Journal of Theoretical and Applied Information Technology
15th January 2012. Vol. 35 No.1

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

SKIN DETECTION IN PORNOGRAPHIC VIDEOS USING THRESHOLD TECHNIQUE

[1]HAJAR BOUIROUGA, [2]SANAA EL FKIHI, [3]ABDEILAH JILBAB, [1]DRISS ABOUTAJDINE

[1]LRIT, Unité associée au CNRST, URAC 29, FSR, Université Mohammed V agdal, Rabat

[2]ENSIASS, Université Mohammed V Souissi, Rabat

[3]ENSET, Université Mohammed V Souissi, Rabat

E-mail:[1]bouirouga_hajar@yahoo.fr

ABSTRACT

Adult classification of images and videos is one of the major tasks for semantic analysis of visual content. A modern approach to this problem is introducing a mechanism to prevent objectionable access to this type of content. In this paper, we propose a real-time system that can classify input videos into adult or non-adult videos. First, vidéos were filtered by skin color model using threshold technique. Then, the output videos are classified by neural network for taken the decision. We notice that the detection of an adult video is based on the detection of the adult images that compose the considered video. The simulation shows that this system achieved 90% of the true rate.

Keywords: *Skin Detection, Probabilistic Models, Neuron Networks, Pornographic Images, Descriptors, Video Filtering.*

1. INTRODUCTION

Detecting objectionable images from websites has been a significant problem in content-based filtering systems. Hence, the use of image analysis techniques to recognize adult images recently has gained much attention. Researchers have developed different algorithms to detect the adult images. James Ze Wang and al. developed the WIPE (Wavelet Image Pornography Elimination) system to block objectionable images from websites [1]. D. A. Forsyth proposed a body plan that marshalled different parts of the body by their geometric constraints for detecting the naked people [2, 3, 4]. Based on the discrete probability distributions obtained from three dimensional skin and non-skin color histograms, Jones and Rehg used five simple features to describe adult images [5]. Soriano treated the detected skin regions as blobs for extracting features representing pornographic images [6]. To some extent, these algorithms can recognize some kinds of pornographic images successfully. However, extracting reliable features [8] that can describe pornographic images has been an open problem. This paper presents a new adult video detection method. It is based on compute vision and pattern recognition algorithms. Images were filtered by skin color model at first, then, they were classified by neural networks.

Our work proposes a method for identifying video adult. The proposed algorithm includes the detection of the human skin based on the colour and motion information. Besides, the proposal takes into account the object motion, skin and non skin areas as well. Experimental results demonstrate the successfulness of the algorithm used and its capability in recognition of adult video.

A brief system overview is given in section2. In section 3, we put forward a subtraction of the background. In section 4 we briefly introduce the skin color model and in section 5, we will talk about the features extraction and its application in adult video detection. In section 6 we present a neural networks algorithm At last, the experiments and the conclusion are given.

2. SYSTEM FRAMEWORK

As illustrated in Figure 1, the system contains four parts. First a motion detector detects mobile objects evolving in the scene. Second, the skin detector can be used as the basis for adult videos detector. Finally, we use images, which are manually classified into adult and non-adult sets to

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

train a neural networks classifier. The neural networks classifier outputs a number Op between 0 and 1 with 1 signifying an adult video.

After extraction of sequence information of the video filtering systems use various algorithms to classify videos as a pornographic video or non-pornographic. Most of these systems have sensitivity settings that allow the user to have a degree of control to adjust the classifier's decision concerning the nature of video processing.

The network consists of a layer of input indexed by i, a hidden layer, indexed by j and output layer indexed by k. In our application, we use 53 nodes in hidden layer, 1 node in output layer.

The number of hidden nodes is chosen empirically (depending on the test case). Each node is associated with a hidden activation function which is a function sigmoïdal corresponding to the entry node. The links between input layer and output are associated with weights denoted by wij while the links between the hidden layer and output are associated with weights denoted by wkj.

The learning process starts with a sequence of random weights. The latter is decomposed into two procedures:

* A forward procedure (before restoration), where each learning pattern p is fed into the network to evaluate the output op.

* A procedure for back propagation (back propagation) in which the error of the output Ep requires changes.

For an input pattern p, the output OP is a real number between 0 and 1. The higher the number is to 1, the more reason to come is a scene adult. Thereafter, it establishes a threshold T, 0 <T <1, for a binary decision.

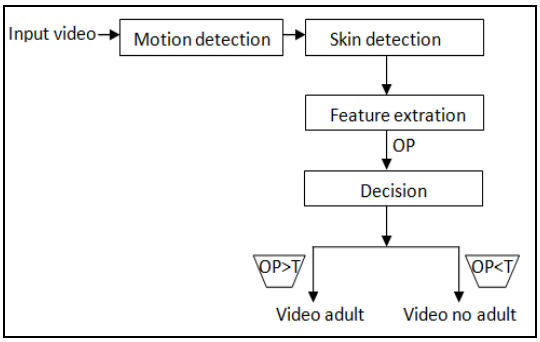

Figure 1. Overview of the proposed adult video filter

3. MOTION DETECTION

Motion detection is very valuable in many vision problems, such as surveillance, traffic monitoring, semantic object segmentation and content-based video retrieval. Although there are a lot of literatures addressing issues about motion detection, it is still a challenging problem.

The simplest method of motion detection is based on the difference of frame to frame [22, 23], including the difference of pixel, edge and area. Since the difference between two frames is sensitive to the noises, the false positives cannot be suppressed effectively [24, 25, and 26]. These methods are more consistent than difference-based methods. However, they cannot be used in real-time application due to their computational complexity.

Some learning-based approaches were also proposed for motion detection, such as in [27, 28]. A learned probability distribution of intensity at each pixel is useful to suppress false positives. But it is difficult to describe motions with one probability distribution model [29, 30], since motions in video are very complex and diverse. Although these approaches are able to detect some specific motion patterns, how to select suitable slices becomes its main obstacle, because a slice only presents a part of motion information. Recently, some research works try to improve traditional methods in many aspects. For example, [31] focuses on detecting shadows of motion object to improving the performance of motion objects detection and segmentation.

Background subtraction is used for finding moving objects in video [32, 33, 34, 35]. This approach provides a more complete set of data describing the characteristics of target motion relative to other approaches to motion detection [36]. Conditionally, the background scene and the viewing sensor are required to be stationary when background subtraction is applied. Usually a motion model of the background is assumed and motion parameters are estimated. Then, the background is ideally recorded and the foreground can be detected pixel by pixel. In this paper, we propose a (SB) model, which is a temporal and spatial description of the background. The foreground detection is carried out based on the SB model. The proposed approach is robust even with approximate motion compensation, noise or environmental changes and is able to detect small moving objects in a highly textured background objects in a highly textured background.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

In our work we suggest using the algorithm for subtraction background this method comprises two major steps: initialization and extraction of motion (foreground).

The first step is to model the background from the first N images (N ≈ 30) to the video sequence. An average intensity is calculated from these images for each pixel and for each channel (R, G and B). The average intensity of a given pixel is then summarized in the following equation:

$$\mu_c(x, y) = \frac{1}{N}\sum_{i=0}^{N} I_{i,c}(x, y) \qquad (1)$$

Where Ii is the ith image initialization, N the amount of images used and c the selected channel. The next step is to calculate a standard deviation for each pixel (for each channel) to be used as the threshold of detection. This operation usually requires the storage of N first images. However, two accumulators are used, S (x, y) to store the sum of intensities of the pixels and SC (x, y) for storing the sum of the squares. Standard deviations can then be calculated using the following equation. It is also interesting to note that S (x, y) can be used to calculate the average, which avoids additional and unnecessary operations.

$$\sigma_c(x,y) = \sqrt{\left(\frac{SC_c(x,y)}{N}\right) - \left(\frac{S_c(x,y)}{N}\right)^2} \qquad (2)$$

4. SKIN DETECTION

Skin detection is the process of finding skin color pixels and regions in an image or a video. This process is generally used as a pretreatment step to find areas that may have human faces and limbs in the images. A sensor skin generally becomes a pixel in a color space appropriate, then use a classifier to label skin pixel if it is a skin pixel or a non-skin. A skin classifier defines a decision limit of the class of skin color in the color space based on a database of training pixels of skin color.

4.1. Related work

In current times various methods of identifying skin inside images have been developed. Soriano et al [6] proposed a face tracking method that is adaptive to varying illumination color. Cho et al [7] presented a new skin color filter that is useful of finding skin regions in images. Yang et al [9] recommended a real-time face tracker using an adaptive skin color model. Jedynak et al [10] suggested a sequence of three models for skin detection built from a large collection of labeled images. Jones and Rehg [11] compared a Bayesian approach to a mixture of Gaussian in RGB space. Storring and Andersen [12] presented a model for detecting human skin under changing lighting conditions. While Caetano et al [13] compared between single model and several Gaussian mixture models used to improve skin detection. This paper presents the impact of adjusting the threshold value in the chromatic skin color model for improving skin detection in videos that contain luminance.

There are many color spaces have been used in earlier works of skin detection, such as RGB, normalized RGB, YCbCr, HIS and TSL [15]. Although the RGB color space is one of the most color spaces used for processing and storing digital images, it is not widely used in detection algorithms because the skin chrominance and luminance components are mixed. Normalized RGB and YCbCr are often used by techniques for detecting the skin. Some work has been performed to compare performance space skin color detection problems of the skin [14]. The conclusion was that the standard color space gives the best results of skin detection.

4.2 Color Representations

In the past, different color spaces have been used in skin segmentation. In some cases, the color grading is done using only pixel chrominance because it is expected that the segmentation of the skin may become more robust to lighting variations if pixel luminance is rejected. In this paper, we examine how the choice of color space and use of chrominance channels affect the skin segmentation. It should be noted that there are many color spaces, but many of them share similar characteristics. Thus, in this study, we focus on the tree representing the color spaces that are commonly used in image processing [37]:

The RGB color space is one of the color spaces widely used for storing and processing digital image.

However, the RGB color space alone is unreliable for identifying pixels of skin color, because 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 skin regions. Chromatic colors are more reliable and are obtained by eliminating luminance using some form of transformation. In RGB space, each color appears in its primary spectral component red, green and blue. The RGB True Color is an additive color

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

system based on tri-chromatic theory. It is one of the most commonly used color spaces, with a lot of research activities being based on it. Therefore, skin color is classified by heuristic rules that take into account two different conditions: uniform daylight or lateral illumination. The color of the skin to sunlight rule uniform illumination is defined as [16]:

$$\begin{cases} (R > 95) \text{et} (G > 40) \text{et} (B > 20) \text{ AND} \\ ((\max[R, G, B]) - (\min[R, G, B])) > 15) \text{ AND} \\ (\text{ABS}(R - G) > 15) \text{ AND } (R > G) \text{ et } (R > B) \end{cases} \quad (3)$$

While the skin color under flashlight or daylight lateral illumination rule is given by [16]:

$$\begin{cases} (R > b) \text{ et } (G > B)) \text{ OR} \\ ((R > 220) (G > 210) \text{ OR} \\ (B > 170) \text{ AND } (\text{ABS}(R - G) <= 15)) \end{cases} \quad (4)$$

RGB values are transformed into YCBCR values using the formulation [17]:

$$Y = 0.299\ R + 0.587\ G + 0.114\ B \quad (5)$$

The other two components of this space represent the color information and are calculated from Luma:

$$\begin{cases} Cr = 128 + 0.5R - 0.418688\,G - 0.081312\,B \\ Cb = 128 - 0.168736\,R - 0.331264 + 0.5B \end{cases} \quad (6)$$

YCbCr is a family of color spaces used in video systems and digital photography. Y is the luminance component and Cb and Cr are the chrominance components of blue and red. The simplicity of processing and explicit separation of luminance components and chrominance makes this area attractive color for modeling the skin color. In this color space, a pixel is classified as skin if the following conditions met:

$$Cr \leq 1.5862 * Cb + 20 \quad (7)$$

$$Cr \geq 0.3448 * Cb + 76.2069 \quad (8)$$

$$Cr \geq -4.5652 * Cb + 234.5652 \quad (9)$$

$$Cr \leq -1.15 * Cb + 301.75 \quad (10)$$

$$Cr \leq -2.2857 * Cb + 432.85 \quad (11)$$

The HSV model space consists in breaking the color according to physiological criteria (hue, saturation and luminance). In HSV space, the intensity information is represented through the V, for this reason, this channel should be overlooked in the process of detection of the skin, we consider only the channels H and S represent the chromatic information.

Hue-Saturation-Value (HSV) space is also a popular color space because it is based on human perception of colors. The intensity, the value and lightness are related to the luminance color. Hue is generally related to the wavelength of light. Saturation is an element that measures the colorfulness in HSV space. The intuition of the color space component and explicit discrimination between the luminance and chrominance properties of these spaces is popular colors.

$$H = \begin{cases} 2 - \cos^{-1}\left\{ \dfrac{[(R-G)+(R-B)]}{2\sqrt{(R-G)^2+(R-B)(G-B)}} \right., B > G \\ \cos^{-1}\left\{ \dfrac{[(R-G)+(R-B)]}{2\sqrt{(R-G)^2+(R-G)(G-B)}} \right., otherwise \end{cases} \quad (12)$$

$$S = 1 - \frac{3*\min(R,G,B)}{I} \quad ; \quad V = \max(R,G,B)$$

(H, S, V) is defined as skin if:

$$25 < H < 50 \quad (13) \quad \text{AND} \quad 0.23 < S < 0.68 \quad (14)$$

The Main target of this paper is introducing new color transform from viewpoint of convex constraint programming. Skin detection is used as a benchmark problem for the proposed algorithm. In the New color space, the skin and non-skin classes are well separated.

Which is the space of color best to use?

To answer on this question, we propose different combinations of existing color space. A set of rules is bounding from all three color spaces, RGB, YCbCr and HSV, based on our observations of training. All rules are derived for the intensity values between 0 and 255 [18].

First we combine the two conditions (3) and (4) by a logical OR, we obtain the rule A

RuleA: Equation (3) ∪Equation (4) (15)

Then we combine the conditions (7) to (11) by a logical AND, we get the rule B

Rule B: Equation (7) ∩ Equation (8) ∩ Equation (9) ∩ Equation (10) ∩ Equation (11) (16)

Finally we combine the conditions (13) and (14) by a logical OR, we get the rule C

Rule C: Equation (13) ∪Equation (14) (17)

For the models containing two color space For example, Rgb-YCbCr, Rgb-Hsv and Hsv YCbCr. A set of rules is bounding from all three

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

rules A, B and C by using a logical AND each pixel which makes the rule A ∩ B, A ∩ C and C ∩ B respectively is classified as a skin color pixel.

5. FEATURE EXTRACTION

In this section we present functions based on grouping of skin regions [19] which could distinguish the adult images of the other images.

Many of these features are based on suitable ellipses [20] calculated on the skin map. These functions are adapted to our demand for their simplicity.

Consequently we calculate for each card skin two ellipses namely Suitable Global Ellipse (GFE) calculated on the card and the entire skin Suitable Local Ellipse (LFE) based only on the largest region on the map skin.

We distinguish 9 functions of the skin map 3 first functions are global namely [20]: large area of the skin map.

- The average probability of skin of the entire image.

$$Moyenne\ -skn = \frac{\sum proba\ (masque\ )}{longueur\ *l\arg ueur} \quad (18)$$

- The average probability of skin inside the GFE.
- The number of areas of skin in the image.

This function is simply the number of gouts of skin in the image. For that we allocate the same number (exp: 1, 2, 3) to pixels belonging to the same region. Afterward, we count the number of regions which exists in the image.

Later, we extract the biggest region and we calculate the following six descriptors:

We calculate the central line r and the central column c by the following formula:

$$\begin{cases} c = \frac{1}{A}\sum_{s\in R} c_s y_s \\ r = \frac{1}{A}\sum_{s\in R} r_s y_s \end{cases} ; \quad A = \sum_{s\in R} y_s \qquad (19)$$

rs and cs are the line and the column braided by the pixel which belongs has all the pixels S of the image at the level of grey and ys the value of this pixel in the interval [0; 1].

Then we calculate the variation between lines, between columns and between lines and columns.

$$\begin{aligned} \mu_{rr} &= \frac{1}{A}\sum_{s\in R}(r_s - r)^2 y_s \\ \mu_{cc} &= \frac{1}{A}\sum_{s\in R}(c_s - c)^2 y_s \qquad (20) \\ \mu_{rc} &= \frac{1}{A}\sum_{s\in R}(r_s - r)(c_s - c) y_s \end{aligned}$$

Then, we calculate 'delta' which is equal to the product of the variations between lines, columns and between both.

$$delta = \mu_{rr}\mu_{cc} - \mu_{rc}^2 \qquad (21)$$

If delta is strictly upper to 0 this means that there is an ellipse otherwise no ellipse is detected.

- Distance from the larger area of skin at the center of the image.

The following formula allows calculating the distance between the center of the biggest region of the skin and the center of the image.

$$Fdist=((r-long2.0)*(r-long2.0)+(c-l\arg 2.0)*(c-l\arg r/2.0))$$

with :long=length and Larg=width

$$Fdist=\sqrt{Fdist}$$

$$Fdist = \frac{Fdist}{\sqrt{l\arg ueur * l\arg eur + longueur * longueur}} \quad (22)$$

with r is the line which coordinates the center of the biggest region of the skin and c is the column which coordinates the center of the biggest region of the skin.

- The angle of the main axis of the LFE of horizontal axis.

We calculate the angle between axes horizontal and vertical of the ellipse. There are four points of law for the calculation of the length of the major axis and that of the minor axis and the angle.

- $\mu_{rc} = 0$ And $\mu_{rr} \geq \mu_{cc}$.

$$\begin{cases} Lmajeur = 4\sqrt{\mu_{rr}} \\ L\ \min\ eur = 4\sqrt{\mu_{rc}} \\ \theta = -90\ ° \end{cases} \qquad (23)$$

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

- $\mu_{rc} = 0$ And $\mu_{rr} \leq \mu_{cc}$.

$$\begin{cases} Lmajeur = 4\sqrt{\mu cc} \\ L\ \min\ eur = 4\sqrt{\mu_{rr}} \\ \theta = -0° \end{cases} \quad (24)$$

- $\mu_{rc} \neq 0$ And $\mu_{rr} \leq \mu_{cc}$.

$$\begin{cases} Lmajeur = \sqrt{8((\mu_{rr}+\mu_{cc}) + \sqrt{4\mu_{rc}{}^2 + (\mu_{rr}-\mu_{cc})^2}} \\ L\min eur = \sqrt{8((\mu_{rr}+\mu_{cc}) - \sqrt{4\mu_{rc}{}^2 + (\mu_{rr}-\mu_{cc})^2}} \\ \theta = \arctan\dfrac{-2\mu_{rc}}{(\mu_{cc}-\mu_{rr}) + \sqrt{4\mu_{rc}{}^2 + (\mu_{rr}-\mu_{cc})^2}} \end{cases} \quad (25)$$

- $\mu_{rc} \neq 0$ And $\mu_{rr} \succ \mu_{cc}$.

$$\begin{cases} Lmajeur = \sqrt{8((\mu_{rr}+\mu_{cc}) + \sqrt{4\mu_{rc}{}^2 + (\mu_{rr}-\mu_{cc})^2}} \\ L\min eur = \sqrt{8((\mu_{rr}+\mu_{cc}) - \sqrt{4\mu_{rc}{}^2 + (\mu_{rr}-\mu_{cc})^2}} \\ \theta = \arctan\dfrac{(\mu_{rr}-\mu_{cc}) + \sqrt{4\mu_{rc}{}^2 + (\mu_{rr}-\mu_{cc})^2}}{-2\mu_{rc}} \end{cases} \quad (26)$$

- The proportion of the secondary axis and the main axis of the LFE. The calculation of this proportion is made by the following relation:

$$fratl = \frac{Lmajeur}{L\ \min\ eur} \quad (27)$$

With Lmajeur: the length of the major axis and Lmineur: the length of the minor axis.

- The proportion of the sector LFE and the image. This proportion is defined by the following formula:

$$frata = \frac{farea}{longueur\ * l\arg ueur} \quad (28)$$

$$farea = 4.0 * \Pi * \sqrt{delta} \quad (29)$$

- The average probability of skin inside the LFE.

The average probability of skin outside the LFE.

After this step we adapt neural networks to classify videos. More specifically, the classifier will act on the vector constructed from the calculated descriptors in the eighth paragraph to decide what kind of video analysis.

6. DECISION

The techniques of classification of image which use statistical methods are improved a lot [38]. Thus, many research groups to study and research in the field of image classification by the technique of mining on the image [39]. The classification of images can be classified as the neural network model of the decision tree and Support Vector Machine. Neural Network based method is the most common technique. This method focuses on the study of the surface of the decision say the limit for adult images from non-adult images by computer classification rule, called perceptron [42, 43]. An artificial neural network (ANN) is a paradigm of information processing inspired by biological nervous systems, such as information processing in the brain. The key element of this paradigm is the new structure of system information processing It consists of a large number of highly interconnected processing elements (neurons) working in unison to solve specific problems. ANNs, like people, learn by examples. An ANN is configured for a specific application, such as pattern recognition or data classification, through a learning process.

Learning in biological systems involves adjustments to the synaptic connections between the model decision tree recursively partitions neurons.The space image data, using variables that can divide the image data to most identical numbers from a given number of variables. This technique can yield incredible results when characteristics and image data are known in advance [41]. The technique of support vector machine is a new method of image classification. The aim of the method is to find the decision lines or surfaces to distinguish the data from other like technique using neural networks. The technique using neural networks is just to find a classification decision surfaces of training data. However, SVM is to find decision surfaces to maximize the distance of two sets. Jiao et al. classifiers tested on adult images using SVM [40].

We employ the Neural Network for the classification module in this paper.

Inputs for the neural network are fed according to the values extracted from descriptors. Since the various descriptors can represent the specific features of a given image, the proper evaluation process should be required to choose the best one for the adult image classification.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

6.1 The structure of the neural network

To study the neural networks you must first become aware of their structure. A neural network is composed of several different elements. Neurons are the most basic unit. The neurons are interconnected. These connections are not equal, that each connection has a weight of connection. Network groups come together to form layers. In this section we will explore each of these topics.

- The Neuron

The neuron is the basic element of the neural network. A neuron is a channel of communication that both accepts input and produces output. Similarly, the neuron sends its output to other neurons [44]. When a neuron produces an output, the neuron is said to activate, or fire. Consider a neuron that is connected to a number of other neurons. The variable w represents the weight between neuron k and the other neuron. We say that this neuron is connected to other neurons k. The variable x represents the input of each neuron to other neurons. Therefore we must calculate the sum of all inputs x multiplied by the corresponding weight w. This is illustrated in the following equation [44].

$$u = \frac{1}{A}\sum_{s\in R} W_k x_k \qquad (30)$$

- Neuron Connection Weights

The previous section already mentioned that the neurons are usually connected. These connections are not equal and can be assigned individual weights. These weights are what gives the ability of the neural network to recognize certain patterns. Adjust the weight and the neural network will recognize a different model [44].

- Neuron Layers

Neurons are usually organized into layers. Layers are groups of neurons that perform similar functions. There are three types of layers. The input layer is the layer of neurons that receive input from the user program. The layer of neurons that send data to the user program is the output layer. Between the input layer and output layer can be zero or more hidden layers. Hidden layer neurons are connected only to other neurons and never directly interact with the user program [44]. Layers of input and output are not just there as interface points. Each neuron in a neural network has the potential to influence treatment. Treatment may occur at any level in the neural network. Not all networks of neurons that has many layers. The hidden layer is optional. Layers of input and output are needed, but it is possible to have a layer act as both an input layer and output. Now that you've seen how a neural network is built, it will be shown how neural networks are used in pattern recognition

6.2 Pattern recognition

The feature extraction steps described in the previous subsection produce a feature vector for each video.

The task is to find the decision rule on this feature vector which optimally separates the images from those of adults do not. Testimony of [6] shows that the MLP classifier provides a significant performance compared to several other approaches such as generalized linear model, the classifier of k nearest neighbor and support vector machine. Our classifier is an MLP net semilinear layers with one hidden layer as in [14]. This network generates a number between 0 and 1, with 1 being the image of adults and 0 not.

The learning process starts with a set of random values of weights. For each pattern p of training, the network evaluates the output of an op order in advance. To decrease the error between the output and the op tp real target, the network calculates the correction weight values using the backpropagation procedure. This procedure is repeated for all models in the training set to give the corrections that result for all weights in this iteration. In a year of successful learning, the system error decreases with the number of iterations, and the procedure converges to a stable set of weights, which show only small fluctuations in value that further learning is attempted. In the test phase for each test pattern, the network calculates the output in one pass. We then set a threshold T, 0 <T <1 to obtain the binary decision.

7. EXPERIMENTS

We conduct two experiments in performance evaluation: one for the detection of skin and one for the classification of videos. Two performance measures used are true positive (TP) and false positives (FP). In the detection of skin, TP is defined as the ratio between the number of ground truth skin pixels detected to the total number of pixels of the skin and FP is the ratio between the number of pixels misclassified as non-skin pixels for skin the total number of non-skin pixels. In the classification of shapes, TP is defined as the ratio

Journal of Theoretical and Applied Information Technology
15th January 2012. Vol. 35 No.1

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

between the number of adult videos ground truth identified for the total number of adult videos and FP is the ratio between the number of videos not misclassified as adults adult videos and the total number of non-adult videos.

In skin detection evaluation, we use 200 videos for training and 100 adult videos for test. Performance comparison between the different color spaces is shown in Figure4 and Figure5.

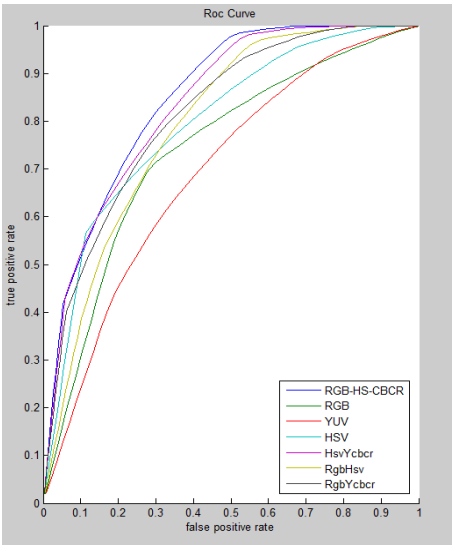

Figure 2. ROC curves for different color spaces without extraction background

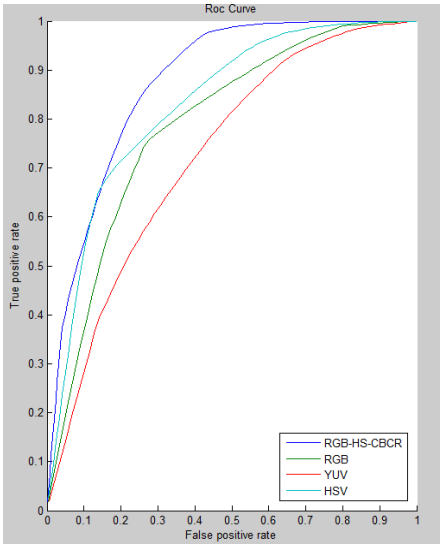

Figure 3. ROC curves for for HSV, RGB, YCbCr and RgbHsvYCbCr color spaces with extraction background

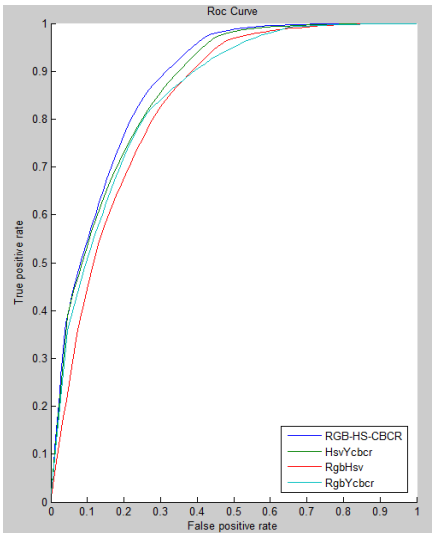

Figure 4. ROC curves for HsvYCbCr, RgbHsv, RgbYCbCr and RgbHsvYCbCr color spaces with extraction background

From Figure 2,3 and Figure 4, we can see that combination of different color space generally provide better classification results than using only single color space.

As a comparison, we also list the performance of corresponding color space with extraction background.

Some example frames after skin detection are shown in Figure 5 with and without extraction background.

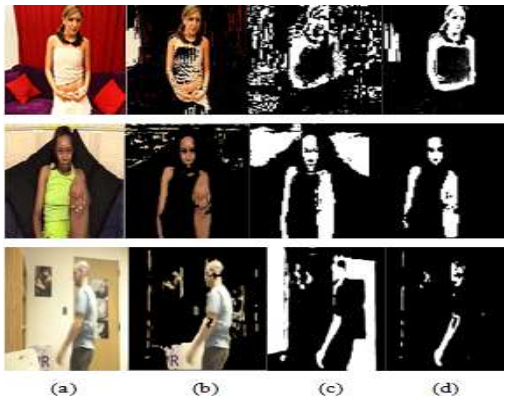

Figure 5. Results of modeling the distribution of skin after extraction background a) Skin Detector Input, b) background extraction, c) Skin detection without background extraction, d)Skin detection with background extraction

Journal of Theoretical and Applied Information Technology
15th January 2012. Vol. 35 No.1

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

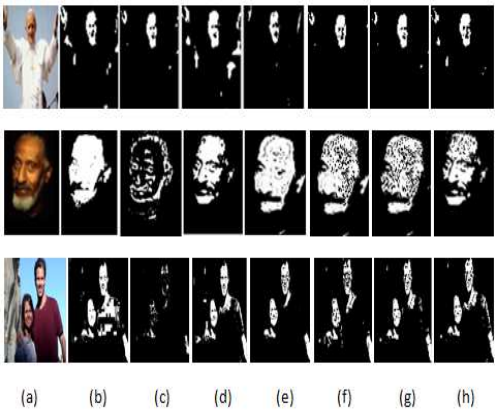

Figure 6. Results of modeling the distribution of skin according to different models with extraction background a) Skin Detector Input, b) RGB model, c) Model YCbCr, d) Model HSV, e) RGB-H-CbCr Skin Color Model, f) HsvYcbcr model, g) Model RgbYCbCr, h) Model RgbHSV

The output of skin detection is a map showing the probability of skin pixels. With the normalization we get a grayscale image on the same grid as the input image and the gray levels proportional to the likelihood of skin. This is called the skin map in this document. Some results of skin detection RGB-YCbCr-H are shown in Figure 6. Most areas of the skin are detected correctly. However, there are also false alarms when certain objects with similar skin color to appear in the background.

The objective is to show that for all color spaces their corresponding optimum skin detectors. In this section some definitions are provided which are useful to proof our statement.

Definition: Let I be an image, P a pixel of the image I and Xp a vector containing the values of the pixel P in the color space C. A function D (Xp) is called a skin color detector in C if:

$$D(Xp) = \begin{cases} 1 & \textit{if Xp is considered skin} \\ 0 & \textit{if Xp is considered no skin} \end{cases} \quad (31)$$

Definition: Given a set of images H and a skin color Detector D(Xp) the performance of D(Xp) can be evaluated with the following two parameters:

$$\text{Detection rate} = \frac{\text{Nbr of segment detected}}{\text{Actual nbr of segment}} \quad (32)$$

$$\text{False detection rate} = \frac{\text{Nbr of false detections}}{\text{Nbr of detection}} \quad (33)$$

Definition: A skin color detector D (Xp) is said to be optimum when it has the highest detection rate for a given fixed false alarm.

Figure 2,3 and 4 shows the ROC (Receiver Operating Characteristic) curves without and with extraction background obtained for each skin detector.

The results in the curve of ROC Figure 3 and 4 show that for a fixed false given alarm FP=0.3 the highest rate TP of detection is plotted on the following table:

Space color	Detection rate without extraction background	Detection rate with extraction background
RGB	72%	78%
YCbCr	59%	64%
HSV	74%	80%
RGB-HCbCr	83%	91%
RGB-HSV	73%	84%
RGB-YCbCr	78%	85%
HSV-YCbCr	79%	88%

Table 1. Results detector skin (Test carried out on 200 videos)

The best rate on the other hand was obtained by the space RGB-HS-CbCr that is 91 % while the lowest score is obtained by the space YCbCr 64 %

- Convex envelope

It is formed by the curves which, at the moment or in the other one, have no curve above of them.

Curves situated on this envelope correspond to the models which are potentially the most successful for a given matrix of cost.

The models which never participate in this envelope can be eliminated.

In our example, the convex envelope is formed by the curves of RGB-HS-CBCR and HsvYCbCr.

The receiver operating characteristics (ROC) curve of the proposed system is shown in Figure. 7. It is noted that the results meet the requirements for the practical application of a naked image detection system, The detection rates the naked videos for

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Rgb-H-CbCr and the HsvYCbCr are 95%, and 90% respectively (Tab2).

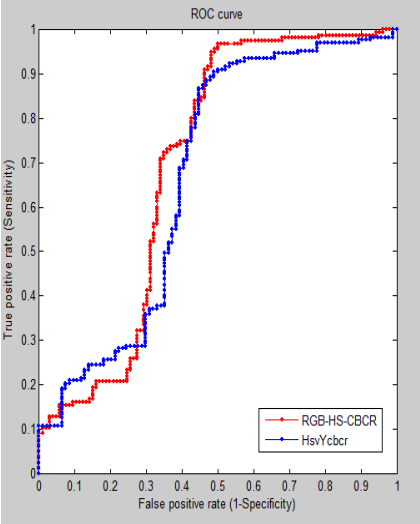

Figure 7. The ROC curve of the RGB-H-YCbCr and the HsvYCbCr adult image detector

On the basis of 200 videos (100 pornographic and 100 not pornographic) we made tests to compare the rate of performance and the medium time of treatment of every method.

A practical system is implemented using MATLAB 7.0 with an Intel® Core™2 Duo processor, Pentium IV1, Microsoft Windows Vista with 2 GB memories.

Color Space	Detection rate	Total CPU time (in seconds)	Total real time detection(in seconds)
RGB-H-YCbCr	97%	558.12	555.62
HsvYCbCr	96%	445.03	423.85

Table 2. Results adult video detector (Test carried out on 100 pornographic and 100 not pornographic)

Testing, have concluded that the method HsvYCbCr is the most effective if we want a rate of acceptable performance while respecting a rather fast processing time. This choice is based on the tests which we made and the found results.

8. CONCLUSION

We are interested in designing a method for screening adult video scenes. To do this, we followed the chain of the following: development of an algorithm for skin detection in image sequences, calculation of descriptors related to adult video scenes. For the first step, we then tested different models for the second step, we calculated nine descriptors on the skin regions detected.

These allow the region to generate the largest and the representation of an ellipse. Finally, the last step, we treated the nine descriptors for a neural network to decide whether a video or not adults.

We found that the model HsvYCbCr gave the best results for still images. Knowing that the filter uses the video images according to the stage under consideration, the percentage has been improved by a vote on various decisions. We have achieved a success rate of 90% for filtering the scenes video.

Then in the next work we can use a new method from the feature porno-sounds recognition is proposed to detect adult video sequences automatically which serves as a complementary approach to the recognition method from images.

REFRENCES:

[1] [1] J. Z. Wang, J. Li, G. Wiederhold and O. Firschein (1998). System for sacreening objectionable images. Computer Communiation Joural, 21(15):1355-1360.

[2] M. Fleck, D. A. Forsyth and C. Bregler (1996). Finding naked people. ECCV, Vol.2, 592-602.

[3] D. A. Forsyth and M. M. Fleck (1997). Body plan. CVPR, 678-863.

[4] D. A. Forsyth and M. M. Fleck (1999). Automatic detection of human nudes. IJCV, 32(1):63-77.

[5] M. J. Jones and J. M. Rehg (1999). Statistical color models with application to skin detection. CVPR, 247-280.

[6] Soriano, m., Huovinen, s., martinkauppi, and laaksonen, (2000) skin detection in video under changing illumination conditions, In Proc. 15th international Conference on pattern Recognition,vol. 1, pp. 839 –842

[7] Cho, k.M., jang, J.H., and Hong , K.S. (2001) Adaptive skin color filter",Pattern recognition,34 (5), pp:1067-1073.

[8] Jinfeng Yang, Zhouyu Fu, Tieniu Tan, Weiming Hu, (2004). A Novel Approach to Detecting Adult Images, icpr, vol. 4, pp.479-482, 17th International Conference on Pattern Recognition (ICPR'04) - Volume 4.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

[9] Yang, J., Lu, W., and Waibel, A. (1998) Skin-color modeling and adaptation. In Chin, R. and Pong, T.C., editors, 3rd Asian Conf. on Computer Vision, volume 1352 of LNCS, pp: 687-694.

[10] B. Jedynak, H. Zheng, M. Daoudi, and D. Barret (2002) Maximum entropy models for skin detection. Technical Report publication IRMA,Volume 57,number XIII, universite' des Sciences et Technologies de Lille,France

[11] Jones, M. J., and Rehg, J. M. (1999) Statistical color models with application to skin detection. In Proc. of the CVPR '99, vol. 1,pp: 274–280.

[12] M. Storring, H. Andersen, E. G., (1999) Skin color detection under changing lighting conditions. In Araujo and j. Dias (ed.) 7th Symposium on Intelligent Robotics systems, pp: 187-195.

[13] Caetano, T.S., Olabarriaga, S.D., and barone, D.A.C. (2003) Do mixture models in chromaticity space improve skin detection Pattern Recognition, 36(12), pp.3019-302.

[14] J. Terrillon and S. Akamatsu, (1999) Comparative Performance of Different Chrominance Spaces for Color Segmentation and Detection of Human Faces in Complex Scene Image".

[15] V. Vezhnevets, V. Sazonov, A. Andreeva, (2003). A Survey on Pixel-Based Skin Color Detection Techniques, In Proceedings Graphicon-2003, pp. 85-92, Moscow, Russia, September.

[16] J. Kovac, P. Peer and F. Solina,(2003), 2D versus 3D color space face detection, 4th EURASIP Conference on Video/Image Processing and Multimedia Communications, Croatia, pp. 449-454.

[17] Dr. John, C. Russ, (2009), The Image Processing and measurement cookbook».Vol.1 978-1448691210

[18] Nusirwan Anwar bin Abdul Rahman, Kit Chong Wei and John See. (2006) RGB-H-CbCr Skin Colour Model for Human Face Detection, In Proceedings of The MMU International Symposium on Information & Communications Technologies (M2USIC).

[19] Vladimir Vezneves Vassali Sazonov , A Survey on Pixel-Based (2002) Skin Color Detection Techniques. Graphics and Media Laboratory, Faculty of Computational Mathematics and Cybernetics Moscow State University, Moscow, Russia, novembre.

[20] HUICHENG ZHENG, (1998) Blocking Objectionable Images: Adult Images and Harmful Symbols, January.

[21] Ricco RAKOTOMALALA, (2006) Receiving Operating Characteristics Another way of estimating a model of prediction 8th International Conference on Enterprise Information Systems

[22] Y.Z. Hsu, H.H. Nagel, and G. Rekers, (1984) "New Likelihood Test Methods for Chang Detection in Image Sequences", Computer Vision, Graphics, and Image Processing, Vol. 26,pp.73-106.

[23] J.M. Letang, V. Rebuffel, and P. Bouthemy, (1993) Motion Detection Robust to Framework, Proc. Int'l Conf. Computer Vision.

[24] A. Lipton, H. Fujiyoshi, and R. Patil, (1998) Moving Target Classification and Tracking from Real-Time Video, Proc. Workshop on Application of Computer Vision.

[25] R. Cutler, and L. Davis, (1998) View-Based Detection and Analysis of Periodic Motion, Proc. Int'l Conf. Pattern Recognition.

[26] L. Wixson, (2000) Detecting Salient Motion by Accumulating Directionally-Consistent Flow, IEEE Trans. Pattern Analysis and Machine Intelligence, Vol.22 no. 8, pp.774-780.

[27] W.E.L. Grimson, C.Stauffer, R. Romano, and L.Lee, (1998) Using Adaptive Tracking to Classify and Monitor Activities in a Site, Proc. IEEE conf. Computer Vision and Pattern Recognition.

[28] C. Wen, A. Azarbayejani, T. Darrell, and A. Pentland, (1997) Pfinder: Real-Time Tracking of the Human Body", IEEE Trans. Pattern Analysis and Machine Intelligence, Vol.19, no.7.

[29] F. Liu and R. Picard, (1998), Finding Periodicity in Space and Time, Proc. Int'l Conf. Computer Vision,

[30] S. Niyogi and E. Adelson, (1994), Analyzing and Recognizing Walking Figures in xyt, Proc. IEEE. Conf. Computer Vision and Pattern Recognition.

[31] J. Stauder, R. Mech, and J. Ostermann, (1999) Detection of Moving Cast Shadows for Object Segentation, IEEE Trans. Multimedia, Vol. 1, no. 1, pp. 65-76, March.

[32] C. R. Wren, A. Azarbayejani, T. Darrell, and A. Pentland. (1997), Pfinder: Real-time

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

tracking of the human body". IEEETrans. on PAMI, 19(7):780–785.

[33] I. Haritaoglu, L. S. Davis, and D. Harwood. (1998) a real time system for detecting andtracking people". In Proc. of ICAFGR, pages 222–227.

[34] A. Elgammal, D. Harwood, and L. Davis. (2000), Non-parametricmodel for background subtraction. In Proc. of ECCV, pages751–767.

[35] T. Kanade, R. T. Collins, and A. J. Lipton. (1998), Advances in coorperative multi-sensor video surveillance. In Proc. Of DARPA

[36] C. Stauffer and W. Grimson, (1999), Adaptive background mixturemodels for real-time tracking. In Proc. of CVPR, pages 246–252.

[37] J.D. Foley, A.v. Dam, S.K. Feiner, and J.F. Hughes, (1990) Computer Graphics: Principles and Practice. New York: Addison Wesley.

[38] Michael J. Jones and James M. Rehg, (1998), Statistical color models with application to skin detection, Technical Report Series, Cambridge Research Laboratory

[39] Yuna Jung, E. Hwang, Wonil Kim, (2004), Sports Image Classifier based on Bayesian Classification, Lecture Note in Artificial Intelligence 3040, 546-555, Springer

[40] Feng Jiao, Wen Giao, Lijuan Duan, and Guoqin Cui, (2001), Detecting adult image using multiple features, IEEE conference, Nov, pp.378 - 383 vol.3.

[41] David Hand, Heikki Mannila, and Padhraic Smyth, (2001) Principles of Data Mining, MIT Press, pp343-347

[42] F. Rosenblatt, (1958), The Perceptron : A probabilistic model for information storage and organization in brain, Psychology Review 65, pp386-408

[43] D.E. Rumelhart, G.E. Hinton, and R.J. Williams. (2002),Learning representations by back- propagating errors, Nature(London), Vol. 323, pp533-536

[44] http://www.heatonresearch.com

AUTHOR PROFILES:

Hajar BOUIROUGA is a PhD student in Computer and Telecommunication at Mohammed V-Agdal University of Rabat in Morocco, her Bachelor's degree (License es Sciences) in statistiques in 2004 She received her Master's degree (DESA) in Computer and Telecommunication from Mohammed V-Agdal University (2005-2007). Her thesis title is "Recognition of Adult Video". She is a member of the laboratory LRIT (Unit associated with the CNRST, FSR, Mohammed V University Agdal, Rabat, Morocco). Mlle Hajar can be contacted at bouirouga_hajar@yahoo.fr

Sanaa EL FKIHI received her Bachelor's degree (License es Sciences) in computer science in 2002 and her Master's degree (DESA) in computer sciences and Telecommunications engineering from the faculty of sciences, university Mohammed V Agdal, Rabat, Morocco, in 2004, and she developed his Masters Project at the TELECOM Bretagne, France. She is received her Ph.D. degree in Computer Sciences and Telecommunications engineering from University of Mohammed V Agdal, Rabat, Morocco, jointly with University of Sciences and Technologies of LILLE-France, in 2008.
Currently, she is an associate professor at ENSIAS (Ecole Nationale Suprieure en Informatique et Analyses de Systmes), Rabat, Morocco. Her current research interests include graph theory, image processing and wireless sensor network.

Abdelilah JILBAB is teatcher at the Ecole Normale Supérieure de l'Enseignement Technique de Rabat, Morocco; He acquired the PhD in Computer and Telecommunication from Mohammed V-Agdal University, Rabat, Morocco in February 2009. His thesis is concerned with the Filtering illegal sites on the Internet: Contribution to the type of image recognition based on the Principle of Maximum Entropy. Since 2003 he is a member of the laboratory LRIT (Unit associated with the CNRST, FSR, Mohammed V University Agdal, Rabat, Morocco).

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Driss ABOUTAJDINE received the Doctorat de 3' Cycle and the Doctorat d'Etat-es-Sciences degrees in signal processing from the Mohammed V-Agdal University, Rabat, Morocco, in 1980 and 1985, respectively.

He joined Mohammed V-Agdal University, Rabat, Morocco, in 1978, first as an assistant professor, then as an associate professor in 1985, and full Professor since 1990, where he is teaching, Signal/image Processing and Communications.

Over 30 years, he developed research activities covering various topics of signal and image processing, wireless communication and pattern recognition which allow him to publish over 300 journal papers and conference communications.

He succeeded to found and manage since 1993 and 2001 respectively the LRIT Laboratory and the Centre of Excellence of Information & Communication Technology (STIC) which gathers more than 30 research laboratories from all Moroccan universities and including industrial partners as well.

Prof. Aboutajdine has organized and chaired several international conferences and workshops.

He was elected member of the Moroccan Hassan II Academy of Science and technology on 2006 and fellow of the TWAS academy of sciences on 2007.

He received several awards such: "Chevalier de l'Ordre des Palmes Académiques" by the French Prime Minister As IEEE Senior Member, he co-founded the IEEE Morocco Section in 2005 and he is chairing the Signal Processing chapter he founded in December 2010.