

Comparison of Performance between Different SVM Kernels for the Identification of Adult Video

Hajar Bouirouga, Sanaa El Fkihi , Abdeilah Jilbab And Driss Aboutajdine1

Abstract—In this paper we propose a method for recognition of adult video based on support vector machine (SVM). Different kernel features are proposed to classify adult videos. SVM has an advantage that it is insensitive to the relative number of training example in positive (adult video) and negative (non adult video) classes. This advantage is illustrated by comparing performance between different SVM kernels for the identification of adult video.

Keywords—Skin detection, Support vector machine, Pornographic videos, Feature extraction, Video filtering, Classification.

I. INTRODUCTION

FELTING the adult objectionable materials is very essential to evade offensive content on the web. There are some behavior to stop naked images and videos arriving on computers, such as blocking unwanted sites, identifying images and sequences depicting naked people. Here, we focus on adult video detection. In general, the process of this latter is composed of three steps that are: skin detection, feature extraction, and decision system. The most important decision systems that offer good results in many image and video processing are support vectors machines (SVMs) and artificial neural networks (ANNs). In this paper we propose SVMs method based adult video recognition; the videos are classified by using a support vector machine 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.

Section 1 gives the most significant skin detection method, while Section 2 presents the proposed features that allow identifying adult videos. Sections 3 and 4 describe respectively the support vector machine and the performance between different SVM kernels. Section 5 compares the results obtained by our approaches. Conclusion is customary in Section 6.

II. SKIN DETECTION REVIEW STAGE

Skin detection plays an important role in various applications such as searching and filtering image content on the web [1], and face detection [2]. In particular, skin detection is the first step for naked detection in videos. However, there are two challenges segmentation skin: variability model and the complexity of the scene.

The first point is summed up in the human skin tone variations while the second point includes uncontrolled illumination, and low quality images in addition to other factors. Several researches have been performed on the detection of human skin pixels in color images to face its challenges using color information as a cue. The skin color segmentation consists of two steps that are: (1) color space selection for skin modeling and (2) building a modelling method. In addition, a third step can be considered. It consists in the use of morphological operations, such as dilation and erosion. The aim of the latter is to improve the skin detection results.

In the next, we will describe the three steps of our proposal skin detector.

A. Color Space Selection

Segmentation with skin may become more robust to lighting variations if pixel luminance is rejected. The HSV color space is usually used to generate high quality computer graphics. Although the colors used in HSV can be evidently distinct by human perception which is not forever the case with RGB or YCrCb [3].

We propose to use HSV color space to deal with the skin detection problem. In this color space, the intensity information is represented through the V, for this reason, this channel should be overlooked in the process of the skin detection. Hence, to describing the skin color the channels H and S which represent the chromatic information are considered [4]. In case of HSV model space a skin pixel is classified as follows [5]:

$$\begin{cases} 0 < H < 50 \text{ AND} \\ 0.23 \leq S < 0.68 \end{cases} \quad (1)$$

B. Skin Detection Techniques

There has been significant recent growth in the research activity. Thus, a considerable number of skin segmentation techniques in color images, has been proposed. They can be classed as physical based, parametric or non-parametric approaches. The most used methods are Gaussian [6], Gaussian mixture [6][7], and the Bayesian classifier with the histogram technique [6]. An exhaustive survey of skin detection can be found in [8].

In this paper we used a Bayes classifier; a bayesian skin color model is very accurate in detecting skin colors. Furthermore, the considered classifier is not a time consuming method. The last characteristic is a so important feature in any considered technique of video analysis.

Unfortunately, pixel-wise color segmentation is not

Bouirouga Hajar is with University Mohammed V, Rabat, 8-mail:bouirouga_hajar@yahoo.fr)

sufficient for skin detection purpose. In addition, we notice that true skin area in a given image, is define by a homogenous region. Therefore, non-homogenous skin colored regions should be removed because they define skin like color zones.

Also, there always exists a boundary between the true skin region and the background. Therefore, we propose detecting such boundary using motion detectors, and subsequently removing boundary pixels from the skin map.

The value of the probability that a given color c is a skin color ($P_{skin}(c)$) is actually a conditional probability $P(c|skin)$ [9][10]; the possibility of observing color c , knowing that we see a skin pixel. A more appropriate calculate for skin detection would be $P(skin|c)$ that is the probability of observe skin and given a concrete c color value. With Bayes rule is:

$$P(skin/c) = \frac{P(c/skin)P(skin)}{P(c/skin)P(skin) + P(c/-skin)P(-skin)} \quad (2)$$

$P(c|skin)$ and $P(c|-skin)$ are directly calculated from skin and non-skin color histograms. Two classes of decisions are compared against the selected threshold $0 < \Theta < 1$; that can be set experimentally. A pixel is said skin pixel if $P(skin/c) > \Theta$, otherwise it is classified as non skin pixel.

C. Morphological Filters

There are many noises, in the classification of pixels like skin and non-skin. These noises are introduced by the false classification of a skin pixel to a non-skin one or by the false classification of a non-skin pixel to a skin one. To denoise the results of the output skin detector morphological operators can be used. In particular, we propose to use the erosion and the dilatation operators.

First, we use a structuring element of dilatation filter. It aims to expand the areas in the skin regions. After that the same structuring element is considered to erode the image and reduce all the imperfections that the dilatation created.

III. FEATURE EXTRACTION

Once the binary images defining the skin areas are detected for each training image, the next step of the adult video recognition process consists in extracting features that discriminate adult video and non-adult ones.

Since skin distribution is of the paramount importance for the detection of adult videos. For the sake of practicality, to classify videos in to objectionable or not-objectionable ones, a speedily process is required. Hence, we are interested to try simpler and easy features to calculate. Many of our features are based on the fit ellipses [11] designed on the skin map, since they could meet our condition for simplicity and capture some significant shape information.

The fit ellipses will expectantly at least help discriminate portraits from adult images. We will evaluate two fit ellipses for every skin map. The Global fit Ellipse (GFE) with the Local fit Ellipse (LFE).

Three features are calculated on the (GFE) and six other features that are calculated on the (LFE).

The nine computed features compose a simple feature vector that will be the input of the decision system to classify a given video in to adult or non adult one.

IV. SUPPORT VECTOR MACHINES

Support vector machines are learning machines that classify data by shaping a set of support vectors [12].

SVMs provide a generic mechanism to robust the surface of the hyper plane to the data through. Another benefit of SVMs is the low expected probability of generalization errors [13]. Moreover, once the data is classified into two classes, an appropriate optimizing algorithm can be used if needed for feature identification, depending on the application [14].

SVM creates a hyper-plane between two sets of data for classification; in our work, we separate the data in to two classes: adult videos and non adult videos. Input data X that fall one region of the hyper-plane, $(X \cdot W - b) > 0$, are labeled as +1 and those that fall on the other area, $(X \cdot W - b) < 0$, are labeled as -1.

Let $\{X_i, Y_i\} \in R^n$ be training data set, $Y_i \in \{1, -1\}$, $i=1,2,\dots,n$. There exists hyper-plan

$$P = \{X \in R^n | X \cdot W + b = 0\} \quad (3)$$

The training data set satisfies the following condition:

$$X_i^T \cdot W + b \geq 1, \quad Y_i = 1 \quad (4)$$

$$X_i^T \cdot W + b \leq -1, \quad Y_i = -1 \quad (5)$$

Where W is a weight vector and b is a threshold

$$W = \sum_{i=1}^n \alpha_i Y_i \phi(X_i) \quad (6)$$

$$\sum_{i=1}^n \alpha_i Y_i = 0$$

The decision boundary of the SVM can then be expressed as follows:

$$f(x) = \text{sign}(W \cdot \phi(X) + b) \quad (7)$$

$$f(x) = \text{sign}\left(\sum_i \alpha_i Y_i (\phi(X) \cdot \phi(X_i) + b)\right) \quad (8)$$

$$f(x) = \text{sign}\left(\sum_i \alpha_i Y_i (K(X, X_i) + b)\right) \quad (9)$$

The presented SVM is used to separate adult and non adult videos patterns. In our case all adult videos are classified as +1, and normal videos are classified as -1.

V. COMPARISON OF PERFORMANCE BETWEEN DIFFERENT SVM KERNELS

The purpose of this part is to study in detail the kernel functions that allow an optimal separation of data and present here some examples of these common features. Any algorithm that uses only the dot product between input vectors can be generalized to an algorithm calculating the value of core vectors. The calculations are made implicitly via the kernel function. In the next, we present some examples of kernels.

A. Polynomial kernel

The calculation of a polynomial kernel of the form $k(x, z) = \langle x, z \rangle^d$, where $x, z \in \mathbb{R}^n$ is equivalent to calculating the dot product of two vectors mapped in the space of all

products (components) of order d . This property is very important practice. Indeed, the cost of such an operation is equivalent to that of a complete exponentiation of the scalar product (in the input space) between two vectors. We conclude this section by mentioning that using a kernel of the form $k(x, z) = (\langle x, z \rangle + 1)^d$ returns to calculate the scalar product in the space of all product components up to order d .

B. Gaussian kernel

The generic form of this kernel is:

$$K(x, z) = \exp\left(-\frac{\|x - z\|^2}{2\sigma^2}\right) \quad (10)$$

The parameter sets the width of the Gaussian. By taking a large σ , the similarity of an example in relation to those around him will be quite high, so that by taking σ tending to 0, the example will be like no other. By tightening strongly Gaussian classifier (making use of the kernel) can happen to learn any example (training set) without committing an error. One senses immediately that the danger of learning by heart is not far. In fact, σ is another parameter to control the ability of a classifier.

C. Sigmoid kernel.

The neuronal sigmoid kernel is defined as:

$$K(x, z) = \tanh(\eta \langle x, z \rangle + \beta) \quad (11)$$

The use of such a kernel is equivalent to a neural network with one hidden layer. This kernel depends on two parameters η and β , which can cause problems during its implementation.

D. linear kernel

The linear kernel define in R^p is

$$k(x, z) = \sum_{i=1}^p x_i z_i = \langle x, z \rangle \quad (12)$$

It has:

$$\forall x, z, k(x, z) = k(z, x)$$

$$\forall a_1, \dots, a_N, x_1, \dots, x_N \in \mathbb{R}^n$$

$$\sum_{i,j} a_i a_j k(x_i, x_j) = \left\| \sum_{i=1}^N a_i x_i \right\|^2 \geq 0$$

So it is clear that encode core functions of basic knowledge about the problem studied, in particular :

- The functional form of the functions possible decisions.
- The type of control on the assumptions made : for example the Gaussian kernel functions penalize the derivatives of all orders and thus promote regular solutions. We must therefore choose carefully when trying to translate with the maximum possible prior knowledge available about the problem and data.

The choice of parameters that best match the problem is not

intuitive. Indeed, in the case of classification we have seen previously, the parameter C is determined. Similarly, if for example we use a Gaussian kernel, the parameter σ is also to be adjusted. To do this, a brute force method is to try different parameter values, and measure the generalization error of SVM by cross validation. Thus, we take the parameters for which the generalization error is lowest.

VI. EXPERIMENTS

In this section, we suggest to evaluate the performances of the proposed skin detector adult video recognition based on the SVM with different kernels. We consider a training database composed of 200 videos (100 adult videos and 100 non adult videos) while the test dataset contains 100 videos (50 adult videos and 50 non adult videos). To evaluate the performances of the proposed skin detector, we present the results of skin detection in case the morphological operators are considered and in case these operators are not considered. In term of quantity estimation, we employ the Receiver Operating Characteristic (ROC) curve. It allows visualizing appearance of a two-class by using the true positive and false positive rates.

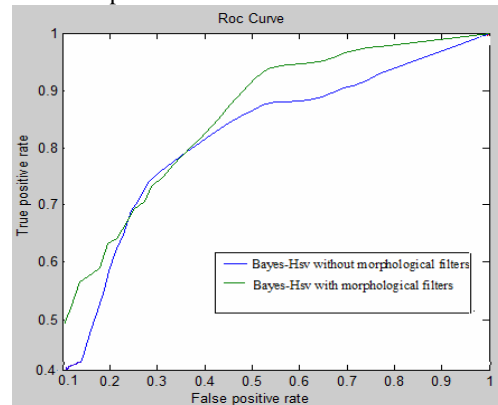


Fig.1 ROC curves for a Fusion between the model Bayesian and the HSV Based Approaches after and before morphological filters

The Figure 1 shows that our skin detector achieves detection rates of 94% and 89% for Bayes-HSV model with and without morphological filters respectively. In the next, we will estimate the performances of the different SVMs kernels of adult video recognition. Ideally an SVM analysis should produce a hyper plane that completely separates the feature vectors into two non-overlapping groups. However, perfect division may not be possible, or it may effect in a model with so many feature vector dimensions that the model does not simplify well to other data. The ROC curves for different kernels of the SVM are given in Figure 2. Here, the true positive (TP) defines an adult video classified as an adult one, while the false positive (FP) determines a non adult video classified as an adult one.

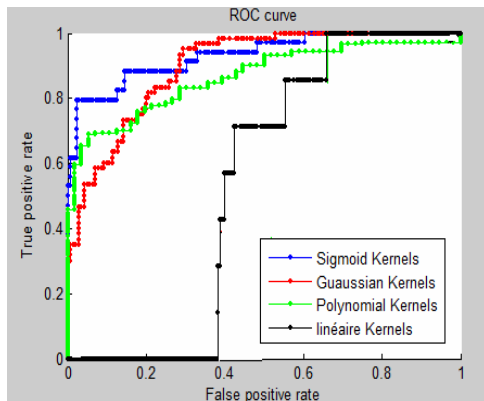


Fig.2 The ROC curves for different SVMs kernels for adult video identification

For a fixed false given alarm $FP=0.4$ the highest rate TP of detection was given by Gaussien kernel (97%) while the lowest score is obtained by Linear kernel (76%).

We demonstrate how different kernel functions contribute to the solution of an adult video problem and we also show how to select and compare such kernel. Even though SVMs are limited to making binary classifications, their superior properties of fast training, scalability and generalization capability give them an advantage in the adult video recognition.

VII. CONCLUSION

This paper aimed to make an overall comparison of different SVMs kernels applied to classify the adult and non adult videos. However, the choice of more appropriate kernel functions may still improve this classification.

A combination of color spaces and bayesian model approach for adapt skin-color decision are able to significantly reduce the number of false positive detection and the classification results become more reliable. Some adult video features are calculated on the both ellipses. Then, based on the obtained feature vector, a decision is made to recognize adult video.

Many experimental results are presented including a ROC curve. Test results taken from SVM's kernels indicate the higher ability of the Gaussien kernel for classifying and recognition of adult videos.

REFERENCES

- [1] J. Z. Wang, J. Li, G. Wiederhold, and O. Firschein, Classifying objectionable websites based on image content, Special issue on interactive distributed multimedia systems and telecommunication services, vol. 21/15, pp. 113–124, 1998.
- [2] J.-C. Terrillon, M. N. Shirazi, H. Fukamachi, and S. Akamatsu, The Comparative of different skin chrominance models for the automatic detection of human faces in color images, Fourth International Conference On Automatic Face and gesture Recognition, pp. 54–61, 2000.
- [3] DAE HYUN KIM and KIM Myoung-Jun, Skin color analysis in HSV color space and rendering with fine scale skin structure, 24th Computer Graphics International Conference, CGI 2006, Hangzhou, China, June 26–28, 2006.
- [4] W. Kelly, A. Donnellan, D. Molloy A Review of Skin Detection Techniques for Objectionable Images, Proc. 4th European Conf. on Computer Vision, 2007.
- [5] Chai, D. Phung, S.L. Bouzerdoud, A. Sch. of Eng. & Math., Edith Cowan Univ., Perth, WA, Australia, A Bayesian skin/non-skin color classifier using non-parametric density estimation, 25–28 May 2003.
- [6] S. L. Phung, A. Sr. Bouzerdoud, and D. Sr. Chai, "Skin segmentation using color pixel classification: Analysis and comparaisn," IEEE Transactions On Pattern Analysis and Machine Intelligence, vol. 27, pp. 148–154, 2005.
- [7] Ming hsuan Yang and Narendra Ahuja, "Gaussian mixture model for human skin color and its application in image and video databases," in Proceedings of SPIE 99 (San Jose CA), pp. 458–466, 1999.
- [8] S. Makrogiannis P. Kakumanu and N. Bourbakis, "A survey of skin-color modeling and detection methods," Pattern Recognition, vol. 40, no. 3, pp. 1106–1122, March 2007.
- [9] S.L. Phung, A. Bouzerdoud, D. Chai, Skin segmentation using colour pixel classification: analysis and comparison, IEEE Trans. Pattern Anal. Mach. Intell, pages 148–154, 2005.
- [10] V.Vezhnevets ,V. Sazonov, A. Andreeva, A Survey on Pixel-Based Skin Color Detection Techniques, Machine Graphics & Vision International Journal 2005.
- [11] HUICHENG ZHENG, Blocking Objectionable Images: Adult Images and Harmful Symbols, January 1998.
- [12] Feng Jiao, Wen Gao, Lijuan Duan, and Guoqin Cui, Detecting adult image using multiple features, IEEE conference, pp.378 - 383 vol.3, Nov. 2001.
- [13] Joachims T, Making Large-Scale SVM Learning Practical. LS8-Report, University of Dortmund, LS 1998
- [14] Vladimir VN, The Nature of Statistical Learning Theory. Springer, Berlin Heidelberg New York, 1995.