Detecting pornographic images by localizing skin ROIs

Sotiris Karavarsamis[a], Nikos Ntarmos[b], Kostantinos Blekas[c], and Ioannis Pitas[a]

[a]AIIA Group, Department of Informatics, Aristotle University of Thessaloniki, Greece
[b]School of Computing Science, University of Glasgow, Scotland
[c]Department of Computer Science and Engineering, University of Ioannina, Greece

Abstract

In this study[1], a novel algorithm for recognizing pornographic images based on the analysis of skin color regions is presented. The skin color information essentially provides Regions of Interest (ROIs). It is demonstrated that the convex hull of these ROIs provides semantically useful information for pornographic image detection. Based on these convex hulls, the authors extract a small set of low-level visual features that are empirically proven to possess discriminative power for pornographic image classification. In this study, we consider multi-class pornographic image classification, where the ”nude” and ”benign” image classes are further split into two specialized sub-classes, namely ”bikini” / ”porn” and ”skin” / ”non-skin”, respectively. The extracted feature vectors are fed to an ensemble of random forest classifiers for image classification. Each classifier is trained on a partition of the training set and solves a binary classification problem. In this sense, the model allows for seamless coarse-to-fine-grained classification by means of a tree-structured topology of a small number of intervening binary classifiers. The overall technique is evaluated on the AIIA-PID challenge of $9,000$ samples of pornographic and benign images collected from the Web. The technique is shown to exhibit state-of-the-art performance against publicly available integrated pornographic image classifiers.

Index Terms

convex hull calculation, multi-class classification, porn detection, random forests, skin ROI localization

I. Introduction

With the galloping evolution of the Internet during the last two decades, pornographic images can be readily accessed even by certain sensitive groups of people, such as adolescents. Moreover, the ability to distribute and share images through public HTTP services without an autonomous content supervision process intervening in the content sharing loop encourages circulation of illicit images. For instance, pornographic content distributors may often exploit unprotected Web services in order to circulate or exchange child pornography and general pornographic content. The impact of pornography to people has long become a rising issue of concern and was under the spotlight of psychologists decades ago. Psychological research dated back to the 80’s stressed out that the exposure of children to pornography effectively impedes the smoothness of their behavioral evolution. In the same direction, similar studies have indicated that intense exposition to pornographic material affects human behavior and mood in adults. For some representative papers dealing with this problem see [BAW76], [Pie84], [CFP71], [LW05], [Mey72], [PBSN89]. The premier source of pornographic information on the Web has traditionally been pornographic images and video. We assert that the semantic load of such images is essentially the primal

Sotiris Karavarsamis can be reached at sokar@aiia.csd.auth.gr.
Nikos Ntarmos can be reached at nikos.ntarmos@glasgow.ac.uk.
Konstantinos Blekas can be reached at kblekas@cs.uoi.gr.
Corresponding author: Prof. Ioannis Pitas. E-mail address: pitas@aiia.csd.auth.gr

[1]This document is a reprint of the homonym article published in the International Journal of Digital Crime and Forensics, 5(1), 39-53, January-March 2013 with minor corrections.

carrier of pornographic information to effectively engage the focus of attention of the end user, e.g., when browsing pornographic web pages. Thus, we contend that pornographic image detection is a crucial aspect in the loop of effectively identifying pornographic web pages.

A system tailored towards identifying pornographic images should exhibit a robust capability in distinguishing between regular benign images and assorted pornographic images. In a real world content filtering scenario, zero-error categorization is still an unrealized goal. That being said, systems proposed in the literature are characterized by certain strengths and weaknesses in detecting pornographic content. Among many challenging problems in this computer vision problem, the difficulty in constructing an accurate algorithmic framework for detecting pornography is often attributed to varied photometric conditions (which often occur in the form of bad illumination), unconstrained clutter, occlusions and variation in the poses of involved human subjects. Thus, it is intrinsically difficult to construct a precise algorithm that encodes accurate prior information about what a pornographic image really is. To the best of our knowledge, many content filtering systems operate satisfactorily in identifying pornographic web pages by means of pornographic image detection. Despite the abundant availability of textual and structural information in common pornographic web pages, a system can exhibit more robust accuracy by being able to tell pornographic and benign images apart. Later, we review some previously proposed systems in the literature aimed at pornographic web page detection.

In previous works [KNB11], [KPN12], we tackled the problem of identifying pornographic web pages. To the best of our intuition, by constructing an algorithm that can tell pornographic and benign images apart at a satisfactory rate (of, e.g., 80% of the time), we can further mine and employ different web page cues in order to come up with a stronger decision mechanism that can reliably tell if a web page is either pornographic or benign. In fact, by naively imposing a threshold on the ratio of the number of pornographic and benign images detected in a given web page, we can readily achieve a meaningful trade-off between detection accuracy and turnaround time [KNB11]. We have observed that the probability of generating a false positive can be significantly lowered down when limiting our focus on web pages that are indeed of pornographic nature. At the same time, true positive can be effectively boosted. However, this naive heuristic decision making criterion can plausibly suffer from artifacts when assessing web pages containing only a limited number of images in their document object model (also being referred to as the *DOM model*). In this class of web pages, the probability of false categorization can be relatively high depending on the underlying image classifier. Obviously, the presence of a large number of candidate images in the categorization pipeline poses a burden towards the user-perceived turnaround time in real-world content filtering systems. In the literature of pornographic content-filtering system design, many systems have employed techniques to reduce the computational burden induced by the high availability of pornographic images in regular web pages. For instance, in the design of the Wavelet Image Pornography Elimination (WIPE) system [WLWF98], a more statistically sound algorithm for sampling only a limited fraction of candidate images is formulated and further investigated analytically.

In the computer vision community, understanding the semantics of an image accurately is still a non-trivial, illusionary task. Previous pioneering works focused on representing visual scenes using low-dimensional representations that extract essential information to capture the nature of a scene. The work of [OT01] proposed the GIST descriptor which attempts to capture certain dominant parameters of the so-called spatial envelope of visual scenes framed by a color image. In the course of describing scenes, direct segmentation of regions and inspection of individual segmented regions in an image is bypassed effectively by probing a holistic representation of the entire image. In the direction of probing low-dimensional representations of pornographic scenes, we study the effect of combining low-dimensional feature extraction (of a limited set of low-level visual features) with categorization ensembles of random forest classifiers.

In a practical content filtering application, a pornographic image classifier must as well meet computational efficiency constraints in terms of the observed turnaround time it attains. In a typical scenario where millions or billions of images are naturally required to be processed in the filtering loop, a classifier that can achieve a meaningful trade-off in assigning correct labels to input and demanding low computational

time towards classification is also preferable. To this end, we discuss an algorithm that can efficiently categorize pornographic images in a meaningful amount of time by localizing regions of interest. The proposed technique for identifying pornographic images is benchmarked against the state-of-the-art open source pornographic image classifier of the Public Open-source Environment for a Safer Internet Access (POESIA) project[2].

II. Related work

The first attempt at devising an autonomous system for screening pornographic images is the work of [FFB96]. In their detection method, skin-colored region segmentation is performed as a pre-processing step. Then, skeletons are found on the segmented skin colored regions. This specialized skeleton identification technique, however, exhibits a combinatorial nature, which significantly slows down the overall image analysis pipeline. It has specifically been reported that "over six minutes are required in order to screen an input image even on a workstation". Although the method in [FFB96] exhibits high recognition accuracy, it cannot satisfy the stringent time-efficiency requirements in real-world systems, where millions or even billions of images may routinely be screened.

A system to recognize pornographic web pages by assessing each image in a web page is presented in [WLWF98], [WWF97]. In order to classify images, a technique employing wavelet and histogram analysis is employed [WWF97]. In this pipeline, an input image is first screened as either being a benign figure or a pornographic image. If the input image is not classified as benign, then more elaborate processing takes place by extracting wavelet domain features. In order to classify an incoming image, its corresponding feature vector is matched against a database of training feature vectors that are computed off-line [WWF97].

In [HZW+11], an integrated pornographic web page detection system capable of detecting nude images and videos is presented. Pornographic image categorization is performed by computing a number of global image features, combined with a random forest classifier. In the case of pornographic video detection, video classification is cast as a problem of identifying pornographic video shots, enhanced by interval frame sampling which reduces processing time. Additionally, prior knowledge acquired by a video sound stream in alignment with the video frame sequence is exploited in order to more accurately categorize a video by means of a Gaussian mixture model. Finally, Web page categorization is achieved by fusing different types of features mined from objects in a web page (i.e., images, video, text and link structure) using a multi-instance learning framework.

More recent works on pornographic image detection employing the widely celebrated idea of visual words, [DPN08], [US11], attempt to identify pornographic images by learning vocabularies of visual words often present in pornographic images.

In this work, we develop a computationally efficient algorithm for recognizing pornographic images. We build up on our previous work on the InFeRno[3] framework [KNB11], [KPN12], which tackles pornographic web page detection by means of pornographic image recognition. A large body of previous research efforts in porn detection relies on the hypothesis that the presence of skin color information in images can provide discriminative information that can assist in the pornographic versus benign image classification pipeline. In our work, we build on the assumption that skin information is an important cue, in order to generate hypotheses for the presence of pornographic content in an image. More specifically, a region splitting scheme is proposed for skin region segmentation, inspired by the geometric localization technique proposed by Yang et al [YSX07]. The criterion to identify skin-colored regions of interest (ROIs) is based both on the accumulation of connected skin-colored pixels and on the spatial coherence of homogeneously textured image patches. By performing feature extraction over the image pixels delimited by an uncovered ROI, we employ a tree-structured ensemble of Random Forest classifiers (called *RF-tree* herein) in order

[2]The source code of the POESIA filter can be obtained at http://sourceforge.net/projects/poesia/

[3]InFeRno source code can be accessed at http://www.github.com/brigr/InFeRno for review, development and use. It is released as open source software under a GPL license.

to categorize the input image. The overall method yields better performance in porn recognition accuracy versus state-of-the-art pornographic image detection systems.

Due to the lack of standard published data sets for pornographic image detection and the difficulty in sharing pornographic data to conduct experiments for porn recognition, we have resorted to compiling the AIIA-PID[4] pornographic data set containing $9,000$ manually annotated pornographic and benign training image samples collected from the Web. In AIIA-PID, the benign and nude categories are split into two semantic categories each, namely "skin"/"non-skin" and "bikini"/"hard porn", respectively. Thus, finer discrimination can be made within the benign and porn image classes. The proposed classification model in the AIIA-POD technique follows a top-down approach in the course of classifying input feature vectors. Our classification scheme receives as input a 15-dimensional feature vector for each convex hull, in contrast to more complex feature extraction-based categorization techniques that compute a large number of features. In a first assessment test, the technique classifies the feature vector as either corresponding to the pornographic or benign semantic category. This is achieved by querying the root node of the tree-structured ensemble (originally being a regressing unit), which provides us with the likelihood p that the input is of pornographic nature. In contrast, the probability that the input feature vector is benign is $1-p$. This step constitutes the coarse-grained phase of the overall technique, in which the categorization model acquires a general first-step assessment on the nature of the input feature vector. The decision control is then delegated to a leaf classification node (called *unit* herein). In the first level of processing in the RF-tree model, an input image feature vector is characterized as either "nude" or "benign". In the intervening refinement step, the input feature vector is passed to either a "bikini"/"porn" or to a "skin"/"non-skin" subclass classifier, in the case of "nude"/"benign" classes, respectively. The overall pornographic image recognition scheme was evaluated on the AIIA-PID image database and attained a high classification accuracy. A discussion on the comparison between the proposed technique and the POESIA pornographic image classifier is also provided herein.

III. SKIN ROI LOCALIZATION

In this section, we describe a geometrical planar decomposition algorithm for detecting skin colored ROIs in a color image. Although the proposed technique is inspired by the work of [YSX07], we impose a simplification over the latter method that results in a less computationally intensive localization algorithm. The algorithm is of recursive design and shares much in common with quad-tree image segmentation [Pit00]. In practice, the algorithm captures sufficiently well an underlying ROI by performing only a limited number of recursive calls, thereby retaining a limited computational overhead.

The skin ROI localization algorithm relies on the uncovering of regions of interest. In this paper, these regions will simply be referred to as ROIs for the sake of simplicity. At the first level, skin pixel detection is performed according to the RGB thresholding technique discussed in [VSA03]. Then, the skin ROI detection algorithm considers the whole image as one skin ROI. The gray-level version of the color image and the gray-level histogram are then computed. Then, the following statistical descriptors of each ROI are computed: a) the ratio r of skin-colored pixels to non-skin pixels, and b) the kurtosis k of the gray-level histogram of the corresponding ROI. Both measures r and k are defined in the following manner. Let I be the normalized gray-level histogram of an image region, that is, the total sum of bin values in histogram I equals 1. We generally assume that I contains a sufficient number N of quantization bins (obviously, N must not exceed the value of 256). In our experimental derivation, we set his parameter to the largest possible value, namely $N = 256$; however, smaller values of N are also capable of representing histogram peaks reasonably well. In this sense, each bin value $p(i)$ of a bin indexed by i can be regarded as the approximate probability mass of intensity value occurrence in the gray-level region. Note that bin indices count from zero. Then, the kurtosis value of this region is expressed as

[4]To obtain a copy of the AIIA-PID data set, please contact the corresponding author.

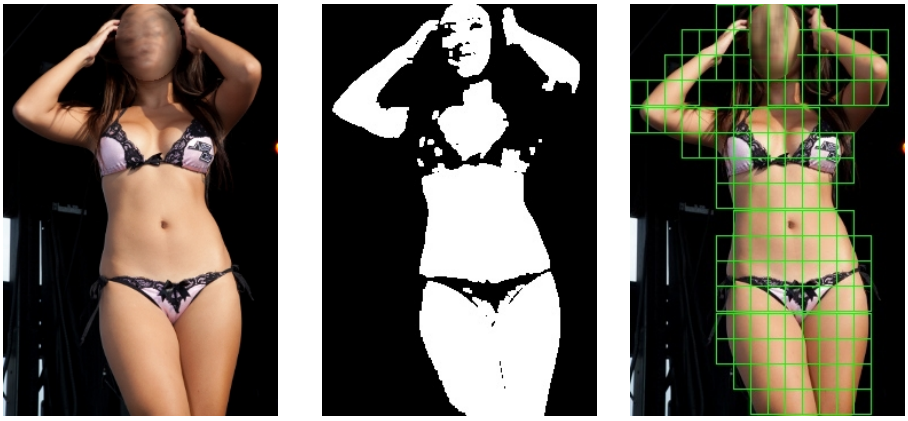
Algorithm 1 Pseudocode of algorithm for uncovering regions of interest on the image plane

procedure DETECTSKIN(I)
  $R \leftarrow$ initialize empty zero integer matrix of size $Width \times Height$
  $maxI \leftarrow widthOf(I)$
  $maxJ \leftarrow heightOf(I)$
  for $i \leftarrow 1, maxI$ **do**
    for $j \leftarrow 1, maxJ$ **do**
      if $isSkinPixel(I(i,j))$ **then**
        $R(i,j) \leftarrow 1$
      else
        $R(i,j) \leftarrow 0$
      end if
    end for
  end for
end procedure
procedure COMPUTEROI($I, currentLevel, maxLevel$)
  $ControlPoints \leftarrow \emptyset$
  $S \leftarrow DetectSkin(I)$
  if $currentLevel = maxLevel$ **then**
    $ControlPoints \leftarrow ControlPoints \cup controlPointsOf(I)$
    return $ControlPoints$
  end if
  while $currentLevel < maxLevel$ **do**
    for $p \in QuadPartitions(I, currentLevel)$ **do**
      $kurtosis \leftarrow computeKurtosis(p)$
      $skinProportion \leftarrow computeSkinToNonskinRatio(S, p)$
      if $kurtosis > T_K$ and $skinProportion > T_S$ **then**
        $computeROI(p, currentLevel + 1, maxLevel)$
      end if
    end for
  end while
end procedure
procedure LOCALIZEREGIONOFINTEREST($I, maxLevel$)
  $M \leftarrow 3$
  $Q \leftarrow computeROI(I, 0, M)$
  $H \leftarrow$ compute convex hull of I
  $P \leftarrow$ stack all pixels in I bounded by H
  return P
end procedure

$$k = \sigma^{-4}(\sum_{i=0}^{255}(i-\mu)^4 P(i)) - 3 \tag{1}$$

where σ and μ are the standard deviation and mean over all bin values in I. Respectively, μ and σ are defined[5] as $\mu = \frac{1}{N}\sum_{i=0}^{N-1} P(i)$ and $\sigma = \frac{1}{N-1}\sum_{i=0}^{N-1}(P(i)-\mu)^2$. Likewise, r is defined as the ratio of

[5]Note that for $N \gg 0$, $\sigma = \frac{1}{N-1}\sum_{i=0}^{N-1}(P(i)-\mu)^2 \approx \frac{1}{N}\sum_{i=0}^{N-1}(P(i)-\mu)^2$. In effect, for a sufficiently large value of N, both multiplicative factors of the finite sum have an insignificant numerical effect to the computation of the statistic, that is only observable in the more insignificant fractional digits.



(a) bikini image (b) skin-colored pixels (c) skin ROI

Fig. 1: Skin ROI detection and localization. We are interested in finding skin regions that have high skin-color pixel ratio and are fairly homogeneously textured, that is, they have sufficiently large values of the kurtosis measure. The testbed image above is courtesy of J. E. Ramspott (reproduced with permission).

skin pixels over the inspected image region to the total number of pixels in the region.

In order to decide on either splitting or ignoring an image region in the region decomposition scheme, we introduce two upper bounds T_S and T_K on the ratio r and the kurtosis k, respectively. We decide to further split the current quadrant into four equal quadrants, if $r > T_S$ and $k > T_K$ holds true. A region that passes this test is expected to possess a strictly homogeneous texture and contain a fair amount of skin-tone pixels. In this sense, such a region can be assumed to correspond to a true body part patch depicted in the image. This recursive operation is repeated up to an explicitly defined level of maximum recurrences. According to our experience on the technique, there is not a single choice in the value of the maximum recurrence level that suits optimal body contour capturing. Therefore, both T_S and T_K and the recurrence level are chosen empirically. However, we impose an indicator step function that maps the actual dimensions of the input image to empirical recurrence level values that we obtain through experimental observation. Figure 1(c) illustrates the results of skin ROI segmentation. The original color image for which skin ROI segmentation is performed is depicted in Figure 1(a), while the skin segmented binary image of the color image is depicted in Figure 1(b). The computations entailed in localizing a skin ROI essentially depend on the raw image data and the segmented image. In Figure 1(c), localized regions in the image are shown as green rectangles overlapping the original image. Note that we are generally interested in obtaining dense segmentations of such rectangular-shaped images patches, thereby ignoring ”holes” exhibited by general background enclosed by homogeneous skin regions. In listing 1, we provide pseudocode detailing the computational process of localizing a skin ROI.

IV. Image feature extraction

In each suspected porn image, we initially calculate the convex hull [DBCVKO08] of the obtained skin ROIs. In our experimental setup, we employed the convex hull computation algorithm of Sklansky [GFY83], which is implemented internally in the Open Computer Vision framework [Bra00]. Given a set of n points (in our case, this set of points lies on a discrete image lattice), Sklansky's algorithm computes the (unique) convex hull in $O(nlog(n))$ time, which in practice is fast enough for our application. In our case, the points to be considered for convex hull construction are the corners of the skin-colored image ROIs obtained during the localization of a skin ROI. The use of convex hulls can meaningfully represent skin regions, without introducing significant noise in the spatial representation of localized skin regions.

The image features that we collect within the convex hull are the following:

1) The ratio of the total skin to non-skin pixels within the convex hull (1 feature). This feature provides significant discriminant information for pornographic image content detection.
2) The arithmetic means (namely, m_R, m_G and m_B) and variances (σ_R^2, σ_G^2 and σ_B^2) of the R, G and B color channels in the convex hull (6 features). These features provide discriminant information on the color properties of the convex hull of the skin ROI.
3) The seven spatially invariant Hu moments [Hu62] of the pixels within the computed convex hull (7 features). These features provide discriminant information on the shape properties of the convex hull corresponding to the uncovered skin ROI.
4) The angle of the principal axis of the convex hull versus the horizontal axis (1 feature). This feature aims at providing discriminant information in order to separate upright human figures (for instance, naked models or frontal facial images) from pornographic images; human bodies tend to be horizontal.

In the abovementioned features, the proportion of the count of skin pixels to the count of non-skin pixels encodes predominant information on the skin tone dominance of an input image. Images of pornographic nature are expected to exhibit a high ratio of skin, while benign images often exhibit low values of this feature. However, this particular feature alone cannot be used effectively in order to classify an image as either pornographic or benign. Intuitively, a close-up image of a human face may contain a large proportion of skin pixels (although actually being benign). In contrast, a pornographic image may exhibit a very large proportion of non-skin pixels but present pornographic information in only a limited fraction of skin-pixels. Thus, pornographic image detection cannot rely on the proportion of skin alone, although it proves important especially when combined along with other visual features. Hu's moments in the feature extraction pipeline encode texture properties of the underlying ROI. It should be pointed out that the seventh Hu's moment in feature (3) encodes little discriminative information, since it typically takes values close to zero and exhibits a small variance about that value. The mean and variances of the RGB channels in the ROI encode information regarding the scatter f skin and non-skin pixels in the ROI. Regions of interest that contain plain skin regions are expected to exhibit mean and variances in their ROI skin pixels that are close to a dominant color of true skin tone.

We used two image classes, namely *nude* and *benign*, for image classification. Each of them has two subclasses, namely *bikini* and *hard porn* (for the porn class), and *skin* and *non-skin* (for the benign class), respectively. The porn label is assigned to real porn images. The bikini subclass is typically assigned to images of women/men wearing swimming suits. The skin label corresponds to images of humans having small exposed skin regions, e.g., face or hands in medium and long shot images. The non-skin label is assigned to images that do not depict humans. During testing, each test image is classified into one of the four subclasses.

V. RF-TREE CLASSIFICATION SCHEME

The employed RF-tree predictor consists of a tree-structured ensemble of strong Random Forest classifiers [Bre01], as shown in Figure 2, that allows to perform coarse-to-fine-grained categorization of input features. The root node classifier assesses the likelihood of an input image feature vector to correspond to either the "nude" or the "benign" class. The three classification nodes in the RF-tree are trained in isolation from each other using the standard Random Forest learning algorithm described in [Bre01].

In our model, the root RF classifier provides a probability p that a test image is assigned to the "nude" class. Then, a feature vector is assigned to the nude class if and only if $p \geq p_n$; otherwise is assigned to the benign class. Based on this decision, we move downwards in the tree to the left or the right leaf node, in order to decide if the image is bikini/porn or skin/non-skin, respectively.

An obvious drawback in the classification pipeline of the RF-tree model is that prediction errors can accumulate in leaft nodes due to false predictions performed in the root node. However, with a proper selection of the p_n threshold, we can achieve a good trade-off between false rejections and false acceptances. In search for a reasonable threshold p_n in the RF-tree pipeline, we compute the normalized

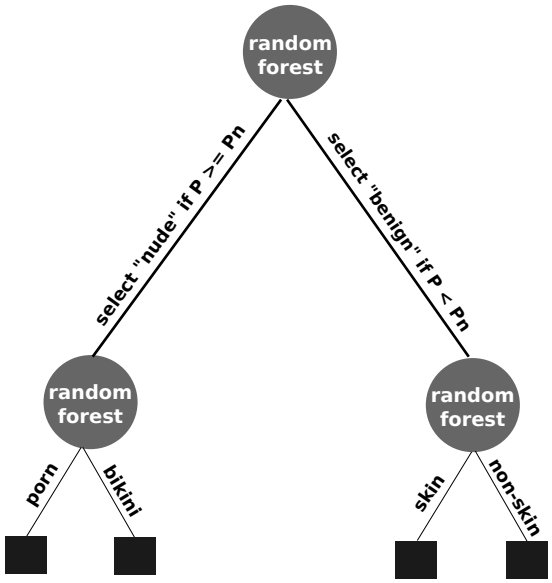
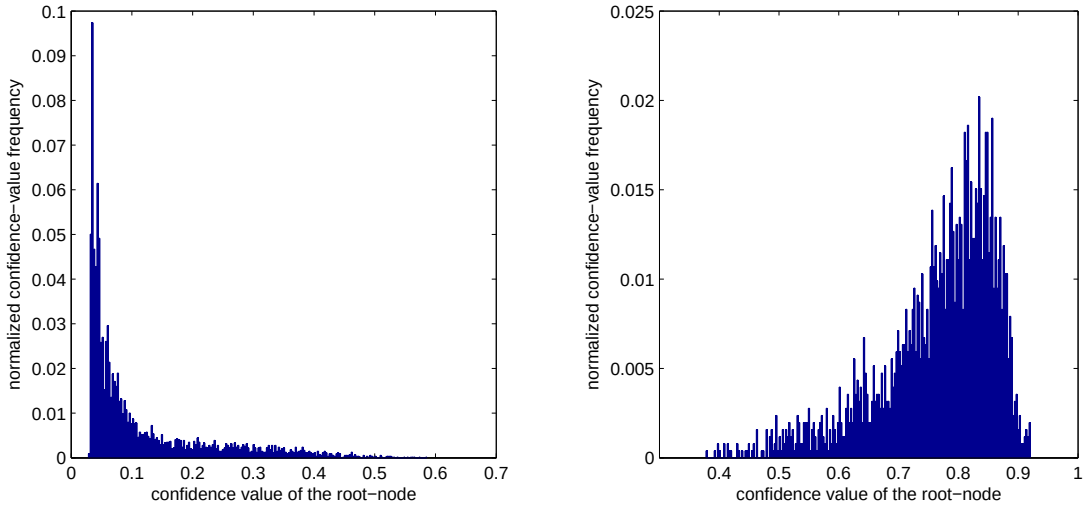


Fig. 2: The RF-tree classification model employing an ensemble of simple binary random forest classifiers.



(a) benign decision confidence frequency histogram

(b) porn decision confidence frequency histogram

Fig. 3: Normalized confidence-value frequency distributions attained by the root-node regressing unit in the RF-tree model over benign and nude image samples in the AIIA-PID dataset.

frequency mass histograms generated by the root-node classifier over the entire AIIA-PID training set (excluding images containing no ROI by the proposed technique); see Figure 3. The response distributions of the benign and pornographic classes follow different distributions overlapping in a small region between the decision frequency values of $p_n = 0.4$ and $p_n = 0.6$. In order to generate the histograms, we use a total of 7879 porn and benign feature vectors from the AIIA-PID data set. 32% of the total samples are labeled against either the pornographic or bikini class, while the remaining features correspond to skin and non-skin samples. From the obtained distributions that are shown in Figure 3, we draw two essential observations that are important for classification: a) the benign class follows a positively skewed probability distribution attaining most of its mass around $p_n = 0.05$. For probability values above that threshold, the mass probability distribution decays rapidly by attaining lower frequencies; and b) the

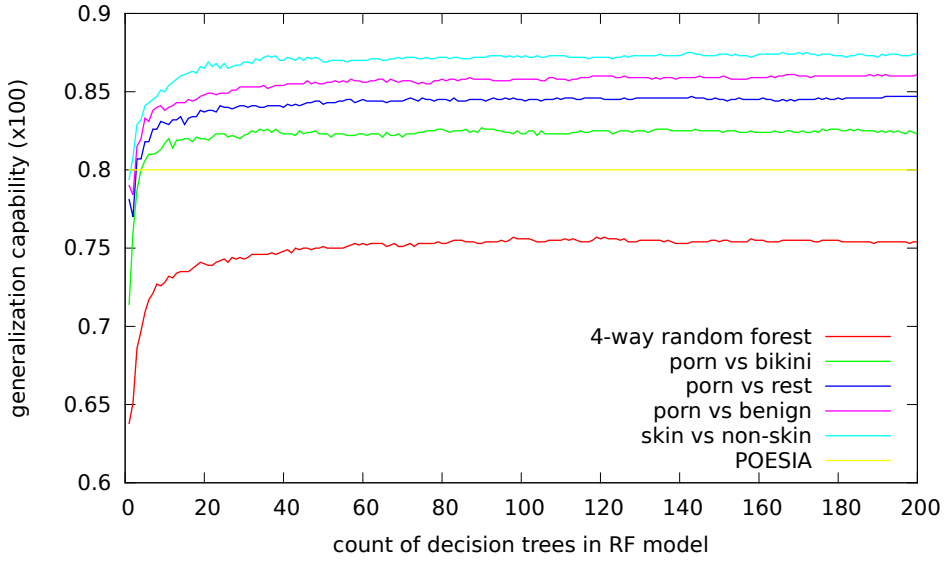


Fig. 4: The figure depicts the results of the 10-fold cross-validated evaluation of the generalization capability of the binary random forest classifiers in the RF-tree model. The 4-way classifier comprising all four semantic classes of the respective binary random forest classifiers exhibits suboptimal performance compared to the respective generalization capabilities of the binary classifiers in the RF-tree model. The generalization capability of the POESIA classifier has been evaluated using the standard training parameters in the POESIA source code bundle, and by using the entire AIIA-PID data set as a testing set.

response frequency mass distribution of the pornographic class follows a negatively skewed Gaussian-like distribution, which approximately attains its mode at $p_n = 0.8$. The latter frequency mass distribution decays rapidly left-wise to that latter response frequency value. However, both distributions attain a small class overlap between the frequency values $p_n = 0.4$ and $p_n = 0.6$. In this discrete interval, few samples correspond to the benign class and most of the overlapped samples belong to the pornographic class. Since false rejections are more costly in porn detection, we greedily choose[6] $p_n = 0.4$ as our decision threshold in the proposed RF-tree model. A more detailed analysis of the attained generalization error of the nodes in the RF-tree considering the AIIA-PID data set is given in the following section.

VI. Experimental results

We have tested the proposed method called AIIA-POD (AIIA porn detector) on a set of 9,000 images of the AIIA-PID database (AIIA Porn Image Database) containing porn, bikini, skin and non-skin images. An image is assigned to a subclass ”porn” / ”bikini” or ”skin” / ”non-skin”, if the probability provided by the corresponding leaf Random Forest classifier is larger than the a-priori subclass probability, as evaluated over the training set. In assessing the binary predictors in the RF-tree model, we performed 10-fold cross validation, using 90% of the total images as training ones and 10% of them as testing ones (non-overlapping sets). In the course of constructing an experimental software implementation of the proposed RF-tree model for assessing generalization performance, we opted for the highly optimized random forest implementation available in the Python scikit machine learning toolkit that implements binary and m-way (multi-class) random forest predictors [PVG+11].

Figure 4 depicts a comparison of the generalization accuracies obtained by our Random Forest classifiers generated across training samples from our 9,000-sample data set over 4 semantic categories. The generalization capability of each binary random forest classifier is evaluated by enforcing 10-fold cross

[6]In fact, this overlap region can form our *reject option* for classification. In this case, a confidence value lying in the reject option can assert model uncertainty about class membership.

validation. Models are trained by varying the number of decision trees integrated. The number of decision trees is varied from 1 tree (resulting in a degenerated random forest) up to 200 decision trees. For a similar discussion on the error progression of Random Forest classifiers, the astute reader is pointed to [FHT01]. In our experimental evaluation, we treat the 4-way random forest classifier as the baseline for our experiments. This type of classifier, as shown in the comparative graph, exhibits sub-optimal generalization compared to the standalone binary classifiers employed in the RF-tree model. As expected, all Random Forest models follow a decreasing pattern of generalization error. In practice, the number of trees employed in the ensemble classifiers impose a model complexity burden which is addressed both in the training and testing assessments of the models. Since training is performed off-line, model testing is more important for our pornographic image detection classifier. In particular, random forest models integrating less than 20 trees are shown to exhibit minimal performance, whereas their cross-validated error estimate is shown to stabilize around particular values as more trees are progressively added to the models. An interesting fact that describes the employed classifier is that optimal generalization can be achieved using a relatively small number of decision trees in the model, thus allowing us to establish a decreased model complexity with an acceptable generalization capability. At the same time, low learning model complexity allows for fast testing evaluations in our RF-tree classifier, as we will describe shortly.

The POESIA image classifier is also tested on our entire data set of pornographic and benign images using the parameters bundled in the POESIA project, achieving 80% of correct classification, as shown in Figure 4. In contrast, our best random forest models exhibit optimal performance compared to POESIA. A generalization capability of approximately $84,9\%$ in the "porn versus other" binary classification problem (equating to the classification problem tackled by the POESIA classifier) is observed to be attained. In our previous work [KPN12], generalization estimates were reported for fixed-parameter binary classifiers in the RF-tree model. As can be seen in Figure 4, as we vary the number of decision trees, the generalization of the models progresses better than the previously reported results.

In addition, an accuracy of approximately 83% is obtained by our best classifier tackling the "porn versus bikini" problem, which signals the ability of the model to accurately distinguish between pornographic and bikini images (e.g., images of people wearing swimsuits). In the same categorization problem, the POESIA image classifier exhibits suboptimal accuracy compared to the AIIA-POD technique by a margin of around 2%. Maximal performance (specifically, of about 88%) is also attained by the best classifier on the "skin versus non-skin" problem of distinguishing between benign images containing no true skin and images containing a limited amount of true skin (e.g., images of athletes with some body parts being exposed, images containing human faces, etc).

Table I provides a summary of the generalization accuracies attained by the binary prediction models and the POESIA classifier. As shown in the same table, the binary models that we trained exhibit small deviations about their attained mean generalization accuracies.

By assessing the turnaround times of both techniques, we opt for selecting the prediction models that exhibit maximal performance in the AIIA-POD technique. For the root node in the RF-tree model (depicted in Figure 2), we integrate 98 decision trees. The porn versus bikini classifier integrates 90 trees in its model, and the best classifier discriminating skin versus non-skin images employs 142 ones. By fixing the count of weak learners in the binary models, as discussed already, we assemble an RF-tree model that is integrated in our AIIA-POD classifier. In general, the observed turnaround time is affected by two major factors. If an image contains a skin ROI, then the total turnaround time is measured as the cumulative time of the computation of feature-vector extraction and classification by the RF-tree model. In the case where no ROIs are present in an input image, the total turnaround time is equal to the time consumed in the feature-extraction procedure. In the latter case, the image is given the "benign" label and no classification is performed. The computations involved in the feature extraction phase amount to deciding non-existence of a skin contour by means of the quad-tree decomposition algorithm, as described in Section III. Typical candidate images that do not contain ROIs are often auxiliary figures, background images, and figures that generally contain no true skin-tone pixels. In case an image contains a skin ROI, then the amount of time consumed in classifying its corresponding feature vector amounts to the time of evaluating two out

AIIA-POD versus POESIA: generalization capability				
Classifier	Test setting	generalization	mean	std
AIIA-POD	porn vs. other (bikini/skin/non-skin)	84,85%	84,05 %	0,0123%
	porn vs. bikini	82,8	82,02%	0,0138%
	nude vs. benign	85,65%	85,05%	0,0119%
	skin vs. non-skin	87,63%	86,9%	0,0130%
POESIA	porn vs. other	80%	N/A	N/A

TABLE I: Comparison of the accuracy attained by the POESIA classifier versus the AIIA-POD method in terms of generalization capability. 10-fold cross-validation has been enforced in estimating the generalization capability of the trained models in the RF-tree ensemble. In the table above, generalization is reported in terms of the best performing predictor that we obtained with respect to a cross-validated estimate. The respective error progressions decrease smoothly and approximately converge about a mean value, as the models exhibit a small deviation about their mean accuracy.

Classifier	0-1M pixels	1-2M pixels	2-3M pixels
AIIA-PID skin	537 ms	999 ms	4972 ms
AIIA-PID non-skin	34 ms	149 ms	290 ms
POESIA skin	753 ms	4455 ms	6744 ms
POESIA non-skin	728 ms	3045 ms	4925 ms

TABLE II: Attained response times in skin and non-skin images in the AIIA-POD and POESIA image classifiers in descending order of magnitude in the number of pixels

of three binary classifiers in the RF-tree model.

In the process of assessing the turnaround response time of our method, we opt for employing the most complex classifiers possible that at the same time attain maximal generalization. As seen in Figure 2, it is evident that we can gain a better speed-up performance by sacrificing little generalization capability (e.g., of no more than 1% according to the obtained 10-fold cross-validated generalization estimates) by only employing fewer decision trees in each respective binary classifier in the RF-tree model. For example, by choosing to employ a less accurate binary model, we can boost the running time of our method by employing a significantly smaller count of trees.

In Table II, we report speed-up comparison results of our technique versus the POESIA classifier. Turnaround times are quantized according to two parameters: a) the total count of pixels in an image (using a 1M pixel quantization factor); b) the presence or absence of a skin colored ROI in the image. In the case of the AIIA-POD technique, reported times refer to the computations including both feature extraction and classification performed by the RF-tree model. Our technique outperforms the POESIA classifier by a large margin in terms of response speed (by an average of a $19\times$ speed-up) when classifying images that are determined to contain no ROI. In the case of images that contain skin ROIs, the AIIA-POD method consumes considerably more time in feature extraction and classification. Images with less than 2 million pixels are in abundance on the Internet (e.g., social media, visual search engines, etc). For this class of images, our method demands on average 1 second of CPU time. Considering images containing between 2 to 3 million pixels, processing time in AIIA-POD gets more demanding. In the latter case, AIIA-POD demands approximately $1.7s$ of less CPU time compared to POESIA (on average). In the AIIA-POD technique, the evaluation of the RF-tree model using the counts of weak learners as described above, adds only 70 milliseconds of extra computational time.

In this experimental derivation, all of the reported turnaround times produced by the AIIA-POD technique enforce serial-only computations without further software optimizations. By enforcing multi-threading or GPU-based hardware acceleration on some of the most computationally intensive operations

in the AIIA-POD technique, we can possibly attain much better turnaround times.

VII. Conclusions

We have presented the AIIA-POD technique for categorizing pornographic and benign images in four semantic classes. The proposed method outperforms the current state of the art POESIA method when discriminating porn image versus all other subclasses (namely, the bikini, skin and non-skin classes) by a margin of 5%. Furthermore, it can discriminate other classes or subclasses, e.g., "porn" versus "bikini", "nude" versus "benign". Our method is much faster than POESIA (attaining approximately a $\times 2$ speed-up against the latter) in processing regular images of less than 2 Mpixels. In the case of images appearing to possess no skin colored ROIs, our method attains very fast response times with an average of a $19\times$ speed-up. Therefore, the AIIA-POD method is much faster than POESIA, has better performance and can provide finer image classification considering the porn, bikini, skin and non-skin subclasses.

VIII. Acknowledgments

The research leading to these results has received funding from the FP7 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.

References

[BAW76] Marvin Brown, Donald M. Amoroso, and Edward E. Ware. Behavioral effects of viewing pornography. *The Journal of Social Psychology*, 98(2):235–245, 1976. PMID: 1256030.

[Bra00] G. Bradski. The OpenCV Library. *Dr. Dobb's Journal of Software Tools*, 2000.

[Bre01] L. Breiman. Random forests. *Machine learning*, 45(1):5–32, 2001.

[CFP71] Royer F Cook, Robert H Fosen, and Asher Pacht. Pornography and the sex offender: Patterns of previous exposure and arousal effects of pornographic stimuli. *Journal of Applied Psychology*, 55(6):503, 1971.

[DBCVKO08] M. De Berg, O. Cheong, M. Van Kreveld, and M. Overmars. *Computational geometry: algorithms and applications*. Springer, 2008.

[DPN08] T. Deselaers, L. Pimenidis, and H. Ney. Bag-of-visual-words models for adult image classification and filtering. In *Pattern Recognition, 2008. ICPR 2008. 19th International Conference on*, pages 1–4. IEEE, 2008.

[FFB96] M. Fleck, D. Forsyth, and C. Bregler. Finding naked people. *ECCV 1996*, pages 593–602, 1996.

[FHT01] Jerome Friedman, Trevor Hastie, and Robert Tibshirani. *The elements of statistical learning*, volume 1. Springer Series in Statistics, 2001.

[GFY83] Ronald L Graham and F Frances Yao. Finding the convex hull of a simple polygon. *Journal of Algorithms*, 4(4):324–331, 1983.

[Hu62] M.K. Hu. Visual pattern recognition by moment invariants. *Information Theory, IRE Transactions*, 8(2):179–187, 1962.

[HZW+11] Weiming Hu, Haiqiang Zuo, Ou Wu, Yunfei Chen, Zhongfei Zhang, and David Suter. Recognition of adult images, videos, and web page bags. *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMCCAP)*, 7(1):28, 2011.

[KNB11] S. Karavarsamis, N. Ntarmos, and K. Blekas. Inferno–an intelligent framework for recognizing pornographic web pages. *Machine Learning and Knowledge Discovery in Databases*, pages 638–641, 2011.

[KPN12] S. Karavarsamis, I. Pitas, and N. Ntarmos. Recognizing pornographic images. In *Proceedings of the 14th ACM Workshop on Multimedia and Security*, pages 105–108. ACM, 2012.

[LW05] Ven-Hwei Lo and Ran Wei. Exposure to internet pornography and taiwanese adolescents' sexual attitudes and behavior. *Journal of Broadcasting & Electronic Media*, 49(2):221–237, 2005.

[Mey72] Timothy P Meyer. The effects of sexually arousing and violent films on aggressive behavior. *Journal of Sex Research*, 8(4):324–331, 1972.

[OT01] Aude Oliva and Antonio Torralba. Modeling the shape of the scene: A holistic representation of the spatial envelope. *International Journal of Computer Vision*, 42(3):145–175, 2001.

[PBSN89] Vernon R Padgett, Jo Ann Brislin-Slütz, and James A Neal. Pornography, erotica, and attitudes toward women: The effects of repeated exposure. *Journal of Sex Research*, 26(4):479–491, 1989.

[Pie84] Robert Lee Pierce. Child pornography: A hidden dimension of child abuse. *Child Abuse and Neglect*, 8(4):483 – 493, 1984.

[Pit00] I. Pitas. *Digital image processing algorithms and applications*. Wiley-interscience, 2000.

[PVG+11] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.

[US11] Adrian Ulges and Armin Stahl. Automatic detection of child pornography using color visual words. In *Multimedia and Expo (ICME), 2011 IEEE International Conference on*, pages 1–6. IEEE, 2011.

[VSA03] V. Vezhnevets, V. Sazonov, and A. Andreeva. A survey on pixel-based skin color detection techniques. In *Graphicon*, volume 3. Moscow, Russia, 2003.

[WLWF98] J. Wang, J. Li, G. Wiederhold, and O. Firschein. Classifying objectionable websites based on image content. In *Interactive Distributed Multimedia Systems and Telecommunication Services*, pages 113–124. Springer, 1998.

[WWF97] J. Wang, G. Wiederhold, and O. Firschein. System for screening objectionable images using daubechies' wavelets and color histograms. In *Interactive Distributed Multimedia Systems and Telecommunication Services*, pages 20–30. Springer, 1997.

[YSX07] J. Yang, Y. Shi, and M. Xiao. Geometric feature-based skin image classification. *Advanced Intelligent Computing Theories and Applications With Aspects of Theoretical and Methodological Issues*, pages 1158–1169, 2007.

AUTHOR BIOGRAPHIES

Sotiris Karavarsamis completed the BSc degree in Computer Science in 2011 at the department of Computer Science and Engineering (formerly Department of Computer Science), University of Ioannina, Greece. Between 2007 and 2008, he was on leave from the undergraduate program at UOI, contracting as a collaborating research programmer with a medical software spin-off firm hosted by the Scientific and Technological Incubator of Epirus (STEP) in Ioannina, Greece. He is currently with the Artificial Intelligence and Information Analysis (AIIA) laboratory, Department of Informatics, Aristotle University of Thessaloniki, Greece, where he is pursuing the MSc degree in Informatics. His research interests lie in intersections between computational vision and cognition, pattern recognition, image processing, and information retrieval in multimedia databases.

Nikos Ntarmos is currently a tenure-track, Lord Kelvin Adam Smith Fellow at the School of Computing Science, University of Glasgow, UK. Previously he served as a postdoctoral researcher at the Computer Engineering & Informatics Dept., University of Patras, Greece, and as an adjunct assistant professor at the Computer Science Dept., University of Ioannina, Greece. He received his Diploma from the Dept. of Electronic & Computer Engineering, Technical University of Crete, Greece in 2001, and his MSc and PhD from the Dept. of Computer Engineering & Informatics, University of Patras, Greece, in 2004 and 2008 respectively. Dr. Ntarmos has an extensive research experience in the areas of distributed computing and large-scale data management systems. He has also worked as a senior systems administrator and security specialist in both the public and private sector, and routinely contributes to open-source projects.

Konstantinos Blekas received the Diploma degree in Electrical Engineering in 1993 and the Ph.D. degree in Electrical and Computer Engineering in 1997, both from the National Technical University of Athens. He is currently on the Faculty of the Department of Computer Science and Engineering, University of Ioannina, Greece. His research interests include Artificial Intelligence, Machine Learning and Pattern Recognition with applications to Computer Vision, Robotics, Bioinformatics and Medical data analysis. He teaches courses in Probability Theory, Applied Statistics, Pattern Recognition and Machine Learning and he has co-authored more than 50 refereed journal and conference articles.

Ioannis Pitas (IEEE fellow, IEEE Distinguished Lecturer, EURASIP fellow) received the Diploma and PhD degree in Electrical Engineering, both from the Aristotle University of Thessaloniki, Greece. Since 1994, he has been a Professor at the Department of Informatics of the same university. He served as a visiting Professor at several universities. His current interests are in the areas of intelligent digital media, image/video processing (2D/3D) and human-centered interfaces. He has published over 690 papers, contributed in 39 books in his areas of interest and edited or (co-)authored another 8 books. He has also been an invited speaker and/or member of the program committee of many scientific conferences and workshops. In the past he served as Associate Editor or co-Editor of eight international journals and general or technical chair of four international conferences (including ICIP2001). He participated in 67 R&D projects, primarily funded by the European Union and has been a principal investigator/researcher in 40 such projects. His work counts 15700+ citations on Publish or Perish, 5500+ on Scopus, and his H-index is 62+ (on Publish or Perish), 37+ (on Scopus).