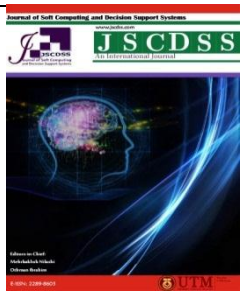
An International Journal
JSCDSS
http://www.jscdss.com
Vol.2 No.6 December 2015: 27-33
Article history:
Accepted 07 October 2015
Published online 12 October 2015

Journal of Soft Computing and Decision Support Systems



E-ISSN: 2289-8603

Fast and Effective Bag-of-Visual-Word Model to Pornographic Images Recognition Using the FREAK Descriptor

S.Hadi Yaghoubyan [a,*], Mohd Aizaini Maarof [a], Anazida, Zainal [a], Mohd Fo'ad Rohani [a], Mahdi Maktabdar Oghaz [a]
[a] Faculty of Computing, Universiti Technologi Malaysia, Johor, Malaysia

*** Corresponding author email address**: yaghoobian.h@gmail.com

Abstract

Recently, the Bag of Visual Word (BoVW) has gained enormous popularity between researchers to object recognition. Pornographic image recognition with respect to computational complexity, appropriate accuracy, and memory consumption is a major challenge in the applications with time constraints such as the internet pornography filtering. Most of the existing researches based on the Bow, using the very popular SIFT and SURF algorithms to description and match detected keypoints in the image. The main problem of these methods is high computational complexity due to constructing the high dimensional feature vectors. This research proposed a BoVW based model by adopting very fast and simple binary descriptor FREAK to speed-up pornographic recognition process. Meanwhile, the keypoints are detected in the ROI of images which improves the recognition speed due to eliminating many noise keypoints placed in the image background. Finally, in order to find the most representational visual-vocabulary, different vocabularies are generated from size 150 to 500 for BoVW. Compared with the similar works, the experimental results show that the proposed model has gained remarkable improvement in the terms of computational complexity.

Keywords: Bag of Visual-Words (BoVW), Pornographic image recognition, Fast descriptor, ROI selection

1. Introduction

In the modern world, the Internet plays an increasingly important role, not only at the level of infrastructure but also in business, culture and society. However, the spread of pornographic images and videos which are planned to absorb user's attention, unfortunately have great side effects to people's mental and physical health, especially for children and teenagers. Therefore, there is a crucial need for detect and filter pornographic images and videos in the Internet (Wang et al., 2012; Wang et al., 2012; Kherfi et al., 2004).

The content-based illicit image recognition has attracted much attention between researchers in the recent years. These techniques relay on extraction features such as color, skin color, texture, shape, face and the human body gesture. Based on the study in (Zhang et al., 2013) generally the existing methods can be grouped into four categories such as: (1) Body structure (2) Content Based Image retrieval (CIBR) (3) Features of skin color region (4) Bag of Visual Words (BoVW).

In the pornographic image recognition method based on body structure, which early proposed by Fleck and Forsyth (Forsyth and Fleck, 1996; Fleck et al., 1996), the skin color regions are, firstly, detected in the image. Then, the detected regions are fed to a predefined grouper, which using geometric constraints on human body structure attempts to group a human figure. Thus, the pornographic image can be recognized, if the grouper finds a predefined structure.

The CIBR technique for recognition is using the visual information by retrieving pools of digital images. The process of retrieval is performed by measuring the similarity between the image in the pre-classified database and query image through similarity measure. If the number of the matched illicit images from the database exceeds a predefined threshold, the query image will be recognized as pornographic (Herwindiati et al., 2010; Wang et al., 1998).

In the pornographic image detection based on skin features, firstly the features on skin color regions are extracted. The features can be texture, color, the area of skin region and etc. Then, the pattern classifier is utilized to classify target image as illicit or benign (Zheng et al., 2004).

It is not easy to find the effective features for the explained methods in the literature somehow distinguish the pornographic image from benign one. Recently the BOVW models, which are inspired by the bag-of-words

(BoW) models in text content analysis, have gained enormous popularity in object retrieval and natural scene analysis, due to its remarkable results and simplicity. This method narrowed the semantic gap between the visual features and image contents.

There is a clearly trade-off between accuracy and computational complexity in the recognition tasks. Since, the performance (processing speed) for real-time applications such as the Internet filtering is critical, the web content filtering methods should rely on accurate identification of pornographic images and videos in a very short time cycle. Most of the existing methods that rely on BoVW (Zhang et al., 2013; Steel, 2012; Deselaers et al., 2008; Lienhart, 2009; Lopes et al., 2009;Lopes and Avila, 2009; Yizhi et al., 2010), are using Scale-Invariant Feature Transform (SIFT) descriptor (Lowe, 2004), which will be discussed in the literature in section 2. Some other researches relay on BoVW, applying the Speeded Up Robust Features (SURF) (Bay et al., 2006) descriptor to identify the pornographic images (Liu and Xie, 2009; Lv et al., 2011). However, employing the SIFT and SURF descriptor in this experiment yields the higher accuracy, but it is very time consuming to process and recognition. The aim of this study is to develop a suitable model that can speed-up pornographic image recognition and present a faster model which is closed to real time recognition by decreasing the computational complexity in keypoint description stage. To cope with the issue, this research adapted Fast Retina Keypoint (FREAK) (Alahi et al., 2012) descriptor that is in general faster to compute keypoint descriptors with lower memory load. Furthermore, keypoints are detected in the region of interest (ROI) per image which eliminates a lots of noise keypoints placed in the image background. Through extensive experiments, it is shown that the proposed model is faster than SIFT-based and SURF-based BoVW representation models.

The rest of the paper is organized as follows. Section 2 presents a comprehensive literature of BoVW-based pornographic image detection models. The framework of the proposed model is illustrated in the section 3. In the section 4, the experimental evaluation of the proposed model is presented. Finally, conclusion is presented in the section 5 of this study.

2. Related Work

R Recent developments in image recognition methods have shown the Bag of Visual Word (BoVW) model, which also called the Bag of Features (BoF) has remarkable performance for numerous image recognition and classification tasks. As the preliminary works in the pornographic field, (Wang et al., 2008) employed the BoVW model to pornographic image recognition. This study includes four stages, descriptors extraction, creating visual-word, image representation and finally recognition. Deselaers et al. (2008) presented a pornographic image classification and filtering method based on BoVW, which is used the principal component analysis to reduce SIFT descriptor dimensions, and to generate visual vocabulary, employed unsupervised training of Gaussian Mixture Model (GMM). Based on SIFT descriptors, Lienhart et al. (2009) proposed a pornographic image recognition model using probabilistic Latent Semantic Analysis (pLSA). In (Lopes and Avila, 2009; Lopes et al., 2009), instead of well-known SIFT descriptor, the Hue-SIFT descriptor, which includes color information, is employed for feature extraction, and in order to obtain the most representative visual vocabulary, different groups of visual vocabularies size are tested. Liu et al. (2009) applied the Speed up Robust Features (SURF) descriptor instead of SIFT and combined with color moments. In (Yizhi et al., 2010), the SIFT patches are extracted in the region of interest (ROI) in order to improve representative power of visual words and then soft-weighting scheme is adopted to detect pornographic images. In (Lv et al., 2011), firstly by using the SURF algorithm, local patches are extracted, and then by fusing the context of visual-vocabulary and spatial-related high-level semantic features of pornographic images a high-level semantic dictionary is constructed. Since, detecting pornographic images in the real-time applications needs to be fast and effective, this research is using the FREAK descriptor for pornographic recognition which is faster than SIFT-based and SURF-based descriptors.

3. Methodology

In the text content categorization called Bag of Word (BoW), using an unsorted set of the contained words, a document could be represented. Analogously, for visual contents, an image is represented by an unsorted set of discrete visual-words called BoVW representation model. In this model, the preliminary step is detecting image keypoints or local interest points using different detectors (Gauglitz et al., 2011). Keypoints are the salient image patches which contain rich local information about an image. The next step, is mapping these patches to the descriptor space and creating the feature vector using various descriptors. In order to carry out a comprehensive comparison, the most popular descriptors such as SIFT and SURF descriptors versus the fast descriptor FREAK are utilized in this research. In the third step, the visual-word vocabulary will be generated using clustering algorithms such as k-means clustering. A visual-word can be considered as a representative of several similar patches in each image. As each cluster center consider as a visual-word in the vocabulary, the parameter “k” determines the size of the visual-word vocabulary. The vocabulary size has enormous effect on the classification accuracy result (Jiang et al., 2007; O’Hara and Draper, 2011). By using various “k” measures, this research is aimed to provide some practical insights on finding the best vocabulary size. Finally, with mapped keypoints into visual words, an image can be represented as a Bag of Visual Words. A normalized histogram then will be created for each image by counting of each visual-word that appears in the images, which is used as feature vector in the classification task.

In order to detect the pornographic images rely on accurate and fast identification, this research is proposed

An International Journal JSCDSS E-ISSN: 2289-8603

the BoVW model based on the FREAK descriptor on the Region of Interest (ROI). There are 4 Steps in the proposed model: 1) detecting the Region of Interest (ROI) which speed up the whole system performance due to eliminating lots of noise keypoints placed in the image background 2) feature extraction using the FREAK which is a simple and fast binary descriptor 3) clustering the extracted patches using k-means clustering algorithm and create the visual vocabulary 4) represent the images with produced visual vocabulary which is called BoVW and generate histograms per image and, 5) applying the SVM classifier to recognize the pornographic image against benign one. Steps 1 and 2 will be discussed in more details in the following sections. Fig. 1 depicted the proposed model using BoVW.

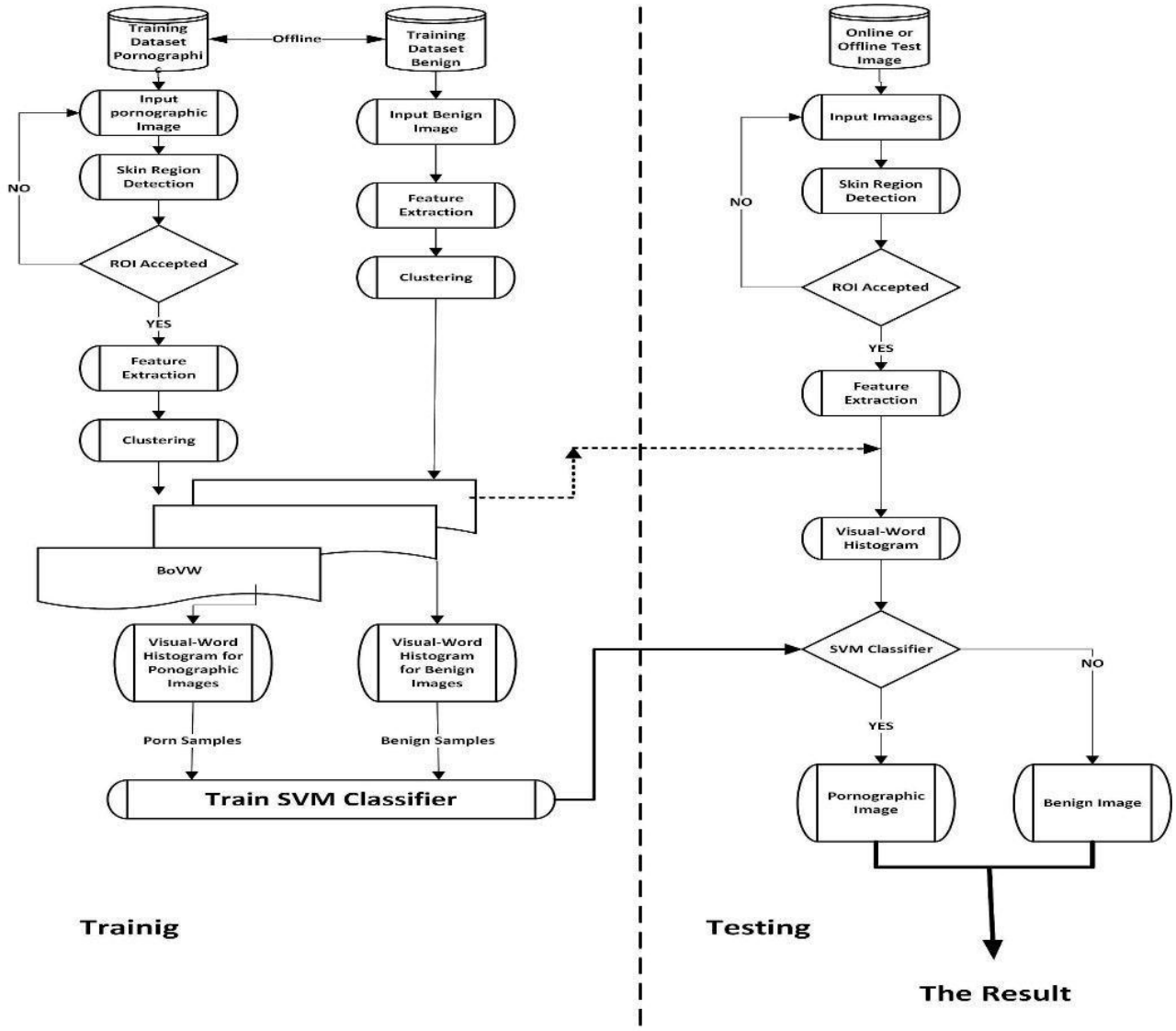


Fig. 1. Framework of the proposed model base on BoVW for pornographic image recognition.

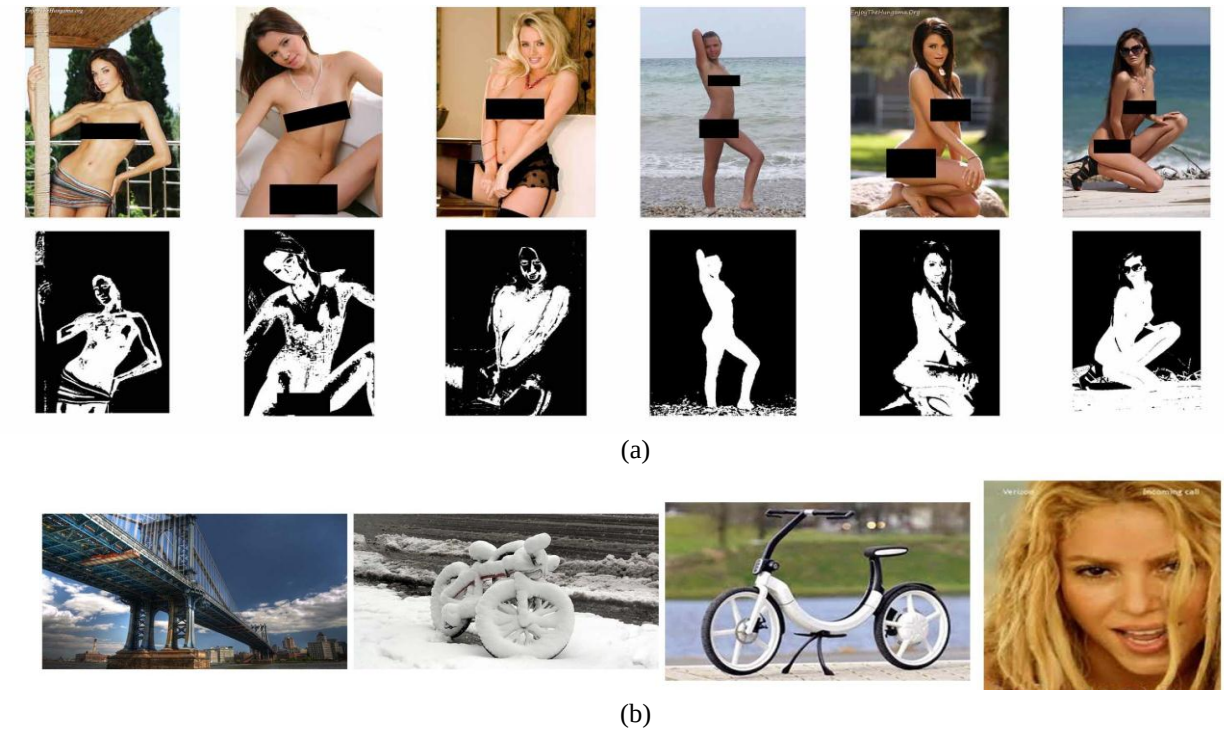
3.1 Selecting the Region of Interest (ROI)

Generally, each pornographic image consists of two types of scenes, image background and some people in the foreground. As the genital parts of the human body could be found in the skin color regions entire the image, the first step is detecting the skin likelihood regions known as ROI, which have the higher probability of exposing these parts. Applying keypoint description bounded to the ROI, improves the speed of the whole pornographic recognition system due to eliminating lots of noise keypoints placed in the background. Fig. 2a shows a few examples of the regular images and the same image with ROI which is selected from dataset.

Selecting the proper color space is very important for skin-like pixel detection. Based on the experiment results in (Maktabdar Oghaz et al., 2015) the SKN color space outperforms the state of the art color space to detect the skin-like regions in the images. This research benefits the SKN color space to detect the ROI. Therefore, the input images transferred from the RGB color space to the SKN. The following Eq. (1) shows the transformation formula of RGB to SKN color space.

An International Journal JSCDSS E-ISSN: 2289-8603

After detecting the ROI per input image, if the ratio of skin color area returned by ROI over the whole area of the target image exceeds a certain predefined threshold, it can be a pornographic image whit higher probability and proceeds to the further stages otherwise it labels as non-pornographic image.



(a)

(b)

Fig. 2. Examples of the image Data-Set a) regular pornographic images and the same images with the ROI in which the white color indicates the skin-like pixels. b) Non-pornographic samples of the Data-Set.

$$S = 0.088\frac{G}{R+G+B} - 58.89G - 30.014R - 11.952B - 7.24$$
$$K = 2.122\ B + 14.859\ G + 6.921R + 0.62\frac{G}{R+G+B} + 1.744$$
$$N = 0.342\frac{G}{R+G+B} - 3.698G - 2.25R - 0.103B - 0.464 \quad (1)$$

3.2 Keypoint Extraction Using the FREAK Descriptor

As respect of discussion in the literature, a fast descriptor plays an important rule to speed up the image recognition system based on BoVW model. This research benefits the FREAK (Alahi et al. 2012) descriptor due to its speed in the keypoint description and matching. The FREAK is a binary descriptor F, inspired by the Human Visual System, and more precisely the retina, which is constructed by a sequence of one-bit Difference of Gaussians (DoG). In this method, based on the retinal sampling grid, several pairs of pixels in the patches are selected to compute the differences of pixel intensities. The density of these pairs is higher near the center and drops exponentially when it goes around. Using this comparison, it can describe and match the image patches. It is formulated as follow:

$$F = \sum_{0 \le a < N} 2^a\, T(P_a), \quad (2)$$

Where, P_a is a pair of respective fields, N is the desired size of the descriptor, and

$$T(P_a) = \begin{cases} 1 & if\ I(P_a^{r_1}) - I(P_a^{r_2}) > 0 \\ 0 & otherwise, \end{cases} \quad (3)$$

With $I(P_a^{r_1})$ is the smoothed intensity of the first receptive field of the pairP_a. The next section describes the evaluation of proposed pornographic image detection base on FREAK descriptor.

4. Experiment and dataset

4.1 Experimental Setup

Since, there is no standard dataset in the field of pornographic image detection, in order to evaluate the proposed model performance, the total numbers of 14000 images are collected from the internet. It includes the different types of scenes such as landscape, building, face/ID image, car, plant, people and etc. as benign images with the total number of 7500 images, and also the number

An International Journal JSCDSS E-ISSN: 2289-8603

of 6500 pornographic images. Furthermore, the most of face/ID images collected from the SFA dataset (Casati et al., 2013). Some example images of the dataset are shown in the Fig. 2. Based on the experiment, it is observed that the performance is sensitive to the picture size. In the other words, with the higher-quality images the process takes more time than images with lower-quality. The main reason is that the number of detected keypoints in the high-quality image are more than the lower one. Excluding the face/ID pictures, the highest quality image used in the dataset is 5616*3744 pixels, and the lowest is 1175*783 pixels. The average size of used images in the collected dataset is 2560*1650 pixels. Furthermore, the average numbers of detected keypoints per image are approximately 2300 keypoints. Used images for the training phase include 3500 pornographic images and 3500 benign images selected from the dataset. Table 1 summarized the collected dataset.

Table 1
The collected data-set including pornographic and benign images

	Train Images	Test Images
Pornographic	3500	3000
Benign	3500	4000
Total images	7000	7000

In the proposed model, the SVM classifier is applied to pornographic image recognition. As it is shown in the Fig. 1 the proposed model constructed in two phases, training and testing. In the training phase a group of the pornographic images is used to generate the visual-word histogram and fed to the SVM classifier as positive samples, and also the benign images as the negative samples. The LIBSVM tool (Chang and Lin 2011) is used to train and test in this research. All of the experiments are executed in the MATLAB 2013b environment, with the machine of AMD Phenom (tm) II x 6, 20 G memory, and Windows 7.

4.2 Evaluation of the proposed model and discussion

In order to compare the proposed model performance, in this subsection two different type of experiments are summarized as below. In the first experiment, the speed of the recognition is compared with two most popular descriptors SIFT and SURF that used in the other researches for pornographic recognition. Additionally, there is a comparison with the model used in (Yizhi et al., 2010) which is similar to the proposed model in this research to show the performance of this model. Beside, several visual vocabulary word size (as measured by k parameter, the number of cluster centers in k-means clustering algorithm) are generated to find the best performance. As the experiment results have shown in Table 2, the proposed model based on the FREAK descriptor is able to remarkably outperform the similar works in the pornographic recognition. Compared to the method in (Yizhi et al., 2010) whit the time 3573ms, the required time to description and match keypoints per image in the proposed model is 1607ms with k=500, which is remarkable improvement in terms of computational complexity. The main reason is that SIFT and SURF descriptors, calculate high dimensional feature vectors per detected keypoints in the images, which are time consuming and therefore they have higher computational complexity. However, the FREAK is a simple binary descriptor with is very faster than the others. Table 2 shows the recognition time (in millisecond) for several vocabulary size from 150 to 500 visual-word. Meanwhile, it is observed that the larger vocabulary sizes perform better for recognition accuracy, within the tested range of 150 to 500.

In the second experiment, the pornographic recognition accuracy is tested. Table 3 shows the results in AUC (Area under Curve in the ROC curve). Due to the multiplicity of experiments i.e. 16 experiments, the results are shown in AUC form. As it is observed in the Table 3, followed by the SIFT-Based models with the AUC rate of 0.822 for k=500, the propose model also shown a remarkable accuracy with amount of 0.744 with the same k size. However, adding color moments (CM) in the method used in (Yizhi et al., 2010) gives better results, but it is more time consuming to do the process.

Table 2
The recognition time comparison for on images with average 2560x1650 pixel where approximately 2300 keypoints are detected per image.

Vocabulary Size	SIFT-Based	Method in [22]	SURF-Based	FREAK
150	3257	3374	1498	**1358**
250	3315	3419	1525	**1442**
350	3381	3488	1587	**1596**
500	3493	3573	1659	**1607**

Table 3
The recognition accuracy are shown in AUC

Vocabulary Size	SIFT-Based	Method in [22]	SURF-Based	FREAK
150	0.735	**0.747**	0.695	0.689
250	0.751	**0.769**	0.719	0.727
350	**0.783**	0.779	0.737	0.731
500	0.822	**0.831**	0.759	0.744

An International Journal JSCDSS E-ISSN: 2289-8603

5. Conclusion

This research is proposed a BoVW based model for pornographic image recognition and filtering in the applications with time constraints such as the web, in which very fast and simple binary descriptor FREAK is adopted. Using this algorithm to descript and match detected patches in the different images, developed the recognition speed. Furthermore, calculating descriptors in the skin-like regions of the images which have the higher probability of exposing the genital parts of the human body, leads to eliminate a lots of noise keypoints placed in the background and thus speed up the pornographic image recognition process. To detect skin area in the images, this research is applied the IHLS color space which is superior in skin detection. Finally, in order to find the most representational visual-vocabulary, the experiments are performed with several sizes from 150-500. The experimental results and comparison with similar researches show the superiority of the proposed model in the term of computational complexity.

Acknowledgment

This paper was supported by a grant from Science Fund MOSTI (01-01-06-SF1167).

References

Alahi, A., R. Ortiz, and P. Vandergheynst. 2012. "FREAK: Fast Retina Keypoint." 2012 IEEE Conference on Computer Vision and Pattern Recognition: 510–17.

Bay, Herbert, Tinne Tuytelaars, and Luc Van Gool. 2006. "Surf: Speeded up Robust Features." Springer berlin Heidelberg, Computer Vision–ECCV 2006: 404–17.

Casati, JPB, DR Moraes, and ELL Rodrigues. 2013. "SFA: A Human Skin Image Database Based on FERET and AR Facial Images." iris.sel.eesc.usp.br.

Chang, CC, and CJ Lin. 2011. "LIBSVM: A Library for Support Vector Machines." ACM Transactions on Intelligent Systems and Technology (TIST): 1–39.

Deselaers, Thomas, Lexi Pimenidis, and Hermann Ney. 2008. "Bag-of-Visual-Words Models for Adult Image Classification and Filtering." 2008 19th International Conference on Pattern Recognition: 1–4.

Fleck, MM, DA Forsyth, and Chris Bregler. 1996. "Finding Naked People." Computer Vision—ECCV'96. .

Forsyth, D A, and M.M. Fleck. 1996. "Identifying Nude Pictures." (1): 103–8.

Gauglitz, Steffen, Tobias Höllerer, and Matthew Turk. 2011. "Evaluation of Interest Point Detectors and Feature Descriptors for Visual Tracking." International Journal of Computer Vision 94(3): 335–60.

Herwindiati, Dyah E, Sani M Isa, and Rahmat Sagara. 2010. "THE NEW NOTION DISTANCE OF CONTENT BASED IMAGE RETRIEVAL (CBIR)." 16(1): 51–67.

Jiang, Yu-Gang, Chong-Wah Ngo, and Jun Yang. 2007. "Towards Optimal Bag-of-Features for Object Categorization and Semantic Video Retrieval." Proceedings of the 6th ACM international conference on Image and video retrieval - CIVR '07: 494–501.

Kherfi, ML, D Ziou, and A Bernardi. 2004. "Image Retrieval from the World Wide Web: Issues, Techniques, and Systems." ACM Computing Surveys (CSUR) 36(1): 35–67.

Lienhart, Rainer. 2009. "FILTERING ADULT IMAGE CONTENT WITH TOPIC MODELS Rainer Lienhart Lehrstuhl F ¨ Ur Multimedia Computing Universit ¨ at Augsburg Augsburg , Germany Rudolf Hauke Advanced US Technology Group , Inc . IEEE International: 1472–75.

Liu, Yizhi, and Hongtao Xie. 2009. "Constructing SURF Visual-Words for Pornographic Images Detection." IEEE Conference, Computers and Information Technology, 2009. ICCIT '09. 12th International Conference on (Iccit): 404–7.

Lopes, Ana P B et al. 2009. "A BAG-OF-FEATURES APPROACH BASED ON HUE-SIFT DESCRIPTOR FOR NUDE DETECTION." (Eusipco): 1552–56.

Lopes, APB, and SEF de Avila. 2009. "Nude Detection in Video Using Bag-of-Visual-Features." IEEE Conference, Computer Graphics and Image Processing (SIBGRAPI), 2009 XXII Brazilian Symposium on: 224–31.

Lowe, David G. 2004. "Distinctive Image Features from Scale-Invariant Keypoints." International Journal of Computer Vision 60(2): 91–110.

Lv, Lintao et al. 2011. "Pornographic Images Detection Using High-Level Semantic Features." 2011 Seventh International Conference on Natural Computation: 1015–18.

Maktabdar Oghaz, Mahdi et al. 2015. "A Hybrid Color Space for Skin Detection Using Genetic Algorithm Heuristic Search and Principal Component Analysis Technique." Plos One 10(8): e0134828.

O'Hara, S, and BA Draper. 2011. "Introduction to the Bag of Features Paradigm for Image Classification and Retrieval." arXiv preprint arXiv:1101.3354 (July): 1–25.

Steel, C. 2012. "The Mask-SIFT Cascading Classifier for Pornography Detection." Internet Security (WorldCIS), 2012 World Congress on: 139–42.

Wang, James Ze, Jia Li, Gio Wiederhold, and Oscar Firschein. 1998. "System for Screening Objectionable Images." (April 1998): 1–12.

Wang, Meng, Richang Hong, Xiao-Tong Yuan, et al. 2012. "Movie2Comics: Towards a Lively Video Content Presentation." IEEE Transactions on Multimedia 14(3): 858–70.

Wang, Meng, Richang Hong, Guangda Li, and ZJ Zha. 2012. "Event Driven Web Video Summarization by Tag Localization and Key-Shot Identification." … , IEEE Transactions on 14(4): 975–85.

Wang, YS, LY Ning, and W Gao. 2008. "Detecting Pornographic Images with Visual Words." Transactions of Beijing Institute of Technology 28(5): 410–13.

An International Journal JSCDSS E-ISSN: 2289-8603

Yizhi, Liu, Lin Shouxun, Tang Sheng, and Zhang Yongdong. 2010. "Adult Image Detection Combining BoVW Based on Region of Interest and Color Moments." IFIP Advances in Information and Communication Technology 340: 316–25.

Zhang, Jing et al. 2013. "An Approach of Bag-of-Words Based on Visual Attention Model for Pornographic Images Recognition in Compressed Domain." Neurocomputing 110: 145–52.

Zheng, Huicheng, Mohamed Daoudi, and Bruno Jedynak. 2004. "Blocking Adult Images Based on Statistical Skin Detection." 4(2): 1–14.

An International Journal
JSCDSS
E-ISSN: 2289-8603