

A Novel Approach for Region-of-Interest Detection in Pornographic Images

Yizhi Liu¹, Dong Zhou, Jianxun Liu, Hongtao Xie, Ying Yang

Abstract. Region-of-interest (ROI) detection plays an important role in filtering out pornographic images. Skin-color models, as a traditional approach, extract skin-color regions as ROI. However, the approach is difficult to differentiate between human skins and other objects with the skin-color. Furthermore, skin-color regions are always larger than the subareas containing pornographic parts. In this paper, we present a novel approach for ROI detection in pornographic images by integrating visual attention and skin-color models. A novel saliency map model is firstly constructed. Then it is integrated with a skin-color model and a face detection model to capture ROI. Experimental results show that our approach is more precisely than using the skin-color model alone.

¹ Yizhi Liu (✉)

Key Laboratory of Knowledge Processing and Networked Manufacturing (KPNM), College of Hunan Province, Xiangtan, China
e-mail: yizhi_liu@sina.cn

Dong Zhou (✉)

Key Laboratory of KPNM, College of Hunan Province, Xiangtan, China
e-mail: dongzhou1979@hotmail.com

Jianxun Liu (✉)

Key Laboratory of KPNM, College of Hunan Province, Xiangtan, China
e-mail: dongzhou1979@hotmail.com

Hongtao Xie (✉)

Institute of Information Engineering, Chinese Academy of Sciences, Beijing, China
e-mail: xiehongtaohao@163.com

Ying Yang (✉)

College of Information and Electrical Engineering, China Agricultural University, Beijing, China
e-mail: hbxtyy@126.com

Keywords: Pornographic image detection • Region-of-interest (ROI) • Visual attention model • Skin-color model • Face detection model

1 Introduction

To prevent people from objectionable information while surfing on the Internet, many researchers and software developers have been focusing on pornographic image detection. Extensive experimentation over the last few years has shown that the region of interest (ROI) based approach is more accurate in describing an image content than using the global features [1,2].

In the field of pornographic image detection, ROI is usually extracted by skin-color models. For instance, Yang et al. divided an image into 16 regions by placing a 4*4 grid and created subregions via an automatic region growing technique [3]. Then ROI is obtained by the votes of neighbor points and grids. Xu et al. captured ROI based on global features and Gaussian mixed models (GMM) after segmenting images into regions with the JSEG algorithm [4]. However, the approach is difficult to differentiate between human skins and other objects with the skin-color.

Skin-color regions are usually larger than the regions drawing the outline of erotic parts. The latter are the most interesting regions in the domain of pornographic image detection. Once the areas containing erotic parts are detected to be salient, an image will be considered pornographic beyond all doubt.

Therefore, we propose a novel approach for ROI detection. This approach regards the intersection of skin-color, salient, and no-face regions as ROI of pornographic images. We firstly conduct a novel saliency map model. Then it is integrated with a skin-color model and a face detection model for ROI detection. Experimental results show that our approach is more precisely than using the skin-color model alone.

2 Related works

Images with salient objects present a complex, inhomogeneous structure. Using global features normally leads to inaccurate content descriptions [1]. How to precisely detect ROI is also challenging for pornographic image detection. Existing methods are mostly based on skin-color detection [3,4].

Visual attention models are seldom used to solve the ROI detection problem. Visual attention is a mechanism filtering out redundant visual information and detecting the most relevant parts of visual field [5]. Attention is a general concept

covering all factors that influence selection mechanisms, whether they be scene-driven bottom-up or expectation-driven top-down [6].

The saliency-based model proposed by Itti et al. [7] and the contrast-based model presented by Ma et al. [8] are two typical kinds of visual attention models. However, there are some shortcomings. On one hand, the saliency-based model analyzes each pixel of the input image via multi-scales. It needs massive computations. Hence it is very time-consuming. On the other hand, the precision of the contrast-based model has limited capability to highlight human-beings in the images.

3 Our approach

Visual attention analysis provides an alternative methodology to detect ROI in pornographic images. In this paper, we present a novel approach of integrating visual attention models and skin-color models.

3.1 Overview

Herein, ROI means the subareas of pornographic parts. Obviously, ROI is the smaller part of skin-color regions. Therefore, we attempt to capture the intersection of salient regions, skin-color regions, and no-face regions.

Figure 1 gives an overview of our approach combined bottom-up with top-down strategies. Firstly, a visual attention model is constructed based on multi-scales and contrast. Then it is integrated with a skin-color model and a face detection model to capture ROI in pornographic images.

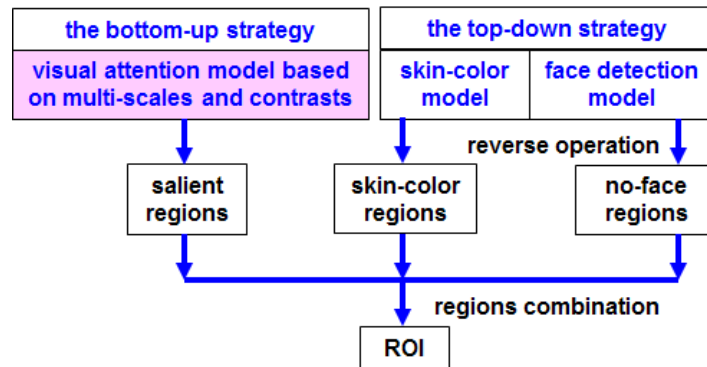


Fig. 1 An overview of our approach

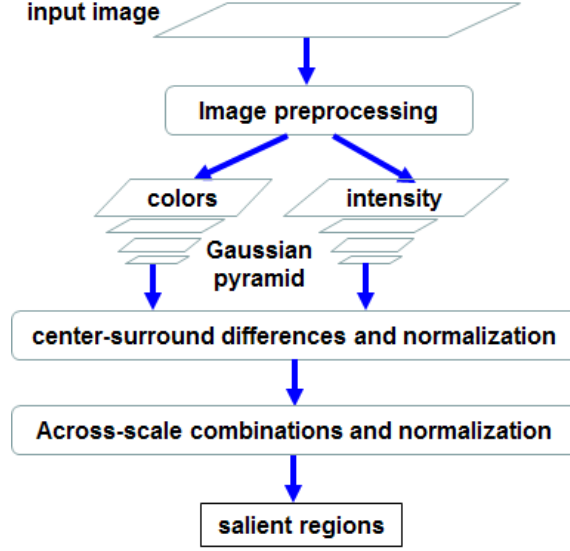


Fig. 2 The architecture of VAMSC

3.2 Visual attention model based on multi-scales and contrasts (VAMSC)

The saliency-based model [7] and the contrast-based model [8] are two typical kinds of visual attention models. The former model is time-consuming because of massive computations and the latter one has limited capability to highlight human-beings in the images.

We present a novel visual attention model VAMSC by combining the two preceding models. The architecture of VAMSC is shown in Figure 2 and its algorithm is given as follows.

Step1: Image preprocessing. First of all, the input image is resized to a uniform image with its aspect ratio unchanged. Then it is quantized to $m \times n$ block images so that the computational complexity can be effectively reduced. We arbitrarily set $m = 640$ and $n = 480$ here.

Step2: Analyzing each block image based on multi-scales and contrasts.

Step2-1: The block image (i, j) is reckoned as the perceivable unit $p_{i,j}$ ($i \in (0, m)$ and $j \in (0, n)$). The contrast value of its neighborhood $C_{i,j}$ is computed according to the equation (2).

$$C_{i,j} = \sum_{q \in \theta} d(p_{i,j}, q) \quad (2)$$

Where θ represents the neighborhood of $p_{i,j}$; q is one of these signals in θ ; d describes the distance between $p_{i,j}$ and q .

Step2-2: With r , g , and b being red, green, and blue channels of the input image, an intensity image I is obtained as $I = (r + g + b) / 3$. I is used to create a Gaussian pyramid I_σ , where $\sigma \in [0, 8]$ is the scale. Four broadly-tuned color channels are created: $R = r - (g + b) / 2$ for red, $G = g - (r + b) / 2$ for green, $B = b - (r + g) / 2$ for blue, and $Y = (r + g) / 2 - |r - g| / 2 - b$ for yellow (negative values are set to zero). Four Gaussian pyramids R_σ , G_σ , B_σ , and Y_σ are created from these color channels.

Step2-3: Each feature is computed by a set of linear “center-surround” operations, denoted as Θ . The goal is to yield three sets of feature maps: respectively concerned with intensity contrast $I(c, s)$, red/green and green/red double opponency $RG(c, s)$, and blue/yellow and yellow/blue opponency $BY(c, s)$. They are computed in terms of the following equations (where $c \in \{2, 3, 4\}$, $s = c + \delta\{3, 4\}$).

$$I(c, s) = |I(c) \Theta I(s)| \quad (3)$$

$$RG(c, s) = |(R(c) - G(c)) \Theta (G(s) - R(s))| \quad (4)$$

$$BY(c, s) = |(B(c) - Y(c)) \Theta (Y(s) - B(s))| \quad (5)$$

Step3: Feature maps are combined into two “conspicuity maps”. The first set of $I(c, s)$ is used for “intensity conspicuity maps \bar{I} ”, and the latter two sets of $RG(c, s)$ and $BY(c, s)$ are used for “color conspicuity maps \bar{C} ”. The two conspicuity maps are normalized and summed into the final input S to the saliency map. In the equation (8), the normalization operation is denoted as $N(\cdot)$, and $C_{i,j}$ of each block image is normalized into $[0, 255]$.

$$\bar{I} = \bigoplus_{c=2}^4 \bigoplus_{s=c+3}^{c+4} N(I(c, s)) \quad (6)$$

$$\bar{C} = \bigoplus_{c=2}^4 \bigoplus_{s=c+3}^{c+4} [N(RG(c, s)) + N(BY(c, s))] \quad (7)$$

$$S = \frac{1}{2} (N(\bar{I}) + N(\bar{C})) \quad (8)$$

3.3 The skin-color model and the face detection model

To extracting skin-color regions, we adopt the skin-color model proposed by Garcia et al. [9]. They proposed that skin-like pixels are more correlated with C_r

and C_b components than Y component. Thus, the input image is transformed from the RGB color model to the YC_bC_r color space.

In the domain of face detection, the Haar-Like algorithm proposed by Viola [10] is one of the most classic algorithms. The algorithm inputs Haar features into the Adaboost classifier. It can be used for real-time face detection.

4 Experiments and Evaluation

Due to the subjectivity of human attention perception, there is not a standardized objective correctness measure for image attention analysis evaluation. Therefore, we adopted the evaluation method proposed in Ma's paper [8]. Twenty non-professional persons were invited to assign one of assessments, GOOD, ACCEPT, or FAILED, to our experimental results. All the experiments were implemented in the visual studio 2003 environment with a machine of 1.86 GHz Duo CPU and 2GB memory.



Fig. 3 Some sample results of ROI detection

We collected pornographic images from the Internet and divided them into three groups, each of which includes 350 images. Our experiments were com-

prised of four parts to capture salient regions, skin-color regions, no-face regions, and final results of ROI respectively. Some of results are shown in Figure 3. Table 1 lists the evaluation results of all the ROI.

According to the results obtained, we can conclude that our approach achieves good performance and the precision reaches 91.33% in average. It is able to detect ROI in pornographic images more precisely than using Itti's model [7], Ma's model [8], and Garcia's model [9] alone.

Table 1 Evaluation of our approach

	GOOD	ACCEPT	FAILED
Group 1	41.71%	49.43%	8.86%
Group 2	41.43%	51.14%	7.43%
Group 3	46.29%	44.00%	9.71%
Average	91.33%		8.67%

5 Conclusions

ROI detection is important to object recognition. In the field of pornographic image detection, skin-color models are traditionally applied to extracting ROI. Nevertheless, they have difficulties in distinguishing human skins from other objects with the skin-color. Visual attention is a mechanism which filters out redundant visual information and detects the most relevant parts of visual field. Visual attention models provide an alternative approach to ROI detection.

In this paper, we present a novel approach for ROI detection in pornographic images. The approach integrates visual attention, skin-color, and face detection models. To accelerate visual attention analysis, a novel saliency map model is constructed. Next, it is combined with the skin-color model proposed by Garcia et al. and the Haar-Like algorithm proposed by Viola to precisely obtain ROI. Experimental results show that our approach achieves good performance and the precision averagely achieves 91.33%. It is able to detect ROI in pornographic images more precisely than Itti's model [7], Ma's model [8], and Garcia's model [9] alone.

6 Acknowledgements

This research has been supported by Hunan Provincial Natural Science Foundation of China (Grant No. 12JJ3059), and partially supported by National Nature Science Foundation of China (Grant No. 61272063, 61202330, 61300129, & 61303171).

7 References

1. Zhou Q, Ma L, Celenk M, et al (2005) Content-based image retrieval based on ROI detection and relevance feedback. *Multimedia Tools and Applications*, 27(2):251-281
2. Donoser M, Bischof H (2007) ROI-SEG: Unsupervised color segmentation by combining differently focused sub results. In: *IEEE Conference on Computer Vision and Pattern Recognition*:1-8
3. Yang J, Fu Z, Tan T, et al (2004) A novel algorithm for content-based image recognition and filtering. *Journal of China Institute of Communication*, 25 (7):93-106
4. Xu Y, Li B, Xue X Y, et al (2005) Region-based pornographic image detection. In: *The 7th IEEE Workshop on Multimedia Signal Processing*:1-4
5. Le Meur O, Le Callet P, Barba D, et al (2006) A coherent computational approach to model bottom-up visual attention. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(5):802-817
6. Borji A, Itti L (2013) State-of-the-art in visual attention modeling. *IEEE Transactions on Pattern Analysis and Machine Intelligence*
7. Itti L, Koch C, Niebur E (1998) A model of saliency-based visual attention for rapid scene analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20 (11):1254-1259
8. Ma YF, Zhang HJ (2003) Contrast-based image attention analysis by using fuzzy growing. In: *The 11th ACM International Conference on Multimedia*:374-381
9. Garcia C, Tziritas G (1999) Face detection using quantized skin color regions merging and wavelet packet analysis. *IEEE Transactions on Multimedia*, 1:264-277
10. Viola P, Jones M J (2004) Robust real-time face detection. *International Journal of Computer*, 57 (2):137-154