A BAG-OF-FEATURES APPROACH BASED ON HUE-SIFT DESCRIPTOR FOR NUDE DETECTION

Ana P. B. Lopes[1,2]*, Sandra E. F. de Avila*[1]*, Anderson N. A. Peixoto*[1]
Rodrigo S. Oliveira[1] *and Arnaldo de A. Araújo*[1]

[1]Computer Science Department – Federal University of Minas Gerais
Av. Antônio Carlos, 6627, Pampulha, CEP 31270–901, Belo Horizonte, MG, Brazil

[2]Exact and Technological Sciences Department – State University of Santa Cruz
Rodovia Ilhéus-Itabuna, km 16 – Pavilhão Jorge Amado, CEP 45600-000, Ilhéus, BA, Brazil

ABSTRACT

Most of previous papers about the detection of nude or pornographic images start by the application of a skin detector followed by some kind of shape or geometric modeling. In this work, these two steps are avoided by a bag-of-features (BOF) approach, in which images are represented by histograms of sparse visual descriptors. BOF approaches have been applied successfully to object recognition tasks, but most descriptors used in that case are based on gray level information. Our approach is based on an extension to the well-known SIFT descriptor – called Hue-SIFT – aimed at adding color information to the original SIFT. Experimental results show recognition rates which are similar to those achieved by other approaches in literature, without the need for sophisticated skin or shape models.

1. INTRODUCTION

Filtering improper visual content coming from the web is a concern in several environments, from homes with children to workplaces. Textual cues are clearly not enough, since inappropriate images can maliciously be mixed to seemingly innocent text. A typical situation would be, for example, to have some search keywords commonly used by children tied to sites with pornographic content.

The ability to filter improper images by visual content instead of text has been the focus of several papers in the last years. Most of them start by detecting skin regions, and then applying some kind of shape or geometric analysis to recognize possible body postures which would be indicative of nudity or pornography. The main problems with those proposals are the complexity of accurate skin detectors and the great variability in shapes and geometry of such images. Moreover, approaches based on skin detection are bound to fail when applied to monochromatic images.

In a sense, the task of nude detection can be seen as an object recognition task. Bag-of-features approaches have been successfully applied to visual recognition or classification [1, 16]. In such approaches, images are represented as histograms built from a sparse set of visual features. No explicit object model is needed, and variability – of shape, scale or illumination, for example – can be addressed by a training set that covers that variability. These characteristics of BOF approaches make them specially suitable for nude detection on web images.

Nevertheless, typical features used to build BOFs – like brightness gradients or the widely known Scale-Invariant Feature Transform (SIFT) descriptor [11] – are mostly computed on gray level information. Their direct application would therefore disregard one of the strongest hints available to find most nude images, which is the presence of skin shades.

In this paper, we propose a BOF-based approach to detect images containing nudity, this time using the Hue-SIFT descriptor, a SIFT extension which includes color information. Experimental results show that this approach is indeed able to distinguish between nude and non-nude images from the web. The main advantage of such approach is that it does not rely on skin detectors or shape modeling. Besides being simpler, it is naturally more generic.

This paper proceeds like this: in Section 2, some related papers are described; in Section 3, our proposition for nude detection is detailed; in Section 4, experimental results are presented and discussed; finally, in Section 5, the paper is concluded.

2. RELATED WORK

The pioneer work for identifying adult images by the analysis of image content is proposed by [6]. 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 [8], a statistical color model for detecting skin and non-skin regions is developed. The set of aggregated features used for adult image identification includes percentage of detected skin pixels, average probability of the skin pixels, size in pixels of the largest connected component in skin regions, number of connected components in skin regions, percentage of colors with no entries in the skin and non-skin histograms, and the height and width of the image.

A system called Image Guarder aimed at detecting adult image content is presented in [18]. To speed up recognition, a two-layer recognition framework is adopted. In the first layer, an illumination adaptive statistical color model is proposed to detect the skin pixels under variant illumination. Only the images that have enough skin color pixels and smooth texture have their color, texture and shape features extracted and submitted to a Support Vector Machines (SVM) classifier.

In [19], an adult image filter and a harmful symbol filter are used to block objectionable images. In adult image filter, a statistical model is adopted for skin detection and a neural network is used for adult image classification. In harmful symbol filter, an edge based Zernike moments method is presented, which can capture the shape feature of symbol object effectively.

In [15], another framework for pornographic image detection based on skin region information is presented. Their approach extracts color and texture features from arbitrary-shaped segmented regions. Then, Gaussian Mixture Models (GMM) are built for skin and non-skin region classification, and the skin map is produced based on the classification result. Finally, eigenregion features (i.e., geometrical features that encompass area, location, and shape properties of the skin region) are used to describe the layout of skin regions on the whole image and pornographic images are detected according to the skin modality.

To overcome the chromatic deviation coming from the unusual lighting conditions, [9] proposed an online skin tone sampling mechanism based on face detection. Three geometrical primitives including area, position, and shape are derived from skin areas as input to a back propagation neural network. Also, a procedure for excluding faces shots is utilized to enhance system performance.

To detect pornographic images, [20] 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. Then, a texture model based on Gray Level Co-Matrix (GLCM) and geometric structure of human beings are used to deal with background regions similar to skin. A combination of constraints on color, texture, and geometric properties are used as features fed to a SVM classifier.

The first step of the system proposed in [10] 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, established by [5] 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 [17], the Adult Image Retrieval and Rating System (AIRS) is presented. AIRS uses a combination of MPEG-7 visual descriptors: the edge histogram descriptor, the color layout descriptor and the homogeneous texture descriptor. Given a query input image, the ten most similar images are retrieved from a database containing both adult and non-adult images. If most of the retrieved images are adult, the query input is regarded as an adult image.

Also in [12], a content-based image retrieval technique is employed for adult image identification. Firstly, background is removed to obtain a rectangular region of interest based on the detection of skin-like pixels. For each input image, the MPEG-7's color, texture, and a proposed shape feature are used to retrieve most similar images from the image database. If the retrieved images contain more than a threshold number of adult images, the input image is identified as an adult one. Otherwise, it is identified as non-adult.

In [2], the SNIF (Simple Nude Image Finder) is presented, which uses a color-based feature only, extracted by an algorithm for image description, the Border/Interior pixel Classification (BIC). SVM is applied for classification.

A color feature representation, which takes the correlation among neighboring components of the conventional color histogram into account and removes the redundant information, is proposed by [14]. To evaluate the performance of the proposed color feature, experiments of nude image detection using SVM and Adaboost algorithm are carried out.

Visual features can also be combined with other feature types, as in the framework for recognizing pornographic web pages presented in [7], in which text and image are both analyzed. In their image classifier, contour-based features are extracted to recognize pornographic images. Text and images features are combined with a fusion algorithm based on Bayes theory.

In [21], a framework for recognizing pornographic movies by fusing the audio and video information is described.

In [4], a method to classify images into different categories of pornographic content is presented, which is based on what is called bag-of-visual-words. This is another common denomination for BOF when applied to visual material. Their proposal is similar to ours, but the features used to build the vocabulary are simply patches around interest points. To the best of our knowledge, this is the only other paper using BOF for nude detection until now.

In this work, we propose a BOF-based approach using a more sophisticated point descriptor – called Hue-SIFT [13] – to capture color information. An extensive experimentation was performed to find the best vocabulary size and SVM model to classify nude and non-nude images using these descriptors. Also, a comparison with a pure SIFT approach is provided, in order to best evaluate the role of color in nude detection.

3. A BAG-OF-FEATURES APPROACH BASED ON HUE-SIFT DESCRIPTOR

The general steps for building a bag-of-features (BOF) representation for images are depicted in Figure 1. It starts by selecting (step a) and then describing (step b) a set of points from the image. The selection can be done in varied ways, the most typical being the application of interest point selection algorithm. Such approach has the advantage of providing both the points and the descriptors together. Also, interest point detectors provide a much sparser set of points to work on, reducing computational complexity.

Typical descriptors normally have a large dimension and are therefore submitted to some dimensionality reduction technique, the most common being Principal Component Analysis (PCA). Feature reduction is not included in Figure 1 because it is not an essential step in the process.

Given a set of – reduced or not – descriptors, the next step (c) is to cluster those points to form the vocabulary by quantization. Then, each point selected in the image is associated to a "word" (i.e., a cluster) in the visual vocabulary (step d) and the BOF histogram counting the occurrences of every word is computed (step e).

Algorithm 1 summarizes this process. In it, the steps of selecting, describing and reducing descriptors dimension is called *feature extraction*. In our BOF implementation, points selection and description is performed by the Hue-SIFT algorithm, proposed by [13]. This descriptor is composed of a concatenation of a hue histogram with the widely used SIFT descriptor [11], this way adding color information to

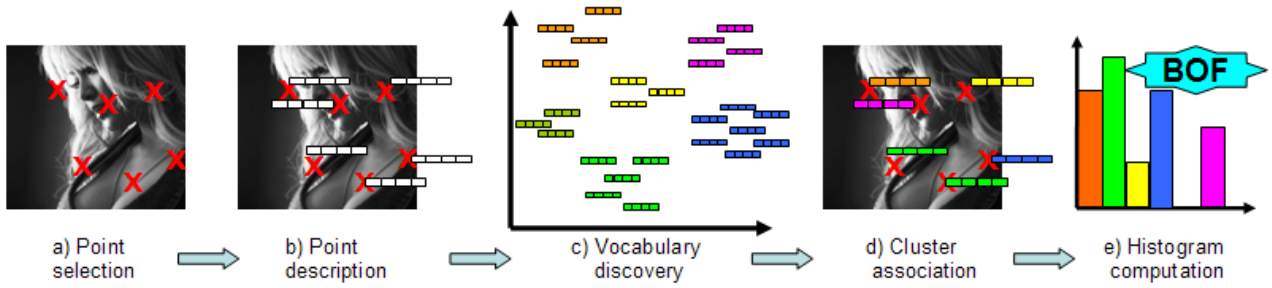

Figure 1: Steps for creating a bag-of-features for images or videos. Details on each step are provided in the text.

Algorithm 1 Pseudo-code for building a BOF representation for images.

```
for all image do
    pointsDesc ← featureExtraction(image)
end for
vocab ← quantizePoints(pointsDesc)
for all image do
    bofs ← computeHistogram(vocab, pointsDesc)
end for

FUNC featureExtraction(image), return descriptors
pointsPos ← selectPoints(image)
descriptors ← describePoints(image, pointsPos)
descriptors ← reduceDimension(descriptors)
return descriptors
```

SIFT. The Hue-SIFT descriptor is scale-invariant and shift-invariant, similar to the hue histogram, which is made by weighing each sample of the hue by its saturation. However, only the SIFT component of this descriptor is invariant to illumination, color changes or shifts; the hue histogram is not. Nevertheless, our experiments show that the classifier is able to capture nudity in different illumination situations, if a similarly diverse set is provided for training. Some examples from our database are provided in Figure 2.

Dimensionality reduction is achieved by PCA and quantization is performed with k-means. For classification, a linear kernel for SVM was preferred, fine-tuned according to the procedure proposed in the documentation of LIBSVM [3] software.

4. EXPERIMENTAL RESULTS

4.1 Experimental Setup

To compose a database for the evaluation of the proposed approach, a set of 180 images was collected from the web, and then manually classified. Examples of selected images are shown in Figure 2. SIFT and Hue-SIFT descriptors were extracted from all images, and then the vocabulary size and SVM model were extensively searched for, using a 5-fold cross validation scheme. For the vocabulary size (k), the experiments spawned values between 50 and 700. For each k value, the penalty of the SVM error term (C) was varied in a logarithmic scale from 10^{-5} to 10^{5}. The k and C values which achieved the best recognition rate were submitted to a finer search for C. This procedure was repeated for SIFT and Hue-SIFT separately.

In order to evaluate the statistical significance of the differences found between the best recognition rates for SIFT and Hue-SIFT, ten new runs of the 5-fold cross validation were performed with fixed k and C. Folds composition changed randomly in each run, keeping the balance between nude and non-nude training examples. The average rates of each cross-validation run were used to compute confidence intervals for both results. Intermediary results for all these computations can be found in http://www.npdi.dcc.ufmg.br/nudedetection/. Their summary and discussions are provided in the following subsection.

4.2 Results and Discussion

In Figure 3, the best recognition rates for all tested values of k are plotted both for Hue-SIFT and SIFT alone. The best recognition rate is the higher value achieved while varying penalty error term of the SVM model (C parameter). It can be seen that SIFT recognition rates are consistently smaller for all tested vocabulary sizes, indicating the importance of color information.

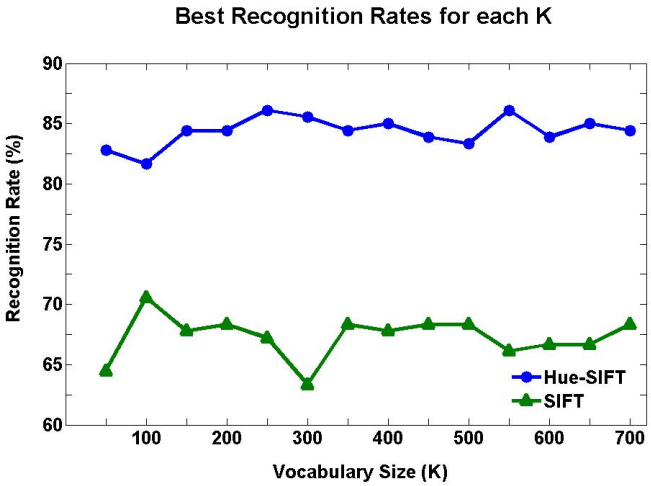

Figure 3: Best recognition rates achieved for each vocabulary size on the model selection experiment.

Refinements were performed on the peaks of this graph – two for Hue-SIFT (at $k = 250$ and $k = 550$) and one for SIFT (at $k = 100$) – to find the best SVM model, which was used in the final experiment. The average recognition rates for the ten runs of the 5-fold cross-validation experiment are in Table 1.

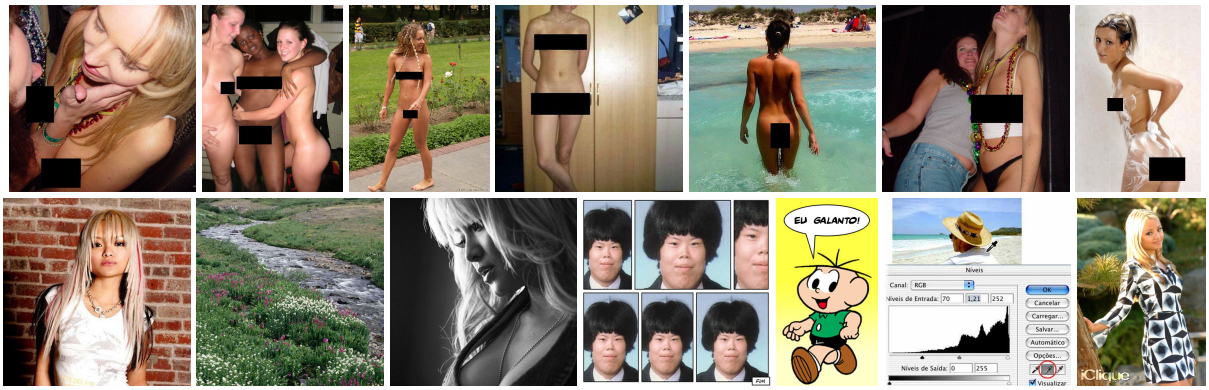

Figure 2: Examples of nude (first row) and non-nude images (second row) from our database.

Best Recognition Rates			
Descriptor	**Average (%)**	**Min. (%)**	**Max. (%)**
SIFT	64.8	61.7	68.0
Hue-SIFT	84.6	82.8	86.5

Table 2: Best recognition rates with Hue-SIFT and SIFT descriptors. Confidence intervals (given by "Min." and "Max." values) have a confidence level of 95%.

Cross-validation Averages		
Run	**SIFT (%)**	**Hue-SIFT (%)**
1	70.0	81.7
2	61.1	87.8
3	66.7	83.3
4	73.9	85.6
5	61.1	87.2
6	61.1	87.2
7	66.1	82.8
8	60.0	81.7
10	63.9	80.0

Table 1: Average recognition rates for all 5-fold cross validation runs.

In Table 2, it is possible to see the final averages and their confidence intervals. These results indicate clearly the importance of color to distinguish between nude and non-nude images, which is in accordance with the basic implicit assumption of most skin-based proposals.

Unfortunately, the lack of standard annotated databases for nude or pornographic images makes it difficult to compare results obtained by different approaches. Nevertheless, it is worth to mention that in a coarse comparison, our approach achieved an overall recognition rate very similar to those ones found in recent literature. This was achieved without the usage of any skin and/or shape models. Additionally, we made use of the simplest possible SVM classifier. This result suggests that BOF-based approaches are indeed a promising path to follow in the pursuit of effective filters for improper visual content.

5. CONCLUDING REMARKS

In this work, we propose a bag-of-features (BOF) approach for the classification of improper visual content collected from the web. This approach has as its main advantage the fact that it does not depend on any skin or shape models to identify nudity. Despite of not using such models, and even using a linear SVM classifier, our results achieved good recognition rates, which are comparable to the ones achieved for more elaborated models.

Future work includes testing on larger databases, adding structural and scale information to the basic BOF representation, experimenting with more sophisticated classifiers and evaluating a similar scheme for nude detection on videos.

6. ACKNOWLEDGMENTS

The authors are grateful to CNPq, CAPES and FAPEMIG, Brazilian research funding agencies, for the financial support to this work.

References

[1] 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*, 26(11):1475–1490, 2004. Member-Dan Roth.

[2] R. J. S. Belem, J. M. B., E. S. de Moura Cavalcanti, and M. A. Nascimento. SNIF: A simple nude image finder. In *Proceedings of the Third Latin American Web Congress (LA-WEB)*, pages 252–258, Washington, DC, USA, 2005. IEEE Computer Society.

[3] 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.

[4] 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.

[5] L. Duan, G. Cui, W. Gao, and H. Zhang. A hierarchical method for nude image filtering. *Journal*

of Computer-Aided Design and Computer Graphics(in Chinese), 14(5):404–409, 2002.

[6] 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)*, pages 593–602, London, UK, 1996. Springer-Verlag.

[7] 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*, 29(6):1019–1034, 2007.

[8] 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)*, 1:1274–1280, 1999.

[9] 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)*, pages 2566–2571. IEEE, July 2006.

[10] 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)*, 0:487–488, 2008.

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

[12] J.-L. Shih, C.-H. Lee, and C.-S. Yang. An adult image identification system employing image retrieval technique. *Pattern Recognition Letters*, 28(16):2367–2374, 2007.

[13] 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.

[14] 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)*, volume 6, pages 353–356. IEEE, October 2007.

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

[16] 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)*, pages 197–206, New York, NY, USA, 2007. ACM.

[17] 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)*, volume 3213 of *Lecture Notes in Computer Science*, pages 164–170. Springer, 2004.

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

[19] 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)*, pages 1223–1226, June 2004.

[20] 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)*, pages 801–804, Washington, USA, 2007. IEEE Computer Society.

[21] 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)*, pages 37–40, April 2008.