Ranking images based on aesthetic qualities

Aarushi Gaur
University of Surrey
United Kingdom
Email: a.gaur@surrey.ac.uk

Krystian Mikolajczyk
University of Surrey
United Kingdom
Email: k.mikolajczyk@surrey.ac.uk

***Abstract*—We propose a novel approach for learning image representation based on qualitative assessments of visual aesthetics. It relies on a multi-node multi-state model that represents image attributes and their relations. The model is learnt from pairwise image preferences provided by annotators. To demonstrate the effectiveness we apply our approach to fashion image rating, i.e., comparative assessment of aesthetic qualities. Bag-of-features object recognition is used for the classification of visual attributes such as clothing and body shape in an image. The attributes and their relations are then assigned learnt potentials which are used to rate the images. Evaluation of the representation model has demonstrated a high performance rate in ranking fashion images.**

I. Introduction

Assessing image quality based on visual perspective has gained momentum in recent years in computer vision, machine learning and image processing [1], [2], [3], [4]. Web based image retrieval is starting to reach maturity where a user not only desires to retrieve images but specify higher quality as a priority. Rank aggregation in recent research is often associated with content-based search systems [5], [6], with specific applications for web image searching [7], [8]. Additional miscellaneous areas include object annotation [9], segmentation [10] and saliency detection [11]. A long established use of ranking is found in preferential voting systems [12], [13]. This is generally a much smaller domain and involves fewer number of candidates.

Fig. 1: Application of the approach to fashion interpretation where a large range of factors generally need to be considered. Some of these include different attributes, such as, clothing and body shape, various rules and additional factors like texture, colour and pattern. In this particular example, an apple body shape with recommended outfit is shown.

In this work we implement an approach for comparing images based on qualitative assessment of image content and aesthetic impression. This is in contrast to ranking based on relevance to well defined image content. In this case the presence or absence of objects in the image is not ambiguous and a similarity measure can be established between images. It is however not clear how to establish such measure between the aesthetic impressions the images make. The aesthetic impression can be considered a hidden variable that is affected by various image attributes and relations between the attributes. We propose to construct an image representation using graphical modelling where the attributes and their relations are learnt from the ranked data. Fashion annotators provide the ratings based on pairwise preferences. From this, the annotated datasets are ranked and the underlying relations are extracted as part of the learning process that constructs the model. Our approach can be applied to various computer vision areas; some examples are retrieval [14], recognition [15], annotating data [14], [15] and qualitative assessments [1]. In this work we apply our approach to fashion interpretation which has recently attracted more attention [17], [18], [19], [20].

The goal is to rank images according to certain criteria. An example of this is illustrated in Fig. 1. In various applications these criteria can be very complex. In fashion there are some general rules for determining the suitability of an outfit. These include a range of categories which are inherently complex. As an example, a fitted skirt can be worn with a specific top by a certain body type where the pattern, texture, fabric and colour aspects may also need to be considered to make reliable decisions. Due to the inter-related nature of the various attributes and rules it becomes difficult to annotate images for training an automatic approach. That is why we adopt a different approach that can learn the influence and relations between many components from a ranked list of images. To produce that ranked list of many images by manual annotation we break the task to comparative scoring between two images at a time and we combine the pairwise preferences into global rankings. These rankings are then used as the reference sets for learning and evaluation. Given this data we train a graphical model that captures all the attributes and relations between them. We first outline the related work in Section I-A. Graph based model, representation of the rankings using our learning approach and attribute recognition are discussed in Section II. The dataset along with how the annotations are used to generate the global ranking is presented in Section III. Finally, experimental results for the evaluations performed are shown in Section IV.

A. Related work

Visual assessment of the quality of images using their proposed regional and global features is done in [2] while [3]

automatically assess the aesthetics of images using generic image descriptors, such as, SIFT and GIST. An image quality metric for auto-denoising is presented in [1]. Bag-of-colour-patterns approach that evaluates the colour harmony of photos with aesthetic quality classification is proposed in [4]. In [8] a re-ranking approach that automatically learns different offline visual semantic spaces is given. A graph-theoretical framework for noise resistant ranking is proposed in [7]. Facial beauty modelling was addressed in [16]. In [15] an effective method for parsing clothing in fashion photographs is presented. They also introduce a novel dataset for garment items and present results on using information about clothing estimates to improve pose identification. Cross-scenario clothing retrieval is addressed in [14] where using a human photo taken from the street they find similar clothing from online shops. Key components proposed here include human/clothing parts alignment and an auxiliary daily photo dataset. Closely related work was recently presented in [17] which discusses approaches to obtain image rankings and learn attribute based models. A cloth recommendation application is considered in [18], [19], [20]. However, [18] uses a common sense reasoning rather than vision based learning. In [19] a graphical model is used that given a cloth part proposes another one. Similar idea is exploited in [20] but introduces attributes and occasion components. Our objective is to learn a model directly from a ranked list of images and to rate outfits to reflect recommendations of fashion experts.

II. Image model: learning and recognition

We first give an overview of our approach for ranking the images based on the aesthetic impression they make. We then describe how the global ranking is utilized to model various attributes and rating criteria. Lastly, we present how the attributes are recognised within our approach.

A. Graph based model

The objective of our approach is to rank the images based on the aesthetic impression they make. This can be simplified to producing an absolute rating where the approach is presented with a single image and generates a score within a normalised range of values. The automatic scoring method should be based on the same attributes and criteria that humans take into account when assessing an image. Building a model requires identifying the essential attributes as well as complex relations between them and then learning the weights with which they influence the score. We propose to model the attributes and their relations with graphical modelling, which is well suited to represent the potentials of attributes as states of nodes of a graph as well as relations between the various attributes represented by edges between the graph nodes. The states of each node and the relations between the states have certain potentials with which they contribute to the overall score of aesthetic appearance. In our fashion assessment application the nodes correspond to body parts and the states of the nodes correspond to cloth and body attributes. Fig. 2 illustrates the model we adopt, where edges between the states of the nodes represent relations between the attributes.

B. Learning image ranking

To facilitate the modelling and rating images we consider the position in the ranking as a joint potential of nodes being

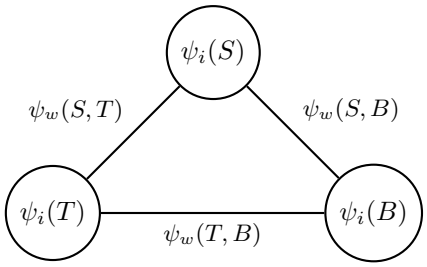

Fig. 2: Object representation model for modelling the ranked lists with 3-nodes each at a range of states. In particular, this figure depicts two nodes for the clothing attributes of top (T), bottom (B) clothing and another for the body shape attribute (S) with the associated node ψ_i and edge ψ_w potentials.

at given states. The higher the individual potentials of the states the higher the position of their configuration in the global ranking. We consider the probabilistic scenario where the dependencies within the graph involve N nodes. The joint probability for this instance is represented using a model based on undirected graphical modelling. The overall rating of an image can therefore be expressed as a product of the attribute potentials and their relations that are present in the image. For all the nodes at states y_i, this is given by the normalized product of non-negative potential functions ψ as:

$$p(y_1, y_2, ..., y_N) = \frac{1}{Z} \prod_{i=1}^{N} \psi_i(y_i) \prod_{w=1}^{W} \psi_w(y_q, y_v), \qquad (1)$$

where potential function ψ_i is associated with node i and ψ_w is associated with edge w connecting nodes y_q and y_v. This distribution is normalised with constant Z given by:

$$Z = \sum_{y_1} \sum_{y_2} \cdots \sum_{y_N} \prod_{i=1}^{N} \psi_i(y_i) \prod_{w=1}^{W} \psi_w(y_q, y_v) \qquad (2)$$

Learning attributes potentials: Learning the model requires estimating all node and edge potentials from a training data. The training data is in the form of a ranked list of images that can be obtained by manual annotation. Providing objective ranking by manual annotation, in particular when there can be hundreds of possible configurations, is not straightforward. We discuss the process of obtaining such ranking in Section III and below we discuss the estimation of the potentials.

For a ranked list R, which may include A examples of the same configuration of nodes at states $y_1, y_2, ..., y_N$, the joint potential of this particular combination is represented as:

$$p(y_1, \bar{y_2}, ..., y_N) = \frac{1}{A} \sum_{i=1}^{A} p(y_{1_i}, y_{2_i}, ..., y_{N_i}) \qquad (3)$$

where $p(y_{1_i}, y_{2_i}, ..., y_{N_i})$ is a rating of an individual example at this particular configuration of states. This allows to accommodate for unbalanced datasets. Once we obtain this estimate for each unique configuration of node states, we can use it to

learn the node potentials ψ_i and edge potentials ψ_w. For $\psi(y_1)$ we average over all configurations that include state $y_1 = z_1$ of node 1 as follows:

$$\psi(y_1 = z_1) = \sum_{y_2} \sum_{y_3} \cdots \sum_{y_N} p(y_1 = z_1, y_2, y_3, ..., y_N) \quad (4)$$

For edge potentials e.g. $\psi(y_1, y_2)$ we use states $y_1 = z_1$ and $y_2 = z_2$:

$$\psi(y_1, y_2) = \sum_{y_3} \sum_{y_4} \cdots \sum_{y_N} p(y_1 = z_1, y_2 = z_2, y_3, y_4, ..., y_N) \quad (5)$$

Fashion aesthetics: In our application of aesthetic assessment we consider a 3-node model with 4 states for the body shape, 5 for top and 6 for bottom clothing attributes. The attributes are listed in Table II. One could add more nodes to represent shoes, jewellery, purse and other accessories as well as more states such as colour and texture but this requires a large training set where each state is included in various configurations of attributes. Furthermore, the study from [17] shows that colour has little impact on the overall dressing attractiveness. For example, in our specific case, the potential $\psi(S_s)$ for body shape at state s is:

$$\psi(S_s) = \sum_{t} \sum_{b} p(S = s, T_t, B_b) \quad (6)$$

Similarly, we can compute the edge potential $\psi(S_s, B_b)$ between the body shape node S and bottom clothing node B at state s and b as follows:

$$\psi(S_s, B_b) = \sum_{t} p(S = s, T_t, B = b) \quad (7)$$

C. Attribute recognition

In order to rank an image with Equation 1 that is based on attributes potentials, we need to recognise all the attributes present in the image. We use SVM classifiers with bags-of-features [9] extracted from grey-value images. There are 15 different attributes including 11 clothing and 4 body shape attributes. The training and testing of the attribute recognition is done using a dataset of images that is discussed in Section III-A and illustrated in Fig. 3. The classification decision can be based on hard threshold of the confidence score that is output by the classifier or by using the label of the classifier with the highest score. We consider both techniques in our experimental evaluation.

III. Ranking based on qualitative assessments

The proposed learning approach requires a set of training images that are ranked by human annotators according to aesthetic impression they make. We first describe the dataset with the different attributes. Next, we present a summary of the method that was used to obtain the ranking for learning the object representation model.

Fig. 3: Some example images from the dataset where the different rows represent the body shape attributes of apple, column, hourglass and pear from top to bottom. From left to right: loose top with fitted skirt, fitted jacket with flared skirt, loose top with flared trousers, fitted jacket with fitted trousers.

A. Dataset

There are several datasets for assessing facial beauty [16] but very few that are related to fashion. A multimodal dataset that includes cloth annotation was collected in [17] but is not publicly available yet and no fashion experts were involved in the annotation. The data from [15] was collected for visual assessments and does not address specific attribute configurations. Therefore, we collect a new dataset from the Internet with images suitable for performing comparative visual assessments. This dataset consists of 1064 images with different clothing attributes worn over a range of body shapes. There are 15 categories altogether with 11 for clothing and 4 for body shapes. The body shapes are generated for each cloth configuration by warping a number of manually selected points on body silhouette to a reference silhouette of a given shape. This has been done very carefully and resulted in a realistic set of examples for different body shapes. The clothing attributes are further divided into top and bottom clothing. The dataset includes fitted, loose and ruffled tops and fitted and loose jackets for the top clothing. Bottom clothing categories are flared, fitted and straight types of both trousers and skirts. The attributes for body shape are apple, column, hourglass and pear. Dataset includes several examples for each of the 120 configurations, which gives 1064 images in total. As an example, an image would have a hourglass body shape with fitted jacket and fitted trousers as shown in Fig. 3.

B. Image ranking using crowdsourcing

The learning process of the representation model requires a list of images that objectively ranks the various configurations of the attributes. Generating a global ranking of images is not straightforward due to the inter-related nature of the attributes and fashion rules. Providing an absolute ranking score within

a certain range of values by an annotator is less reliable than comparing very few images at a time. K–wise rating is proposed in [17] where the annotators have to rank 10 images at a time and the individual rankings are then assembled in a global list. We argue that pairwise preference score is much more efficient to carry out by an annotator and more reliable in terms of following the objective fashion rules as the annotator only has to indicate which of the two presented images is more aesthetically pleasing. We remove faces and convert images to grey-values to avoid bias. Given $N = 1064$ images, $(N^2 - N)/2 = 565516$ unique pairs can be formed. This problem has been extensively studied in the area of electoral voting for which Kemeny–Young method [12], [13] was developed. This method is viewed as a voting algorithm and it not only computes the top voted candidate but also determines an entire ranked list of candidates. It makes use of relative ordering and minimises the disagreements amongst the voters in their pairwise preferences between all the candidates. It uses the mean rank as an initial estimate and does $Ntry$ independent greedy minimizations to produce a ranked list. We observe that with $Ntry > 500$ the produced rankings change very little, we use $Ntry = 2000$.

A fashion expert as well as 10 annotators who have knowledge of fashion and its principles provided binary scores for 57400 out of the 565516 total paired-images. Each annotator compared 7000 paired-images comprising of 5600 unique pairs and 1400 pairs that overlap with the expert's subset for verification. The ranked pairs can be used to generate a global ranking for each annotator independently or for a subset of annotators. This is done with an implementation of the Kemeny–Young preference aggregation [21]. Given a global ranking of all images we can train the node and edge potentials as discussed in Section II-B.

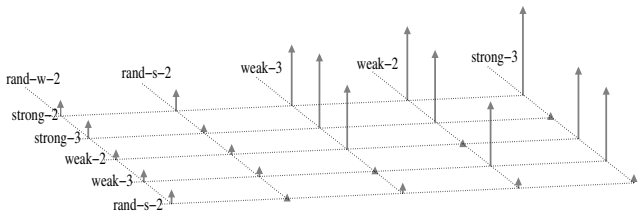

Fig. 4: Kendall's τ [22] correlation between strong, weak and random annotators. The values vary from 0 to 0.45.

IV. Experimental Results

In this section we evaluate the performance of our approach. We first assess the rankings of images and the representation model trained from the rankings. Next, performance for the attributes recognition is discussed. Finally, we investigate the accuracy of produced rankings when the attribute recognition is incorporated in the representation model.

A. Performance evaluation

Data: In order to generate training and test rankings for evaluating the proposed approach we split the 10 annotators into four groups: two strong and two weak ones, with two or three annotators in each group (strong-2, strong-3, weak-2, weak-3). This is done by comparing the annotators rankings with the expert's data using measures of agreement and correlation between the ranked lists, such as Kendall's τ [22] measure.

Kendall's τ measure: It evaluates the agreement and correlation between two global rankings of images. It returns 1 if two input rankings are identical, 0 if there is no correlations, and -1 if the rankings are in inverse order.

Accuracy measure: Any pair of images can be ranked based on scores that are output by the method for each of the images. We can evaluate the performance of the method by measuring the fraction of all pairs that were correctly ranked. The output of this measure is correlated with Kendall's τ but it is more intuitive. The accuracy of 0.5 corresponds to random ranking and 1 indicates that all pairs were correctly ranked. Also Kendall's τ takes into account larger subsets in the global ranking as opposed to a pair of images.

B. Rankings correlations

In this section we measure the correlation between rankings that resulted from different groups of annotators. By using Kendall's τ [22] measure we evaluate every pair of groups i.e. strong-2, strong-3, weak-2, weak-3. In addition, we generate random annotations for the paired-images presented to annotators in group strong-2 and weak-2 which give two random rankings rand-s-2, rand-w-2. Random rankings will serve as a baseline in the experiments. The correlations between groups is displayed in Fig. 4. Diagonal self-correlations were removed to reduce clutter. As expected strong groups have better correlation than weak groups. The random rankings have significantly lower scores as they are not correlated with any of the annotators groups. Interestingly, there is a slight correlation between strong-2 and rand-s-2. This may be due to the fact that these two global rankings were generated from the same subset of pairwise scored images and any random correlation was amplified when generating global rankings. In summary, the agreements amongst strong as well as weak annotators indicate that they apply common fashion criteria and it should be possible to automatically learn these criteria from the provided rankings.

C. Attributes potentials

We use the rankings generated with Kemeny-Young method [21] to learn the node and edge potentials, that is the attribute and relations potentials in the graph model as discussed in Section II-B. In order to better visualise the learnt potentials we subtract from each estimated potential the corresponding potential learnt from a random ranking. Thus negative potentials in Fig. 5 indicate lower than random influence of a body shape or a cloth part on the overall rating of the image. For example, apple body shape, loose jacket and straight skirt have the lowest potentials. In addition, we observe that some potentials differ for strong and weak annotators groups e.g. loose top, which indicates slightly different criteria used by these groups. The overall rating consists of individual node potentials and edge potentials that correspond to the relations between certain clothing and body shapes. The relation potentials are illustrated in Fig. 6. Some relations are particularly strong in both negative and positive impact on the

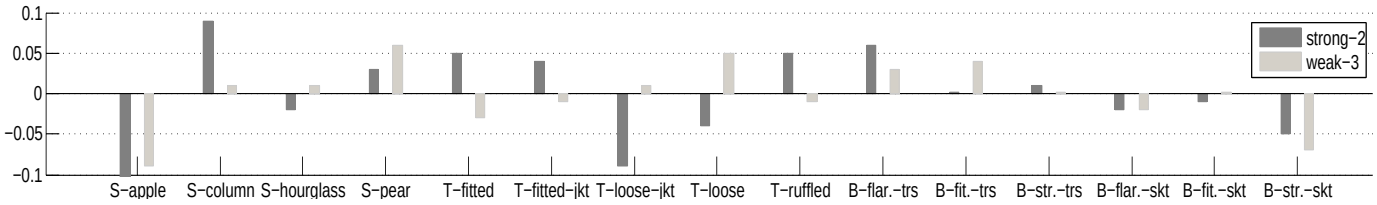

Fig. 5: Attributes potentials learnt from rankings: strong-2 w.r.t. rand-s-2 and weak-3 w.r.t. rand-s-2.

rating e.g. loose jackets or tops in combination with apple shape in contrast to fitted jacket with column shape. These observations have been validated by expert annotator.

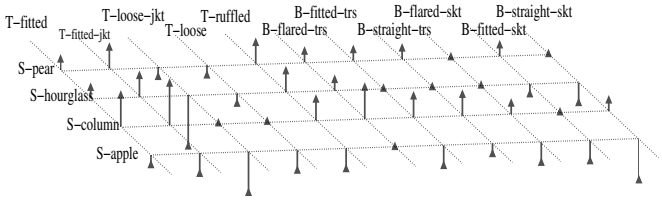

Fig. 6: Attribute relations potentials between body shape and cloth, learnt from rankings: strong-2 w.r.t rand-s-2.

D. Ranking based on learnt attribute potentials

In order to validate the model we train it on image ranking that resulted from one group of annotators and test it on another one. This is measured with the average accuracy of pairwise preference ratings of images. In this experiment we assume that all the states of the nodes, that is the attributes present in the image are known. In this way the disagreements between training and test are only due to the limitations of the proposed model and differences between test and training data. Note that not all pairwise constraints given by the annotators can be satisfied in one global ranking as some of them may contradict each other i.e. same configurations can be scored differently by different annotators or even by the same annotator. Table I shows the percentage of pairs correctly ranked for different training and testing sets. The highest score is obtained when trained and tested on rankings provided by expert i.e. strong-2 and strong-3. The score of 0.91 indicates that the level of contradictions within the pairwise rankings is low ($<$10%). These results also show that the model captures the annotation criteria very well and reflects the ranking of image pairs with high accuracy. We observe that the results gradually decrease when training and testing on weak sets with the lowest results for randomly generated rankings. The random chance score for all train/test combinations is 0.5.

E. Attribute recognition

Previous experiments were assuming that all attributes can be recognised without an error. For an automatic ranking of images we recognise the attributes using bags-of-features approach [9]. We report the performance for the clothing and

test \train	strong-2	strong-3	weak-2	weak-3	rand-s-2
strong-2	0.91	-	-	-	-
strong-3	0.76	0.87	-	-	-
weak-2	0.75	0.76	0.88	-	-
weak-3	0.62	0.66	0.66	0.84	-
rand-s-2	0.58	0.59	0.58	0.55	0.72
rand-w-2	0.55	0.58	0.54	0.52	0.56

TABLE I: Accuracy of ranked pairs of images using the representation model and assuming the node states are known.

body shapes with recall and precision measures. For each category, we split the data randomly into training and test sets. The positive training images for one category are used as the negative examples for all the other categories. Similarly the positive test images for one category are used as the negative test images for all the other categories. The results reported in Table II are for averaged 5 runs of random splits using different thresholds as well as the maximum response of the set of classifiers (cf. Section II-C). A high performance for the 11 clothing categories is obtained with different threshold settings and in particular when taking the label of the maximum prediction value $pmax^*$. Both, precision and recall are very high for the clothing attributes due to their distinctive shape characteristics but much lower for body shapes with an average recall and precision of 0.30. The reason for this performance decline comes from the very subtle differences in features extracted from different body shapes. In addition, the quantisation of SIFT features and spatial bins of the pyramid do not allow to capture these variations. We experimented with different variants of feature extractors but no significant improvement was observed. General shape features are not designed for such task as a body shape recogniser requires much more accurate measurements from the images focussed on the mid body regions and global shape proportions. The best performance of 0.43 was observed for apple body where the shape differences are the largest compared to column or hourglass. We leave the design of such feature for future work as the attribute recognition system is not a contribution in this paper.

F. Image ranking with attribute recognition

The attribute recognition error has an impact on the performance of the entire ranking system. To assess this impact we carry out a controlled experiment where the percentage of the misclassified attributes in the individual classifiers is increased by a constant value for every consequent test. Previous misclassification are kept and used with added error

Category		Perf	$0.5th^*$	$0.7th^*$	$1th^*$	$pmax^*$
top	fitted	rec	0.91	0.90	0.89	0.97
		pre	0.98	1.00	1.00	0.97
	loose	rec	0.73	0.64	0.53	0.92
		pre	0.99	0.99	1.00	0.92
	ruffled	rec	0.75	0.74	0.74	0.90
		pre	0.98	0.99	1.00	0.90
jkt	fitted	rec	0.90	0.87	0.81	0.97
		pre	1.00	1.00	1.00	0.97
	loose	rec	0.83	0.80	0.76	0.94
		pre	1.00	1.00	1.00	0.94
trous	flared	rec	0.83	0.82	0.79	0.93
		pre	0.98	0.98	1.00	0.93
	fitted	rec	0.82	0.77	0.72	0.96
		pre	1.00	1.00	1.00	0.96
	straight	rec	0.72	0.66	0.58	0.92
		pre	0.98	0.98	1.00	0.92
skirt	flared	rec	0.80	0.78	0.73	0.95
		pre	0.98	0.99	1.00	0.95
	fitted	rec	0.84	0.81	0.77	0.97
		pre	0.99	0.99	1.00	0.97
	straight	rec	0.67	0.66	0.63	0.85
		pre	0.96	0.97	1.00	0.85
bshape	all	rec				0.30
		pre				0.30

TABLE II: Recall and precision averaged over five runs of random splits for the clothing and body shape attributes where th^* is the threshold estimate for each individual category and $pmax^*$ is the maximum prediction estimate over the categories that are part of the same region (top, bottom).

for the next test. For every test, we estimate the performance by comparing the training and testing paired-configurations as in Section IV-D. The results are presented in Table III. We make several observations from these results. The performance is only slightly lower compared to results with no error in attribute classification. Moreover, the rate of decline is lower than the actual error induced. For example, for the strong data, the performance decline is from 0.76 to 0.73 at 10% attribute recognition error, that is 73% of image pairs were correctly ranked. Our classification error is far below 10% for most attributes except body shape and such error is realistic for state-of-the-art visual classification systems. Even better performance can be achieved in certain application scenarios e.g. no pose or viewpoint variations in front of a mirror.

train/test	0%	10%	40%	70%
strong-2/strong-3	0.76	0.73	0.63	0.54
weak-2/weak-3	0.66	0.64	0.58	0.53
strong-2/rand-s-2	0.58	0.57	0.53	0.51
weak-2/rand-w-2	0.54	0.53	0.51	0.50

TABLE III: Accuracy of ranked pairs of images using the representation model and attribute recognition with increasing % error for each of the 15 influencing categories.

V. Conclusions and Future Work

This paper proposes an effective approach for learning the ranking of images using qualitative assessments of visual aesthetics. We proposed a graph based representation where node and edge potentials capture the importance of visual attributes and their relations. We also presented a method for learning the model from pairwise preference scores via global ranking of images. We have demonstrated the effectiveness of our approach on a collection of fashion images that include different combinations of clothing and body shapes. The results show that the proposed model can generate rankings very similar to those provided by expert annotators. This approach is not limited to fashion only, it is applicable to other domains where a ranking of data can represent human preferences and the assessment criteria can be defined. One of the future directions is to develop more reliable body shape classifier and extend the model with other attributes such as shoes, colour, and different accessories. A comparison to [17], [20] will also be interesting once the dataset is released.

Acknowledgement. This work was supported by EU Chist-Era EPSRC EP/K01904X/1.

References

[1] X. Kong, K. Li, Q. Yang, L. Wenyin, and M.-H. Yang, "A new image quality metric for image auto-denoising," in *ICCV*, 2013.

[2] W. Luo, X. Wang, and X. Tang, "Content-based photo quality assessment," in *ICCV*, 2011.

[3] L. Marchesotti, F. Perronnin, D. Larlus, and G. Csurka, "Assessing the aesthetic quality of photographs using generic image descriptors," in *ICCV*, 2011.

[4] M. Nishiyama, T. Okabe, I. Sato, and Y. Sato, "Aesthetic quality classification of photographs based on color harmony," in *CVPR*, 2011.

[5] J. He, J. Feng, X. Liu, T. Cheng, T.-H. Lin, H. Chung, and S.-F. Chang, "Mobile product search with bag of hash bits and boundary reranking," in *CVPR*, 2012.

[6] M. Rastegari, C. Fang, and L. Torresani, "Scalable object-class retrieval with approximate and top-k ranking," in *ICCV*, 2011.

[7] W. Liu, Y.-G. Jiang, J. Luo, and S.-F. Chang, "Noise resistant graph ranking for improved web image search," in *CVPR*, 2011.

[8] K. Liu and X. Wang, "Query-specific visual semantic spaces for web image re-ranking," in *CVPR*, 2011.

[9] P. Koniusz, F. Yan, and K. Mikolajczyk, "Comparison of mid-level feature coding approaches and pooling strategies in visual concept detection," *CVIU*, vol. 117, pp. 479–492, 2013.

[10] P. Yadollahpour, D. Batra, and G. Shakhnarovich, "Discriminative re-ranking of diverse segmentations," in *CVPR*, 2013.

[11] C. Yang, L. Zhang, H. Lu, X. Ruan, and M.-H. Yang, "Saliency detection via graph-based manifold ranking," in *CVPR*, 2013.

[12] J. Kemeny, "Mathematics without numbers," *Daedalus*, vol. 88, pp. 577–591, 1959.

[13] H. P. Young and A. Levenglick, "A consistent extension of Condorcet's election principle," *SIAM Journal on Applied Mathematics*, vol. 35, pp. 285–300, 1978.

[14] S. Liu, Z. Song, G. Liu, C. Xu, H. Lu, and S. Yan, "Street-to-shop: Cross-scenario clothing retrieval via parts alignment and auxiliary set," in *CVPR*, 2012.

[15] K. Yamaguchi, M. H. Kiapour, L. E. Ortiz, and T. L. Berg, "Parsing clothing in fashion photographs," in *CVPR*, 2012.

[16] D. Gray, K. Yu, W. Xu, and Y. Gong, "Predicting facial beauty without landmarks," in *ECCV*, 2010.

[17] T. V. Nguyen, S. Liu, B. Ni, J. Tan, Y. Rui, and S. Yan, "Sense beauty via face, dressing, and/or voice," in *ACM Multimedia*, 2012.

[18] E. Shen, H. Lieberman, and F. Lam. "What am I gonna wear?: Scenario-oriented recommendation," in *IUI*, 2007.

[19] T. Iwata, S. Watanabe, and H. Sawada. "Fashion coordinates recommender system using photographs from fashion magazines," in *IJCAI*, 2011.

[20] 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 Multimedia*, 2012.

[21] W. H. Press, "C++ program for kemeny-young preference aggregation," 2012. http://www.nr.com/whp/ky/kemenyyoung.html

[22] M. Kendall, "A new measure of rank correlation," *Biometrika*, vol. 30, pp. 81–89, 1938.