SNIF: A Simple Nude Image Finder

Ruan J.S. Belém
Federal University of Amazonas
Department of Computing Science
Manaus, Brazil
ruan@ufam.edu.br

João M.B. Cavalcanti
Federal University of Amazonas
Department of Computing Science
Manaus, Brazil
john@dcc.ufam.edu.br

Edleno S. de Moura
Federal University of Amazonas
Department of Computing Science
Manaus, Brazil
edleno@dcc.ufam.edu.br

Mario A. Nascimento
University of Alberta
Department of Computer Science
Edmonton, AB, Canada
mn@cs.ualberta.ca

Abstract

The lack of control of the content published is broadly regarded as a positive aspect of the Web, assuring freedom of speech to its users. On the other hand, there is also a lack of control of the content accessed by users when browsing Web pages. In some situations this lack of control may be undesired. For instance, parents may not desire their children to have access to offensive content available on the Web. In particular, accessing Web pages with nude images is among the most common problem of this sort. One way to tackle this problem is by using automated offensive image detection algorithms which can filter undesired images. Recent approaches on nude image detection use a combination of features based on color, texture, shape and other low level features in order to describe the image content. These features are then used by a classifier which is able to detect offensive images accordingly. In this paper we propose SNIF - Simple Nude Image Finder - which uses a color based feature only, extracted by an effective and efficient algorithm for image description, the Border/Interior pixel Classification (BIC), combined with a machine learning technique, namely Support Vector Machines (SVM). SNIF uses a simpler feature model when compared to previously proposed methods, which makes it a fast image classifier. The experiments carried out depict that the proposed method, despite its simplicity, is capable to identify up to 98% of nude images from the test set. This indicates that SNIF is as effective as previously proposed methods for detecting nude images.

1. Introduction

The Web is a wealthy source of information with different purposes and applications. The popularization of digital cameras and mobile phones, among other devices capable of producing digital images, combined with little difficulty for online publishing are largely responsible for the large amount of multimedia content available on the Web.

The lack of control of the content published on the Web is a positive aspect, assuring freedom of speech. However this can also be a problem in some situations. For instance, offensive content accessed inadvertently by children or non-authorized content viewed by employees in companies or other institutions [10]. While offensive content can present itself in a number of different formats and media, offensive images are particularly problematic as they can be easily embedded on Web pages. Since Web browsers are typically set up to automatically display images, they can be seen involuntarily.

Here we address the problem of finding images related to pornographic content, which will be referred as nude images. One way to tackle this problem is by using automated nude image detection algorithms capable of filtering and/or alerting about the existence of nude images. This can also be regarded as a classification problem [8, Ch. 7]. There are two types of classification approaches, namely supervised and unsupervised classification. Recent supervised classification approaches for nude image detection [13, 15] use a combination of features based on color, texture, shape and other low level features to describe image content and perform comparisons between those features and features extracted from an image training set. Most techniques referred here achieve good accuracy results (higher than 90%). How-

Proceedings of the Third Latin American Web Congress (LA-WEB'05)
0-7695-2471-0/05 $20.00 © 2005 IEEE

ever, the use of many features for image description likely yield an undesired increase in complexity and processing time.

In this paper we propose SNIF - Simple Nude Image Finder, an approach for nude image detection and filtering. Our solution reuses a simple but yet effective and efficient technique for image description named Border/Interior pixel Classification (BIC) [12], which uses only color information of images, combined with a well known machine learning classification technique, namely Support Vector Machines (SVM) [6]. BIC and SVM combined allow our method to achieve similar results to the techniques described in the literature but in a much simpler and therefore more efficient way.

There are many possible applications for this approach on the web, such as filters for Web portals and proxy servers. This sort of technology is used by UOL[1] (censorship) and Google[2] (SafeSearch) on its image search engines.

The next section provides an overview of some recent approaches on nude image detection. The SNIF components and the classification process are explained in Section 3. Section 4 presents the details and a discussion on the the experiments carried out. Finally the conclusions are presented in Section 5.

2. Related Work

In this section we discuss some existing detection methods of nude pictures. In [13], it was proposed two evaluation metrics which are adopted in most of the previous related work: *sensitivity* and *specificity*. Sensitivity denotes the ratio of the nude images identified to the total number of nude images (and is similar to the concept of recall for nude images) whereas specificity denotes the ratio of non-nude images identified to the total number of hits on non-nude images (which is similar to the recall for nun-nude images). Besides these two metrics, we here will also adopt the accuracy, which is usually adopted to evaluate the quality of results in classification problems [2].

Forsyth and Fleck approach [7] is based on the combination of color and texture properties to identify skin regions in images. The algorithm uses geometric constraints based on the human structure trying to group skin regions. As described in [7] the obtained results were 43% of sensitivity and 96% of specificity for a test set with 565 nude images and 4289 non-nude images.

Wang et al [13] proposed the Wavelet Image Pornography Elimination (WIPE $_{TM}$) which uses a combination of filters organized in a pipeline. In this approach faster filters are used first in order to pass non-nude images more quickly and to reduce the overall classification time. After the query image passes all initial filters, the algorithm produces feature vectors using moment analysis, texture analysis, color histogram analysis and statistics to match the query image with a pre-marked training data set containing 500 nude images and 8,000 non-nude images. A query image is classified based on the 15 closest matches in the training database. This system processes images at a rate close to 0.5 per second on a Pentium Pro PC with 96% of sensitivity and 91% of specificity. The experiments carried out used a test set of 1,076 objectionable images and 10,809 non-offensive images.

Arentz and Olstad presented in [1] a method to classify adult Web sites based on image content. The image classifier algorithm starts with a filter that removes non-skin pixels from an image by rejecting pixels outside a given range previously defined. The remaining pixels are grouped into objects composed by connected regions of pixels. The objects are labeled, and their labels and number of pixels are stored in distinct arrays. These arrays are sorted out so that the largest objects are analyzed to extract feature vectors composed by color, texture, shape, size and placement of each object. The importance of the different features is then computed by a genetic algorithm. The probability of an image to be nude is calculated based on the gathered information. The test data was composed by 500 nude and 800 non-nude images. Experimental results presented an average of 11 images processed in a second with 89,4% of accuracy, 95% of sensitivity and 88% of specificity.

Zhu et al [15] proposed an adaptive skin-detection method and analyzed the impact of the proposed algorithm applying it in a nude image image classification. The skin detection is performed in two stages. First, a generic skin model is used to identify the image skin-similar space, which contains skin pixels and some non-skin pixels due to overlap in color space. In a second stage, the standard Expectation-Maximization (EM) algorithm [3] is used to derivate a Gaussian Mixture Model (GMM) specific to the considered image. A Support Vector Machine (SVM) classifier uses shape and spatial information of the skin pixels to identify skin Gaussian in the skin-similar space.

This method was used to replace the generic skin model used in MORF [14], a personalized information system that helps the users to avoid objectionable information by filtering offensive content. One of the main components of MORF is the image classifier, based on a SVM classifier where each image is converted in a vector with 144 dimensions, representing information about color histograms, means, variances, enlogation and spreadness, extracted from 12 natural colors channels and texture features in three orientations extracted at different resolutions. The same image training set is used by the

1 `http://www.uol.com.br`
2 `http://www.google.com`

multiple classifiers, but with different features. The classification procedure is performed in two levels. In the first level, a classifier verifies if the SVM score is inside a previously defined range and, in this case, the two classifiers of the second level are used and the final result is given by the majority vote of the three classifiers. If the score is out of range the classification finishes in the first level. Experiments carried out used 13,500 offensive images and 13,500 non-offensive images for training. Tests used 1,500 offensive and 1,500 non-offensive images, obtaining 94.63% of accuracy.

The main difference between the previous proposals and our approach is the simplicity. For instance, the previous proposals use a larger number of features, which results in more complexity and processing time. The most similar work is MORF with the adaptive skin-detection algorithm [15], which pre-processes the image in order to detect skin and uses different features to represent an image, while our approach uses a single feature, namely color. Further, we using vectors of 64 dimensions to represent an image, we obtained almost the same accuracy obtained by MORF using vectors of 144 dimensions. That means it is possible to detect nude images with high accuracy using less computational efforts than proposed in previous work.

3. SNIF

The combination of a set of features is the basis of most recent approaches for identifying nude pictures. According to Arentz and Olstad [1], the key element in identifying nudity is finding skin regions on images. Despite these evidences, in our approach we try to obtain good performance using only the color feature, extracted with a very efficient algorithm combined with a robust classifier based on machine learning. Our assumption is that as simpler the method is, better is its performance. Next we present details about the SNIF components and the classification process.

3.1. BIC

In order to extract color features from digital images we use BIC (Border / Interior pixel Classification), an efficient approach for image content based retrieval [12]. BIC is composed of a simple image analysis algorithm, a logarithmic distance (*dLog*) and a compact representation of the visual feature extracted.

The image analysis rely on the RGB color-space quantized in Q colors. After quantization, each pixel is classified as (1) *border*, when it is located at the border of the image or if one of its 4-neighbors (top, bottom, left and right) has a different quantized color or (2) *interior* if they have the same color. Based on this classification, two color histograms are computed. One considers only the border pixels and the second refers to the interior pixels, each with Q bins. A bin contains initially the total amount of pixels normalized to integer values in the interval [0,255].

To avoid the problem of high values in a single histogram bin dominate the distance between two histograms, the dLog distance function is used. This function uses a logarithm scale reducing 28 times the range of distances between the smallest and the largest histogram bin values, given that in the log-scale a bin is normalized in the interval [0,9]. As a result of this process, we have only one histogram with $2 \times Q$ bins where each bin contains integer values between 0 and 9. Notice that using this histogram, each image can be represented through a vector where each position represents a bin. This histogram Figure 1 shows an example of an image and its corresponding border and interior pixels images obtained with BIC.

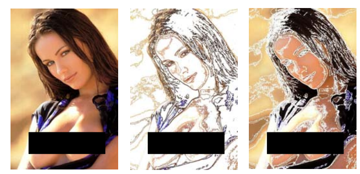

Figure 1. BIC: Original image, border image, interior image

Note that a large amount of interior pixels refer to areas of human skin. This occurs to the vast majority of nude images and is determinant information for detecting offensive images.

3.2. SVM Classifier

Support Vector Machines (SVM) [6] is a supervised machine learning technique for classification, where labeled examples are necessary to build a model. In this method, each element is represented by a vector which summarizes its main characteristics by numeric values. These element vectors are the input for SVM and are mapped into a space. A linear decision function is defined by the optimal hyperplane constructed in this feature space. The margin between the vectors of the classes are also maximized to allow the generalization. The vectors that define the maximal margins are called support vectors. These elements are exemplified in Figure 2, where squares and circles represent vectors of class members in a linear separable problem. Only 2 dimensions were used to simplify the example. Finally, the class of unlabeled feature vectors can be predicted applying the learned model. Further details about SVM can be found in [6].

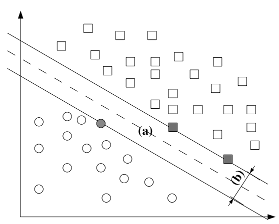

Figure 2. SVM separating two class in a 2-dimensional space. Filled squares and circles are the support vectors. (a) optimal hyperplane (b) optimal margin

Two parameters must be considered when we work with SVM classifiers, the kernel function and the regularization parameter C [11]. The kernel function computes the similarity of two element vectors and is also used to map them into the space created by SVM. It allows separating the element vectors in different classes, what is frequently not possible in a straightforward representation of the input space defined by the element vectors. The basic kernel types are linear, polynomial, sigmoid and radial basis function (RBF) [4]. The regularization parameter is used to adjust rigidity of the optimization procedure. The value must be balanced to prevent poor accuracy, when very low values are selected or overfitting, in the case of very high values. In [9] arguments are presented to consider RBF kernel as the most indicated choice when the user is not familiarized with all SVM details. This kernel type requires only two parameters: the regularization parameter C and a function parameter γ. More detailed information about SVM parameters can be found in [9].

We use BIC to extract from each image a vector where each position represents a bin. Based on the image vectors extracted by BIC, SVM is able to "learn" what sort of features belong to nude images and which ones refer to non-nude images. A training image set is used to perform this learning phase. After training other images can be classified accordingly. Given that other nude image detection methods also use some classification procedure with similar algorithms complexity, the performance gain of our approach rely on the BIC algorithm.

3.3. How SNIF Works

The classification process, as showed in Figure 3, is divided in three phases:

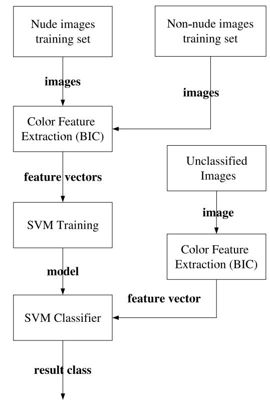

Figure 3. An overview of the SNIF process

1. *Feature Extraction*: The content feature extraction is performed by BIC as described earlier. The output format of BIC, a $2 \times Q$ bin histogram, is easily adapted to serve as input to the SVM classifier. Another important characteristic is memory space savings obtained by this method. As each bin contains an integer value in the interval [0-9], two values can be stored in one byte. Thus, the content information of one image occupies only Q bytes of storage space. Experiments described in Section 4 shows that the border pixels histogram can even be discarded without considerable decrease on the accuracy obtained by the classifier. This halves the storage requirement, and as a result the histogram used has $Q/2$ bins and memory space required is reduced to just $Q/4$ bytes. In particular we have obtained good results using 128 quantized colors, which requires only 32 bytes/image of storage space.
2. *Training*: During the training phase, a pre-classified image set is processed by BIC and the resulting output is converted in feature vectors. Using positive (nude) and negative (non-nude) examples of images, SVM is trained and produces a model that will be validated in the next phase.
3. *Test and classification*: In this phase, an independent image test set is classified in order to evaluate accuracy and validate the model. After validation, the vector representation extracted from unclassified images is submitted to the SVM classifier that predicts its class.

4. Experiments

Both training and test sets were extracted from Web pages collected using the Cade[3] Internet directory. Nude

Colors	Specificity	Sensitivity	Accuracy
16	83.8%	94.4%	88.8%
32	84.7%	96.3%	90.1%
64	89.8%	96.4%	93.2%
128	91.0%	97.6%	94.4%
256	85.6%	97.5%	91.1%

Table 1. Results for nude image classification using different color quantizations

images were extracted from erotic categories and non-nude images from categories as sports, including aquatic sports, arts, animals and others with clothed people pictures. In total 1,000 nude images and 1,000 non-nude images were used as training set and the test was composed by 1,000 nude images and 1,135 non-nude images. All these pictures were manually selected and evaluated. In this evaluation, it was considered as nude images, those portraying people wearing no clothing or wearing significantly less clothing than expected by the conventions of the local culture and exposing the bare skin of intimate parts.

The color feature was extracted from all images using the BIC algorithm, where Q was set to 128, and the obtained histogram converted to be used as input to the SVM classifier. Our SVM classifier was implemented using LIBSVM [5]. Two files containing the class labels and the image vectors for each image were generated, one for the training set and the other for the test set.

During the training phase, the model selection was done by cross validation and using a RBF kernel. In a cross validation, the training set is divided into subsets of equal size and each one is tested using the classifier trained with all the remaining subsets. For each division adopted we perform experiments with different parameters, varying C and γ. At the end, the best result found for C and γ was adopted as the parameters for the experiments using the whole training set.

A first experiment set was carried out with the objective of verifying how color quantization affects the classification results and, therefore, identifying the most appropriate color quantization. Tests with different color quantizations (16, 32, 64 128 and 256 colors) were carried out, considering both the border and interior pixel histograms. For each color quantization we executed the classifier, using LIBSVM to find the best parameters for C and γ. Results are presented in Table 1. As can be seen the best results were achieved when using 128 colors. It is interesting to notice that results are worse than 128 when using 256 colors, which gives more color information.

3 `http://www.cade.com.br`

Colors	Specificity	Sensitivity	Accuracy
32i	87.0%	94.9%	91.0%
128i	87.4%	98.0%	92.5%

Table 2. Nude image classification using only the interior pixels

In another experiment, we have investigated the possibility of discarding part of color information reducing memory space required for classifying images, while keeping acceptable image classification accuracy. Based on empirical observation, we realized that that the majority of skin pixels are located on the interior pixels. Therefore, discarding the border pixel histogram, the classification results would probably remain stable.

To confirm this, the BIC pixel classifier was modified to consider a proximity threshold between colors, instead of an exact match as originally proposed in [12]. As a result, a pixel is classified as border if one of its 4 neighbors is of a color out of this threshold. Different threshold values (10, 20, 30) were experimented, showing that the nude image recall decreases as the threshold increases. This confirms that all the border pixels identified by the standard BIC algorithm can be discarded. As only the interior pixel histogram is used, half of the original storage space is saved. This characteristic is useful to reduce the time to compute the model, which depends on the number of dimensions in the vectors. Further, it is also useful for reducing the memory requirements when classifying the images, since vectors are represented with half size.

We carried out tests with 32 and 128 colors considering only the interior pixel histogram. The results are presented in Table 2, where the letter "i" beside the quantization number means only interior pixels were used. Comparing these results with the ones presented in Table 1 we notice that using the interior pixels we achieved an improvement in the sensitivity. For instance, looking at the results for 128i compared against the 64 colors method, we achieved an improvement of 1.6% in sensitivity using the same storage space. In some applications, it is important to achieve higher sensitivity values because it minimizes the number of nude images that will be presented as non-nude for the users. The drawback of using just interior pixels was that specificity was reduced, which means more non-nude images would not be presented to the users due to a wrong conclusion that they are nude. Therefore, decision about using only interior pixels will depend on the final application given to the method. Table 2 also shows that sensitivity for 128i was slight better than sensitivity for 128, which means using interior pixels we can reach higher recall values.

Finally, we have carried out experiments testing different image class proportion in the training set. This experiment aims at verifying how the classification accuracy is related to the proportion of positive (nude) and negative (non-nude) image examples in the training set. A 128 color quantization was used, discarding the border pixels histogram (128i). Three different configurations are tested with the respectively proportions of nude and non-nude images: (75%-25%), (50%-50%) and (25%-75%).

The results in Table 3 show that even using relatively few positive (nude images) examples it is possible to get good results. Our assumption is that this occurs due to the fact that the class of nude images is well defined. On the other hand, the class of non-nude images is actually a group including very different types of images.

	(75%-25%)	(50%-50%)	(25%-75%)
Sensitivity	98.6%	98.0%	97.8%
Specificity	81.67%	87.58%	87.49%
Accuracy	86.5%	92.5%	91.8%

Table 3. Sensitivity and specificity achieved with different proportions used in the image training set

4.1. Evaluation and Discussion

One of the main results, which can be noted in Table 1, is that even using only 16 colors the sensitivity and specificity remain acceptable. Note that using 16 color quantization storage space requirements are reduced and performance is improved.

The experiment configurations using quantization with 128 colors present higher values of accuracy, being also the best in specificity and almost the best in sensitivity. The configuration with only interior pixels is slightly better in identifying nude pictures, however the specificity value decreases a little in relation to other configurations. If the main objective of possible applications of this method is the detection of nude images, this high values of sensitivity (nude image recall) is a better option when compared against to the options using all the pixels.

The achieved results are compatible with the results of related work presented in Section 2 as can be observed in Table 4, where two SNIF results are presented, one considering both histograms of border and interior pixels (SNIF-128) and the other with only the interior pixel histogram (SNIF-128i). The others results are from Forsyth and Fleck [7], Wang et al [13], Arentz and Olstad [1] and Zhu et al [15]. This is an indirect comparison because the nude nature of the images makes very difficult to have access to the image test set used in other work. The commercial characteristic of the possible applications of this type of classifier/detector also makes difficult the access to the algorithms and implementations. Further, details presented in related work turns the implementation a difficult task. An important contribution of this work is that SNIF is quite simple and implementation details were not omitted here, which means it can be easily adopted as a baseline in future work.

Method	Specificity	Sensitivity	Accuracy
Forsyth and Fleck	96.0%	43.0%	-
Wang et al	91.0%	96.0%	-
Arentz and Olstad	88.0%	95.0%	89.4%
Zhu et al	-	-	94.6%
SNIF-128	91.5%	97.6%	94.4%
SNIF-128i	87.6%	98.0%	92.5%

Table 4. Comparison between SNIF and methods previously proposed in literature

Most of the non-nude images falsely classified as nude image are of portrait type containing large areas of skin color. In case of nude images, the few nude images wrongly classified as non-nude contained many blue color pixels, usually depicting people in a beech or swimming pool. This happens due to the presence of swimming images in the non-nude image training set. In future work, the aggregation of text information can increase accuracy on non-nude image classification because generally, the context of this type of image is different of the document (Web page) context of nude images.

5. Conclusion

In this paper we presented SNIF, a method for nude image detection. Ours is a simple but efficient method based on color information to describe an image. The BIC algorithm was used to extract this information resulting in two histograms: one for border pixels and the other for interior pixels. This information is coded in feature vectors which are used by SVM to perform training and classification.

The simplicity of our approach comes from the BIC algorithm, which is capable of extracting color information efficiently. Although our method uses only one feature (color information) whereas most other work use several features, the results obtained in the experiments carried out have shown that accuracy, sensitivity and specificity are compatible with all related work presented in Section 2.

This confirmed our assumption that a simple approach for image feature extraction is enough for achieving good results. In addition, other experiments have also demonstrated that BIC can be used in a more simpler way, as discarding the border pixel histogram. As a result, memory space needs are reduced for storing the feature vector and the classification procedure can be processed more efficiently.

Acknowledgment

This work was partially supported by the GERINDO project - grant MCT/CNPq/CT-INFO 552.087/02-5. Ruan Belém is supported by the Amazonas State Government through FAPEAM grant 386/2003. Edleno Moura is supported by CNPq Individual grant 303576/2004-9. Mario A. Nascimento was supported in part by NSERC Canada and the Canadian Cultural Content Management Research Network, a Network financed through Heritage Canada's New Media Research Networks Fund.

References

[1] W. A. Arentz and B. Olstad. Classifying offensive sites based on image content. *Comput. Vis. Image Underst.*, 94(1-3):295–310, 2004.

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

[3] J. Bilmes. A gentle tutorial on the em algorithm and its application to parameter estimation for gaussian mixture and hidden markov models, 1997.

[4] C. J. C. Burges. A tutorial on support vector machines for pattern recognition. *Data Mining and Knowledge Discovery*, 2(2):121–167, 1998.

[5] C.-C. Chang and C.-J. Lin. *LIBSVM: a library for support vector machines*, 2001. Software disponvel em `http://www.csie.ntu.edu.tw/~cjlin/libsvm`.

[6] C. Cortes and V. Vapnik. Support-vector networks. *Machine Learning*, 20(3):273–297, 1995.

[7] D. A. Forsyth and M. M. Fleck. Identifying nude pictures. In *Proceedings of the 3rd IEEE Workshop on Applications of Computer Vision (WACV '96)*, page 103. IEEE Computer Society, 1996.

[8] J. Han and M. Kamber. *Data Mining: Concepts and Techniques)*. Morgan Kauffman, 2000.

[9] C.-W. Hsu, C.-C. Chang, and C.-J. Lin. A practical guide to support vector classification.

[10] M. L. Kherfi, D. Ziou, and A. Bernardi. Image retrieval from the world wide web: Issues, techniques, and systems. *ACM Comput. Surv.*, 36(1):35–67, 2004.

[11] D. Mazzoni, K. Wagstaff, and R. Castano. Using trained pixel classifiers to select images of iterest. In *The Interplanetary Network Progress Report*.

[12] R. O. Stehling, M. A. Nascimento, and A. X. Falcao. A compact and efficient image retrieval approach based on border/interior pixel classification. In *Proceedings of the eleventh international conference on Information and knowledge management*, pages 102–109. ACM Press, 2002.

[13] J. Wang, J. Li, G. Wiederhold, and O. Firschein. System for screening objectionable images. *Computer Communications Journal*, 1998.

[14] Y.-L. Wu, E. Y. Chang, K.-T. Cheng, C. Chang, C.-C. Hsu, W.-C. Lai, and C.-T. Wu. Morf: A distributed multimodal information filtering system. In *IEEE Pacific Rim Conference on Multimedia*, pages 279–286, 2002.

[15] 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*, pages 56–63. ACM Press, 2004.