



Visual Comparison Based on Multi-class Classification Model

Hanqin Shi^{1,2} and Liang Tao¹(✉)

¹ School of Computer Science and Technology,
Anhui University, Hefei 230601, China
taoliang@ahu.edu.cn

² School of Computer Science and Technology,
Huaibei Normal University, Huaibei 235000, China

Abstract. Visual comparison is that given two images, we can not only predict which one exhibits a particular visual attribute more than the other, but also predict whether a visual attribute of one image is equal to that of another image. Most existing methods for visual comparison relying on ranking Support Vector Machine (SVM) functions only distinguish which image in a pair exhibits an attribute more or less in test time. However, it is significant to distinguish which image in a pair exhibits an attribute more, less or equal in test time. To address this issue, we propose a multi-class classification model based on one-versus-one method for visual comparison, which can be formulated by learning mapping functions between any two different classes in image pairs. With regard to the mapping functions, we choose the linear regression functions. Experimental results on the three databases of UT-Zap50K-1, OSR and PubFig demonstrate the advantages of the proposed method.

Keywords: Visual comparison · Multi-class classification
Linear regression model · Ranking SVM · Relative attributes

1 Introduction

Visual comparison is a significant technique in computer vision, which is defined as that given two images and a special attribute, we can predict which one exhibits the attribute more, less or equal compared to the other. As is shown in Fig. 1, we could predict the relation of the right image pair according to prior relation of the left image pairs. Obviously, you are likely to conclude that given some indistinguishable image pairs there are meaningless to only predict the ordered results.

Attributes, which are visual properties describable in words, can capture anything from material properties ('plastic', 'wooden'), shapes ('pointy', 'round') to facial expressions ('serious', 'smiling'). Since their emergence, attributes have inspired a lot of work in image search [1–4], biometrics [5,6], and language based supervision for recognition [7–10]. Those attribute models are mainly



Fig. 1. Visual comparison with attribute ‘Smile’

divided into two forms: binary attributes and relative attributes. Whereas binary attributes are suitable only for clear-cut predicates, such as boxy, relative attributes can show ‘real-valued’ properties that inherently exhibit a series of strengths, such as ‘comfort’. Relative attributes [8] were first proposed by learning the global ranking Support Vector Machine (SVM) functions, followed by much recent work for visual comparison based on ranking SVM functions [3, 11–15]. With relative attributes, originally introduced in [8, 16], images could be compared in terms of how strongly they exhibit a nameable visual property. Given an image pair, relative attributes could indicate which image in a pair exhibits an attribute more or less, while the Just Noticeable Differences (JND) method introduced in [15] could indicate one image in a pair exhibits an attribute equal or not to another image in test time. Now we propose a novel method to indicate which image in a pair exhibits an attribute more, less or equal in test time.

In order to obtain both the ordered pairs and equal pairs in test time for visual comparison, we propose one-versus-one multi-class classification with relative attributes by training the linear regression model for visual comparison, which can be formulated by learning a mapping function between a vector-formed feature input and a scalar-valued output.

Due to the wide existence of multi-class classification problems in different areas, many different methods have been developed to solve such problems. A wide variety of empirical studies have reported the decomposition and ensemble methods can increase the performance on multi-class classification problems. Most existing research shows that the design or selection of decomposition and ensemble strategies play an important role in the performance of decomposition and ensemble methods. With regard to decomposition strategies, One-vs-One (OVO) [17], One-vs-All (OVA) [18], and Error-Correcting Output Coding (ECOC) [19] are the most widely used. Due to the intrinsic of relative attribute, visual comparison with relative attribute could be casted into 3 class classification problem. Compared to OVA, OVO strategy is competent for the case the category number is pretty small.

The main contribution of this paper is the idea to learn OVO multi-class classification by linear regression for visual comparison, which to our knowledge has not been explored for visual comparison in any prior work. The other

contribution is that we not only predict the ordered pairs but also predict equal pairs in test time. Tests on three challenging datasets show that the proposed approach obtains promising results for visual comparison.

2 Related Work

Comparing attributes has gained a lot of interest recently. The relative attributes approach learned a global linear ranking function for each attribute [8], which was extended to non-linear ranking functions in [20,21] by training a hierarchy of rankers and normalizing predictions at the leaf nodes. Aside from learning to rank formulations, researchers have applied the Elo rating system for biometrics [5], and a local learning method based on the ranking SVM [16] was proposed for fine-grained visual comparison. Most of the prior methods produce a ranking function based on SVM for each attributes, whereas we propose multi-class classification with relative attributes and produce a mapping function based on linear regression for each attribute in visual comparison. In contrast to the proposed approach, all those prior methods are only able to predict the ordered image pairs.

Regression is one of the critical techniques for visual attribute application. A number of computer vision problems such as human age estimation could be formulated as a regression problem by learning a mapping function between a high dimensional vector-formed feature input and a scalar-valued output [22–26]. A locally adjusted regression method [24] to search local regions for adjusting was proposed, and followed by bio-inspired features (BIF) for regression [25] in human age-estimation. Most of these regression methods have achieved better performance.

Besides, much prior work [8,16,21] predicting ordered image pairs in test time for visual comparison has been proposed in visual applications. JND method [15] proposed to identify equal image pairs in test time could predict ordered pairs according to the learned ranks of image pairs when the image pairs are not equal, but it cost a lot of time to compute the prior probability due to the local learning. Therefore, we propose a multi-class classification method to conduct visual comparison so that we can not only predict both ordered pairs and equal pairs in test time, but also can save a lot of computational time.

3 Approach

We use the OVO multi-class classification to conduct relative attributes for visual comparison, and apply linear regression to efficiently train the OVO models for visual comparison. In the following, we first introduce OVO multi-class classification for visual comparison, and then present the linear regression model to realize OVO multi-class classification for visual comparison.

3.1 OVO Multi-class Classification for Visual Comparison

Relative attributes are generally obtained from the ranking SVM functions for only predicting ordered pairs, while OVO multi-class classification are obtained from the regression model for predicting both ordered pairs and equal pairs in test time in visual comparison. The regression model will be introduced in the following section.

The Multi-class classification model aims at assigning a class label for each input observation. Given a training data set $\{(X_1, y_1), \dots, (X_n, y_n)\}$, where $X_i \in R^r$ denotes the i th observation feature vector, and $y_i \in \{1, \dots, K\}$ is the class label of the i th observation. It is a mapping function $f : X \rightarrow \{1, \dots, K\}$ inferred from the labeled training data set through a training process. Therefore, visual comparison problem could be casted into the following multi-class classification problem. Given a certain attribute a_m and a set of images $I = \{u_i\}$, each of which is described by the image feature $u_i \in R^d$, and a set of image pairs $P_m = \{(s, t)\}$. P_m is a set of image pairs with attribute a_m and the corresponding class labels can be defined as $l_m \in \{1, 2, 3\}$. Among them, $l_m = 1$ or $l_m = 2$ denote that image s has the attribute a_m more or less than image t respectively, while $l_m = 3$ denotes image s has the attribute a_m as much as image t . We wish to learn a multi-class classification to successfully identify the relation between image s and image t given the attribute a_m . In particular, visual comparison could be categorized to 3 classes (more, less or equal) according to relations of image pairs. To this end, we define pairwise vector between image s and t as follows:

$$x_{st} = p(u_s, u_t), u_s, u_t \in I \quad (1)$$

where p is an entry-wise function that outputs a pairwise vector between u_s and u_t . Therefore, the multi-class training set for visual comparison can be represented as $\{(x, l_m)\}_{st}, \forall (s, t) \in P_m$.

The OVO [27] approach is to divide the multi-class problem with K classes into $C_2^K = K(K-1)/2$ binary classification problems. One binary classifier is constructed for each binary classification problem for discriminating each pair of classes. Let the binary classifier that discriminates the classes of i and j be denoted by f_{ij} , the output of binary classifier f_{ij} , denoted by y_{ij} , is defined as follows.

$$y_{ij} = f_{ij}(x) \quad (2)$$

where $f_{ij}(\cdot)$ is realized by linear regression introduced by the following section. More specifically, y_{ij} is the confidence score denoting that images pair x belongs to i th class, while $1 - y_{ij}$ is the confidence score denoting x belongs to j th class. The class selected by the weighted voting strategy for OVO is the class with the largest total confidence score from all binary classifiers and is defined as [18]:

$$class = \underset{i=1, \dots, K}{argmax} \sum_{1 \leq j \neq i \leq K} (y_{ij}) \quad (3)$$

3.2 OVO by Linear Regression Model

Given the pairwise visual comparison training set $\{(x, l_m)\}$ and the attribute a_m , OVO multi-class classification is to train $N = K(K - 1)/2$ binary classification functions by linear regression. Specifically, we select image pairs of class i and j to train linear regression model. Therefore, we need to learn the mapping relationship between x and l_m by a regression function for binary classification. Most existing relative attribute learning methods aim to establish a mapping by SVM. However linear ridge regression [28] is a classical statistical problem that aims to find a linear function that models the dependencies between vectors $\{x\}$ in R^r and label variables $\{l_m\}$ in R . In this paper, we learn the mapping by a multivariate linear ridge regression function, our goal is to learn N regression functions for each attribute:

$$f_{ij}(x_{st}) = w_m^T x_{st} + b_m, \forall (s, t) \in P'_m \quad \text{and} \quad P'_m \subset P_m \quad (4)$$

The objective functions by Ridge Regularization [29] can be written as:

$$\min \frac{1}{2} \|w_m\|_2^2 + C \sum_{(s,t) \in P'_m} \text{loss}(f_{ij}(x_{st}), l_m(x_{st})) \quad (5)$$

where the constant C is a balanced parameter between minimizing error function and regularization, and $\text{loss}(\cdot)$ denotes the loss function. To simplify the above objective functions without losing generality, quadratic loss function is considered as the loss function. The objective functions are then written as:

$$\min \frac{1}{2} \|w_m\|_2^2 + C \sum_{(s,t) \in P'_m} (l_m(x_{st}) - (w_m^T x_{st} + b_m))^2 \quad (6)$$

To further simplify the above objective functions, we set $z_k = x_{st}$, $N = |P'_m|$, and then the objective functions can be written as

$$\min \frac{1}{2} \|w_m\|_2^2 + C \sum_{k=1}^N (l_m(z_k) - (w_m^T z_k + b_m))^2 \quad (7)$$

where $z_k \in R^r$ is a training vector after the feature reduction, w_m is also a weight vector with r dimensions and $b_m \in R$ is a bias term respectively. The model parameters are estimated by solving an equality-constrained Quadratic Programming Problem, which has a closed-form global optimal solution as follows [6]:

$$\begin{bmatrix} w_m \\ b_m \end{bmatrix} = -(Q^T Q)^{-1} Q^T p \quad (8)$$

where positive semi-definite matrix Q and vector p are given by

$$Q = \begin{bmatrix} 2C \sum_{k=1}^N z_k z_k^T + E & 2C \sum_{k=1}^N z_k \\ 2C \sum_{k=1}^N z_k^T & 2CN \end{bmatrix} \quad (9)$$

$$p = \begin{bmatrix} -2C \sum_{k=1}^N l_m(z_k) z_k \\ -2C \sum_{k=1}^N l_m(z_k) \end{bmatrix} \quad (10)$$

where E is an identity matrix.

Therefore, given a test pair (s, t) and an attribute a_m , we can compute $x_{st} = p(u_s, u_t) = \text{concat}(u_s, u_t)$ and infer $y_{ij}(x_{st})$, and then we can obtain the class of the image pair by Eq. (3). At last we can predict image s exhibits the attribute a_m more, less or equal compared to image t through the obtained class label.

4 Experiments

To validate the advantages of the proposed method, we compare it with several state-of-the-art methods on three datasets: UT-Zap50K-1 [16], the Outdoor Scene Recognition dataset [30] (OSR), and a subset of the Public Figures faces dataset [31] (PubFig). All methods run for 10 random train/test splits on all pairs, in which we select 300 pairs for testing and the remaining for training. In all methods, we simply fix it at $C = 1$ and use the same labeled data as in [8], and then report the accuracy of the percentage of correctly pairs and macro-Average measure (maA) commonly used in evaluating performance on multi-class problems respectively. We will compare the following methods on the above datasets:

- JND [15]: The JND method which develops a Bayesian local learning strategy to infer whether images are indistinguishable or not for a given attribute. If the images are distinguishable, the ordered relation could be obtained by the learned ranks in pairs.
- RSVM + OVA: The method which develops a one-versus-all multi-class classification method by ranking svm (RSVM) for visual comparison.
- RSVM + OVO: The approach which develops a one-versus-one multi-class classification method by ranking svm (RSVM) for visual comparison.
- LRM + OVA: The one-versus-all method which develops a one-versus-all multi-class classification method by linear regression model (LRM) for visual comparison.
- LRM + OVO: The proposed approach which first develops a one-versus-one multi-class classification method by linear regression model (LRM) for visual comparison.

4.1 Experiments on Three Benchmark Datasets

Experiment on UT-Zap50K-1. UT-Zap50K-1 contains 50025 images with 4 attributes (‘Open’, ‘Pointy’, ‘Sporty’, ‘Comfort’) [16]. The image descriptors kindly provided by the authors of each dataset are 960-dim GIST and 30-bin Lab color histograms. We reduce their dimensionality to 30 with PCA to prevent overfitting. For a fair comparison, we take the same feature reduction as in the other methods. Table 1 demonstrates the test results of the proposed method compared to the other methods on UT-Zap50K-1.

Table 1. Accuracy of visual comparison tested on UT-Zap50K-1

Methods	Open	Pointy	Sporty	Comfort
JND [15]	70.88	66.13	63.90	65.83
RSVM + OVA	64.77	65.53	67.98	67.05
RSVM + OVO	66.03	65.93	67.55	67.52
LRM + OVA	70.77	69.28	67.52	69.70
LRM + OVO	74.22	69.72	67.08	71.20

Obviously as seen in Table 1, the accuracy of LRM is far higher than the accuracy of RSVM, which demonstrates the LRM has an advantage over the RSVM in visual comparison. This just validates RSVM method is not optimal because the model used in the method may be more sensitive to training samples. By the same token, the OVO method is superior to the OVA for most attributes in visual comparison. Only for the attribute ‘Sporty’ there are the approximate accuracy. More importantly, from the Table 1 we see the accuracy of LRM + OVO is far higher than that of JND [15], which shows the proposed method is more effective and significant for visual comparison.

The maA measure is another performance measure in multi-class problem, which is defined as follows [32]:

$$maA = \frac{1}{K} \sum_{i=1}^K \frac{n_{ii}}{n_i} \quad (11)$$

where n_{ij} denotes the number of observations of the i class which are predicted as the j class ($i = 1, \dots, K, j = 1, \dots, K$) and $n_i = \sum_{j=1}^K n_{ij}$.

Table 2 shows the maA measure on UT-Zap50K-1. Obviously, the proposed method outperforms the other methods in all attributes except the JND method. The maA measure of the JND is a little more than that of the proposed method with attributes ‘Pointy’ and ‘Sporty’, but the accuracy of the proposed method is a lot more than that of the JND method. Therefore, the proposed method is still effective for visual comparison.

Table 2. The maA measure on UT-Zap50K-1

Methods	Open	Pointy	Sporty	Comfort
JND [15]	67.34	69.67	63.95	64.73
RSVM + OVA	59.20	58.47	62.73	58.33
RSVM + OVO	61.42	59.18	61.53	60.54
LRM + OVA	71.52	67.41	62.84	66.60
LRM + OVO	73.47	68.17	63.18	67.95

Experiment on OSR. The Outdoor Scene Recognition dataset [30] (OSR) consists of 2,688 images with 8 categories and 6 attributes (‘natural’, ‘Open’, ‘perspective’, ‘size-large’, ‘diagonal-plane’ and ‘depth-close’, the corresponding abbreviations are ‘Natr’, ‘Open’, ‘Persp’, ‘LgSi’, ‘Diag’ and ‘ClsD’). The image pairs are those based on category-wise comparisons such that there are about over 20,000 pairs per attribute when we select 30 images in each category. Without loss of generality, we randomly select 1000 pairs used for training and 300 pairs for testing. Tables 3 and 4 respectively show the experimental accuracy and the maA measure on the OSR dataset for all attributes. Similar to the results on UT-Zap50K-1, the proposed method outperforms the other state-of-art methods on OSR. This just demonstrates both LRM and OVO are the effective approaches for visual comparison.

Table 3. Accuracy of visual comparison tested on OSR

Methods	Natr	Open	Persp	LgSi	Diag	ClsD
JND [15]	78.60	73.40	74.25	75.90	76.10	73.20
RSVM + OVA	71.85	68.90	73.10	73.15	73.80	64.15
RSVM + OVO	74.30	68.90	75.95	76.05	76.15	67.00
LRM + OVA	78.75	72.30	76.00	76.75	73.50	71.30
LRM + OVO	81.25	74.75	78.25	77.90	76.40	79.60

Table 4. The maA measure on OSR

Methods	Natr	Open	Persp	LgSi	Diag	ClsD
JND [15]	74.18	71.89	56.50	71.56	69.15	72.54
RSVM + OVA	68.35	67.11	58.37	66.08	63.37	65.16
RSVM + OVO	70.78	66.84	65.46	72.88	70.38	67.27
LRM + OVA	74.50	71.18	61.67	71.92	64.17	71.25
LRM + OVO	78.53	73.72	67.98	74.54	71.34	80.18

Experiment on PubFig. We select a subset of the Public Figures faces dataset [31] (PubFig), which includes 772 images with 8 categories and 11 attributes (‘Masculine_looking’, ‘White’, ‘Young’, ‘Smiling’, ‘Chubby’, ‘Visible_Forehead’, ‘Bushy_Eyebrows’, ‘Narrow_Eyes’, ‘Pointy_Nose’, ‘Big_Lips’, ‘RoundFace’, and the corresponding abbreviations are ‘Male’, ‘White’, ‘Young’, ‘Smil’, ‘Chub’, ‘Foreh’, ‘Eyebrow’, ‘Eye’, ‘Nose’, ‘Lip’, ‘Face’). The method of generating the image pairs is similar to that on OSR. Tables 5 and 6 respectively report the experimental results on the accuracy comparison and the maA measure. Obviously, like the results on OSR, the proposed method almost achieves best results compared to the other methods for all attributes. This further validates the proposed method is an effective method for visual comparison.

Table 5. Accuracy of visual comparison tested on PubFig

Methods	Male	White	Young	Smil	Chub	Foreh
JND [15]	71.70	66.60	75.20	69.25	69.75	69.95
RSVM + OVA	72.20	65.25	74.40	69.50	69.65	66.35
RSVM + OVO	72.65	67.35	75.25	72.60	68.90	69.25
LRM + OVA	73.75	68.25	75.30	72.30	70.75	72.55
LRM + OVO	74.05	70.80	75.95	74.95	71.50	77.10
Methods	Eyebrow	Eye	Nose	Lip	Face	
JND [15]	69.60	70.20	61.70	73.30	74.95	
RSVM + OVA	68.90	69.25	61.25	72.60	75.05	
RSVM + OVO	68.80	69.80	65.55	72.30	75.25	
LRM + OVA	72.40	68.30	65.35	73.40	73.45	
LRM + OVO	73.10	70.35	69.40	75.15	75.05	

Table 6. The maA measure on PubFig

Methods	Male	White	Young	Smil	Chub	Foreh
JND [15]	52.78	46.90	58.02	58.16	54.85	65.35
RSVM + OVA	55.46	46.64	59.05	61.66	57.68	64.44
RSVM + OVO	58.79	53.09	66.41	68.74	60.01	67.18
LRM + OVA	53.29	47.31	59.80	65.33	54.62	73.06
LRM + OVO	60.29	54.84	65.51	71.89	61.94	76.60
Methods	Eyebrow	Eye	Nose	Lip	Face	
JND [15]	51.12	52.87	49.18	57.25	58.73	
RSVM + OVA	52.86	55.21	53.73	58.59	61.30	
RSVM + OVO	57.32	61.51	59.20	62.46	65.60	
LRM + OVA	55.21	54.36	56.64	60.10	56.54	
LRM + OVO	61.92	62.29	62.52	66.42	65.81	

Therefore, from experimental results on the above datasets, we can conclude that the proposed method is an effective method by applying one-versus-one multi-class classification and LRM for visual comparison.

4.2 Time Complexity Analysis

All algorithms are implemented by Matlab on a PC with an Intel i5-4670 CPU @3.40 GHz. Without loss of generality, Tables 7 and 8 respectively shows the train time and test time of the proposed method and JND method in a train/test split pairs on the UT-zap50k-1 dataset under the same setup. From these tables we can conclude that the proposed method significantly reduces the

computational time compared to the JND method. For train time, the reason is that the JND is trained by RSVM which is solved by an optimized iterative approach, while the proposed method is realized by linear regression model which is solved by a closed form. For so much test time of the JND method, it is mainly because it is solved by a local learning strategy by finding K nearest pairs in order to obtain the prior probability of each test pair in test time and the method used to calculate the distance between the pairs is Information Theoretic Metric Learning (ITML) [33] method which is solved by an optimized iterative approach.

Table 7. Train time (s)

Methods	Time
JND [15]	1.7
LRM + OVO	0.018

Table 8. Test time (s)

Methods	Time
JND [15]	78.17
LRM + OVO	0.008

5 Conclusion

In this paper, we have proposed a novel visual comparison method, which applies one-versus-one multi-class classification method and linear regression model with relative attributes for visual comparison. The comprehensive experimental results on three benchmark datasets verified that the proposed method is an effective approach for visual comparison. Meanwhile, the proposed method can save a lot of computational time.

Acknowledgement. This study was funded by information assurance technology collaborative innovation center of anhui university.

References

1. Kumar, N., Belhumeur, P., Nayar, S.: FaceTracer: a search engine for large collections of images with faces. In: Forsyth, D., Torr, P., Zisserman, A. (eds.) ECCV 2008. LNCS, vol. 5305, pp. 340–353. Springer, Heidelberg (2008). https://doi.org/10.1007/978-3-540-88693-8_25
2. Siddiquie, B., Feris, R.S., Davis, L.S.: Image ranking and retrieval based on multi-attribute queries. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 801–808 (2011)

3. Kovashka, A., Parikh, D., Grauman, K.: Whittlesearch: image search with relative attribute feedback. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 2973–2980. IEEE (2012)
4. Kovashka, A., Grauman, K.: Attribute pivots for guiding relevance feedback in image search. In: IEEE Conference on International Conference on Computer Vision, pp. 297–304 (2013)
5. Reid, D.A., Nixon, M.S.: Using comparative human descriptions for soft biometrics. In: International Joint Conference on Biometrics, pp. 1–6 (2011)
6. Chen, K., Gong, S., Xiang, T., Loy, C.C.: Cumulative attribute space for age and crowd density estimation. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 2467–2474 (2013)
7. Lampert, C.H., Nickisch, H., Harmeling, S.: Learning to detect unseen object classes by between-class attribute transfer. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 951–958. IEEE (2009)
8. Parikh, D., Grauman, K.: Relative attributes. In: 2011 International Conference on Computer Vision, pp. 503–510. IEEE (2011)
9. Shrivastava, A., Singh, S., Gupta, A.: Constrained semi-supervised learning using attributes and comparative attributes. In: Fitzgibbon, A., Lazebnik, S., Perona, P., Sato, Y., Schmid, C. (eds.) ECCV 2012. LNCS, vol. 7574, pp. 369–383. Springer, Heidelberg (2012). https://doi.org/10.1007/978-3-642-33712-3_27
10. Biswas, A., Parikh, D.: Simultaneous active learning of classifiers and attributes via relative feedback. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 644–651 (2013)
11. Parkash, A., Parikh, D.: Attributes for classifier feedback. In: Fitzgibbon, A., Lazebnik, S., Perona, P., Sato, Y., Schmid, C. (eds.) ECCV 2012. LNCS, vol. 7574, pp. 354–368. Springer, Heidelberg (2012). https://doi.org/10.1007/978-3-642-33712-3_26
12. Liang, L., Grauman, K.: Beyond comparing image pairs: setwise active learning for relative attributes. In: Proceedings of the IEEE conference on Computer Vision and Pattern Recognition, pp. 208–215 (2014)
13. You, X., Wang, R., Tao, D.: Diverse expected gradient active learning for relative attributes. IEEE Trans. Image Process. **23**(7), 3203–3217 (2014)
14. Qian, B., Wang, X., Cao, N., Jiang, Y.G., Davidson, I.: Learning multiple relative attributes with humans in the loop. IEEE Trans. Image Process. Publ. IEEE Signal Process. Soc. **23**(12), 5573–5585 (2014)
15. Yu, A., Grauman, K.: Just noticeable differences in visual attributes. In: IEEE International Conference on Computer Vision, pp. 2416–2424 (2015)
16. Yu, A., Grauman, K.: Fine-grained visual comparisons with local learning. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 192–199 (2014)
17. Sez, J.A., Galar, M., Luengo, J., Herrera, F.: Analyzing the presence of noise in multi-class problems: alleviating its influence with the one-vs-one decomposition. Knowl. Inf. Syst. **38**(1), 179–206 (2014)
18. Hüllermeier, E., Vanderlooy, S.: Combining predictions in pairwise classification: an optimal adaptive voting strategy and its relation to weighted voting. Pattern Recogn. **43**(1), 128–142 (2010)
19. Allwein, E.L., Schapire, R.E., Singer, Y.: Reducing multiclass to binary: a unifying approach for margin classifiers. In: Seventeenth International Conference on Machine Learning, pp. 9–16 (2000)

20. Li, S., Shan, S., Chen, X.: Relative forest for attribute prediction. In: Lee, K.M., Matsushita, Y., Rehg, J.M., Hu, Z. (eds.) ACCV 2012. LNCS, vol. 7724, pp. 316–327. Springer, Heidelberg (2013). https://doi.org/10.1007/978-3-642-37331-2_24
21. Li, S., Shan, S., Yan, S., Chen, X.: Relative forest for visual attribute prediction. *IEEE Trans. Image Process.* **25**(9), 1 (2016)
22. Yan, S., Wang, H., Tang, X., Huang, T.S.: Learning auto-structured regressor from uncertain nonnegative labels. In: *IEEE Conference on International Conference on Computer Vision*, no. 7, pp. 1–8 (2007)
23. Fu, Y., Huang, T.S.: Human age estimation with regression on discriminative aging manifold. *IEEE Trans. Multimed.* **10**(4), 578–584 (2008)
24. Guo, G., Fu, Y., Dyer, C.R., Huang, T.S.: Image-based human age estimation by manifold learning and locally adjusted robust regression. *IEEE Trans. Image Process.* **17**(7), 1178–1188 (2008)
25. Guo, G., Mu, G., Fu, Y., Huang, T.S.: Human age estimation using bio-inspired features. In: *2009 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2009*, pp. 112–119 (2009)
26. Chan, A.B., Vasconcelos, N.: Counting people with low-level features and Bayesian regression. *IEEE Trans. Image Process.* **21**(4), 2160–2177 (2012)
27. Galar, M., Ndez, A., Barrenechea, E., Bustince, H., Herrera, F.: An overview of ensemble methods for binary classifiers in multi-class problems: experimental study on one-vs-one and one-vs-all schemes. *Pattern Recogn.* **44**(8), 1761–1776 (2011)
28. An, S., Liu, W., Venkatesh, S.: Face recognition using kernel ridge regression. In: *IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2007*, pp. 1–7 (2007)
29. Haitovsky, Y.: On multivariate ridge regression. *Biometrika* **74**(3), 563–570 (1987)
30. Oliva, A., Torralba, A.: Modeling the shape of the scene: a holistic representation of the spatial envelope. *Int. J. Comput. Vis.* **42**(3), 145–175 (2001)
31. Kumar, N., Berg, A.C., Belhumeur, P.N., Nayar, S.K.: Attribute and simile classifiers for face verification. In: *2009 IEEE 12th International Conference on Computer Vision*, pp. 365–372. IEEE (2009)
32. Zhou, L., Wang, Q., Fujita, H.: One versus one multi-class classification fusion using optimizing decision directed acyclic graph for predicting listing status of companies. *Inf. Fusion* **36**, 80–89 (2016)
33. Davis, J.V., Kulis, B., Jain, P., Sra, S., Dhillon, I.S.: Information-theoretic metric learning. In: *International Conference on Machine Learning*, pp. 209–216 (2007)