Street-to-Shop: Cross-Scenario Clothing Retrieval via Parts Alignment and Auxiliary Set

Si Liu*+, Zheng Song*, Guangcan Liu*, Changsheng Xu+, Hanqing Lu+, Shuicheng Yan*

* ECE Department, National University of Singapore

+ NLPR, Institute of Automation, Chinese Academy of Science

{dcslius, zheng.s, eleliug, eleyans}@nus.edu.sg,{csxu, luhq}@nlpr.ia.ac.cn

Abstract

In this paper, we address a practical problem of cross-scenario clothing retrieval - given a daily human photo captured in general environment, e.g., on street, finding similar clothing in online shops, where the photos are captured more professionally and with clean background. There are large discrepancies between daily photo scenario and online shopping scenario.

We first propose to alleviate the human pose discrepancy by locating 30 *human parts detected by a well trained human detector. Then, founded on part features, we propose a two-step calculation to obtain more reliable one-to-many similarities between the query daily photo and online shopping photos: 1) the within-scenario one-to-many similarities between a query daily photo and the auxiliary set are derived by direct sparse reconstruction; and 2) by a cross-scenario many-to-many similarity transfer matrix inferred offline from an extra auxiliary set and the online shopping set, the reliable cross-scenario one-to-many similarities between the query daily photo and all online shopping photos are obtained.*

We collect a large online shopping dataset and a daily photo dataset, both of which are thoroughly labeled with 15 *clothing attributes via Mechanic Turk. The extensive experimental evaluations on the collected datasets well demonstrate the effectiveness of the proposed framework for cross-scenario clothing retrieval.*

1. Introduction

Nowadays, online clothing shopping is becoming an increasingly popular shopping model. In many online shopping websites such as Amazon.com, eBay.com, and shopstyle.com, customers can conveniently find their favorite clothing by typing some keywords, such as "black, sleeveless cocktail dress".

In this paper, we consider a more interesting and practical shopping model: given a human photo captured on street or saw occasionally, finding similar clothing from online shops. This shopping model can be integrated into various platforms, e.g. in mobile applications as shown in Figure 1. One can take a photo of any fashionably dressed lady with mobile device, and our system can parse the photo and search for the clothing with similar styles from online shopping websites. In social network platform, one may occasionally browse a photo in a friend's album and is interested in the clothing. With one click over the photo, similar clothing from online shops can be returned.

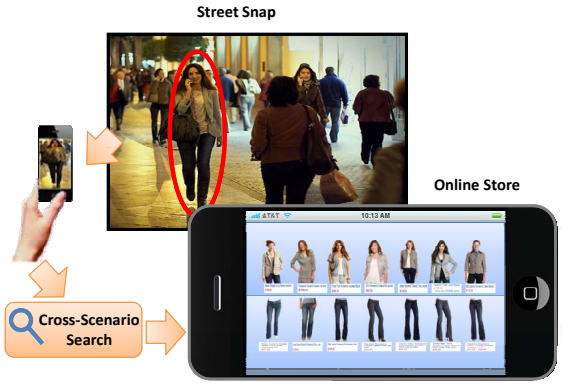

Figure 1. The "street-to-shop" clothing retrieval model: user takes a photo of any person, then similar clothing from online shops are retrieved using the proposed cross-scenario image retrieval solution to facilitate online clothing shopping. For better viewing of all images in this paper, please see original color pdf file.

The proposed shopping model rises a challenging research problem which is not well studied yet: given a clothes/human image captured in general environment (i.e. daily photo clothings), finding similar clothes/human images from a dataset taken in special environment (i.e. embellished photos used in online clothing shops).

To investigate this problem, we collect an Online Shopping dataset (OS) and a Daily Photo dataset (DP). Obviously, there exist large discrepancies between these two scenarios as shown in Figure 2. Consequently large variance will exist in direct feature description of the clothes. Human pose variation is the first aspect causing the data variance.

978-1-4673-1228-8/12/$31.00 ©2012 IEEE

Figure 2. Comparison of (a) the collected online shopping dataset and (b) the daily photo dataset. It is illustrated that (b) contains more appearance variance (left part) and occlusions (right part) than (a).

More specifically, the clothing models in OS are generally in a similar professional pose. On the contrary, people in DP can take more flexible poses. Secondly, the background of OS is generally clean, while the background is more cluttered and diverse in DP since they are taken in different environments. The diverse background can be buildings, stairs or even other people. The large discrepancy will result in the inaccurate image similarity estimation in the street-to-shop image retrieval model.

To handle the aforementioned large discrepancies between these two scenarios, we propose two subsequent strategies for effective cross-scenario clothing retrieval. First, to handle the human pose discrepancy between professional model and ordinary people, human parsing technique [21] is used to align human parts. Extracting features from aligned human parts, such as neck, shoulders, hands, can reduce the feature discrepancy caused by human pose variation.

Although describing clothes image using aligned parts, there still exist background discrepancies inside the detection rectangles between the two scenarios. The background may dominate similarity calculation between two parts and bias the real clothing similarity. We observe that regularity exists in the background area of detected rectangles, and the structure of the discrepancy can be statistically modeled using an auxiliary daily photo set and online shopping dataset. An offline learning process is proposed for such a purpose.

The key assumption is that the discrepancy observed between the auxiliary set and OS dataset have common structures, i.e., the discrepancies often occur in similar spatial positions. To mine this structure, all the auxiliary photos are collaboratively and sparsely reconstructed by the OS dataset, where the reconstruction error corresponds to the cross-scenario discrepancy. With group sparsity regularization upon the reconstruction errors, discrepancy regularity is encouraged. Then the offline sparse reconstruction coefficients are used to construct the cross-scenario similarity transfer matrix.

For online retrieval, given a captured daily photo, the similarities between the query daily photo and all the auxiliary photos are first derived by direct sparse representation. By integration of the online within-scenario similarities with offline calculated cross-scenario similarity transfer matrix, more reliable similarities between a single query photo and all online shopping photos are obtained, which can be used for final cross-scenario image retrieval. The whole framework is unsupervised, and thus quite practical.

The main contributions of this paper can be summarized as follows:

1. To our best knowledge, this work is the first attempt to tackle the task of cross-scenario online shopping clothing retrieval queried by a daily photo.
2. We collect a large online shopping dataset and daily photo dataset for investigating the cross-scenario clothing retrieval task. Additionally, these datasets are thoroughly labeled with complete attributes via Mechanical Turk.
3. We propose a two-step calculation to handle the cross-scenario discrepancies. Human parts are first aligned using state-of-the-art human part detection method. Then, an auxiliary set is utilized to discover cross-scenario discrepancy by the proposed collaborative sparse reconstruction with group sparsity regularized reconstruction errors. Experimental evaluations show that the retrieval performance is boosted by these two strategies.

2. Related Work

Clothing Study : There is a large body of research literature on clothing segmentations, modeling and recognition. Hasan et al. [12] and Wang et al. [17] proposed to use different priors to segment clothing. One representative work for clothing modeling is from Chen et al. [7], which used an And-Or graph representation to produce a large set of composite graphical templates accounting for the wide variabilities of cloth configurations. Yu et al. [20] proposed to integrate face detection, tracking and clothing segmentation to recognize clothing in surveillance video.

Clothing retrieval problem has not been extensively studied yet. Two related works are from Wang et al. [18] and Chao et al. [6]. However, these two approaches are designed only within one scenario.

Parts Appearance based Human Attribute Analysis: Attribute has received much attention in these years from the early pioneers [9, 13] to the most recent achievement [2]. In the field of fashion design, Berg et al. [1] proposed to automatically discover attributes from noisy web data [1]. The key difference between our work and theirs is that we focus on the clothing-specific attribute learning, which is more fine-grained.

Previous research on human attributes analysis tend to first align human parts [5, 16] due to the large pose variation and background noise in daily photos. The part-based detection [21, 4] is proven to be able to assist in matching human parts and thus facilitate the appearance modeling for attribute classification. In this paper, we follow this line and design a part-based detection scheme for the clothing retrieval task.

Unsupervised Transfer Learning : Cross-domain learning problem has been encountered in many applications [15]. Many previous research work have tried to reduce the cross-domain mismatch between the data distributions. In our work, however, we observe regular structures on the discrepancy itself and hence design a direct learning scheme to model the discrepancy structure; also we may inevitably encounter corrupted training data, which to our best knowledge cannot be explicitly handled in traditional transfer learning scheme.

3. Dataset Construction

Table 1. Number of labels for each upper-body attribute.

	color	plain	material	collar	sleeve	pattern
OS	7226	7528	4811	6809	8159	2142
DP	3485	3783	2081	2989	4180	922
	front	button	zip	belt	all	
OS	8258	6844	7483	7044	8293	
DP	4265	3303	3503	3548	4321	

Table 2. Number of labels for each lower-body attribute.

	color	plain	material	pattern	length
OS	7286	7824	5540	1599	7852
DP	3153	3585	1910	467	3570
	drape	pants shape	skirt shape	curling	all
OS	8050	6536	7949	3877	8343
DP	3488	3259	3414	2796	4068

There are several existing clothing datasets but none of them is suitable to evaluate the cross-scenario clothing retrieval task. The resolution of the dataset collected by Yang et al. [20] is 200 $\times$ 300 pixels, and thus is not enough for detailed clothing attribute prediction. Another dataset constructed by Bourdev et al. [5] only contains 5 attributes related with clothing. So we construct a new database for cross-scenario clothing retrieval.

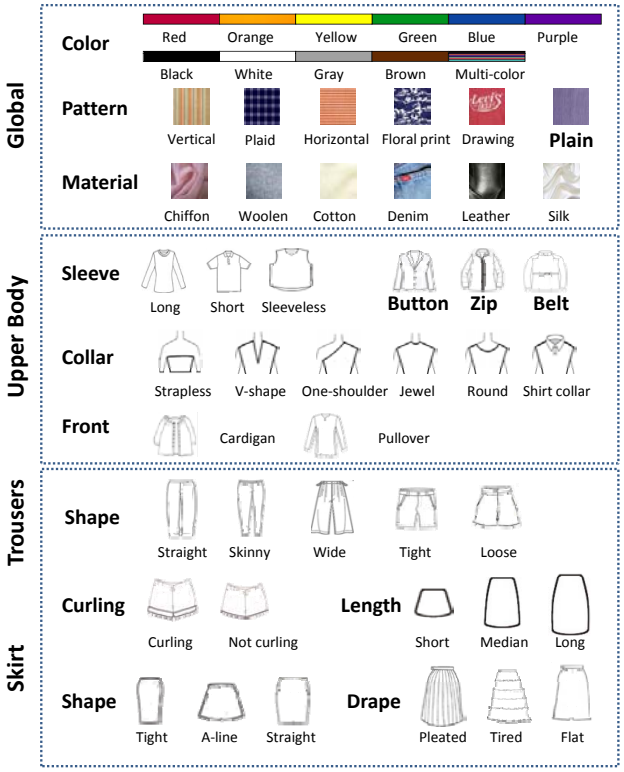

Figure 3. The annotated attributes. An example or line drawing is shown to explain each attribute value.

Clothing Image Collection : We collect two datasets in this work: Online Shopping (OS) dataset and Daily Photo (DP) dataset. The OS dataset is collected by crawling images from several online shopping websites, such as Amazon.com, using keywords about clothing categories such as "T-shirts", "suits". We also collect the DP dataset from Flickr.com using queries such as "street shot", "shopping girls", etc. A well-trained human part detector [21] is applied on all these images and only the high-confidence detection outputs are kept. Then each detected human/clothing is cropped as one image sample. Figure 2 shows several examples from the two datasets.

Clothing Attribute Labeling : We propose to measure the groundtruth similarity between two pieces of clothing by the number common attributes. As shown in Figure 3, we manually define a set of clothing-specific attributes according to empirical study on clothing catalog. The defined attributes can be summarized into three classes, i.e., global, upper-body and lower-body attributes, while lower-body attributes can be further divided into trousers and skirts related attributes. To the best of our knowledge, this dataset has the most complete clothing attribute annotations among all current benchmark datasets.

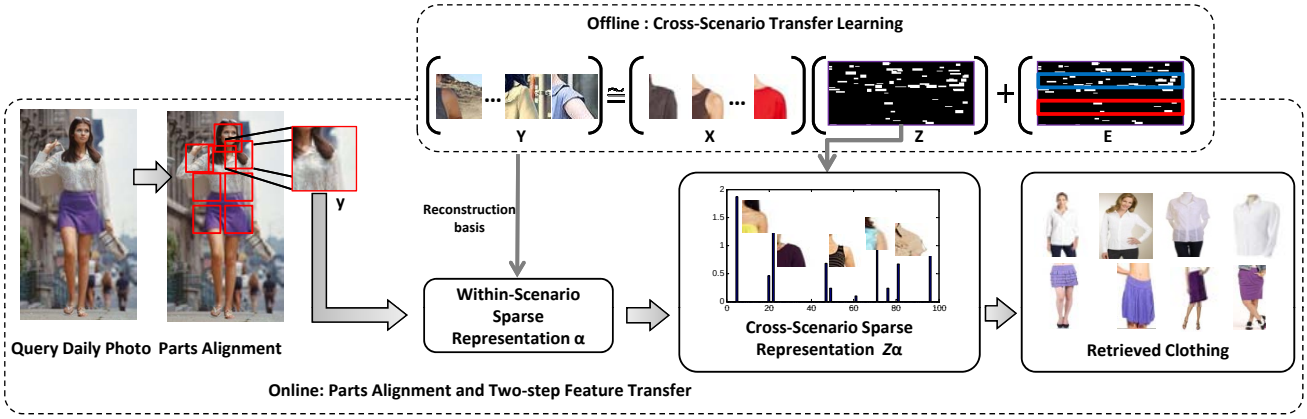

Figure 4. The whole framework of the street-to-shop clothing retrieval system.

Table 3. Number of samples in each subset: "_U" denotes upper-body numbers and "_L" denotes lower-body numbers.

	training_U	query_U	training_L	query_L
OS	6293	2000	6343	2000
	auxiliary_U	query_U	auxiliary_L	query_L
DP	3321	1000	3068	1000

We obtain manually labeled clothing attributes of these two datasets by crowd sourcing via Amazon Mechanical Turk website. Regarding to the difficulty in distinguishing the attribute categories, different number of annotators are assigned to different labeling task. A label was considered as ground truth if at least more than half of the annotators agreed on the value of the label. Totally, we collect 8293 upper body images and 8343 lower body images in OS. And the numbers in DP are 4321 and 4068, respectively. Note that if a photo contains both upper body and lower body clothing, for example, a cocktail dress, the two parts are cropped and included into the upper body and lower body dataset respectively. Table 1 and Table 2 show the distribution of each attribute. We randomly split OS into a training subset and a test query subset, and randomly split DP into an auxiliary subset and a test query subset for further use. Table 3 shows the number of samples in each subset.

4. Framework

Our framework is shown in Figure 4. First, given a daily photo, 20 upper-body parts and 10 lower-body parts are located. Then for each human part in the daily photo, e.g. the left shoulder part in Figure 4, its features y is linearly reconstructed by Y, the reference feature samples of this part extracted from the auxiliary daily photo set, and a sparse coefficient vector α can be obtained.

The auxiliary samples Y are also offline collaboratively reconstructed by the samples from corresponding OS dataset X and the sparse reconstruction coefficient matrix Z is obtained by multi-task sparse representation with the constraint that the reconstruction errors corresponding to the same spatial position (several rows of the error term E) are activated or turned off simultaneously. Finally, feature representation of the daily photo y is refined by the integration of the online calculated within-scenario sparse coefficient α and the offline cross-scenario collaborative reconstruction coefficient Z by $y' = XZ\alpha$. Consequently, a nearest neighbor search in OS dataset can be implemented based on the reconstructed new feature representation y'.

The online processing part of our framework brings two extra procedures than traditional image retrieval framework, i.e, the human part alignment and the within-scenario sparse coding. The two procedures take around 0.07 and 0.1 second CPU time per part on an Intel 2.83 GHz CPU in average. Note that the processing for multiple parts can be paralleled. Hence the extra processing can be further accelerated.

5. Towards Cross-Scenario by Human Parts Alignment

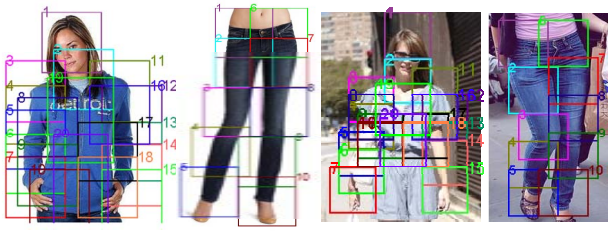

Figure 5. The detected upper and lower body parts from the OS dataset (a,b) and the DP dataset (c,d) are illustrated.

We use the annotated key points in human photos [4] and train one human upper body and one human lower body detector [21]. Figure 5 shows several human detection results, which demonstrates the necessity of human parts alignment. Taking the lower-body parts as an example, in Figure 5(b), the lady's legs occupy the whole image, while the leg parts only cover the lower part in Figure 5(d). Moreover, the

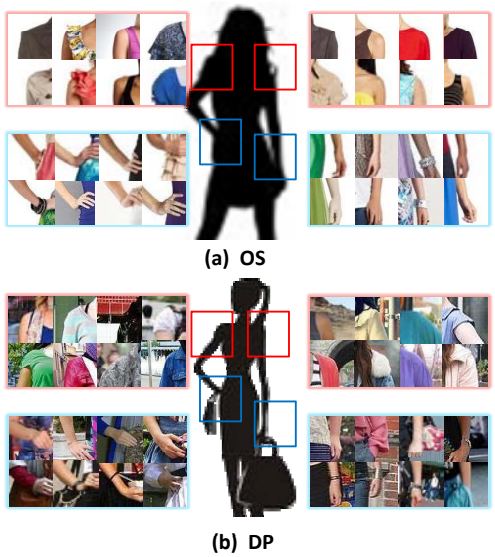

Figure 6. Exemplar of aligned parts of two datasets. The models in OS dataset (a) are professionally posed while people are in casual pose and may take many personal belongings in DP dataset (b).

poses of legs are changeable: the lady in Figure 5(d) is walking so the leg parts are positioned differently from the static model in Figure 5(b). Thus only after human parts alignment, each human part can share similar feature representation which is critical for appearance modeling.

Following [16, 5], we extract 5 kinds of features from the 20 upper-body parts and 10 lower-body parts. The features include HOG, LBP, Color moment, Color histogram and skin descriptor. More specifically, each human part is first partitioned into several smaller, spatially evenly distributed regular blocks. 5 features are extracted from each block and features from all blocks are finally concatenated to represent a human part. The block based features can roughly preserve relative position information inside each human part, which facilitates the following feature refinement process.

6. Towards Cross-Scenario by Auxiliary Set

The human part detection step can handle the human pose discrepancy and partially filter out background clutter. Figure 6 gives some exemplar images for the shoulder and hand parts. From this figure, it can be seen that even the parts are well-aligned, large discrepancies still exist, e.g. for the left shoulder part, there are various background clutter outside the human contour.

The background clutter can bias the true clothing similarity and affect the similarity based image retrieval between online shopping clothing and daily photo clothing. Therefore, further feature refinement towards more reliable similarly measure is necessary. Our aim is to find a new representation of daily photo, so that it can be directly compared with clothing in the online shopping dataset.

However, it is difficult to robustly estimate the cross-scenario discrepancies only based on one query daily photo and an online shopping dataset. Therefore, we collect auxiliary daily photo clothing set. Intuitively, the discrepancies inferred by simultaneously considering the entire auxiliary set and online shopping set, are much more reliable. Noted that this procedure does not require any annotations for the auxiliary daily photo set, and thus this set can be collected beforehand.

To mine the cross-scenario discrepancies, we propose to collaboratively reconstruct the auxiliary set by online shopping clothing. The reconstruction is performed at a set-to-set manner and thus is more robust and reliable.

Essentially, only a small subset of the over complete OS set are similar with each image in the auxiliary set, therefore, the reconstruction coefficient is constrained to be sparse with an ℓ_1 norm constraint, which has been widely applied in many computer vision tasks [19]. From Figure 6, we can observe the cross-scenario discrepancies often lie in similar positions after the parts alignment. Based on this observation, all features (5 kinds of features) extracted from the same region are considered as a group and enforced to have similar sparsity properties by a group sparsity regularization.

Formally, denote $X = [x_1, x_2, ..., x_n] \in R^{d\times n}$ as image features from the OS dataset and $Y = [y_1, y_2, ..., y_m] \in R^{d\times m}$ as image features from the auxiliary DP dataset, where each column is an image feature. We propose to learn a sparse matrix $Z \in \mathbb{R}^{n\times m}$ to reconstruct Y using X with reconstruction error $E \in \mathbb{R}^{d\times m}$ constrained by group sparsity. The objective function is formulated as:

$$\min_{Z,E} \frac{1}{2}\|Y - XZ - E\|_F^2 + \lambda_1 \sum_{g=1}^{G} \|E_g\|_F + \lambda_2 \|Z\|_1, \qquad (1)$$

Algorithm 1 Solving Problem (2) by Inexact ALM

Input: matrices X and Y; parameters λ_1 and λ_2.
Initialize: $Z = J = 0, E = 0, W = 0, \mu = 10^{-6}, \mu_{max} = 10^6, \rho = 1.1$, and $\varepsilon = 10^{-8}$.
while not converged **do**
1. Fix the others and update E by

$$E = \arg\min_E \lambda_1 \sum_{g=1}^{G} \|E_g\|_F + \frac{1}{2}||E - (Y - XJ)||_F^2.$$

2. Fix the others and update Z by

$$Z = \arg\min_Z \frac{\lambda_2}{\mu}||Z||_1 + \frac{1}{2}||Z - (J - W/\mu)||_F^2.$$

3. Fix the others and update J by

$$J = (X^T X + \mu I)^{-1}(X^T(Y - E) + \mu Z + W).$$

4. Update the multiplier W by

$$W = W + \mu(Z - J)$$

5. Update the parameter μ by $\mu = \min(\rho\mu, \mu_{max})$.
6. Check the convergence condition $||Z - J||_\infty < \varepsilon$.
end while

where $\|\cdot\|_F$ is the Frobenius norm, $E = [E_1; E_2; \cdots; E_G]$, G is the number of blocks (groups) in the feature vector and $\|\cdot\|_1$ is the ℓ_1-norm, $\lambda_1 > 0, \lambda_2 > 0$ are two trade-off parameters to control the strength of the regularization.

The above optimization problem is convex and can be solved by various methods. For efficiency, we adopt in this paper the Augmented Lagrange Multiplier (ALM) [14] method. We first convert (1) into the following equivalent problem:

$$\min_{Z,J,E} \quad \|Y - XJ - E\|_F^2 + \lambda_1 \sum_{g=1}^{G} \|E_g\|_F + \lambda_2 \|Z\|_1, \quad \text{s.t. } Z = J, \tag{2}$$

and then minimize the following augmented Lagrange function:

$$\mathcal{L} = \|Y - XJ - E\|_F^2 + \lambda_1 \sum_{g=1}^{G} \|E_g\|_F + \lambda_2 \|Z\|_1 + Tr(W^T(Z - J)) + \frac{\mu}{2}\|Z - J\|_F^2,$$

where $Tr(\cdot)$ is the trace of a matrix, and $\mu > 0$ is a penalty parameter.

The above problem can be alternatively optimized with respect to J, Z and E, respectively, and then update the Lagrange multiplier W. The inexact ALM method, is used for efficiency and outlined in Algorithm 1. Step 1 is solved via Lemma 3.2 of [11] and 2 are solved via the shrinkage operator [14]. Step 3 of the algorithm has closed-form solution. The convergency of the ALM has been generally discussed in [14, 3].

7. Experiments

7.1. Experimental Setting

Evaluation Criterion : We follow the evaluation criterion of [8] using a ranking based criteria for evaluation. Given a query image q, all the n images in a dataset can be assigned a rank by the retrieval procedure. Let $Rel\,(i)$ be the groundtruth relevance between q and the i^{th} ranked image. We can evaluate a ranking of top k retrieved datum with respect to a query q by a precision,

$$Precision@k = \frac{\sum_i^k Rel\,(i)}{N},$$

where N is a normalization constant to ensure that the correct ranking results in an precision score of 1.

The precision calculation is extended to multiple attributes according to [10]. Specifically, if we consider one particular attribute of the query image, then the value of $Rel\,(i)$ is binary. And if we evaluate on multiple attributes of a query image, then $Rel\,(i)$ will have multiple levels of relevance values. For example, if the query image q is labeled as "short-sleeve, round-neckline and blue" and one retrieved image is annotated as "long-sleeve, shirt-neckline and blue", the relevance value is 2.

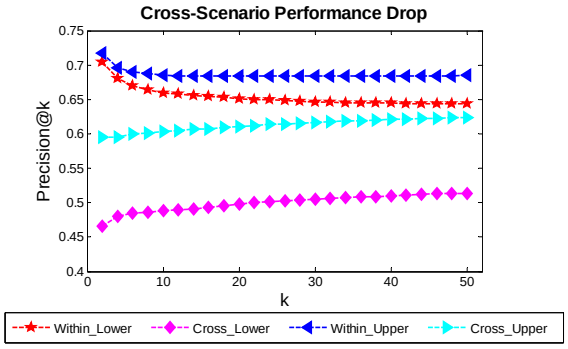

Figure 7. The performances of within-scenario and cross-scenario settings. The experiments are performed on upper body and lower body sets separately.

7.2. Within-Scenario Vs. Cross-Scenario

We simulate the within-scenario and cross-scenario occasion respectively by using the OS query set and the DP query set to retrieve images in the OS training set and the retrieval performance is shown in Figure 7. It can be seen that in upper-body and lower-body cases, the performances drop from within-scenario to cross-scenario setting at about 20%. It shows that large discrepancies exist between the online shopping and daily photo scenarios.

Note that the evaluation method of [8] will results in different trends of the $Precision@k$ vs k curve regards different task difficulty. Hence the performance comparison normally forms $\succ$ shaped curves, i.e. better-performed tasks decrease and worse-performed tasks increase when k increases from 0. However, both kinds of curve increase to near 1.0 when k is large enough.

7.3. Performances of Different Features and Parts

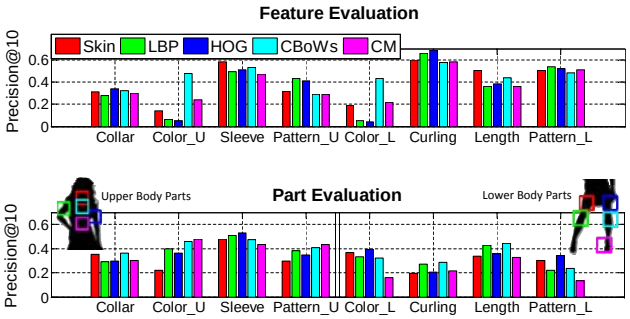

Figure 8. The performances of representative attributes with respect to different features and parts.

In this experiment, we evaluate how the concerned features and body parts affect the retrieval performances of the defined clothing attributes. We evaluate the top 10 performances of each attribute with respect different features and body parts. The experiment is performed using the cross-scenario configuration.

Generally different features and body parts contribute to different clothing attributes. Several representative results

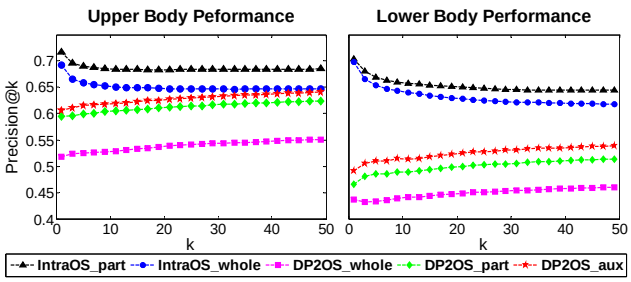

Figure 9. The cross-scenario retrieval performance comparison. We compare the within-scenario ("IntraOS") and cross-scenario ("DP2OS") configurations. The suffix "_whole" denotes using global features without parts alignment, "_part" denotes using features with parts alignment and without auxiliary set and "_aux" denotes using both parts alignment and auxiliary set.

are shown in Figure 8. As illustrated, HOG feature and body parts near the neck are most effective to retrieve similar collar shape. CBoWs and CM features and the central body parts are most useful for color identification. Precise sleeve type prediction highly relies on skin feature from shoulder parts. Finally, HOG and LBP features can produce better results for clothing pattern recognition. Considering all clothing attributes, all of these features from all body parts are respectively normalized and concatenated together to boost the overall performance of predicting all clothing attributes.

7.4. Parts Alignment for Cross-Scenario Retrieval

To validate the effectiveness of parts alignment for the cross-scenario retrieval task, we compare our method with a baseline using global features. To implement the global feature baseline, we extract the same features from the clothing area with 4×4, 2×2 and 1×1 image pyramids. The performances shown in Figure 9 illustrate that features with parts alignment are more discriminative than the global image feature about 8% percent for the upper body set and about 5% for the lower body set. By observing more in detail, we find the improvement is more significant in upper body case because: 1) upper body images contain finer part structures which makes the parts alignment more important, and 2) upper body images have more pose variation and thus parts alignment is more effective.

7.5. Auxiliary Set for Cross-Scenario Retrieval

We implement the Algorithm 1 with $\lambda_1 = 0.1$ and $\lambda_2 = 0.01$ to learn the transfer matrix. In our experiment, 50 to 80 iterations are required for convergence for the upper body and lower body set respectively.

The results after feature refinement with auxiliary set is shown in Figure 9. It can be observed that the auxiliary set can further improve clothing retrieval. The ultimate performance at top rank 50 on upper body set is already near the within-scenario performance, which proves the feature refinement is effective.

Figure 10. Some typical imperfect retrieval results and the possible reasons are illustrated.

7.6. Exemplar Retrieval Results

Figure 12 provides several exemplar retrieval results. Overall, even though people in images take handbags, plastic bags, or at a profile pose, most of the found clothing shares many similar attributes with queries. The color attribute is most reliably retrieved, which is very beneficial since color is one of the most salient attributes when people search for favorite clothing. Other attributes, such as sleeve, trousers length can also achieve promising results. Some typical confused retrieval results are shown in Figure 10.

7.7. Extension: Interactive Clothing Retrieval

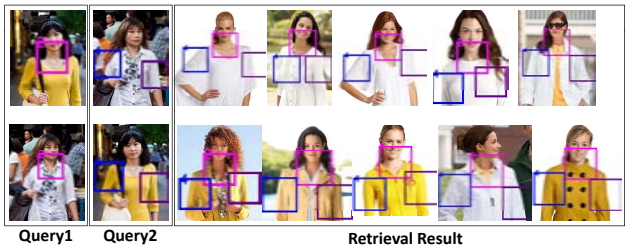

Figure 11. The first two columns show user-specified parts (neck part for query 1 and sleeve parts for query 2). From the third to seventh columns illustrate the retrieval results.

In certain cases, user may have several favourite clothing parts to emphasize in the retrieval task. Since our cross-scenario transfer learning is performed in the human part level, our proposed system can be easily adapted to user specified queries by retrieving with only the features from the selected parts. Figure 11 demonstrates two examples: if one expects to search clothing with similar collar area with query image 1 and similar sleeve area with query image 2, we only need calculate the similarity based on features from corresponding collar part and sleeve parts.

8. Conclusions and Future Work

We are the first to address an important but challenging problem: find similar clothing photos across different photo capturing scenario. We propose a solution including two key components, i.e., human/clothing parts alignment to handle human pose variation and bridging cross-scenario discrepancies with an auxiliary daily photo dataset. Promising results are achieved on our collected Online Shopping and Daily Photo datasets.

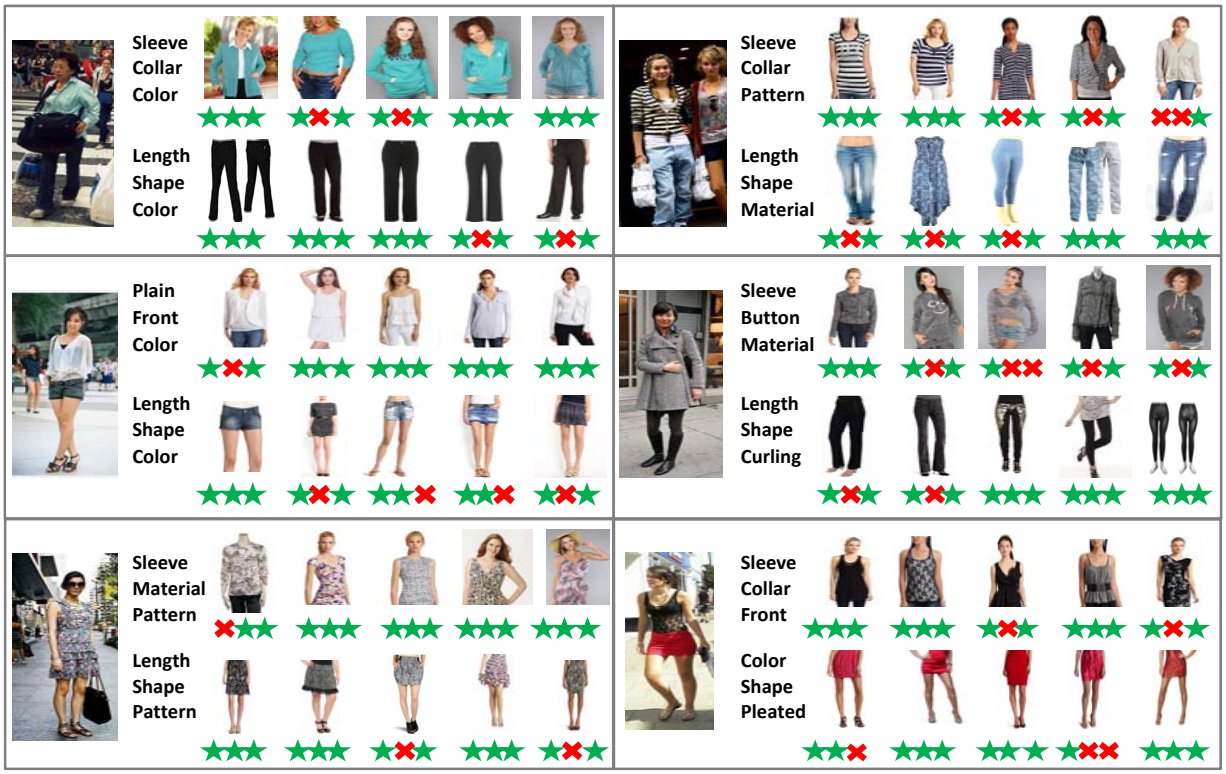

Figure 12. Six samples of flickr daily photos are used as queries, and the retrieved upper-body and lower-body online shopping clothing photos are displayed in separate rows. Correctly and wrongly retrieved clothing attributes are marked by green stars and red crosses respectively.

In the future, we plan to add more supervision information to assist clothing search, e.g. our system can be extended as an online learning system. More specifically, since user clicks indicate the recommended online shopping photos and query images are similar, the image pairs can be recorded and used to update the system. Efficient large-scale computing is also our future focus.

Acknowledgement

This work was supported by 973 Program No. 2010CB327905, the National Natural Science Foundation under Grant No. 60833006 & 60905008 of China and NExT Research Center funded under the research grant WBS. R-252-300-001-490 by MDA, Singapore.

References

[1] T. Berg, A. Berg, and J. Shih. Automatic attribute discovery and characterization from noisy web data. In *ECCV*, 2010. 3

[2] M. Berkowitz and G. Haines Jr. Relative attributes. In *ICCV*, 2011. 3

[3] D. Bertsekas. Constrained optimization and lagrange multiplier methods. *Computer Science and Applied Mathematics*, 1982. 6

[4] L. Bourdev, S. Maji, T. Brox, and J. Malik. Detecting people using mutually consistent poselet activations. In *ECCV*, 2010. 3, 4

[5] L. Bourdev, S. Maji, and J. Malik. Describing people: A poselet-based approach to attribute classification. In *ICCV*, 2011. 3, 5

[6] X. Chao, M. Huiskes, T. Gritti, and C. Ciuhu. A framework for robust feature selection for real-time fashion style recommendation. In *ICME*, 2009. 3

[7] H. Chen, Z. Xu, Z. Liu, and S. Zhu. Composite templates for cloth modeling and sketching. In *CVPR*, 2006. 2

[8] J. Deng, A. Berg, and L. Fei-Fei. Hierarchical semantic indexing for large scale image retrieval. In *SIAM J. Scientific Computing*, 2010. 6

[9] A. Farhadi, I. Endres, D. Hoiem, and D. Forsyth. Describing objects by their attributes. In *CVPR*, 2009. 3

[10] R. Feris and L. Davis. Image ranking and retrieval based on multi-attribute queries. In *CVPR*, 2011. 6

[11] Z. L. Guangcan Liu and Y. Y. Robust subspace segmentation by low-rank representation. In *ICML*, 2010. 6

[12] B. Hasan and D. Hogg. Segmentation using deformable spatial priors with application to clothing. In *BMVC*, 2010. 2

[13] C. Lampert, H. Nickisch, and S. Harmeling. Learning to detect unseen object classes by between-class attribute transfer. In *CVPR*, 2009. 3

[14] Z. Lin, M. Chen, L. Wu, and Y. Ma. The augmented lagrange multiplier method for exact recovery of corrupted low-rank matrices. *Arxiv preprint*, 2010. 6

[15] S. Pan and Q. Yang. A survey on transfer learning. *TKDE*, 2010. 3

[16] Z. Song, M. Wang, X. Hua, and S. Yan. Predicting occupation via human clothing and contexts. In *ICCV*, 2011. 3, 5

[17] N. Wang and H. Ai. Who blocks who: Simultaneous clothing segmentation for grouping images. In *ICCV*, 2011. 2

[18] X. Wang and T. Zhang. Clothes search in consumer photos via color matching and attribute learning. In *ACM MM*, 2011. 3

[19] J. Wright, A. Yang, A. Ganesh, S. Sastry, and Y. Ma. Robust face recognition via sparse representation. *TPAMI*, 2009. 5

[20] M. Yang and K. Yu. Real-time clothing recognition in surveillance videos. In *ICIP*, 2011. 2, 3

[21] Y. Yang and D. Ramanan. Articulated pose estimation with flexible mixtures-of-parts. In *CVPR*, 2011. 2, 3, 4