Fang and Yuan *EURASIP Journal on Image and Video Processing* (2018) 2018:44
https://doi.org/10.1186/s13640-018-0282-x

EURASIP Journal on Image and Video Processing

RESEARCH Open Access

Attribute-enhanced metric learning for face retrieval

Yuchun Fang* and Qiulong Yuan

Abstract

Metric learning is a significant factor for media retrieval. In this paper, we propose an attribute label enhanced metric learning model to assist face image retrieval. Different from general cross-media retrieval, in the proposed model, the information of attribute labels are embedded in a hypergraph metric learning framework for face image retrieval tasks. The attribute labels serve to build a hypergraph, in which each image is abstracted as a vertex and is contained in several hyperedges. The learned hypergraph combines the attribute label to reform the topology of image similarity relationship. With the mined correlation among multiple facial attributes, the reformed metrics incorporates the semantic information in the general image similarity measure. We apply the metric learning strategy to both similarity face retrieval and interactive face retrieval. The proposed metric learning model effectively narrows down the semantic gap between human and machine face perception. The learned distance metric not only increases the precision of similarity retrieval but also speeds up the convergence distinctively in interactive face retrieval.

Keywords: Metric learning, Attribute learning, Hypergraph learning, Face retrieval

1 Introduction

The rapid increase of available face images in media, security, and Internet comes up with enormous requirements for retrieval applications. As in most media retrieval tasks, the similarity measure is an essential step in face retrieval. Traditionally, image similarity is measured with distance metrics between the feature vectors of a pair of images. Such measurements rely mainly on the feature extraction strategies, which cannot incorporate sufficient semantic information such as attribute labels. Hence, it is hard to use simple distance metrics to annotate complex semantic correlations in media database and accomplish high-level media retrieval tasks. As a more advanced technique, metric learning takes advantage of more supervision information to refine the general distance metric and reveal the hidden correlations in the retrieval set of media [1–3]. Metric learning has made great achievements in image classification [4, 5] and pedestrian re-identification [6].

Graph-aided metric learning attracts special concentration in the domain of media retrieval. Pourdamghani et al. [7] proposed semi-supervised metric learning to build up the nearest neighbor graph. Baya and Granitto [8] presented a penalized K-nearest neighbor graph metric by minimizing the average silhouette. Graph-based metric learning strategies are usually helpful in annotating the topological structure of high-dimensional image spaces.

In the case of complex attribute correlation of multiple media data, the simple graph is not sufficient to represent complex conceptions. Hypergraph, a more complex model, has been proved more applicable in retrieval tasks. Hypergraph bears good structure to reflect complex semantic correlations [9–11] and proves to be advantageous in various tasks such as image re-ranking [12], clustering [13, 14], classification [15, 16], and content-based image retrieval [9]. Gao et al. [10] used multiple hypergraphs to unify the similarity measurements among the 3-D objects at different granularities. Liu et al. [11] proposed a novel image retrieval framework based on soft hypergraph to better utilize the correlation of image information.

As a very special media content, face image bears very complex semantic conceptions such as identity, demographics, and decorations. Normally, such information is annotated as attribute labels. Multiple facial attributes form very complex conceptional relations in the image database. Hence, it is very natural to model facial attributes with the hypergraph learning framework.

*Correspondence: ycfang@shu.edu.cn
School of Computer Engineering and Science, Shanghai University, Shanghai, China

An example of facial attribute hypergraph is illustrated in Fig. 1, in which each ellipse denotes a hyperedge corresponding to a facial attribute and all images in the same hyperedge share the same facial attribute.

Attribute analysis has received significant attention in recent years and resulted in very promising applications such as object classification [17], image re-ranking [18], image retrieval [19, 20], and face verification [21].

In general cross-media systems, it is a very natural thought to combine attribute labels and image content for retrieval tasks. Unfortunately, the facial attributes are not standardly labeled across various databases and various media. It is hard to design a universal model to adapt to multiple databases and application scenario. Hence, as an extension of our previous work [22], we propose to separately handle the label information and image contents. The attribute labels are utilized to establish a hypergraph learning model, which bears the information of attribute correlation. For any general similarity measure of image contents, the attribute hypergraph serves as a metric learning model to reform the distance metrics with reinforced attribute information. The proposed attribute hypergraph framework for metric learning is shown in Fig. 2. The right block shows the hypergraph model built-up with the attribute labels. For any image database, low-level features can be extracted and mapped into attribute features with general machine learning methods.

Besides its easy adaptation to cross-database applications, another advantage of the metric learning model is to incorporate the semantic annotation for the scenario of human-computer interaction. The visual variance of face image semantics is normally too rich to be annotated with the general low-level features. The resulted semantic gap is often regarded as an obstacle to interactive retrieval. With the attribute-enhanced hypergraph model, the high-level semantic information is embedded into the learned metrics. We validate the proposed model in interactive face retrieval as well as in similarity retrieval. As shown in Fig. 3, the learned attribute-enhanced metrics are utilized in relevance feedback model for interactive retrieval. The coherence between human and machine face perception is promoted with the reformed metrics.

2 Methods

We propose an attribute hypergraph framework for metric learning and apply it in various face image retrieval tasks. The proposed framework can be adapted to face databases labeled with different attribute protocols. In interactive face retrieval, the framework serves to narrow down the semantic gap between human and computer face perception.

2.1 Attribute-based hypergraph learning

The idea of hypergraph originally appeared in [23]. Since Zhou et al. [24] proposed the normalized Laplacian method, hypergraph learning has gradually become popular in various applications. We develop it in matrix level adjustment and adopt the improved learning framework in metric learning.

To describe the method, some important notations and corresponding definitions about hypergraph used throughout this paper are summarized in Table 1.

Different from other hypergraph learning tasks, we intend to transform the source similarity matrix Y into a target similarity matrix F through embedding the attribute information. Therefore, the regularization

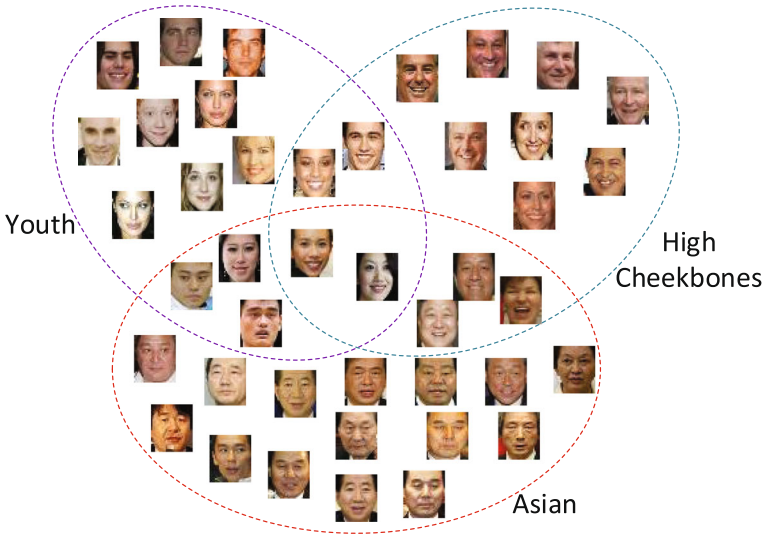

Fig. 1 Illustration of a facial attribute hypergraph

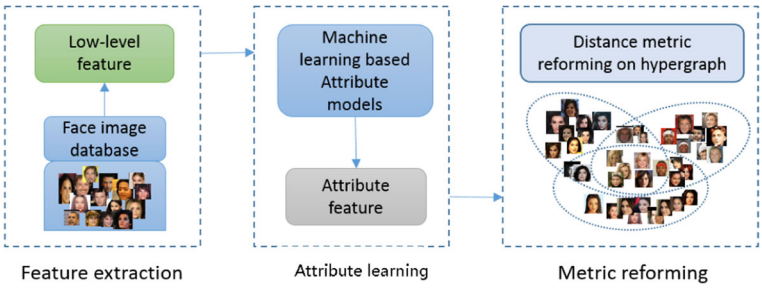

Fig. 2 The flowchart of improved hypergraph learning framework for metric learning

framework of hypergraph learning can be formulated as Eq. (1),

$$\underset{F,\omega}{\operatorname{argmin}}\,\Omega(F,\omega)+\lambda R_{emp}(F)+\mu\sum_{i=1}^{m}\omega^2(e_i) \tag{1}$$

where λ and μ are regularization coefficients.

Let

$$\Theta=D_v^{-\frac{1}{2}}HWD_e^{-1}H^TD_v^{-\frac{1}{2}} \tag{2}$$

The normalized cost function $\Omega(F)$ can be calculated according to Eq. (3).

$$\begin{aligned}&\Omega(F)\\&=\frac{1}{2}\sum_{k=1}^{n}\sum_{e\in E}\sum_{u,v\in e}\frac{\omega(e)h(u,e)h(v,e)}{\delta(e)}\left(\frac{F_{uk}}{\sqrt{d(u)}}-\frac{F_{vk}}{\sqrt{d(v)}}\right)^2\\&=\sum_{k=1}^{n}\mathbf{f_k^T}(I-\Theta)\mathbf{f_k}\\&=\sum_{k=1}^{n}\mathbf{f_k^T}\triangle\mathbf{f_k}\end{aligned} \tag{3}$$

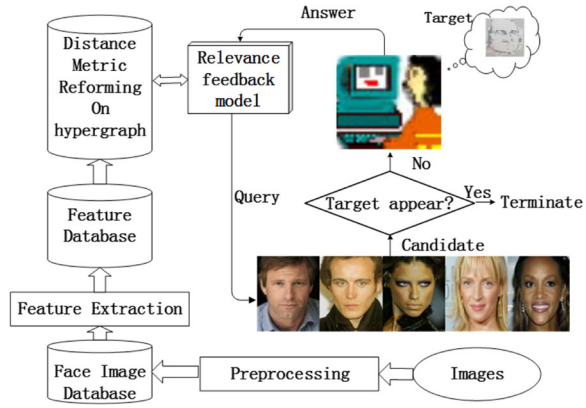

Fig. 3 The flowchart of interactive face retrieval

where I is the identity matrix and the positive semi-definite matrices $\triangle$ is the hypergraph Laplacian.

The empirical loss function in Eq. (1) can be computed according to Eq. (4),

$$R_{emp}(F)=||F-Y||^2=\sum_{k=1}^{n}||\mathbf{f_k}-\mathbf{y_k}||^2 \tag{4}$$

The item $\sum_{i=1}^{m}\omega^2(e_i)$ in Eq. (1) is corresponding to the selection of the hyperedges, considering not all the attributes are available in the hypergraph. The initial values of the hyperedge weights $\omega(e_i)$ can be set flexibly according to the availability of attribute labels corresponding to different databases, for which the details and examples are described in the next Section.

Table 1 Notations and definitions in hypergraph learning model

Notation	Definition
$G=(V,E,\omega)$	Hypergraph of a face image set containing n images and m attributes.
$V=\{v_1,v_2,\ldots,v_n\}$	The set of vertices.
$E=\{e_1,e_2,\ldots,e_m\}$	The set of hyperedges.
$\omega(e_i)$	The weight of the hyperedge e_i. $\sum_{i=1}^{m}\omega(e_i)=1$ and $\omega(e_i)\geq 0$.
W	The diagonal matrix of the hyperedge weights.
$h(v_i,e_i)$	The incidence between a pair of vertex v_i and hyperedge e_i.
H	The incidence matrix of the hypergraph.
$\delta(e_i)$	The degree of the hyperedge e_i.
D_e	The diagonal matrix of the hyperedge degrees.
$d(v_i)$	The degree of the vertex v_i.
D_v	The diagonal matrix of the vertex degrees.
$Y=\{Y_{ij},1\leq i,j\leq n\}$	The source similarity matrix. Y_{ij} represents the distance between image i and image j. $\mathbf{y_k}$ denotes a column of Y.
$F=\{F_{ij},1\leq i,j\leq n\}$	The target similarity matrix. F_{ij} represents the reformed distance between image i and image j. $\mathbf{f_k}$ denotes a column of F.

The solution to Eq. (1) is realized via alternative optimization. With fixed ω, we can obtain the adjustment to F as Eq. (5),

$$F = \left(I + \frac{1}{\lambda}\triangle\right)^{-1} Y \tag{5}$$

When fixing F, we can obtain the adjustment to ω as Eq.(6) ,

$$\omega(e_i) = \frac{\sum_{k=1}^{n} \mathbf{f_k^T} D_v^{-\frac{1}{2}} H \cdot R \cdot D_e^{-1} H^T D_v^{-\frac{1}{2}} \mathbf{f_k}}{\sum_{k=1}^{n} \mathbf{f_k^T} D_v^{-\frac{1}{2}} H D_e^{-1} H^T D_v^{-\frac{1}{2}} \mathbf{f_k}} \tag{6}$$

where R is an $m \times m$ matrix with all entries equal to zero, except $R(i,i) = 1$.

Each iteration of the alternative optimization contributes to the decrease of the value of the objective function until reaching the minimum value 0. The convergence of the iterative process is thus guaranteed [10]. A more detailed deduction about the hypergraph learning theory can be referenced in [22].

With the similarity matrix Y as input, the calculation of the reformed similarity matrix F is summarized in Algorithm 1.

Algorithm 1 Attribute Hypergraph Learning

Input: A distance metric computed through feature matrix $Y = \{Y_{i,j}, 1 \le i,j \le n\}$

Output: Reformed distance metric $F = \{F_{i,j}, 1 \le i,j \le n\}$

1: Initialize the incidence matrix H, the vertex degree matrix D_v, the hyperedge degree matrix D_e and the initial weight matrix W.
2: Construct the hypergraph Laplacian $\triangle = I - \Theta$.
3: **repeat**
4: Compute F according to Eq. (5).
5: Update W according to Eq. (6).
6: Update Θ according to Eq. (2).
7: **until** convergence
8: **return** Reformed distance metric F

2.2 Attribute adaptation for metric

When introducing attribute information, several problems are specially considered in this work. The first problem is to represent facial attributes. The second problem is to incorporate the attribute representation in the metric learning model. Since the attributes are labeled according to different protocols across databases and problems, the third problem is to adapt and transfer the attribute-enhanced metric learning model for general application scenario.

As to face images, the aim of similarity retrieval or interactive retrieval is to find a target with specified personal identity. Hence, an attribute related to personal identification is used for the task. The facial attributes defined in [21] are very typical in facial analysis applications. Besides the identity and demographic labels, the attribute labels such as Oval Face and Chubby are very close to the semantic description of human perception. Except for manual labels of facial attributes, semi-supervised machine learning models are very helpful in annotating facial attribute labels. Any facial image can be transformed into a signature s with any classic feature extraction models such as the Uniform Local Binary Pattern (ULBP) [25], or with deep learning models such as the Very Deep Convolution Networks proposed by the Visual Geometry Group (VGG) [26] and the Deep Residual Networks (Resnet) [27]. In the attribute space, the general binary classifier such as Support Vector Machine (SVM) can be trained to map the signature vector into a scalar as output. The scalar output can either be used as the attribute value [21] or concatenated into an attribute vector a. For each attribute, the scalar output can be thresholded into a binary value o to denote the status of being with or without an attribute.

The setting of the initial value for the hyperedge weights $\omega(e_k)$ serves as incorporating the attribute representation in the hypergraph metric learning model in the section above. Let each entry in $A(i,j) \in [0,1]$ of the similarity matrix A represents the distance between v_i and v_j. We define the initial value of the hyperedge weight $\omega(e_k)$ as in Eq. (7).

$$\omega(e_k) = \frac{\sum_{v_i \in e_k} \sum_{v_j \in e_k} A(i,j)}{\sum_{r=1}^{m} \sum_{v_i \in e_r} \sum_{v_j \in e_r} A(i,j)} \tag{7}$$

Since both the raw image representation r and the learned attribute vector s bear useful identity information, we combine both to define the similarity matrix A in Eq. (8).

$$A(i,j) = \alpha \cdot exp(-D(r_i, r_j)) + \beta \cdot exp(-D(s_i, s_j)) \tag{8}$$

where $D(\cdot)$ denotes the distance metrics and α and β are the balance coefficients of the two parts in Eq. (8).

Besides setting the initial values of the hyperedge weights, the attribute information is utilized to establish the hypergraph model. The available attribute labels or the learned binary values o are utilized to build the incidence matrix H, the diagonal matrix of the hyperedge degrees D_e, and the vertex degrees D_v.

The above measures can be easily transferred to cross-database situations in several aspects. With the attribute provided database, the raw hypergraph model can be constructed directly. For those attribute unavailable or partly available databases, we can first use the classic feature extraction models to obtain signature s. Using the raw features as inputs to the attribute learning models, we can obtain the attribute vector a and the binary attribute value o as well.

2.3 Interactive face retrieval

Face retrieval is a hot topic that has a very close connection to both face recognition and content-based image retrieval. Besides the feature extraction and distance metrics in similarity retrieval, human factor should also be involved in the user feedback during retrieval. Hence, interactive face retrieval is developed to address the interdisciplinary problem of face cognition and image retrieval [28–30]. Interactive face retrieval has wide applications in personal identification, human resource administration, and criminal detection.

In this paper, we use the interactive face retrieval model mentioned in [31, 32] as the experimental platform. The retrieval process is illustrated in Fig. 3. The system aims to search a target face in a database through relevant feedback when the target has no physical form but exists in the memory of a user. During retrieval, the user is required to select the most similar image to the target mental face among the candidates. Then, the probabilistic relevance feedback model provides a new group of candidates based on the one selected as the response by the user. The retrieval process is an iterative process, ending when the target appears or the user abandons the search. The retrieval terminates when the number of iterations exceeds a certain threshold set according to the actual conditions. The probabilistic relevance feedback model updates the candidates based on the posterior calculated from both the user feedback and the distance metric defined in the feature space. Hence, it is challenging to validate and assess the proposed metric learning model with interactive face retrieval.

Two measures are adopted to evaluate the retrieval performance. For K retrieval tests, the iteration number of each test is recorded as $T_i, i = 1, \ldots, N$. The average iteration number $E(T)$ [32] measure with Eq. (9) is a statistic of multiple tests.

$$E(T) = \sum_{i=1}^{K} \frac{T_i}{K} \tag{9}$$

The smaller the number of $E(T)$ is, the smaller the average number of feedback iterations is. Hence, $E(T)$ indicates the retrieval speed of interactive retrieval. Another measurement is the cumulative probability $P(T \leq t)$ [32] as shown in Eq. (10).

$$P(T \leq t) = \frac{L}{K} \tag{10}$$

where $L = |\{T_i | T_i \leq t, i \in \{1, \ldots, K\}\}|$ is the number of targets found in fewer than t iterations among K retrieval tests. When t is fixed, larger value of $P(T \leq t)$ means higher probability of targets found in less than t iterations and better performance of retrieval algorithm.

Another role that the interactive retrieval can play is to measure the representation ability of the models. Since the convergence of the interactive retrieval process relies heavily on the cognition coherence between human and computer in representing semantic information. The coherence measurement is used to evaluate the effectiveness of the reformed distance metric in narrowing the semantic gap between human and computer face perception. The computer selects the most similar image among candidates according to the distance metric, while the human user makes the choice by the memory. As mentioned in [31, 32], cognition coherence is largely influenced by the semantic gap measured by the coherence distribution $P(r)$, which is the percentage of the selection of the user is the r-th ($r = 1, 2, \ldots, |X|$) closest to the target. Higher coherence leads to faster retrieval.

3 Experimental

We conduct the tests on the public dataset LFW [33] and the CFW [20] to assess the performance of the proposed model. The LFW dataset consists of 13,233 face images from 5749 different subjects. The CFW [20] face data set is a large collection of celebrity face images. It contains 200,000 images from 1500 subjects labeled with 14 attributes. Various subsets are selected from LFW and CFW to fit the requirement for further experiments. For parameter selection of hypergraph learning, we use a small subset of LFW with 33 attribute labels, which includes 1680 subjects, 4 images per subject. The obtained results are validated on a subset of CFW for similarity face retrieval to validate the ability of attribute adaptation. To evaluate the coherence between human and machine face perception, the reformed metrics are evaluated on relevance feedback dataset collected on LFW. In the simulation of interactive retrieval, we compare the effect of the metric learning model on several popular metrics. Based on the results of parameter setting experiments, we evaluate the proposed metric learning model in real user experiments of interactive face retrieval.

The attribute feature learning process contains two stages for each image in the datasets. The first stage is extracting the raw image feature with the ULBP [25] and the fine-tuned VGG [26] neural network model. The obtained 3304-dimensional ULBP feature and 1024-dimensional VGG feature are very popular in face recognition or facial attribute recognition. These raw features are usually very high-dimensional and the semantic information is 'hidden' inside them. We project the raw feature into a scalar with a learning model supervised by each attribute. The concatenation of the outputs of these models forms the attribute feature vector. For ULBP feature, a 33-dimensional attribute feature vector is composed of the SVM values as in [21]. For VGG feature, a 14-dimensional attribute vector is composed of the outputs of the fine-tuned VGG.

To evaluate the performance of the proposed attribute hypergraph learning framework in reforming the topology of the image distance metric, four distance measures, including L2, L1, SCD [34] and Chi2 [35], are used in the experiment. The union of the distance measure in Eq. (8) on the raw image features and attribute features is adjusted with the balance coefficients.

According to Eq. (8), the similarity matrix is calculated by the raw image representation and the learned attributes. The coefficients α and β respectively indicate the weight of their contributions to the result of the similarity matrix. We set $\alpha = 0.5$ and $\beta = 0.5$ in the experiments to ensure the balance of raw image information and image attribute information. A series of experiments are conducted in both similarity face retrieval and interactive face retrieval.

4 Results and discussion

In this section, in order to select the appropriate hyperparameter, we first carry out the experiments of parameter selection. On this basis, we apply the proposed attribute-enhanced metric learning model in similarity face retrieval, coherence analysis, simulation of interactive face retrieval, and real user interactive face retrieval experiments in the LFW and CFW datasets.

4.1 Parameter selection for hypergraph learning

For utilizing the proposed metric learning model in retrieval, we need to determine parameter λ in the hypergraph framework. The other parameters can be directly computed from the experimental data. As in [24], we vary λ from 0 to 1 to evaluate its effect in metric learning. We adopt similarity retrieval in parameter setting of λ since it is easier to observe the variation of the parameter with regard to algorithm performance.

In similarity face retrieval experiments, we need each subject to have multiple images. Hence, a subset is selected from LFW containing 1680 subjects and 4 images per subject. The regular measure of precision in information retrieval is adopted to assess the performance of the metric learning model with regard to the parameter λ. Figure 4 illustrates the precision of the original and reformed distance metrics. By the precision curves obtained with 4 general metrics L2, L1, SCD, and Chi2, we can observe that the performance first increases with the growth of λ and then decreases gradually or remains stable. By the measure of precision, it can be concluded that the reformed distance metrics can achieve better performance compared with the original distance metrics. On average, $\lambda \in [0.7, 0.9]$ shows superiority to other values.

4.2 Similarity face retrieval

We first do experiments on similarity face retrieval. To verify the easy adaptation of the proposed metric learning model, we carry out the face retrieval experiments in a subset of the CFW with 400 subjects and 20 images per subject. Based on the analysis in the above section, we

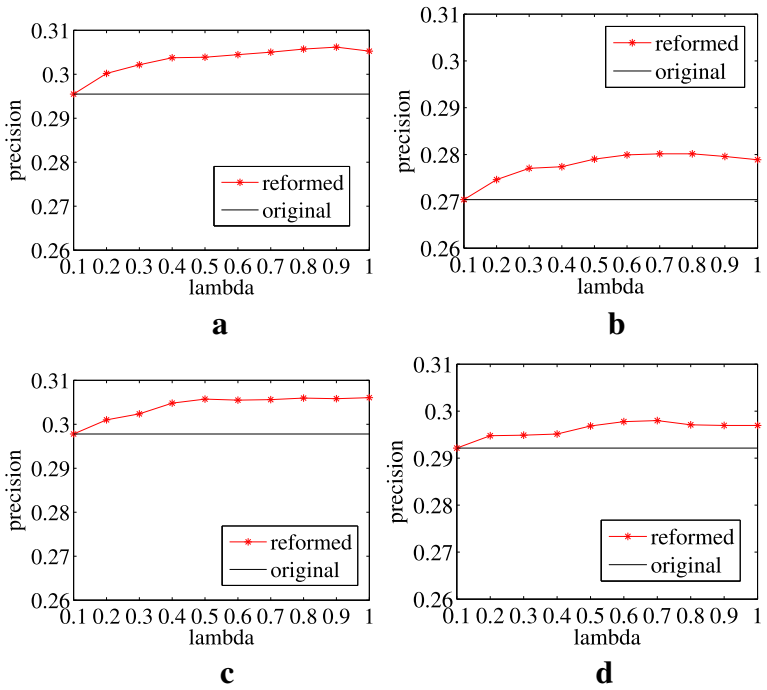

Fig. 4 The precision of the reformed and original distance metrics. **a** chi2. **b** L2. **c** L1. **d** SCD

Table 2 Comparison of the reformed metrics in similarity face retrieval

	Original metrics				Reformed metrics			
Methods	Chi2	L2	L1	SCD	Chi2	L2	L1	SCD
Precision	0.375	0.250	0.313	0.125	0.438	0.250	0.375	0.125
Recall	0.300	0.182	0.238	0.095	0.350	0.200	0.300	0.100
F-measure	0.333	0.211	0.270	0.108	0.389	0.222	0.333	0.111

fix the value $\lambda = 0.8$. The raw feature is the VGG feature and the attribute feature is the output value of VGG for 14 kinds of face attributes such as gender, race, and age. For the four general metrics L2, L1, SCD, and Chi2, we use the proposed hypergraph metric learning model to learn the reformed metrics respectively. In the evaluation, we select top 16 retrieved images to compare with the index image. For one aspect, 16 closest images are sufficient to obtain the precision in the setting of 20 images per subject. For another, we choose top-16 retrieval results in similarity retrieval since the number of displays in interactive retrieval is also 16. Such choice makes it comparable for the two types of retrieval experiments. The performance is evaluated with several general similarity retrieval measurements, i.e., precision, recall, and F1-measure. The experimental results are shown in Table 2.

For all compared distance metrics, the reformed distance metrics result in by par or better performance. The experiments reveal that the metric learning model with the attribute hypergraph is effective in similarity face retrieval. The reformed topology of the image similarity relationship can be effectively transferred across face databases, general image representation schemes, and similarity metrics.

4.3 Coherence analysis

In interactive face retrieval experiments, we evaluate the effectiveness of the reformed distance metric not only by the retrieval convergence speed but also by evaluating the semantic gap between human and computer face perception.

The data for measure the semantic gap is collected under the framework of interactive retrieval. A total of 2838 user feedback records are collected from 15 users in 107 tests. During the retrieval, the number of display images is set to 16 in each query and the user is required to select only 1 from the 16 as the most similar to the target. For the four general metrics, L1, L2, SCD and Chi2, Figs. 5 and 6 show that the reformed metrics have higher $P(r = 1)$ than the original metrics for both ULBP and VGG features. A higher $P(r)$ means more effective in narrowing the semantic gap.

4.4 Simulation experiments of interactive face retrieval

As the process of the real user interactive retrieval is time-consuming, we first conduct simulation experiments to enrich the comparisons. The simulation experiments make use of a designed computer model to simulate a real user retrieval process according to the coherence distribution obtained in the above subsection. In simulation tests, we conduct experiments on the same 200 targets, which are randomly selected from the experimental dataset. We take $E(T)$ as the measurement of convergence speed in the simulation experiment. In simulation retrieval, we especially compare the effect of SCD metrics in raw feature (ULBP) and attribute feature space, as well as the mixture metrics, are shown in Eq. (8).

Figure 7 shows curves of the cumulative probability with respect to iteration numbers. As a comparison reference, the straight line corresponds to the case of retrieval with a random display, in which the candidates are randomly selected. The reformed SCD metrics obtain higher cumulative probability than the mixture metrics and the SCD metrics in both feature spaces. The higher cumulative probability of the reformed metrics leads to faster retrieval. In addition, the proposed metric learning model obtains the best performance on the proposed mixture metrics.

The coherence distribution at $r = 1$ and the average iteration number $E(T)$ of the above comparisons are summarized in Table 3. It can be observed that the reformed distance metrics have higher $P(r = 1)$, and the reformed Mixture metrics has the highest coherence distribution at

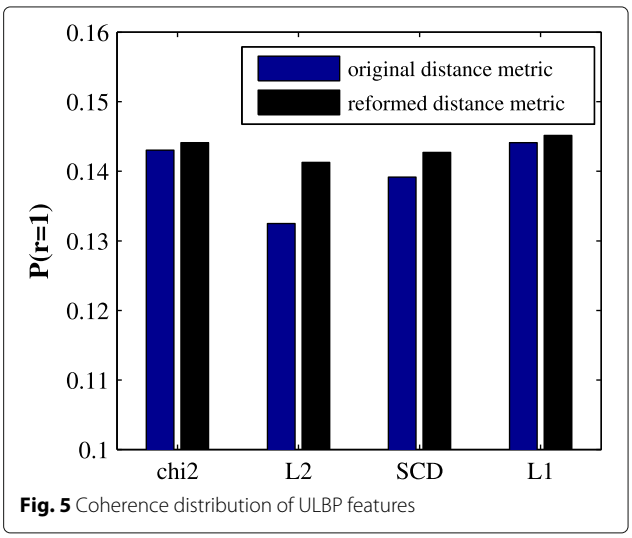

Fig. 5 Coherence distribution of ULBP features

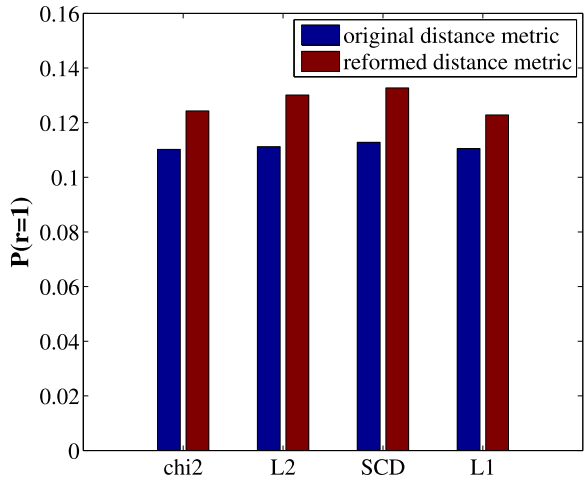

Fig. 6 Coherence distribution of VGG features

$r = 1$. The capability of the reformed metric on the ULBP feature is the most prominent, since the $E(T)$ shows that 8 iterations are saved on average. Based on the above comparison results, we use the reformed mixture metrics in real user retrieval.

4.5 Real users experiments in interactive face retrieval

To further validate the ability of the proposed metric learning, we conduct interactive retrieval with real users. We collect 150 tests from 24 users in the experiments. Generally, most users feel it hard to concentrate on the search after over 60 iterations [31]. Therefore, the users are allowed to give up retrieval when the maximum iteration number exceeds 60. In this experiment, there are 122 tests among all 150 tests that users actually have the target found before the abortion. A simulation test is also performed as the comparison reference.

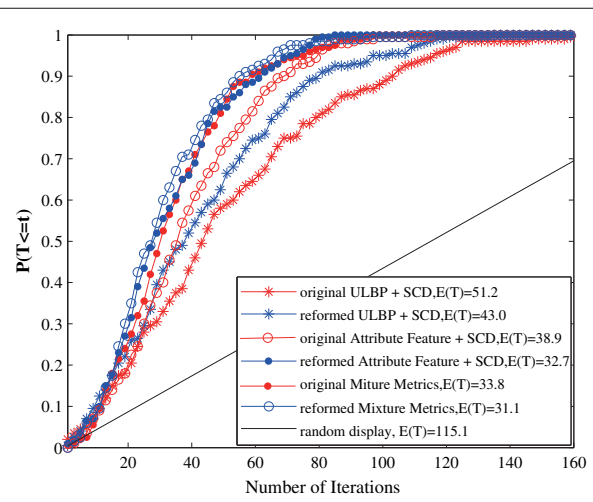

Fig. 7 Simulation tests (best viewed in color)

Table 3 Comparison across metrics and features

Methods	P(r=1)	E(T)	P(r=1)	E(T)
	Original Metrics		Reformed Metrics	
ULBP+SCD	0.139	51.2	0.143	43.0
Attribute feature+SCD	0.153	33.8	0.155	32.7
Mixture metrics	0.164	38.9	0.167	31.1

Figure 8 shows the statistical results, including the simulation experiment, real user experiment, and random display test. According to the curves of cumulative probability with respect to iteration numbers in Fig. 8, the performance of the simulation experiment is close to that of the real user experiment. With the reformed metrics, a cumulative precision approaches 94% in less than 50 iterations. This demonstrates the effectiveness of the reformed metrics with the attribute hypergraph in interactive retrieval.

5 Conclusions

In this paper, we propose an attribute-enhanced metric learning model for face retrieval. The model combines the strength of attribute and hypergraph in a unified framework. The attribute labels are expressed as hypergraph model to reform the distance metrics to incorporate the semantic information. The proposed model combines the attribute semantic in both feature and decision level. Metrics reforming is formulated as learning tasks with the regularization framework on attribute hypergraph. The framework can be easily adapted to various databases, low-level features, attribute learning models, and general metrics. The reformed metrics can promote the coherence of face cognition between human and computer. The effectiveness of the proposed metric learning model is validated with both similarity retrieval and interactive retrieval.

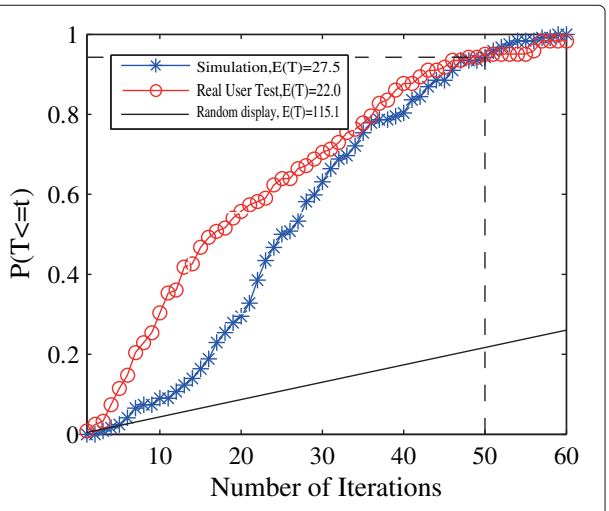

Fig. 8 Real user tests

Abbreviations
CFW: Celebrity faces in the wild; LFW: Labeled faces in the wild; SVM: Support vector machine; ULBP: Uniform local binary pattern; VGG: Visual geometry group

Funding
The work is funded by the National Natural Science Foundation of China (No. 61170155) and the Shanghai Innovation Action Plan Project (No. 16511101200).

Authors' contributions
YF proposed the framework of this work, carried out major experimental research, and drafted the manuscript. QY helped to modify the manuscript and carried out some experiments. Both authors read and approved the final manuscript.

Authors' information
Yuchun Fang, Associate Professor. She gained her Ph.D. from the Institute of Automation, Chinese Academy of Sciences, in 2003. From 2003 to 2004, she worked as a post-doctoral researcher at the France National Research Institute on Information and Automation (INRIA). Since 2005, she has worked at the School of Computer Engineering and Sciences, Shanghai University. She is a member of IEEE, ACM, and CCF (Chinese Computer Federation). Her current research interests include multimedia, pattern recognition, machine learning, and image processing.
Qiulong Yuan, the graduate student at Shanghai University, China. His research interests lie in the field of machine learning and pattern recognition.

Competing interests
The authors declare that they have no competing interests.

Publisher's Note

Received: 7 January 2018 Accepted: 23 May 2018
Published online: 07 June 2018

References
1. JV Davis, B Kulis, P Jain, S Sra, IS Dhillon, in *Proceedings of the 24th international conference on Machine learning*. Information-theoretic metric learning (ACM, 2007), pp. 209–216
2. M Guillaumin, J Verbeek, C Schmid, in *Computer Vision, 2009 IEEE 12th international conference on*. Is that you? Metric learning approaches for face identification (IEEE, 2009), pp. 498–505
3. Z Xu, Y Liu, L Mei, C Hu, L Chen, Semantic based representing and organizing surveillance big data using video structural description technology. J. Syst. Softw. **102**(C), 217–225 (2015)
4. X Han, T Leung, Y Jia, R Sukthankar, AC Berg, in *Computer Vision and Pattern Recognition (CVPR), 2015 IEEE Conference on*. Matchnet: Unifying feature and metric learning for patch-based matching (IEEE, 2015), pp. 3279–3286
5. E Hoffer, N Ailon, in *International Workshop on Similarity-Based Pattern Recognition*. Deep metric learning using triplet network (Springer, 2015), pp. 84–92
6. S Liao, Y Hu, X Zhu, SZ Li, in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Person re-identification by local maximal occurrence representation and metric learning, (2015), pp. 2197–2206
7. N Pourdamghani, HR Rabiee, M Zolfaghari, in *Pattern Recognition (ICPR), 2012 21st International Conference on*. Metric learning for graph based semi-supervised human pose estimation (IEEE, 2012), pp. 3386–3389
8. AE Bayá, PM Granitto, in *Ibero-American Conference on Artificial Intelligence*. Improved graph-based metrics for clustering high-dimensional datasets (Springer, 2010), pp. 184–193
9. Y Huang, Q Liu, S Zhang, DN Metaxas, in *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*. Image retrieval via probabilistic hypergraph ranking (IEEE, 2010), pp. 3376–3383
10. Y Gao, M Wang, D Tao, R Ji, Q Dai, 3-d object retrieval and recognition with hypergraph analysis. IEEE Trans. Image Process. **21**(9), 4290–4303 (2012)
11. Q Liu, Y Huang, DN Metaxas, Hypergraph with sampling for image retrieval. Pattern Recog. **44**(10-11), 2255–2262 (2011)
12. J Cai, ZJ Zha, M Wang, S Zhang, Q Tian, An attribute-assisted reranking model for web image search. IEEE Trans. Image Process. A Publ. IEEE Signal Process. Soc. **24**(1), 261–72 (2015)
13. SR Bul, M Pelillo, A game-theoretic approach to hypergraph clustering. IEEE Trans. Pattern. Anal. Mach. Intell. **35**(6), 1312 (2013)
14. Q Liu, Y Sun, C Wang, T Liu, D Tao, Elastic net hypergraph learning for image clustering and semi-supervised classification. IEEE Trans. Image Process. **26**(1), 452–463 (2017)
15. M Wang, X Liu, X Wu, Visual classification by hypergraph modeling. Knowl. Data Eng. IEEE Trans. **27**(9), 2564–2574 (2015)
16. Q Liu, Y Sun, R Hang, H Song, Spatialspectral locality-constrained low-rank representation with semi-supervised hypergraph learning for hyperspectral image classification. IEEE J. Sel. Top. Appl. Earth Obs. Sremote Sens. **PP**(99), 1–12 (2017)
17. W Kusakunniran, S Satoh, J Zhang, Q Wu, in *Multimedia and Expo (ICME), 2013 IEEE International Conference on*. Attribute-based learning for large scale object classification (IEEE, 2013), pp. 1–6
18. B Siddiquie, RS Feris, LS Davis, in *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*. Image ranking and retrieval based on multi-attribute queries (IEEE, 2011), pp. 801–808
19. BC Chen, YY Chen, YH Kuo, WH Hsu, Scalable face image retrieval using attribute-enhanced sparse codewords. IEEE Trans. Multimed. **15**(5), 1163–1173 (2013)
20. Y Li, R Wang, H Liu, H Jiang, S Shan, X Chen, in *Proceedings of the IEEE International Conference on Computer Vision*. Two birds, one stone: Jointly learning binary code for large-scale face image retrieval and attributes prediction, (2015), pp. 3819–3827
21. N Kumar, AC Berg, PN Belhumeur, SK Nayar, in *Computer Vision, 2009 IEEE 12th International Conference on*. Attribute and simile classifiers for face verification (IEEE, 2009), pp. 365–372
22. Y Fang, Y Zheng, in *Image Processing (ICIP), 2017 IEEE International Conference on*. Metric learning based on attribute hypergraph (IEEE, 2017), pp. 3440–3444
23. C Berge, Graphs and hypergraphs. **34**(8), 1307–1315 (1973)
24. D Zhou, J Huang, B Schölkopf, in *Advances in neural information processing systems*. Learning with hypergraphs: Clustering, classification, and embedding, (2007), pp. 1601–1608
25. T Ahonen, A Hadid, M Pietikainen, Face description with local binary patterns: Application to face recognition. IEEE Trans. Pattern. Anal. Mach. Intell. **28**(12), 2037–2041 (2006)
26. K Simonyan, A Zisserman, Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556 (2014)
27. K He, X Zhang, S Ren, J Sun, in *Proceedings of the IEEE conference on computer vision and pattern recognition*. Deep residual learning for image recognition, (2016), pp. 770–778
28. P Shih, C Liu, Comparative assessment of content-based face image retrieval in different color spaces. Int. J. Pattern Recognit. Artif. Intell. **19**(07), 873–893 (2008)
29. G Sun, J Liu, J Sun, S Ba, in *Innovative Computing, Information and Control, 2006. ICICIC'06. First International Conference on*. Locally salient feature extraction using ICA for content-based face image retrieval, vol. 1 (IEEE, 2006), pp. 644–647
30. M Srikanth, A Ramamurthy, KTV Subbarao, To improve content based face retrieval by creating semantic code words. Ijseat. **2**(12), 956–959 (2014)
31. Y Fang, D Geman, in *International Conference on Audio-and Video-Based Biometric Person Authentication*. Experiments in mental face retrieval (Springer, 2005), pp. 637–646
32. Y Fang, Y Tan, C Yu, in *International Conference on Multimedia Modeling*. Coherence analysis of metrics in LBP space for interactive face retrieval (Springer, 2014), pp. 13–24
33. GB Huang, M Ramesh, T Berg, E Learned-Miller, Labeled faces in the wild: A database for studying face recognition in unconstrained environments (2007). Technical Report 07-49, University of Massachusetts, Amherst
34. ZG Fan, J Li, B Wu, Y Wu, in *Image Processing, 2008. ICIP 2008. 15th IEEE International Conference on*. Local patterns constrained image histograms for image retrieval (IEEE, 2008), pp. 941–944
35. D Gorisse, M Cord, F Precioso, Locality-sensitive hashing for chi2 distance. IEEE Trans. Pattern. Anal. Mach. Intell. **34**(2), 402 (2012)