eVision

Content Filtering

An Overview

1 South 450 Summit Ave.
Suite 210
Oakbrook Terrace, IL 60181
Ph: 630.932.8920 ■ Fax: 630.932.8936
www.evisionglobal.com

AN OVERVIEW

Content Filtering

Printed: June, 2001

Mailing Address
eVision LLC
1 South 450 Summit Ave., Suite 210
Oakbrook Terrace, IL
60181

Introduction

The explosive growth of Internet has been fueled by two main factors: technology convergence and media convergence.

There are three main factors that have contributed to the technology convergence: increasing processor speed, affordable primary and secondary memory (RAM), and widening data pipes (large bandwidth). These factors have led to the development of tools that make it easy to create and deliver online content at a low cost. Web developers and content managers are now able to combine text, images, video, and audio to produce and deliver rich, multimedia content via the World Wide Web.

While the growing ease of use for posting and accessing complex, online media is changing the way we teach, learn, conduct business, govern, and live our every day lives, it also provides a convenient platform for the less mainstream portions of our society to advertise their views, such as pornography and messages of hate and violence. Because the majority of content on the Internet is unregulated, such objectionable material has become widely available at the click of a mouse. To make matters worse, recent government initiatives have made the Internet and its content accessible to unsuspecting children.

Most of us agree that some online material is inappropriate not only for children, but also for adults in certain environments such as workplaces, libraries, etc. This document describes a method for automatically filtering such images and videos based on their content, thus enabling people to block or "reject" images that are considered inappropriate. The same method can also be extended to include audio.

The Current State of Filtering

The filtering technologies that exist today are strictly text based. While their intent is to block inappropriate material, they often end up denying access to acceptable information as well and thus have become subject to much debate and criticism. Because there has been no better alternative to these text-based filtering solutions up until now, they have enjoyed moderate success in the market place.

To understand how filtering works, you must understand the two general approaches that today's methods use: inclusion filtering and exclusion filtering.

Inclusion Filtering

Inclusion Filtering grants Internet users access to specified, pre-screened sites only. The screening process consists of a person or group of people who decide which sites are appropriate, and the system is programmed to allow access to only those sites. This list of approved sites is then updated periodically on an as needed basis. This list is also referred to as the "White List". Religious organizations and some large corporations typically use this method of filtering.

Unfortunately, inclusion filtering often blocks much of the valuable information available on the Web as well. For most individuals and organizations, this method is far too limiting and frequently overreaches its intended function.

Exclusion Filtering

Exclusion Filtering uses an opposite approach to inclusion filtering. Instead of keeping a list of "safe" sites, it maintains a list of objectionable sites and grants access to any site that is not included on that list. An advantage of this filtering method over inclusion filtering is that it does not block a site until it is proven objectionable. However, as the number of objectionable sites on the Web grows on a daily basis, maintaining this type of list can become unwieldy.

Three methods are used to achieve site exclusion filtering: keyword blocking, packet filtering, and URL blocking. These can be used individually or in combination. Brief descriptions of these methods are as follows:

- **Keyword Blocking** is a process by which a site is scanned for keywords as it is downloaded. If any of the material downloaded contains the designated keywords, that site is blocked. The main problem with this method is that it works only on text and without any regard to context. Much of the content on the Web consists of images, which usually have no text on which to scan for keywords.

- **Packet Filtering** controls access by blocking requests to specific IP addresses that define individual sites. While this method is fast and simple, it does not allow for fine-grained control. It can also defeated by some of the newer technologies such as IP-independent virtual hosts.

- **URL Blocking** is process by which access is controlled by URL address. Since most sites are a collection of embedded URLs, this method provides fine-grained control. Most commercial software uses URL blocking.

Content-Based Image Filtering (CBIF)

As previously mentioned, text-based filtering methods use some knowledge of the domain and text scanning to impose restrictions. This makes it very hard to filter non-text information, such as visual and audio media. Because of this, the three objectives of any good filtering mechanism, **Accuracy, Scalability, and Maintainability,** have not been met by existing methods. The problem for filtering methods that provide accurate blocking is that they are hard to scale and maintain. Conversely, filtering methods that are easily scalable and maintainable are not as accurate.

The Internet filtering field has attracted a lot of attention over the last few years due to the growing number of web sites with objectionable content. One of the areas that has received the most attention is content-based filtering. Content-based filtering consists of examining an image (rather than text) for patterns to detect objectionable material and block the hosting site.

The process of CBIF consists of three specific steps.

1. Determine if an image on a site consists of large areas of skin-colored pixels

2. Segment the image

3. Match it against a reference set of objectionable images and object regions

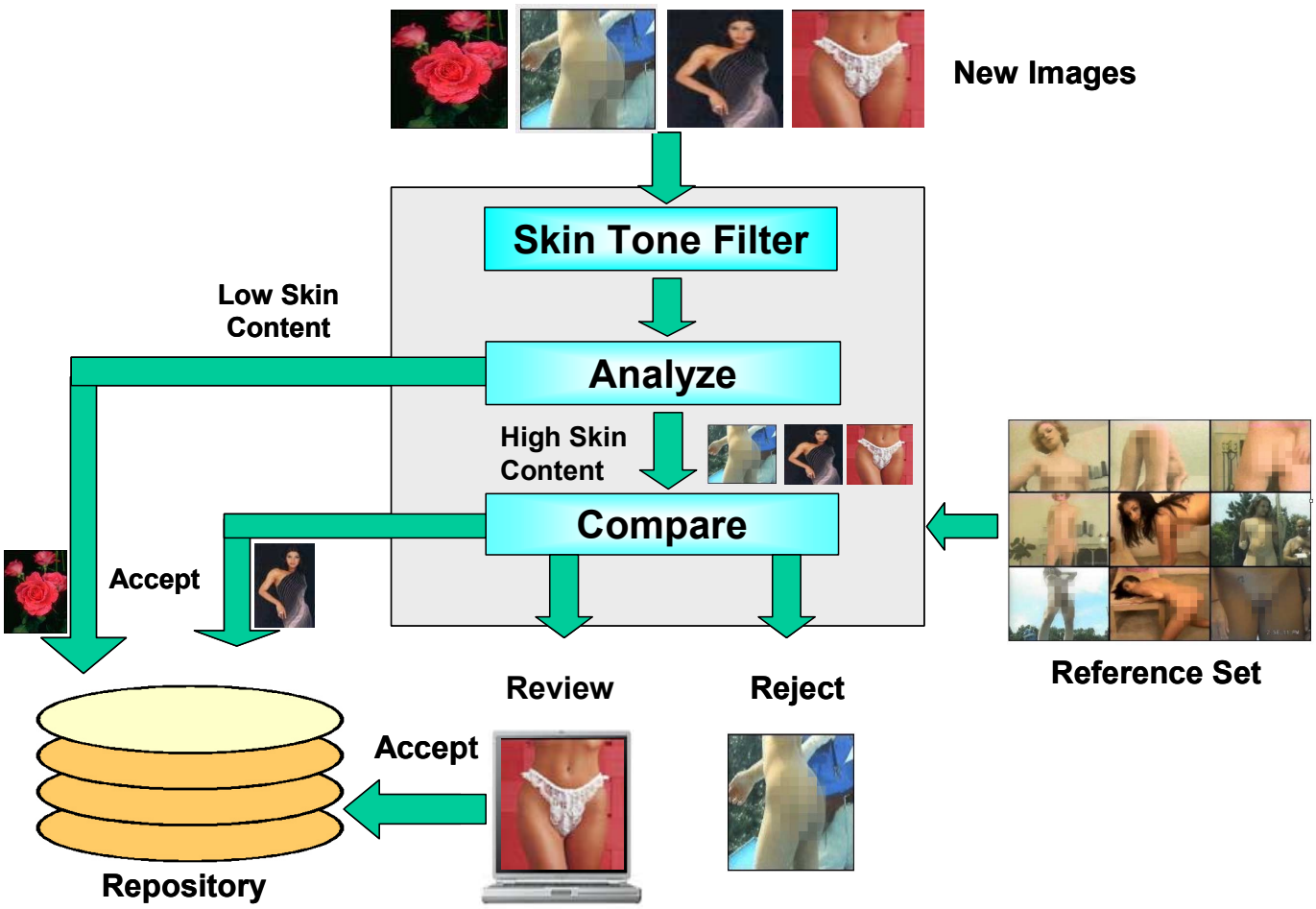
By its very nature this is a non-deterministic process, so the output of this filter has three states:

- An image is rejected outright if it matches an objectionable image with a similarity of 70% or better

- An image is queued for manual resolution if the similarity is between 40 - 70%

- Any similarity matches of <40% are accepted and access is allowed

Note: Please note that these percentage threshold values are arbitrary and can be adjusted based on filter accuracy.



The above figure illustrates the filtering process. This filter can be applied to local databases, corporate databases, and the Internet in general. The three steps within this process are described below:

Step 1

Images are filtered based on the presence of skin tones. The color of human skin is created by a combination of blood (red) and melanin (yellow, brown). These combinations restrict the range of hues that skin can possess. In addition, skin has very little texture. These facts allow us to ignore regions with high amplitude variations, and design a skin tone filter to separate images before they are analyzed.

Step 2

Because the images have already been filtered for skin tones, any images that have little to no skin tones will be accepted. They can be analyzed to generate signatures and added to the repository. Images that have an abundant presence of skin tones go through the third step. They are analyzed to generate signatures.

Step 3

The analyzed images are then compared to a pre-determined reference data set of representative images (such as pornographic images). If the analyzed images match any of the images in the reference set with over 70% similarity, then the image is rejected. If the similarity falls in the range of 40-70%, then those images are set aside for manual intervention. An operator can look at these images and decide to accept or reject. Images that fall below 40% similarity are accepted and added to the repository.

Conclusion

The currently available methods for filtering objectionable web sites are based strictly on text searches and manual list compiling, and thus are faced with the following problems:

- They are unable to accurately filter out inappropriate sites without also blocking access to acceptable sites
- Even when they can block out inappropriate sites accurately without affecting acceptable sites, there is no way to easily maintain and update them

The content-based filtering scheme described in this document solves both of these problems. It does this by recognizing regions of pixels within images that consist mostly of skin tones. Once these regions are extracted, they are compared with a reference set of objectionable images to filter for pornographic content. Three decisions are possible based on the degree of similarity between the new image and the reference set: an image can be accepted, rejected, or set aside for manual review. These threshold values can be tuned to specific data sets to minimize manual intervention.