Identification of Fashion Trends

Gunjan Batra
Rutgers Business School,
Newark, NJ

***Abstract* -** Fashion is a reflection of the society. The objective of the study is to understand the various methods of identifying patterns and trends in fashion using multimedia technology. It is interesting to see how various authors have used different methods to study fashion trends. The method of automatic segmentation of clothing proposed an approach that partitioned individuals from clothing, and also identified individual pieces of clothing. Next, the research on fashion trends in New York presents an algorithm that automatically discovers visual style elements representing fashion trends for a particular season. Taking it to the next level, the third research paper, 'Runway to Realway: Visual analysis of Fashion' used visual representation and human judgment to train a classifier to find similarity in trends on the ramp and the streets. This research has taken the task of image segmentation to a next level giving it a new objective and direction.

1. Introduction

Fashion industry employs 4.2 million people and has global annual revenue of 1.2 trillion dollars. Over 2 million online blogs and stores devoted to fashion, millions of images readily available for computer vision applications. Fashion is heavily driven by trends, similar aesthetics followed by people that appear/reappear cyclically e.g. Florals for spring and oranges for fall. Fashion trends have been analyzed by fashion designers and analysts but this is a novel topic in multimedia science. Fashion is an excellent domain for applying computer vision and moreover, understanding change in fashion is a good way to know society and time. Hence, there is a need for fashion trends spotting.

2. Approaches Proposed

We surveyed three approaches for understand the method to determine the fashion trends.

1. Automatic Segmentation of Clothing for the Identification of Fashion Trends Using K-Means Clustering, 2013
2. What are the fashion trends in New York? , 2014
3. Runway to Realway : Visual Analysis of Fashion, 2015

These three approaches came one after another annually and showed improvement in their methodology with time.

2.1. Automatic Segmentation of Clothing for the Identification of Fashion Trends Using K-Means Clustering, 2013

For Automatic Image Segmentation, the authors proposed an approach that partitioned individuals from clothing, and also identified individual pieces of clothing. They used a combination of following algorithms:

1. Simple Segmentation Utilizing Global K-means Analysis
2. Segmentation through Combinatorial Computer Vision and Machine Learning Algorithms
3. Segmentation through Image-Specific Combinatorial Computer Vision and Machine Learning Algorithms
4. Segmentation using a combination of the algorithms
5. Clustering of Clothing Segments

Data Description – The experiment was performed on 153 images that were selected from a set of 2500 images of two large US clothing retail stores. The selected images had a single individual that displayed the face, torso and, at least, a portion of the leg and contained individuals wearing one to four pieces of clothing. 35 images were of men and the remaining 118 were of women. Finally, image sizes ranged between 6 and 12 kilobytes, with approximate dimensions of 167 x 204.

Global K-means used Matlab Image Processing Toolkit on images for segmentation of images. Following K-means segmentation, the algorithm used face and skin detection, in conjunction with segment pixel ranges to resolve the layer that contained clothing, background and skin. Using this method, 102 of 153 images were segmented properly.

Segmentation through Combinatorial Computer Vision and Machine Learning Algorithm [1] uses knowledge of the input image criteria to optimize the segmentation.

First, Canny edge detection was applied to the image, thus a new image contour was made of component boundaries and person outline. A background mask was created by setting all pixels inside the external edge to white. The detected background was sampled to create a representative RGB histogram. Each pixel in the remaining image was compared to the background histogram. All high occurrence pixels were included in the background mask by back-projection. This step was done to eliminate background that is not discovered in the scanning step. The result was stored in a foreground image.

Figure 1: Results from cluster analysis based on standard deviation of clothing using k=2

Figure 2: Results from cluster analysis based on standard deviation of color (histogram) using k=4

Figure 3: Results from cluster analysis based on standard deviation of color (histogram) and color using k=4

Second, the face was located using Haar classification. The pixels of the facial region were sampled and put into an H-S histogram for the skin, with the V component omitted to reduce the influence of lighting variations. Foreground image pixels are compared to the skin histogram, and pixels with high occurrences are marked as skin.

Next, the background and skin masks were subtracted from the original image, leaving behind the clothes.

Finally, the remaining items in the mask went through a polygon approximation algorithm. This eliminated jagged contours along the clothing.

Later, the images and masks were sent to the clustering algorithm.

Using this method, 75 of 153 images were segmented properly.

Segmentation through Image Specific Combinatorial Computer Vision and Machine Learning Algorithm The algorithm defined the five critical values for a quality image's segmentation: background probability threshold, background histogram bin size, skin probability threshold, skin histogram bin size, background sample size.

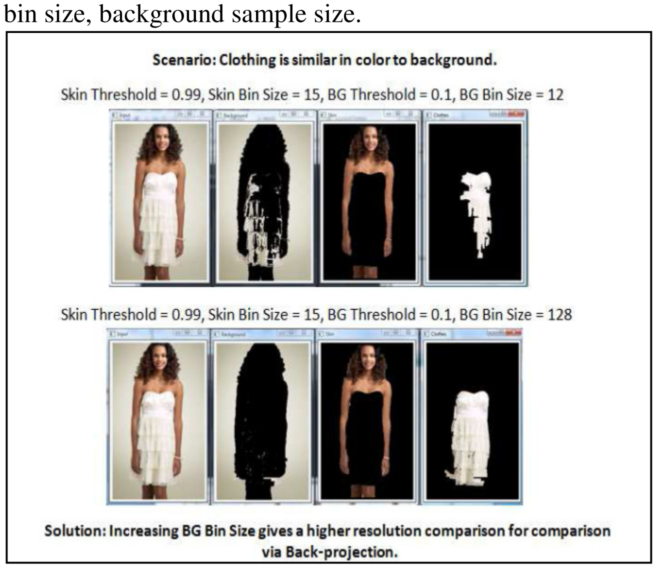

Figure 4

The probability thresholds and histogram bin sizes are applied during back-projection. Any pixel above/below a certain threshold probability defines the group of a pixel. This is mostly useful to distinguish clothes from shin when they are for similar color. The histogram bin sizes determine the resolution of the histogram comparison. A large bin size signifies an accurate grouping whereas a small bin size signifies that the close intensity values are grouped together as one color. The background sample size signifies the background is that is used to construct the histogram during back-projection. The larger the sample area is, the more accurate the background segmentation.

Segmentation using combination of the aforementioned algorithms helped to effectively segment clothing tops from clothing bottoms and hence into two primary components.

Clustering of Clothing Segments

The 7 Hu moments was used as the feature vector to extract the shape approximations. For example, a shirt and pants set have a near-rectangular shape while a gown will have a more triangular shape.

The average standard deviation of the red, green, and blue channels was extracted as a way to represent the complexity of the outfit.

Outfits with patterns and multiple-pieces had higher standard deviations than clothing with solid colors and fewer pieces.

Next, the R,G,B standard deviations were kept as separate entries in the feature vector, so that the algorithm could differentiate between different levels of complexity across the primary colors.

Next, H and S histograms were used as a feature vector to allow differentiation based on colors in the clothing. The H and S histograms each had bin sizes of 256, giving a total of 512 points. Also the similarity clothing color between tops and bottoms was used as a feature vector to understand contrasts. These values were represented as R, G, B.

Now, the algorithm segmented based on the occurrence of color combinations in an outfit, versus just discriminating based on primary color.

2.2. What are the fashion trends in New York? , 2014

The paper presents an algorithm that automatically discovers visual style elements representing fashion trends for a particular season with initial focus on New York Fashion Week (NYFW). The catwalk model is detected from the fashion show videos. Then, with a combination of supervised and unsupervised learning, patterns are identified that are coherent (frequently occurring within the fashion week event) and unique (for the fashion week).After detecting the catwalk model the visual style element initialization is done followed by iterative clustering.

Dataset: Images were extracted from Fashion TV [9] to build a fashion show image dataset of 3276 different catwalk models. These had been presented by nine fashion design companies for the past 10 seasons of NYFW.

Season / Designer	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									

Table 1: Number of Fashion Styles in the dataset

Catwalk Model Detection identified the catwalk model who demonstrated clothing and accessories on the stage at the fashion shows.

Let $v_v^{w,x}$ denote the fashion show video of the designer *x* who has an exhibition at NYFW season *w*. Assuming M frames in $v_v^{w,x}$ that is $v_v^{w,x} = \{\mu_{v1}^{w,x}, \mu_{v2}^{w,x}, \ldots, \mu_{vM}^{w,x}\}$, dataset containing fashion videos was represented as

$$D_v = \{v_v^{1,1},\ v_v^{1,2}, \ldots, v_v^{1,G}, v_v^{2,1}, v_v^{2,2}, v_v^{E,G}\}$$

Where E is the number of NYFW events and G is the number of designers.

Using [10] the presence of faces is tracked over the video frames in $v_v^{w,x}$ of a particular designer who has an exhibition at NYFW and then it is checked if the identified faces belong to the same person. Also a constraint for frontal faces is used as models show their front faces and audiences show side faces. Next, using [11], an algorithm that estimates the spatial layout of humans via pictorial structure model, the full body image with a clean background is obtained from the detected faces. The set containing full body image of catwalk model is represented as $D_i = \{v_i^{1,1}, v_i^{1,2}, \ldots, v_i^{1,G}, v_i^{2,1}, v_i^{2,2}, \ldots v_i^{E,G}\}$

Visual Style Element Initialization

The visual style element of catwalk models was represented by square patches of size 128×128. The collection of full body image of models was divided into 2 sets

1. Positive Set, D+ - Containing catwalk models from the selected fashion week event whose visual style elements of the fashion trends yet to be discovered, e.g. catwalk models at NYFW S/S 2014
2. Negative Set, D- - Containing catwalk models from other NYFW events, e.g. all catwalk models at NYFW except from NYFW S/S 2014).

The full body images of catwalk model from both the sets were partitioned into non overlapping square patches.

The 4 elements of fashion trends are captured using color and texture information.

Color – HSV color histogram, color correlogram [12], color moments

Texture – HOG[13], Haar wavelet transform, 2D Gabor wavelets

Each patch is represented using 3 features: Color, texture and both color and texture

Next, the candidate patches that might represent the visual style elements were determined by finding out nearest neighbors of the patches in entire data set Di (D+ and D-) and filtering out the ones who had more than 10 nearest neighbors in top 20 NN exist in D+. Further, the standard k-means clustering was used to cluster similar style elements and groups of similar style elements were obtained as initial cluster

Iterative Clustering

Iterative clustering was performed on the initial visual elements obtained from the previous step in order to improve the stylistic coherence and uniqueness of patches. This was done in order to remove any irrelevant visual representation in the candidate patches so that it does not reflect in the retrieved matches post clustering.

A linear SVM classifier [10] with C fixed to 0.1 was trained for each cluster, treating patches within the cluster as positive examples and all negative-set patches as negative examples.

However, on iteration the SVM classifier directly by utilizing the top t detectors from previous round as positive set, it just improved the top t detectors a little bit due to the overfitting problem [16].

To overcome overfitting, cross-validation was applied at each round and both positive and negative sets were divided into two equal, non-overlapping subsets. New clusters were formed from the top t firings of each detector by considering all SVM scores above -1 to be firings. This process was repeated for 4 iterations, in which most of top t patches in a cluster converged. The top ten detectors from the list of detectors of top 30 firings in the positive set were chosen as the visual style elements that represented fashion trends at a particular season of New York Fashion Week.

Result

The fashion trends identified were both coherent and unique

Season S/S 2014 - Pastel green was identified as the most representative color trend for the investigated nine designers.

Season A/W 2010 - High bun together with a neck scarf is identified as the most representative fashion trend in terms of texture information.

Figure 5: The most representative fashion trend elements at NYFW

Based on high average accuracy of fashion trend detector for each season, the framework was very effective

NYFW with higher average accuracy has more coherent and unique fashion styles compared to fashion week with lower average accuracy

E.g. the low average accuracy of A/W 2012 implies the fact that the designers had less fashion elements in common compared to the other seasons).

The highest average accuracy among three types of features (color, texture, color+texture) for each fashion week are shown in Table 2.

Eg. 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.

		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 %

Table 2: The average accuracy of the fashion trend element detector

2.3 Runway to Realway: Visual Analysis of Fashion, 2015

The paper presents the first attempt to provide a quantitative analysis of fashion both on the runway and in the real world, computationally, and at large scale, using computer vision.

Data Set:

Two types of data sets are used

a. Runway Dataset – This contained 348,598 runway fashion photos collected from *style.com* that represented 9,328 fashion show collections over 15 years from 2000 to 2014. These were collected together with the metadata which included season, brand name, category, date of event, and a short text description.

b. Paper Doll dataset [43] – This contained 339,797 images collected from the Chictopia social network, based on fashion, where interested people post and annotate pictures of their everyday outfits.

The combination of these two datasets allows us to make initial steps toward analyzing how fashion trends transfer from runway collections to the clothing people wear in real life.

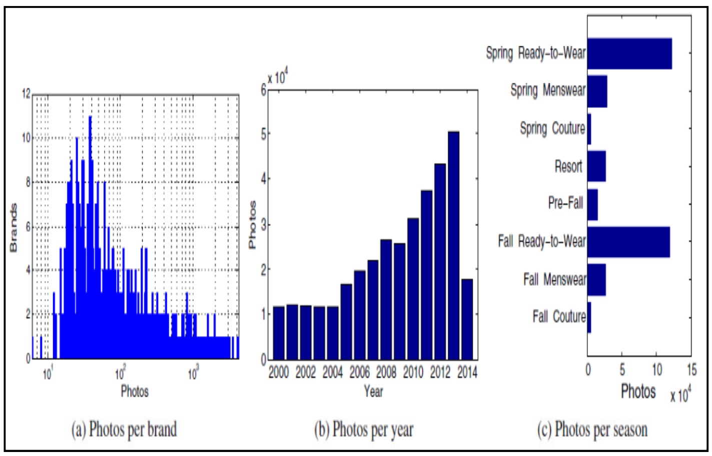

Figure 6: Style dataset statistics

Figure 7: Sample images from Runway Dataset with estimated clothing parses

Methodology:

Visual Representation

The images were first preprocessed by resizing each person detection window to 320 × 160 pixels and extracting 9 sub-regions for the head, chest, torso, left/right arm, hip, left/right/between legs, as shown in Figure 8.

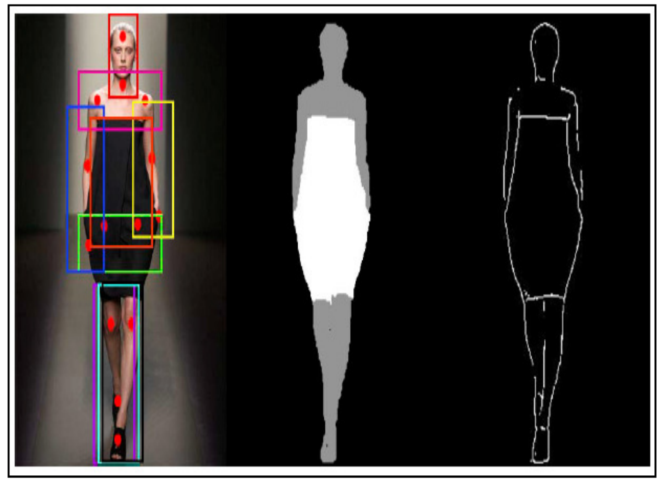

Figure 8: (a) Pose estimation and corresponding bounding boxes (b) Foreground binary mask and (c) Binary edge maps

The features that were measured were: color, texture, shape, and the clothing parse's estimates of clothing items worn and an existing global style descriptor.

Color: From the pixels that are parsed as foreground (clothes, skin, hair), two 512 dimensional histograms in RGB and Lab color spaces are extracted for each sub region.

Texture: From the pixels parsed as foreground two bags of words histograms are extracted for each sub region. The first one is a histogram of MR8 responses [40] quantized into 256 visual words and other is a histogram of HOG descriptors [25] (8x8 blocks, 4 pixel step size, 9 orientations) quantized into 1000 words.

Shape Two features are extracted for each sub-region: a binary mask of the foreground pixels estimated by the clothing parsing algorithm, and an edge map of the foreground region estimated using structured forest efficient edge detection [43]. The threshold, t, was selected to binarize the edge map, in an image-dependent manner.

Parse: The individual itemmasks for each sub region is extracted for all the 56 different clothing categories (e.g. dress, shirt, shoes, etc) and then a 56-dimensional descriptor of the percentage of each item present in the sub-region is formed.

Style descriptor: The style descriptor of [43] is computed over the entire body. It also includes RGB, Lab, MR8, Gradients, HOG, Boundary Distance and Pose Distance. The dimensions are reduces to 441 by using PCA.

These five features are concatenated to form a vector for each subregion used for learning the outfit similarity models later.

Human Judgment

Amazon Mechanical Turk (MTurk) is used to collect human judgment of "similarity" between images of clothing outfits. Labelers were asked to select the most similar outfit from a set

of 5 pictures, or none in case there were no similar outfits. The 5 outfits were selected basis the cosine similarity using each individual feature in isolation (e.g., color, texture), or to an equally weighted combination of all features.
Human judgments were obtained for 2000 random query images from the runway dataset.
The 5 similar candidates were chosen for each query under two scenarios; by selecting the images from the runway dataset, and selecting the images from the Paper Doll, Realway dataset. The similarity annotations were collected for each query image and scenario from 5 labelers.
The majority of labelers agreed for 74.8% of the queries in the within-runway scenario and for 73.9% of the runway-to-realway queries.

Learning to compare outfits
A linear SVM [15] was trained to classify a pair of pictures as similar or dissimilar. Using the three strategies of Majority, Unanimity and Some human judgments were converted to positive/negative labels for training. Majority meant that if an image-pair is marked as positive when the pair gets the majority of labeler clicks. Unanimity meant that any query-image pair for which all-five labelers agree on the best match are treated as positive. Some meant that image pairs marked by any of the five labelers are treated as positive. And any query for which all the labelers clicked "none" is used to form 5 negative pairs with each of its 5 potentially similar images.

a) Qualitative Evaluation Example results are shown in Figure 8 a) The first image is the query image i.e. a runway outfit and the retrieved results (images 2-5) are also sampled from runway collections. In 3 b) the query image is a runway image and the images 2-5 images are the retrieved outfits from the realway Chictopia dataset. In both cases the model usedis the majority training approach. Outfits retrieved both from the runway and from the realway images look quite promising, colors are matched well, and shape and pattern also tend to be similar. We are even able to match challenging visual characteristics quite well such as neckline shape.
(E.g. white strapless dress, bottom row left), and overall garment shape (e.g. red a-line dress, 2nd row right).

b) Quantitative Evaluation results are shown in Table 3. The performance of the learning model is evaluated by the area under the precision recall curve (AUC), using 6-fold cross validation. The proposed framework of learned similarity model achieves 73-76% AUC in Runway to Runway task and 53-55% in Runway to Realway tasks. This is an increase in 10% performance compared to the style descriptor which has AUC of 62-66% in Runway to Runway tasks and 42-45% in Runway to Realway tasks.

	Runway to Runway		Runway to Realway	
Method	our feature	Style descriptor	our feature	Style descriptor
Majority	0.76 ± 0.11	0.66 ± 0.11	0.54 ± 0.03	0.45 ± 0.02
Unanimity	0.73 ± 0.08	0.62 ± 0.07	0.53 ± 0.02	0.42 ± 0.01
Some	0.73 ± 0.14	0.63 ± 0.12	0.55 ± 0.01	0.43 ± 0.03

Table 3: AUC for predicting outfit similarity

3. Comparison

We see that with every passing year more meaningful research was done on the topic. While the first paper attempted to discover fashion trends from two US retail stores by K-means clustering, the second paper determined fashion trends in New York City using fashion show videos taken from Fashion TV by K-Means Clustering and subsequently trained a linear SVM for clustering of similar patches. The third paper used computer vision to learn measures for similarity of outfits from both runway and street images and trained a linear SVM to classify pictures as similar or dissimilar. This paper studied the Color, Texture, Shape, Parse, Style Descriptor of the images compared to only color and texture in the previous papers. It also gave us a vision where similarity in outfits can be used for prediction tasks. A detailed comparison with different aspects has been provided in Table 10.

Figure 8: a) Runway to Runway b) Runway to Realway

	Automatic Segmentation of Clothing for Identification of Fashion Trends	What are the Fashion Trends in New York?	Runway to Realway : Visual Analysis of Fashion
Year of Publication	2013	2014	2015
Outcome	Clustering based on color complexity and outfit shape	Algorithm for fashion trend detection	Approach to studying fashion on runway and real world setting
Classification/ Clustering Used	K-Means Clustering	K-Means Clustering Simple Vector Machine	Simple Vector Machine
Data Set	153 images from 2 large US clothing Retail store	3276 images extracted from Fashion TV	348,598 Runway fashion photos
Feature Measures Used	Standard deviation of clothing, Standard deviation of color(histogram), Std dev of color(histogram) and color	Color, Texture, Color+Texture	Color, Texture, Shape, Parse, Style Descriptor

Table 10

4. Conclusion/Future Scope

The study in the papers successfully presented new and different ways to judge similarity between outfits and identify the trends in fashion. This trends analysis in fashion models can be used to develop clothing recommender systems to pair clothing, give style suggestions for different types of events, product suggestions on e-commerce websites. This method of study can be linked with the online social media on real time and can be used to study fashion trends among the users worldwide.

These models can also be used to mine large datasets for similar styles/ trends and predict fashion trends based on past data and improving market analysis of fashion experts.

5. References

[1] Hu, Z., Yan, H., and Lin, X. Clothing segmentation using foreground and background estimation based on the constrained Delaunay triangulation. Pattern Recogn. 41, 5 May. 2008, Pages 1581-1592.

[2] Borràs Agnès, Tous Francesc, Lladós Josep, Vanrell Maria, "High-Level Clothes Description Based on Colour - Texture and Structural Features", 1st. Iberian Conference on Pattern Recognition and Image Analysis IbPRIA 2003.

[3] Mena, J.B., Malpica, J.A., 2003. Color image segmentation using the Dempster sharer theory of evidence for the fusion of texture. Pattern Recognition. Internat. Arch. Photogrammet. Remote Sensing 34, Part 3/W8.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

[18] T. Berg and P. N. Belhumeur. Poof: Part-based one-vs.-one features for fine-grained categorization, face verification, and attribute estimation. In CVPR, pages 955–962, 2013.

[19] T. L. Berg, A. C. Berg, and J. Shih. Automatic attribute discovery and characterization from noisy web data. In ECCV, pages 663–676. Springer-Verlag, 2010.

[20] L. Bossard, M. Dantone, C. Leistner, C. Wengert, T. Quack, and L. Van Gool. Apparel classification with style. ACCV,pages 1–14, 2012.

[21] L. Bourdev, S. Maji, and J. Malik. Describing people: A poselet-based approach to attribute classification. In ICCV, pages 1543–1550, 2011.

[22] C.-C. Chang and C.-J. Lin. LIBSVM: A library for support vector machines. ACM Trans. Intell. Syst. Technol., 2:27:1–27:27, 2011.

[23] H. Chen, A. Gallagher, and B. Girod. Describing clothing by semantic attributes. In ECCV, pages 609–623. Springer-Verlag, 2012.

[24] C. T. Corcoran. The blogs that took over the tents. In Women's Wear Daily, Feb. 2006.

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

[26] P. Dollar and C. L. Zitnick. Structured forests for fast edge detection. In ICCV, 2013.

[27] R. Farrell, O. Oza, N. Zhang, V. I. Morariu, T. Darrell, and L. S. Davis. Birdlets: Subordinate categorization using volumetric primitives and pose-normalized appearance. In ICCV, 2011.

[28] E. Gavves, B. Fernando, C. Snoek, A. Smeulders, and T. Tuytelaars. Fine-grained categorization by alignments. ICCV, pages 1–8, 2013.

[29] Y. Kalantidis, L. Kennedy, and L.-J. Li. Getting the look: clothing recognition and segmentation for automatic product suggestions in everyday photos. In ICMR, pages 105–112. ACM, 2013.

[30] G. Kim and E. P. Xing. Discovering pictorial brand associations from large-scale online image data. ICCV Workshops, 2013.

[31] A. Kovashka, D. Parikh, and K. Grauman. Whittlesearch: Image search with relative attribute feedback. In CVPR, pages 2973–2980. IEEE, 2012.

[32] I. S. Kwak, A. C. Murillo, P. N. Belhumeur, D. Kriegman, and S. Belongie. From bikers to surfers: Visual recognition of urban tribes. In BMVC, 2013.

[33] Y. J. Lee, A. A. Efros, and M. Hebert. Style-aware mid-level representation for discovering visual connections in space and time. In ICCV, 2013.

[34] 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.

[35] S. Liu, Z. Song, G. Liu, C. Xu, H. Lu, and S. Yan. Street-toshop: Cross-scenario clothing retrieval via parts alignment and auxiliary set. In CVPR, pages 3330–3337, 2012.

[36] F. Palermo, J. Hays, and A. A. Efros. Dating historical color images. In ECCV, pages 499–512. Springer, 2012.

[37] D. Parikh and K. Grauman. Relative attributes. In ICCV, pages 503–510. IEEE, 2011.

[38] M. Shao, L. Li, and Y. Fu. What do you do? Occupation recognition in a photo via social context. ICCV, 2013.

[39] Z. Song, M. Wang, X.-s. Hua, and S. Yan. Predicting occupation via human clothing and contexts. In ICCV, pages 1084–1091, 2011.

[40] M. Varma and A. Zisserman. A statistical approach to texture classification from single images. Int. J. Comput. Vision, 62(1-2):61–81, Apr. 2005.

[41] N.Wang and H. Ai. Who blocks who: Simultaneous clothing segmentation for grouping images. In ICCV, pages 1535–1542, 2011.

[42] K. Yamaguchi,M. H. Kiapour, and T. L. Berg. Parsing clothing in fashion photographs. In CVPR, pages 3570–3577, 2012.

[43] K. Yamaguchi, M. H. Kiapour, and T. L. Berg. Paper doll parsing: Retrieving similar styles to parse clothing items. In ICCV, 2013.

[44] Y. Yang and D. Ramanan. Articulated pose estimation with flexible mixtures-of-parts. In CVPR, pages 1385–1392, 2011.

[45] B. Yao, G. Bradski, and L. Fei-Fei. A codebook-free and annotation-free approach for fine-grained image categorization. In CVPR, pages 3466–3473. IEEE, 2012.