Adult Image Detection Combining BoVW Based on Region of Interest and Color Moments

Liu Yizhi[1,2,3], Lin Shouxun[1], Tang Sheng[1], Zhang Yongdong[1]

[1] Laboratory of Advanced Computing Research, Institute of Computing Technology, Chinese Academy of Sciences, Beijing 100190, China
[2] Graduate University of the Chinese Academy of Sciences, Beijing 100039, China
[3] Institute of Computer Science and Engineering, Hunan University of Science and Technology, Xiangtan 411201, China
{ liuyizhi, sxlin, ts, zhyd } @ ict.ac.cn

Abstract. To prevent pornography from spreading on the Internet effectively, we propose a novel method of adult image detection which combines bag-of-visual-words (BoVW) based on region of interest (ROI) and color moments (CM). The goal of BoVW is to automatically mine the local patterns of adult contents, called visual words. The usual BoVW method clusters visual words from the patches in the whole image and adopts the weighting schemes of hard assignment. However, there are many background noises in the whole image and soft-weighting scheme is better than hard assignment. Therefore, we propose the method of BoVW based on ROI, which includes two perspectives. Firstly, we propose to create visual words in ROI for adult image detection. The representative power of visual words can be improved because the patches in ROI are more indicative to adult contents than those in the whole image. Secondly, soft-weighting scheme is adopted to detect adult images. Moreover, CM is selected by evaluating some commonly-used global features to be combined with BoVW based on ROI. The experiments and the comparison with the state-of-the-art methods show that our method is able to remarkably improve the performance of adult image detection.

Keywords: Adult image detection, bag-of-visual-words (BoVW), region of interest (ROI), soft-weighting, color moments

1 Introduction

With the rapid penetration of the Internet into every part of our daily life, it is crucial to protect people, especially children, from exposure to objectionable information. Content-based adult image detection is one of the most powerful approaches of filtering pornography. It is mostly based on global features, whereas the false positive rate (FPR) is higher than people's expectation.

Bag-of-visual-words (BoVW) based adult image detection [1, 2] has been applied to reduce FPR because it is robust to within-class variation, occlusion, background clutter, pose and lighting changes. Its goal is to automatically mine the local patterns

of adult content, such as pornographic parts or poses, by certain clustering algorithm. These patterns are described as visual words. Therefore, visual words are the kernel of BoVW.

The usual BoVW method clusters visual words from the patches in the whole image and adopts the weighting schemes of hard assignment. Another name of patches is keypoints in some literatures [3, 4]. Hard assignment means hardly assigning the patches to the nearest visual words while generating the BoVW histogram for training or testing. However, there are many background noises in the whole image. And soft-weighting scheme is better than hard assignment.

Aiming at detecting adult images accurately, we propose a novel method of adult image detection which combines BoVW based on region of interest (ROI) and color moments (CM). There are two differences between the method of BoVW based on ROI and the usual BoVW method. Firstly, we propose to create visual words in ROI for adult image detection. The patches in ROI are more indicative to adult contents than those in the whole image. So it can improve the representative power of visual words for adult image detection. Secondly, soft-weighting scheme is adopted to improve the performance of BoVW further. Soft-weighting scheme, recently proposed by Jiang et al. [3], is better than hard assignment on both PASCAL-2005 and TRECVID-2006 datasets because of assigning a patch to $top-N$ nearest neighbors. Moreover, CM is selected by evaluating some commonly-used global features. The experiments and the comparison with the state-of-the-art methods show that our method is able to remarkably improve the performance of adult image detection.

The remainder of the paper is organized as follows: section 2 introduces related works, section 3 illustrates our method in detail, section 4 shows the experiments and section 5 concludes the paper.

2 Related Works

The traditional approach of content-based adult image detection is based on the global low-level features which include color, shape, texture and etc. Forsyth et al. [5] construct a human figure grouper after detecting skin regions, but consuming too much time and low detection accuracy are the two shortcomings. Zeng et al. [6] implement the image guarder system to detect pornographic images by different kinds of global features. Kuan and Hsieh [7] use image retrieval technique and extract visual features from skin regions. Q. F. Zheng et al. [8] use an edge-based Zernike moment method to detect harmful symbol objects. Rowley et al. [9] adopt 27 visual features for Google to filter adult-content images, including color, shape, texture, face and etc. Tang et al. [10] employ latent Dirichlet allocation to cluster images into some subsets and then combine SVM classifiers on one subset to rate adult images. Nevertheless, it is difficult to detect adult images in the presence of within-class variation, occlusion, pose and lighting changes.

To cope with the difficulty, BoVW approaches have been applied. Wang et al. [1] explore an algorithm to reduce the number of visual words and integrate it with spatial distribution to detect adult images. Deselaers et al. [2] combine BoVW with

color histogram to classify images into different categories of adult content. Both of them use difference of Gaussian (DoG) detector and scale-invariant feature transform (SIFT) descriptor.

Visual words are usually clustered from the patches in the whole image. The patches are always assigned to the nearest visual words, as it called hard assignment. We name the preceding procedures "the usual BoVW method". However, visual words are clustered from the patches in the whole image and these patches are full of background noises. Furthermore, it has not been discussed in-depth that the effects of weighting schemes and combination with global features on adult image detection.

3 Our Method

To detect adult images accurately, we combine BoVW based on ROI with color moments (CM). In this section, we will illustrate our method in detail.

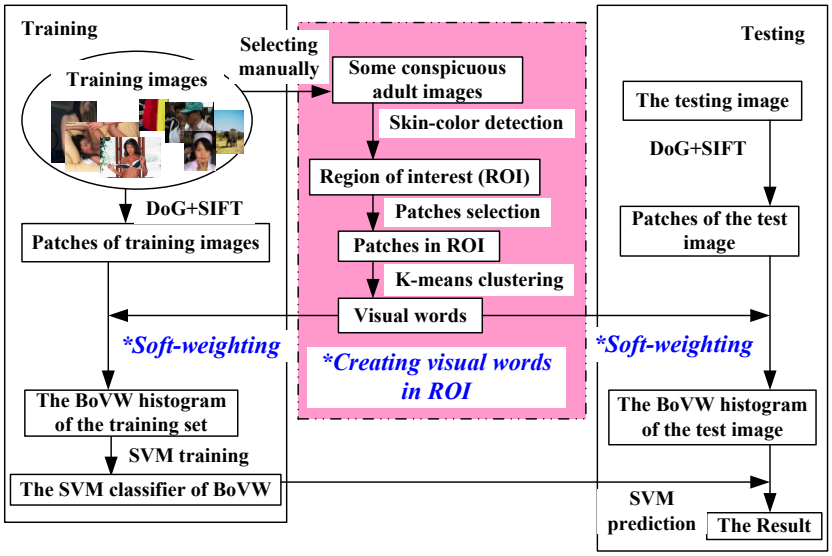

Fig. 1. BoVW based on ROI

3.1 BoVW Based on ROI

As shown in Fig. 1, we propose the method of BoVW based on ROI to improve the performance. We use DoG detector and SIFT descriptor, too. Two differences between BoVW based on ROI and the usual BoVW method are marked with the blue and boldfaced words and with a bar nearby. Firstly, visual words are created in ROI. Secondly, soft-weighting scheme is adopted. The training and testing procedures are

depicted respectively on the left and the right column in Fig. 1. The details of ROI detection and soft-weighting scheme are introduced in the subsections respectively.

3.1.1 ROI Detection

ROI means the subareas containing pornographic parts or poses in the field of adult image detection. Skin-color regions include these pornographic subareas. Thus, we apply the kind of skin-color models [11] to capture ROI.

In Fig. 2, some examples of SIFT patches in ROI are given. To avoid the objectionable information, we transform the skin-color region into white and other regions into black. SIFT patches are represented by the red points in the images. We can observe clearly that many background noises are removed after selecting patches from ROI.

(a) SIFT patches in the whole image

(b) SIFT patches in ROI

Fig. 2 Some examples of SIFT patches in ROI

Garcia and Tziritas [11] have shown that skin-like pixels are more correlated with C_r and C_b components than Y component. Thus, the input image is transformed from the RGB color model to the YC_bC_r color space. A pixel is considered to be skin-like if its C_r and C_b components meet the following constraints:

$$C_r = \max\{-2(C_b + 24), -(C_b + 17), -4(C_b + 32), 2.5(C_b + \theta_1), \theta_3, 0.5(\theta_4 - C_b)\}; \quad (1)$$

And

$$C_r = \min\{(220 - C_b)/6, 4(\theta_2 - C_b)/3\}. \quad (2)$$

There are two constraints for θ_1, θ_2, θ_3, and θ_4:

If $Y > 128$, then

$$\begin{aligned} \theta_1 &= -2 + (256 - Y)/16; \\ \theta_2 &= 20 - (256 - Y)/16; \\ \theta_3 &= 6; \\ \theta_4 &= -8. \end{aligned} \quad (3)$$

Otherwise,

$$\begin{aligned}\theta_1 &= 6;\\ \theta_2 &= 12;\\ \theta_3 &= 2 + Y/32;\\ \theta_4 &= -16 + Y/16.\end{aligned} \tag{4}$$

3.1.2 Weighting Scheme

Weighting schemes play an important role in generating the BoVW histogram and thus have great effects on the performance of BoVW. Generally speaking, the patches are hardly assigned to the nearest visual words. These schemes are named hard assignment, such as binary weighting, term frequency (TF) and the product of term frequency and inverse document frequency (TF×IDF). Among them, binary weighting always produces top or close-to-top performance on the datasets of PASCAL-2005 and TRECVID-2005 [4]. Binary weighting indicates the presence and absence of a visual word with values 1 and 0 respectively.

Soft-weighting scheme, recently proposed by Jiang et al. [3], outperforms the preceding schemes of hard assignment on both PASCAL-2005 and TRECVID-2006 datasets. Instead of assigning directly a patch to its nearest neighbor, soft-weighting assigns a patch to $top-N$ ($N=4$ empirically) nearest neighbors to weight the significance of visual words. Suppose that there are M visual-words in a vocabulary and the $top-N$ nearest visual-words are selected for each patch in an image, each component v_k of a M -dimensional vector $V=[v_1, v_2, \mathrm{L}, v_k, \mathrm{L}, v_M]$ represents the weight of a visual word k in an image such that

$$t_k = \sum_{i=1}^{N} \sum_{j=1}^{L_i} \frac{1}{2^{i-1}} sim(j,k) \tag{5}$$

where L_i represents the number of patches whose i th nearest neighbor is visual word k . The nearest $sim(j,k)$ means the similarity between the patch j and the visual word k .

Weighting schemes are also relative to the number of visual words. Jiang et al. conclude that the performances of BoVW are similar using soft-weighting on TRECVID-2006 when the number of visual words ranges from 500 to 10,000 [3]. If the number of visual words becomes large, the cost increases in clustering, computing the BoVW histogram and running the classifier.

Therefore, we adopt soft-weighting as the weighting scheme and employ the K-means algorithm based on DBSCAN [12] to create visual words whose number is around 500. The clustering algorithm has some advantage, such as rapidness, robustness to noises, finding any shape of clusters in spaces, and adjusting clusters adaptively.

3. 2 Classification and Combination

To integrate the advantage of global features, we combine BoVW based on ROI with color moments (CM). After evaluating some global color features — CM, color correlogram and color histogram, we find that CM is the best one. Then we concatenate it with some global texture features, such as texture co-occurrence, Haar wavelet and edge histogram. But the performances of its concatenation are not as good as CM alone. CM provides a measurement for color similarity between images. In CM, we calculate the first 3 moments of 3 channels in Lab color space over 5×5 grid partitions, and aggregate the features into a 225-dimension feature vector.

The supported vector machines (SVM) classifier has been one of the most popular classifiers for BoVW-based image classification [2, 3, 4] and adult image detection based on global features [7, 9, 10, 11]. We employ SVM with Gaussian radial basis function (RBF) kernel to obtain good performance. The form is as follows:

$$g(x) = \sum_i \alpha_i y_i k(x_i, x) - b = \sum_i \alpha_i y_i e^{-\rho d(x_i, x)} - b = \sum_i \alpha_i y_i e^{-\rho \sum_i |x_i - x|^2} - b \qquad (6)$$

In the formula (6): $k(x_i - x)$ is the response of a kernel function for the training sample x_i and the test sample x, which measures the similarity between the two data samples; y_i is the class label of x_i; α_i is the learned weight of the training sample x_i, and b is a learned threshold parameter.

We combine the classification results of BoVW based on ROI and CM with "average fusion". Average fusion is one of the commonly-used "late fusion" methods.

4 Experiments

In this section, our method is evaluated step by step. (1) Our dataset and the baseline are reported. (2) We do experiments to show the effect of soft-weighting scheme on BoVW in subsection 4.2. (3) The improvement of creating visual words in ROI is evaluated in subsection 4.3. (4) Subsection 4.4 compares some commonly-used global features. (5) In subsection 4.5, the combination of BoVW based on ROI and color moments is compared with the baseline and many previous works.

4.1 Our Dataset and The Baseline

As Table 1 shows, we provide statistics of our dataset. We collect 90,000 images from Internet. The training set is made up of 10,000 adult images and 40,000 non-adult images. The testing set has 10,000 adult images and 30,000 non-adult images. Both sets include 10,000 non-adult images containing body parts, such as faces, hands, feet and trunks. We do all these experiments in the visual studio 2003 environment with the machine of 1.86 GHz Duo CPU and 2GB memory. We evaluate our method with receiver operating characteristic (ROC) curves. A ROC space is defined by false

positive rate (FPR) and true positive rate (TPR) as x and y axes respectively. There is no common dataset in the field of adult image detection. Consequently, we build up the baseline named CH+EH. The baseline uses the features concatenating color histogram (CH) and edge histogram (EH) and is classified by the SVM classifier with the Gaussian RBF kernel.

Table 1. Statistics of our dataset

	The training set	The testing set
Adult images	10,000	10,000
Non-adult images	40,000	30,000

4.2 The Usual BoVW Method with Different Weighting Schemes

In this subsection, we estimate the effect of soft-weighting scheme on BoVW. ROC curves of BoVW with binary weighting and soft-weighting are respectively abbreviated to SIFT-BW and SIFT-SW. The numbers in the parentheses are the size of visual words. According to Fig. 3, soft-weighting scheme is a little better than binary weighting.

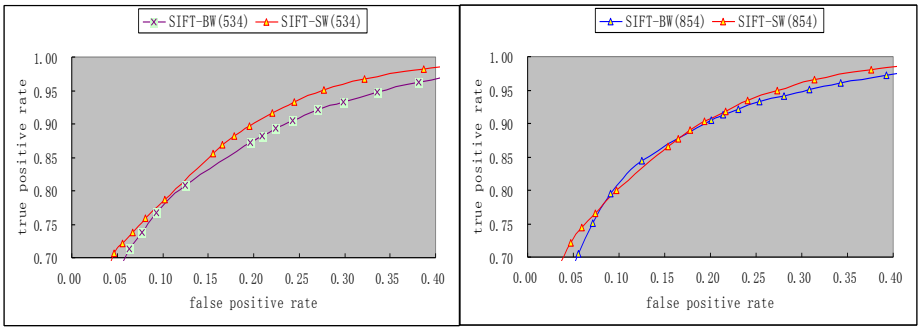

Fig. 3 The ROC curves of the usual BoVW method

4.3 Evaluation of BoVW Based on ROI

We evaluate the advantage of creating visual words in ROI. SIFT-ROI-SW is on behalf of the ROC curve of BoVW based on ROI. All the curves in Fig. 4 adopt soft-weighting scheme. So we can conclude that the method of BoVW based on ROI can remarkably improve the performance of BoVW based adult image detection.

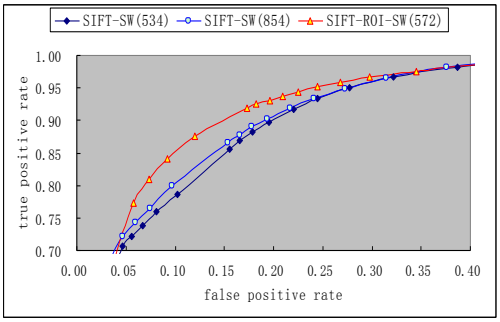

Fig. 4 The performance of BoVW based on ROI

4.4 The Performance of Color Moments

After evaluating some commonly-used global color features — color moments (CM), color correlogram (CC) and color histogram (CH), we find that CM is the best one. Then we concatenate it with some global texture features, such as texture co-occurrence (TC), Haar wavelet (HW) and edge histogram (EH). We can infer from Fig. 5 that the performances of the concatenations of CM are worse than CM alone.

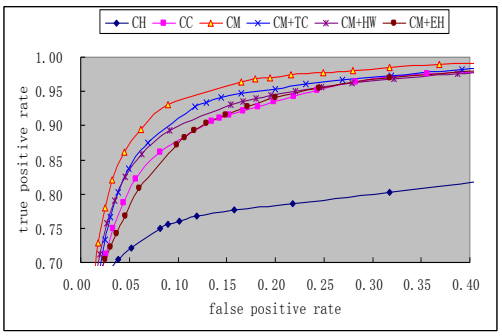

Fig. 5 Adult image detection based on the global features

4.5 Evaluation of Our Method

To detect adult images accurately, we use “average fusion” to combine BoVW based on ROI and color moments. As shown in Fig. 6, our method outperforms the baseline and the state-of-the-art method (Deselaers et al. [2]) on the same dataset. [2] combines BoVW with color histogram to classify images into different categories of adult content. The ROC curves of our method, CH+EH and Deselaers et al. [2] are represented respectively as the red, blue and purple curve. The points in Fig. 6 are on

behalf of the performance of some previous works in section 2. The results experimentally show that our method is able to remarkably improve the performance and outperforms many previous works, including the state-of-the-art method (Deselaers et al. [2]).

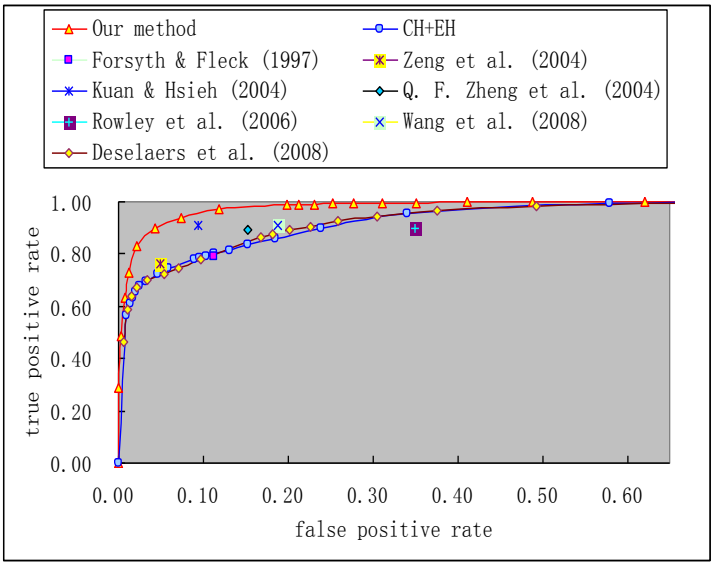

Fig. 6 Performances compared with many previous works.

5 Conclusions

To filter adult contents proliferating on the Internet effectively, we propose a novel method of adult image detection which combines BoVW based on ROI and CM. We create visual words in ROI and adopt soft-weighting scheme to improve the performance of BoVW. The patches in ROI are more indicative than those in the whole image. Furthermore, soft-weighting scheme is better than that of hard assignment and CM is selected by evaluating some commonly-used global features. The experiments and the comparison with the state-of-the-art methods show that our method is able to remarkably improve the performance.

Acknowledgments. This work is supported by the National Basic Research Program of China (973 Program, 2007CB311100); National High Technology and Research Development Program of China (863 Program, 2007AA01Z416); National Nature Science Foundation of China (60873165, 60802028); Beijing New Star Project on Science & Technology (2007B071); Co-building Program of Beijing Municipal Education Commission.

References

[1] Y. S. Wang, Y. N. Li, W. Gao, “Detecting pornographic images with visual words”, Transactions of Beijing Institute of Technology, vol. 28, pp. 410-413, May 2008 (in Chinese).
[2] T. Deselaers, L. Pimenidis, H. Ney, “Bag-of-visual-words models for adult image classification and filtering”, 19th International Conference on Pattern Recognition, Tampa, USA, pp. 1-4, 2008.
[3] Y. G. Jiang, C. W. Ngo, J. Yang, “Towards optimal bag-of-features for object categorization and semantic video retrieval”, 6th ACM International Conference on Image and Video Retrieval, Amsterdam, Netherlands, pp. 494-501, 2007.
[4] J. Yang, Y. G. Jiang, A. G. Hauptmann, et al., “Evaluating bag-of-visual-words representations in scene classification”, 9th ACM SIGMM International Workshop on Multimedia Information Retrieval, Augsburg, Germany, pp. 197-206, 2007.
[5] M. M. Fleck, D. A. Forsyth, C. Bregler, “Finding naked people”, 4th European Conference on Computer Vision, Cambridge, UK, pp. 593-602, 1996.
[6] W. Zeng, W. Gao, T. Zhang, et al., “Image guarder: an intelligent detector for adult”, 6th Asian Conference of Computer Vision, Jeju Island, Korea, pp. 198-203, 2004.
[7] Y. H. Kuan, C. H. Hsieh, “Content-based pornography image detection”, International Conference on Imaging Science, System and Technology, Las Vegas, USA, 2004.
[8] Q. F. Zheng, W. Zeng, W. Gao, et al., “Shape-based adult images detection”, 3th International Conference on Image and Graphics, Hong Kong, China, pp. 150-153, 2004.
[9] H. A. Rowley, J. Yushi, S. Baluja, “Large scale image-based adult-content filtering”, 1st International Conference on Computer Vision Theory and Applications, pp. 290-296, 2006.
[10] S. Tang, J. Li, Y. Zhang, et al., "PornProbe: an LDA-SVM based Pornography Detection System", ACM International Conference on Multimedia, Beijing, China, 2009.
[11] C. Garcia, G. Tziritas, “Face detection using quantized skin color regions merging and wavelet packet analysis”, IEEE Transactions on Multimedia, vol.1, pp. 264-277, 1999.
[12] M. Ester, H. P. Kriegel, J. Sander, et al., “A density based algorithm for discovering clusters in large spatial databases with noise”, 2nd International Conference on Knowledge Discovery and Data Mining, Portland, USA, pp. 226-231, 1996.