Nude Detection in Video using Bag-of-Visual-Features

Ana Paula B. Lopes[*†], Sandra E. F. de Avila[*], Anderson N. A. Peixoto[*],
Rodrigo S. Oliveira[*], Marcelo de M. Coelho[*‡] and Arnaldo de A. Araújo[*]

[*]Computer Science Department, Federal University of Minas Gerais – UFMG
31270–010, Belo Horizonte, MG, Brazil
[†]Exact and Technological Sciences Department, State University of Santa Cruz – UESC
45662–000, Ilhéus, BA, Brazil
[‡]Preparatory School of Air Cadets – EPCAR
36205–900, Barbacena, MG, Brazil

{paula, sandra, andenap, rsilva, mcoelho, arnaldo}@dcc.ufmg.br

Abstract—The ability to filter improper content from multimedia sources based on visual content has important applications, since text-based filters are clearly insufficient against erroneous and/or malicious associations between text and actual content. In this paper, we investigate a method for detection of nudity in videos based on a bag-of-visual-features representation for frames and an associated voting scheme.

Bag-of-Visual-Features (BoVF) approaches have been successfully applied to object recognition and scene classification, showing robustness to occlusion and also to the several kinds of variations that normally curse object detection methods. To the best of our knowledge, only two proposals in the literature use BoVF for nude detection in still images, and no other attempt has been made at applying BoVF for videos. Nevertheless, the results of our experiments show that this approach is indeed able to provide good recognition rates for nudity even at the frame level and with a relatively low sampling ratio. Also, the proposed voting scheme significantly enhances the recognition rates for video segments, achieving, in the best case, a value of 93.2% of correct classification, using a sampling ratio of 1/15 frames. Finally, a visual analysis of some particular cases indicates possible sources of misclassifications.

Keywords-Nude detection; Bag-of-Visual-Features; Video classification;

I. INTRODUCTION

According to [1], one important challenge to be approached by Computer Scientists nowadays is to find effective ways to deal with the vast amount of multimedia data being generated every day. Among the various subtasks involved in that generic one, filtering of improper multimedia material by its visual content has reasonably obvious importance and applications, since text-based filters are clearly insufficient against erroneous and/or malicious associations between text and actual content. A filter able to deal with visual content could, for example, help to identify the existence of adult material associated with childish tags or segments of adult videos inserted in seemingly unsuspected ones.

In this paper, we investigate a method for detection of nudity in videos based on a bag-of-visual-features representation for frames and an associated voting scheme. A number of papers have been proposed in recent years for identifying nudity in still images. Most of these proposals start by the application of skin detectors and then try to apply geometrical models of body postures to the detected skin areas, in the hope of finding nude images. The main problems with such approaches are: a) precise skin detectors are themselves far from trivial to build and b) the immense variability in the body poses of such images makes it difficult to define a generic geometrical model.

The nude recognition task can be seen as a special kind of object recognition task. Bag-of-Visual-Features (BoVF) approaches have been applied to object recognition very successfully [2], [3], showing robustness to occlusion and also to the several kinds of variations that normally curse object detection methods, such as those derived from lighting conditions, scale, shape and rotations.

BoVF approaches – which are going to be detailed in Section III – typically use local descriptors based on gray-level values. Nevertheless, for nude detection, the color is a fundamental information, as we have already confirmed by the experiments described in [4]. In that work, a BoVF approach based on the HueSIFT [5] is applied to still images classification between nude and non-nude images. In this paper, we expand that proposal with a voting scheme, aiming at detection of nudity in videos. To the best of our knowledge, no other attempt has been made at applying BoVF to videos, and [6] is the only other paper using a BoVF representation for nude detection in still images.

Additionally, for this work to be done, an annotated database has been created and made available for evaluation (details are provided in Section V). Some snapshots from the videos collected for this database can be seen in Figure 1.

This paper proceeds as follows: Section II describes some related work; Section III details how a BoVF representation

Figure 1: Examples of non-nude (first and second rows) and nude frames (third and forth rows) from the database built for this work.

is built; Section IV shows how we propose to use BoVF representations for the task of nudity detection in videos; in Section V, numerical and visual results of a set of experiments are presented and discussed; finally, Section VI presents some concluding remarks for the paper.

II. RELATED WORK

The pioneer work for identifying adult images by the analysis of image content is proposed by [7], [8], [9], [10]. Their approach combines color and texture properties to obtain a mask for skin regions, which are then fed to a specialized grouper, which attempts to group a human figure using geometric constraints derived from human body structure.

Most subsequent proposals on nude detection are also based on this general idea of searching for skin regions and then describing their geometry. In [11], [12], [13], [14], a statistical color model for detecting skin regions is developed.

In [15], a framework for nude image detection based on skin region information is presented. Their approach extracts color and texture features from arbitrary-shaped segmented regions. Gaussian Mixture Models (GMM) are built for skin and non-skin region classification.

To filter objectionable images, [16] proposed an adaptive skin color detection model. The skin-similar pixels are used to train a GMM with several Gaussian kernels using the standard Expectation-Maximization (EM) algorithm. Support Vector Machine (SVM) is then used to identify the skin component using the trained GMM.

To overcome the chromatic deviation coming from the unusual lighting conditions, [17] proposed an online skin tone sampling mechanism based on face detection. A similar approach is proposed by [18]. The face detection and skin filter are used to filter out the non-nude images. Then, a nonlinear-SVM is used with shape features to determine whether images are nude or not. In [19], an algorithm based on the face and trunk detection is developed to classify pornographic images. To identify adult images, [20] analyzed the shape of human body trunk.

A naked image detection algorithm is proposed by [21]. A learning-based scheme for matching chromatic distribution is proposed to determine the image skin chroma distribution. Texture features are used to acquire accurate skin segmentation. Low-level geometrical constraints and the mug shot exclusion procedure are employed to further examine the skin regions.

To detect pornographic images, [22] proposed a skin model based on the combination of YIQ, YUV, and HSV color models. A white balance algorithm is applied to better detection of skin areas and a combination of constraints on color, texture, and geometric properties are used as features fed to a SVM classifier.

In [23], the first step of the system proposed is to use content-based image retrieval to determine whether the image contains humans in it. This retrieval step is based on color and shape features. Then a skin color model is performed on the image to judge whether the image is pornographic or benign.

Some attempts at avoiding the need for a fine-tuned skin detector originated color-based approaches (usually combined with shape and/or texture features). In [24], [25], a combination of MPEG-7 visual descriptors are used for adult image identification. Visual features can also be combined with other types of features, as in the framework for

recognizing pornographic web pages presented in [26], [27], [28], in which text and image are both analyzed.

In [29], human-skin blobs are used for extracting features representing potentially pornographic images. A two-class cascade SVM is employed for classifying the images in three classes: porno, nude and non-adult. These classes are characterized by their faces content.

In [30], a framework for recognizing pornographic movies by fusing audio and video information together is developed. A one-class GMM is used to recognize porno-sounds and a generalized pornographic image recognition algorithm, based on contours, is used to detect pornographic frames from a video shot. Then, a fusion algorithm based on the Bayes theory is employed to combine the recognition results from audio and video.

In [31], a method and a device for determining obscenities in videos and blocking them is described. By extracting changes in (generic) feature values for each type of video within a lapse of time and comparing changes in those feature values from an input video, obscenities are determined.

In [32], a method and *apparatus* for generating discriminant functions for distinguishing obscene videos is proposed. Frames of input video data are extracted, and visual features for each frame and each group frame are extracted from the video data to be compared with the generated discriminant function.

In [6], a method to classify images into different categories of pornographic content is presented, which is based on what is called a bag-of-visual-words. This is another denomination for BoVF when applied to visual material. Their proposal is similar to [4], but the features used to build the vocabulary are simply patches (gray-level values) around interest points, while [4] applies the more descriptive HueSIFT [5].

In our work, we propose an extension to the work described in [4]. There, a BoVF-based approach using the HueSIFT descriptor was used to detect nudity in still images. In our work, individual frames are represented by BoVFs and classified by a linear SVM model. Then, a voting scheme is applied to enhance the classification for videos segments.

III. REPRESENTING IMAGES AS BOVF

Bag-of-Visual-Features (BoVF) are inspired in the well-known bag-of-words from Information Retrieval [33]. There, every text document is represented by a histogram counting words' occurrences. Normally the vocabulary is defined from a collection of training documents (the *corpus*). In order to reduce the dimensionality of the final histograms, non-significant words like articles and prepositions, called *stop words*, are removed from the vocabulary; in addition, words are clustered into families by their root. Therefore, the so called bag-of-words – the histogram representing each text – actually counts occurrences inside each word family.

The equivalent to those word families in the visual scenario are clusters of small patches taken from a number of training images. The idea here is that the statistics of such patches are able to represent the original images in a compact yet discriminative way. The whole process of creating a BoVF representation is summarized in Figure 2.

In step (a), a number of points from which to take the patches are selected. Point selection can be made in several forms: it is possible to collect these points densely, using a grid to define patches boundaries, but this approach is not computationally efficient. Sparser samplings can be achieved both by taking some points randomly or, what is more common, by applying an interest point detector. Algorithms for detection of interest points are able to detect points in the image which have specified visual characteristics, considered as "interesting". One interest point detector widely used to several different tasks, including object recognition is the Scale-Invariant Feature Transform (SIFT) [34], which searches for points invariant to several transformations.

Point selection is followed by point description – step (b). Actually, this step consists of the visual description of the regions around the points (the patches themselves). Here again, it is possible to describe such regions in varied ways, like gray-level values or gray-level gradients, for example. Interest point detectors normally deliver their own descriptors for the points found. This way, in those cases where interest point detectors are used to point selection, their own descriptors are commonly chosen for point description.

Given the points for a number of training images, they are clustered by visual similarity, as it can be seen in step (c). The idea in this step is to quantize the points feature space, so to create the visual vocabulary. Typically, the k-means clustering algorithm [35] and Euclidean distance are applied.

A critical point in this step is to establish the vocabulary size. While in text retrieval the vocabulary size emerges naturally from the training *corpus* and the subsequent dimensionality reduction techniques which are applied, in the visual case, how to choose the best vocabulary size is an open research issue. In our work, the vocabulary size is chosen empirically, and an explicit effort to keep it small was done, since large vocabulary sizes have an important impact on processing time.

Once a visual vocabulary has been defined, it is possible to associate each previously selected point to a word in this vocabulary, as it is seen in step (d). In practice, this means to associate every descriptor to that cluster whose centroid is the nearest to the descriptor.

After that, the BoVF histogram is built by counting the occurrences of every word, as it is shown in step (e). Normally, some normalization is applied to the histogram. In the present work, this is done by computing the relative frequencies, which is equivalent to normalizing the histogram to one. Such a simple histogram is used because there is some evidence in the literature [36] that more sophisticated
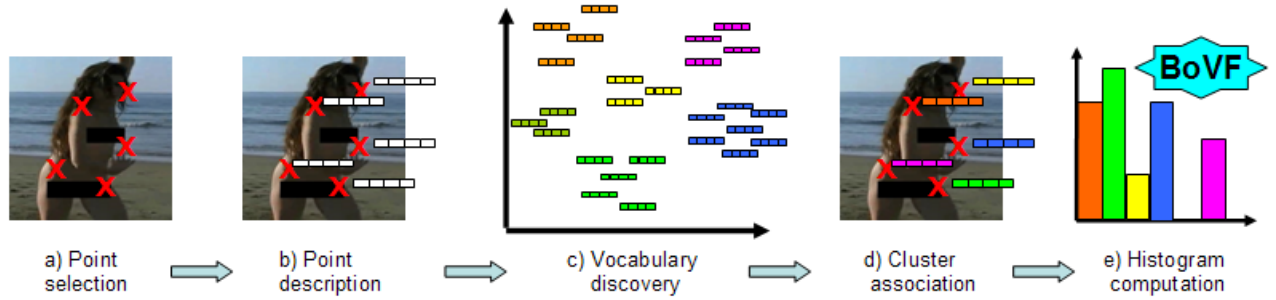
Figure 2: Creating a Bag-of-Visual-Features for images. Details on each step are provided in the text (this image is better appreciated in color).

schemes for the final BoVF histogram, like TF/IDF [33] for example, which is widely used in text retrieval, have a small influence in the final recognition results.

The BoVF histogram obtained by the process just described represents the image content, and can be used for comparisons among images. In our case, a linear SVM classifier was applied to distinguish between nude and non-nude images. The choice of a linear kernel was motivated by the fact that in high-dimensional spaces, it normally suffices. Additionally, since the results of this classifier depend strongly on the penalty error parameter, another extensive empirical search for the best value for this parameter was performed, based on the process described in [37].

IV. Detecting Nudity from Videos

Normally, BoVF approaches rely on descriptors based on gray-level values, but in case of nude recognition, the first intuition – pursued by most previous authors – is that the skin color is an important clue. This supposition is confirmed by our experiments reported in a previous recent work [4]. In that work, the BoVF representation created by the process described in Section III was applied to nude detection in still images, both by using the gray-level SIFT descriptor [34] and a variation called HueSIFT [38], which generates a point descriptor that takes into account the color hue. The correct recognition rates achieved by them – using a linear SVM classifier – were $65 \pm 3\%$ for a pure (gray-level based) SIFT and $85 \pm 2\%$ for HueSIFT.

In this work, we propose to explore that BoVF representation – based on HueSIFT – to detect nudity in videos. Working with videos has some advantages over still images. Instead of a unique image, videos provide entire sequences of frames to collect information from. However, this creates a need to specify sampling rates for the frames which are at the same time "safe" (not ignoring entire nude sequences, for example) and computationally feasible.

A. The proposed voting scheme

To take advantage of the existence of several similar frames, a voting scheme is defined as follows: first, the videos are separated into frames, and some of them are selected according to a previously chosen sampling rate. The BoVFs for the selected frames are built and then each frame is classified between nude and non-nude. For each video segment, every frame classification is counted as *votes* for that class, and the class receiving the higher number of votes is considered the video class. Because false-negatives are potentially harmful in typical scenarios, ties are considered as nude. Finally, in order to verify the recognition rates, the final video classification was compared to the manually built ground-truth.

V. Experimental Results

A. Database Creation

Despite the importance of the nude detection task, there is no public available standard database for a significant comparison among different algorithms. Then, to be able to test the proposed algorithm, a collection of 179 video segments was created. Such segments were collected from the following movies: *Alpha Dog*, *Basic Instinct*, *Before The Devil Knows You're Dead*, *Cashback*, *Eros*, *Les Anges Exterminateurs*, *Loner*, *Original Sin*, *Primer*, *Striptease* and *The Bubble*. Some example snapshots from the nude and non-nude classes can be seen in Figure 1. The entire segments database are available online[1].

For nude samples, longer sequences were partitioned into shorter ones, so the sequences submitted for classification were between 43 frames and 308 frames long.

The sequences for the non-nude class were collected by randomly selecting the initial time and length. Random selections which felt inside a nude scene were discarded. These random selections were performed on the same movies above listed for the nude class, and resulted in segments whose lengths are between 50 and 278 frames.

B. Experiment Setup

The experiments were designed to evaluate the ability of discriminating between nude and non-nude videos from

---

[1]http://www.npdi.dcc.ufmg.br/nudeDetection

BoVFs representations for the selected frames. Two different sample ratios were tested: 1/15 frame (2 frames per second) and 1/30 frames (1 frame per second). These sample ratios provided 2021 and 1011 frames to be classified, respectively. Sampling ratios lower than those ones were considered not desirable, to avoid missing short nude sequences between non-nude material.

After that, the process described in Section III was applied to create BoVFs representation for each selected frame. A random subset of 10,000 HueSIFT points, established through preliminary experimental tests, was selected for vocabulary creation.

Three different vocabulary sizes were tested for classification of those BoVF vectors between nude and non-nude: 60, 120 and 180. A linear SVM classifier was used, and its penalty error parameter (C) was refined by the procedure described in [37], extensively testing values from 10^{-5} to 10^5 with a multiplicative step of 10. In each step, the classification rate was measured in a 5-fold cross validation scheme. This way, for each k (vocabulary size), the C value which provided a higher recognition rate was found and fixed in the following runs.

In order to compute de confidence intervals for the recognition rates of each k (and its best C previously determined), 30 new 5-fold cross validation runs were performed, randomly varying the folds separation.

Finally, no comparison with alternative proposals was possible due to the fact that, in previous papers, nor the databases neither the codes for the algorithms were made available.

C. Results and Discussion

The recognition rates achieved in the above experiments are summarized in Table I. In them, the recognition rate for frames (second column) are those expected if only a keyframe was selected to represent the entire video segment. In the third column, are the recognition rates using the proposed voting scheme. From those tables, it is possible to see that, in all cases, applying the voting algorithm causes a statistically significant increase in the overall recognition rate for the videos, when compared to the recognition rate of the individual frames. Such results indicate that, indeed, the voting scheme is able to take advantage of the existence of several similar frames to solve some dubious cases.

The results indicate also that, the greater the vocabulary size, the greater the enhancement obtained. Nevertheless, it is worth to point out that this is a very specific result for these experimental setting and database. In this work, greater values for the vocabulary size were not tested because of the computational overhead which would result.

The increases shown in the third column of Table I(a) are smaller than those ones shown in Table I(b). This means that taking a higher sampling ratio improves even more the correct recognition of nudity in videos, which is in

Table I: Comparing recognition rates for keyframe and voting based classification.

Voc. Size	Keyframe (%)	Voting (%)	Increase
60	76.4 ± 0.2	77.1 ± 0.4	0.7
120	80.2 ± 0.3	80.9 ± 0.4	0.7
180	83.9 ± 0.2	88.4 ± 0.6	4.5

(a) 1/30 frames

Voc. Size	Keyframe (%)	Voting (%)	Increase
60	79.1 ± 0.1	80.5 ± 0.4	1.4
120	83.7 ± 0.2	87.3 ± 0.4	3.6
180	85.9 ± 0.1	93.2 ± 0.4	7.3

(b) 1/15 frames

accordance with the intuition that with more frames to vote on (i.e., more information to analyze), one would be able to better solve the ambiguities.

The less desirable cases of wrong classification are the false-negatives, i.e., the cases of nude scenes being erroneously classified as non-nude. In Table II, the decreases in the false-negative rates are presented, showing that only in one case, where the smaller vocabulary size was used, no improvement is achieved.

Table II: False-negative rates for keyframe and voting based classification.

Voc. Size	Keyframe (%)	Voting (%)	Decrease
60	12.2 ± 0.2	10.4 ± 0.3	1.8
120	11.0 ± 0.2	9.1 ± 0.2	1.9
180	8.0 ± 0.2	4.2 ± 0.3	3.3

(a) 1/30 frames

Voc. Size	Keyframe (%)	Voting (%)	Decrease
60	10.7 ± 0.1	10.7 ± 0.3	0.0
120	10.0 ± 0.1	8.5 ± 0.2	1.5
180	7.5 ± 0.1	4.2 ± 0.2	3.3

(b) 1/15 frames

In Figures 3 and 4, some particular results are presented. Figure 3 shows examples in which the voting resulted in a tie, meaning a strong ambiguity in the scene from the classifier point of view (in other words, for the classifier, those images are near the decision boundary).

In examples labeled as (a) to (f), the video segment is misclassified, while in those labeled as (g) to (l), the voting is able to solve the ambiguity correctly.

In Figure 4, in cases (a) to (d) and (i) to (l), the final classification was non-nude, and in cases (e) to (h) and

(m) to (p), the final classification was nude. In other words, cases (i) to (p) are examples in which the voting was a fundamental step for overcoming occasional frames misclassifications, since although there were frames classified in both classes, the correct class prevailed in the voting. The remaining cases are examples in which the ambiguity could not be solved.

Finally, the examples presented in Figures 3 and 4 indicate that background colors near to skin tones, the presence of large skin areas and illumination variations are the most probable sources of misclassifications.

VI. CONCLUSION

In this paper, we presented an approach for the recognition of nudity in video segments, based on a Bag-of-Visual-Features (BoVF) representation for the frames extracted from those segments. This work is an extension of a previous one [4] aimed at still images. As an extension, a voting scheme was added to the original approach in order to make use of the availability of several similar images to solve ambiguous cases. Additionally, a test database was produced and made available for research purposes.

Although BoVF approaches have been successfully used in object and scene categorization by several authors, this is – to the best of our knowledge – the first use of BoVF for nude detection in videos, and only other two proposals applied similar techniques to still images.

The experimental results discussed along the paper show that the algorithm is indeed able to provide good recognition rates for nudity even at the frame level, achieving, in the best case, a value of 93.2% of correct classification. The results also indicate that the proposed voting scheme significantly enhances the recognition rates for video segments, solving some ambiguities and frames misclassifications. Finally, the analysis of some particular examples indicate that background colors near to skin tones, the presence of large skin areas and illumination variations are the most probable sources of misclassifications.

Future work includes the validation of the proposed scheme using a larger database, including low-resolution videos from sites focused on user-generated content, like YouTube[2] and similar ones. Also, in a more generic line of research, the issue of vocabulary formation in BoVF approaches is currently under investigation. As we pointed out in Section III, this is an open research issue, and we hope that a better formed vocabulary can enhance even more our results, possibly allowing to distinguish among more than two classes.

ACKNOWLEDGMENTS

[2]http://www.youtube.com

REFERENCES

[1] A. de Carvalho, A. Brayner *et al.*, "Grand Challenges in Computer Science Research in Brazil – 2006-2016," Brazilian Computer Society, Tech. Rep., May 2006.

[2] S. Agarwal and A. Awan, "Learning to detect objects in images via a sparse, part-based representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 26, no. 11, pp. 1475–1490, 2004, member-Dan Roth.

[3] J. Yang, Y.-G. Jiang, A. G. Hauptmann, and C.-W. Ngo, "Evaluating bag-of-visual-words representations in scene classification," in *ACM Multimedia Information Retrieval (MIR)*. New York, NY, USA: ACM, 2007, pp. 197–206.

[4] A. P. B. Lopes, S. E. F. de Avila, A. N. A. Peixoto, R. S. Oliveira, and A. de A. Araújo, "A bag-of-features approach based on hue-SIFT descriptor for nude detection," in *Proceedings of the 17th European Signal Processing Conference (EUSIPCO)*, Glasgow, Scotland, August 2009.

[5] K. van de Sande, T. Gevers, and C. Snoek, "Evaluation of color descriptors for object and scene recognition," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, 2008.

[6] T. Deselaers, L. Pimenidis, and H. Ney, "Bag-of-visual-words models for adult image classification and filtering," in *International Conference on Pattern Recognition (ICPR)*, Florida, USA, December 2008.

[7] M. M. Fleck, D. A. Forsyth, and C. Bregler, "Finding naked people," in *Proceedings of the 4th European Conference on Computer Vision-Volume II (ECCV)*. London, UK: Springer-Verlag, 1996, pp. 593–602.

[8] D. A. Forsyth and M. M. Fleck, "Identifying nude pictures," in *Proceedings of the 3rd IEEE Workshop on Applications of Computer Vision (WACV)*. Washington, DC, USA: IEEE Computer Society, 1996, pp. 103–108.

[9] ——, "Body plans," in *Proceedings of the Conference on Computer Vision and Pattern Recognition (CVPR)*. Washington, DC, USA: IEEE Computer Society, 1997, pp. 678–683.

[10] ——, "Automatic detection of human nudes," *International Journal on Computer Vision (IJCV)*, vol. 32, no. 1, pp. 63–77, 1999.

[11] M. J. Jones and J. M. Rehg, "Statistical color models with application to skin detection," *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, vol. 1, pp. 1274–1280, 1999.

[12] W. Zeng, W. Gao, T. Zhang, and Y. Liu, "Image guarder: An intelligent detector for adult images," in *Asian Conference on Computer Vision*, Jeju Island, Korea, January 2004, pp. 198–203.

[13] H. Zheng, H. Liu, and M. Daoudi, "Blocking objectionable images: adult images and harmful symbols," in *Proceedings of the IEEE International Conference on Multimedia and Expo (ICME)*, June 2004, pp. 1223–1226.

(a) Non-Nude (b) Nude (c) Non-Nude (d) Nude

(e) Non-Nude (f) Nude (g) Non-Nude (h) Nude

(i) Non-Nude (j) Nude (k) Non-Nude (l) Nude

Figure 3: Images of video segments in which the voting process has resulted in a tie, and therefore a final classification as nude. The classes provided by the frame classifier are indicated. For segments represented by frames (a) to (f), this yielded false-positive results, and for those ones represented by frames (g) to (l), the ambiguity was correctly solved to provide true-positives.

(a) Non-Nude (b) Nude (c) Non-Nude (d) Nude

(e) Non-Nude (f) Nude (g) Non-Nude (h) Nude

(i) Non-Nude (j) Nude (k) Non-Nude (l) Nude

(m) Non-Nude (n) Nude (o) Non-Nude (p) Nude

Figure 4: Some more cases in which some ambiguity was found, but not to the point of a tie: in cases (a) to (d) and (i) to (l), the final classification was non-nude, and in cases (e) to (h) and (m) to (p), the final classification was nude. This means that in the first two lines, the voting was not able to solve the ambiguity, while the third and forth lines show examples in which the voting was fundamental.

[14] Q.-F. Zheng, W. Zeng, G. Wen, and W.-Q. Wang, "Shape-based adult image detection," in *Proceedings of the Third International Conference on Image and Graphics (ICIG)*. Washington, DC, USA: IEEE Computer Society, 2004, pp. 150–153.

[15] Y. Xu, B. Li, X. Xue, and H. Lu, "Region-based pornographic image detection," *IEEE 7th Workshop on Multimedia Signal Processing (MMSP)*, pp. 1–4, November 2005.

[16] Q. Zhu, C.-T. Wu, K.-T. Cheng, and Y.-L. Wu, "An adaptive skin model and its application to objectionable image filtering," in *Proceedings of the 12th Annual ACM International Conference on Multimedia*. New York, NY, USA: ACM, 2004, pp. 56–63.

[17] J.-S. Lee, Y.-M. Kuo, and P.-C. Chung, "The adult image identification based on online sampling," in *Proceedings of the International Joint Conference on Neural Networks (IJCNN)*. IEEE, July 2006, pp. 2566–2571.

[18] C.-Y. Jeong, J.-S. Kim, and K.-S. Hong, "Appearance-based nude image detection," in *Proceedings of the 17th International Conference on Pattern Recognition (ICPR)*, vol. 4. Washington, DC, USA: IEEE Computer Society, August 2004, pp. 467–470.

[19] X. Shen, W. Wei, and Q. Qian, "The filtering of internet images based on detecting erotogenic-part," in *Proceedings of the Third International Conference on Natural Computation (ICNC)*. Washington, DC, USA: IEEE Computer Society, 2007, pp. 732–736.

[20] J. Yang, Z. Fu, T. Tan, and W. Hu, "A novel approach to detecting adult images," in *Proceedings of the 17th International Conference on Pattern Recognition (ICPR)*. Washington, DC, USA: IEEE Computer Society, August 2004, pp. 479–482.

[21] J.-S. Lee, Y.-M. Kuo, P.-C. Chung, and E.-L. Chen, "Naked image detection based on adaptive and extensible skin color model," *Pattern Recognition*, vol. 40, no. 8, pp. 2261–2270, 2007.

[22] H. Zhu, S. Zhou, J. Wang, and Z. Yin, "An algorithm of pornographic image detection," in *Proceedings of the Fourth International Conference on Image and Graphics (ICIG)*. Washington, USA: IEEE Computer Society, 2007, pp. 801–804.

[23] B.-B. Liu, J.-Y. Su, Z.-M. Lu, and Z. Li, "Pornographic images detection based on CBIR and skin analysis," *International Conference on Semantics, Knowledge and Grid (SKG)*, vol. 0, pp. 487–488, 2008.

[24] J.-L. Shih, C.-H. Lee, and C.-S. Yang, "An adult image identification system employing image retrieval technique," *Pattern Recognition Letters*, vol. 28, no. 16, pp. 2367–2374, 2007.

[25] S.-J. Yoo, "Intelligent multimedia information retrieval for identifying and rating adult images," in *Proceedings of 8th International Conference Knowledge-Based Intelligent Information and Engineering Systems (KES)*, ser. Lecture Notes in Computer Science, vol. 3213. Springer, 2004, pp. 164–170.

[26] M. Hammami, Y. Chahir, and L. Chen, "Webguard: A web filtering engine combining textual, structural, and visual content-based analysis," *IEEE Transactions on Knowledge and Data Engineering*, vol. 18, no. 2, pp. 272–284, 2006, member-Chen,, Liming.

[27] W. Hu, O. Wu, Z. Chen, and Z. Fu, "Recognition of pornographic web pages by classifying texts and images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 6, pp. 1019–1034, 2007.

[28] J. Polpinij, C. Sibunruang, S. Paungpronpitag, R. Chamchong, and A. Chotthanom, "A web pornography patrol system by content-based analysis: In particular text and image," in *Proceedings of the IEEE International Conference on Systems, Man and Cybernetics (SMC)*, October 2008, pp. 500–505.

[29] J. R. del Solar, V. Castaneda, R. Verschae, R. Baeza-Yates, and F. Ortiz, "Characterizing objectionable image content (pornography and nude images) of specific web segments: Chile as a case study," in *Proceedings of the Third Latin American Web Congress (LA-WEB)*. Washington, DC, USA: IEEE Computer Society, 2005, pp. 269–278.

[30] H. Zuo, O. Wu, W. Hu, and B. Xu, "Recognition of blue movies by fusion of audio and video," *Proceedings of the IEEE International Conference on Multimedia and Expo (ICME)*, pp. 37–40, April 2008.

[31] S. M. Lee, H. G. Lee, T. Y. Nam, and J. S. Jang, "Method and device for discriminating obscene video using time-based feature value," United States Patent US 2007/0101354 A1, May 2007.

[32] S. M. Lee, T. Y. Nam, J. S. Jang, and H. G. Lee, "Method and apparatus for distinguishing obscene video using visual feature," United States Patent US 2007/0098267 A1, May 2007.

[33] R. A. Baeza-Yates and B. A. Ribeiro-Neto, *Modern Information Retrieval*. ACM Press / Addison-Wesley, 1999.

[34] D. Lowe, "Object recognition from local scale-invariant features," *IEEE International Conference on Computer Vision (ICCV)*, vol. 2, pp. 1150–1157 vol.2, 1999.

[35] T. M. Mitchell, *Machine Learning*. New York: McGraw-Hill, 1997.

[36] Y.-G. Jiang, C.-W. Ngo, and J. Yang, "Towards optimal bag-of-features for object categorization and semantic video retrieval," in *Proceedings of the International Conference on Image and Video Retrieval (CIVR)*, 2007, pp. 494–501.

[37] C.-C. Chang and C.-J. Lin, *LIBSVM: a library for support vector machines*, 2001, software available at http://www.csie.ntu.edu.tw/ cjlin/libsvm.

[38] S. L. Wang and A. W. C. Liew, "Information-based color feature representation for image classification," in *Proceedings of the IEEE International Conference on Image Processing (ICIP)*, vol. 6. IEEE, October 2007, pp. 353–356.