Joint Latent Space and Multi-view Feature Learning

Kailing Guo, Xiangmin Xu(✉), Bolun Cai, and Tong Zhang

South China University of Techonology, 381 Wushan Road, Guangzhou, China
eecollinguo@gmail.com, caibolun@gmail.com,
{xmxu,tony}@scut.edu.cn

Abstract. GoDec+ shows its robustness in low-rank matrix decomposition but only deals with single-view data. This paper extends GoDec+ to multi-view data by jointly learning latent space and multi-view fusion feature. The proposed method factorizes the low-rank matrix in GoDec+ into the product of a basis matrix of the latent space and a shared representation given by a transformation matrix. By constraining the basis matrix to be group sparse, the proposed method treats the effects of different views differently. Extensive experiments show that the proposed method learns a good fusion feature and outperforms the compared methods in image classification and annotation.

Keywords: Low-rank · Multi-view · Correntropy

1 Introduction

In real applications, multi-view data are common since data are usually collected from different domains or obtained from various feature extrators. For examples, images of one human face taken in different directions, and shape, texture, and color properties of one image. Multi-view learning aims to make good use of the information from different views and is a classic problem in machine learning [1–3].

Canonical correlation analysis (CCA) is commonly used for multi-view data analysis [4]. For two observation vectors x_1 and x_2, CCA finds transformation matrix B_1 and B_2 such that the transformed data $B_1^T x_1$ and $B_2^T x_2$ are maximally correlated. Its probablistic interpretation is that there exists a latent variable z satisfying $x_1 = B_1 z + \epsilon_1$ and $x_2 = B_2 z + \epsilon_2$, where ϵ_1 and ϵ_1 are Gaussian noise [5]. It means that CCA intrinsically finds a common latent representation. Some recent multi-view learning methods also adopt a common latent intrinsic representation and show great success [6,7].

Low-rank is a good property for capturing the intrinsic representation. GoDec+ [8] is a robust and fast low-rank approximation method that maximizing the sum of correntropy of the difference between the original data X and the low-rank approximation $\widetilde{X}$. Discriminative GoDec+ (D-GoDec+) [9]

© Springer Nature Singapore Pte Ltd. 2018
B. Huet et al. (Eds.): ICIMCS 2017, CCIS 819, pp. 23–31, 2018.
https://doi.org/10.1007/978-981-10-8530-7_3

extends GoDec+ for classification by replacing $\widetilde{X}$ with a matrix factorization form BW^TX. Since it is easy to add more information for learning by enforcing constraints on the factor matrices B and W, D-GoDec+ successfully incorporates label information for classification. Motivated by this, we replace $\widetilde{X}$ in GoDec+ by BW^TX and add constraints on B and W for multi-view learning in this paper. Here B is treated as the view generation matrix (i.e., the basis of the latent space) and W is treated as the transformation matrix that extracts the latent intrinsic representation.

Comprehensive experiments on face recognition, digit classification and image annotation demonstrate the effectiveness of the proposed multi-view learning method.

2 Problem Formulation

2.1 Brief Review of GoDec+

Given a data matrix $X \in \mathbb{R}^{m\times n}$, GoDec+ represents the data by a low-rank matrix $\widetilde{X}$ and the error E modeled by a nonlinear similarity measurement correntropy [10]. The definition of correntropy is given as $C(E) = \sum_i^m \sum_j^n g_\sigma(E_{i,j})$, where g_σ is Gaussian kernel $g_\sigma(x) = \exp(-x^2/\sigma^2)$. Maximizing correntropy is equivalent to minimizing the sum of the Welsch M-estimator, which is defined as

$$w(E) = \sum_i^m \sum_j^n [1 - g_\sigma(E_{i,j})]. \tag{1}$$

The model of GoDec+ is given as

$$\min_{\widetilde{X}} w(X - \widetilde{X}), \quad s.t. \quad \mathrm{rank}(\widetilde{X}) \le r, \tag{2}$$

where $\widetilde{X}$ is the low-rank matrix and r is the given rank.

2.2 The Proposed Model

Following [6], group sparsity is enforced on the view generation matrix B to achieve view specific generation sub-matrices. The proposed model is given as follows.

$$\min_{B,W} w(X - BW^TX) + \alpha \sum_{v=1}^{V} \|B_v\|_{2,1} + \frac{\beta}{2}\|W\|_F^2, \tag{3}$$

$$\min_{B,W} w(x_i^v - B_vW^Tx_i^v) + \frac{\alpha}{2}\|B\|_F^2 + \frac{\beta}{2}\|W\|_F^2, \tag{4}$$

$$\min_{B,W} w(x_i^v - B_v(W_s^Tx_i + W_v^Tx_i^v) + \frac{\alpha}{2}\|B\|_F^2 + \frac{\beta}{2}\|W_s\|_F^2 + \frac{\gamma}{2}\sum_{v=1}^{V}\|W_v\|_F^2, \tag{5}$$

where $B = [B_1; B_2; \cdots; B_V]$ with B_i's are the view-specific generation sub-matrices, α and β are positive trade-off parameters, $\|\cdot\|_{2,1}$ denotes $\ell_{2,1}$ norm, and $\|\cdot\|_F$ is Frobenius norm. The $\ell_{2,1}$ norm is defined as

$$\|A\|_{2,1} = \sum_j \sqrt{\sum_i A_{i,j}^2} = \sum_j \|a_{:,j}\|_2. \tag{6}$$

The $\ell_{2,1}$ norm encourages group sparsity and the columns of the matrix tend to be zeroed-out. Thus, each view depends on only a subset of the latent dimensions. By the competitions of the views for data reconstruction, we can learn view-specific generation sub-matrices. The last term in the objective function is regularization term for stable solution.

2.3 Optimization

Since half-quadratic (HQ) optimization is a commonly used optimization method for dealing with correntropy, we give a short review of the main ideas of HQ. Let $\phi(v)$ be a objective function of v that satisfies the preliminary facts [11] of HQ. Then, we have

$$\phi(v) = \min_p \frac{1}{2}(v\sqrt{c} - \frac{p}{\sqrt{c}})^2 + \varphi(p). \tag{7}$$

where c is a constant satisfying that $c > 0$ and $cv^2 - \phi(v)$ is convex, p is an auxiliary variable determined, and $\varphi(.)$ is the dual potential function of $\phi(.)$. It follows that

$$\min_v \phi(v) = \min_{v,p} \frac{1}{2}(v\sqrt{c} - \frac{p}{\sqrt{c}})^2 + \varphi(p). \tag{8}$$

Although the exact formulation of $\varphi(p)$ is often unknown, the minimizer of Eq. (7) can be determined by a specific function $\delta(.)$ only related to $\phi(.)$ with the form

$$p = \delta(v) = cv - \phi'(v). \tag{9}$$

With this solution, minimizing $\phi(v)$ can be solved by iteratively optimizing v and p. When p is given, the sub-problem of minimizing v is a quadratic problem. This is why this method is called half-quadratic optimization. We refer interested readers to [11] for more details.

For our specific problem, $w(v)$ is $\phi(v)$ in (7). According to (9), the function $\delta(\cdot)$ for the Welsch M-estimator is

$$\delta(v) = cv - \frac{2}{\sigma^2} v \exp(-\frac{v^2}{\sigma^2}). \tag{10}$$

In this case, $c = \frac{2}{\sigma^2}$. Define $\alpha_1 = \alpha/c$, $\beta_1 = \beta/c$ and $\hat{X} = X - \frac{T}{c}$, problem (5) changes into

$$\min_{B,W,T} \frac{1}{2}\|\hat{X} - BW^T X\|_F^2 + \frac{\varphi_s(T)}{c} + \alpha_1 \sum_{i=1}^{V} \|B_i\|_{2,1} + \frac{\beta_1}{2}\|W\|_F^2, \tag{11}$$

where T_i is the auxiliary variable introduced by HQ and $\varphi_s(T)$ is defined as $\varphi_s(T) = \sum_{i,j} \varphi(T_{i,j})$. When the other variables are given, it is easy to obtain T by

$$T = cE - \frac{2}{\sigma^2} E \circ g_\sigma(E), \tag{12}$$

where $\circ$ denotes the Hadamard product and $E = X - BW^T X$. Thus, problem (5) can be solved by alternately optimizing the variables.

In order to deal with the term $BW^T X$, the inexact augmented Lagrange multiplier (ALM) method [12] is adopted. Auxiliary variables D and K are introduced and the problem (11) changes into

$$\begin{aligned}
&\min_{\substack{B,W,T\\D,K}} \frac{1}{2}\|\hat{X} - BK\|_F^2 + \frac{\varphi_s(T)}{c} + \alpha_1 \sum_{i=1}^{V} \|D_i\|_{2,1} + \frac{\beta_1}{2}\|W\|_F^2, \\
&s.t. \quad B = D, \quad W^T X = K.
\end{aligned}$$

The augmented Lagrange function of this new optimization problem is

$$\begin{aligned}
&L(B, W, T, D, K, Y_1, Y_2, \mu) \\
&= \frac{c}{2}\|X - BK - \frac{T}{1}\|_F^2 + \frac{\varphi_s(T)}{c} + \alpha_1 \sum_{i=1}^{V} \|D_i\|_{2,1} + \frac{\beta_1}{2}\|W\|_F^2 \\
&\quad + \langle Y_1, B - D\rangle + \frac{\mu}{2}\|B - D\|_F^2 + \langle Y_2, W^T X - K\rangle \\
&\quad + \frac{\mu}{2}\|W^T X - K\|_F^2,
\end{aligned}$$

where Y_1 and Y_2 are the Lagrange multipliers and μ is a positive scalar. When the other variables are fixed, the solutions of B, W, K are given as

$$B = (\hat{X}K^T + \mu(D - \frac{1}{\mu}Y_2))(KK^T + \mu I)^{-1}, \tag{13}$$

$$W = (\mu X X^T + \beta_1 I)^{-1}(\mu X(K - \frac{1}{\mu}Y)^T), \tag{14}$$

and

$$K = (B^T B + \mu I)^{-1}(B^T \hat{X} + \mu W^T X + Y_1). \tag{15}$$

When the other variables are fixed, the minimization of L with respect to D_i is to solve the following problem

$$\alpha_2\|D_i\|_{2,1} + \frac{1}{2}\|B_i - D_i + \frac{1}{\mu}Y_{1,i}\|_F^2, \tag{16}$$

where $\alpha_2 = \alpha_1/\mu$. Define $Q_i = B_i + \frac{1}{\mu}Y_{1,i}\|_F^2$. Following [13], the jth column of the optimal solution is given by

$$[D_i]_{:,j} = \begin{cases} \frac{\|[Q_i]_{:,j}\|_2 - \alpha_2}{\|[Q_i]_{:,j}\|_2}[Q_i]_{:,j}, & \text{if } \|[Q_i]_{:,j}\|_2 > \alpha_2; \\ 0, & \text{otherwise.} \end{cases} \tag{17}$$

Y_1, Y_2 and μ are updated following [12]. Algorithm 1 summarizes the solution to problem (5).

Algorithm 1. The proposed multi-view learning method

```
Input: $X \in \mathbb{R}^{m \times n}, r, \sigma, \alpha, \beta, \rho, \mu_0, \mu_{max}$
Output: $B, W$
Initialize $Y_1 = 0, Y_2 = 0, E = 0, k = 0$.
Generate standard Gaussian matrix $B, W \in \mathbb{R}^{m \times n}$.
Compute $D = B$ and $K = W^T X$.
while not converged do
    Update $W$ as (14);
    for $i = 1$ to $v$ do
        Update $D_i$ as (17);
    end for
    Update $K, B$ and $E$ as (15), (13), and (12).
    Update the Lagrange multipliers as follows:
    $Y_{1,k+1} = Y_{1,k} + \mu_k(B - D)$
    $Y_{2,k+1} = Y_{2,k} + \mu_k(W^T X - K)$
    Update $\mu$ as follows: $\mu_{k+1} = \min(\mu_{\max}, \rho\mu_k)$.
    Update $k$ :    $k \leftarrow k + 1$.
end while
```

3 Experiments

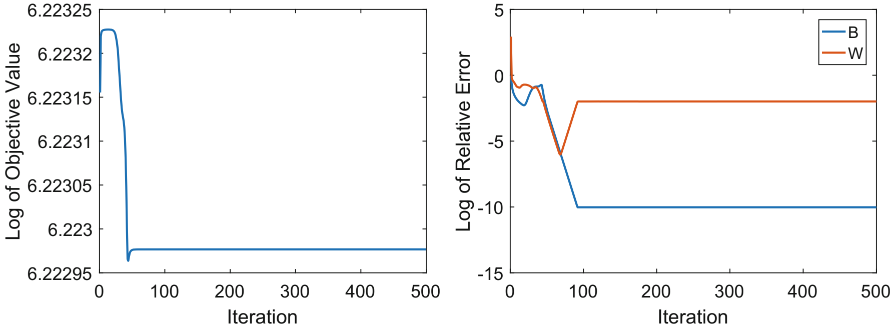
Here we conduct experiments on several popular datasets to verify the effectiveness of the proposed method. It is compared with multi-view intact space learning (MISL) [7], multi-view embedding (MSE) [14], and GoDec+ [8]. For GoDec+, the fusion feature is obtained by projecting the concatenated multi-view data onto the column space of the low-rank matrix learned by GoDec+. The data of each view are rescaled to range in [0,1]. The parameters are tuned for optimal performance and all the experiments are repeated for ten times.

3.1 Face Recognition

The CMU PIE face images dataset [15] contains 68 individuals under 13 different poses, 42 illumination and four expressions. We select two near frontal poses (C9 and C29) as two views to construct the multi-view setting. Each image is reshaped to 32 × 32. K-nearest neighbor (KNN) method based on the Euclidean distance is used for face recognition. We randomly select 50 percent of one individual for training and the rest for test. To study the effectiveness of multi-view learning, experiments results with various combinations are summarized in Table 1. Here C9 and C29 mean using single view, and $\mathrm{C9} + \mathrm{C29}$ means multi-view learning. All methods achieve improvement when combining the two views and the proposed method is the best. The recognition rate is illustrated in Fig. 2 with varying dimension. Even the worst case of the proposed method is better than the best cases of the other methods. The convergence of the variables and the objective value of the proposed method is shown in Fig. 1, which shows that Algorithm 1 converges quite well.

Table 1. Face recognition rate (%) on CMU PIE

Views	C9	C29	C9 + C29
MISL	72.62	72.95	75.75
MSE	79.57	78.49	80.16
GoDec+	70.38	72.50	74.42
Model1	**86.03**	**84.36**	**88.65**



(a) Objective Value

(b) Relative Errors

Fig. 1. Convergence plot of the proposed method on CMU PIE

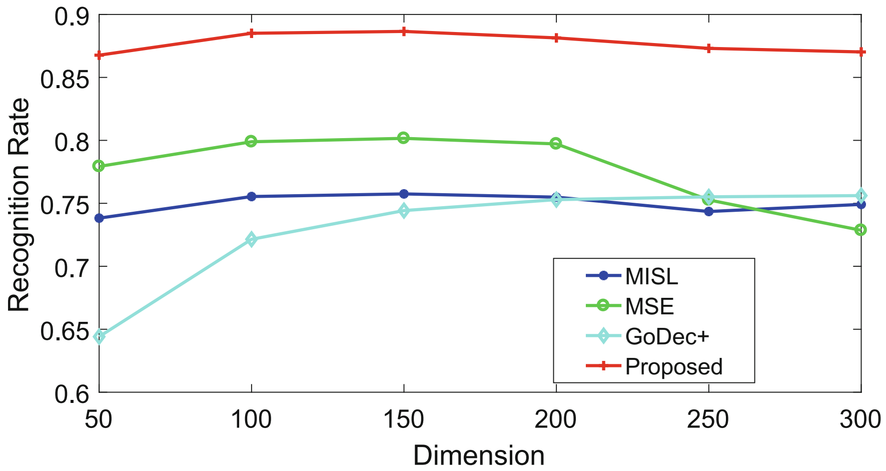


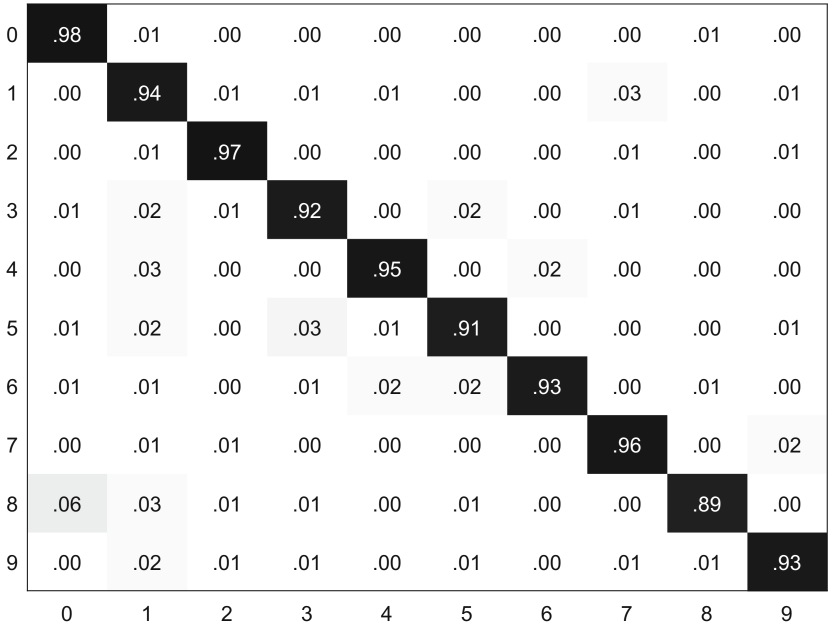
Fig. 2. Recognition rate on CMU PIE with varying dimension

3.2 Digit Classification

The multiple features (MFeat) dataset [16] is a handwritten numeral dataset with 10 categories (i.e., “0–9”) and 200 samples per category. The samples are represented by six kinds of features and the total dimensions of all the features are 649. For each category, 20% of the data are selected for training and the rest for testing. The classification accuracy is reported in Table 2. The proposed method outperforms the other compared method. The confusion matrix of the proposed method is given in Fig. 3. It shows that most categories are classified with high accuracy.

Table 2. Digit classification accuracy (%)

Methods	Accuracy	Methods	Accuracy
MISL	92.67	GoDec+	93.78
MSE	85.01	Model1	**93.83**



	0	1	2	3	4	5	6	7	8	9
0	.98	.01	.00	.00	.00	.00	.00	.00	.01	.00
1	.00	.94	.01	.01	.01	.00	.00	.03	.00	.01
2	.00	.01	.97	.00	.00	.00	.00	.01	.00	.01
3	.01	.02	.01	.92	.00	.02	.00	.01	.00	.00
4	.00	.03	.00	.00	.95	.00	.02	.00	.00	.00
5	.01	.02	.00	.03	.01	.91	.00	.00	.00	.01
6	.01	.01	.00	.01	.02	.02	.93	.00	.01	.00
7	.00	.01	.01	.00	.00	.00	.00	.96	.00	.02
8	.06	.03	.01	.01	.00	.01	.00	.00	.89	.00
9	.00	.02	.01	.01	.00	.01	.00	.01	.01	.93

Fig. 3. Confusion matrix

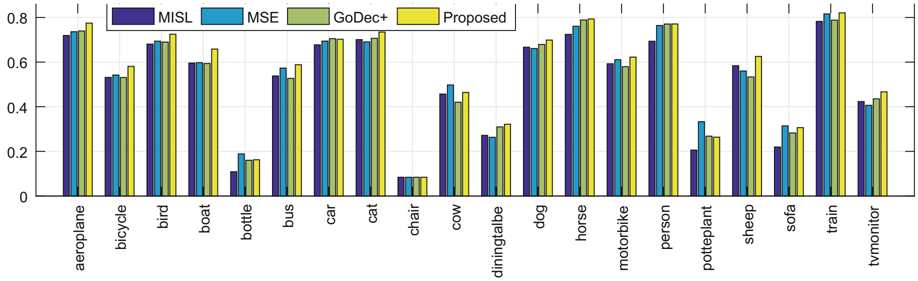


Fig. 4. AP scores of different algorithms for Pascal VOC'07

3.3 Image Annotation

The Pascal VOC'07 dataset [17] contains 9963 images of 20 classes. To simulate a multi-view setting, we choose four types of features including 1000-dimensional "DenseSift", 512-dimensional "Gist", 100-dimensional "DenseHue" and 804-dimensional "Tag" from [18]. Following common setting [17], the images are spited into a training set of 5,011 images and a test set of 4,952 images. A support vector machine (SVM) classifier is trained for the fusion feature of each class. We utilize average precision (AP) for evaluating the performance for each class and mean AP (mAP) for all classes [19]. The AP scores of different algorithms are illustrated in Fig. 4. Table 3 shows the mAP scores. The proposed method performs significantly better than the other methods.

Table 3. Performance on Pascal VOC'07

Methods	mAP	Methods	mAP
MISL	51.19	GoDec+	52.97
MSE	53.94	Model1	**55.76**

4 Conclusion and Future Work

Based on correntropy and matrix factorization, this paper extends GoDec+ to multi-view learning that jointly learns latent space and multi-view fusion feature. Experiment results show that the proposed method is efficient and provides a good feature fusion method in practice. We will extend this method with kernel function in the future.

Acknowledgments. This work is supported in part by the National Natural Science Founding of China (61170192, U1636218), Science and Technology Planning Project of Guangzhou (201704020043), the Fundamental Research Funds for the Central Universities (2017MS045), and Guangzhou Key Lab of Body Data Science (201605030011).

References

1. Zhao, J., Xie, X., Xu, X., Sun, S.: Multi-view learning overview: recent progress and new challenges. Inf. Fusion **38**, 43–54 (2017)
2. Hong, R., Zhang, L., Zhang, C., Zimmermann, R.: Flickr circles: aesthetic tendency discovery by multi-view regularized topic modeling. IEEE Trans. Multimedia **18**(8), 1555–1567 (2016)
3. Hong, R., Hu, Z., Wang, R., Wang, M., Tao, D.: Multi-view object retrieval via multi-scale topic models. IEEE Trans. Image Process. **25**(12), 5814–5827 (2016)
4. Podosinnikova, A., Bach, F., Lacoste-Julien, S.: Beyond CCA: moment matching for multi-view models. In: Proceedings of 33rd International Conference on Machine Learning (2016)
5. Bach, F.R., Jordan, M.I.: A probabilistic interpretation of canonical correlation analysis (2005)
6. Guan, N., Tao, D., Luo, Z., Shawe-Taylor, J.: MahNMF: Manhattan non-negative matrix factorization. arXiv preprint arXiv:1207.3438 (2012)
7. Xu, C., Tao, D., Xu, C.: Multi-view intact space learning. IEEE Trans. Pattern Anal. Mach. Intell. **37**(12), 2531–2544 (2015)
8. Guo, K., Liu, L., Xu, X., Xu, D., Tao, D.: Godec+: fast and robust low-rank matrix decomposition based on maximum correntropy. IEEE Trans. Neural Netw. Learn. Syst. (2017)
9. Guo, K., Xu, X., Tao, D.: Discriminative Godec+ for classification. IEEE Trans. Sig. Process. (2017)
10. Liu, W., Pokharel, P.P., Príncipe, J.C.: Correntropy: properties and applications in non-Gaussian signal processing. IEEE Trans. Sig. Process. **55**(11), 5286–5298 (2007)
11. Nikolova, M., Ng, M.K.: Analysis of half-quadratic minimization methods for signal and image recovery. SIAM J. Sci. Comput. **27**(3), 937–966 (2005)
12. Lin, Z., Chen, M., Ma, Y.: The augmented Lagrange multiplier method for exact recovery of corrupted low-rank matrices. arXiv preprint arXiv:1009.5055 (2010)
13. Liu, G., Lin, Z., Yan, S., Sun, J., Yu, Y., Ma, Y.: Robust recovery of subspace structures by low-rank representation. IEEE Trans. Pattern Anal. Mach. Intell. **35**(1), 171–184 (2013)
14. Xia, T., Tao, D., Mei, T., Zhang, Y.: Multiview spectral embedding. IEEE Trans. Syst. Man Cybern. Part B (Cybern.) **40**(6), 1438–1446 (2010)
15. Sim, T., Baker, S., Bsat, M.: The CMU pose, illumination, and expression (PIE) database. In: Proceedings of IEEE International Conference on Automatic Face and Gesture Recognition, pp. 46–51 (2002)
16. Lichman, M.: UCI machine learning repository (2013)
17. Everingham, M., Van Gool, L., Williams, C.K., Winn, J., Zisserman, A.: The Pascal visual object classes (VOC) challenge. Int. J. Comput. Vis. **88**(2), 303–338 (2010)
18. Guillaumin, M., Verbeek, J., Schmid, C.: Multimodal semi-supervised learning for image classification. In: 2010 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 902–909. IEEE (2010)
19. Hong, R., Yang, Y., Wang, M., Hua, X.S.: Learning visual semantic relationships for efficient visual retrieval. IEEE Trans. Big Data **1**(4), 152–161 (2015)