On the Image Content of the Chilean Web

A. Jaimes[1], J. Ruiz-del-Solar[2], R. Verschae[1,2], D. Yaksic[1,2], R. Baeza-Yates[1], E. Davis[1], C. Castillo[1]

[1]Center for Web Research, Department. of Computer Science, Universidad de Chile, CHILE.
[2]Department of Electrical Engineering, Universidad de Chile, CHILE.

Abstract

In this paper we perform a study of the image contents of the Chilean web (.cl domain) using automatic feature extraction, content-based analysis and face detection algorithms. In an automated process we examine all .cl websites and download a large number of the images available (approx. 83,000). Then we extract several visual features (color, texture, shape, etc.) and we perform face detection using novel algorithms. Using this process we semi-automatically characterize the image content of the web in Chile in terms of the detected faces and the visual features obtained automatically. We present statistics of use to anyone concerned with the image content of the web in Chile. Our study is the first one to use content-based tools to determine the image contents of the web.

1. Introduction

The web is growing at an increasingly rapid pace. More importantly, faster computers and network connections are allowing creators of web content more freedom to add, with fewer constraints, larger quantities of images, graphics, and video. At the same time, people's interest in using images from the web has also increased. The web in Latin America is no exception, and many internet websites in countries like Chile are full of multimedia content. In a recent study [3], for instance, it was found that in 2001 the keyword "fotos" (photos) was the second most searched keyword in the Chilean search engine TodoCL [3][12].

Given the trend to enrich websites with multimedia, it becomes increasingly important to be able to characterize a given segment of the web according to the multimedia elements that it contains. This type of information is of great importance for internet service providers (who can determine required levels of regional service), for content producers, and for web search application developers.

Characterizing the multimedia contents of the web, however, is a challenging technical problem. First, one must deal with huge amounts of distributed data. Second, it is necessary to use media-specific content-based[1] analysis tools to be able to determine the content of the multimedia elements. With images and video, this means developing tools to automatically determine their visual characteristics: color, texture, shape, etc. More interestingly, it implies using algorithms to automatically detect objects of interest (e.g., faces). Obviously, given the large amounts of data, manual classification is not an option.

In this paper we investigate the image content of the web in Chile using an efficient web-crawling architecture, novel face detection algorithms, and content-based analysis tools (to extract low-level features such as color, shape and texture). Our goals are to propose a general framework to characterize the multimedia content of the web, and to provide some statistics specific to the web in Chile.

Our web-crawling architecture is based on a long-term schedule for collecting sites and a short-term schedule that worries about network politeness and use of resources (CPU, bandwidth) [1][4].

Automatic detection of faces in the images of interest is carried out by sequentially combining a skin detection algorithm and a state of the art face detector. The skin detection algorithm uses color analysis for determining if a given image pixel corresponds to skin or not. As many similar algorithms it is based on the fact that skin colors form a cluster in color space. The employed face detection algorithm corresponds to a cascade asymmetrical Adaboost detector [14]. The algorithm uses simple, rectangular feature face detectors (a kind of Haar wavelets), the integral image [13] for fast computation of these feature detectors, asymmetrical Adaboost as a boosting strategy for the training of the classifiers, and a cascade structure for combining successively more complex classifiers.

Finally, we extract low level visual features using standard techniques from content-based retrieval. Several feature extraction algorithms for color, shape and texture are used to compute content-dependent statistics across the collection.

We also summarize and discuss the results of face detection and visual feature extraction.

[1] We will use the term content-based to refer to *visual* content.

Proceedings of the First Latin American Web Congress (LA-WEB 2003)
0-7695-2058-8/03 $17.00 © 2003 IEEE

The article is structured as follows. Related work is presented in section 2. An overview of the framework is presented in section 3. In section 4 we discuss our face detection algorithm. In section 5 we describe the extraction of visual features and in section 6 we present statistics of the image contents in the Chilean web. We discuss our results in section 7 and conclude in section 8.

2. Related work

In this section we discuss related work in the areas of content-based analysis, face detection, and finally on studies of image content of the web (for details on previous work on web crawlers please see [1]).

The authors of [11] presented a system for automatically indexing images collected from the WWW. Images are automatically collected and assigned to categories based on text surrounding the images. In addition, visual features are extracted from the images to construct a search engine that allows search by visual content. The authors of [5] implemented a similar system, which in addition uses automatic face detection to index images on the web. Although our approach could be used to construct a search engine, our goal is instead to explore the use of content-based techniques to characterize a particular segment of the web (in this case .cl domain). In addition, our specific algorithms for web crawling, face detection, and feature extraction differ from those of [11][5].

A very comprehensive review of face detection can be found in [6] and [15]. The main approaches can be classified as feature-based (low-level analysis, feature analysis and active shape models) and image-based (linear subspace methods, neural networks and statistical approaches) [6]. Image-based approaches have shown better performance than feature based. Among them, systems like the ones developed by “Sung&Poggio”, “Rowley”, “Schneiderman” and SNoW have shown very good results (for references and exact performance information please see [6] and [15]). The system proposed by Viola and Jones [14] outperforms previous systems in terms of processing speed. This system uses simple, rectangular features (a kind of Haar wavelets), a cascade of filters that discard non-face images, the integral image for fast computation of these filters and asymmetrical Adaboost as a boosting strategy for the training of the detectors. Our face detector differs from previous work in that we combine the work of Viola and Jones with a simple and fast skin color filter, when color information is available.

Over the last few years many approaches have been developed to index images by content (what appears in the images) [7][10][9]. A large number of such approaches compare images using similarity measures between low-level visual features such as color, shape, and texture. Many approaches also seek to classify images at the scene (e.g., indoor, outdoor, etc.) and object levels (e.g., face, sky, etc.).

In spite of the large amount of work in this area, to our knowledge, there have not been any studies of web content that use content-based features to characterize the images on the web. In the work of [11] for example, over 500,000 images and videos were catalogued, but general statistics on the visual content of the images in the entire collection (or a subset of the collection using a pre-defined criteria such as our .cl domain) were not presented.

Finally, two previous studies have been performed on the contents of the web in Chile [2][3]. Neither one of them has focused on the image content of the web, only on the image’ types (GIF vs. JPG, etc.).

3. The proposed system

3.1 System Overview

We start the process by automatically collecting web pages in the .cl domain and extracting the links to images in each of the pages. Although it is possible to collect all of pages available, in practice, it is often necessary to reduce the amount of data to be analyzed. The general process, therefore, consists of a series of steps at which different filters can be applied to select subsets of the data. This is illustrated in figure 1.

We use the same terminology of [3]. A *page* is a document indexed by the crawler. A *site* is a logical web server identified by a sub-domain (e.g., *dcc.uchile.cl* which belongs to the *uchile.cl* domain). A *domain* is any name of the form *x.y* where, in this paper *y=cl*.

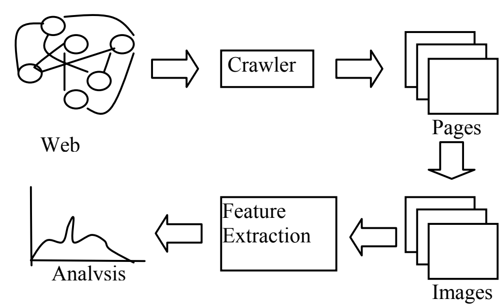


Fig. 1. Overall architecture.

First we obtain a list of the domains of interest. Then we use a crawler [1][4] to obtain the web pages in each of the selected domains. The next step consists of automatically extracting the links to images, downloading the images, and analyzing them.

3.2. The Crawler

The web crawler we used is described in detail in [1][4]. This crawler was chosen specifically because it is designed for high-performance, large-scale information gathering on the web.

3.3. Processing of Web Pages

The text associated with images on the web can undoubtedly be useful for their analysis. At this stage, therefore, decisions must be made regarding the amount of text that is extracted with each image. Similarly, decisions must be made about the selection of images to be downloaded. Although text information is not used in our current analysis, it has been found to be extremely useful [11].

3.4. Analysis of the images

After downloading each of the images we apply the face detection algorithms we have developed (section 4). In addition, we extract visual features (section 5) and use these results to characterize the overall visual content of the segment of the web of interest (the .cl domain).

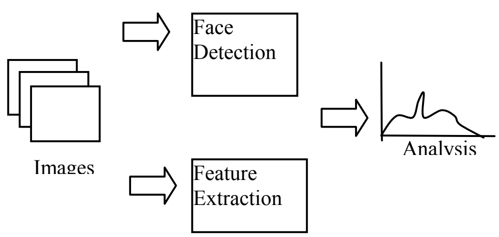


Fig. 2. Content-based analysis.

4. FACE DETECTION

4.1. Overview

Automatic detection of faces in the images of interest is carried out sequentially using a skin detection algorithm and a state of the art face detector.

4.2. Skin Detection

To reduce the search area for face detection (increasing the processing speed and decreasing the false detection rate), we used simple rules to verify if a given pixel corresponds to skin or not, using the pixels' normalized RG color space information. Normalized RG color space is computed as:

$$Rn \quad \frac{R}{R \quad G \quad B}, \qquad Gn \quad \frac{G}{R \quad G \quad B} \tag{1}$$

In the normalized RG color space, skin colors create a cluster that can be delimited by five straight lines (figure 3), whose parameters (m_i, c_i) were chosen by inspection. The skin pixels of the cluster were collected from several images in which skin pixels were manually selected. We choose a point (x,y) as a candidate of skin as:

$$SO(x,y) \quad \begin{matrix} 1 \quad m_i \quad (x,y).Gn \quad c_i \quad p_i \quad (x,y).Rn \\ i \quad 1,2,3,4,5 \\ p_i \quad 1,1 \\ o \quad otherwise \end{matrix} \tag{2}$$

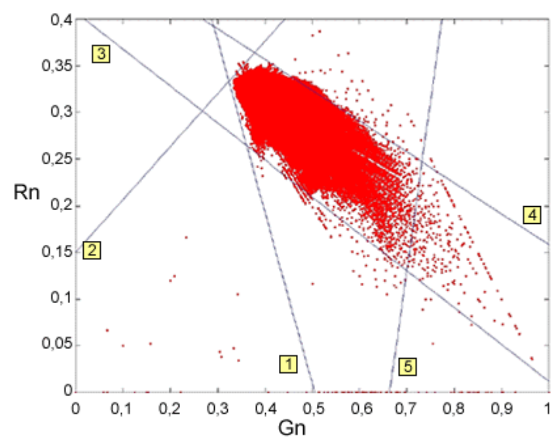


Fig. 3. Skin cluster en the RG color space.

Using this filter we obtain an image mask containing the skin color information.

4.3. Face Detection

The implemented detection subsystem detects frontal faces with small in-plane rotations and it is based mainly on [14]. This face detector corresponds to a cascade of filters that discard non-faces and let face candidates pass to the next stage of the cascade. This architecture seeks to have a fast face detector, considering the fact that only a few faces are to be found in an image, while almost all of the image area corresponds to non-faces. The fast detection is achieved in two ways: (i) having a small complexity in the first stages of the cascade (filters composed of few detectors, 2 to 5) and greater complexity in the later stages of the cascade (filters composed by many detectors, 100 to 200), and (ii) using simple features called rectangular features (the detectors), which are quickly evaluated using a representation of the image called the integral image [13]. In the integral image each pixel value corresponds to the accumulated sum of all pixel values between the current pixel location and the top left corner of the image.

Each of the filters of the cascade is trained using an asymmetric version of Adaboost (see explanation in [14]), which gives more importance to errors occurring during the training process when classifying faces as non-faces than to errors of non-faces classified as faces. Adaboost sequentially trains and selects a small number of rectangular features. The main problem of this face detector is training time, which can extend for weeks or even months when a single computer is used.

The final cascade Adaboost detector implemented in this work has 49 layers and was trained in about one month. To train each layer 1,500 face images and 4,000 images that did not contain any face were used. The training images were obtained mainly from the Internet using the Google image search engine, and from personal image collections. For training the first 2 filters, 4,000 and 3,000 non-face sub-window images were randomly chosen from our dataset. For training the remaining filters of the cascade, 1,500 non-face sub-window images wrongly classified by the already trained cascade were collected (a kind of bootstraping). In order to reduce the training time, a randomly chosen subset of the set of rectangular features was used in each iteration of the Adaboost algorithm. Each time that a decision rule was trained, only 50% of the training examples (faces and non-faces) were employed. As a result of the training, the final number of rectangular features used at each of the 21 stages of the cascade was 2, 5, 20, 20, 50 and 50 in the first 6 stages of the cascade. After that five stages of 100 and thirty eight stages of 200 rectangular features were used. The detector was trained using 24x24 pixel windows. For detecting faces at different scales, the detector was scaled multiple times using a scale factor of 1.2.

For each window of the image being considered we use the skin classifier as follows. The number of skin pixels is counted, and if this number is greater than 50% of the pixels of the window, the window is classified as a possible face region. Every possible face region is passed to the face classifier described above and all other windows are discarded. To count the number of skin pixels in a given window the integral image representation [13] of the skin mask is used. This representation allows us to get the sum of the pixels of the skin mask in constant time. With this procedure, face detection time was reduced by a factor of 2 and the number of false detections was reduced considerably with an increase in the face detection rate. The increase in the detection rate was achieved by reducing the number of stages in the cascade when the detector was applied to color images. In gray scale images 21 stages of the cascade were used, but in color images only 18 were used.

4.4 Processing of Overlapping Detections

Face windows obtained in the face detection module are processed and fused for determining the size and position of the final detected faces. Overlapping detections are processed for filtering false detections and for merging correct ones. All detections (detected face regions) are separated in disjoint sets using the following heuristic. Considering the inscribed circumference of each square face region, two detections belong to the same set if the sum of their circumference radius is smaller than 0.4 times the distance between their centers, and if each radius is not larger than twice the other. If a set contained only one element, this detection is discarded. Detections belonging to each set are merged by averaging the coordinates of the corners of all square face regions. Figure 4 shows an example of applying this heuristic on an image that has two close faces.

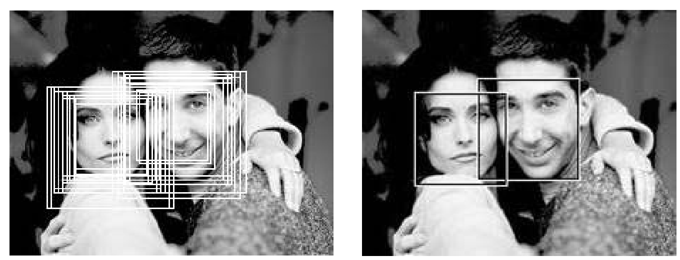


Fig. 4. Results from the use of the heuristic that separates and merges overlapping detections. (a) Overlapping detection, (b) Final detections

5. Content-based Analysis

We are investigating the use of content-based tools to characterize the images found on the Chilean web. Next, we describe the visual features employed.

5.1. Visual Features

We extract a set of visual features to represent color, shape and texture. Although some of these features are fairly simple, they are useful in giving us a snapshot of the visual content of images in the web.

Proceedings of the First Latin American Web Congress (LA-WEB 2003)
0-7695-2058-8/03 $17.00 © 2003 IEEE

Table 1. Visual Features (72). In brackets number and type of features (g: general; c: color; s: shape; t: texture).

Feature	Description
Area (1, g)	Number of pixels in the image
Aspect ratio (1, g)	Image aspect ratio.
First and second HSV dominant colors (6, c)	Quantized HSV dominant color [11].
Average SV components (2, c)	Average S and V in HSV space.
Number of most frequent color intensities (1, c)	Color intensities with number of pixels larger than 2% of image area [7].
Average and standard deviation in the R, G and B histograms (6, c)	Average and standard deviation in R, G and B histograms.
Percentiles (2%, 5%, 10%, 50%, 90%) in the R, G and B accumulated histograms (15, c)	Percentiles in the R, G and B accumulated histograms [8].
Edge Histogram in 0°, 45°, 90° and 135° (4, s)	Number of edge pixels in the horizontal, vertical, and diagonal (2) directions [7].
Extension of shape primitives (5, s)	Shape primitives (short, long, etc.) [7].
Texture features from co-occurrence matrices (20, t)	Energy, entropy, contrast, homogeneity and first moment calculated in 0°, 45°, 90° and 135° co-occurrence matrices.
Percentiles (2%, 5%, 10%, 50%, 90%) in the LBP accumulated histogram (5, t)	Percentiles in the LBP (Local Binary Partition) accumulated histogram [8].
Texture feature MMD (Mean Maximal Difference) (1, t)	Mean value of the maximal difference between each pixel and its neighbors.
Texture feature MTV (Mean Total Variation) (1,t)	Mean value of the minimal difference between groups of directional neighbor pixels [7].
Texture feature LD (Local Deviation) (4, t)	Mean value of standard deviation of 5 direct neighbor pixels, in four different directions (0°, 45°, 90° and 135°) [7].

We extracted color features in the RGB and in the HSV color space. Mean values, standard deviations and histogram percentiles were computed in each color band. The color space itself was quantized to 166 colors (quantization deemed sufficient to represent the majority of important colors). Edge direction histograms were computed by extracting edges from the images in different directions, thresholding to keep only the strongest edges, and counting the number of resulting edge pixels. These serve as a coarse measure of shape and have been used effectively in the construction of several classifiers (such as indoor and outdoor) [10]. Texture features were derived from co-occurrence matrices and LBP (Local Binary Partition) [8]. In Table 1 are displayed de 72 features (general, color, texture and shape) employed in this study,

5.2. Processing

For all images under analysis we extracted a feature vector containing each of the 72 features described in the previous section. The next step, described in section 6 consisted of obtaining statistics over the different features.

6. Statistics of the Image Content

In this section we describe the parameters used to collect the images from the web as well as the statistics obtained.

6.1 Domains and pages

In the most recent official study of the .cl domain [3] 1,988,706 pages were found in 38,307 sites in 34,867 domains. Current estimations of TodoCL [12] point out that the Chilean Web has 4.5 millions of pages +/- 10% and that the number of sites and domains is 80,000 +/- 10%.

For the study presented in this paper we obtained a list of domains from NIC Chile. In May 2003 we downloaded 800,000 pages to obtain enough representative images for developing our algorithms and performing a draft version of this study. For the final version of this study we downloaded 450,896 pages in August 2003. We did not consider secure pages (https) and pages were collected to a depth of up to 15 levels of links for static pages and 6 levels for dynamic pages. Pages marked with "non-index" were not considered.

6.2. Images

From the 450,896 pages used for this final version of our study we obtained 7,601,929 links to images. After filtering wrong links, as well as missing and repeated images (more than one link) we obtained 3,851,503 links to images. From this set of images we randomly selected 83,461 for performing this study. These images have different sizes and after downloading them, their total size in disk was 841,1 Mbytes.

We believe that taking into account the large number of pages and images of the Chilean Web and also considering the homogeneity of images and sites, these 83,461 images correspond to a representative sample of the total of 38 million images that we estimate are currently contained in the Chilean Web. The distribution of image types within this set of images is shown in Table 2.

Table 2. Type distribution of the images contained in the Chilean Web.

Type	Percentage of total
GIF	74.29
JPG (JPEG, JPE)	24.87
PNG	0.71
BMP	0.13

6.3. Faces

Using the automatic face detection algorithm we found that 3.18% of the images contained faces. Probably the reason of this small number is that most of the images correspond to graphics (button, icons, etc.) and not to pictures, where human faces are rather common. This aspect will be analyzed in the next section.

As figure 5 shows, the distribution of the number of faces in the images (considering only the ones that contain faces) is close to a Power law.

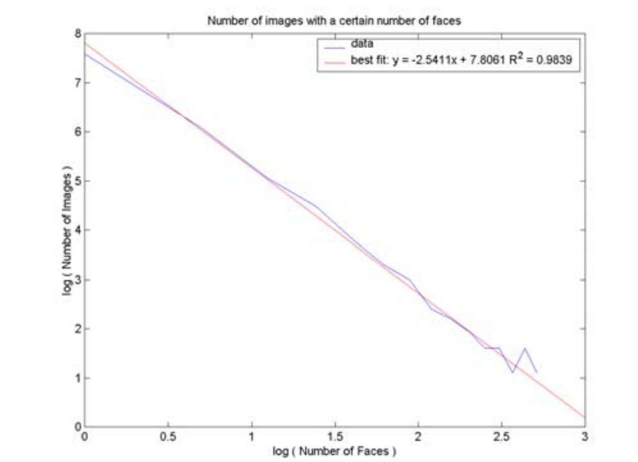


Fig. 5. A log – log graph of the number of images containing a certain number of faces.

Within the images containing faces the average number of faces per image is 1.81, and the maximum number of faces found in a single image is 39. Figure 6 shows a histogram of the number of images containing faces with a certain size (face area). As it can be seen, most of the images contain middle-size faces and no large faces.

Figure 7 shows the average number of faces contained on an image with a certain size. Figure 7.a considers all images and figure 7.b only images containing faces. It is interesting to notice that the number of faces in an image is not correlated with the image size

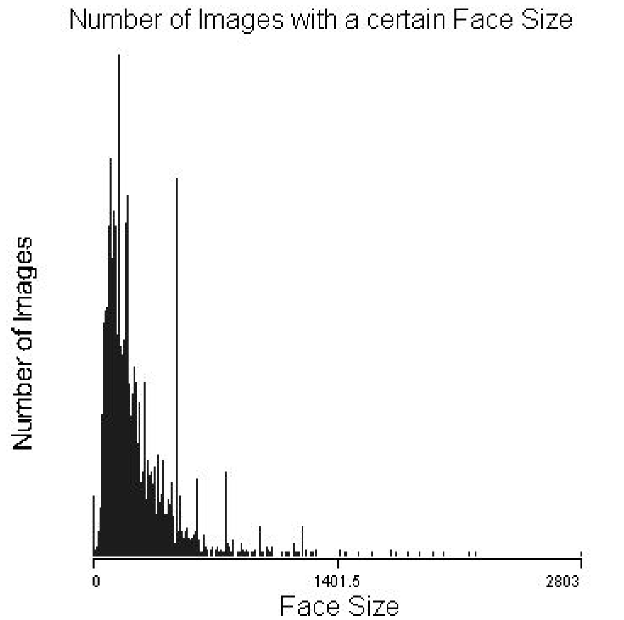


Fig 6. Histogram of the number of images containing faces with a certain size (face area).

The average size of the faces is 49.3x49.3 pixels and their average position, upper-left corner, is (225.6, 152.0). In figure 8 we present some relations between face area, as a percentage of the image area, and position, as a percentage of the image size (x or y). As it can be seen in these graphs faces are normally centered in the images, and they have a rather symmetrical distribution (position x vs. position y).

6.4. Visual Features

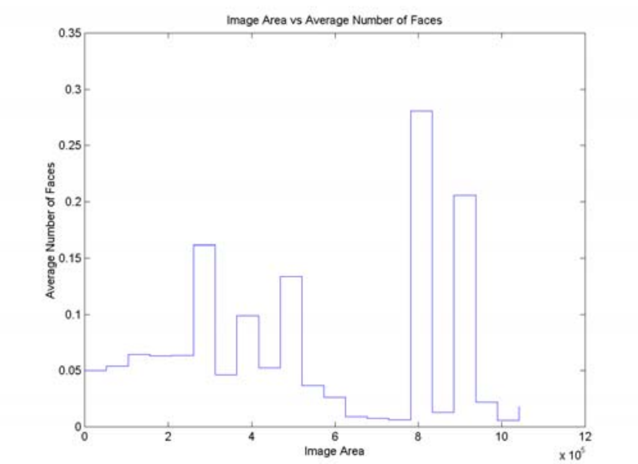
We extracted all 72 visual features presented in Table 1. Figure 9 shows statistics about the area of the analyzed images (log-log graph). The graph shows that the distribution of the number of images with a certain area is close to a Power law.

By analyzing the images under study (83,461) we realized that many of them correspond to graphics. For this reason we split our images set between large images (size larger than 50x50 pixels) and small images (size smaller than 50x50 pixels). We found that the percentage of large/small images is 21.10% / 78.89%. This means that for each large image we have four small images.

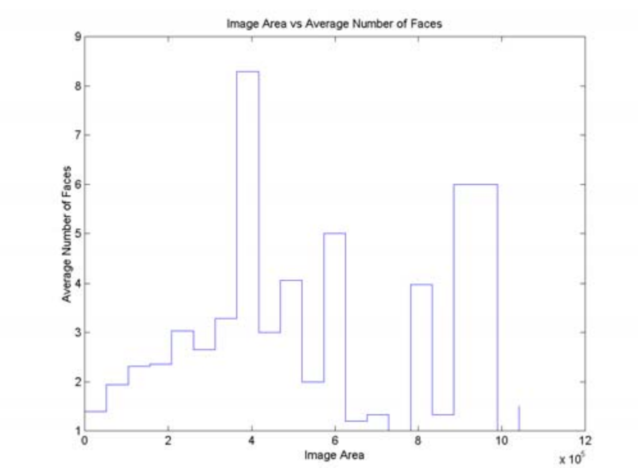
Figures 10-12 display statistics of some selected general, color, shape and texture features for both groups of images.

It is important to analyze the shape of these curves and not their amplitude values, which depend on the number of images under analysis in each category. In the displayed curves it is interesting to note that the features of large images are very different from the ones of small images. This fact confirms our hypothesis that the content of small images, mainly graphics, differs from the content of large images, mainly pictures. If we analyze the statistics presented in section 6.3 we will also see that most of small images do not contain faces.

Taking into account the featural difference between graphics and pictures, it is possible to automatically classify a given image into one of these two groups (see [11]).

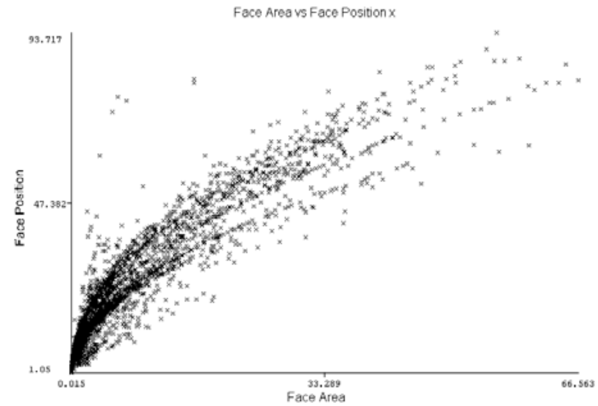


(a)

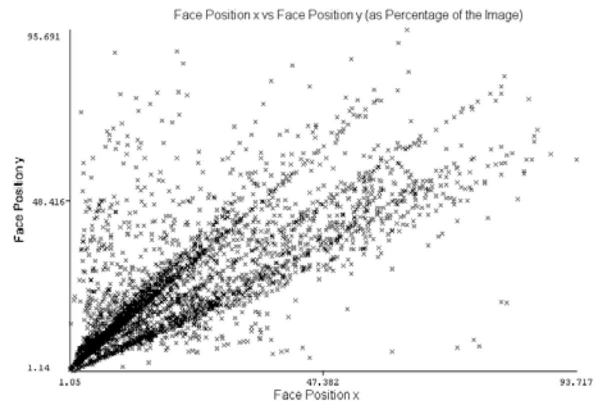


(b)

Fig 7. Image area against average number of contained faces. In (a) all images are considered, while in (b) only the ones containing faces.



(a)



(b)

Fig 8. Relations between face area, as a percentage of the image area, and position, as a percentage of the image size (x or y). Face area against face position in the image (a). Face position x against face position y.

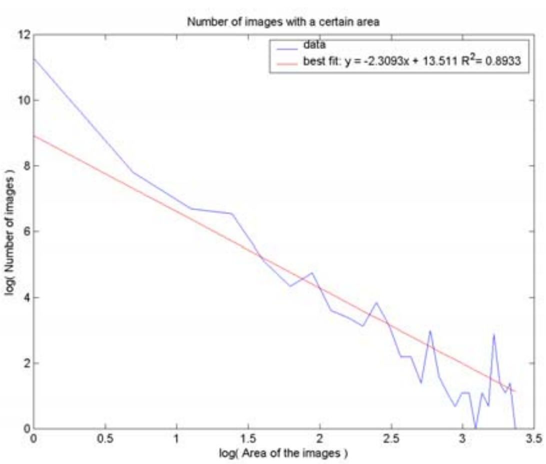


Fig. 9. Area statistics. A log-log graph of the number of images with a certain area.

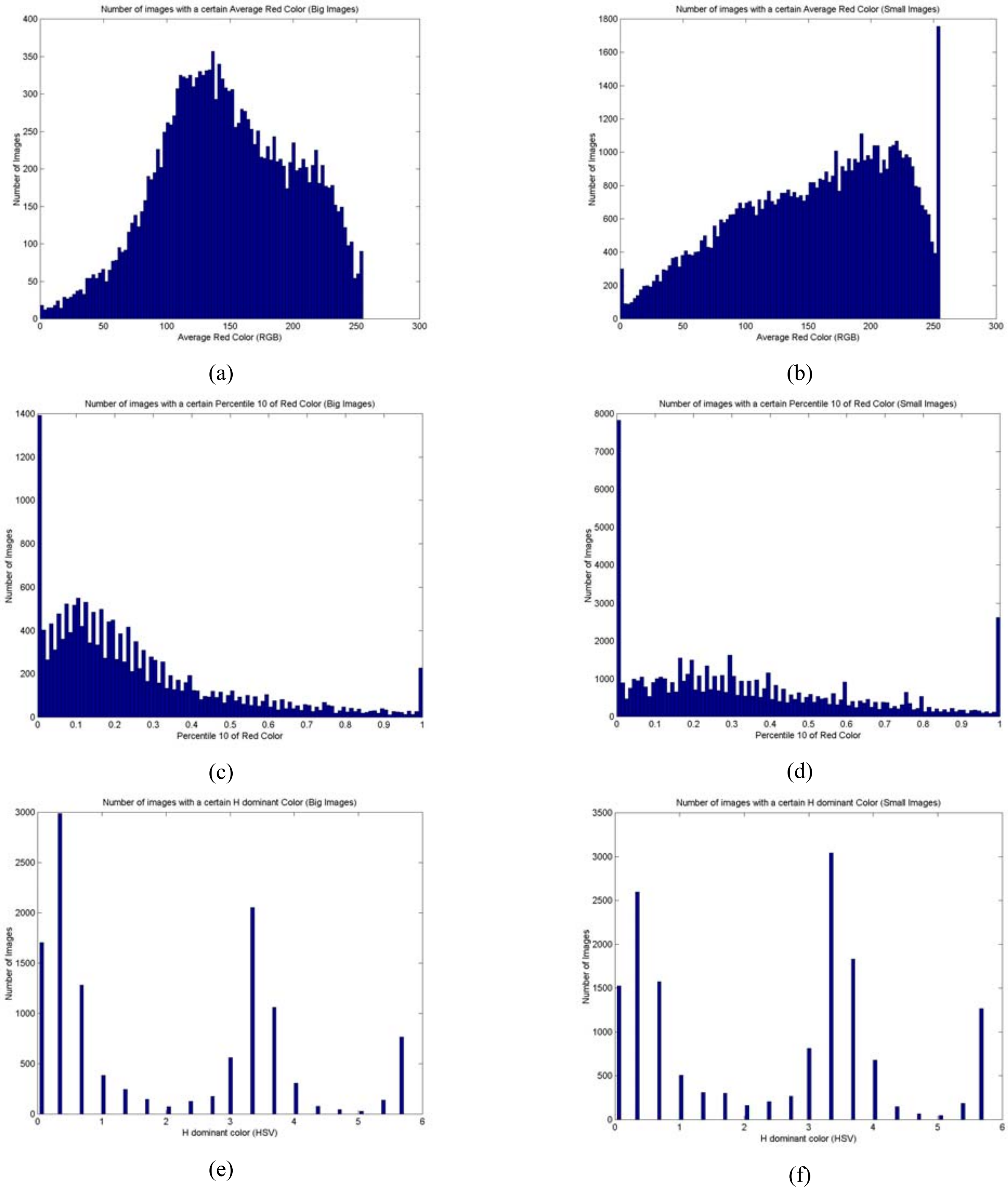
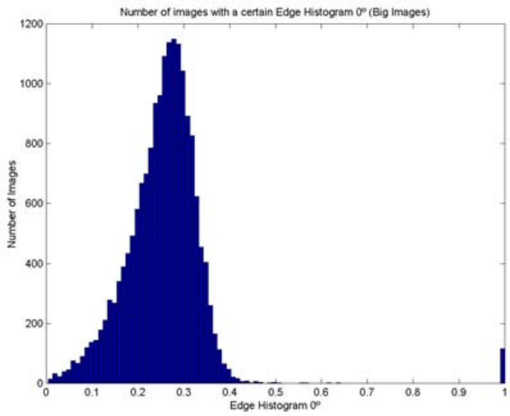
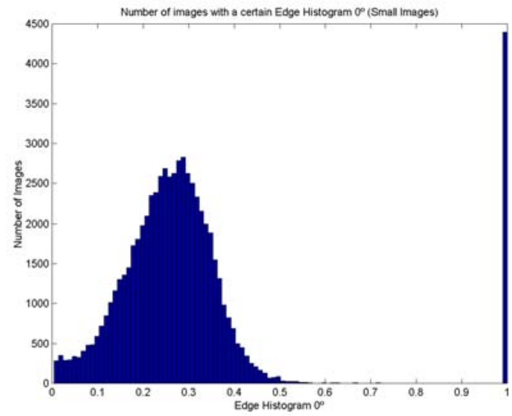


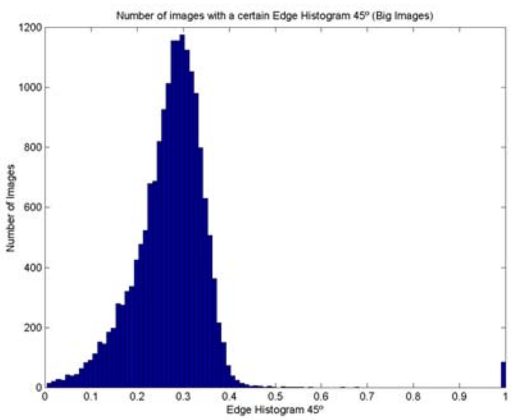
Fig. 10. Selected color features (see Table 1): Average in R histogram (a-b), 10 % percentile in red histogram (c-d), and first dominant H color (e-f). Left/Right graphs are for larger/smaller than 50x50 pixels images.



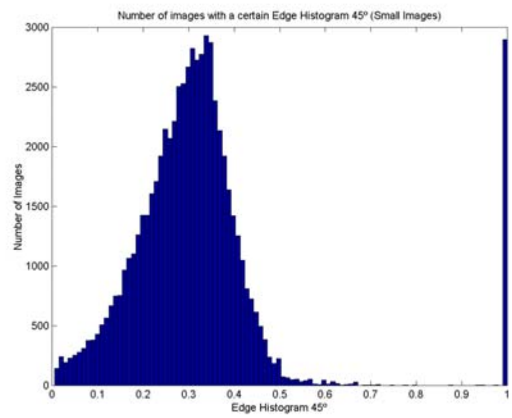
(a)



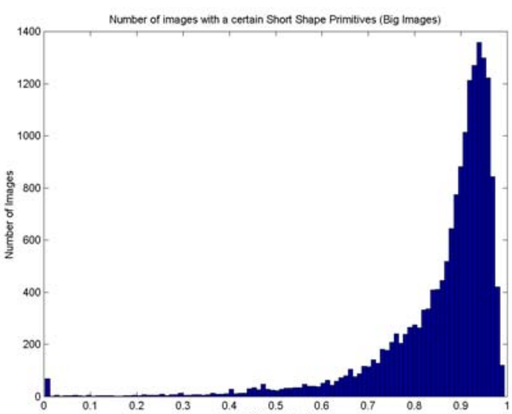
(b)



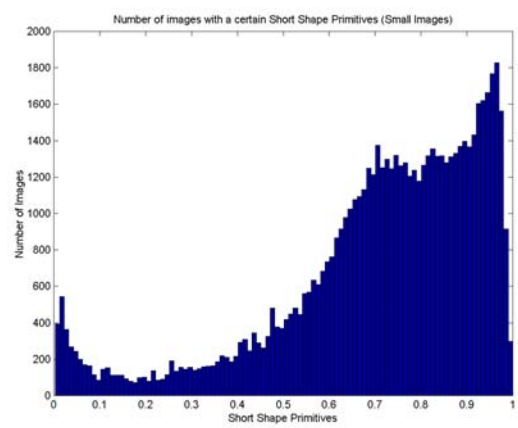
(c)



(d)

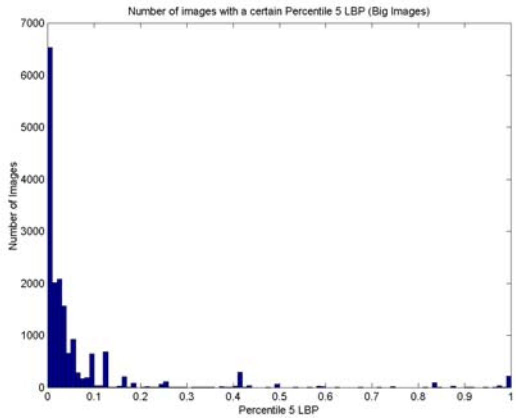


(e)

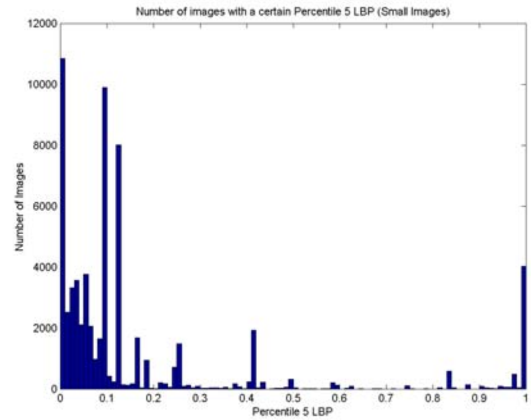


(f)

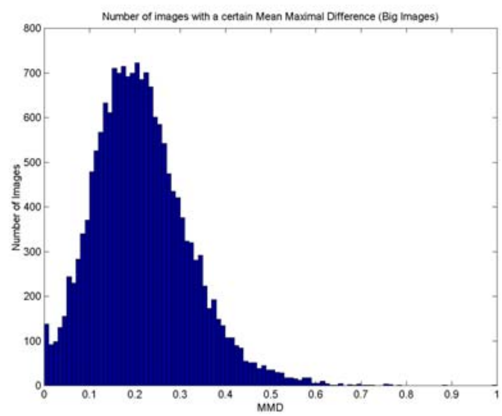
Fig. 11. Selected shape features (see Table 1): Edge histogram 0° (a-b), Edge histogram 45° (c-d), and shape primitives (short) (e-f). Left/Right graphs are for larger/smaller than 50x50 pixels images.



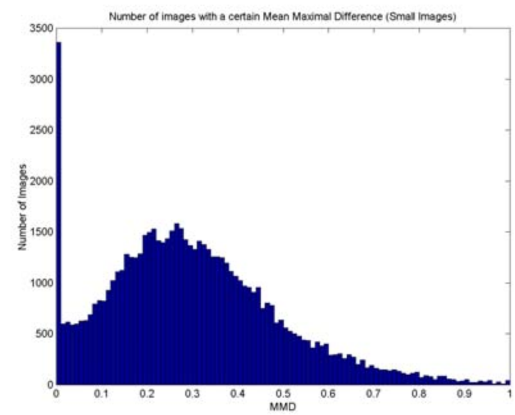
(a)



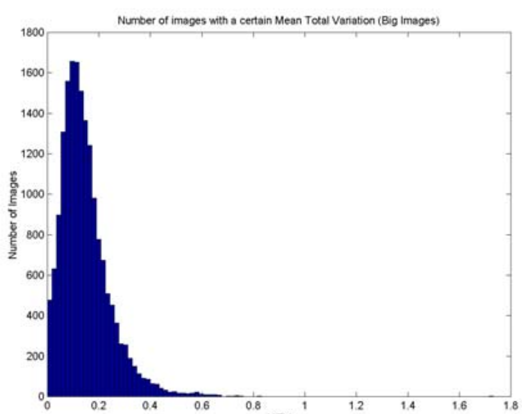
(b)



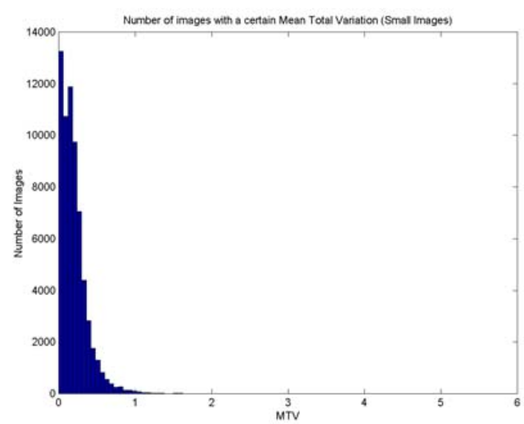
(c)



(d)



(e)



(f)

Fig. 12. Selected texture features (see Table 1): Percentile 5% LBP (a-b), MMD feature (c-d), and MTV feature (e-f). Left/Right graphs are for larger/smaller than 50x50 pixels images.

6.5. Processing Time

Page Gathering: The process of page gathering or collection takes about 72 hours for the first million of pages using a single, standard PC running Linux. For the remaining images (over one million), the process slows down a bit, but it is difficult to quantify the exact required time, because it depends in the web topology. In any case we can say that using a single PC, the page recollection of the whole Chilean web takes about two weeks. Obviously, this process can be speeded up using more PCs.

Feature Extraction: The process of automatic extraction of the 72 visual features on the 83,461 images under analysis takes about 47 hours on a single, standard PC running Linux with feature extraction implementations in C. The total amount of processed information was 841,1 Mbytes (remember that images have different sizes).

Face Detection: The process of face detection on the 83,461 images takes about 24 hours on a single, standard PC running Linux with face detection implementations in C. As mentioned before, this time can be reduced using more than one PC.

7. Discussion

We processed a total of 83,461 images and obtained some statistics to characterize their content. We found that filtering decisions (to reduce the amount of data to be processed) are crucial and must be made carefully. Obtaining the basic statistics is trivial once the images have been downloaded, but analysis of the results requires visual inspection of the content (or further automatic analysis). It is interesting to determine the types of images that contribute to certain statistics (e.g., determine the percentage of "small" images that are photographs and not graphics, or that contain faces).

Another interesting issue related to the analysis is the accuracy of automatic results. In most approaches to automatic classification of scenes or detection of objects (e.g., faces), a small database is used as a test base. The accuracy of the results can be easily measured once the test set is manually labeled. When automatic detection algorithms are applied to large sets of data it becomes more difficult to determine their accuracy. In any case we can say that for the detection of faces in this kind of images (images present in the Web), the detection rate of our face detector is about 80% (4 from 5 faces are detected). The rate of false-positives (wrong detected faces) is about 1/1.000.000 for each processed window (for analyzing an image a lot of windows are processed), for black and white images (the skin filter is not used). For color images (skin filer is used), we obtain about one false positive every 5-7 images, for typical 320x240 images. Altogether (color and non-color images), we obtain about one false positive every 20 images, for typical 320x240 images.

8. Conclusions & Future Work

We presented a general framework for characterizing the image content of a segment of the web. The proposed process consists of automatically collecting web pages and downloading images from the links contained in the pages. For content-based analysis we use face detection algorithms we have developed and use standard content-based retrieval tools. We presented statistics of the web in Chile (.cl domain) using the results of automatic face detection and the values obtained from visual feature extraction algorithms on a set of 83,461 images obtained automatically from the Chilean web.

Although we computed interesting statistics further analysis is necessary. Future work includes the implementation of more complex feature extraction algorithms, the use of text to characterize the contents of a segment of the web (i.e., the text surrounding an image), and expansion of this work to perform automatic clustering and classification of the multimedia elements (graphics vs. pictures, etc.). This is a first step to develop a image search engine based in real image content and not in the surrounding text or other attributes like in Google or Fast.

Acknowledgements

This research was funded by Millenium Nucleus Center for Web Research, Grant P01-029-F, Mideplan, Chile.

9. References

[1] R. Baeza-Yates, and C. Castillo, "Balancing collection volume, quality and freshness in a web crawler," in A. Abraham. J. Ruiz-del-Solar, M. Köppen (Eds.), *Soft-Computing Systems: Design, Management and Applications*, *Frontiers in Artificial Intelligence and Applications* 87, IOS Press, pp. 565 – 572, 2002.

[2] R. Baeza-Yates and C. Castillo, "Relating Web Characteristics with Link Based Web Page Raking," *Proc. of SPIRE 2001*, IEEE CS Press, Laguna San Rafael, Chile, pp. 21-32, Nov. 2001.

[3] R. Baeza-Yates, B.J. Poblete, and F. Saint-Jean, "Evolución de la Web Chilena 2001-2002 (Evolution of the Chilean Web 2001 - 2002)," Center for Web

Research, Department. of Computer Science, Universidad de Chile, January 2003 (in Spanish).
[4] C. Castillo, and R. Baeza-Yates, “A new model for web crawling,” WWW2002, Honolulu, USA, 2002.
[5] C. Frankel, M.J. Swain and V. Athitsos, “WebSeer: An Image Search Engine for the World Wide Web,” *University of Chicago Technical Report TR-96-14*, July 31, 1996.
[6] E. Hjelmås, B. K. Low, “Face detection: A survey,” *Computer Vision and Image Understanding* 83, pp. 236-274, 2001.
[7] A. Jaimes, “Conceptual Structures and Computational Methods for Indexing and Organization of Visual Information," Ph.D. Thesis, Department of Electrical Engineering, Columbia University, February 2003.
[8] M. Niskanen, O. Silven, and H. Kauppinen, “Color an Texture based Wood Inspection with non-supervised Clustering,” *Proc. of the 12th Scandinavian Conf. on Image Analysis* - SCIA 2001, 336 - 342, Bergen, Norway, June 11-14, 2001.
[9] Y. Rui, T.S. Huang, and S.-F. Chang, “Image Retrieval: Current Directions, Promising Techniques, and Open Issues,” *Journal of Visual Communication and Image Representation*, No. 10:1-23, 1999.
[10] A.W.M. Smeulders, M. Worring, S. Santini, A. Gupta, and R. Jain, “Content-Based Image Retrieval at the End of the Early Years”, *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 22, No. 12, pp. 1349-1380, Dec. 2000.
[11] J.R. Smith and S.-F. Chang, “An Image and Video Search Engine for the World-Wide Web,” *Proc. of SPIE Storage & Retrieval for Image and Video Databases V*, Vol. 3022, pp. 84-95, San Jose, CA, Feb. 1997.
[12] TodoCL (www.todocl.cl)
[13] P. Viola, M. Jones, “Robust Real-time Object Detection,” *Second Int. Workshop on Statistical and Computational Theories of Vision* (ICCV), Vancouver, Canada, July 2001.
[14] P. Viola and M. Jones, “Fast and Robust Classification using Asymmetric AdaBoost and a Detector Cascade,” *Advances in Neural Information Processing System* 14, MIT Press, Cambridge, MA, 2002.
[15] M. Yang, D. Kriegman, N. Ahuja, “Detecting Faces in Images: A Survey”, *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol. 24, No 1, pp. 34-58, Jan. 2002.