Finding Suits in Images of People

Lei Huang[1,2], Tian Xia[1], Yongdong Zhang[1], and Shouxun Lin[1]

[1] Advanced Computing Research Laboratory, Institute of Computing Technology, Chinese Academy of Sciences, Beijing 100190, China
[2] Graduate University of Chinese Academy of Sciences, Beijing 100049, China
{huanglei,txia,zhyd,sxlin}@ict.ac.cn

Abstract. Clothing style is a salient feature for understanding images of people. To automatically identify the style of clothing that people wear is a challenging task. Suit as one of the clothing style is a key element in many important activities. In this paper, we propose a novel suits detection method. By analyzing the style of clothing, we propose the color features, shape features and statistical features for suits detection. Experiments with five popular classifiers have been conducted to demonstrate that the proposed features are effective and robust. Comparative experiments with Bag of Words (BoW) method demonstrate that the proposed features are superior to BoW which is a popular method for object detection. The proposed method has achieved promising performance over our dataset, which is a challenging web image set with various styles of clothing.

Keywords: Suits Detection, Photo Ranking, People Search, Clothing Style.

1 Introduction

People always wear different styles of clothing when participate in different activities. As shown in Fig.1, when people attend a meeting they always wear suits and when people do sports they always wear sportswear. Therefore, detecting the style of clothing is very beneficial to understand the content of images with people. Meanwhile, clothing provides significant information in people search in consumer photo albums and surveillance videos, e.g., Daniel et al. [1] used the color feature of clothing to perform attribute-based people search in surveillance environments. Besides the color feature, the style is also an important factor to reflect the characteristic of clothing. As shown in Fig. 2, in consumer photo selection systems, finding photos of people wearing suits from an increasing amount of personal photos is also very useful.

Many researchers have noted that clothing feature is important in two computer vision tasks [2-7], including human detection [2] and human recognition [3-7]. Sprague et al. [2] used the segmentation results of clothing to detect human in still images. In [3-7], researchers took the clothing feature as context information to aid human recognition. Song et al. [3] used the trained code-words to represent the clothing region, and this is an effective method concerning clothing's different types of color and texture. Gallagher et al. [4, 5] represented the clothing via three color features and two texture features. For the color features, they used the values of luminance-chrominance

K. Schoeffmann et al. (Eds.): MMM 2012, LNCS 7131, pp. 485–494, 2012.

(a) Meeting (b) Playing Football & Baseball (c) Cooking

Fig. 1. Clothing style is a salient feature to understand images of people. (a) When people wear suits, they may be attending a meeting. (b) When people wear sportswear, they may be playing. (c) When people wear chef apparel, they may be cooking.

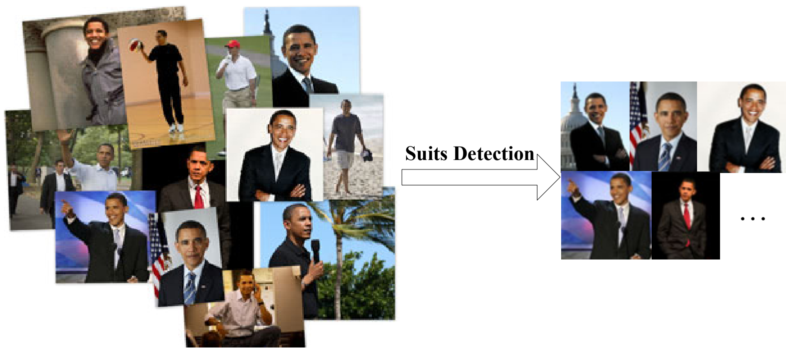


Fig. 2. Finding Obama in suits through suits detection, which is useful in consumer photo selection systems and attribute-based portrait search

space (LCC) and for the texture features, they used the responses to a horizontal and vertical edge detector. Zhang et al. [6] exploited the spatial relationship of features. They used the color signature, color pyramid and texture pyramid to represent the clothing. Khoury et al. [7] employed 3D histogram of the dominant color for clothing matching. In addition, there are many other studies on clothing analysis. Chen et al. [8] proposed an And-Or graph to represent the wide diversity of cloth configurations for clothing sketching. Tian et al. [9] proposed a clothing matching method to help the blind or color blind people.

All the above researches concentrated on extracting clothing features for clothing matching, using the segmentation of clothing for human detection and sketching. There are few researches focusing on recognizing style of clothing automatically.

There are various styles of clothing, e.g. suit, t-shirt and skirt. In this paper, we focus on finding suits in images of people for that suits have the uniform appearance and are generally a key element in many important activities. However, the features proposed in [3-7] [9] cannot be used in suits detection for the colors and textures are different in suits. Furthermore, there are some popular local features which have received a lot of research attention in recent years [3]. These features have been successfully used in many applications. However, the detectors of most of the local features are based on local extreme which cannot work if the clothing regions don't have textures (e.g., a suit always has a single color). Thus some more effective features are desired. The overview of our approach is shown in Fig. 3. To the best of our knowledge, we are the first one exploring visual features to find suits in images of people.

The contribution of the paper is that, by analyzing lots of suits and non-suits images, we discover several regular patterns for suits and propose three color features, one shape feature and two statistical features for suits detection (section 2). The proposed features will also be helpful for other clothing style detection (section 4).

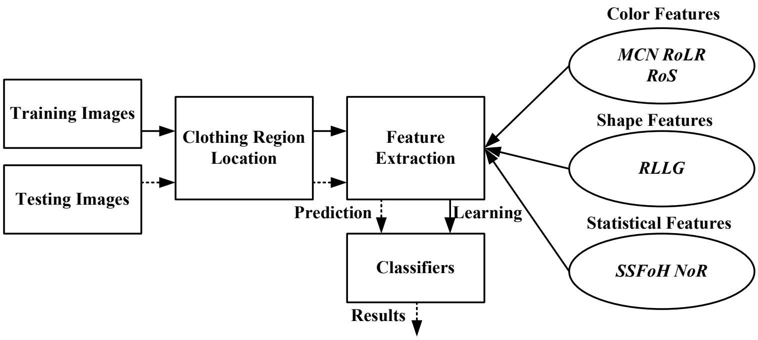


Fig. 3. Overview of the proposed method

2 Visual Features for Suits Detection

By analyzing the appearance of clothing, color features, shape features and statistical features are proposed for detecting photos of people wearing suits. The details of these features are described in this section.

2.1 Clothing Region Location

Before extracting clothing features, the clothing region should be located. A modified version of [10] which can handle rotated faces is performed first to get the face regions. And the clothing regions are acquired in rectangular regions below the face [3]. All images are normalized by the face size. In our experiments, the face region is normalized to 100*80 pixels. The clothing regions are located 10 pixels below the faces. And the size of clothing regions is 200*180. Fig.4 shows the results of the clothing region location. All the features presented below are computed in the clothing region CR . Since all features are computed in CR , our method focuses on finding photos of frontal view people wearing suits.

2.2 Color Features

Color is an important factor of clothing. Since the colors are different for different suits, conventional color features (e.g., values of luminance-chrominance space [4], color pyramid [6]) are not appropriate for suits detection, some novel color-related features should be explored. By observing lots of clothing images, we find that there are some

Fig. 4. Results of clothing region location. Face regions are shown in red rectangles and clothing regions are shown in blue rectangles.

rules of suits on the aspect of color: (1) Suits usually have no more than 3 main colors. (2) In the case of frontal view portrait, the distribution of the main color in the left region and right region of the suits is symmetric. (3) When people wear suits, there will be small skin region exposed in the suits location. Thus we define three features: Main Color Number (MCN), Ratio of Left-Right ($RoLR$) and Ratio of Skin(RoS).

To solve the problem caused by the lighting changes and self-shadow we use the color quantization method proposed in [9]. The clothing region CR_{rgb} is quantized to 11 colors image CR_{11} in HSI color space. First, “white”, “black” and “gray” are defined based on saturation S and luminance I. Then, “red”, “orange”, “yellow”, “green”, “cyan”, “blue”, “purple” and “pink” are defined based on the hue information. The results of color quantization are shown in Fig.5. After color quantization, MCN, $RoLR$ and RoS are defined as follows:

MCN : MCN is the number of colors which occupy larger than 10% of clothing region CR_{11}.

$RoLR$: $RoLR$ is the ratio of $ratio\_left$ to $ratio\_right$ which is computed by Eq. (1).

$$RoLR = \frac{ratio\_left}{ratio\_right} , \qquad (1)$$

where $ratio\_left = \frac{\#MC\_LCR}{\#LCR}$, $ratio\_right = \frac{\#MC\_RCR}{\#RCR}$, $\#MC\_LCR$ is the area of the main color regions in left clothing region, $\#LCR$ is the area of left clothing region. $\#MC\_RCR$ is the area of the main color regions in right clothing region. $\#RCR$ is the area of right clothing region. The left clothing region (LCR), middle clothing region (MCR) and right clothing region (RCR) are defined in Fig.5.

This feature also can get rid of profile-clothing images, e.g. Fig.5 (g).

RoS : RoS is the ratio of the skin area in the clothing region. Skin detection [11] is performed on CR_{rgb} to get the skin mask CR_{skin} . Morphology processing is

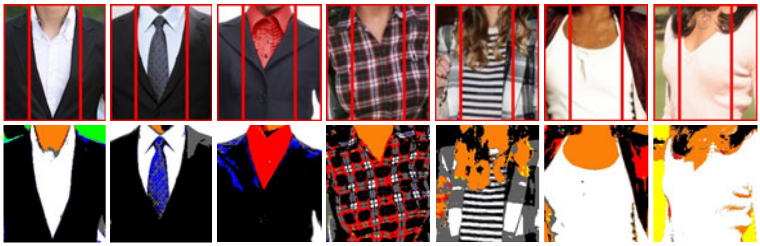


Fig. 5. Three portions of clothing region (top row) and the results of color quantization (bottom row). (a)-(c) Suits. (d)-(g) Non-suits. The left region and right region occupy 0.25 the width of the clothing region each and the middle region occupy 0.5 the width of the clothing region.

implemented on CR_{skin} to remove small regions. The results of skin detection are shown in Fig.6. RoS is computed as follows:

$$RoS = \frac{\#skin}{\#CR} \quad , \tag{2}$$

where $\#skin$ is the area of skin region in the skin mask CR_{skin}, and $\#CR$ is the area of clothing region.

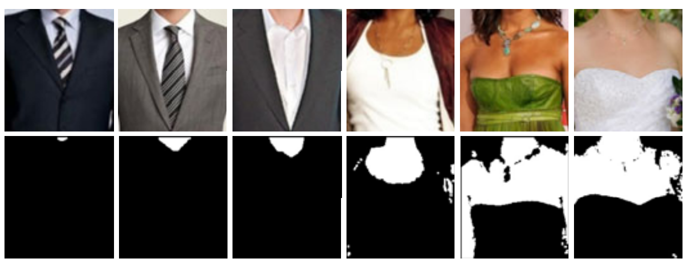


Fig. 6. Results of skin detection in clothing regions

2.3 Shape Features

There are some regular shape patterns of the suits. The most significant pattern is the line feature as shown in Fig.7. The lines at the middle of the clothing region form an inverted triangle. Meanwhile, the two lines are not intersecting in some cases as shown in the fourth column of Fig.7. Therefore, we use the ratio of the longest lines to the clothing region girth ($RLLG$) to represent this feature. $RLLG$ can be obtained by Eq.(3).

$$RLLG = \frac{L_{\max}}{Girth_{CR}} , \tag{3}$$

where $L_{\max} = \max_{l \in LineSet} |l|$, $Girth_{CR}$ indicates the girth of the clothing region CR, $|l|$ means the length of l. $LineSet$ denotes a line set acquired as follows: First, the segmentation method JSEG [12] is implemented to get the edges in clothing region.

JSEG is more robust to clothing images than canny edge detection which will cause more false edges induced by self-shadow and self-fold of the clothing. Then, the Hough line detection [13] is carried out on the edge set acquired by JSEG to get the candidate line set. There are still some useless lines in the candidate line set, such as too short lines, horizontal lines, lines induced by the self-fold and self-shadows of the clothing and lines induced by the boundary of the skin region. Therefore, the line filter algorithm is carried out. Following lines are filtered: with length shorter than one-fifth of the face width, with slope smaller than 0.7, with same color on both sides, with too much skin region on one side. The results of $LineSet$ are shown in Fig.7.

(a)

(b)

(c)

Fig. 7. Results of line detection in clothing regions. (a) Results of JSEG; (b) Results of Hough line detection; and (c) Results of our line filter algorithm.

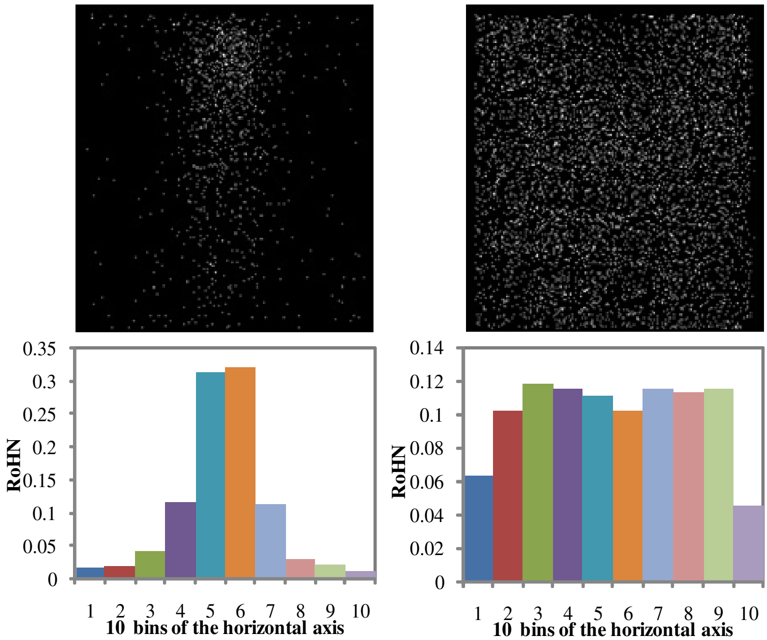
2.4 Statistical Features

There are some salient statistical features in suits. We compute a map which comes from the statistics of the location distribution of the Harris corners in the clothing regions. The map is shown in Fig.8. We use 100 suits images and 100 non-suits images (training data set in section 3.1) to get the map. For the suits images we can see, the corners mainly concentrate in the middle of the clothing regions. For non-suits images, the corner distribution is uniform. Therefore we define the spatial statistical feature of Harris Corner ($SSFoH$) to represent this regular pattern. Clothing regions are divided into three regions as shown in Fig.5. $SSFoH$ can be calculated by Eq.(4).

$$SSFoH = \frac{HNoM}{HNoLR} \quad , \tag{4}$$

where $HNoM$ is the number of Harris Corners in the middle clothing region, $HNoLR$ is the number of Harris Corners in the left and right clothing region.

Another significant statistical feature of suits is the blob statistical features. Suits always have simple texture and a small number of blobs (connected regions), e.g. Fig.5 (a)-(c). But for some non-suits clothing, the texture is complex and the number of connected regions is large, e.g. Fig.5 (d)-(e). Thus, we define Number of Connected Regions (NoR). NoR can be obtained on the quantized image CR_{11}. Morphological processing is processed on CR_{11} first to remove small regions.



(Suits) (Non-Suits)

Fig. 8. Statistical results of Harris corner location for suits and non-suits. The point in the images at top row means the location of Harris corners in clothing region and the brightness of the point indicates the corner number, the whiter the more. The bar charts are the statistical results projecting to horizontal axis. The horizontal axis is divided into 10 bins, each bar indicates the ratio of the corner points falling in the bin to the total corner number (RoHN).

3 Data Set and Experiments

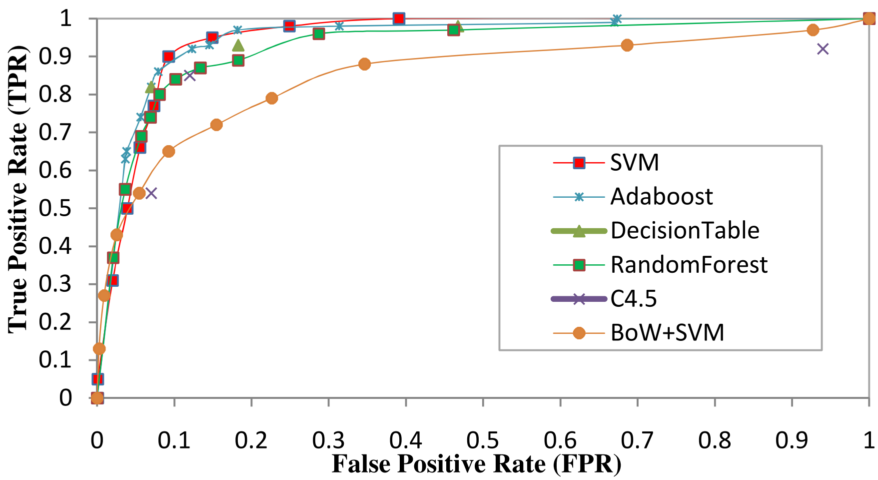
3.1 Data Set

A consumer image collection of 1500 images of people is constructed to evaluate the proposed features. All the images come from "flickr", "baidu" and "google". We use the keywords like *"portrait"*, *"portrait+suits"*, *"people+suits"* to search these images. The image collection contains 200 suits images and 1300 non-suits images. 100 suits and 200 non-suits are used for training, other 100 suits and 1100 non-suits are used for testing. Some samples of our dataset are shown in Fig.9.

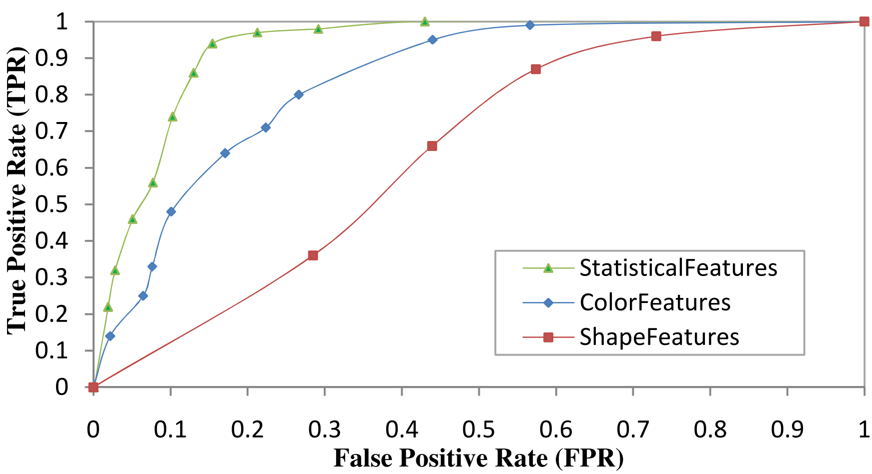
3.2 Experimental Results

To demonstrate the effectiveness and robustness of the features, we use five popular classifiers including SVM [14], Random Forest [15], Adaboost [16], Decision Table [17] and C4.5 [18]. True Positive Rate (TPR) and False Positive Rate (FPR) are used

Fig. 9. Some samples in our dataset. The first line is suits images and the second line is non-suits images. Most of the suits images are frontal. The non-suits images contain various style of clothing, e.g. skirt, t-shirt, shirt and sweater.



(a)



(b)

Fig. 10. Performance comparison. (a) Performance of five classification and BoW method. (b) Performance of each type features based on SVM.

for evaluation. The results are shown in Fig.10(a). We can see that all the five classifiers get TPR 80% under FPR less than 10%. This shows the effectiveness and robustness of our features. The SVM method gets the best result with TPR of 90% under FPR of 9.2%. To illustrate that our features are superior to state of the art, comparison to Bag of Words (BoW) [19] is performed. In the BoW method, we use harris corner detector since we have normalized the images based on face regions. The descriptor we use is SIFT descriptor [20]. SVM is used as the classifier. The performance of BoW method is shown in Fig.10(a): BoW+SVM. We can see that our features outperform BoW method.

To further evaluate the contribution of each type of features, comparative experiments are performed. SVM which gets the best performance in the above experiment is used as the classifier. The results are shown in Fig.10(b). From the results we can see that every type of our features is effective. In low True Positive Rate, False Positive Rate of all three types of features is under 20%. The statistical features are the most discriminative of all the features. The shape features give the minimum contribution which doesn't meet our intuition. By analyzing this issue, we find that this is caused by the non-rigid deformation of clothing which reduces the performance of Hough line detection.

Some images are falsely detected by our method. As for the suits images that are not detected by our method, most of them are profile-suits images or suits images with occlusion. As for the non-suits images incorrectly detected as suits images, most of them have similar appearance with suits.

4 Conclusions and Future Work

Clothing style is an important feature to understand images of people including what he/she is doing and what is his/her profession etc. Actually, recognizing the style of clothing is a very hard task due to two reasons: (1) no-rigid deformation and (2) various styles of clothing. In this paper, we proposed a novel method to detect suits in images of people. This method is useful in photo selection systems and attribute-based people search task. We proposed six novel features including color features, shape features and statistical features for this task. The experimental results show the effectiveness of the proposed features.

Also these features will be helpful for other clothing style detection. For example, skin feature (one of our color features) is useful in t-shirt detection. The limitation of our method is that the results for suits images with occlusion and profile-suits images are not satisfactory. In the future, we will find ways to solve these problems and expand our work to other clothing style detection. We will also investigate on how to further combine the multiple features in a multiple graph framework [21][22] to learn a better detector for a particular style of clothing.

Acknowledgments. This work is supported by the National High Technology and Research Development Program of China (863 Program, 2009AA01A403); National Nature Science Foundation of China (61100087); Beijing Natural Science Foundation (4112055); Beijing New Star Project on Science & Technology (2007B071); Co-building Program of Beijing Municipal Education Commission.

References

1. Daniel, A.V., Rogerio, S.F., Duan, T., Lisa, B., Arun, H., Matthew, T.: Attribute-Based people search in surveillance Environments. In: WACV, pp. 1–8 (2009)
2. Sprague, N., Luo, J.: Clothed people detection in still images. In: ICPR, pp. 585–589 (2002)
3. Song, Y., Leung, T.: Context-Aided Human Recognition – Clustering. In: Leonardis, A., Bischof, H., Pinz, A. (eds.) ECCV 2006, Part III. LNCS, vol. 3953, pp. 382–395. Springer, Heidelberg (2006)
4. Gallagher, A.C., Chen, T.: Clothing cosegmentation for recognizing people. In: CVPR, pp.1–8 (2008)
5. Gallagher, A.C., Chen, T.: Using context to recognize people in consumer images. IPSJ Transactions on Computer Vision and Applications 1(5), 115–126 (2009)
6. Zhang, W., Zhang, T., Tretter, D.: Clothing-based person clustering in family photos. In: ICIP, pp. 4593–4595 (2010)
7. Khoury, E.E., Senac, C., Joly, P.: Face-and-Clothing based people clustering in video content. In: MIR, pp. 295–304 (2010)
8. Chen, H.Z., Xu, J., Liu, Z.Q., Zhu, S.C.: Composite templates for cloth modeling and sketching. In: CVPR, pp. 943–950 (2006)
9. Tian, Y., Yuan, S.: Clothes Matching for Blind and Color Blind People. In: Miesenberger, K., Klaus, J., Zagler, W., Karshmer, A. (eds.) ICCHP 2010, Part II. LNCS, vol. 6180, pp. 324–331. Springer, Heidelberg (2010)
10. Viola, P., Jones, M.: Rapid object detection using a boosted cascade of simple features. In: CVPR, pp. 511–518 (2001)
11. Jones, M.J., Rehg, J.M.: Statistical color models with application to skin detection. International Journal of Computer Vision 46(1), 81–96 (2002)
12. Deng, Y., Manjunath, B.S.: Unsupervised segmentation of color-texture regions in images and video. IEEE Transactions on Pattern Analysis and Machine Intelligence, 800–810 (2001)
13. Duan, R.O., Hart, P.E.: Use of the hough transformation to detect lines and curves in pictures. In: Communications of the Association for Computing Machinery, pp. 11–15 (1972)
14. Chang, C.C., Lin, C.J.: LIBSVM: a library for support vector machines. ACM Transactions on Intelligent Systems and Technology 2, 27:1–27:27 (2011)
15. Breiman, L.: Random forests. Machine Learning 45(1), 5–32 (2001)
16. Freund, Y., Schapire, R.E.: Experiments with a new boosting algorithm. In: Proceedings of the Thirteenth International Conference on Machine Learning, pp. 148–156 (1996)
17. Kohavi, R.: The Power of Decision Tables. In: Lavrač, N., Wrobel, S. (eds.) ECML 1995. LNCS (LNAI), vol. 912, pp. 174–189. Springer, Heidelberg (1995)
18. Quinlan, J.R.: Induction of decision trees. Machine Learning 1(1), 81–106 (1986)
19. Sivic, J., Zisserman, A.: Video Google: a text retrieval approach to object matching in videos. In: ICCV, pp. 1470–1478 (2003)
20. Lowe, D.G.: Distinctive image features from scale-invariant keypoints. International Journal of Computer Vision 60(2), 91–110 (2004)
21. Wang, M., Hua, X.S., Hong, R.C., Tang, J.H., Qi, G.J., Song, Y.: Unified Video Annotation via Multi-Graph Learning. IEEE Trans. on Circuits and Systems for Video Technology 19(5) (2009)
22. Xia, T., Tao, D.C., Mei, T., Zhang, Y.D.: MultiView Spectral Embedding. IEEE Trans. on Systems, Man, and Cybernetics: Part B 40(6), 1438–1446 (2010)