System for Screening Objectionable Images

James Ze Wang [1], Jia Li, Gio Wiederhold, and Oscar Firschein

Stanford University, Stanford, CA 94305, USA

Abstract

As computers and Internet become more and more available to families, access of objectionable graphics by children is increasingly a problem that many parents are concerned about. This paper describes WIPE$_{TM}$ (Wavelet Image Pornography Elimination), a system capable of classifying an image as objectionable or benign. The algorithm uses a combination of an icon filter, a graph-photo detector, a color histogram filter, a texture filter, and a wavelet-based shape matching algorithm to provide robust screening of on-line objectionable images. Semantically-meaningful feature vector matching is carried out so that comparisons between a given on-line image and images in a pre-marked training data set can be performed efficiently and effectively. The system is practical for real-world applications, processing queries at the speed of less than 2 seconds each, including the time to compute the feature vector for the query, on a Pentium Pro PC. Besides its exceptional speed, it has demonstrated 96% sensitivity over a test set of 1,076 digital photographs found on objectionable news groups. It wrongly classified 9% of a set of 10,809 benign photographs obtained from various sources. The specificity in real-world applications is expected to be much higher because benign on-line graphs can be filtered out with our graph-photo detector with 100% sensitivity and nearly 100% specificity, and surrounding text can be used to assist the classification process.

Key words: Internet, Pornography, Content-based Image Retrieval, Human Recognition

1 Introduction

With the rapid expansion of the Internet, every day large numbers of adults and children use the Internet for searching and browsing through different

[1] Partially supported by the Stanford University Libraries and Academic Information Resources. Correspondence to: wangz@cs.stanford.edu

multimedia documents and databases. Convenience in accessing a wide range of information is making the Internet and the World-Wide Web part of the everyday life of ordinary people. Because there is freedom of speech, people are allowed to publish various types of material or conduct different types of business on the Internet. However, due to this policy, there is currently a large amount of domestic and foreign objectionable images and video sequences available for free download on the World-Wide Web and usenet newsgroups. Access of objectionable graphic images by under-aged "netters" is a problem that many parents are becoming concerned about.

1.1 Related Work in Industry

There are many attempts to solve the problem of objectionable images in the software industry. Pornography-free web sites such as the *Yahoo! Web Guides for Kids* have been set up to protect those children too young to know how to use the web browser to get to other sites. However, it is difficult to control access to other internet sites.

Software programs such as *NetNanny*, *Cyber Patrol*, or *CyberSitter* are available for parents to prevent their children from accessing objectionable documents. However, the algorithms used in this software do not check the image contents. Some software stores more than 10,000 IP addresses and blocks access to objectionable sites by matching the site addresses, some focus on blocking websites based on text, and some software blocks all unsupervised image access. There are problems with all of the approaches. The Internet is so dynamic that more and more new sites and pages are added to it every day. Maintaining lists of sites manually is not sufficiently responsive. Textual matching has problems as well. Sites that most of us would find benign, such as the sites about breast cancer, are blocked by text-based algorithms, while many objectionable sites with text incorporated in elaborate images are not blocked. Eliminating all images is not a solution since the Internet will not be useful to children if we do not allow them to view images.

1.2 Related Work in Academia

Academic researchers are actively investigating alternative algorithms to screen and block objectionable media. Many recent developments in shape detection, object representation and recognition, people recognition, face recognition, and content-based image and video database retrieval are being considered by researchers for use in this problem.

To make such algorithms practical for our purposes, extremely high sensitivity (or recall of objectionable images) with reasonably high speed and high specificity is necessary. In this application, *sensitivity* is defined as the ratio of the number of objectionable images identified to the total number of objectionable images downloaded; *specificity* is defined as the ratio of the number of benign images passed to the total number of benign images downloaded. A perfect system would identify all objectionable images and not mislabel any benign images, and would therefore have a sensitivity and specificity of 1. The "gold standard" definition of objectionable and benign images is a complicated social problem and there is no objective answer. In our experiments, we use human judgment to serve as a gold standard.

For real-world application needs, a high sensitivity is desirable, i.e., the correct identification of almost every objectionable image even though this may result in some benign images being mislabeled. Parents might be upset if their children are exposed to even a few objectionable images.

The following properties of objectionable images found on the Internet make the problem extremely difficult:

- mostly contain non-uniform image background;
- foreground may contain textual noise such as phone numbers, URLs, etc;
- content may range from grey-scale to 24-bit color;
- some images may be of very low quality (sharpness);
- views are taken from a variety of camera positions;
- may be an indexing image containing many small icons;
- may contain more than one person;
- persons in the picture may have different skin colors;
- may contain both people and animals;
- may contain only some parts of a person;
- persons in the picture may be partially dressed.

Forsyth's research group [7,8] has designed and implemented an algorithm to screen images of naked people. Their algorithms involve a skin filter and a human figure grouper. As indicated in [7], 52.2% sensitivity and 96.6% specificity have been obtained for a test set of 138 images with naked people and 1401 assorted benign images. However, it takes about 6 minutes on a workstation for the figure grouper in their algorithm to process a suspect image passed by the skin filter.

1.3 Overview of Our Work

Our approach is different from previous approaches. Instead of carrying out a detailed analysis of an image, we match it against a small number of feature vectors obtained from a training database of 500 objectionable images and

8,000 benign images, after passing the images through a series of fast filters. If the image is close in content to a threshold number of pornographic images, e.g., matching two or more of the marked objectionable images in the training database within the closest 15 matches, it is considered objectionable. To accomplish this, we attempt to effectively code images based on image content and match the query with statistical information on the feature indexes of the training database. The foundation of this approach is the content-based feature vector indexing and matching developed in our multimedia database research. Image feature vector indexing has been developed and implemented in several multimedia database systems such as the IBM QBIC System [5] developed at the IBM Almaden Research Center. Readers are referred to [9,13,17,18,21] for details on this subject. However, for WIPE we use quite specialized features.

In the WIPE project, we developed a new algorithm to efficiently and accurately compare the semantic content of images mainly consisting of objects such as the human body. Using moment analysis, texture analysis, histogram analysis and statistics, the algorithm produces feature vectors that provide excellent accuracy in matching images of relatively isolated objects such as the human body. We use a novel multi-step metric to screen objectionable images. A training database of about 500 objectionable images and about 8,000 benign images has been indexed using such an algorithm. When a query comes in, we first process the image using a series of fast filters. If the image passes all the filters, we compute the feature vector and use it to match with the training database. If it matches with objectionable images in the training database, we classify it as an objectionable image. Otherwise, we classify it as a benign image. Promising results have been obtained in experiments using a test set of 1,076 objectionable images and 10,809 benign images.

2 Related Background

In this section, we review the basic background for wavelet analysis and moment analysis used in our algorithm.

2.1 Daubechies' Wavelets

Theoretical details on wavelet analysis, wavelet basis and Daubechies' wavelets can be found in [6,15,3,4,14,21]. Daubechies' wavelets give remarkable results in image analysis and synthesis due to the efficiency of the transformation coefficients in representing images.

Daubechies' wavelet transform separates the image into clean distinct low frequency and high frequency parts. Although some other wavelets may be better for certain applications [16,20], various experiments [23] and studies have shown that Daubechies' wavelets are highly suitable for dealing with general-purpose images.

For the problem of abstracting image content based on shape, we want to represent the object shape in the image as exactly as possible by the coefficients of the feature vector. When using the simpler Haar wavelet [15,19], the extra noise we obtain in the high-pass bands makes the matching very difficult to perform. Traditional edge detection algorithms have the same problem. Daubechies' wavelets offer a multiresolution analysis, which has the potential for high speed algorithm design, and a wide range of flexibility. For example, we may select the appropriate wavelet basis to obtain the exact amount of fluctuation we desire in the high-frequency bands to represent the object shape.

2.2 Moments

Moments are descriptors widely used in shape and region coding [10] because a moment-based measure of shape can be derived that is invariant to translation, rotation, and scale. For a 2-D continuous surface $f(x,y)$ embedded on the xy-plane, the *moment of order* $(p+q)$ is defined as

$$m_{pq} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x^p y^q f(x,y)dxdy \tag{1}$$

for $p,q \in \mathbb{N}\bigcup\{0\}$. The theory of moments has shown that the moment sequence $\{m_{pq}\}$ is uniquely determined by $f(x,y)$ and vice versa.

The *central moment* is defined as

$$\mu_{pq} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \left(x - \frac{m_{10}}{m_{00}}\right)^p \left(y - \frac{m_{01}}{m_{00}}\right)^q f(x,y)dxdy. \tag{2}$$

For discrete cases such as a digitized image, we define the *central moment* as

$$\mu_{pq} = \sum_x \sum_y \left(x - \frac{m_{10}}{m_{00}}\right)^p \left(y - \frac{m_{01}}{m_{00}}\right)^q f(x,y). \tag{3}$$

Then the *normalized central moments* are defined as

$$\eta_{pq} = \frac{\mu_{pq}}{\mu_{00}^{\gamma}} \quad where \quad \gamma = \frac{p+q+2}{2} \tag{4}$$

for $p + q = 2, 3, 4, ...$

A set of seven *translation, rotation, and scale invariant moments* can be derived from the 2nd and 3rd moments. A detailed introduction to these moments can be found in [10,11]. These moments can be used to match two objectionable images containing people having the same posture but taken from different camera angles.

3 Screening Algorithm in WIPE

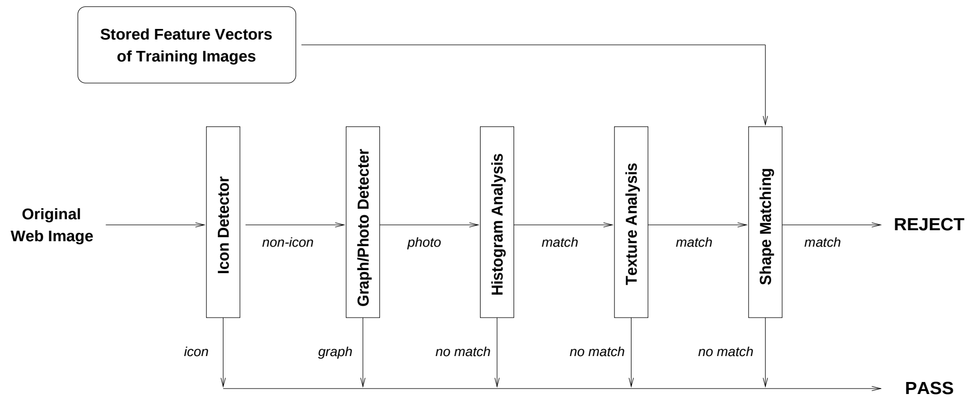

Fig. 1. **Basic structure of the algorithm in WIPE.**

3.1 Overview

We have developed a new shape-based indexing scheme using forward and backward Daubechies' wavelet transforms, variant and invariant normalized central moments, and color histogram analysis that is able to capture the object posture. The screening algorithm uses several major steps, as shown in Figure 1. The layout of these filters is a result of a cost-effectiveness analysis. Faster filters are placed earlier in the pipeline so that benign images can be quickly passed.

Our design has several immediate advantages.

(1) It does not rely too much on color when detecting sharp edges. That means that naked people of different races can be detected without bias. It also has the potential for shape-based matching of benign images. Image background does not affect the querying results unless the background has sharp features. Also, the query image can be of different color quality.
(2) We used multi-resolution analysis rather than a traditional edge detector to capture the shape information in the images. This reduces the dependence on the quality or the sharpness of the images.

(3) We used a combination of variant and invariant normalized central moments to make the querying independent of the camera position.

3.2 Icon Filter and Image Normalization

We first apply an icon filter to the image downloaded from the Internet. The current implementation of the icon filter is rather simple. If the length of any side of the image is small, we consider the image an icon image, and hence benign.

Many color image formats are currently in use, e.g., GIF, JPEG, PPM and TIFF are the most widely used formats on the Internet. Because images can have different formats and different sizes, we must first normalize the data for histogram computation. For the wavelet computation and moment analysis parts of our algorithm, any image size is acceptable. To save computation time, we rescale images using bi-linear interpolation so that the length of the longest side is 256 pixels. Red-Green-Blue (i.e., RGB) color space is used for histogram computation.

3.3 Graph/Photo Classification

We apply a classifier to decide whether an image is a photograph, i.e., a continuous-tone image, or, a graph image, i.e., a image containing mainly text, graph and overlays. If the image is a graph image, it is very likely that the image is a benign image map commonly used on the web pages.

The classifier breaks an image into blocks and segments every block into either of the two classes. If the percent of blocks classified as a photograph is higher than a threshold, the image is marked as photograph; otherwise, it is marked as text. The algorithm we used to segment image blocks is based on a probability density analysis for wavelet coefficients in high frequency bands. For every block, two feature values, which describes the distribution pattern of the wavelet coefficients in high frequency bands, are evaluated. Then the block is marked as a corresponding class according to the two feature values. Details of the algorithm are given in [12]. We do not count pure black or pure white blocks because index images usually have black or white backgrounds.

Classification as a graph causes the WIPE algorithm to exit. Misclassifying a photographic image as graph image leads to false classification of objectionable images. However, misclassifying a text image as photograph simply means that the image is sent to the next stage filter in the whole WIPE screening system. Consequently, we sacrifice the accuracy in classifying graph images to obtain very high sensitivity in classifying photographic images. In this step, we achieved 100% sensitivity for photographic images and higher than 95% specificity. This result was obtained on a database of 12,000 photographic images and a database of 300 randomly downloaded graph-based image maps from the web.

3.4 Color Histogram Analysis and Texture Analysis

Examination of color histograms revealed that objectionable images have a different color distribution than benign images [22]. We use a total of 512 bins to compute the histogram. An efficient histogram analysis was designed and implemented. We manually define a certain color range in the color spectrum as human body colors. Then we define a weight for each of these colors based on the probability of being a human body color. While this assessment was done manually, in the future version of WIPE this will be done statistically. Finally a weighted amount of human body colors that the given image contains can be obtained by summing over the entire histogram. If we set a threshold of, say, 0.15, about one half of the benign images are then classified correctly, while only a small number of skin images are classified incorrectly.

The texture analysis part of the WIPE algorithm is rather simple due to the simple texture shown by the human body in objectionable images. We statistically analyze the histogram of the high frequency bands of the wavelet transform. If the areas of human body colors contains much high frequency variations, we consider the area a non-human body area.

3.5 Edge and Shape Detection and Matching Using Wavelets and Moments

Clearly the color histogram approach alone is not sufficient. Sometimes two images may be considered very close to each other using this measure when in actuality they have completely unrelated semantics.

We apply the wavelet transform to perform multidirectional and multiscale edge detection. Readers are referred to [1] for the theoretical arguments on the effectiveness of a similar algorithm. Our purpose is not to obtain a high quality edge detection algorithm for this application. Rather, since the goal here is to effectively extract the conceptual shape information for objects and textural information for areas from the image; it is not necessary to produce a perceptually pleasant edge image. Consequently, we kept the algorithm simple to achieve a fast computation speed.

We start the edge detection process by transforming the image using the

Daubechies-3 wavelet basis. The image is decomposed into four frequency bands with corresponding names LL, HL, LH and HH. The notation is borrowed from the filtering literature [19]. The letter 'L' stands for low frequency and the letter 'H' stands for high frequency. The left upper band is called 'LL' band because it contains low frequency information in both the row and column directions. An even number of columns and rows in the querying image is required due to the downsampling process of the wavelet transform. However, if the dimensions of the image are odd, we simply delete one column or one row of pixels from the boundaries.

The LH frequency band is sensitive to the horizontal edges, the HL band is sensitive to the vertical edges, and the HH band is sensitive to the diagonal edges [4,1]. We detect the three types of edges separately and combine them at the end to construct a complete edge image. To detect the horizontal edges, we perform an inverse Daubechies-3 wavelet transform on a matrix containing only the wavelet coefficients in the LH band. Then we apply a zero-crossing detector in vertical direction to find the edges in the horizontal direction. The mechanism for using zero-crossing detector to find the edges can be found in [1]. Similar operations are applied to the HL and HH band, but different zero-crossing detectors are applied. For the HL band, we use a zero-crossing detector in the horizontal direction to find vertical edges and for the HH band, we use zero-crossing detector in the diagonal direction to find diagonal edges.

After we get the three edge maps, we combine them to get the final edge image. To numerically show the combination, let us denote [2] the three edge maps by $E_1[1:m,1:n]$, $E_2[1:m,1:n]$ and $E_3[1:m,1:n]$. The image size is $m \times n$. Then the final edge image, denoted by $E[1:m,1:n]$, can be obtained from $E[i,j] = (E_1[i,j]^2 + E_2[i,j]^2 + E_3[i,j]^2)^{\frac{1}{2}}$.

Once the edge image is computed, we compute the normalized central moments up to order five and the translation, rotation, and scale invariant moments based on the gray scale edge image using the definitions in Section 2.2. A feature vector containing these $21 + 7 = 28$ moments is computed and stored for each image in the training database. When a query comes in that has passed the histogram matching step, a moment feature vector is computed and a weighted Euclidean distance is used to measure the distance of the query and an image in the training database. The weights are determined so that matching of the 21 normalized central moments has higher priority than the matching of the 7 invariant moments. In fact, many objectionable images are of similar orientation. If the query matches with objectionable images in the training database, we classify it as an objectionable image, otherwise we classify it as a benign image.

[2] Here we use MATLAB notation. That is, $A(m_1 : n_1, m_2 : n_2)$ denotes the submatrix with opposite corners $A(m_1, m_2)$ and $A(n_1, n_2)$.

4 Experimental Results

This algorithm has been implemented on a Pentium Pro 200MHz workstation. We selected about 500 objectionable images from news groups and 8,000 benign images from various sources such as the Corel Photo CD-ROM series for our training database. When we downloaded the objectionable images, we tried to eliminate those from the same source, i.e., those of extremely similar content. To compute the training feature vectors for the 8,000 color images in our database requires approximately one hour of CPU time.

We also selected 1,076 objectionable photographic images and 10,809 benign photographic images as our queries in order to test WIPE. The matching speed is very fast. It takes less than one second to process a query and select the best 100 matching images from the 8,500 image database using our similarity measure. Once the matching is done, it takes almost no extra CPU time to determine the final answer, i.e., if the query is objectionable or benign.

(a) *(b)* *(c)* *(d)* *(e)*

Fig. 2. **Typical benign images being marked mistakenly as objectionable images by WIPE.** (a) areas similar to human body (b) fine-art (c) some dog images are difficult to classify (d) partially undressed (e) hard to tell (bathing) .

(a) *(b)* *(c)* *(d)* *(e)*

Fig. 3. **Typical objectionable images being marked mistakenly as benign images by WIPE.** (a) undressed part too small (b) frame (c) hard to tell (d) image too dark and of extremely low contrast (e) dressed but objectionable . Some areas of objectionable images are blackened and blurred.

Besides the fast speed, the algorithm has achieved remarkable accuracy. It has demonstrated 96% sensitivity and 91% specificity. Figure 2 and 3 show typical images being mistakenly marked by the WIPE system. Most of the few failures for marking objectionable images happen when the query contains textual noise and/or a frame, as in Figure 3.a. We expect the specificity in real-world applications to be much higher than we reported here because there are many graph images in web pages. These images can be classified as benign images without any error. Also, we did not experiment on methods for assisting WIPE by processing surrounding text. We expect the performance to be much improved once image and textual information is combined.

5 Conclusions and Future Work

In this paper, we have demonstrated an efficient integration of various filters and a shape-based indexing and matching system using Daubechies' wavelets developed by us for practical screening of on-line objectionable images. We have obtained a performance which already appears satisfactory for practical applications.

It may be possible to improve the search accuracy by fine-tuning the algorithm, e.g., using a neural network, using a perceptually-comparable color space, adjusting weights for different matching steps, or adding more complicated preprocessing to eliminate textual noise or frames. The accuracy may also be improved by fine-tuning the training database. It is also possible to make the searching faster by developing a better algorithm for storing and matching the feature vectors. Significant speed-up is also possible if a more extensive statistical analysis can be utilized. The algorithm can also be modified to execute in parallel on multi-processor systems. Experiments with our algorithm on a video database system could be another interesting study.

References

[1] Turgut Aydin et al, Multidirectional and Multiscale Edge Detection via M-Band Wavelet Transform, *IEEE Trans. on Image Processing*, 5(9):1370-1377, 1996.

[2] D.H. Ballard, C.M. Brown, *Computer Vision*, Prentice-Hall, Inc., NJ, 1982.

[3] Ingrid Daubechies, Orthonormal bases of compactly supported wavelets, *Communications on Pure and Applied Mathematics*, 41(7):909-996, Oct. 1988.

[4] Ingrid Daubechies, *Ten Lectures on Wavelets*, CBMS-NSF Regional Conference Series in Applied Mathematics, 1992.

[5] C. Faloutsos et al, Efficient and Effective Querying by Image Content, *J. of Intelligent Information Systems*, 3:231-262, 1994.

[6] G. B. Folland, *Fourier Analysis and Its Applications*, Pacific Grove, Calif., 1992.

[7] Margaret Fleck, David A. Forsyth, Chris Bregler, Finding Naked People, *Proc. 4'th European Conf on Computer Vision*, UK, 2:593-602, 1996.

[8] David A. Forsyth et al, Finding Pictures of Objects in Large Collections of Images, *Proc. Int'l Workshop on Object Recognition*, Cambridge, 1996.

[9] Amarnath Gupta and Ramesh Jain, *Visual Information Retrieval*, Comm. of the ACM, 40(5):69-79, 1997.

[10] Rafael C. Gonzalez, Richard E. Woods, *Digital Image Processing*, Addison-Wesley Publishing Co., 1993.

[11] M. K. Hu, Visual Pattern Recognition by Moment Invariants, *IRE (IEEE) Trans. Info. Theory*, IT(8):179-187, 1962.

[12] Jia Li and Robert M. Gray, Text and Picture Segmentation by the Distribution Analysis of Wavelet Coefficients, submitted to ICIP, 1998.

[13] C. E. Jacobs et al., Fast Multiresolution Image Querying, *Proc. of SIGGAPH 95 Computer Graphics*, pp.277-286, August 1995.

[14] Gerald Kaiser, *A Friendly Guide to Wavelets*, Birkhauser, Boston, 1994.

[15] Yves Meyer, *Wavelets: Algorithms & Applications*, SIAM, Philadelphia, 1993.

[16] J. E. Odegard and C. S. Burrus, Smooth biorthogonal wavelets for applications in image compression, *Proc. of DSP Workshop*, Loen, Norway, September 1996.

[17] M. Ortega, Y. Rui, K. Chakrabarti, S. Mehrotra, T. Huang, Supporting Similarityh Queries in MARS, *ACM Multimedia '97*, Seattle, Nov 1997.

[18] J. R. Smith and S.-F. Chang, VisualSEEk: A Fully Automated Content-Based Image Query System, *ACM Multimedia Conference*, Boston, Nov 1996.

[19] Martin Vetterli, *Wavelets and Subband Coding*, Prentice Hall, N.J., 1995.

[20] J. Villasenor, B. Belzer, J. Liao, Wavelet Filter Evaluation for Image Compression, *IEEE Transactions on Image Processing*, 2:1053-1060, August 1995.

[21] James Ze Wang et al., Wavelet-Based Image Indexing Techniques with Partial Sketch Retrieval Capability, *Proc. 4th ADL Forum (ADL'97)*, Washington D.C., May 1997.

[22] James Ze Wang et al., System for Screening Objectionable Images Using Daubechies' Wavelets and Color Histograms, *Proc. IDMS'97*, Springer-Verlag LNCS 1309, Sept. 1997.

[23] Special Issue on Wavelets and Signal Processing, *IEEE Trans. Signal Processing*, Vol.41, Dec. 1993.