Recognizing Pornographic Images

Sotiris Karavarsamis, Ioannis Pitas*
Artificial Intelligence & Information Analysis Lab
Department of Informatics
Aristotle University of Thessaloniki, Greece
{sokar, pitas}@aiia.csd.auth.gr

Nikos Ntarmos
Network-Centric Information Systems Lab
Computer Engineering & Informatics Dept.
University of Patras, Greece
ntarmos@ceid.upatras.gr

ABSTRACT

We present a novel algorithm for discriminating pornographic and assorted benign images, each categorized into semantic subclasses. The algorithm exploits connectedness and coherence properties in skin image regions in order to capture alarming Regions of Interest (ROIs). The technique to identify ROIs in an image employs a region-splitting scheme, in which the image plane is recursively partitioned into quadrants. Splitting is achieved by considering both the accumulation of skin pixels and texture coherence. This processing step is proven to significantly boost the accuracy and reduction of running time demands, even in the presence of sparse noise due to errors attributed to skin segmentation. For detected ROIs, we extract 15 rough color and spatial features computed from the pixels residing in the ROI. A novel classification scheme based on a tree-structured ensemble of strong Random Forest classifiers is also proposed. The method achieves competitive performance both in terms of response time and accuracy when compared to the state-of-the-art.

Categories and Subject Descriptors

I.5.4 [**Pattern Recognition**]: Applications—*Computer Vision*

Keywords

Porn image detection, skin segmentation, Gamma correction, region of interest, geometrical region splitting, texture analysis, random forest classifier

1. INTRODUCTION

In the last decade, the proliferation of pornographic images on the World Wide Web along with the systematic exposure of children to pornography have provided alarming statistical evidence about their negative impact on behavioral functioning, often being manifested by means of aggressive behavior. Furthermore, intense exposure to pornography by adults has been associated with specific problematic behavioral patterns. Moreover, exposure to pornographic material can occur involuntarily; e.g., it is very common for pornographic ads or utterances with references of pornographic nature to repetitively spawn without the consent of the user, or actually being presented without being requested explicitly. From a user experience point of view, dissemination and acquisition of explicit pornographic material is most commonly achieved by video and imagery. Text annotations can occur as a supplementary piece of meta-information aimed at achieving comprehension by the user. These often contain rich information about the thematic concept of a web page, and can be treated as useful prior knowledge (i.e. treated as indicator information on the nature of the underlying visual signals in imagery). We believe that an image is the primal carrier of pornographic information to engage the user's focus of attention. A recognition system attaining a pattern of high precision in predicting the correct labels of images in web pages can essentially provide reliable indication about the underlying semantic category of the content being presented. However, other useful features, e.g. those describing the structure (or template) of a web page, can be fused with the outputs provided by an accurate pornographic image detector in order to demonstrate stronger decision making in integrated content filtering systems.

Several works have focused on the problem of automatically identifying patterns in images indicating the presence of pornographic information. The first of these attempts was the seminal work by Forsyth et al. [2] attempting to discriminate porn images by identifying common naked human figure skeletons. The inferred skeleton obtained by analyzing the skin segmentation of an input image was used to predict the presence of either a pornographic scene or a general benign image. However, the required computational effort for identifying skeletons hindered any potential real-world applications. In a real-world content filtering challenge (e.g. image search), billions of images may routinely be required to be processed. Although the precision of the proposed method was high, it required approximately six minutes per input image. In contrast, the WIPE system proposed by Wang et al. [10, 11] was a complete framework to capture pornographic web sites based on the images they refer to. The image analysis core of this system employs a step-wise filtering pipeline: first, input images are screened as either general figures or thumbnails (or "icons"); next, if the image is not positive with respect to these cat-

*Corresponding author

egories, feature extraction based on wavelet and histogram analysis is employed. The image categorization process aims at determining whether the extracted image features match closely to either a positive or negative semantic class, based on training features stored in a feature database. The authors report promising results regarding the performance of their system, both in terms of generalization capabilities and user-perceived response time.

In this work, we extend our previous work for recognizing pornographic web pages by exclusively exploiting the visual signals in imagery[4]. The contribution of this work is twofold. First, we present a new fast and simplified geometrical technique of recursive design, inspired by the work by Yang et al. [12] on capturing and extracting Regions of Interest (ROIs) in pornographic images. Second, we propose (and experimentally evaluate) a novel multi-class tree-structured ensemble categorization scheme employing strong Random Forest classifiers [1] in the tree nodes. In this model, predictions are induced in a top-down and coarse-to-fine-grained manner, by walking a path from the root categorization node to a leaf node according to the decisions made by the intermediate nodes. The proposed technique is evaluated on a challenging manually collected dataset of 9,000 images obtained from the web. Our experimental results show that our solution attains high recognition accuracy and a significant reduction in processing time, compared to the state-of-the-art POESIA project[6].

The training set of porn and assorted benign images is structured hierarchically. The samples are first coarsely categorized (by human annotators) as either being pornographic or benign. Each of these classes is further split into two respective semantic subclasses. The porn class is split into a "porn-scene" class, which contains images depicting pornographic scenes involving one or more human subjects in naked postures or sexual intercourse. Additionally, a subclass of the pornographic category is the "bikini" class, which contains color images of human subjects (e.g., art models) wearing bikinis (or swimsuits, etc.), but having no "sensitive" body parts exposed. Considering these four class labels, we train and test the discriminative capabilities of the proposed detection method. For the coarse class of benign images, we further provide two subclasses considering the amount of true skin and falsely identified skin, commonly attributed to false alarms in skin detection techniques. The first subclass contains benign images with dressed or lightly dressed people (for instance, athletes, etc.) where some of the body parts (e.g. head, limbs, hands, waist, etc.) are being exposed. Last, we introduce a subclass which contains benign images with no true skin. In our setting, the coarse class labels are being referred to as "porn" and "benign", while the labels of the fine-grained subclasses are "porn scene", "bikini", "non-skin", and "skin", respectively.

2. IMAGE ANALYSIS & CLASSIFICATION

The image processing steps that illustrate our general approach are designed with two basic principles in mind: a) image analysis is required to exhibit robustness by relying only on rough visual features that demand low computational effort on their extraction, and b) the overall observed response of the pipeline in our recognition core should on average be able to assess an average-sized image in the order of a few hundred milliseconds or even less. In our evaluation of the proposed detection technique, we enforce pure serial computations in the recognition pipeline in order to study the effectiveness of the method. The main image analysis concept is that a pornographic image can be effectively discriminated by exploiting the inherent statistical properties of a properly selected region of interest (referred to as ROIs) having special characteristics (e.g. skin color) expressed by relevant image features. In the algorithmic design of the method, the existence of such a ROI provides strong evidence on its potential to be of pornographic nature. It is important to note that non-existence of such a ROI is observed to provide reliable evidence that the image is of benign nature. In the latter case, the most demanding computations involving feature extraction and classification can often be safely discarded. We stress, however, that a major factor that impedes the reliability of this ROI capturing step is the inability of the employed skin detection process to be robust against noise, namely false alarms. For instance, large populations of accumulated adjacent pixels covering a large area on the image can, under certain circumstances, be falsely treated as ROIs. On the other hand, such ROIs often tend to possess characteristics that are atypical of pornographic ROIs, a fact harnessed by our predictive scheme which empirically proved to be robust in discriminating such occurrences.

2.1 Skin color detection and localization

Skin detection is a heavily studied research topic; see [9] for a survey. Our selected skin detection algorithm operates on the RGB color space of reference and defines explicit rules on the RGB intensity channels [9]. This technique essentially aims at capturing the true skin color cluster in RGB explicitly. However, it is empirically found to suffer from many false alarms in the presence of either excessive illumination, or objects with both coherent texture and skin-like color [4]. To alleviate the illumination artifact to some extent, we apply an adaptive Gamma correction filter [8] on the entire image in a preprocessing step.

In a preliminary localization phase, the image plane is partitioned into four quadrants of equal dimensions. The algorithm proceeds by iterating recursively over each of the four incident quadrants on the plane. At each iteration, the normalized luminance histogram of the skin colored pixels of the corresponding quadrant $P(i) \in [0, ..., 1]$ for $i \in [0, ..., 255]$ is computed. By normalizing the entire histogram by the maximally valued bin, we can force the expression $\sum_{i=0}^{255} P(i) = 1$ to hold. In this sense, $P(i)$ can be interpreted as the approximate discrete probability mass distribution of gray-level intensity occurrences in the quadrant. Based on these, we further compute the following features:

$$SR = \frac{number\ of\ skin\ pixels\ in\ quadrant}{total\ number\ of\ pixels} \quad (1)$$

$$KR = \sigma^{-4}(\sum_{i=0}^{255}(i-\mu)^4 P(i)) - 3 \quad (2)$$

where σ is the standard deviation of the histogram bin values in $P(i)$ and μ is the mean histogram bin value. Coherent skin regions are observed to exhibit a tendency to form grey-level histograms with sharp peaks. This is a consequence of

(a) (b) (c)

Figure 1: Main image processing steps: (a) original image, (b) skin detection, (c) ROI localization

the fact that pixel intensity tends to deviate slightly about an intensity of reference (i.e. the intensity attaining the highest frequency). Thus, for such textures, increased values of kurtosis are expected, especially when compared to other non-skin non-coherently textured image regions [5].

The algorithm proceeds by deciding whether the SR and KR quantities exceed two predefined thresholds, namely τ_1 and τ_2. These thresholds are constant at each splitting level in the localization procedure, and are estimated by relaxing a locally optimal quadratic discriminant in the $SR-KR$ feature space. This feature space is populated by tuples computed from small image patches obtained from the original positive and negative training images, as determined by the region splitting mechanism. More specifically, if $SR \geq \tau_1$ and $KR \geq \tau_2$, then the quadrant is recursively split into four quadrants, and the same splitting test is applied to the new occurring quadrants. The splitting process can essentially be regarded as a quad-tree 2D planar decomposition of the skin ROIs; see Figure 1(c) for a visualization of this procedure. Our experiments indicate that three levels of recursion are enough in order for the capturing technique to adequately localize the shape of incident objects in the image. The number of quadrants can be fixed adaptively in terms of a step function with respect to the dimensions of the image, so that more quadrants are formed in each splitting level for larger images and finer localization be enforced. The larger the number of partitions performed at each splitting level, the less splitting levels are required, in order to produce finer adaptation to the true shape of the objects in the image.

2.2 Feature extraction

If the localization process outlined above does locate a ROI, we then collect the following fifteen features by iterating over the pixels inside the convex hull representing the ROI:

1. Mean and variances of the R, G and B intensities of the pixels residing inside the determined convex hull (6 features). Considering pornographic and benign images alike, these features tend to deviate about certain values of reference.
2. Seven spatially invariant Hu's moments [3] (7 features). These moments are able to capture the geometrical properties of the ROI.
3. Ratio of the total skin and non-skin pixels delimited by the ROI (1 feature). This feature tends to exhibit significant contribution to discriminating pornographic and benign images.
4. The angle of the diameter of the convex hull established by the vertices of the quadrants in the deepest splitting level of the ROI capturing process (1 feature). This feature aims at providing adequate discriminative information in order to separate upright human figures (for instance, naked models or frontal face images) from assorted pornographic scenes. Measurements of this feature in positive and negative training samples in our data set follow a slightly altered distribution.

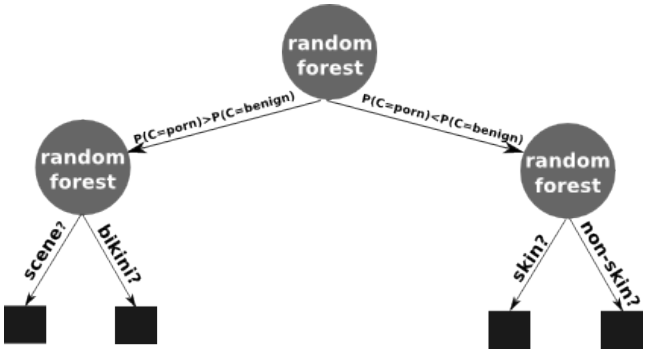

Figure 2: Decision-tree classifier employing random forests in the nodes

All features above are aggregated in a 15-dimensional real feature vector, which is next fed into our classification engine (described shortly). With a stack-based flood fill iterative method, the computation of the above features takes $O(n)$ time, with n being the number of pixels in the ROI.

2.3 RF-tree categorization model

The underlying idea behind the proposed predictive model is to combine the properties of decision tree classifiers (as in Quinlan's $C4.5$ decision tree or CART [7]) along with strong classifiers that can effectively discriminate pornographic and benign images but also their interclass counterparts as defined previously. To the best of our knowledge, the classifier that achieves the finest accuracy on our challenging dataset is the Random Forest classifier [1]. One impressive fact about this type of classifier is that it attains high accuracy without excessive fine-tuning; only an adequate number of trees in the model must be determined. The purpose of our tree-based random forest classifier is to tackle the 4-subclass classification task in a hierarchical top-down process. One interesting fact regarding this ensemble model is that it does not exhibit greater complexity than the inherent complexity in obtaining a separate regression or classification model for each node in the tree. In this sense, each predictive unit in the tree is trained off-line separately from the other intervening nodes. Classification is achieved by fist providing a coarse discrimination on an input feature vector and then performing categorical categorization by fine-grained classification units. At the root node of the complete binary decision tree, shown in Figure 2, the regressing Random Forest classifier outputs the likelihood of the input vector being an indicator of either a pornographic or benign image. In order to make the transition from the root-node classification unit to a deeper fine-grained classifier, we exploit the prior probabilities p_P and p_B of the pornographic and benign classes. These probabilities are estimated by an evaluation of the root-node classifier trained over the entire dataset.

Classifier	Test setting	CCR	FCR
Proposed method	Porn vs. benign	88,2%	11,8%
	Porn-scene vs. bikini	87%	13%
	Skin vs. non-skin	90%	10%
POESIA	Porn vs. benign	82,4%	17,6%

Table 1: Comparison of the recognition accuracy of POESIA vs. our method, in terms of Correct Classification Rate (CCR=TP+TN) and False Classification Rate (FCR=FP+FN)

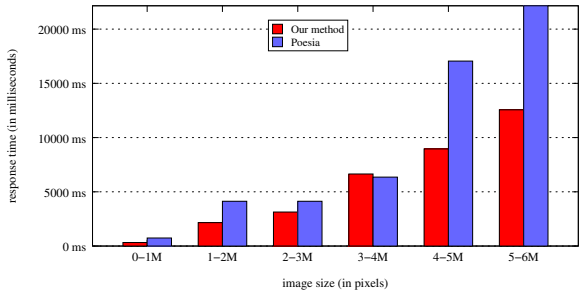

Figure 3: Comparison of the attained response times of our method to those of the POESIA porn classifier

Formally, they are estimated by the following formulas

$$p_P = \frac{CCP}{T},\ p_B = \frac{CCB}{T} \tag{3}$$

where T denotes the total number of features (in our case T=9,000), and CCP and CCB denote the number of correctly classified porn and benign features respectively. The values of the CCP and CCB parameters are estimated from our training set by training the root-node classifier and estimating its generalization by enforcing 10-fold cross validation (in fact, over 50% of all porn-benign features imposing an equal number of features in the positive and negative classes are used for training; the remaining features are used for testing purposes). The classification algorithm proceeds to the next level of classifiers based on the relationship between the predicted joint probability of the input feature vector and the prior probabilities of the respective coarse classes. In the same sense, the feature is then assigned to a specific subclass of the pornographic category (namely, "porn-scene" and "bikini" classes) and broader benign category (namely, "skin" and "non-skin" classes).

3. EXPERIMENTAL RESULTS

Here we provide experimental results on our porn image recognition algorithm against our 9,000-sample data set of pornographic and benign images (comprising subclasses). Our recognition system is compared against the open-source POESIA project [6] both in terms of processing speed and accuracy. In order to compare the processing speed of these two systems, the pornographic image filter of the POESIA project is used in a standalone fashion. The classification accuracy of both algorithms is summarized in Table 1.

Moreover, Figure 3 depicts the classification wall-clock turnaround times of both our technique and the POESIA image classifier. As seen in Figure 3, the classification time is progressively dependent on image size. The proposed method attains similar accuracy to that of POESIA (in the porn versus benign classification challenge), but it exhibits an almost 2× speedup against the latter.

4. CONCLUSIONS

In this paper, a novel technique for categorizing pornographic images based on skin colored ROIs is presented. The method is evaluated on a hierarchically categorized dataset comprising four semantic classes. The proposed predictive model manages to achieve high accuracy on our data set, while outperforming state-of-the-art approaches in classification processing time. It is thus deemed as a promising pornographic image categorization model for real-world applications, in which both adequate speed and high precision demands are imposed.

5. ACKNOWLEDGMENTS

The research leading to these results has received funding from the European Union Seventh Framework Programme (FP7/2007-2013) under grant agreement no 287674 (3DTVS) and from the COST project IC1106. The publication reflects only the authors' views. The EU is not liable for any use that may be made of the information contained therein.

6. REFERENCES

[1] L. Breiman. Random forests. *Machine learning*, 45(1):5–32, 2001.

[2] M. Fleck, D. Forsyth, and C. Bregler. Finding naked people. In *Proc. ECCV*, pages 593–602, 1996.

[3] M. Hu. Visual pattern recognition by moment invariants. *IRE Transactions on Information Theory*, 8(2):179–187, 1962.

[4] S. Karavarsamis, N. Ntarmos, and K. Blekas. InFeRno–an intelligent framework for recognizing pornographic web pages. In *Proc. ECML PKDD*, pages 638–641, 2011.

[5] A. Materka and M. Strzelecki. Texture analysis methods-a review. Technical report, Tech. Univ. of Lodz, Inst. of Electronics, 1998.

[6] POESIA project. http://www.poesia-filter.org/.

[7] J. Quinlan. *C4.5: Programs for machine learning*. Morgan kaufmann, 1993.

[8] Y. Shi, J. Yang, and R. Wu. Reducing illumination based on nonlinear gamma correction. In *Proc. IEEE ICIP*, volume 1, pages 529–532, 2007.

[9] V. Vezhnevets, V. Sazonov, and A. Andreeva. A survey on pixel-based skin color detection techniques. In *Proc. Graphicon*, volume 3, 2003.

[10] J. Wang, J. Li, G. Wiederhold, and O. Firschein. Classifying objectionable websites based on image content. In *Proc. IDMS*, pages 113–124, 1998.

[11] J. Wang, G. Wiederhold, and O. Firschein. System for screening objectionable images using Daubechies' wavelets and color histograms. In *Proc. IDMS*, pages 20–30, 1997.

[12] J. Yang, Y. Shi, and M. Xiao. Geometric feature-based skin image classification. In *Proc. ICIC*, pages 1158–1169, 2007.