogram of each patch into $8 \times 2 \times 2$ color bins and quantized color correlogram of each patch in RGB space into $4 \times 4 \times 4$ bins. For HOG feature, the cell size is set as 8×8 with a stride of 8 pixel/cell. During iterative learning, we employed a linear SVM [12] with C fixed to 0.1.

3.2 Results and Validation

Figure 3 shows the most representative fashion trends of different NYFW identified from the proposed algorithm, in which they are both coherent (frequently occurring within a fashion week) and unique (recognizable from other fashion weeks). For each subfigure of Figure 3, the element inside the dotted line rectangle indicates the identified visual trend at a certain NYFW event. For example, pastel green is identified as the most representative color trend at NYFW based on the investigated nine designers exploited for the season S/S 2014. That is pastel green is widely applied for the season S/S 2014 at NYFW, but it is unlikely to be observed elsewhere. As for the season A/W 2010, the high bun together with a neck scarf is identified as the most representative fashion trend in terms of texture information. Note that though the style (high bun with a neck scarf) occurs more frequently for the season A/W 2010, this element might not be beneficial to help identify color trend in this season since the colors of bun and scarf varied and hence it does not satisfy “stylistic coherence” which is one of the key fashion trend criteria. Besides, based on the discovered fashion patch from our proposed framework, we are capable of telling if the fashion trend of a certain season focuses on a specific location or spreads to several parts of the clothing and/or accessories. A practical prototype of the proposed system is demonstrated in Figure 4.

As a quantitative evaluation, for each top 10 fashion trend elements identified from our proposed algorithm for each season, we computed the percentage of instances corresponding to the positive set out of the top 30 extracted fashion styles. This evaluation is performed on an unseen dataset, containing 50% from positive set and 50% from negative set. Table 2 reports the average accuracy of our fashion trend detector for each season. Based on the high accuracy numbers showed in Table 2, we conclude that our proposed framework is very effective to discover both the stylistic coherence and the uniqueness of visual style elements over a variety of fashion shows. Besides, it can be learned from this table that fashion week with higher average accuracy has more coherent and unique fashion styles comparing to fashion week with lower average accuracy (e.g. the average accuracy of A/W 2012 is low and this implies the fact that the designers of A/W 2012 had less fashion elements in common comparing to the other seasons). Moreover, considering the highest average accuracy among three types of features (color, texture, color+texture) for each fashion week shown in Table 2, it is also interesting to note that from S/S 2010 to S/S 2013, the NYFW designers had put more concern in designing new textures rather than employing new color, while for S/S 2014 and A/W 2014 they had put more efforts in finding new colors and color schemes.

Figure 3: The most representative fashion trend element at NYFW.

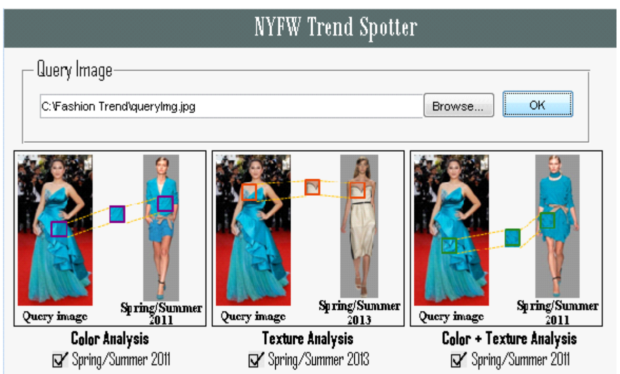

Figure 4: Interface of the proposed system: recognized fashion elements and their corresponding fashion events given a photo query.

4. CONCLUSIONS

In this paper, we presented a novel approach for automatic fashion trends analysis. As research pioneers in this field, we collected a fashion style image dataset that are extracted from fashion show videos of several designers that have presented their collections at ten different seasons of New York Fashion Week (NYFW). By investigating the coherence and uniqueness of four major style elements, i.e. color, cut, head decoration, and pattern, with iterative clustering techniques, we showed what are the fashion trends in New York for the past ten seasons. Our experiments confirmed that the proposed algorithm offers a satisfactory fashion trends identification through subjective and objective evaluations.

5. REFERENCES

[1] S. C. Hidayati *et al.* Clothing Genre Classification by Exploiting the Style Elements. In *ACM MM, 2012*.

[2] Y. Kalantidis *et al.* Getting the look: clothing recognition and segmentation for automatic product suggestions in everyday photos. In *ICMR, 2013*.

[3] S. Liu *et al.* Hi, magic closet, tell me what to wear!. In *ACM MM, 2012*.

[4] Vogue. Fashion show — Vogue, Fashion magazine. 2013. `http://www.vogue.co.uk/fashion/trends`

[5] CNN.com International. 2013. `http://edition.cnn.com/interactive/2013/09/` newblock `living/fashion-week-trends/`

[6] Fashion TV Network's Official Websites. 2013. `http://www.fashiontv.com/videos/fashion-weeks`

[7] J. Sivic *et al.* "Who are you?" – Learning Person Specific Classifiers from Video. In *CVPR*, 2009.

[8] M. Eichner *et al.* 2D articulated human pose estimation and retrieval in (almost) unconstrained still images. *IJCV*, vol. 99, 2012.

[9] J. Huang *et al.* Image indexing using color correlogram. In *CVPR*, 1997.

[10] N. Dalal and B. Triggs. Histograms of oriented gradients for human detection. In *CVPR*, 2005.

[11] T.S. Lee. Image representation using 2D gabor wavelets. *IEEE TPAMI*, 1996.

[12] C.-C. Chang and C.-J. lin LIBSVM: A library for support vector machines. *ACM TIST*, vol. 2, 2011.

[13] S. Singh *et al.* Unsupervised discovery of mid-level discriminative patches. In *ECCV*, 2012.

[14] C. Doersch *et al.* What makes Paris look like Paris. *ACM TOG (SIGGRAPH)*, vol. 31, 2012.