International Journal of Informatics and Information System ISSN 2579-7069
Vol. 2, No. 3, December 2019, pp. 121-125

A Classification of Internet Pornographic Images

Chetneti Srisa-an [1,*]

[1] Faculty of Information Technology Rangsit University, Thailand
[1] chetneti@rsu.ac.th*;
* corresponding author

Abstract

According to Pornography Statistics,more than 34 percent of Internet users exposeto pornography. There are 12 percent of the total number of websites and 72 million monthly visitors.Internet pornography (Internet Porn) is addictive to teenagers and kids around the world. The normal practice is to block those websites or filter out pornographyfrom kids.In order to do so, researchers has to find a way to detect and classify first. The pixel features including YCbCr range, area of human skin are chosen as pornographyfeatures because of their easy acquisition. C4.5 (Data mining technique)is applied to construct a decision tree. The purpose of this paper is to classify pornography images in a simple if-then rule. The accuracy of experimental result is 85.2%.

Keywords: Pornographic images; Data Classification, Skin Detection; Internet Porn;

1. Introduction

According to Pornography Statistics, more than 34 percent of Internet users expose to pornography and age-inappropriate content. There are 12 percent of the total number of websites and 72 million monthly visitors. Internet pornography is addictive to teenagers and kids around the world. A threat of the internet is pornography and age-inappropriate content. There is no or little control over the distribution of pornographic. The victims become children and their parents. In this paper, we focus on pornography images. The normal practice is to block those websites or filter out nude images from the web page. In this paper, if nudity means that the state of being without or little covering of clothing then we propose to use range of YCbCr to detect a human skin. The skin color of human [12] consists of a combination of blood (red) and melanin (yellow, brown). This paper is aim to detect and classify pornography images by calculate a total human skin area first and then detecting face. After performing faced detection, if the human skin and the largest face area is greater than the specified threshold, the image is classified as a non-nude image [10]. In this paper, we classify this type of picture as a “close-up face” type.

A decision table is constructed based on those three features along with a decision attributes as shown in Table 2. This program was developed by MATHLAB software, Version 7.1 on 5,000 sample images of the average size 405,318 pixels or 86 Kband then by WEKA. The experimental results showed that the accuracy result is 85.2%.

2. Literature Review

Many research papers [3], [4], [5], [6] have been published how to detect a nude image using color histograms, texture or shape measures. Most of them are to detect a human skin [8]. Since Most photographers usually use a special lighting and color altering in order to make an image look more attractive [9], YCbCr is the better feature to detect a human skin. Skin tone is formed by the interaction between skin color and light. Warm lighting can help skin tone more attractive, while human skin color deviates from the normal case in the same time [9]. Chai and Ngan [2] proposed in their studies pointed out that pixels belonging to the skin region exhibit similar Cb and Cr values. The skin color model based on Cb and Cr values can provide good coverage of all human races [11]. In this paper, the range of Cb and Cr will be defined in order to specify the human skin color.

3. Methodology

3.1. Identify a human skin

To detect human skin color, we examine YCbCr property. All images are loaded in Matlab and then convert from RGB to YCbCr. Y values represent luminance while CbCr represent chrominance. Color or Black-and-white images are loaded in Matlab and then convert RGB color into YCbCr color

Step 1: From Wikipedia, the formula to convert from RGB to YCbCr is as shown.

Y = 0.299R+0.587G+0.114B

R-Y = -1.72(G-Y)-0.678(B-Y)

B-Y = -2.53(G-Y)-1.47(R-Y)

G-Y = -0.581(R-Y)-0.394 (G-Y)

V = (R-Y)/ 1.14 = 0.877(R-Y)

U = (B-Y)/ 2.03 = 0.493(B-Y)

Cb = 1.144U

Cr = 0.813V

From 5000 samples, we found that the range of Y> 100 and 100 <Cb <120 or128 <Cb <130 and130 <Cr <160 indicate Human Skin. If not in this range, it is not Human Skin. No matter where any part of the body such as arm, leg, chest, back or arm, the range is applied to the same range. It is also true from any race including Asian, Europe, and African. Figure 1 shows a range of human skin tone.

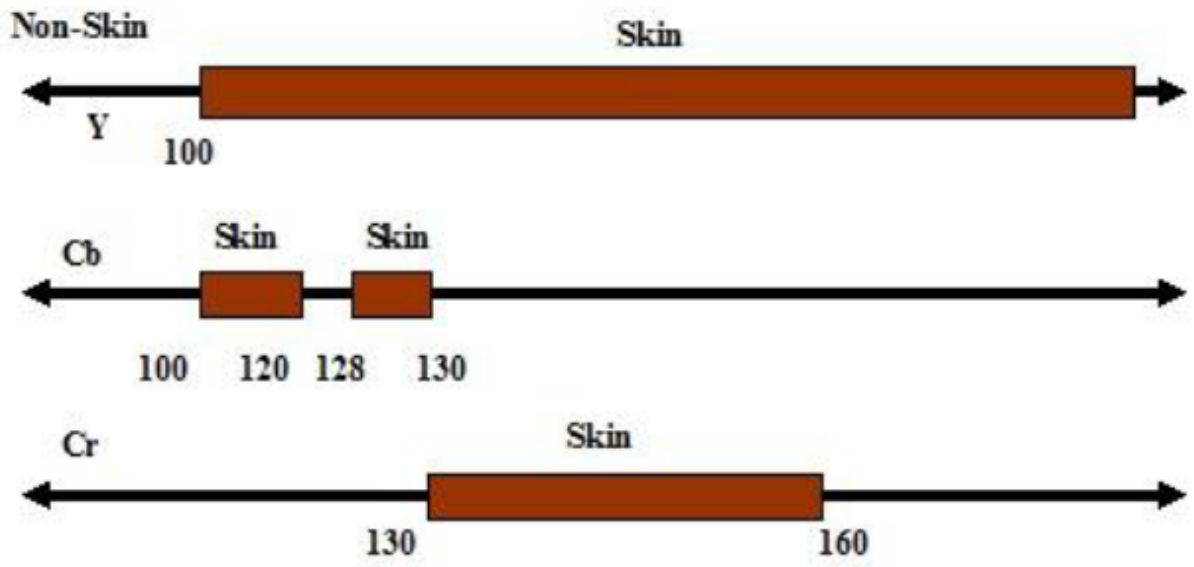

Figure. 1. Human Skin Range

3.2. Calculate a percent of human skin per image

The higher percent, the higher potential to be pornography images. Calculate Area of Human Skin (% AHS) which is a ratio of the total skin to all pixels.

%AHS = Ns/ Nr

Ns = Number of pixels in Human Skin tone

Nr = Total of pixels

Figure. 2. Natural Scenery view (RGB)

Figure. 3. % AHS = 0.03805 (Not Nude)

Figures 2 and 3 illustrate that low% AHS leads to “not nude”.

Figure. 4. % AHS = 0.8318 (Nude% AHS> = 0.4)

Figure 4 illustrate that low% AHS lead to “nude”.

3.3. Calculate a human face area

After performing face detection, if the human skin and the largest face area is greater than the specified threshold, the image is classified as a non-nude image [10]. In this paper, we classify this type of picture as a “close-up face” type. The area feature is used to quantify the size of the examined area. Edge detection is applied to distinguish human

image from non-skin image. With edge detection, segments are created. % Face_Area and for an area of the biggest segment.

% Face_Area = Face skin pixels in segment / total number of pixels

If% Face_Area > 0.3 then an image is Not nude.

If% Face_Area <= 0.3 then an image is nude.

Figure. 5. % Face_Area = 0.1348

3.4. Apply C4.5 to construct a decision tree

C4.5 is an algorithm used to generate a decision tree developed by Ross Quinlan [7]. Its nodes, except leaf-nodes, represent decision conditions and the edges stand for different values of corresponding conditions [7]. We apply C4.5 to determine the cutting point of % AHS and % Face_Area. Fig. 1 shows an example of a decision tree for our model.

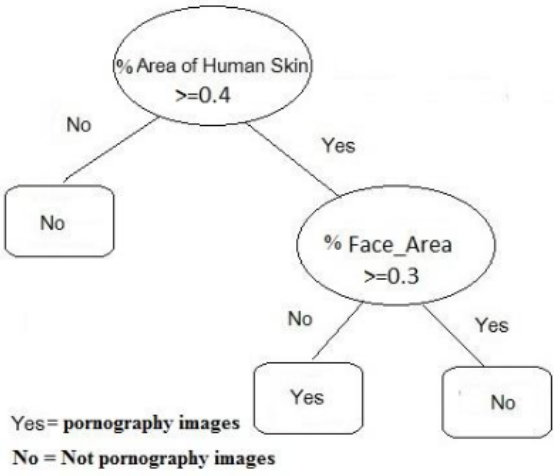

Figure. 6. Decision Tree

The experimental results showed that the accuracy result is 85.2% and average processing time is 213.14 msec. per image. From figure 7, Decision rules are as follows:

If% AHS >= 0.4 and (% Face_Area> = 0.3) then "Not Pornography images"

If% AHS >= 0.4 and (% Face_Area < 0.3) then “Pornography images”

If% AHS < 0.4 and % Face_Area < 0.3 then “Not Pornography images”

References

[1] Arentz, Will Archer, & Olstad, Bjorn. (2004). Classifying offensive sites based on image content. Computer Vision and Image Understanding-Special issue on color for image indexing and retrieval, 94, 295-310

[2] Chai, Douglas, & Ngan, King N. (1998). Locating facial region of a head-and-shoulders color image. Paper presented at the Third IEEE International Conference on Automatic Face and Gesture Recognition (FG '98), Nara, Japan.

[3] Zheng, Huicheng, Daoudi, Mohamed, & Jedynak, Bruno. (2004). Blocking Adult Images Based on Statistical Skin Detection. Electronic Letters on Computer Vision and Image Analysis, 4 (2), 1-14.

[4] FLECK, M.M., & FORSYTH, D.A. (1999). Automatic detection of human nudes. International Journal of Computer Vision, 32 (1), 63-77.

[5] Forsyth, David, & Fleck, Margaret. (1996). Identifying nude pictures. IEEE Workshop on the Applications of Computer Vision, 103-108.

[6] Jedynak, Bruno, Zheng, Huicheng, & Daoudi, Mohamed. (2003). Statistical models for skin detection. IEEE Workshop on Statistical Analysis in Computer Vision, in conjunction with CVPR 2003.

[7] Quinlan, J. R. C4.5: Programs for Machine Learning. Morgan Kaufmann Publishers, 1993.

[8] Jiao, Feng, Gao, Wen, Duan, Lijuan, & Cui, Guoqin. (2001). Detecting adult image using multiple features. Paper presented at the IEEE International Conference on Info-tech and Info-net.

[9] Leea, Jiann-Shu, Kuob, Yung-Ming, Chungb, Pau-Choo, & Chenc, E-Liang. (2007). Naked image detection based on adaptive and extensible skin color model: Elsevier, Science Direct.

[10] Jeong, Chi-Yoon, Kim, Jong-Sung, & Hong, Ki-Sang. (2004). Appearance-Based Nude Image Detection. Pattern Recognition, 2004.ICPR 2004. Proceedings of the 17th International Conference on, 4, 467-470.

[11] Phung, Son Lam, Bouzerdoum, Abdesselam, & Chai, Douglas. (2002). A novel skin color model in YCBCR color space and its application to human face detection. Paper presented at the IEEE International Conference on Image Processing.

[12] Rossotti, Hazel. (1983). Color: Why the World Isn't Gray: Princeton, NJ: Princeton University Press.