Neural Network Adult Videos Recognition using Jointly Face Shape and Skin Feature Extraction

Hajar Bouirouga[1], Sanaa Elfkihi[2], Abdelilah Jilbab[3] and Driss Aboutajdine[1]
[1]LRIT, unité associée au CNRST, FSR, Mohammed V, Rabat, Morocco
[2]ENSIAS, Mohammed V , University Souissi, Rabat, Morocco
[3]ENSET, Madinat Al Irfane, Rabat-Instituts, Rabat, Morocco

Keywords: Skin Detection, Activation Function, Neuron Networks, Pornographic Images Descriptors, Video Filtering, Face Detection.

Abstract: This paper presents a novel approach for video adult detection using face shape, skin threshold technique and neural network. The goal of employing skin-color information is to select the appropriate color model that allows verifying pixels under different lighting conditions and other variations. Then, the output videos are classified by neural network. The simulation shows that this system achieved 95.4% of the true rate.

1 INTRODUCTION

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 the literature, different adult image filtering methods are presented. A skin color is used in combination with other features such as texture and color histograms. Most of these systems build on neural networks or Support Vector Machines (Duda R.O et al., 2001) as classifiers. One of the pioneering works is done by Forsyth et al. (Fleck et al., 1996) where they combine tightly tuned skin filter with smooth texture analysis. Another work is conducted by (Duan et al., 2002). Their study is based purely on skin color detection and SVM. The images are first filtered by skin model and outputs are classified. (Rowley et al., 2006) propose a system that includes skin color and face detection where they utilize a face detector to remove the special property of skin regions. In this paper we propose ANN method based adult video recognition; the videos are classified by using a 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.

A brief system overview is given in section 2. 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

The real-time system is based on motion detection and segmentation of skin tone. The movement in each image is detected by comparing images taken at progressively stream video to each other. The next step identifies skin tones, and then the current image is converted into binary image, which are manually classified into adult and non-adult sets to train a neural networks classifier. For an input pattern p, the output OP is a real number between 0 and 1, with 1 for adult image and 0 for no-adult image. Thereafter, it establishes a threshold T, 0 <T <1, for a binary decision.

3 MOTION DETECTION

Detecting motion, carried out immediately after the acquisition of an image, represents a very advantageous for a digital vision system. Indeed, a considerable performance gain can be achieved when the interest-free zones are eliminated before

Bouirouga H., Elfkihi S., JIlbab A. and Aboutajdine D..
Neural Network Adult Videos Recognition using Jointly Face Shape and Skin Feature Extraction.
DOI: 10.5220/0004210904220425
In *Proceedings of the International Conference on Computer Vision Theory and Applications* (VISAPP-2013), pages 422-425
ISBN: 978-989-8565-47-1

the analysis phases. The fundamental principle of this method is based on a statistical estimate of the observed scene. The movement is detected by comparing a test image with the model background calculated earlier (Letang et al., 1993).The algorithm used for subtraction of the background in statistical modeling has two major steps: initialization and extraction of foreground.

3.1 Initialization

The first step is to modelling the background from the first N frames (N ≈ 30) for a sequences of videos. An average intensity is calculated from these images for each pixel and for each channel (R, G and B).

The next step is to calculate a standard deviation for each pixel (for each channel) to be used as the threshold of detection.

3.2 Extraction of Foreground

To extract the motion in an image, the model of the background it must first be subtracted. A mask of motion can then be generated for each channel.

Therefore, if motion is detected for a pixel in a single channel, this will be enough to change the state.

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. 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 colours spaces have been used in earlier works of skin detection, such as RGB, normalized RGB, YCbCr, HIS and TSL (Vezhnevets et al., 2003), but many of them share similar characteristics. The question now is: Which is the space of color best to use? To answer on this question, we propose different combinations of existing color space. Thus, in this study, we focus on the tree representing of the color spaces that are commonly used in image processing:

In RGB space, each color appears in its primary spectral component red, green and blue. Therefore, skin colour 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 (Kovac et al., 2003):

$$\begin{cases} (R>95)\mathrm{et}(G>40)\mathrm{et}(B>20)\mathrm{AND} \\ ((\max[R,G,B])-(\min[R,G,B]))>15)\mathrm{AND} \\ (\mathrm{ABS}(R-G)>15)\mathrm{AND}(R>G)\mathrm{et}(R>B) \end{cases} \quad (1)$$

$$\begin{cases} (R>b)\mathrm{et}(G>B))\,\mathrm{OR} \\ ((R>220)(G>210)\,\mathrm{OR} \\ (B>170)\,\mathrm{AND}(\mathrm{ABS}(R-G)<=15)) \end{cases} \quad (2)$$

While the skin color under flashlight or daylight lateral illumination rule is given by (Kovac et al., 2003):

RGB values are transformed into YCbCr (Kovac et al., 2003) values using the formulation:

$$Y = 0.299R + 0.587G + 0.114B \quad (3)$$

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

$$Cr = R - Y \text{ and } Cb = B - Y \quad (4)$$

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 skin detection, we consider only the channels H and S represent the chromatic information.

$$0 < H < 50 \quad (5)$$

$$0.23 < S < 0.68 \quad (6)$$

In this paper 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 (HUICHENG et al., 1998).

5 FACE SHAPE AND FEATURE EXTRACTION

After choosing the model of the skin, we propose a new method to identify adult video based on face detection. The category of the shot was considered to be "Adult", only if there is at least one image with more than one face within that shot. It can be concluded that most common way in video adult detection is via detecting human face. Human face is the most unique part in human body, and if it is accurately detected it leads to robust human existence detection. Identifying the presence of face in video streams is one of the most important features that must be extracted. For each image of the video containing more than one face, we calculate the number of existing faces in each frame of video then removes the region face, and calculate the rate of correct detection of the skin. In order to separate the region face, we scan the segmented image in search of pixels that match the label of the region. The result will be a binary image that does not contain the region.

We must first determine the number of regions of skin in the image, by associating with each region an integer value called a label. We performed measurements by testing different sets of 100 and averaging the results. All of the results are represented by the following figure.

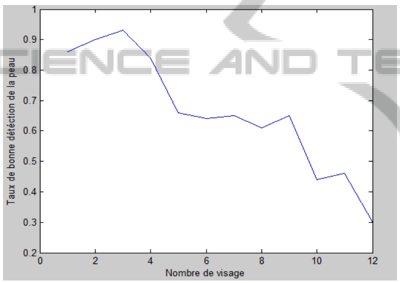

Figure 1: Rate of good detection based on the number of face.

We assume that an image will contain an adult material if the image contains at max four persons and one person at least. Normally this is where we find the most actually. Our way proves to be able to correctly online determine the skin and effectively distinguish naked videos from non-naked videos by integrating texture, features extraction and face detection. 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 next paragraph to decide what kind of video analysis. After we present functions based on grouping of skin regions which could distinguish the adult images of the other images. Many of these features are based on suitable ellipses 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) and Local Ellipse (LFE) based only on the largest region on the map skin. We distinguish 8 functions of the skin map 3 first functions are global.

- The average probability of skin of the entire image.
- The average probability of skin inside the GFE.
- The number of areas of skin in the image.
- Distance from the larger area of skin at the center of the image.
- The angle of the main axis of the LFE of horizontal axis.
- The average probability of skin inside the LFE.
- The average probability of skin outside the LFE.
- Number of dominant face in the video to analyze.

6 NEURAL NETWORK

In this step, we suggest to use the Artificial Neural Network (ANN) classifier which is considered as the majority common technique used of a decision support system in image processing. In particular we use a Multi Layer Perceptron (MLP) neural network. Hence, the used network concentrates on the study of decision-boundary surface telling adult videos from non-adult ones. It is composed of a large number of vastly interconnected processing elements (neurons) working in unison to solve the adult video recognition problem. The decision tree model recursively partitions an image data space, using variables that can divide image data to most identical numbers among a number of given variables. This technique can give incredible results when characteristics and features of image data are known in advance (BOUIROUGA et al., 2011). The inputs of our neural network are fed from the feature 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. Our MLP classifier is a semi-linear feed forward net with one hidden layer. The MLP output is a number between 0 and 1; with 1 for adult image and 0 for no-adult image.

7 EXPERIMENTS

We conduct two experiments in performance evaluation: one for the detection of skin and one for the classification of videos. In skin detection evaluation, we use 200 videos, 130 for training and 70 adult videos for test. Performance comparison between the different color spaces is shown in Figure 2.

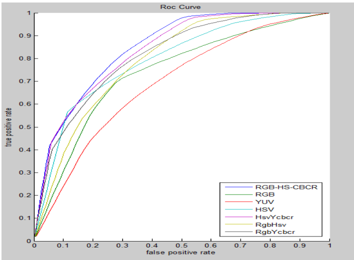

Figure 2: ROC curves for different color spaces.

From Figure 2, 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. The objective is to show that for all color spaces their corresponding optimum skin detectors.

The best rate on the other hand was obtained by the space RGB-HS-CbCr that is 97 % while the lowest score is obtained by the space YCbCr 64 %. After skin detection, two fit ellipses are used for each skin map. The fit ellipse of all skin regions and the fit ellipse of the largest skin region. Some example frames are shown in Figure 3.

Figure 3: First row: a) Original frame, b) Skin detection of the whole image, c) Skin detection inside the GFE. Second row: a) Original frame, b) Large area of the skin map, c) Skin detection inside the LFE.

After this step we adapt neural networks to classify videos.

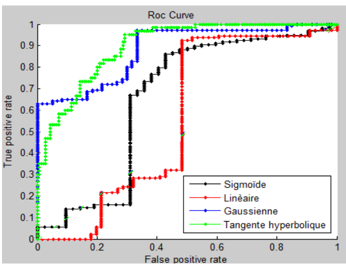

Figure 4: ROC curves for different functions activation for adult video identification.

For a fixed false given alarm FP=0.3 the highest rate TP of detection was given by hyperbolic tangent activation function (95.4%) while the lowest score is obtained by Linear function (25.3%). We demonstrate how different functions activation contributes to the solution of an adult video problem.

8 CONCLUSIONS

This article describes a filtering system of video, which aims to automatically detect and filter out adult content. Our system combines skin detection with motion information, face detection and uses neural network techniques to classify the videos. We found that the model RGB-H-CbCr gave the best results for still images. Many experimental results are presented including a ROC curve. Experimental results show that hyperbolic tangent activation function is more efficient compared to sigmoid and gaussian activation function. The simulation shows that this system achieved 95.4% of the true rate. 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.

REFERENCES

Duda R. O., Hart P. E., Stork D. G. *(2001)* Pattern Classification. *John Wiley & Sons, USA.*

Fleck, M., Forsyth, D.A., Bregler, C. *(1996)* Finding Naked People. pp. 593--602. *Springer.*

Duan L., Cui G., and Zhang H., *(2002)* Adult Image Detection Method Base-on Skin Color Model and SVM. In *5th Conference on Computer Vision. pp. 780--797.*

Rowley H. A., Jing Y., Baluja S., *(2006)* Large Scale Image-Based Adult-Content Filtering. In *1st International Conference on Computer Vision Theory. pp. 290--296.*

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

V. Vezhnevets, V. Sazonov, A. Andreeva,. (2003) A Survey on Pixel-Based Skin Color Detection Techniques, *Graphicon-2003, pp. 85-92 .*

J. Kovac, P. Peer and F. Solina, (2003) 2D versus 3D color space face detection, *EURASIP , pp. 449-454.*

Huicheng Zheng, (1998) Blocking Objectionable Images: Adult Images and Harmful Symbols, *January.*

H.Bouirouga,S. Elfkihi, A. Jilbab, D. Aboutajdine, *(2011)* Skin Detection in pornographic Videos using Threshold Technique, *JATIT, Vol 35 Issue 1, 15th January.*