The Mask-SIFT Cascading Classifier for Pornography Detection

XXXXXXX
XXXXXXXX
XXXXXXX
XXXXXXX

***Abstract*—Pornography detection using the Scale Invariant Feature Transform (SIFT) has been shown effective in identifying pornographic images. By including automated Gaussian skin masking for feature isolation, classifier performance is significantly improved. Similarly, utilizing a cascading classifier that pre-filters images based on size and skin percentage further improves precision and recall with a substantial increase in classification speed.**

***Index Terms*—Pornography detection, feature recognition, skin detection**

I. Introduction

The accurate detection of pornography is of interest in both academic and commercial settings. Applications such as Internet content filters rely on accurate identification of pornographic images in near-real time. With the advent of high-bandwidth Internet connections and multi-terabyte home storage arrays, forensic analysis of in-situ drive images has surpassed the ability of an analyst to manually review all image content in a reasonable timeframe. Additionally, the transition from static images to movie content has increased the problem space further.

Classic techniques to classify images as pornographic rely on skin detection [1], shape detection[2], or feature extraction[3]. Multiple enhancements have been made to each of the techniques individually, and at least one major research endeavor used a cascading classifier [3].

Mask-SIFT makes use of a modified scale invariant feature transformation (SIFT) classifier [4] in a cascading classification system. The classification tool improves upon existing classification systems by using skin tone, shape, and image metadata to determine if an image is pornographic. The Mask-SIFT is tested against a corpus of real-world images and shown to outperform the bag-of-visual-words algorithm (using SIFT) [5] and an optimized skin-based classification algorithm. Additionally, the implementation in a cascading classifier further improves recall and precision while increasing processing throughput on images.

For the purposes of this paper, pornography is defined as any image that contains nudity, including male or female genitalia, female breasts, and/or male or female buttocks.

II. Contributions

Our preliminary research into a Mask-SIFT cascading classifier provides several contributions to the field of pornography detection.

- First, a new variant of SIFT is created which automatically pre-filters pornographic images to isolate features of interest.
- Second, a cascading classifier is developed which outperforms existing classifiers.
- Finally, a more realistic dataset is created for testing pornographic image detection approaches.

III. Prior Art

The simplest pornography detectors use basic skin classification. Skin detectors have been used across all color spaces, and rely on the concept that pornographic images have more skin-centric pixels than non-skin-centric [6]. Results of basic skin detection are effective at eliminating non-pornographic images (by definition, an image with no skin cannot be pornographic), but not effective discriminators amongst images with high skin-colored pixel densities.

Shape and feature detection have shown promise in incorporating areas of skin pixels with classifiers that attempt to discern orientation, contour, and makeup of skin-based areas. Wavelet-based approaches provided strong early results, with current iterations focusing on reducing error rates [7]. Similarly, feature-based techniques have used automated learning classifiers to identify image features specific to pornographic images, with the application of content based image retrieval methods [8]. While these approaches produce reasonable results, they have limitations in that they either require manual identification of relevant features a priori, or they detect features in an automated fashion which may incorporate features correlated with, but not always representative of, pornographic images.

Three enhanced classifiers using SIFT to detect poronography have been showing promising results. Chen et al use a Bayesian network to improve on the bag-of-visual-words approach [5], though without isolating skin features before the classification. Lopez et al make use of color information present in the image to enhance SIFT[6]. Mask-SIFT improves on general color-based SIFT approaches by using only those colors that are relevant, and by isolating only person-related features. Liu et al present the closest in concept, using SIFT to detect the presence of people then skin percentage to rate the image as pornographic[7] or not. Unlike Mask-SIFT, their use of a cascading classifier places the slower discriminator (SIFT) first, and does not make use of any image metadata.

One problem of all of the testing on the above classifiers is the use of a dataset skewed toward non-pornographic images, mimicing a web filtering scenario. Most of the research datasets use a ratio of 5:1 or 10:1 of non-pornographic images to pornographic images. It is easier to detect the edge cases of non-pornographic images, resulting in a higher true negative rate (and accordinging skewed ROC results). Mask-SIFT is tested with an equal number of pornographic and non-pornographic images, representing a forensic scenario more likely to occur in the analysis of a hard drive of an individual extensively viewing or collecting pornography, and providng a more realistic ROC curve.

IV. Methodology

SIFT is recognized as one of the most effective general pattern recognition algorithms, but it is most effective when the individual elements to be identified can be easily isolated from a background image. Our implementation of SIFT, Mask-SIFT, uses a pre-filter to automatically remove all non-skin pixels from an image for training purposes. Specifically, a Gaussian projection of likely skin pixels in the RGB colorspace is quantized into a lookup table (for speed). Each pixel in an image is classified as skin or non-skin based on a naïve Bayesian classifier. The image is then post-processed using a median filter to fill-in missing pixels and remove salt-and-pepper noise in the image.

Once the masked image is created, SIFT is used to generate features from the person portions of the image. By masking out the skin-based features, other features common to pornographic images which may occur in innocuous image (e.g. a bed) are not considered. Finally, a set of representative features is generated by grouping the features into visual words common to the training corpus (pornographic or non-pornographic), using the approach outlined in [8]. The Mask-SIFT normalized aggregate histograms are then generated for each of the pornographic and non-pornographic images.

The cascading classifier itself has three stages, as shown in Figure 1. First, a size-based classifier is used to eliminate images that could not be classified as pornographic. Specifically, any image with either a height or a width of less than 32 pixels is classified as non-pornograpic (in the training dataset, no identifiable pornographic images has less than 50 pixels, but 32 pixels was used to provide a buffer). Discrete dimension discrimination was used instead of total pixels to eliminate high width, low height separators common on web downloads. The size discriminator relies solely on image metadata, and is O(c).

The second classification stage uses the number of skin pixels present in the image. The same algorithm used to mask images in the training stage is used to identify skin pixels. Based on the training dataset, any images with fewer than 33% skin pixels could be safely classified as non-pornographic images (Recall: .99), with a complexity of O(n), where n is the number of image pixels.

The third stage calculates the SIFT features of the unmasked image, then compares them to the histograms from the training set. The class of the histograms of the five nearest images from the training corpus (measured by Euclidian distance) are identified and a simple voting algorithm is used to identify the test image as pornographic or non-pornographic.

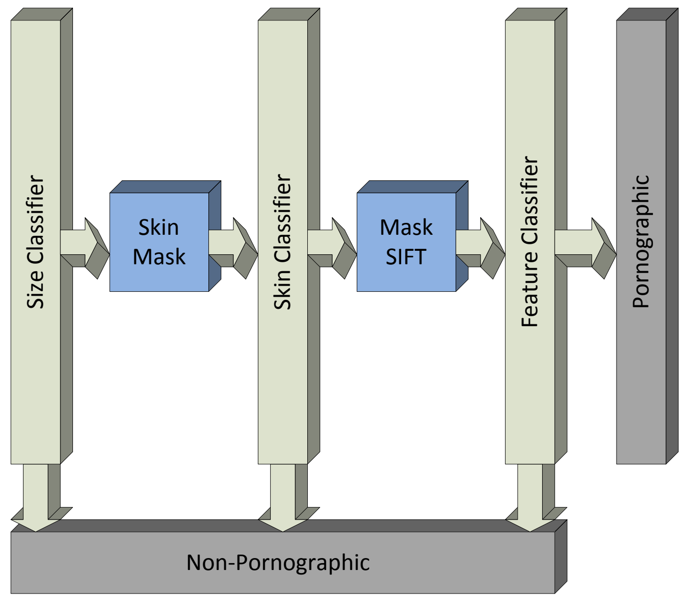

Figure 1 – Cascading classification model

V. Experimental Procedures

A. Data Set

There are no standardized datasets for pornography detection, primarily due to copyright issues and the potential legal limitations on distributions of large quantities of pornographic material. As such, a representative dataset of internet images, both pornographic and non-pornographic, was created for this experiment.

A bulk downloader was used to collect the source images returned from Google images using both pornographic (e.g. “xxx”) and non-pornographic (e.g. “picture”) terms. The resultant images were manually reviewed for proper classification. A total of 1500 pornographic and 1500 non-pornographic images were collected, and randomly split into a training set (1000 images from each class) and a test set (the remaining 500 images from each class). A larger training set than test was used to provide a more representative set of features for classification.

B. Method 1: Skin Detection

The first method used a basic skin detector to classify each image in the test set as pornographic or non-pornographic. The skin classifier noted above was used to classify pixels as skin or non-skin, then the classifier was run through the possible classifications given a linear separation of classes based on the aggregate number of skin pixels.

C. Method 2: Bag-of-Visual-Words (SIFT)

For the second method, the bag-of-visual-words approach based on SIFT and detailed in [8] was used to classify the training images as pornographic or non-pornographic. The same training dataset was used, and the same simple voting algorithm used in the Mask-SIFT implementation was performed to classify the images in the test corpus.

D. Method 3: Mask-SIFT

A single-stage classifier based solely on the Mask-SIFT algorithm proposed above was used to develop a corpus of pornographic and non-pornographic histograms from the training dataset. The simple voting algorithm above was performed on the test dataset to classify each image as pornographic or non-pornographic.

E. Method 4: Cascading Classifier

The full cascading classifier implementation detailed in Figure 1 above was performed. Each image in the test set was run through the classification process, with a subset of the total images passing along to the next stage. Images eliminated as each stage were classified as non-pornographic, and those making it through to the Mask-SIFT stage used the technique in method 3 above for final classification.

VI. Results

The ROC curve for the results of each method are shown in Figure 2 below. The bag-of-visual words approach outperformed a basic skin classifier, as per past research results. Overall, masking improved performance over bag-of-visual-words, and overall the cascading classifier approach performed the best. Specific comparisons at false positive rates of .1 of .2 are shown in Table 1 below, with the cascading classifier having .81 and .87 true positive rates, respectively.

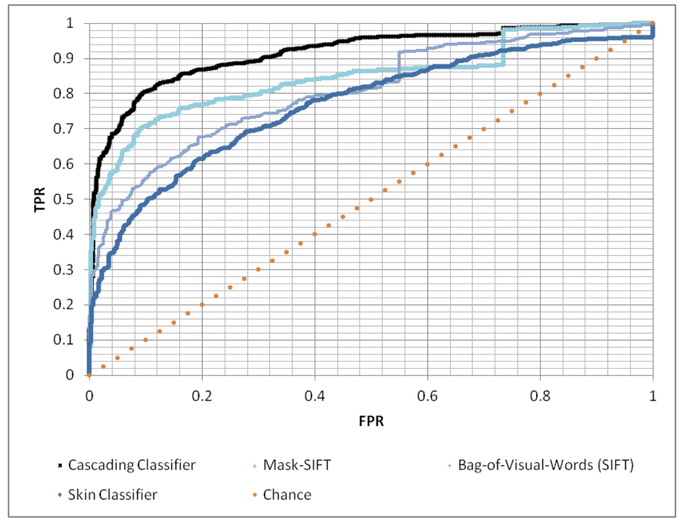

Figure 2 – Cascading classifier results

	FPR = 0.1	FPR = 0.2
Cascading Classifier	*0.81*	*0.87*
Mask-SIFT	0.71	0.77
Bag-of-Visual-Words	0.58	0.66
Skin Detection	0.49	0.61

Table 1 – Classifier true positive rates

In terms of average processing time per image, skin detection performed the best. Because Mask-SIFT and bag-of-visual-words used essentially the same algorithm with different comparison histograms (the extra time with Mask-SIFT is in the training), there was no difference in performance. The performance of the cascading classifier was second-best. The initial pass (removing images that did not meet size requirements) was extremely rapid, and the skin detection was equivalent to the top performing algorithm. While the Mask-SIFT analysis was slow, only a subset of images required the extra processing. Most importantly for forensic review, a large number of images can be removed quickly from potential candidacy of pornography, leading to a potential for an iterative display of images found and rapid ranking of potential pornographic images that can be refined as time constraints permit. The detailed performance per-image is shown in Table 2 below.

	Avg. Normalized Time (per image)
Skin Detection	0.01
Cascading Classifier	0.602
Mask-SIFT	1
Bag-of-Visual-Words	1

Table 1 – Classifier performance

The images misidentified as pornographic had features consistent with pornographic images, but could quickly be reviewed and eliminated by a human analyst. The highestranking innocuous images misclassified as pornographic are shown in Figure 3 below.

Figure 3 – Misclassified Images

VII. Future Work

This research represents the first iteration of a new classifier for detecting pornography on hard drives. Mask-SIFT compared favorably to previous techniques that have been shown to be effective on a common dataset. Future research would compare Mask-SIFT to additional algorithms that take into account color or image features on the same dataset.

The Mask-SIFT cascading classifier can be implemented in an iterative-display fashion, where all of the images on a drive are identified and an analyst is presented with those most likely to be pornographic in ranked order as thumbnails. As the classifier refines the ranking, the likelihood of pornographic images being displayed first will increase over time. This allows for a rapid determination of the likelihood that pornographic content is present on the drive, and a mathematical probability based on the number of images viewed in relation to the number present on the drive could be calculated.

In addition to an iterative display, Mask-SIFT can be improved by clustering images, then performing pornography detection on a representative image from each cluster. This would eliminate the need to review duplicate and near-duplicate images.

Finally, for the detection of child pornography, age detection algorithms can be run on the images classified as pornographic to assist law enforcement in identifying potential contraband present on a given hard drive.

VIII. Conclusions

The problem of pornography detection has been approached from many angles, incorporating skin detection, shape detection, and feature extraction. Previous algorithms have been shown to be effective, but the results have been skewed by the use of testing corpuses heavily biased toward non-pornographic images. An unbiased corpus representative of web-based images was created to engender fair comparison amongst pornography detection algorithms.

The use of Mask-SIFT in a cascading classifier highlights the effectiveness of automated feature selection as a precursor for SIFT feature extraction, with the added benefit of incorporating fast, simple classifiers with strong discriminators in early classification stages. This initial implementation is a work-in-progress, but shows strong performance compared to a leading feature-based detector and an optimized skin detector, beating both in terms of precision and recall.

References

[1] M J Jones and J M Rehg, "Statistical color models with application to skin detection," *International Journal of Computer Vision*, vol. 1, p. 2 vol. (xxiii+637+663), 1999.

[2] D A Forsyth and M M Fleck, "Automatic detection of human nudes," *International Journal of Computer Vision*, vol. 32, pp. 63--77, 1999.

[3] James Z Wang, Jia Li, Gio Wiederhold, and Oscar Firschein, "System for Screening Objectionable Images," *Computer Communications Journal*, vol. 21, pp. 1355--1360, 1998.

[4] David G Lowe, "Distinctive Image Features from Scale-Invariant Keypoints," *International Journal of Computer Vision*, 2003.

[5] T Deselaers, L Pimenidis, and H Ney, "Bag-of-visual-words models for adult image classification and filtering," in *Proceedings of the International Conference on Pattern Recognition*, 2008, pp. 1 -4.

[6] Vladimir Vezhnevets, Vassili Sazonov, and Alla Andreeva, "A Survey on Pixel-Based Skin Color Detection Techniques," in *IN PROC. GRAPHICON-2003* , 2003, pp. 85--92.

[7] H M Sun, "Pornographic image screening by integrating recognition module and image black-list/ white-list subsystem," *IET-IPR*, vol. 4, pp. 103-113, April 2010.

[8] Liu Yizhi, Lin Shouxun, Tang Sheng, and Zhang Yongdong, "Adult Image Detection Combining BoVW Based on Region of Interest and Color Moments," *IFIP Advances in Information and Communication Technology*, vol. 340, pp. 316-325, 2010.

[9] Siqi Chen, "Pornographic Pictures Detection Based on SIFT Algorithm," in *International Symposium on Information Engineering and Electronic Commerce, 3rd (IEEC 2011)*, Three Park Avenue New York, NY 10016-5990, Jan. 2011, pp. 425--428.

[10] Ana P B Lopes et al., "A bag-of-features approach based on hue-SIFT descriptor for nude detection," in *Eusipco*, 2009, pp. 1559-1556.

[11] Bei-bei Liu, Jing-yong Su, Zhe-ming Lu, and Zhen Li, "Pornographic Images Detection Based on CBIR and Skin Analysis," *Semantics, Knowledge and Grid, International Conference on*, vol. 0, pp. 487-488, 2008.