A Highly Efficient Content Based Approach to Filter Pornography Websites

Tarek M. Mahmoud, Minia University, Egypt
Tarek Abd-El-Hafeez, Minia University, Egypt
Ahmed Omar, Minia University, Egypt

ABSTRACT

With the ever-growing Web, the Websites with objectionable contents like pornography, violence, racism, etc. have been augmented rapidly during recent years. Among the offensive contents, the pornography is the most harmful one affecting children safety and causing many destructive side effects. A content filter is one or more pieces of software that work together to prevent users from viewing material found on the Internet. This paper presents an efficient content based software system for detecting and filtering pornography images in Web pages. The proposed system runs online in the background of Internet Explorer (IE) for the purpose of restricting access to pornography Web pages. Skin and face detection techniques are the main components of the proposed system. Because the proposed filter works online, the authors propose two fasting techniques that can be used to speed up the filtering system. The results obtained using the proposed system is compared with four commercial filtering programs. The success rate of the proposed filtering system is better than the considered filtering programs.

Keywords: Color Models, Content Based Filtering, Face Detection, Skin Detection, URL

1. INTRODUCTION

Since the mid-1990s, greater Internet accessibility and advances in file-sharing technology spurred a proliferation in child pornography. The possession of child pornography is extremely troubling when you consider that in each image a child is being forced to engage in a sexual act (Sylvia, 2011). Pornography uses a variety of media, ranging from books, magazines, film and video. However, currently, one of the most prevalent means of distributing pornography is the Internet, by simply typing a few keywords into a search engine (e.g., Google) it is extremely easy to search for pornography on the Internet. The number one search term used on search engine sites is “sex”. Users searched for “sex” more than other terms, such as “games, travel, music, jokes, cars, weather, health,” and “jobs” combined (Luke, 2010).

DOI: 10.4018/ijcvip.2012010105
According to ConvententEyes report (Luke, 2010) and Internet Pornography Statistics (Ropelato, 2006):

- The largest group of viewers of Internet porn is children between ages 12 and 17.
- More than 11 million teenagers view Internet pornography on a regular basis.
- One senior executive spent at least 331 days looking at pornography and chatting online with partially clad or nude women, this cost tax payers anywhere from $13,800 to $58,000.
- One worker perused hundreds of pornographic websites during work hours in a three week time frame in June 2008; that employee received a 10-day suspension.
- 12% (4.2 million websites) of the websites on the Internet are pornographic.
- 25% of daily search engine requests are pornography related.
- 35% of all Internet downloads are pornographic.
- 34% of Internet users received unwanted exposure to sexual material.
- Worldwide Pornography Revenues ballooned to $97.06 billion.
- Every second - 28,258 Internet users are viewing pornography.
- Every 39 minutes: a new pornographic video is being created in the United States.

Different researches and efforts have been carried out recently on how to block the pornography Websites among them content-based filtering is the most effective one (Ahmadi, Fotouhi, & Khaleghi, 2011).

Also, many software packages have been developed (Hammami, Chahir, & Chen, 2006) which mainly employ two kinds of approaches for classifying Web pages: static filtering and dynamic filtering. Static filtering is based on blocking a specific Web address via searching it in a reference list of black URLs. Although this method has high speed of processing, but its shortcoming is the requirement for instantly update of the URL list. This updating is a very hard task in the rapidly improving Web. Another problem is the high rate of over-blocking the usual pages such as pages with medical, sports, or arts topics, or blocking a Website because of only one immoral page on it. In dynamic filtering, the classification is performed based on the content analysis. First, the content of the pages is analyzed by using learning models and then the page is classified based on the content features. Static and dynamic filters give the parents different degrees of control over the censorship depending on the program.

The outline of the paper is as follows: Section 2 gives a brief overview of some filtering techniques. Section 3 gives a brief overview of skin detection techniques. Section 4 describes face detection techniques. In Section 5 the Proposed Filtering System is introduced. Experimental results are presented in Section 6. Finally, conclusion and future work are presented in Section 7.

2. RELATED WORK

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. Generally, there are two main approaches for pornography filtering software classification: classification based on textual content features, and classification based on both visual and textual features. The first group uses textual analysis mainly by searching a list of indicative keywords over the text such as sex, naked, etc. Some representative companies as NetNanny and SurfWatch (Girgis, Mahmoud, & Abd-El-Hafeez, 2010), operate by maintaining lists of URL’s and newsgroups and require constant manual updating. In the second methods, textual content-based analysis together with visual features are used to get a more robust classification. Visual features (skin and non-skin pixels) are extracted from the images in the Web pages by utilizing effective skin
Related Content

Variational Problems in Image Segmentation and I-Convergance Methods
www.igi-global.com/chapter/variational-problems-image-segmentation-convergance/4836?camid=4v1a

A Marked Point Process Model Including Strong Prior Shape Information Applied to Multiple Object Extraction From Images
www.igi-global.com/article/marketed-point-process-model-including/55096?camid=4v1a
3D Shape Compression Using Holoimage
[www.igi-global.com/chapter/shape-compression-using-holoimage/60261?camid=4v1a](www.igi-global.com/chapter/shape-compression-using-holoimage/60261?camid=4v1a)

Object Analysis with Visual Sensors and RFID
[www.igi-global.com/chapter/object-analysis-visual-sensors-rfid/77607?camid=4v1a](www.igi-global.com/chapter/object-analysis-visual-sensors-rfid/77607?camid=4v1a)