



A New Effective System for Filtering Pornography Images from Web Pages and PDF Files



Moheb R. Girgis¹, Tarek M. Mahmoud², Tarek Abd-El-Hafeez³

¹Professor

Department of Computer Science

Faculty of Science

Minia University, El-Minia

Egypt

moheb_girgis@eun.eg

²Associate Professor

Department of Computer Science

Faculty of Science

Minia University, El-Minia

Egypt

tarek@minia.edu.eg

³Lecturer

Department of Computer Science

Faculty of Science

Minia University, El-Minia

Egypt

tarek_1_2@yahoo.com

ABSTRACT: One of the more rapidly growing areas in search technology is image search. With this availability comes the natural need to filter offensive content, to prevent Pornography images from reaching the wrong eyes. Filtering and blocking software is one of the most frequently touted prevention devices. As any user of these services is aware, they often fail to remove offensive images. The reasons are clear in that “current Internet image search technology is based upon words, rather than image content”, as images are obtained by using the image filename or text that surrounds the image on a webpage. This paper presents an automatic software system for detecting and filtering Pornography images of any format (JPG, PNG, etc.) in web pages by using skin recognition. The proposed system is an online client-side filtering system that allows the user to choose skin detection filtering, keyword filtering, URL filtering, domain filtering, or combination of two or more of these filtering techniques. A hybrid skin color detection technique is proposed to overcome the failure of detecting complex image’s background. To the best of our knowledge, there is no tool to block pornographic images included in PDF documents. So, we augmented our proposed system with the capability of online and offline filtering of PDF documents based on skin color detection. The experimental results showed that our proposed filtering system is more effective and accurate compared to some of the established commercial filtering software.

Keywords: Skin detection, Keyword filtering, URL filtering, Domain filtering, Color spaces.

Received: Received: 18 May 2009, Revised 19 July 2009, Accepted 1 August 2009

© 2009 D-Line. All rights reserved.

1. Introduction

There is a vast amount of undesirable explicit material (Pornography) available via the Internet. Pornography images are easy to access using any public search engine such as Google or Yahoo. Controlling access to the Internet by means of filtering software has become a growth industry in every where in the world. Also, with the wide spread use of Portable Document

Format (PDF), a large number of documents are presented on the web as PDF files. These PDF files may contain Pornography images. For the best of our knowledge there is no any software tool for filtering these images from PDF files.

To block adult content, many solutions exist, such as filters that are part of image search engines like those of Google or Yahoo!. As any user of these services is aware, they often fail to remove Pornography images. The reasons are clear in that “current Internet image search technology is based upon words, rather than image content”, as images are obtained by using the image filename or text that surrounds the image on a webpage. Even though attempts have been made to search using actual image semantics, it remains a largely unsolved problem and thus, it is no surprise that traditional image filtering software also use the same methods as search, in that they screen for the text and web pages to do with the image, but not the image itself. Some representative companies as NetNanny and SurfWatch, operate by maintaining lists of URL’s and newsgroups and require constant manual updating.

Image content analysis based on skin detection [1] has the advantage of processing equally all the images without the need for updating, so will produce more effective filtering. By taking advantage of the fact that there is a strong correlation between images with large patches of skin and adult images we have to develop a skin detector. Skin color offers an effective and efficient way to detect the adult image content.

In this paper, we present an effective and robust automatic software system for detecting and filtering Pornography images in web pages and PDF files. The proposed system is a client-side filtering system and allows the user to choose keyword filtering, URL filtering, domain filtering, skin detection filtering, or combination of two or more of these filtering techniques. Our System is a Browser Helper Object, which runs in the background whenever the Internet Explorer is run. It reads the web page as HTML, or the PDF document, and extracts all images and links (URLs). Then, the system analyzes Text, URL, Domain, and all images. The output is the web page/PDF document without all images if the site contains Pornography images. Otherwise, the web page/PDF document is displayed as it is.

This paper is organized as follows: Section 2 presents skin detection techniques. The description of the proposed filtering system is given in section 3. Section 4 presents PDF filtering. Experimental results followed by conclusions are presented in sections 5 and 6, respectively.

2. Skin Detection Techniques

The simplest methods in skin detection define or assume skin color to have a certain range of values in some coordinates of a color space. A good skin color model must have a high detection rate and a low false positive rate. That is, it must detect most skin pixels while minimizing the amount of non-skin pixels classified as skin. In general, the fairness or darkness of skin depends on the amount of melanin in the skin. Due to the emerging importance of skin detection in computer vision several studies have been made on the behavior of skin chromaticity at different color spaces such as RGB Color Space, the Hue-Saturation-Value (HSV) /Intensity (HSI) /Luminance (HSL) Color Space, the Normalized RGB Color Space, YUV , YIQ, ..., etc. [1], [2], [3], [11], [12], [13].

Duan et al. [4] found that the combination of YUV and YIQ color space is more robust than each other. The most people’s skin varies in the range from 20 to 90 in the term of the intensity (I). The drawback of Duan et al.’s method is that, if the image background contains pixels similar to human skin pixels (i.e. $20 \leq I \leq 90$) but don’t belong to the skin region, the method fails to detect them. Girgis et al. [5] used the saturation parameter to overcome this drawback. We implemented some algorithms for detecting and extracting skin regions from images. This step is essential for building the proposed filtering system because the proposed system depends on the image content to prevent Pornography images from displaying.

3. The Description of The Proposed Filtering System

The proposed system is a Browser Helper Object, which runs in the background whenever the Internet Explorer (IE) runs. It reads the web page as an HTML document, extracts all images and links (URLs) from it. Using the skin color detection technique, the extracted image is analyzed to classify it as Pornography or not before displaying the web page.



Figure 1. The “Filter Tool Bar” buttons

As shown in Figure 1, the Filter Tool Bar of the proposed system contains six buttons with captions “Extract Links”, “Thumbnail”, “Skin Regions”, “PDF Filtering”, “View Source”, “Image Data”, and “Options”. The purpose of the “Extract links” button is to extract all URLs in the current web page and record these links after loading the web page. The purpose of “Thumbnail” button is to display all the extracted images in the web page as a photo album. The purpose of “Skin Regions” button is to recognize the skin in every image of the current web page. We use this button to determine the skin percentages of each image of the current web page. The purpose of the “view source” button is to display the HTML source code of the current web page. The purpose of the “Image Data” button is to display the Image Data window. Finally, the purpose of the “Options” button is to display the Internet Options window.

Figure 2 gives the steps of the proposed system which runs in the background of the IE public void Run()

```
{  
  1- Start the browser.  
  2- Set the browser events.  
  3- Go to the specified home page and execute the extract images phase.  
  4- Execute the Skin detection phase.  
  5- Execute the Filter decision phase.  
  6- Start navigating to different URL.  
}
```

Figure 2. Proposed filtering system steps

These phases are described below.

3.1 Extracting Images Phase

Ilan Assayag [6] used Google patterns that accurately allow us to locate interesting image information such as, URL, width, height, and its location. The problem with these patterns is that if Google changes the format of the

Returned HTML code the parsing will fail!. Girgis et al. [5] introduced three new methods to extract images before loading the web page, which overcome this problem.

3.2 Filter Decision Phase

After extracting the relevant features of an image, filtering systems use various algorithms to classify the image as Pornography or otherwise. Since pornography is a relative term, most filtering systems have sensitivity settings to allow the user to have some degree of control over what images are classified as Pornography or not. Nevertheless, some images may be wrongly classified. Pornography images classified as non Pornography are referred to as false-negatives while non Pornography images classified as Pornography are called false-positives.

The efficiency of filtering systems can be measured in terms of the amount of false-negatives or false-positives. In classifying images, most systems use skin color models to process and characterize the images. In these systems, the color model is trained on a number of examples under varying degrees of illumination conditions [7]. In this phase, we use two methods namely “The Nudity Detection Algorithm (NDA)” and “Automatic Online Porn Detection and Tracking (AOPDT)” to detect Pornography images found in the web page based on skin color detection algorithms. The first method computes the skin color regions found in each image and the percentages of the largest regions relative to the image. The second method calculates only the skin percentage of each image.

3.2.1 The Nudity Detection Algorithm (NDA)

The Nudity Detection Algorithm [8] is based primarily on observations that in general, Pornography images contain large amounts of skin, people have different skin tones, and skin regions in Pornography images are relatively close to each other. These observations require the identification of skin and non-skin pixels in the image to be classified. The identified skin pixels are analyzed to determine which of them are connected or form continuous regions. These skin regions are further analyzed for clues of Pornography.

In general, the Nudity Detection Algorithm consists of the following steps:

1. Detect skin-colored pixels in the image.
2. Locate or form skin regions based on the detected skin pixels.
3. Analyze the skin regions for clues of Pornography.
4. Classify the image as Pornography or not.

The first step in the algorithm requires a skin color distribution model, which is the basis for classifying pixels as skin or non-skin. Most skin color distribution models use a single color space to classify pixels as skin-colored or not [9]. However, based on the researcher's experiments, the performance of a model can be improved by using two or more color spaces.

The next step in the algorithm is the analysis of detected skin pixels in terms of adjacency or continuity to identify or locate skin regions. The number of skin regions, the sizes of the three largest skin regions and the relative positions of these skin regions are important features that aid in the classification. The results show that the performance can be improved by combining skin percentage feature and region properties. We modified this technique by ignoring the small regions to speed up this technique and get more accurate results.

In summary, the Nudity Detection Algorithm works in the following manner:

1. Scan the image starting from the upper left corner to the lower right corner.
2. For each pixel, obtain the RGB component values.
3. Calculate the corresponding HSV or any color space values from the RGB values.
4. Determine if the pixel color satisfies the parameters for being skin established by the skin color distribution model.
5. Label each pixel as skin or non-skin.
6. Calculate the percentage of skin pixels relative to the size of the image.
7. Identify connected skin pixels to form skin regions.
8. Count the number of skin regions.
9. Identify pixels belonging to the three largest skin regions.
10. Calculate the percentage of the largest skin region relative to the image size.
11. Ignore the small regions with size less than 0.09 % from the total image.
12. Classify the image as follows:
 - a. If the percentage of skin pixels relative to the image size is less than 15%, the image is not nude. Otherwise, go to the next step.
 - b. If the number of skin pixels in the largest skin region is less than 35% of the total skin count, the number of skin pixels in the second largest region is less than 30% of the total skin count, and the number of skin pixels in the third largest region is less than 30% of the total skin count, the image is not nude.
 - c. If the number of skin pixels in the largest skin region is less than 45% of the total skin count, the image is not nude.
 - d. If the total skin count is less than 30% of the total number of pixels in the image, the image is not nude.
 - e. If the total skin of the largest three regions relative to the image is less than 25% and the total skin of the largest three regions relative to the total skin count of the image is less than 45%, the image is not nude.
 - f. Otherwise, the image is nude.

3.2.2 Automatic Online Porn Detection and Tracking (AOPDT)

Distinguishing between an accurate description of reasonable human body exposure and excessive exposure is a contentious issue. Therefore to be reliable, the technology should be able to isolate skin-like pixels from other background colors and provide a reliable measure of how much skin content is in the photo and how much of it should be considered as pornography. Human skin detection technologies have largely been used in image processing and compression, and one of the most promising approach involves the use of color histograms and hue-saturation modeling of the human body colors.

The system presented in [10] calculates the percentage of human skin color in each image extracted from web page and produces an assessment that indicates if it is predominantly skin (pornography) or has an acceptable level of skin color content. The percentage of skin content pixels is used as the criterion for judging if the image contains significant human skin pixels to be classified as Pornography or non-Pornography. If skin color pixels dominate the image, the image is ranked as Pornography. The algorithm of this technique can be described as follows:

1. Extract Images found in the web page.
2. Execute the Skin Detection Algorithm.
3. Determine Pornography Decision as follows:
 - If skin color pixels are up to the level of between 5% and 20% it indicates a human being is most probably in the photo.
 - If the percentage of skin pixels is between 20% and 25%, it is more than likely Pornography image.
 - If the percentage of skin pixels is more than 25%, it is a Pornography image.

3.2.3 The Filter Setting and Decision

Our System allows the user to choose one of the following filtering techniques:

- Skin detection Filtering.
- Keyword filtering.
- URL filtering.
- Domain Filtering.
- Combination of two or more of the above filtering techniques.

In our filtering system, the blacklist filtering is divided into URL and Domain filtering. The steps of first three filtering algorithms are summarized in Figures 3, 4, and 5, respectively.

If the TextAnalysis (), URL_Analysis (), and/or Domain_Analysis () method returns a non-zero number then:

1. Stop the web browse (display a window like the case of Internet disconnection), or

```

static int TextAnalysis ( )
{
    1- Read the contents of the web page.
    2- Read the BadWords from the TextDataBase.
    3- foreach (string badword in TextDataBase)
        {
            3.1 if the Document contains this badword except educational/
                medical then increase BadWordCounter.
        }
    4-return BadWordCounter.
}

```

Figure 3. The steps of Keyword filtering algorithm.

```

static int URL_Analysis(string url)
{
    1- Read the BadURL from the URLDataBase.
    2- foreach (string BadURL in URLDataBase)
        {
            2.1 if the url (argument) equals to BadURL then
                return 1.
        }
    3-return 0.
}

```

Figure 4. The steps of URL filtering algorithm

```

static int Domain_Analysis ( )
{
    1- Read the contents of the web page.
    2- Read the BadDomain from the DomainDataBase.
    3- Obtain the Domain name of the current URL.
    4- foreach (string BadDomain in DomainDataBase)
        4.1 if the Domain equals the BadDomain then return 1.
    5-return 0.
}

```

Figure 5. The steps of Domain filtering algorithm

2. Redirect the user to the Error page built by us, or
3. Hide all images from the site and add the marquee text about us. (we apply this option).

Once the System classifies the web page as Pornography, it adds this web page to the blacklist automatically, if it is not already there. Figure 6 shows the structure of the filtering system.

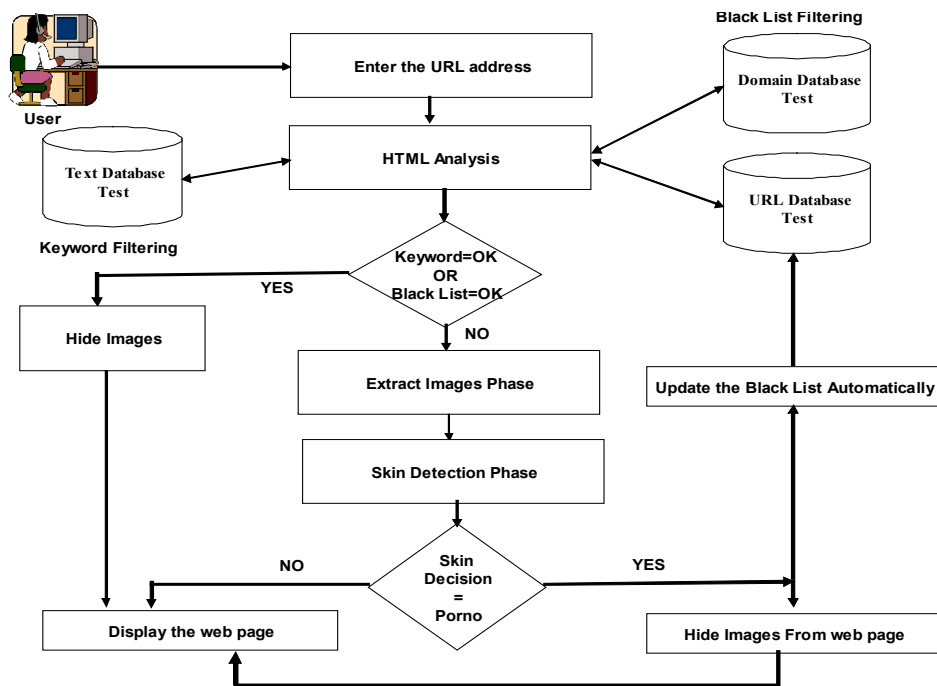


Figure 6. Filtering system structure

4. Filtering PDF Documents Based on Skin Color Detection

With the wide spread use of Portable Document Format (PDF), a large number of documents are presented on the web as PDF files. These PDF files may contain Pornography images. For the best of our knowledge there is no any software tool for filtering these images from PDF files. Our proposed system has the ability to detect and filter Pornography images that exist in PDF files. It extracts all image formats from the PDF document, and converts them to the JPEG format. Then it performs skin color detection. The system calculates the percentage of human skin color in each image and produces an assessment that indicates whether it is predominantly skin (pornography) or has an acceptable level of skin color content.

The percentage of skin content pixels is used as the criterion for judging if the image contains significant human skin pixels to be classified as Pornography or non-Pornography. If skin color pixels dominate the image, the image is ranked as Pornography and all images are hidden from displaying in the PDF document. The PDF Filtering system has the ability to analyze the PDF files offline and online.

The proposed PDF filtering system steps can be summarized as follows:

- 1- Extract the images from the PDF Document.
- 2- Detect the skin color regions in the extracted images.
- 3- Apply the offline/online filtering before displaying the PDF document.

4.1 The Offline PDF Filtering

This feature of the filtering system allows the user to browse a storage device (e.g. hard disk) and choose a PDF document to be filtered. The algorithm for filtering Pornography images from offline PDF files is summarized as follows:

- 1- Browse and open PDF document.
- 2- Extract Images found in the PDF document.
- 3- Execute the Skin Detection Algorithm.
- 4- Pornography Decision as follows:
 - If skin color pixels are up to the level of between 5% and 20% it indicates a human being is most probably in the image.
 - If the percentage of skin pixels is between 20% and 25%, it is more than likely porn image.
 - If the percentage of skin pixels is more than 25%, it is porn.

Figure 7 shows a sample PDF document which contains some Pornography images. We presented this document to our proposed system. Our system hides all images of this document as shown in Figure 8.

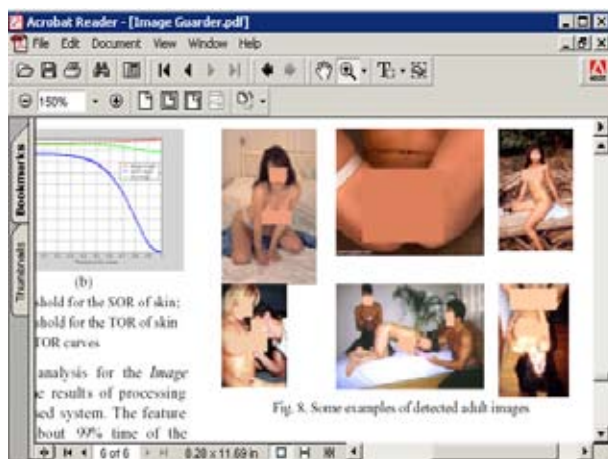


Figure 7. A sample offline PDF document contains some Pornography images

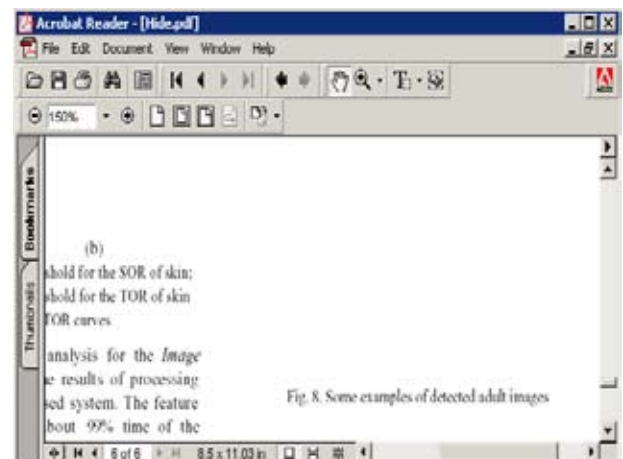


Figure 8. A filtered PDF document

4.2 The Online PDF Filtering

When the user clicks any PDF document link in a web page, the online PDF Filtering function is automatically activated at the background of the web browser and picks up the requested PDF document. Figure 9 shows the online PDF Filtering interface. This interface will be displayed on top of the web browser window after analyzing the PDF file and performing the image and skin extraction.

The Online PDF Filtering steps:

1. If the web site under test contains Pornography images, all links to images, Videos, PDF documents, ... etc are blocked.
2. If the user tries to open a PDF document in the web browser:

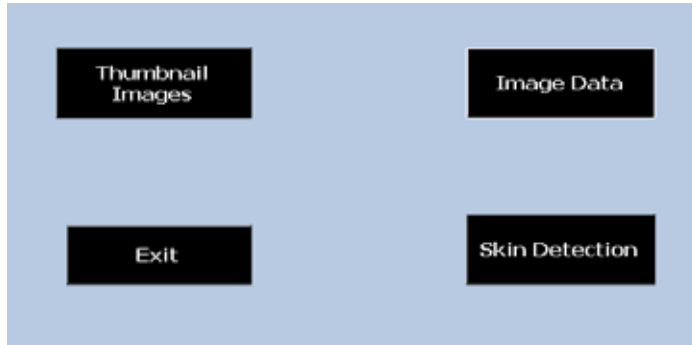


Figure 9. The online PDF Filtering interface

- The PDF Document is retrieved and saved on the hard disk with name “in.pdf”.
- The PDF Filtering function is called as described above.
- The skin detection technique is applied.
- If the PDF document contains Pornography images, the system will hide all images in the document.
- The filtered PDF document is saved as “out.pdf”.
- The file “out.pdf” is displayed in the web browser and delete “in.pdf” file if it contains Pornography images otherwise the file “in.pdf” is displayed.
- The online PDF filtering interface is displayed to allow the user to select another capability.

We included the PDF Filtering system in the Filter Tool Bar of our web filtering system and assigned the “PDF Filtering” button to call the offline PDF filtering system to filter any PDF document found anywhere on the hard disk (or any storage device).

Figure 10 shows the structure of the proposed online and offline PDF filtering system.

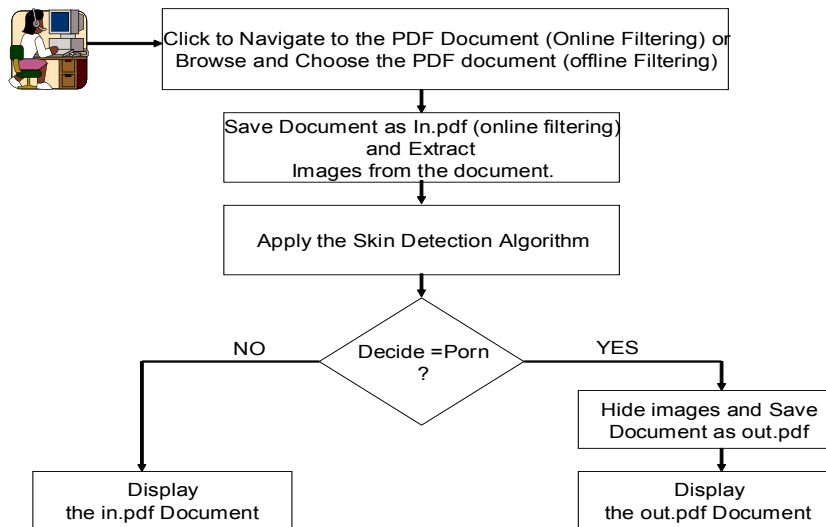


Figure 10. The structure of the proposed online and offline PDF filtering system

5. Experimental Results

This section presents the results of the experiments, which were carried out to evaluate the effectiveness of our system in filtering Pornography images. These experiments also helped us to modify and improve the system. Three types of experiments were conducted.

The aim of the first experiment was to evaluate the ability of the system to extract skin color regions from the extracted images. The second experiment was conducted to evaluate the effectiveness of the automatic filtering of Pornography images and updating the black list. The aim of the last experiment was to compare our filtering system with some other existing filtering software. The implementation of the proposed filtering system is based on C#. Net.

5.1 Skin Recognition Results

We implemented some algorithms for detecting and extracting skin regions from images. Table 1 shows sample results of the implemented techniques. The first column shows the original images, and the next 8 columns show the skin regions obtained using log opponent technique, YUV-YIQ technique, HIS technique, normalized RGB technique, YCrCb technique, RGB technique, HSV technique, and hybrid technique, respectively.

As shown in Table 1, the hybrid technique that combines the HSI, HSV, RGB, and log opponent techniques together obtains more accurate results and overcome the complex backgrounds problem. Accordingly, we employed the hybrid technique for skin color detection in our proposed filtering system. We modified the Automatic Online Porn Detection and Tracking (AOPDT) technique discussed in section 3.2.2 by applying our hybrid skin detection technique instead of the skin detection technique used by Agbinya et al [10].

5.2 The Proposed Filtering System Results

Two image classification algorithms based on skin color detection, described in sections 3.2.1 and 3.2.2, are used in the proposed filtering system to detect the Pornography images.

| Original | Log Op-ponent | YUV - YIQ | HIS | Normal RGB | YCrCb | RGB | HSV | Hybird |
|----------|---------------|-----------|-----|------------|-------|-----|-----|--------|
| | | | | | | | | |
| | | | | | | | | |
| | | | | | | | | |
| | | | | | | | | |
| | | | | | | | | |
| | | | | | | | | |
| | | | | | | | | |
| | | | | | | | | |
| | | | | | | | | |

Table 1. Sample results of the implemented techniques

Table 2 shows the results of applying the proposed filtering system on some web pages found on different servers. Each row contains the results of a web page. The first column shows the URL of the web page under test. The second column shows the frequency of bad words found in the web page (Keyword Filtering). Blacklist filtering (URL filtering, and Domain filtering) results are presented as Boolean in columns 3 and 4, where F means that the URL under test does not exist in the black list, T means that it exists in the blacklist. The total number of images found in the web page under test are listed in column 5. The results of the first method (NDA) that computes the skin color regions found in each image and the percentages of the largest regions relative to the image are presented in columns 6 and 7. The F column shows the number of images that classified Pornography/human false and the T% column shows the percentage of true image classification. The last two columns show the results of the second method (AOPDT) that calculates only the skin percentage of all images.

| URL | No. of bad words | BlackList T/F | Domain T/F | No. of Images | NDA | | AOPDT | |
|---|------------------|------------------|---------------|---------------|-----|-------|-------|-------|
| | | | | | F | T % | F | T % |
| www.sexcity.com/sexhome.htm | 220 | F | F | 68 | 2 | 97.05 | 3 | 95.58 |
| www.bravoteens.com | 36 | F | F | 175 | 7 | 96 | 8 | 95.42 |
| 00videosex.com | 160 | F | T | 57 | 3 | 94.73 | 3 | 94.47 |
| Anonib.com/_bisex | 0 | T | F | 22 | 1 | 95.45 | 1 | 95.45 |
| Handsexpics.com | 7 | F | F | 171 | 5 | 97.07 | 6 | 96.49 |
| sexZool.com | 27 | F | F | 32 | 2 | 93.75 | 2 | 93.75 |
| www.magiclevel.com/ | 201 | F | F | 170 | 4 | 97.64 | 5 | 97.05 |
| http://bestporndb.com/myhandjob/ agneshandjob/ | 0 | T | F | 34 | 1 | 97.05 | 2 | 94.11 |
| girlsex.net | 0 | F | T | 12 | 1 | 91.66 | 1 | 91.66 |
| www.clitlist.com/ | 151 | F | F | 25 | 2 | 92 | 2 | 92 |
| www.freegirlspictures.com/ | 64 | F | T | 61 | 4 | 93.44 | 5 | 91.80 |
| www.sinnerscorner.com/ | 109 | F | T | 240 | 7 | 97.08 | 8 | 96.66 |
| www.asianpic.org/ | 63 | F | T | 193 | 6 | 96.89 | 8 | 95.85 |
| 18better.com/01xxx | 0 | T | F | 8 | 1 | 87.50 | 1 | 87.50 |
| celebslip.com/blog | 57 | T | T | 3 | 0 | 100 | 1 | 100 |
| chaossex.com/adult-dvds | 13 | F | T | 134 | 7 | 94.77 | 8 | 94.02 |
| . | . | . | . | . | . | . | . | . |
| . | . | . | . | . | . | . | . | . |
| TOTAL | | | | 11113 | 485 | 95.63 | 530 | 95.23 |

Table 2. Some Servers filtering results

As can be seen in Table 2:

1. In our filtering system, we combine both blacklists and image content information to improve the classification accuracy. The classification accuracy of some servers results is 95.63 ± 0.5 %.
2. The results of the proposed filtering system to filter the Pornography images using the NDA method is more accurate than the AOPDT method.

5.3 Comparing Our Filtering System with Other Commercial Filters

Filtering software companies divide their control lists into multiple categories for which they have created unique definitions. SurfControl uses 40 such categories, N2H2 uses 35 categories (and seven “exception” categories), Websense uses 30 categories, and Secure Computing uses 30 categories. In this section, we compare our filtering system with four filtering software companies which are: CYBERSitter, CyberPatrol, NetNanny, and SurfWatch, to evaluate its effectiveness.

We tested 124 web pages in different categories and the comparison results are shown in Tables 3, 4, and 5. Table 6 summarizes the efficiency percentages of the proposed filtering system compared to other filtering products in web sites classification. The second row contains the total number of the erroneously classified web sites for each product. The last row contains the success percentage of each product and the proposed filtering system.

| # | URL | CYBERSitter | CyberPatrol | NetNanny | SurfWatch | Our System |
|---|---|----------------------|-------------------------------|--------------------------------|----------------------|--|
| | http://www.xtacy.com/ | language = 4 blocked | nudity = 4 not blocked | sex = 4 not blocked | violence = 0 blocked | Text=0 Blacklist=1 Skin=1 blocked |
| | http://www.trailerparktrash.com/ | language = 4 blocked | nudity = 4 blocked | sex = 4 _not blocked | violence = 0 blocked | Text=0 Blacklist=1 Skin=0 blocked |
| | http://www.sexis.com/russia/info.htm | language = 0 blocked | nudity = 3 not blocked | sex = 0 not blocked | violence = 0 blocked | Text=0 Blacklist=1 Skin=1 blocked |

Table 3. Sex Search Results

| # | URL | CYBERSitter | CyberPatrol | NetNanny | SurfWatch | Our System |
|---|---|--------------------------|---------------------------|-------------------------------|------------------------------------|--|
| | http://www.citylights.com/ | language = 0 not blocked | nudity = 0 not blocked | sex = 0 not blocked | violence = 0 not blocked | Text=0 Blacklist=1 Skin=0 not blocked |
| | http://www.sexualhealth.com/ | language = 4 blocked | nudity = 3 blocked | sex = 1 not blocked | violence = 0 not blocked | Text=1 Blacklist=1 Skin=0 blocked |

Table 4. COPA litigants

| # | URL | CYBERSitter | CyberPatrol | NetNanny | SurfWatch | Our System |
|---|---|-----------------------------|---------------------------|------------------------|-----------------------------|--|
| | http://www.now.org/ | language = 0 blocked | nudity = 0 not blocked | sex = 0 not blocked | violence = 0 not blocked | Text=0 Blacklist=0 Skin=0 not blocked |
| | http://www.feminist.org/ | language = 0 not blocked | nudity = 0 not blocked | sex = 0 not blocked | violence = 0 not blocked | Text=0 Blacklist=0 Skin=0 not blocked |

| | | | | | |
|--------------------------------|-----------------------------|------------------------|---------------------|--------------------------|--|
| http://wlo.org/ | language = 0 not blocked | nudity = 0 not blocked | sex = 0 not blocked | violence = 0 not blocked | Text=0 Blacklist=0 Skin=0 not blocked |
| http://www.guerrillagirls.com/ | language = 0 blocked | nudity = 1 not blocked | sex = 0 not blocked | violence = 0 not blocked | Text=0 Blacklist=0 Skin=0 not blocked |
| http://www.FeministUtopia.com/ | language = 1 not blocked | nudity = 1 not blocked | sex = 0 not blocked | violence = 0 not blocked | Text=0 Blacklist=0 Skin=0 not blocked |

Table 5. Feminist Sites

| | CYBERsitter | CyberPatrol | NetNanny | SurfWatch | Our System |
|---------------------|---------------|---------------|---------------|---------------|---------------|
| Total No. of Errors | 17 | 6 | 12 | 7 | 4 |
| success % | 86.29% | 95.16% | 90.32% | 94.35% | 96.77% |

Tables 6. The efficiency comparison of our system and other filtering products

As can be seen in Table 6, it is clear that the success percentage of our system in detecting and preventing adult images from displaying is better than the other filtering products software. It should be noted that the 4 sites erroneously classified by our system are due to that the system concentrates only on adult images and related words, which did not appear in these sites.

6. Conclusions

This paper presented an online client-side filtering system that allows the user to choose skin detection filtering, keyword filtering, URL filtering, domain filtering, or combination of two or more of these filtering techniques. We implemented some established skin detection techniques for color images. A hybrid skin color detection technique is proposed to overcome the failure of detecting complex image's background. Two methods are used to detect Pornography images found in the web page based on skin color detection algorithms. To the best of our knowledge, there is no tool to block pornographic images included in the PDF documents. So, we enhanced our filtering system with the capability of an online and offline filtering of PDF documents based on skin color detection. The experimental results showed that our proposed filtering system is more effective and accurate compared to some of established commercial filtering software.

Since the detection techniques employed in the proposed system fails to correctly classify face images, further work is required to mix image content based analysis and face detection algorithms.

References

- [1] Kakumanu, P., Makrogiannis, S., Bourbakis, N. (2007). A Survey of Skin-Color Modeling and Detection Methods, *Pattern Recognition*, V. 40, pp. 1106-1122. Available at www.sciencedirect.com.
- [2] Vezhnevets, V., Sazonov, V., Andreeva, A. (2003). A survey on pixel-based skin color detection techniques. In GraphiCon, Moscow, Russia, Sept.
- [3] Chai, D., Bouzerdoum, A. (1999). A Bayesian Approach to Skin Color Classification in YCbCr Color Space. *In: Proc. Of IEEE Region Ten Conference*, V. 2, pp. 421- 4124.
- [4] Duan, L., Cui Gao, G. W., Zhang, H. (2002). Adult Image Detection Method Base-on Skin Color Model and Support Vector Machine. *In Asian Conference on computer Vision*, pp. 797-800, Melbourne, Australia.

- [5] Girgis, M. R., Mahmoud, T. M., Abd-El-Hafeez, T. (2007). An Approach to Image Extraction and Accurate Skin Detection from Web Pages, *International Journal of Computer Science And Engineering*, V.1, ISSN 1307-3699, Number.
- [6] Ilan, A. (2008). An API for Google Image Search, <http://www.codeproject.com>, last access: October.
- [7] Jones, M. J., Rehg, J. M. (1999). Statistical Color Models with Application to Skin Detection, *In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, V. 1, pp. 274–280.
- [8] Rigan, A. (2005). An Algorithm for Nudity Detection. *In: Proceedings of the 5th Philippine Computing Science Congress*, ISSN1908-1146, March 4-5.
- [9] Lin, Y., Tseng, H., Fuh, C. (2003). Pornography Detection Using Support Vector Machine. 16th IPPR Conference on Computer Vision, Graphics and Image Processing, CVGIP, Kinmen, ROC.
- [10] Agbinya, J. I., Lok, B., Wong, Y. S., Da Silva, S. (2005). Automatic Online Porn Detection and Tracking, *In: Proceedings ICT 2005 - 12th International Conference on Telecommunications*, Cape Town, South Africa, 3 – 6 May.
- [11] Jedynek, B., Zheng, H., Daoudi, M. (2003). Statistical models for skin detection, *IEEE Workshop on Statistical Analysis in Computer Vision*, in conjunction with CVPR 2003 Madison, Wisconsin, June 16–22.
- [12] Albiol, A., Torres, L., Delp, E. J. (2001). Optimum color spaces for skin detection, *In: Proceedings of the IEEE International Conference on Image processing*, pp. 122–124.
- [13] Jayaram, S., Schmugge, S., Shin, M. C., Tsap, L. V. (2004). Effect of color space transformation, the illuminance component, and color modeling on skin detection, *In: Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, V. 2, pp. II-813–II-818.