Multimed Tools Appl (2018) 77:22173–22184
https://doi.org/10.1007/s11042-018-5705-2

Adaptive recommendation for photo pose via deep learning

Tong Hao[1] · Qian Wang[1] · Dan Wu[1] · Jin-Sheng Sun[1,2]

Received: 9 October 2017 / Revised: 4 December 2017 / Accepted: 22 January 2018 /
Published online: 15 February 2018

Abstract With the development of image acquisition devices and the popularity of smart phones, more and more people would like to upload their photos to diverse social networks. It is hard to guarantee the quality and artistry of these photos because of not everyone is a professional photographer. In order to handle this problem and further help each common user to improve the beauty of photos, we propose an intelligent photo pose recommendation method to recommended professional photo pose according to everyone's posture in viewfinder. Firstly, the CNN model (VGG-16) is utilized to extract the global features for each photo. Secondly, the salient region detection method is leveraged to extract the regions of interest in each photo. To represent the edge distribution in the local regions, we extract the histogram of oriented gradients. Finally, we propose an effective feature fusion method based on CCA to generate the global visual features for each photo. We implement the Euclidean distance to handle the similarity measure between uploaded photos and the professional photos. The most similar professional photo will be utilized to guide user photo composition. In order to evaluate the performance of the proposed method, we collected a set of professional photos form some professional photography websites. The comparison experiments and user study demonstrate the superiority of the proposed approach.

Keywords Photo pose recommendation · Deep learning · Feature fusion

✉ Tong Hao
joyht2001@163.com

[1] Tianjin Key Laboratory of Animal and Plant Resistance/College of Life Science, Tianjin Normal University, Tianjin 300387, China

[2] Tianjin Aquatic Animal Infectious Disease Control and Prevention Center, Tianjin 300221, China

1 Introduction

With the development of internet technology, people spend more and more time on virtual social networks, such as Facebook, Weibo, Twitter and so on. They are prone to show themselves by different medias. With the development of visual acquisition technology, the visual information is easy to be recorded by cheap visual acquisition equipments. Therefore, diverse visual information has been uploaded by different users to different social networks [10, 23, 32]. However, not everyone is a professional photographer. It is hard to guarantee that each photo uploaded by users has high quality. As shown in Fig. 1, Fig. 1b shows the photos with satisfactory layout while Fig. 1a is not professional enough. In order to attract more attention, users hope to improve the artistry of photos.

To recommend ideal photo pose, the current fancy retrieval and recognition algorithms can be utilized [6, 11, 12, 32, 36]. Babenko et al. [2] investigated the use of neural codes for image retrieval. This descriptor can be used to handle similarity measure between pairwise images. Liu et al. [25] proposed the color difference histograms for image representation. This feature descriptor can effectively represent the structure characteristics. Furthermore, the similarity measure methods can be used to compute the distance between user's photos and professional photos. The structure learning methods and structure matching methods [13, 21, 28] can also be used to handle this problem. The retrieved feedback can be considered as references to guide users to take one picture. Fei et al. [8] proposed an efficient graph feature which can utilize graph model to represent image structure. Krissinel et al. [19] described the SSM algorithm of protein structure comparison in three-dimensional space. However, structure information of images is hard to be represented with human knowledge. These is no one specific method proposed to help users to improve the artistry of photos.

In this paper, we propose an intelligent photo pose recommendation method by global and local feature fusion. First, the popular CNN model (VGG-16) is leveraged to extract globla visual features of photos. Second, the salient region detection method is used to find the regions of interest within each photo and the histogram of oriented gradients is used to represent the local structure information. Third, an modified Canonical Correlation Analysis (CCA) method is used to fuse global and local visual features for united representation. Finally, Euclidean distance is implemented to compute the similarity between user's photos and professional photos. The most similar professional photo will be recommended to guide user's photo composition.

Fig. 1 Some example of user's photo and professional photos. **a** some photos of poor composition; **b** some professional photos

The contributions of this paper are as follows:

- To the best of knowledge, this is the first photo pose recommendation method to help users to improve the artistry of photos;
- We propose to utilize the modified CCA method to handle global and local feature fusion.
- We contribute one professional photo database, which can help other researchers to develop the related applications.

The rest of this paper is structured as follows. Section 2 introduces the related works. The proposed method will be detailed in Section 3. In section 4, we present the comparison experiments and user study to evaluate the performance of this method. Finally, we conclude this paper in Section 5.

2 Related work

Till now, few works have been done for photo pose recommendation. Since the critical step for this task is feature extract and fusion, we will first introduce the related work from this viewpoint. Furthermore, this task is closely related to image retrieval. Therefore, we further introduce the related works on image retrieval.

Feature fusion is a popular research topic, which can effectively improve the robustness of feature representation [9, 24]. Wu et al. [29] proposed a novel feature extraction model, called Feature Fusion Net(FFN), for pedestrian image representation. The model jointly utilizes CNN features and hand-crafted features to form a new deep-feature representation which is more discriminative and compact. Pong et al. [26] proposed a face recognition approach by utilizing the image features at higher and lower resolutions to enhance the information content of the features. It employed the cascaded generalized canonical correlation analysis (GCCA) to fuse the information to form a single feature vector for face recognition. Bai et al. [3] proposed a SoftMax regression-based feature fusion method by learning distinct weights for different features, which enables the estimation of object-to-class similarity measure and the conditional probabilities that each object belongs to different classes. Chen et al. [5] proposed a new feature descriptor called Histogram of Oriented Gradients from Three Orthogonal Planes (HOG-TOP) to extract the dynamic visual features from video sequences. It adopts Multiple Kernel Learning (MKL) to find an optimal feature fusion and trains an SVM with multiple kernels fore motion expression.

Image retrieval is a classic computer vision problem for effective visual information management. Till now, many methods have been proposed to handle this problem [34]. Xia et al. [30] proposed a watermark-based protocol, which can directly embed watermark into the encrypted images by the cloud server to deter illegal distributions. Zhou et al. [35] proposed the cascaded scalar quantization scheme in dual resolution. It formulated the visual feature matching as a range-based neighbor search problem and realized it by identifying hyper-cubes with a dual-resolution scalar quantization strategy. Zhang et al. [33] proposed a graph-based query specific fusion approach. Multiple graphs are merged and re-ranked by conducting a link analysis on a fused graph. Qian et al. [27] proposed an effective sketch-based image retrieval approach with re-ranking and relevance feedback schemes. Yu et al. [31] proposed a ranking model based on large margin structured output learning. Lai et al. [20] proposed a deep architecture that learns instance-aware image representations for multi-label image data. Gordo et al. [15] proposed a novel approach for instance-level image retrieval. It produces a global and compact fixed-length representation for each image by

aggregating many region-wise descriptors and leverages a deep architecture trained for the specific task of image retrieval.

3 Our approach

The proposed method aims to measure the similarity between user's photos and the professional photos. Then, the similar professional photos can be recommended to guide user for photo taking. The framework (Fig. 2) includes four steps: 1) Global feature extraction: we implement the CNN model to extract the global visual feature, which can be considered as the context information of entire photo; 2) Local feature extraction: we implement the salient region detection method [1, 16] to find the regions of interest. Then the histogram of oriented gradients is used to represent the local structure information; 3) Feature fusion: the CCA algorithm is utilized to fuse the global features and the local features for discriminative representation; 4) Similarity measure: specific distance-based metric can be used to compute the similarity between user's photos and the professional photos. we will detail these four steps as follows.

3.1 Global feature extraction

We utilizes deep learning methods to extract visual features from each photo with the popular CNN model, VGG-16, [7]. VGG is a convolutional neural network model that achieved 92.7% top-5 test accuracy in ImageNet [7], which contains over 14 million images, belonging to 1000 classes. However, this dataset mainly includes living pictures, which are quite different from professional photos. To make the CNN model suitable for professional photos, we selected 10% of the data from our dataset to fine-tune the previously trained network to make it adapt to the new dataset. We implemented model fine-tuning with Caffe [18], which is one of the most popular open-source deep learning frameworks.

3.2 Local feature extraction

This step aims to extract local discriminative information to represent the characteristics of user's photos [22]. Firstly, the salient region detection method is utilized to find the key regions in each photo. Then, the popular histogram of oriented gradients (HoG) is used to represent the local structural information. In this work, each specific region in the photo is divided into 8×8 pixel units, and the gradient direction is divided into 9 bins. The histogram of oriented gradients can be computed with respect to each pixel in each cell. Then, we obtain a 9-dimensional feature vector, where each adjacent 4 cells constitute a block

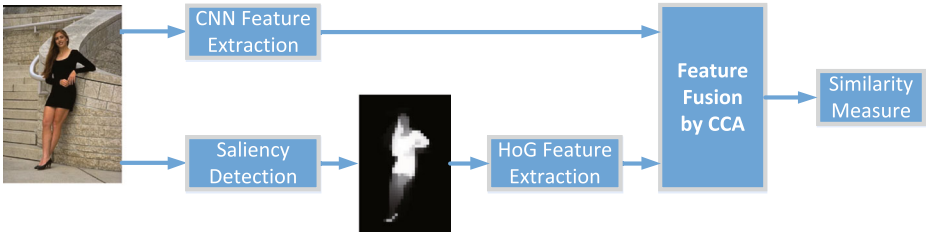

Fig. 2 The framework of the proposed approach

(16 × 16 pixels), and the feature vectors in a block are connected in series to obtain 36-dimensional feature vectors. The blocks are used to scan the image patches (50 × 50 pixels), and the scanning step is 8 units. There are 5 horizontal and vertical scanning windows. Consequently, we obtain a 900-dimensional HoG feature vector for individual image patch.

3.3 Feature fusion & simialrity measure

To take advantage of both global context and local saliency, we propose to leverage Canonical Correlation Analysis (CCA) for diverse feature fusion. Consider pair-wise feature representation, including CNN features, $X = [x_1, ..., x_N] \in R^{D_x \times N}$, and HoG features, $Y = [y_1, ..., y_N] \in R^{D_y \times N}$. They are firstly preprocessed by subtracting the mean value. Then, CCA can transfer both global and local features of the same image to compute the projection pairs for two sets of vectors such that the transformed vectors are maximally correlated in the low dimensional space. In other words, CCA computes pairs of projection vectors, $w_x \in R^{D_x}$ and $w_y \in R^{D_y}$, such that the correlation coefficient can be maximized:

$$\rho = \frac{{w^T}_x X Y^T w_y}{\sqrt{({w^T}_x X X^T w_x)({w^T}_y Y Y^T w_y)}}. \tag{1}$$

With this algebraic operation, ρ can be invariant to the scaling of w_x and w_y, (1) can be transformed into the following maximization problem:

$$\begin{aligned} &arg\max_{w_x, w_y} {w^T}_x X Y^T w_y \\ &s.t. {w^T}_x X X^T w_x = 1 \\ &{w^T}_y Y Y^T w_y = 1 \end{aligned} \tag{2}$$

.

The objective function (2) can be solved by using the Lagrangian multiplier. The objective function L can be formulated by

$$L = {w^T}_x X Y^T w_y + \lambda_x (1 - {w^T}_x X X^T w_x) + \lambda_y (1 - {w^T}_y Y Y^T w_y) \tag{3}$$

By setting the derivative of L with respect to w_x and w_y to zero, we have

$$\begin{cases} \dfrac{\partial L}{\partial w_x} = X Y^T w_y - \lambda_x X X^T w_x = 0 \Longrightarrow X Y^T w_y = \lambda_x X X^T w_x \\ \dfrac{\partial L}{\partial w_y} = Y X^T w_x - \lambda_y Y Y^T w_y = 0 \Longrightarrow Y X^T w_x = \lambda_y Y Y^T w_y \end{cases} \tag{4}$$

With the constraints ${w^T}_x X X^T w_x = 1$ and ${w^T}_y Y Y^T w_y = 1$, we have $\lambda_x - \lambda_y = 0$. The solution of (2) can be obtained by solving the following generalized eigenvalue problem:

$$\begin{bmatrix} 0 & X Y^T \\ Y X_T & 0 \end{bmatrix} \begin{bmatrix} w_x \\ w_y \end{bmatrix} = \lambda \begin{bmatrix} X X^T & 0 \\ 0 & Y Y^T \end{bmatrix} \begin{bmatrix} w_x \\ w_y \end{bmatrix} \tag{5}$$

The objective function (2) can be further formulated as follows:

$$\begin{aligned} &arg\min_{w_x, w_y} \| {w_x}^T X - w_y Y \|^2 \\ &s.t. \| {w_x}^T X \|^2 = 1 \\ &\| w_y Y \|^2 = 1. \end{aligned} \tag{6}$$

From (6), we can see that CCA is a generalized linear regression in nature. w_x and w_y will be learned by (6). Finally, the fused visual feature can be written as:

$$f(I) = [{w_x}^T X; w_y Y]. \tag{7}$$

After previous operations, we can obtain the fusion feature f for each image. Then, the Euclidean distance can be used to handle similarity measure between user's photos and the professional photos. The related function can be formulated as:

$$S(I_i, I_j) = \frac{1}{\sqrt{||f(I_i) - f(I_j)||_2}} \tag{8}$$

According to the similarity measure $S(I_i, I_j)$, we can recommend the similar professional photos to users. These photos can guide user to capture the photos with high quality.

4 Experiment

4.1 Dataset

To the best of our knowledge, this is the first work toward intelligent photo pose recommendation. There is no specific dataset to evaluate the related algorithms. To deal with this problem, we collected the professional photos from several image websites, including Flickr, Weibo and Foursquare. We developed a crawler to collect 50000 images as shown in Fig. 3. These images contain the scenarios of sky, river, sea, mountain, grassland, park, building, and so on. Therefore, this professional photo dataset contains most of common scenarios and can be utilized for user guidance. Different from the classic image dataset classified based on the semantic concepts of the images, we manually labeled these images according to the number of persons in each image. The reason is that photos usually contain persons and the surrounding scenes and the user is guided to take photos by properly setting the location of persons and the ratio of their sizes with respect to the entire scenes. The statistics of this dataset is shown in Table 1.

Fig. 3 Prepared dataset consisting of professional photos

Table 1 Statistics of the prepared dataset

Persons	1	2	3	4	>5
Number	1037	873	1201	1003	1377

4.2 Evaluation of feature fusion

The first important component of the proposed method is feature fusion, which can take advantage of both global context and local saliency for visual representation. Although the popular deep learning features can capture the global characteristics by training on the large scale image dataset, it can not highlight the characteristics of specific regions. On the other hand, although the local appearance and structural features can capture the local characteristics, it loses the context information. For example, it highlights the characteristics of individual persons while ignoring the surrounding sky and sea. Therefore, the local feature is still not discriminative enough for similar photo recommendation. Therefore, it is reasonable to expect the fused feature can benefit improving the performances.

In our work, we evaluated the performances with 3 comparison experiments: 1) we only extract the CNN feature with respect to the entire photo and utilized the Euclidian distance for similarity measure; 2) we only extract the HoG feature with respect to the local salient region and utilized the Euclidian distance for similarity measure; 3) we fuse both features and measure the similarity based on Euclidian distance. The experimental comparison are shown in Table 2. From Table 2, the proposed feature fusion method can consistently outperform the comparison methods. It demonstrated that combining both global and local features can benefit visual representation of individual photos.

4.3 Comparison with the state of the arts

In general, the proposed method can be regarded as the image retrieval method. Thus, some well-known image retrieval methods are selected as the comparison methods. The comparison methods contains:

- Gonde et al. [14] proposed the innovative approach for image retrieval. Especially, it can measure the similarity between pairwise images in the transform domain with the proposed 3D local transform pattern (3D-LTraP).
- Chandrasekhar et al. [4] researched on the problem of neural network model compression for image instance retrieval. They proposed the methods for quantization, coding, pruning and weight sharing to reduce model size for instance retrieval.
- Huang et al. [17] proposed discriminative extreme learning machine (DELM) to replace the classic SVM classifier. Both within-class and between-class scatter matrices are leveraged in DELM to enhance the discrimination capacity for relevance feedback.
- Zhou et al. [35] proposed the cascaded scalar quantization scheme in dual resolution. It formulated the visual feature matching as a range-based neighbor search problem

Table 2 Comparison of different visual features

Methods	Precison(%)	Recall(%)
CNN	87.3	90.1
HoG	73.4	80.3
Our Approach	91.2	92.5

and realized it by identifying hyper-cubes with a dual-resolution scalar quantization strategy.

The comparison experiments are shown in Table 3. From Table 3, we have several key observations:

- The literatures [14] [35] mainly work on novel visual feature formulation. Although both features have specific superiority with respect to specific applications, they can not achieve ideal performances in the generalized application. However, the proposed feature fusion method by taking advantage of both global context and local saliency can well represent the characteristics of photos for image retrieval. Therefore, the proposed method can outperform both methods.
- The literature [4] also works on the deep neural network for feature representation. Although this method can reduce the dimension of feature vector, it lose information with the compact representation. Furthermore, it ignores the local discriminative information. Consequently, it works worse than the proposed method.
- The literature [17] proposed the novel similarity measure method by DELM. Although this new machine learning technique is superior to the Euclidian distance-based similarity measure, this method ignores feature representation, which is critical for this task. Therefore, the proposed method can still outperform this complicated machine learning method by effective feature fusion.

4.4 User study

The quality and artistry of photos is extremely hard to be judged objectively. There is no quantitative standard to evaluate the related algorithms in this field. To evaluate the performance of this method, we conducted a subjective experiments by user study. We invited 20 people to use the developed photo pose recommended system guide photo taking. These 20 people came from different regions: 5 people from Singapore, 5 people from the United States, and 10 people from China. Considering different habits of men and women, 10 men and 10 women were selected. Finally, each people was required to provide 20 photos captured with the guidance of the developed recommended system.

Then, we invited the other 20 people to judge the quality and artistry of these photos. These 20 people also came from different regions: 5 people from the United States, 5 people from Singapore, and the last 10 people from China. In order to ensure the profession of the judgment, all of them are professional photographers. The judgments are based on two criteria:

- The level of similarity of the professional photo for each user's photo
- How satisfying the recommended photos are.

Table 3 Comparison against the state of the arts

Methods	Precison(%)	Recall(%)	Top-5	Top-10
Gonde [14]	85.7	83.4	93.2	91.8
Chandrasekhar [4]	86.3	74.9	94.7	90.3
Huang [17]	80.3	84.6	89.3	87.7
Zou [35]	83.8	80.2	87.0	85.4
Our Approach	91.2	92.5	95.1	94.4

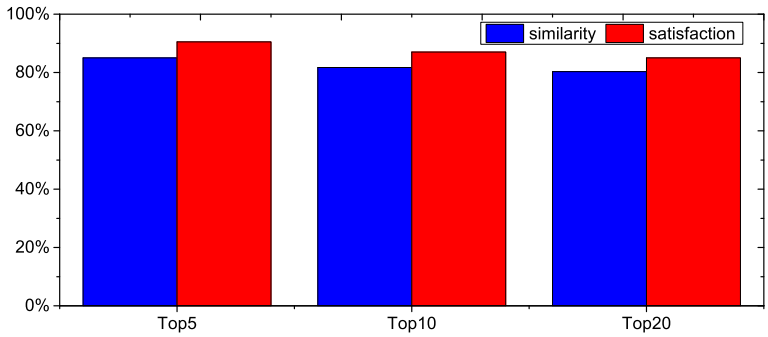

Fig. 4 The experiment result of user study

The final evaluation results are shown in Fig. 4. From these result, we can find that the evaluation of similarity criteria is more than 80%, which demonstrate the accuracy of the feature description. The evaluation of the satisfaction standard is over 83%, which demonstrates that our system improves the quality of the photo, which also meets most of the requirements.

5 Conclusion

In this paper, we propose a novel method for photo pose recommendation. Although we did not develop specific novel visual features and only fused the popular global context feature and local salient features, this method can achieve the best performance comparing against the other competing methods. We further quantitatively demonstrated that the proposed feature fusion method based on CCA can outperform the single CNN/HoG-based method. We employed the user study to subjectively evaluate the proposed method. The user feedback shows that the proposed method can benefit helping user improve the artistry and quality of captured photo.

Acknowledgments This work was funded by National High-Tech Research and Development Program of China (863 programs, 2012AA10A401), Grants of the Major State Basic Research Development Program of China (973 programs, 2012CB114405), National Natural Science Foundation of China (31770904,21106095), National Key Technology R & D Program (2011BAD13B07, 2011BAD13B04), Tianjin Applied Basic and Advanced Technology Research Program (15JCYBJC30700), Project of introducing one thousand high level talents in three years(5KQM110003), Tianjin Normal University Academic Innovation Promotion Program for Young Teachers (52XC1403) and Tianjin Innovative Talent Training Program (ZX110170).

References

1. Achanta R, Hemami S, Estrada F, Susstrunk S (2009) Frequency-tuned salient region detection. In: IEEE conference on Computer vision and pattern recognition, 2009. cvpr 2009. IEEE, pp 1597–1604
2. Babenko A, Slesarev A, Chigorin A, Lempitsky V (2014) Neural codes for image retrieval. In: European conference on computer vision, Springer, pp 584–599
3. Bai X, Liu C, Ren P, Zhou J, Zhao H, Su Y (2015) Object classification via feature fusion based marginalized kernels. IEEE Geosci Remote Sens Lett 12(1):8–12

4. Chandrasekhar V, Lin J, Liao Q, Morere O, Veillard A, Duan L, Poggio T Compression of deep neural networks for image instance retrieval. arXiv:1701.04923
5. Chen J, Chen Z, Chi Z, Fu H (2014) Emotion recognition in the wild with feature fusion and multiple kernel learning. In: Proceedings of the 16th International Conference on Multimodal Interaction, ACM, pp 508–513
6. Cheng Z, Shen J (2016) On very large scale test collection for landmark image search benchmarking. Signal Process 124:13–26
7. Deng J, Dong W, Socher R, Li L, Li K, Feifei L (2009) Imagenet: A large-scale hierarchical image database
8. Fei H, Huan J (2008) Structure feature selection for graph classification. In: Proceedings of the 17th ACM conference on Information and knowledge management, ACM, pp 991–1000
9. Gao Y, Zhen Y, Li H, Chua T-S (2016) Filtering of brand-related microblogs using social-smooth multiview embedding. IEEE Trans Multimedia 18(10):2115–2126
10. Gao Z, Li SH, Zhu YJ, Wang C, Zhang H (2017) Collaborative sparse representation leaning model for RGBD action recognition. J Vis Commun Image Represent 48:442–452
11. Gao Z, Zhang H, Xu GP, Xue YB, Hauptmann AG (2015) Multi-view discriminative and structured dictionary learning with group sparsity for human action recognition. Signal Processs 112:83–97
12. Gao Z, Zhang L, Chen M, Hauptmann AG, Zhang H, Cai A (2014) Enhanced and hierarchical structure algorithm for data imbalance problem in semantic extraction under massive video dataset. Multimedia Tools Appl 68(3):641–657
13. Gens R, Domingos PM (2013) Learning the structure of sum-product networks, pp 873–880
14. Gonde AB, Murala S, Vipparthi SK, Maheshwari R, Balasubramanian R (2017) 3d local transform patterns: A new feature descriptor for image retrieval. In: Proceedings of International Conference on Computer Vision and Image Processing, Springer, pp 495–507
15. Gordo A, Almazán J., Revaud J, Larlus D (2016) Deep image retrieval: Learning global representations for image search. In: European Conference on Computer Vision, Springer, pp 241–257
16. Guo J, Ren T, Bei J (2016) Salient object detection in RGB-D image via saliency evolution. In: IEEE International Conference on Multimedia and Expo, IEEE, pp 1–6
17. Huang X, Sun L, Guo H, Liu S (2016) Discriminative extreme learning machine to content-based image retrieval with relevance feedback. In: 2016 12th World Congress on Intelligent Control and Automation (WCICA), IEEE, pp 3056–3060
18. Jia Y, Shelhamer E, Donahue J, Karayev S, Long J, Girshick R, Guadarrama S, Darrell T (2014) Caffe: Convolutional architecture for fast feature embedding
19. Krissinel E, Henrick K (2004) Secondary-structure matching (ssm), a new tool for fast protein structure alignment in three dimensions. Acta Crystallogr Sec D Biol Crystallogr 60(12):2256–2268
20. Lai H, Yan P, Shu X, Wei Y, Yan S (2016) Instance-aware hashing for multi-label image retrieval. IEEE Trans Image Process 25(6):2469–2479
21. Li A, Morariu VI, Davis LS (2014) Planar structure matching under projective uncertainty for geolocation. In: European Conference on Computer Vision, Springer, pp 265–280
22. Liu A, Su Y, Jia P, Gao Z, Hao T, Yang Z (2015) Multipe/single-view human action recognition via part-induced multitask structural learning. IEEE Trans Cybern 45(6):1194–1208
23. Liu A, Nie W, Gao Y, Su Y (2016) Multi-modal clique-graph matching for view-based 3d model retrieval. IEEE Trans Image Process 25(5):2103–2116
24. Liu A-A, Su Y-T, Nie W-Z, Kankanhalli M (2017) Hierarchical clustering multi-task learning for joint human action grouping and recognition. IEEE Trans Pattern Anal Mach Intell 39(1):102–114
25. Liu G-H, Yang J-Y (2013) Content-based image retrieval using color difference histogram. Pattern Recogn 46(1):188–198
26. Pong K-H, Lam K-M (2014) Multi-resolution feature fusion for face recognition. Pattern Recogn 47(2):556–567
27. Qian X, Tan X, Zhang Y, Hong R, Wang M (2016) Enhancing sketch-based image retrieval by re-ranking and relevance feedback. IEEE Trans Image Process 25(1):195–208
28. Vogelstein JT, Park Y, Ohyama T, Kerr RA, Truman JW, Priebe CE, Zlatic M (2014) Discovery of brain-wide neural-behavioral maps via multiscale unsupervised structure learning. Science 344(6182):386–392
29. Wu S, Chen Y-C, Li X, Wu A-C, You J-J, Zheng W-S (2016) An enhanced deep feature representation for person re-identification. In: 2016 IEEE Winter Conference on Applications of Computer Vision (WACV). IEEE, pp 1–8
30. Xia Z, Wang X, Zhang L, Qin Z, Sun X, Ren K (2016) A privacy-preserving and copy-deterrence content-based image retrieval scheme in cloud computing. IEEE Trans Inf Forensics Secur 11(11):2594–2608

31. Yu J, Tao D, Wang M, Rui Y (2015) Learning to rank using user clicks and visual features for image retrieval. IEEE Trans Cybern 45(4):767–779
32. Zhang H, Shang X, Luan H, Wang M, Chua T (2016) Learning from collective intelligence: Feature learning using social images and tags. TOMCCAP 13(1):1:1–1:23
33. Zhang S, Yang M, Cour T, Yu K, Metaxas DN (2015) Query specific rank fusion for image retrieval. IEEE Trans Pattern Anal Mach Intell 37(4):803–815
34. Zhao S, Yao H, Gao Y, Ji R, Ding G (2017) Continuous probability distribution prediction of image emotions via multitask shared sparse regression. IEEE Trans Multimedia 19(3):632–645
35. Zhou W, Yang M, Wang X, Li H, Lin Y, Tian Q (2016) Scalable feature matching by dual cascaded scalar quantization for image retrieval. IEEE Trans Pattern Anal Mach Intell 38(1):159–171
36. Zhu L, Shen J, Xie L (2016) Unsupervised visual hashing with semantic assistant for content-based image retrieval. IEEE Trans Knowl Data Eng 29(2):472–486

Tong Hao is with the College of Life Sciences, Tianjin Normal University, Tianjin 300387, China.

Qian Wang is a master student of the College of Life Sciences, Tianjin Normal University, Tianjin 300387, China.

Dan Wu is a master student of the College of Life Sciences, Tianjin Normal University, Tianjin 300387, China.

Jin-Sheng Sun is with the College of Life Sciences, Tianjin Normal University, Tianjin 300387, China.