Adult Image Detection System

Bibek Raj Dhakal
062BCT506
bibekraj9@gmail.com

Biru Charan Sainju
062BCT507
thugliffe@gmail.com

Suvash Sedhain
062BCT548
mesuvash@gmail.com

Mar 3,2010

Abstract

Adult imagery or as they put it pornography has been there for a long time and with the advent of web it has only just grown. Pornography is considered as obscene material whose intention is to provoke sexual arousal. Pornography which was then a small publication has emerged as one of the highly distributed information on the internet. Pornography which is by no means unnatural may certainly be harmful to children and affect the efficiency of the workers. So, it should be managed and certainly shouldn't be popping where it is not meant to.

So we propose to design a system that takes an image and detects if it is a pornographic image or not. We intend to be using SVM (Support Vector Machines) to exploit primitive information from different images and use the knowledge to determine whether given image is pornographic or not.

Our system will be using the skin detection model to do the prediction based on the machine learning of the images.

List of Tables

List of Figures

Objectives:

Some of the major objectives of our project can be listed as

- To develop a system can learn the patterns of the feature extracted from the images
- To use SVM to predict whether a given image Is pornographic or not based on the leanings from the previous data of the images
- To distinguish adult content from the non adult ones and server as a tool for filtering images

Contents

Chapter 1

Introduction

Our human civilization has been influenced and intoxicated by the web revolution. People of every age use the web for different needs and purposes. Some use it for fun; some use it for their studies and find information while some live on it. Images are essentially part of the modern web and we all agree to the fact that a single picture is worth thousand words. We can find all sort of information on the web and it has been a part of our daily life. However, it also has abundance of images and contents that may be unsuitable for certain age groups. Finding pornographic images posted on social sites and links on study groups is not a new thing in todays world. Pornographic images certainly need to be managed and unavailable to children and men at work.

1.1 Background

Our human civilization has been influenced and intoxicated by the web revolution. People of every age use the web for different needs and purposes. Some use it for fun; some use it for their studies and find information while some live on it. Images are essentially part of the modern web and we all agree to the fact that a single picture is worth thousand words. We can find all sort of information on the web and it has been a part of our daily life. However, it also has abundance of images and contents that may be unsuitable for certain age groups. Finding pornographic images posted on social sites and links on study groups is not a new thing in todays world. Pornographic images certainly need to be managed and unavailable to children and men at work.

At the same time, features of the current version of LaTeX (LaTeX 2ε) are illustrated — such as mathematical expressions, numbering and cross-referencing, bibliography and citations, graphics and tables. Comparison of the source files with the printer-ready document will answer a few FAQs: '*How can I do . . . in LaTeX?*'.

However, this is *not* a textbook on LaTeX — for that, use the '*Not so Short*' notes by Oetiker *et al.* [**?**]. They are written for novices, and are a pleasure to read. They are available free on-line, and are kept up-to-date. The LaTeX book at *Wikipedia* [?] includes the *Not so Short* material and is good for reference too. Access these via the LaTeX resources page [?].

For more advanced features see *eg.* '*The Companion*' [?].

Well-meant advice on LaTeX for report-writing and poster-making is available[1] in room CM315, where there are reference copies of both *The Companion* [?] and *The Graphics Companion* [?].

Even if you are misguided enough [?] to prepare your report in *Word*, this template at least exemplifies a good structure — and gives advice about references and help with typography.

1.2 Contents

The main body of this report is divided as follows.

Chap. **??** has some examples of mathematics, then Chap. **??** deals with graphics and includes Sec. **??** about tables. The Conclusion, in Chap. **??**, summarises what's been achieved, the open questions and what could be done next.

Then comes the Bibliography, listing all sources of material, data and computer programs used, *etc.* Its construction is explained in [?, Sec. 4.2] and there's more about it in App. **??**.

Otherwise appendices typically hold basic background theory, or additional or similar examples, or longer proofs (App. **??**) — anything you need but which would hold up the main flow of the story. You could also use an appendix for listings of any computer programs that you've written (App. **??**).

Here there's information about using a PC (App. **??**) plus brief advice on grammar and typography (App. **??**).

---

[1] From `bob.johnson@dur.ac.uk`

Chapter 2

Literature Review

Current adult image filtering techniques can be classified into three categories: keyword based, blacklist based and content based. Our proposed system is content based i.e. images will be classified on the basis of their content. The classification approach we applied can be broadly classified into two categories- Skin based and non-skin based.

2.1 Skin Based Approach

Our human civilization has been influenced and intoxicated by the web revolution. People of every age use the web for different needs and purposes. Some use it for fun; some use it for their studies and find information while some live on it. Images are essentially part of the modern web and we all agree to the fact that a single picture is worth thousand words. We can find all sort of information on the web and it has been a part of our daily life. However, it also has abundance of images and contents that may be unsuitable for certain age groups. Finding pornographic images posted on social sites and links on study groups is not a new thing in todays world. Pornographic images certainly need to be managed and unavailable to children and men at work.

2.1.1 Skin Color Detection Method

The most important feature that provides clues to image content is color. Color is a low level feature, which makes it computationally inexpensive and therefore suitable for real-time object characterization, detection and localization. Generally pornographic images show a lot of skin and thus skin color is a basic feature used in detecting pornographic images. The main goal of skin color detection or classification is to build a decision rule that will discriminate between skin and non-skin pixels. Identifying skin colored pixels involves finding the range of values for which most skin pixels would fall in a given color space

Color Spaces

The purpose of a color space is to facilitate the specification of colors in some standard, generally accepted manner. A color space is a specification of a coordinate system and subspace within a system where each color is represented by a single point. Various color spaces are used for processing digital images. For some purposes, one color space may be more appropriate than others.

The RGB Color Space

The RGB color space originated from CRT display applications. In the RGB space each color appears in its primary spectral component of red, green, and blue. Images represented in the RGB space consist of three component images, one for each primary color. When fed into an RGB monitor, these images combine on the phosphor screen to produce a composite color image (Gonzalez and Woods, 2002).The RGB color space is one of the most widely used color spaces for storing and processing digital image. However, the RGB color space alone is not reliable for identifying skin-colored pixels since it represents not only color but also luminance. Skin luminance may vary within and across persons due to ambient lighting so it is not dependable for segmenting skin and non-skin regions. Chromatic colors are more reliable and these are obtained by eliminating luminance through some form of transformation. The color spaces Normalized RGB, HSV, and YCbCr are transformations commonly used by studies on skin color (Waibel et al., 1999).

The HSV Color Space

The HSV (Hue, Saturation, Value/Intensity/Luminance) color space describes color with intuitive values, based on the artists idea of tint, saturation and tone. This was introduced when there was a need to specify color properties numerically. Hue defines the dominant color as described by wavelength, for instance the distinction between red and yellow. Saturation measures the colorfulness of an area in proportion to its brightness such as the distinction between red and pink. Value refers to the color luminance, the distinction between a dark red and a light red. For skin detection, the value component is discarded to eliminate the undesirable effect of uneven illumination. The transformation is defined by

$$H = \arccos \frac{\frac{1}{2}(R-G)+(R-B)}{\sqrt{(R-G)^2+(R-B)(G-B)}} \tag{2.1}$$

$$S = 1 - \frac{3min(R,G,B)}{R+G+B} \tag{2.2}$$

$$V = \frac{R+G+B}{3} \tag{2.3}$$

Some studies show that HSV is invariant to highlights at white light sources, to matte surfaces, and ambient lighting. However, hue discontinuities and the computation of the luminance component conflict badly with the properties of color vision.

The YCbCr Color Spacr

YCbCr or YCbCr is a family of color spaces used as a part of the color image pipeline in video and digital photography systems. Y is the luma component and CB and CR are the blue-difference and red-difference chroma components. Y' (with prime) is distinguished from Y which is luminance, meaning that light intensity is non-linearly encoded using gamma.

Y'CbCr is not an absolute color space, it is a way of encoding RGB information. The actual color displayed depends on the actual RGB colorants used to display the signal. Therefore a value expressed as YCbCr is only predictable if standard RGB colorants or an ICC profile are used.The transformation from RGB to YCbCr is defined by

$$Y' = 16 + (65.481R + 128.553G + 24.966B) \tag{2.4}$$

$$Cb = 128 + (-37.797R - 74.203G + 112B) \tag{2.5}$$

$$Cr = 128 + (112R - 93.786G - 18.214B) \tag{2.6}$$

Procedure

We have implement skin based classifier for nudity detection in Images. Skin based implementation composed of

- Skin segmentation algorithm
- Feature selection and SVM training

Skin Segmentation Algorithm

Skin segmentation is widely used in human and face detection applications. Histogram based algorithm are very popular due to its proved efficiency.

Histogram density for skin and non-skin pixels is calculated from the dataset (Compaq Dataset and manually collected images).The color of skin in the visible spectrum depends primarily on the concentration of melanin and hemoglobin. Distribution of the skin colors is affected by the ethnicity an illumination conditions. However, under arbitrary conditions of illumination the variation in skin color will be less constrained. Given a sufficiently large collection of labeled images captured under a wide variety of imaging conditions, we can model the distribution of skin and non-skin colors accurately.

We constructed skin and non-skin histogram models using our classier training set of images. Given skin and non-skin histograms we can compute the probability that a given color value belongs to the skin and non-skin classes:

$$P(rbg|skin) = \frac{s[rgb]}{T_s} \tag{2.7}$$

$$P(rbg|\neg skin) = \frac{n[rgb]}{T_n} \tag{2.8}$$

where s[rgb] is the pixel count contained in bin rgb of the skin histogram, n[rgb] is the equivalent count from the non-skin histogram, and T_s anmd T_n are the total counts contained in the skin and non-skin histograms, respectively.

From the skin and non-histogram models we can construct skin pixel classifier. Skin pixel classifier can be mathematically expressed in terms of likelihood ratio as:

$$\frac{P(rgb|skin)}{P(rgb|\neg skin)} \geq \theta \tag{2.9}$$

Where $0 \leq \theta \geq 1$ is the threshold and can be adjusted to trade-off between correct detections and false positives. We can write as θ a function of the priors and the cost of false positives and false negatives:

$$\theta = \frac{c_p P(\neg skin)}{c_n P(skin)} \tag{2.10}$$

Where c_p and c_n are the application-dependant cost of false positives and false negatives, respectively. One reasonable choice of priors is $P(skin) = T_s/(T_s + T_n)$. Using the likelihood equation each pixel in and image can be classified as skin or non-skin.

We also implemented gaussian mixture model for skin detection. we used precomputed mixture parameter [3].It was found that gaussian mixture model is computationally expensive and doesn't fit into our performance requirement.

Feature selection and SVM training

We computed feature vector from the output of the skin detector and trained a classifier on these features to determine whether the image is adult or not. We tested for different set of features vector consisting of Percentage of pixels detected as skin, Average probability of the skin pixels, size in pixels of the largest connected component of skin, Number of connected components of skin etc. However, the system didnt meet our desired accuracy.

2.2 Non Skin Based Approach

2.2.1 BIC

BIC is a pixel classification algorithm that labels pixels into two categories Border and Interior. BIC is based upon Point image processing algorithm.

The BIC approach basically composed of three main components:

- a simple and powerful image analysis algorithm that classifies image pixels as border or interior
- a logarithmic distance to compare histograms
- a compact representation for the visual features extracted from images.

Image Analysis

BIC image analysis algorithm relies on the RGB color-space uniformly quantized in 4X4X4=64 colors. BIC can be implemented using other color-space quantization for eg. YCbCr, HSV etc. After the quantization step, image pixels are classified as border or interior pixels. A pixel is classified as border if it is at the border of the image itself or if at least one of its 4-neighbors (top, bottom, left and right) has a different quantized color. A pixel is classified as interior if its 4-neighbors have the same quantized color. It is important to observe that this classification is mutually exclusive (either a pixel is border or it is interior) and it is based on a inherently binary visual property of the images. We choose 4-neighbors instead of 8-neighbors because, given the simplicity and generality of the problem, the use of 4-neighbors is able to reduce the image analysis complexity without perceptual losses in terms of retrieval effectiveness.

After the image pixels are classified, one color histogram is computed considering only border pixels, and another color histogram is computed considering only interior pixels. In this way, we have the border/interior classification represented for each quantized color.

Logarithmic Distance function

Image analysis algorithm describes an image by the means of two color histogram with 64 bins. Two histogram can be combined into a single histogram with 64 bins so that vectorical distance function like L1-Norm and L2-Norm can be used to compare the visual features.

Vectorical distances have also well-known limitations. One of such limitations is that a high value in a single histogram bin dominates the distance between two histograms, no matter the relative importance of this single value.To avoid the problem of high values in a single histogram bin dominate the distance between two histograms, the dLog distance function is used. This function uses a logarithm scale reducing 28 times the range of distances between the smallest and the largest histogram bin values, given that in the log-scale a bin is normalized in the interval [0,9].

The dLog function compares histograms in a logarithmic scale, and is defined as:

$$dLog(q,d) = \sum_{i=0}^{M} |f(q[i]) - f(d[i])| \tag{2.11}$$

$$f(x) = \begin{cases} 0 & \text{if } x = 0 \\ 1 & \text{if } 0 < x \leq 1 \\ |\log_2 x| & \text{,otherwise} \end{cases} \tag{2.12}$$

In the given equation 2.11, q and d are two histograms with M bins each. The value q[i] represents the i^{th} bin of histogram q and d[i] represent the i^{th} bin of histogram d. The histogram bins are normalized between 0 and 255.

The comparison of histograms with the dLog function does not solve the problem of histogram bins with very high values, but diminishes its effects in most of the situations. In a log-scale, the difference between the largest and the smallest distances between histogram bins becomes smaller than in the original scale. In the original scale, the smallest distance between histogram bins is zero (both images have the same amount of a particular color) and the largest distance is 255 (when the images have just one color and they are different). In our log-scale, the smallest distance is 0 and the largest distance is just 9. The range of distances in the original scale is thus 255=9 = 28 times larger than in the proposed log-scale.

Compact Visual Features Representation

After applying the DLog function we obtain one histogram with 2XQ bins where each bin contains integer values between 0 and 9.Each image is compactly represented by this histogram vector where each position represents a bin. Thus produced compact representation of the images can be directly plugged into SVM as a features vector.

2.2.2 MPEG-7

The MPEG-7 standard[4] is an international standard since September 2001 which species metadata for describing multimedia content. The interest- ing part for our project is this part of the standard which denes visual descriptors. These are structures to describe multimedia data. Their exact extraction methods are not standardized. shows an overview of MPEG-7 visual descriptors which are suitable for still images. Since 2001, lots of research has been conducted on making use of these standardized visual descriptors in the eld of Computer Vision, especially in Content-Based Image Retrieval Systems

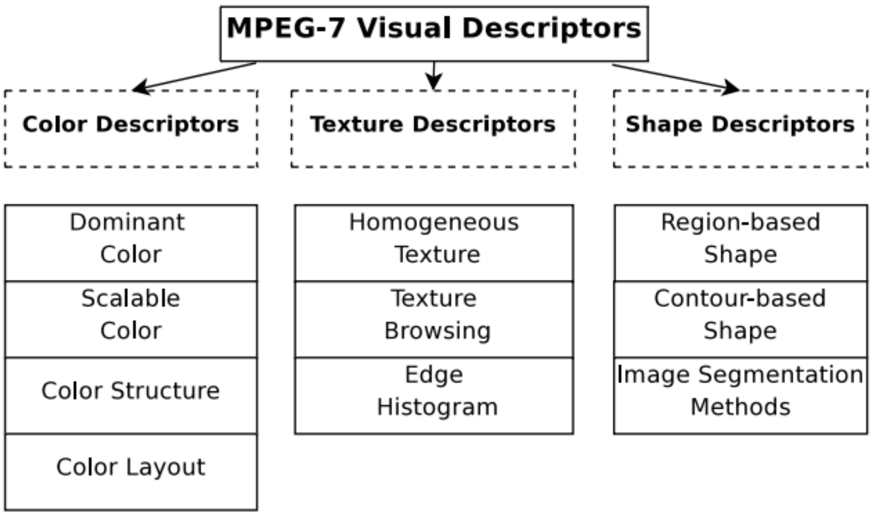


Figure 2.1: MPEG-7 Visual Descriptors For Low-level Features.

Dominant ColorDescriptor

Dominant Color Descriptor (DCD) provides an effective, compact and intuitive description of the representative colors in an image or region . The descriptor consists of the Color Index (ci), Percentage (pi), Color Variance (vi) and Spatial Coherency (s); the last two parameters are optional. Then the DCD is defined by:

$$F = (c_i, p_i, v_i), s, i = 1, ..., N \tag{2.13}$$

where N is the number of the colors and . For dominant color extraction, the generalized Lloyd algorithm is used for color clustering. There is one overall Spatial Coherency (SC) value for the whole image and several groups of (c_i, p_i, v_i) for the corresponding dominant colors. The Perceptual colors to represent images based on dominant colors 2.1

S/No.	Red	Green	Blue
Manchester United	6	4	0
Celtic	6	3	0
FC Porto	6	2	1
FC Copenhagen	6	2	1

Table 2.1: Perceptual colors to represent images based on dominant colors

DCD Extraction

The extraction procedure for the dominant color uses the Generalized Lloyd Algorithm (GLA)[1] to cluster the pixel color values. After defining the colors, for each image implement the following steps:

- Read in the image and create an image array that contains the RGB components of each pixel in the image
- For each pixel in the image do:
  - Search color table for the nearest color by finding the distance between the pixel color I represented as (P_r, P_g, P_b) and the color in the color table C_i represented as (C_{iR}, C_{iG}, C_{iB}) using the distance formula (3):

    $$C_d = (\sqrt{(P_r - C_{ir})^2 + (P_g - C_{ig})^2 + (P_b - C_{ib})^2}),$$

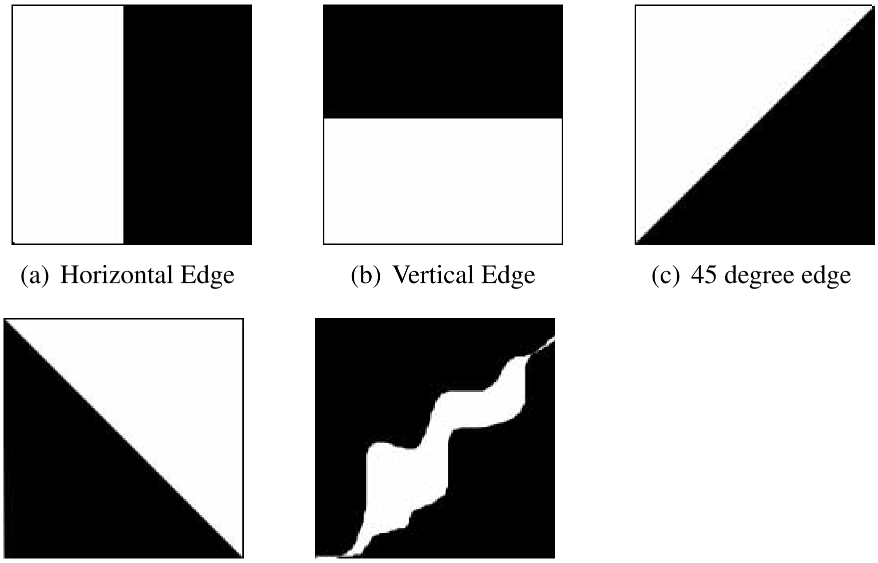
    i=1,2....18
  - Assign to the pixel the RGB entry in color table for which C_d is the minimum
- Create a frequency table for each assigned color
- Sort the frequency table in descending order MPEG-7 DCD allows at most eight colors to be represented.The highest four frequent colors are then selected with their percentages to create the description of the image.

Edge Histogram Descriptor

The EHD represents the spatial distribution of edges in an image. The extraction process of the EHD consists of the following stages:

- The edges in each image-block is categorized into one of the following six types: vertical, horizontal, 45 diagonal, 135 diagonal, nondirectional edge and no-edge. (See Figure 8)
- Now a 5-bin edge histogram of each subimage can be obtained. (See 2.2)

- Each bin value is normalized by the total number of image-blocks in the image.
- The normalized bin values are nonlinearly quantized.



(a) Horizontal Edge (b) Vertical Edge (c) 45 degree edge (d) 135 degree edge (e) non-directional edge

Figure 2.2: Five types of Edges

2.2.3 Edge and Moment Appraoch

Edge Detection

An edge in an image is a contour across which the brightness of the image changes abruptly. In image processing, an edge is often interpreted as one class of singularities. In a function, singularities can be characterized easily as discontinuities where the gradient approaches infinity. However, image data is discrete, so edges in an image often are defined as the local maxima of the gradient.
Edge detection is an important task in image processing. It is a main tool in pattern recognition, image segmentation, and scene analysis. An edge detector is basically a high- pass filter that can be applied to extract the edge points in an image.

Classical Edge Detectors

Many classical edge detectors have been developed over time. They are based on the principle of matching local image segments withspecificedgepatterns. Theedgedetection is realized by the convolution with a set of directional derivative masks.The popular edge detection operators are Roberts, Sobel, Prewitt, Frei-Chen, and Laplacian operators Creating a footnote is easy.[1] an example footnote.

Sobel Edge Detector

$$G_x = \begin{bmatrix} +1 & +2 & +1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{bmatrix} * A$$

$$G_y = \begin{bmatrix} +1 & 0 & -1 \\ +2 & 0 & -2 \\ +1 & 0 & -1 \end{bmatrix}$$

Multiscale Edge Detector

The resolution of an image is directly related to the proper scale for edge detection. High resolution and small scale will result in noisy and discontinuous edges; low resolution and large scale will result in undetected edges.Because image data is always discrete, the practical scale in images is usually integer.
The scale controls the significance of edges to be shown. Edges of higher significance aremorelikelytobekeptbythewavelettransformacross scales. Edges of lower significance are more likely to disappear when the scale increases.

Image Moments

Spatial and central moments are important statistical properties of an image.An image moment is a certain particular weighted average (moment) of the image pixels' intensities, or a function of such moments, usually chosen to have some attractive property or interpretation. Image moments are useful to describe objects after segmentation. Simple properties of the image which are found via image moments include area (or total intensity), its centroid, and information about its orientation.

For a 2-D moment of order (p+q) of a digital image f(x,y) is defined as

$$m_{pq} = \sum_x \sum_y x^p y^q f(x,y) \tag{2.14}$$

for p,q=0,1,2,....., where the summations are over the values of the spatial coordinates x

[1] An example footnote.

and y spanning the image. The corresponding *central moment* is defined as

$$\mu_{pq} = \sum_x \sum_y (x-\bar{x})^p (y-\bar{y})^q f(x,y) \tag{2.15}$$

where

$$\bar{x} = \frac{m_{10}}{m_{00}} \quad and \quad \bar{y} = \frac{m_{01}}{m_{00}}$$

The *normalized central moment* of order (p+q) is defined as

$$\eta_{pq} = \frac{\mu_{pq}}{\mu_{00}^{\gamma}}$$

for p,q=0,1,2,...., where

$$\gamma = \frac{p+q}{2} + 1$$

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

A set of seven 2-D *moment invariants* that are insensitive to translation,scale change , mirroring and rotation can be derived from these equations. They are

$$\begin{aligned}
\phi_1 &= \eta_{20} + \eta_{02} && (2.16)\\
\phi_2 &= (\eta_{20} - eta_{02})^2 + 4\eta_{11}^2 && (2.17)\\
\phi_3 &= (\eta_{30} - 3\eta_{12})^2 + (3\eta_{21} - \eta_{03})^2 && (2.18)\\
\phi_4 &= (\eta_{30} + \eta_{12})^2 + (\eta_{21} + \eta_{03})^2 && (2.19)\\
\phi_5 &= (\eta_{30} - 3\eta_{12})(\eta_{30} + \eta_{12}) && \\
&\quad [(\eta_{30} + \eta_{12})^2 - 3(\eta_{21} + \eta_{03})^2] + && \\
&\quad (3\eta_{21} - \eta_{03}(\eta_{21} + \eta_{03}) && \\
&\quad [3(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03}^2] && (2.20)\\
\phi_6 &= (\eta_{20} - \eta_{02})[(\eta_{30} + \eta_{12}^2 - (\eta_{21} + \eta_{03})^2] && \\
&\quad + 4\eta_{11}(\eta_{30} + \eta_{12})(\eta_2 1 + \eta_{03}) && (2.21)\\
\phi_7 &= (3\eta_{21} - \eta_{03})(\eta_{30} + \eta_{12})[(\eta_{30} + \eta_{12})^2 && \\
&\quad - 3(\eta_{21} + \eta_{03})^2] + (e\eta_{12} = \eta_{30})(\eta_{21} + \eta_{03}) && \\
&\quad [3(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2] && (2.22)
\end{aligned}$$

These sets of equations are commonly referred as Hu set of invariant moments.[2]

Procedure

Multiscale Edge detection since sobel is just not good enough Process is started with the multiscale edge detection . Once the edge image is computed. we compute the normalized central moments upto order five and the translation,rotation and scale invariant based on the grayscale edge image using the definitions.A feature vector containing these 21+7=28 moments is computed and stored in the training database.[5]

2.3 SVM(Support Vector Machine)

The central idea of SVM is the adjustment of a discriminating function so that it optimally uses the separability information of the boundary cases. Given a set of cases which belong to one of two classes, training a linear SVM consists in searching for the hyperplane that leaves the largest possible number of cases of the same class on the same side, while maximizing the distance of either class from the hyperplane. If the training set is linearly separable, then a discriminant hyperplane will satisfy the inequalities:

$$v_i(w.x_i + b) \geq 1 \tag{2.23}$$

where $x_i \in \Re^d$ is a vector of the training set, d being the dimension of the input space, and $y_i \in \{-1, +1\}$ is the corresponding class. Among the separating hyperplanes, the SVM approach selects the one for which the distance to the closest point is maximal. Since such a distance is $1/||w||$, finding the hyperplane is equivalent to minimizing $||w||^2$ under constraints (1). The points closest to the hyperplane are called Support Vectors, and the quantity $2/||w||$ is called the margin (see Figure 2); it can be considered a measure of the generalization ability of the SVM: the larger the margin, the better the generalization is expected to be.

Bibliography

[1] A. Gersho and R.M. Grey. *Vector Quantization and Signal Compression.* Dover, Massachusetts, 1991.

[2] Ming K. Hu. Visual pattern recognition by moment invariants. *IRE Transactions on Information Theory*, IT-8:179–187, February 1962.

[3] Michael J. Jones and James M. Rehg. Statistical color models with application to skin detection. *International Journal of Computer Vision*, 46(1):81–96, January 2002.

[4] Jos Mara Martnez Sanchez. Mpeg-7: Overview of mpeg-7 description tools, part 2. *IEEE MultiMedia*, 9(3):83–93, 2002.

[5] James Z. Wang, Gio Wiederhold, and Oscar Firschein. System for screening objectionable images using daubechies' wavelets and color histograms. In *IDMS '97: Proceedings of the 4th International Workshop on Interactive Distributed Multimedia Systems and Telecommunication Services*, pages 20–30, London, UK, 1997. Springer-Verlag.