Skin Sheriff: A Machine Learning Solution for Detecting Explicit Images

Christian Platzer, Martin Stuetz, Martina Lindorfer
Secure Systems Lab, Vienna University of Technology
{cplatzer,mstuetz,mlindorfer}@iseclab.org

ABSTRACT

Digital forensics experts are increasingly confronted with investigating large amounts of data and judging if it contains digital contraband. In this paper, we present an adaptable solution for detecting nudity or pornography in color images. We combine a novel skin detection approach with machine learning techniques to alleviate manual image screening. We upgrade previous approaches by leveraging machine learning and introducing several novel methods to enhance detection rates.

Our nudity assessment uses skin detection and positioning of skin areas within a picture. Sizes, shapes and placements of detected skin regions as well as the total amount of skin in an image are used as features for a support vector machine that finally classifies the image as non-pornographic or pornographic. With a recall of 65.7% and 6.4% false positive rate, our approach outperforms the best reported detection approaches.

Categories and Subject Descriptors

I.4.6 [**Segmentation**]: Pixel classification; K.4.2 [**Social Issues**]: Abuse and crime involving computers

Keywords

Image Processing; Skin Detection; Pornography Detection; Digital Forensics

1. INTRODUCTION

According to the *Internet Pornography Statistics* [37] 12% of all Internet websites contain pornography, 42.7% of Internet users view pornography, while 34% of average users receive unwanted pornographic exposure. Furthermore, the use of the Internet to exchange illegal content causes enormous public concerns [2]. Forensics and law enforcement agencies are hard-pressed to find and rate huge amounts of data based on its content. Thus, techniques to automatically assess pornographic content are needed for content filters, parental control software, and criminal investigators.

One possibility to solve these issues is to completely block specific web addresses by searching a blacklist for matching URLs or querying special keywords in the text of the website. Another approach is to block the offensive images themselves. Recently, Facebook began combating the sharing of illegal pornography by integrating a tool from Microsoft called PhotoDNA [32, 36]. PhotoDNA rates images based on unique signatures generated from reference images, that are resistant to manipulations. A more flexible approach would support automatic classification of suspicious images based on their content.

Currently, the task of deciding whether images contain pornography is still largely manual and tedious because existing automatic techniques deliver inexact results. Furthermore, the separation of pornographic and non-pornographic images is not always possible. Even for humans it can be a subjective decision. Moreover, the cognitive capacity of humans suffers over time. After hours of browsing pictures, investigators are prone to miss crucial elements because they simply grow tired. Even though full automation in this field might not be possible, assistance in the classification through software is indispensable and there is definite room for improvement among the currently available detection algorithms for pornographic content. When it comes to illegal pornography, tools mostly distinguish these pictures from conventional pornography by image metadata such as filenames or image hashes [17, 18].

Another motivation for this work is grounded in the intrinsic differences when judging pornographic material. Picture material may be offensive, depending on the culture, a country's legal framework, a company's policy of use, or simply the perception of different people.

To address these shortcomings, we present SKIN SHERIFF, a trainable tool to automatically detect pornographic content in images with high precision and recall. SKIN SHERIFF incorporates novel skin detection mechanisms combined with a highly dynamic support vector machine (SVM) to rate unknown, arbitrary images. The detection engine can be trained to target images of specific domains.

We performed an evaluation of SKIN SHERIFF on 13,633 images from the *Compaq* dataset [25]. SKIN SHERIFF outperformed the best-performing reference algorithm [3] with 6.8% higher recall, 8.9% higher precision and 2.2% less false positives on the above mentioned dataset. Furthermore, we tested our approach on a total of 15,776 images from various sources to assess its overall performance.

SFCS'14, June 3, 2014, Kyoto, Japan

http://dx.doi.org/10.1145/2598918.2598920.

In summary, we make the following contributions:

- We present an extensive evaluation of existing skin detection techniques. Based on these results, we introduce a novel method to enhance pixel-based skin detection and extract skin regions from color images.
- We describe a collection of shapes, geometric rules and image filters to create a rich set of features for each individual image.
- We discuss the development of an SVM-based machine learning approach which is capable of rating images based on the available features and decide which are the most descriptive.
- We finally introduce and evaluate SKIN SHERIFF, a comprehensive solution for detecting pornography in color images.

2. RELATED WORK

Before introducing our approach, it is necessary to discuss basic principles and related work in the field of image processing.

2.1 Skin Detection

The basis for almost every pornographic filter is skin detection. The simplest method for separating skin pixels from non-skin pixels is to define thresholds for each channel in a color space. These thresholds or functions can work in all three color coordinates [16, 27, 33, 42, 46], in the ratios of different color channels [21, 30], in the chromatic coordinates (hue and saturation or red-green and yellow-blue channels) [5,11,12,19,20,41,43], in the hue channel [1,26] or in the luminance [39, 45]. Additionally, several color spaces can be combined to improve skin segmentation performance. Examples are combinations of RGB color space and hue channels [35] or R and G coordinates of RGB with the entire HSV color space [40]. These combination can improve skin detection results under various illumination conditions and yield fewer false positive classifications than skin detection with RGB or HSV color space alone. Other approaches use trainable classification systems [25], leverage shape detection [34], or use texture [47] to get a better classification of skin areas [22]. Strictly speaking, they employ a combination of basic, pixel-based skin detection and pattern recognition. Therefore, they are not directly comparable to each other. Since our approach also utilizes combinations of pixel-based detection metrics and shape detection, we decided to develop both components independently and choose the best-performing combination for the final implementation.

2.2 Classification

According to Ruiz del Solar et al. [38], one of the first systems for content-based pornography detection was proposed in 1996 [20]. Skin was detected by using a variation of the RGB color space. Afterwards, the detected skin areas were grouped by humans. Edge detection, symmetry measures and the *Hough transform* were additionally used. They achieved a recall of 52% and a precision of 60% on a set with 1,539 images of which 138 images contained nudity.

Nudity detection based on skin-derived features and a k-nn classifier was proposed by Chan et al. [13]. A text analysis for the detection of pornographic websites was proposed as well. The system was tested on a relatively small dataset of 140 images. A recall of 55% was achieved (65% when hand-segmented skin data was used).

Bosson et al. [6] computed skin-likelihood ratios for all values of the HSV color space. They used five features for the classification of images: the fractional area of the largest skin segment, the number of skin segments, the fractional area of the largest skin segment, the number of colors in the image and the fractional face area. For face detection, a commercial face finder was used. The classification also discriminated between pornographic and nude images. Out of the four different classifiers that were tested, the best classifier achieved an accuracy of 87.2%.

Arentz et al. [4] proposed an image classification system based on skin detection in YCbCr color space. For the image classification, color-, texture-, contour-, placement-, and relative size information of detected skin regions was used. Images were classified as offensive or not offensive.

The approach of Zheng et al. [48] detected skin with a Bayes classifier. Detected skin regions were refined by repetitions of erosion and dilation operations. From the refined skin regions compactness, eccentricity and rectangularity were calculated as features. Image classification was evaluated with three different classifiers on a dataset consisting of 897 offensive and 732 benign images. Best performance was a recall of 89.2% in combination with a false positive rate of 15.3%, which is close to the results delivered by our approach, but on a closed dataset.

The approach most closely related to ours was presented by Rigan Ap-apid [3]. He proposed an *Algorithm for nudity detection* which mainly operates on detected skin areas. One new feature was to calculate a bounding polygon around the biggest three skin areas of the image. The algorithm was tested on a set of 421 pornographic images and 635 non-pornographic images and achieved a recall of 94.32% and a false positive rate of 5.98%. We were, however, not able to reproduce this performance in our experiments.

A forensic tool utilizing Rigan Ap-apid's algorithm for the detection of pornography is NuDetective [18]. In addition to the original algorithm, a filename analysis verified if filenames were suspicious for containing child pornography. All images classified as suspicious by the tool were checked by forensic examiners again. As a result, a low false positive rate was not considered that important in this case. Furthermore, the integrated metadata evaluation (filename, hidden thumbnails, etc.) is especially tailored to the law-enforcement domain.

One of the few trainable approaches is presented by Karavarsamis et al. [28, 29]. Here the authors use fixed RGB thresholds for skin detection and group their results in ROI's (regions of interest) which they later use to train a random forest decision tree. Although closely related to our approach, our evaluation shows that skin regions within the convex hull of a skin map alone are not descriptive enough to use as training features.

3. APPROACH

As mentioned above, the basis for our approach is a well-performing detection mechanism for exposed skin in a picture. Although it is possible to use texture, structural irregularities, images signatures or even trainable classifiers for pixel-based detection, we decided to use such advanced methods only later in the classification step, where they can actually act as features.

Figure 1: Autocontrast example.

Figure 2: Histogram equalization example.

3.1 Skin

In general, SKIN SHERIFF evolved around the basic problem of skin detection in color images. This, in turn, comes with a set of problems that have to be addressed first.

3.1.1 Preprocessing

Not all images have the same quality. Some are over- or underexposed, may contain compression artifacts or are simply of a small size and therefore not rich in detail. In order to use a pixel-based skin detection mechanism, it is important to equalize input pictures as good as possible. Naturally, a lot of possibilities to equalize images exist. To evaluate which combination performs best, we permutated the following preprocessing operations before applying our skin detection mechanism:

- Gamma correction
- Histogram equalization
- Gray world normalization
- Autocontrast
- Max-RGB

After applying each method, we created a matrix of all results, and performed a factor analysis to evaluate which preprocessing functions are beneficial to the skin detection algorithm. According to this evaluation, the autocontrast function alone proved to be sufficient as a preprocessing method. At first glance this may come as a surprise. However, for a pixel-based detection approach it produces exactly the desired effect: This function essentially stretches a picture's color range to its maximum values. In other words, in underexposed images, where skin appears darker than under normal conditions, autocontrast shifts this color range into a detectable, brown range. The same is true for overexposed images.

Figure 1 shows the effect of the autocontrast preprocessing. Less saturated color channels like blue are stretched over the whole spectrum and therefore shift the skin tone towards a more natural color with less red. In comparison, Figure 2 shows the same picture with histogram equalization. Judging from the picture alone it seems obvious that Figure 2 produces better results since the picture looks more natural. But since this method tries to achieve a uniform color distribution throughout the whole picture, it distorts colors such that they are harder to detect as skin than with autocontrast. To substantiate this claim with numbers, we calculated the F-measure for both filters and 30 different color spaces resulting in scores between 0.396 and 0.381 for autocontrast and between 0.259 and 0.241 for histogram equalization.

As further preprocessing steps, we also resized and rotated our input images according to their EXIF information. However, these operations are only important for shape detection and processing speed. We provide a more detailed description of our preprocessing steps in Section 4.

3.1.2 Color Space

After the preprocessing phase, each pixel can independently be classified as skin or not skin. To this end, we transform our source picture from the RGB color space to HSV, YUV, TSL and LCCS. For each color space, we define a range of values corresponding to skin areas, and test their detection rate on a specifically labeled test dataset. This dataset is provided by Compaq [25] and contains thousands of pictures with manually labeled skin pictures. This allowed us to create a factor analysis of all color spaces and choose the optimal range for selecting skin pixels.

3.1.3 Postprocessing

Finally, the result is a skin map, showing a black and white representation of image areas containing skin. To get homogenous areas, we employ postprocessing filters to eliminate single pixels, fill holes in overexposed areas and smoothen the skin map as a whole. Similar to the previous sections, we applied several morphological filters and chose the best-performing based on the factor analysis. We tested the following filters:

- *Dilation and erosion:* This filter can either be used to eliminate single pixels (erosion) or close gaps (dilation) in skin areas.
- *Opening and Closing:* Erosion followed by delation (opening) and vice versa.
- *Canny edge detection*
- *Hough transformation:* Detects straight lines in images.
- *Sobel operator:* Evaluates the strength of edge points and their orientation.

Since edge-finding morphological filters reduce the area of a skin map, we only use a closing function for postprocessing skin maps.

It is important to note here, that our skin detection approach comes with some limitations. There are, for instance, objects that have the same color as skin and therefore produce false positives in the detected skin areas. These excep-

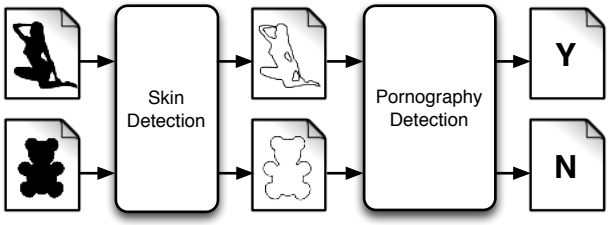


Figure 3: Overview of SKIN SHERIFF.

tions have to be handled by the pornography detection algorithm and cannot be avoided in all cases. Secondly, this approach is naturally blind to black and white pictures, a limitation it shares with all color-based detection approaches.

3.2 Pornography Detection

As illustrated in Figure 3, SKIN SHERIFF consists of two stages: First, the *Skin Detection* component identifies continuous patches of skin as described above and in detail in Section 4. Second, the *Pornography Detection* component (described in Section 5) identifies necessary features to provide the input for our SVM-based image classification. The final result is a boolean assessment if the image contains pornography or not.

4. SKIN DETECTION

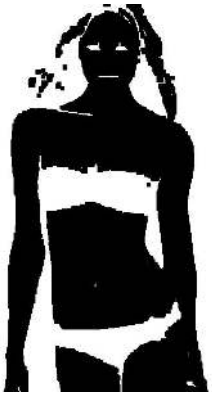
The goal of the *Skin Detection* component is to extract all skin areas from an image. We classify and label all pixels separately and mark them in a binary image. We label pixels that we classify as skin in grey, and pixels we classify as non-skin pixels in white. Thus, we extract a binary or grey leveled image with labels for skin and non-skin pixels, a so-called *skin map*, from each image under investigation. An example of a skin map is shown in Figure 4.

We already mentioned in Section 2 that a threshold on each color channel is the simplest way to distinguishing between skin pixels and non-skin pixels. Existing approaches claim a very high precision/recall ratio, most of them above 90% total. However, for an objective performance comparison of different skin color modeling methods, identical testing conditions are needed. Since many skin detection methods measure their performance on their own, publicly unavailable datasets, their performance is hard to verify. To amend this shortcoming, we utilized the previously introduced *Compaq* dataset [25, 44].

In Table 1, we list the most important pixel-based skin detection algorithms on the same dataset and compare their performance to our method. While many approaches actually exhibit a decent true positive rate, their false positive ratio is unacceptable for many use cases. False positive rates between 20% and 30% have a severe negative impact on the subsequently executed classification algorithm.

We evaluated several thresholds in different color spaces as well as combinations of different thresholds in order to increase the precision and reduce the false positive rate of our skin detection algorithm. We combined all thresholds, defining skin in their respective color space, with a logical AND: We only labeled a pixel as skin, if it was labeled as skin by both thresholds separately. As we considered recall an important metric for skin detection we selected RGBHSV$_2$, a combination of RGB [33] and HSV$_2$ [40], as the best performing threshold combination. This decision is based on the evaluation results presented in Table 4 in Section 6.

(a) Original image.



(b) Extracted skin map.

Figure 4: Skin Detection example.

Method	Color space	TPR	FPR
Bayes SPM [25]	RGB	80% 90%	8.5% 14.2%
Bayes SPM [8]	RGB	93.4%	19.8%
Maximum Entropy model [23]	RGB	80%	8%
Gaussian Mixture models [25]	RGB	80% 90%	∼9.5% ∼15.5%
SOM [9]	TS	78%	32%
Elliptical boundary model [31]	CIE-xy	90%	20.9%
Single Gaussian [31]	CbCr	90%	33.3%
Gaussian Mixture [31]	IQ	90%	30%
Thresholding of I axis [8]	YIQ	94.7%	30.2%
Our approach (RGB + HSV)	RGBHSV$_2$	82.3%	11.4%

Table 1: Performance of different skin detection methods on the *Compaq* dataset [44].

Algorithm 1 finally depicts our skin detection algorithm. As a first preprocessing step, we scale the image to a width of less than 1000 pixels. This threshold has no measurable impact on our detection results but improves performance tremendously. Afterwards we execute an autocontrast function to maximize the image's contrast. As a result, the darkest pixel becomes black and the lightest pixel becomes white. We then calculate the coordinates of each pixel in the RGB and HSV color spaces and detect skin pixels based on the thresholds listed in Equations 1 and 2. As a last step, we perform the previously mentioned closing operation on the extracted skin map in order to close holes and fissures in the detected skin areas. Finally, we output a skin map of the detected skin pixels for further processing by the *Pornography Detection* component.

$$\begin{aligned}(R > 220 \wedge G > 210 \wedge B > 170 \wedge |R-G| > 15 \wedge R > B \wedge G > B) \vee \\ (R > 95 \wedge G > 40 \wedge B > 20 \wedge \max(R,G,B) - \min(R,G,B) > 15 \wedge \\ |R-G| > 15 \wedge R > G \wedge R > B)\end{aligned} \quad (1)$$

$$(0 \leq H \leq 50 \vee 340 \leq H \leq 360) \wedge 0.2 < S \wedge 0.35 < V \quad (2)$$

Algorithm 1 Skin Detection.

```
1: function SkinDetect(img, img_width, img_height)
2:     Scale(img, width < 1000px)
3:     AutoContrast(img)
4:     skinmap ← NewImage(img_width, img_height, white)
5:     for all pixel in img do
6:         R, G, B ← pixel
7:         H, S, V ← ConvertRGBtoHSV(R, G, B)
8:         if IsSkin(R, G, B, H, S, V) then
9:             skinmap[pixel_x, pixel_y] ← grey
10:        else
11:            skinmap[pixel_x, pixel_y] ← white
12:        end if
13:    end for
14:    grey_closing(skinmap, size ← (6, 6))
15:    return skinmap
16: end function
```

5. PORNOGRAPHY DETECTION

Once the major skin areas are identified, SKIN SHERIFF must decide whether an image contains pornographic content or not. One of the most popular methods to do this on arbitrary skin maps was developed by Rigan Ap-apid [3]. This algorithm calculates a bounding polygon around the biggest three skin areas of an image. The percentage of skin pixels inside the polygon area is then used for classification. Other important features are the total amount of skin in the image, the number of skin areas and the sizes of the three largest skin areas. An example of such a skin map with skin areas and bounding polygon is shown in Figure 5. The idea behind this set of heuristics is that skin maps from pornographic content are large, connected areas, while ordinary pictures show these areas disconnected by clothes. An extension of this work [18] also leverages skin area sizes but relies on a convex hull instead of the bounding polygon to get a better idea of how connected areas really are.

Our experiments show, that these heuristics have a decent detection rate. There are, however, certain cases where they are not delivering the expected results. Since most of the detection mechanism relies on the amount and percentage of *connected* skin areas, any case where large skin areas get separated leads to false negatives. Examples are, for instance, necklaces, ribbons, handrails or simply local over-/underexposure from flashlights. Furthermore, these algorithms are hardcoded and cannot adapt to varying illumination conditions.

As a result, we use a more versatile classification method utilizing SVM-based machine learning. Here, a multi dimensional vector space is spanned by arbitrary features of an image. These features must be chosen carefully to describe the classified category as precise as possible. The vector space can then be populated with training data (e.g. from pornographic and non-pornographic images) and subsequently used to binary classify other images. With this method it is possible to train the classification engine to specifically detect pornography in indoor lighting conditions, for example.

Obviously, a very integral part of our approach is represented by the features we use to train our SVM. As a first

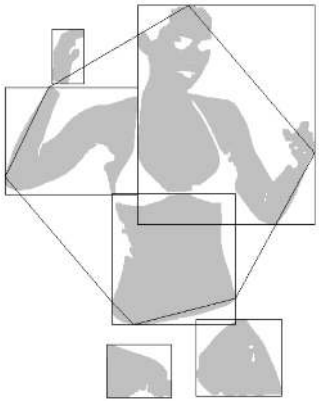


Figure 5: Skin map with area polygons and bounding box.

step we create an image's skin map according to the method described in Section 4. Then, we extract all connected skin region in the skin map with a flood fill algorithm. For each skin region we create the following features:

- *Rectangularity:* Ratio of a skin region's area to the area of its minimum bounding rectangle.
- H_{mean}, S_{mean} *and* V_{mean}*:* Mean values of the pixels belonging to the skin region in the HSV color space
- *Mean grey value*
- *Eccentricity, ellipticity, orientation:* We compute these features from the central moments [10] of a skin region. The central moment is calculated as

$$\mu_{pq} = \sum_{x,y \in \mathcal{R}} (x - \overline{x})^p \cdot (y - \overline{y})^q \qquad (3)$$

  and essentially represents a shift of the origin coordinates to this region's center.
- *Amount of border touching pixels:* We count the number of pixels of a skin region touching the borders of the whole image.
- *Number of touched corners:* We count the number of image corners touched by the skin region.
- *Amount of skin:* Percentage of skin compared to the size of the whole image.
- *Hue standard deviation:* The standard deviation of the hue component of the HSV_2 color space.
- *Perimeter:* We calculate the perimeter of the skin region by a contour tracing algorithm.
- *Compactness:* The relation between the skin region's area and its perimeter.
- *Centroid* of the skin area.

As with the skin detection mechanism, we included as many features as possible, rated their corresponding Fisher scores (F-scores) [15] and only kept the most descriptive features for the final implementation. We use some of these features (rectangularity, eccentricity, orientation, compactness, mean grey value, centroid and H_{mean}) for classification of the image with our SVM. As our focus lies on optimizing the precision of SKIN SHERIFF we experimented with and implemented several measures to reduce the false positive rate. Thus, we use the other features (amount of border touching pixels, the number of touched corners, the hue standard

Figure 6: Skin map of a crib (detected by *EliminateStraightAreas*).

Figure 7: Skin maps of a red record disc, yellow rubber duck and brown case (detected by *ShapeElimination*).

deviation, S_{mean} and V_{mean}) for the following false positive prevention measures.

CheckSpatial() This function analyzes the spatial distribution of the detected skin region. It fragments the image into nine parts and calculates the percentage of skin present in the central part. If this percentage is below a certain threshold (in our experiments we evaluated the optimal threshold as 29%), the image is rated non-pornographic.

CheckFace() The aim of this function is to prevent portrait shots being falsely classified as pornographic due to the high amount of skin present. First, we use the face detection of *OpenCV* [7] to flag faces by a bounding rectangle. We further detect frontal and profile faces with *Haar classifier cascades*. If the percentage of skin present in the face is greater than a specified threshold (in our experiments 38%), the image is rated as non-pornographic.

EliminateStraightAreas() This function aims at eliminating parts of an image detected as skin, that exhibit contours that are too straight for a human shape such as the example shown in Figure 6. We detect edges of skin areas with the Canny edge detector and lines with a probabilistic Hough transform. Subsequently, we remove skin areas containing at least one end of a straight line.

ShapeElimination() The main goal of this function was eliminating shapes like the examples in Figure 7. This function eliminates skin areas based on four different rules that check whether an area possibly belongs to a human subject or not. Rules 4 and 5 remove skin areas which were too compact, while rule 6 eliminates areas that are not compact enough. Finally, rule 7 eliminates sundown images that are often detected as pornographic due to their color.

$$\text{Rectangularity}(\mathcal{R}) > 0.81 \vee \text{Compactness}(\mathcal{R}) > 0.8 \quad (4)$$

$$\begin{aligned} &\text{Rectangularity}(\mathcal{R}) > 0.75 \wedge \\ &\text{Compactness}(\mathcal{R}) > 0.75 \wedge \\ &\text{Ellipticity}(\mathcal{R}) > 0.75 \end{aligned} \quad (5)$$

$$\text{Compactness}(\mathcal{R}) < 0.1 \quad (6)$$

$$\begin{aligned} &\text{Length}(\mathcal{R}) * 1.1 > \text{Length}(\mathcal{I}) \wedge \\ &\text{Area}(\mathcal{R}) * 2 < \text{Area}(\mathcal{I}) \wedge \\ &\text{Rectangularity}(\mathcal{R}) > 0.60 \end{aligned} \quad (7)$$

Figure 8: Skin maps illustrating the *EliminateDifferentAreas* function (normally grey leveled skin map, edge reinforced skin map and skin map after removal of areas).

Figure 9: Skin map of a plane in front of a sunset before and after *BorderTouchingAreasElimination*.

EliminateDifferentAreas() This function utilizes differences between areas which were detected as skin in order to eliminate areas that are likely not skin such as illustrated in Figure 8. We observed, that the detected area located in the center of the image has the highest probability to represent skin. Therefore, we first select the most central skin area. Then, we reinforce the edges to separate different skin areas. Finally, we remove skin areas, whose mean hue and mean saturation values deviated too much from the central skin area. Since most images are taken with a person in or around its center, this method is capable of eliminating large portions of falsely detected skin.

BorderTouchingAreasElimination() This function aims at eliminating skin-like backgrounds. We remove a skin area if it touches at least two corners of the image and at least $image_{length} + image_{width}$ image border pixel.

After eliminating certain skin areas or even whole pictures, that are unlikely to be pornographic, we subject the rest of the image to classification by an SVM. For each image, we extracted 43 features, with 40 of those features being derived from the five largest skin areas:

- Size in percent from the whole image
- Compactness
- Ellipticity
- Rectangularity
- Eccentricity
- Orientation
- H_{mean}

To avoid overfitting the SVM, we selected the most decisive features based on their F-score. Table 2 shows the 7 most distinctive features as computed by the feature selection tool of LIBSVM [14] that we used for the final classification. $Size_1$ describes the percentage of the largest skin area, $Size_2$ the percentage of the second largest skin area. $Compactness_1$ describes the compactness of the largest skin

Feature	F-score
Skinfillrate	0.251
Size_1	0.207
Polygonfillrate	0.126
Compactness_1	0.111
Rectangularity_1	0.0620
Size_2	0.044
Rectangularity_2	0.005

Table 2: Selected features for the pornography detection with their F-score.

Algorithm 2 Pornography Detection.

```
1: function ClassifySkinmap(skinmap)
2:     skinareas ← FloodFill(skinmap)
3:     EliminateSmallAreas(skinareas)
4:     features ← []
5:     for skinarea_n in skinareas do
6:         features_n ← ExtractFeatures(skinares)
7:     end for
8:     skinareas ← ShapeElimination(skinareas)
9:     if CheckSpatial(skinareas) then
10:        return false
11:    end if
12:    if CheckFace(skinareas) then
13:        return false
14:    end if
15:    ConvexHull(max(skinareas, 3))
16:    ConvexHullFillRate
17:    ScaleFeatures(features)
18:    isNude ← SVMClassification(features)
19:    return isNude
20: end function
```

area, Rectangularity_1 and Rectangularity_2 the rectangularity of the largest and second largest skin areas.

The resulting F-score describes how well a feature describes a certain class. Our evaluation (see Section 6) shows, that EliminateStraightAreas, EliminateDifferentAreas and BorderTouchingAreasElimination did not noticeably increase our classifier's performance. Therefore, these strategies were removed from the final pornography detection.

With these elements together, we can finally depict our pornography detection approach in Algorithm 2.

6. EVALUATION

With all components of our approach in place, we now provide a detailed evaluation of the algorithms introduced in the previous sections.

6.1 Datasets

To test our skin detection approach we used the *Compaq* [25] dataset. This collection of images is publicly available and contains a multitude of various images with a hand-labeled set of skin maps. This set was our ground truth to evaluate retrieval performance of both, our own approach and the most commonly used skin detection approaches from related work.

To evaluate the classification performance of different approaches for pornography detection we collected a comprehensive dataset of pornographic and non-pornographic images that also incorporates the *Compaq* dataset. We manually classified all images of the *Compaq* dataset as offensive or ordinary images. The whole dataset was organized in six different categories amounting to 15,776 different images.

- 8,964 images of the *Compaq* dataset, which did not contain any skin.
- 3,841 images of the *Compaq* dataset, which did not contain pornography but showed skin.
- 828 images of the *Compaq* dataset, which contained pornography.
- 990 images without pornography. 812 of them were collected from the *INRIA Holidays* dataset [24].
- 151 swimwear images, displaying male and female models with a high amount of skin. These images were collected from different websites. All images featured good quality and similar illumination conditions.
- 1,002 images containing pornography, which were collected from different pornographic websites.

Overall, the dataset contained 13,946 non-pornographic images and 1,830 images showing pornography. For experiments with different SVM configurations, the dataset was split into a training and a test dataset. For the training dataset, the following images were used:

- 372 images of the *Compaq* dataset, which did not contain any skin.
- 363 images of the *Compaq* dataset, which did not contain pornography but showed skin.
- 441 images of the *Compaq* dataset, which contained pornography.
- 645 images without pornography.
- 42 swimwear images, displaying male and female models.
- 578 images showing pornography.

In total, the training dataset contained 1,019 images with pornographic content and 1,422 images without. Because images in which no skin areas were found were skipped during the training process, exactly 1,005 non-pornographic and 1,005 pornographic images were used for training. Consequently, the SVM was trained on a balanced dataset. This ensures the best separation of both classes. For the test dataset all remaining images were used:

- 8,592 images of the *Compaq* dataset, which did not contain any skin.
- 3,478 images of the *Compaq* dataset, which did not contain pornography but showed skin.
- 387 images of the *Compaq* dataset, which contained pornography.
- 345 images without pornography.
- 109 swimwear images, displaying male and female models.
- 424 images showing pornography.

In total the test dataset consisted of 811 pornographic images and 12,524 non-pornographic images.

Method	Recall	Precision	FPR	Accuracy	F-measure
RGB_{1D}	0.91	0.204	0.295	0.721	0.333
RGB	0.929	0.197	0.315	0.704	0.325
HS	0.78	0.2	0.259	0.744	0.319
YUV-YIQ	0.874	0.191	0.308	0.706	0.313
LCCS	0.641	0.188	0.229	0.761	0.291
HSV_1	0.855	0.163	0.364	0.653	0.274
$YCbCr_1$	0.776	0.163	0.331	0.678	0.269
HSV_2	0.926	0.144	0.459	0.571	0.249
MDL	0.988	0.124	0.582	0.461	0.22
HSI	0.995	0.112	0.653	0.397	0.202
TSL	0.326	0.128	0.184	0.779	0.184
$YCbCr_2$	0.526	0.1	0.393	0.601	0.168
YUV	0.564	0.087	0.492	0.512	0.15

Table 3: Results of skin detection in different color spaces.

Method	Recall	Precision	FPR	Accuracy	F-measure
RGBHS	0.757	0.263	0.176	0.819	0.39
RGB_{1D}RGB	0.87	0.241	0.227	0.78	0.378
$RGBHSV_2$	0.893	0.239	0.236	0.774	0.377
$RGBHSV_1$	0.813	0.236	0.219	0.784	0.366
RGBYUV-YIQ	0.854	0.23	0.237	0.77	0.363
RGBHCbCr	0.61	0.233	0.166	0.816	0.338
$RGBYCbCr_1$	0.73	0.21	0.228	0.769	0.327
$RGB_{1D}YCbCr_1$	0.74	0.204	0.24	0.758	0.319
RGB_{1D}HCbCr	0.639	0.199	0.214	0.775	0.303
$RGBYCbCr_2$	0.48	0.188	0.172	0.802	0.27
$RGB_{1D}YCbCr_2$	0.491	0.184	0.181	0.794	0.267

Table 4: Results of skin detection in combinations of color spaces.

6.2 Skin Detection

We evaluated our skin detection approach separately on our datasets without any image processing in order not to falsify the classification results. Table 3 shows the classification performance for different color spaces. The results exhibited high recall and high false positive rates. Also, precision was quite low for all thresholds. In order to improve the skin detection rates, we combined several different color spaces. We focused on combining color spaces with high recall (such as RGB and HSV_2) to improve precision and false positive rate.

The results for color space combinations are shown in Table 4. Again, no image processing was used. The best performing combination was RGBHS with an F-measure of 39%. The top six combinations outperformed all skin detection approaches which were executed in only one color space in terms of F-measure. Highest recall was achieved with the thresholds for $RGBHSV_2$ with 89.3%, highest precision with RGBHS with 26.3%. Lowest false positive rate was achieved with RGBHCbCr with 16.6%. Both, RGB and HS individually achieved a high recall. The recall achieved through the combination of both was only slightly lower than the recall achieved through application of HS alone. However, precision increased drastically: RGB delivered a precision of 19.7%, HS delivered 20.0% but the combination of both achieved 26.3%. Similar results were achieved for $RGBHSV_2$ and $RGBHSV_1$: thresholds of HSV_2 and HSV_1 only achieved F-measures of 24.9% and 27.4%, the respective combination with RGB resulted in F-measures of 37.7% for $RGBHSV_2$ and 36.6% for $RGBHSV_1$.

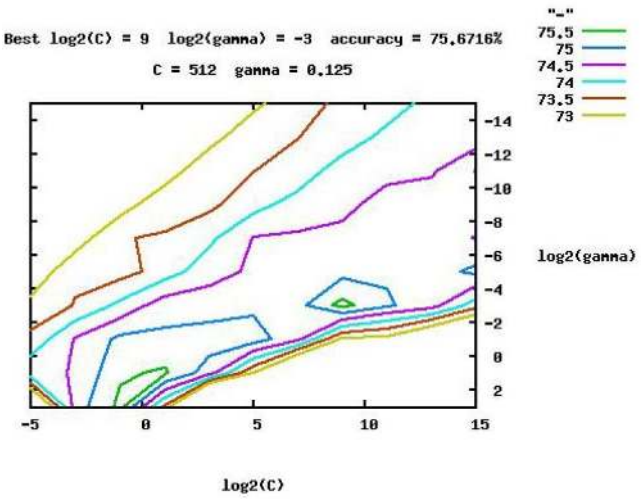


Figure 10: Different accuracies on the training dataset, computed by *LIBSVM's easy.py*-script by cross-validation.

We further evaluated the best skin detection results of related pixel-based skin detection approaches on the *Compaq* dataset alone. In Table 1, we show our proposed $RGBHSV_2$ skin detection approach in comparison to these other methods. Although different methods use slightly different separation of the dataset into training and test dataset, the table gives an overall picture of the classification performance of the respective methods. The overall performance depends on the evaluation criteria of the desired application of skin detection. For pornography detection, we consider the false positive rate very important. In this respect, $RGBHSV_2$ was outperformed by the Maximum Entropy Model [23], which delivered the lowest reported false positive rate (8%). However, this method also shows one of the lowest true positive rates with 80%.

6.3 Pornography Detection

Analogous to our skin detection approach, we also evaluated our SVM-based pornography detection. We selected the 21 features with the best F-score to avoid overfitting our SVM. After training it with our labeled training set, we adjusted the SVM parameters for maximum accuracy. Figure 10 shows different accuracies on the training dataset which were achieved with different SVM training parameters by cross-validation. The best reported accuracies were achieved with $C = 512$ and $\gamma = 0.125$ for the SVM parameters.

Table 5 finally compares the performance of our SVM with the performance of the *Algorithm for nudity detection* on the same test dataset. Depicted is the classification performance of the SVM with and without the false positive prevention. It clearly shows, that the SVM-based pornography detection is superior in all rates in both cases. The better performance is not based on differences in the skin detection method, because the $RGBHSV_2$ method was used in both cases to produce the skin map. Therefore, with a final recall of 65.7% and 6.4% false positive rate, our approach outperforms the best reported, image-based pornography detection approach.

Classifier	Recall	Precision	FPR	Accuracy	F-measure
Algorithm for nudity detection	0.589	0.309	0.086	0.895	0.405
SKIN SHERIFF (linear SVM with FP prevention)	0.657	0.398	0.064	0.919	0.496
SKIN SHERIFF (linear SVM)	0.702	0.353	0.083	0.904	0.47

Table 5: Comparison of the classification performance of the *Algorithm for nudity detection* and the selected SVM configuration on the test dataset, with and without false positive (FP) prevention.

7. LIMITATIONS

While SKIN SHERIFF outperformed current state-of-the-art pornography detection approaches, there is still room for improvement. The most fundamental limitation of the pornography detection is caused by the skin detection itself. Nudity or pornography can only be detected if skin is present. Furthermore, color-based nudity detection is naturally blind to black and white pornography images [47]. Also, skin detection will not work for skin which is too dark or too bright. In many of our test images, this was at least partially the case. Figure 11 shows a test image from the *Compaq* dataset as an example for overexposed skin. The left portion of visible skin is very close to absolute white. The precise skin map, however, marks it as a valid skin tone. The same borderline cases exist for underexposed skin areas and black, of course. As a result, skin detection results can never be perfect and a certain amount of false positives are unavoidable. Furthermore, nudity or pornography also cannot be detected, if only a small amount of total skin is present in an image.

Another limitation is the runtime of SKIN SHERIFF, even though it is reduced through scaling of images. Some image processing operations were implemented with functionality in mind rather than speed. Suboptimal data structures used for storing images also increase the runtime.

Another point worth mentioning is the subjective definition of *offensiveness*. Even in our research group, we were not really clear on how to define a separation between pornography, nudity and unoffensive images. These concepts vary from country to country and even between individuals. An advantage of our trainable system is, that it can be aligned to reflect individual subjective views, or, in the case of law enforcement, legal boundaries.

8. FUTURE WORK

With our pornography detection, we took the first step to create a tool that helps law enforcement forensics to automatically screen files for digital contraband. In a next step, we will devise a method to estimate the age of individuals on contraband images and categorize them in adults and underage persons. In collaboration with a Wisconsin Police Department, we are currently assessing face recognition algorithms and their potential to solve this problem. Ultimately, the work presented here, paired with an age detection method, constitutes a tool for investigating digital contraband.

A logical evolution of image-based skin detection is video analysis. The conceptual problem is the same as in the image domain. Therefore, we can use our existing approach and apply it on images captured from videos. To improve skin and shape detection, other frames of the same video could be used.

Finally, we plan to enhance the overall performance of our approach. The current throughput of our system is roughly

Figure 11: Example for overexposed skin making skin detection more difficult.

2,000 images per hour on a single core 2GHz Machine. Since our prototype is neither tweaked for execution time nor distributed on multiple cores, we are confident that throughputs of 10,000 and more images per hour are possible without heavy modifications.

9. CONCLUSION

In this paper, we presented SKIN SHERIFF, a novel approach to detect pornography in arbitrary images. We evaluated the best performing skin detection algorithms and compared them to our approach. Our evaluation shows, that our method delivers a 82.3% true positive and 11.4% false positive rate on the public *Compaq* dataset. To the best of our knowledge, this performance is unmatched by any of the other approaches. Our SVM-based pornography detection delivers an accuracy of 91.9% and outperforms the most common approaches. We received very positive feedback from law enforcement, where our tool was used to preselect digital contraband from ordinary images.

Acknowledgments

The research leading to these results has received funding from the European Union Seventh Framework Programme under grant agreement n° 257007 (SysSec) and from the FFG – Austrian Research Promotion under grant COMET K1.

10. REFERENCES

[1] Skin detection in digital images. http://popscan.blogspot.com/2012/08/skin-detection-in-digital-images.html, 2012.

[2] Y. Akdeniz, D. Wall, and C. Walker. *The Internet, Law and Society*. Longman, New York, USA, 2000.

[3] R. Ap-apid. An Algorithm for Nudity Detection. In *Proceedings of the 5th Philippine Computing Science Congress (PCSC)*, 2005.

[4] W. A. Arentz and B. Olstad. Classifying Offensive Sites Based on Image Content. *Computer Vision and Image Understanding*, 94, 2004.

[5] S. Baskan, M. M. Bulut, and V. Atalay. Projection Based Method for Segmentation of Human Face and Its Evaluation. *Pattern Recognition Letters*, 23(14), 2002.

[6] A. Bosson, G. C. Cawley, Y. Chan, and R. Harvey. Non-retrieval: Blocking Pornographic Images. In *Proceedings of the International Conference on Image and Video Retrieval (CIVR)*, 2002.

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

[8] J. Brand and J. S. Mason. A Comparative Assessment of Three Approaches to Pixel-Level Human Skin-Detection. In *Proceedings of the 15th International Conference on Pattern Recognition (ICPR)*, 2000.

[9] D. A. Brown, I. Craw, and J. Lewthwaite. A SOM Based Approach to Skin Detection with Application in Real Time Systems. In *Proceedings of the British Machine Vision Conference (BMVC)*, 2001.

[10] W. Burger and M. J. Burge. *Principles of Digital Image Processing: Core Algorithms*. Springer, 2009.

[11] D. Chai and K. N. Ngan. Locating Facial Region of a Head-and-Shoulders Color Image. In *Proceedings of the 3rd International Conference on Face & Gesture Recognition (FG)*, 1998.

[12] D. Chai and K. N. Ngan. Face Segmentation Using Skin-Color Map in Videophone Applications. *IEEE Transactions Circuiuts and Systems for Video Technolology*, 9(4), 1999.

[13] Y. Chan, R. Harvey, and D. Smith. Building Systems to Block Pornography. In *Challenge of Image Retrieval*, 1999.

[14] Y. W. Chen. LIBSVM Tools: Feature selection tool. http://www.csie.ntu.edu.tw/~cjlin/libsvmtools/#feature_selection_tool.

[15] Y. W. Chen and C. J. Lin. Combining SVMs with Various Feature Selection Strategies. In *Feature Extraction, Foundations and Applications*. Springer, 2006.

[16] K.-M. Cho, J.-H. Jang, and K.-S. Hong. Adaptive skin-color filter. *Pattern Recognition*, 34(5), 2001.

[17] P. M. da Silva Eleuterio and M. de Castro Polastro. Identification of High-Resolution Images of Child and Adolescent Pornography at Crime Scenes. *The International Journal of Forensic Computer Science*, 5(1), 2010.

[18] M. de Castro Polastro and P. M. da Silva Eleuterio. NuDetective: A Forensic Tool to Help Combat Child Pornography through Automatic Nudity Detection. In *Proceedings of the Workshop on Database And Expert Systems Applications (DEXA)*, 2010.

[19] M. M. Fleck, D. A. Forsyth, and C. Bregler. Finding Naked People. In *Proceedings of the 4th European Conference on Computer Vision (ECCV)*, 1996.

[20] D. A. Forsyth and M. M. Fleck. Automatic Detection of Human Nudes. *International Journal of Computer Vision*, 32(1), 1999.

[21] T. Gevers, R. Aldershoff, and A. W. M. Smeulders. Classification of images on the internet by visual and textual information. In *Proc. SPIE 3964, Internet Imaging*, 1999.

[22] W. Hu, H. Zuo, O. Wu, Y. Chen, Z. Zhang, and D. Suter. Recognition of Adult Images, Videos, and Web Page Bags. *ACM Trans. Multimedia Comput. Commun. Appl.*, 7S(1):28:1–28:24, Nov. 2011.

[23] B. Jedynak, H. Zheng, M. Daoudi, and D. Barret. Maximum Entropy Models for Skin Detection. In *Proceedings 3rd Indian Conference on Computer Vision, Graphics and Image Processing*, 2002.

[24] H. Jégou, M. Douze, and C. Schmid. Hamming Embedding and Weak Geometric Consistency for Large Scale Image Search. In *Proceedings of the 10th European Conference on Computer Vision (ECCV)*, 2008.

[25] M. J. Jones and J. M. Rehg. Statistical Color Models with Application to Skin Detection. In *International Journal of Computer Vision*, volume 46, 2002.

[26] L. Jordao, M. Perrone, J. P. Costeira, and J. Santos-Victor. Active Face and Feature Tracking. In *Proceedings of the 10th International Conference on Image Analysis and Processing (ICIAP)*, 1999.

[27] A. M.-B. Jorge, A.-T. Gualberto, S.-P. Gabriel, T.-M. L. Karina, and H. M. Perez-Meana. Detection of Pornographic Digital Images. *International Journal of Computers*, 5(2), 2011.

[28] S. Karavarsamis, N. Ntarmos, , K. Blekas, and I. Pitas. Detecting Pornographic Images by Localizing Skin ROIs. *International Journal of Digital Crime and Forensics (IJDCF)*, 5(1), 2013.

[29] S. Karavarsamis, I. Pitas, and N. Ntarmos. Recognizing Pornographic Images. In *Proceedings of the 14th ACM Workshop on Multimedia and Security (MM&Sec)*, 2012.

[30] B. Khanal and D. Sidibé. Efficient Skin Detection under Severe Illumination Changes and Shadows. In *Proceedings of the 4th International Conference on Intelligent Robotics and Applications (ICIRA)*, 2011.

[31] J. Y. Lee and S. I. Yoo. An Elliptical Boundary Model for Skin Color Detection. In *Proceedings of the International Conference on Imaging Science, System and Technology*, 2002.

[32] S. Lohr. Microsoft Tackles the Child Pornography Problem. http://bits.blogs.nytimes.com/2009/12/16/microsoft-tackles-the-child-pornography-problem/, 2009.

[33] P. Peer, J. Kova, and F. Solina. Human Skin Colour Clustering for Face Detection. In *The IEEE Region 8 EUROCON 2003. Computer as a Tool.*, 2003.

[34] S. L. Phung, D. Chai, and A. Bouzerdoum. Adaptive Skin Segmentation in Color Images. In *Proceedings of the International Conference on Acoustics, Speech,*

and Signal Processing (ICASSP), 2003.

[35] N. Rahman, K. Wei, and J. See. RGB-H-CbCr Skin Colour Model for Human Face Detection. In *Proceedings of the MMU International Symposium on Information & Communications Technologies (M2USIC)*, 2006.

[36] R. Richmond. Facebook's New Way to Combat Child Pornography. `http://gadgetwise.blogs.nytimes.com/2011/05/19/facebook-to-combat-child-porn-using-microsofts-technology/`, 2011.

[37] J. Ropelato. Internet Pornography Statistics. `http://internet-filter-review.toptenreviews.com/internet-pornography-statistics.html`.

[38] J. Ruiz del Solar, V. Castaneda, R. Verschae, R. A. Baeza-Yates, and F. Ortiz. Characterizing Objectionable Image Content (Pornography and Nude Images) of Specific Web Segments: Chile as a Case Study. In *Proceedings of the 3rd Latin American Web Congress (LA-WEB)*, 2005.

[39] O. Severino Jr. and A. Gonzaga. HSM: A New Color Space used in the Processing of Color Images. *Revista de Informatica Teorica e Aplicada*, 16(2), 2010.

[40] B. Smolka, K. Czubin, J. Y. Hardeberg, K. N. Plataniotis, M. Szczepanski, and K. Wojciechowski. Towards Automatic Redeye Effect Removal. *Pattern Recognition Letters*, 24(11), 2003.

[41] K. Sobottka and I. Pitas. Extraction of Facial Regions and Features using Color and Shape Information. In *Proceedings of the International Conference Of Pattern Recognition (ICPR)*, 1996.

[42] M. R. Tabassum, A. U. Gias, M. M. Kamal, H. M. Muctadir, M. Ibrahim, A. K. Shakir, A. Imran, S. Islam, M. G. Rabbani, S. M. Khaled, M. S. Islam, and Z. Begum. Comparative Study of Statistical Skin Detection Algorithms for Sub-Continental Human Images. *Information Technology Journal*, 9(4), 2010.

[43] F. Tomaz, T. Candeias, and H. Shahbazkia. Improved Automatic Skin Detection in Color Images. In *Proceedings fo the 7th International Conference on Digital Image Computing: Techniques and Applications (DICTA)*, 2003.

[44] V. Vezhnevets, V. Sazonov, and A. Andreeva. A Survey on Pixel-Based Skin Color Detection Techniques. In *Proceedings of GraphiCon*, 2003.

[45] C. Wang and M. S. Brandstein. Multi-Source Face Tracking With Audio And Visual Data. In *Proceedings of the 3rd IEEE Workshop on Multimedia Signal Processing (MMSP)*, 1999.

[46] J. Yang, Z. Fu, T. Tan, and W. Hu. Skin Color Detection Using Multiple Cues. In *Proceedings of the 17th International Conference on Pattern Recognition (ICPR)*, 2004.

[47] T. Yang. Applications of Computational Verbs to Effective and Realtime Image Understanding. *International Journal of Computational Cognition*, 4(1), 2006.

[48] Q.-F. Zheng, W. Zeng, G. Wen, and W.-Q. Wang. Shape-Based Adult Image Detection. In *Proceedings of the 3rd International Conference on Image and Graphics (ICIG)*, 2004.