2011 International Conference on Computer Science and Information Technology (ICCSIT 2011)
IPCSIT vol. 51 (2012) © (2012) IACSIT Press, Singapore
DOI: 10.7763/IPCSIT.2012.V51.60

Adult Image Detection Using Wavelet Trans form and Support Vector Machine

Chun Liu[a +], Changsheng Xie [a,b], Guangxi Zhu [a,b], Qingdong Wang [a,b]

[a]Dep. of computer science, HUST, Wuhan Hubei, P.R. China

[a]Dep. of optoelectronic storage, WNLO, Wuhan Hubei, P.R. China

Abstract. An efficient method of adult image based on wavelet transform incorporating with Support Vector Machine (SVM) is proposed. Firstly, an image template matching is used to coarse filter for locating and speeding up, then a Haar wavelet transform of template matching constrained subspace is performed to decrease the dimension and extract feature for SVM classify according to the eigenvector trained and built by given samples. The test results show that the system not only achieves high detection rate, but also reduces the computational complexity, has practice ability to be applied in embedded platform.

Keywords: image detection, wavelet transform, support vector machine

1. Introduction

The high profits of sex industry and the interconnectivity, opening of internet, large numbers of pornographic information and erotica videos had already seriously interfered the normal network living, harmed the teenager’s mind and the body’s health. Accompany with overall transition of digital TV in more and more countries, and the internet function integrated in the STB devices, the pornographic videos or pictures used almost only appeared in PC domain have penetrated to and appeared at home public domain like lobby. So constructing identification and filter technology based on embedded platform had became a kind of strong requirement.

The difficulty in such application is compromise between accuracy and response time. Hung-Chih Lai, and Junguk Cho etc[1][2][3] had presented a SOC chipset architecture based on FPGA used to face detection. They analyzed and trained the face feature with Haar wavelet and ANN (Artificial Neural Network) algorithm, and accelerate the computation with pipeline and parallel process. The realtime performance reached to 625frame/s. While Zhang [4] also utilized the wavelet and SVM (Support Vector Machines) to analyze face feature. But compared to face recognition, if there includes pornographic content in a image is a high-level semantic characteristic, moreover because of the indeterminacy of object size and complexity of picture’s background, it’s very hard to filter the picture or video in short time. So the key to identifying sex picture is to decrease the gap between high-level semantic characteristic and low-level vision characteristic, in another word, it’s much easy to convert the identification to judge if there includes specific sexual organs in a picture. In this paper we transition the key frame of a video with Haar wavelet, and then classify with the eigenvector of SVM based on the trained specific sex organ’s samples, combined with fast template matching, and time-sharing identifying strategy of different sex organs, we can catch the optimum effect between detection effective and time.

2. Pornographic Video Detection Based on Wavelet Transform and SVM

[+] Corresponding author.. Chun Liu
E-mail address: Chun_liu@tom.com.

2.1. Key frame Extract and Pre-process

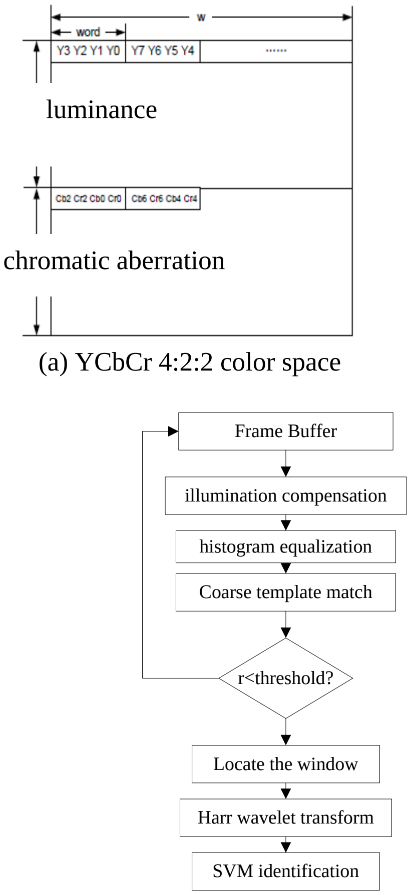
To meet the run-time claim before the video output to screen a key frame need to be extracted and pre-process, including color space transforming, build pyramid tile layer, coarse template matching, then a gray scalar graphic of skin area is building to perform the Haar wavelet transforming and SVM analysis.

The aim of extracting key frame is to convert the analysis of motion video to static picture. The fastest strategy to decrease system load and accelerate detection in embedded SOC decoder is to pick up I frame of input video every 1 second or 0.5s on time. Considered that color render and illumination intensity of TV program, are quite different from natural color and illumination, illumination compensation and histogram equalization are necessary, while even we extend the color space to: Cr ∈ [100；220] and Cb∈ [70：220], the skin detection model based on color range is still not satisfied, especially in some close-up shot image. So we have to skip the skin and texture detection, and extract key frame from YCbCr color space used in output frame buffer of current video decoder, compensate illumination and equalize histogram, then transform it to gray level graphic and build pyramid tile layer accordingly. The YCbCr architecture in frame buffer and image process flow are illustrated in Figure 1.

A suit of 64×64 pixels size breast gray level graphic templates shown in Figure 2 are used to match and detect then. These templates are picked up from pictures and ranks with different directions. Each time we scan pyramid tile layer with one template. Assumed the gray level matrix of template is T(x, y) ($0 \le x < w$, $0 \le y < h$), the mean and variance are μ_T and σ_T^2 ; the gray matrix of input image area is R(x, y) , the mean and variance are μ_R and σ_R^2. So the coefficient of correlation is:

$$r(R,T) = \frac{1}{w \cdot h \cdot \sigma_T \cdot \sigma_N} \sum_{y=0}^{h-1} \sum_{x=0}^{w-1} (R(x,y) - \mu_R)(T(x,y) - \mu_T) ; \qquad (1)$$

A coarse filter threshold value T is used to speed up the detection, and assure that potential breast area can pass the filter, when r(R, T)>T the area is considered to match the template. After locating the area, the area is transformed by wavelet and extracts the eigenvector.



(a) YCbCr 4:2:2 color space

(b) process flow

Fig. 1: The YCbCr architecture in frame buffer and process flow

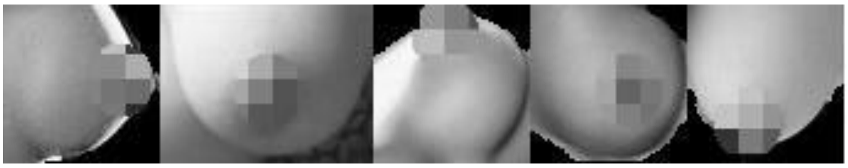
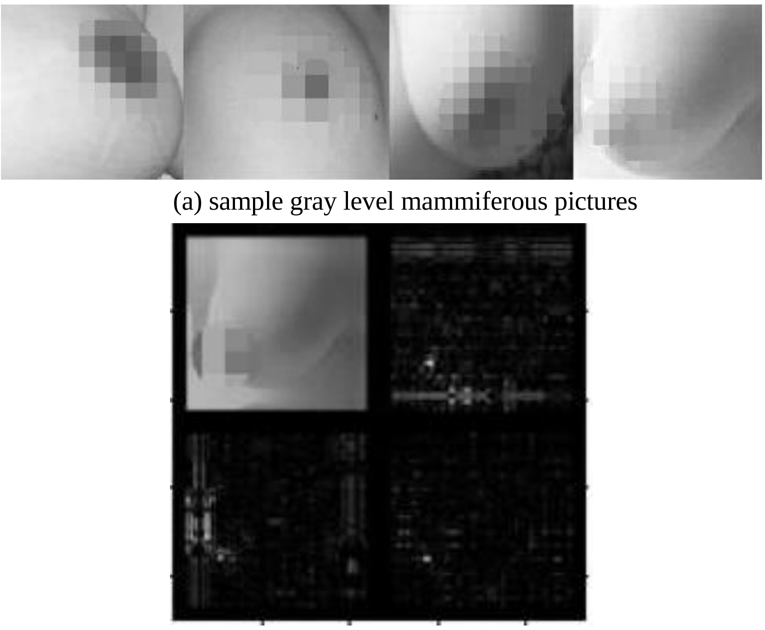


Fig. 2: The YCbCr architecture in frame buffer and pre-process flow

2.2. Susceptive Characteristic Extraction

Wavelet Transform is the signal process technique grown up in these years which has excellent local characteristic in both time domain and frequency domain, also has scale variation feature and directivity property, and widely used for all kinds of research field like image process, pattern recognition, data fusion and etc on. An image can be decomposed to one low-frequency component and a series of high frequency components by wavelet transform. The low-frequency component is the approximate image of original image, while the detail feature such as edge, brightness line, and border of different areas are distributed in high frequency components. The absolute value of wavelet coefficient reflects the intensity of variety.

After transforming an 2D image with wavelet, it results in 4 different frequency bands: LL, LH, HL, HH, each has only 1/4 size of original image size[5]. LL brings the original content information; the energy of entire image is concentred on these frequency bands. HL band bring the high frequency edge information at horizontal direction, LH keep the high frequency edge information at vertical direction, and HH keep the high frequency edge information at diagonal direction. In this paper we collect mammiferous pictures with different directions and different poses and transform them to 82×82 pixels size, and then they were delete edge and corner pixel to decrease noise, form an 80 dimensions vector as an input train vector of support vector machine introduced later. Figure 2(a) shows the mammiferous pictures and 2(b) shows the effect after Haar wavelet transform.



(a) sample gray level mammiferous pictures

(b)The LL sub picture after Haar wavelet transform

Fig. 3: The sample picture and LL sub picture

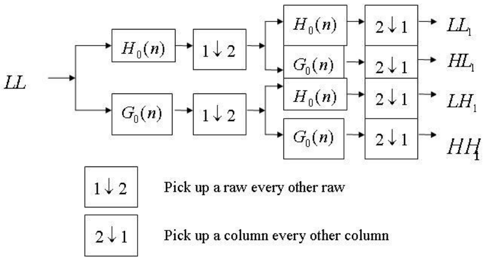


Fig. 4: The two dimension wavelet decompose process. H0(n) and G0(n) are low-pass filter and high-pass filter respectively.

After trained, the test picture are also transformed by Haar wavelet transform with Mallat fast algorithm, and then the LL sub-picture are input to SVM and perform detection. The Mallat algorithm filter the input signals with a series decompose low-pass filter H and high-pass filter G, then down sample every other sample from output set to realize the wavelet transform.

2.3. Support Vector Machine

SVM is a statistic learning method presented by Vapnik and based on SRM (Structural Risk Minimization principle), has special advantages when solve the recognition problems at small samples, non-linear and high dimension conditions[6].

Assumed that a train sample set with given size l is {(xi, yi), i=1,2,…,l}, and composed by 2 classes. If xi∈RN belongs to the first class, and marked as +(yi=1); If xi∈RN belongs to the second class, then marked as -(yi=1).If the train sample set is linear separable, then a classify hyperplane w·xi+b=0 exists and makes:

$$y_i\,(\boldsymbol{w}\cdot\boldsymbol{x}_i+b)\geqslant 1,\quad i=1,2,\cdots,l \tag{2}$$

in which w∈RN. From the statistics for machine learning theory, if all samples can be correctly separated by hyperplane, and the distance from the nearest sample to hyperplane is the longest, then the hyperplane is the optimum hyperplane, and the classic distance is 2/||w||. So to maximize the classic distance is equivalent to minimize ||w||2/2, thus the solution to optimum hyperplane is converted to quadratic programming problem like equation (2):

$$\min_{\mathbf{w},b}\frac{1}{2}\|\boldsymbol{w}\|^2,\qquad s.t.:y_i\,(\boldsymbol{w}\cdot\boldsymbol{x}_i+b)\geqslant 1,\quad i=1,2,\cdots,l \tag{3}$$

In our research the identification of obscene picture belongs to non-linear classification problem. When the train sample set is non-linear separable, the train data x can be mapped to a high dimension space, and a an optimum hyperplane $\boldsymbol{w}\,\phi(\boldsymbol{x}_i)+b=0$ can be constructed even the dimension maybe infinite.

If considered the slack variable l and punishment parameter C, the optimum hyperplane problem (2) is equivalent to equitation (3):

$$\min_{w,b}\left(\frac{1}{2}\|\boldsymbol{w}\|^2+C\sum_{i=1}^{l}\xi_i\right),\qquad s.t.:y_i\,(\boldsymbol{w}\cdot\phi(\boldsymbol{x}_i)+b)\geqslant 1-\xi_i\,,i=1,2,...,l,\xi_i\geqslant 0 \tag{4}$$

In which ξ_i, i=1,2,⋯,l is the nonnegative slack variable. With Lagrange multiplier method the solution to this quadratic programming problem restricted by linear constraint condition is given in (4):

$$\boldsymbol{w}^*=\sum_{i=1}^{l}\alpha_i^*y_i\,\phi(\boldsymbol{x}_i),\;\; s.t.:\sum_{i=1}^{l}\alpha_i\,y_i=0,0\leqslant\alpha_i\leqslant C \tag{5}$$

So the discriminant function is:

$$y(x)=\operatorname{sgn}\left\{\sum_{i=1}^{l}\alpha_i^*y_i\,K(\boldsymbol{x}_i,\boldsymbol{x})+b^*\right\} \tag{6}$$

While $K(\boldsymbol{x}_i,\boldsymbol{x})=\phi(\boldsymbol{x}_i)\phi(\boldsymbol{x})$ is kernel function. We need only compute the kernel function thus to avoid the disaster resulted by too high dimension eigenspace. In this paper we choose Gaussian RBF (radial basis function) with kernel width σ as our kernel function:

$$K(\boldsymbol{x}_i,\boldsymbol{x})=\exp(-\|\boldsymbol{x}-\boldsymbol{x}_i\|^2/2\sigma^2) \tag{7}$$

The training data set are built by hand and training method is descripted as following:

- Firstly pick up 64×64 pixels object picture just like breast from image library.
- Transform it to gray level figure and then deal with it by hand; means clear other pixel value to 0 except object.
- Scan the processed gray level figure, if value=0, then the corresponding group value in SVM training data structure is 0; otherwise the group value is 1. Then a group matrix with 4096 dimension is obtained.
- Substitute the group matrix and gray level figure in step 2 to SVMtrain, to find support vector.

- According to different objects and directions, all kinds of SVMstruct can be trained in advanced.

3. Simuliation Experiment and Analysis

Because there is no standard test picture set currently, we collect pictures from internet and capture screenshot, the pictures are good quality and all uniform lighting. Totally 80 pictures include 65 obscene scenes and 146 exposed breast characteristics, others include scenic and seaside resort scenes. The system judgment method is listed in Table 1. The system performance indexes can be descripted as (8):

$$\begin{cases} \text{Abandon True Rate: } R_a = C/(A+C) \\ \text{Take fasle Rate: } R_t = B/(B+D) \\ \text{Correct Recognition Rate: } R_c = A/(A+C) \\ \text{Error Rate: } R_e = (C+B)/(A+B+C+D) \end{cases} \tag{8}$$

Tab. 1: The system judgment Table

	obscene picture	***normal picture***
correct recognition No.	A(obscene&identified)	B(normal but confused)
error recognition No.	C(obscene but omit)	D(normal&identified)

The test pictures' classified results are shown in Table 2.

Tab.2:Test Result

Time	Value(s)	Rate	Value
pre-process time	<2s	R_a	0.0923
haar transfrom time	<1s	R_t	0.375
train time	230s	R_c	0.907
SVM classify time	<1s	R_e	0.115

The simulation test results indicate that Haar transform and coarse template matching are both time consumption actions, but because we cannot operate the output frame buffer in PC and consider that the picture format is bmp or jpg which need decoding time and transferring time, so the actual time is approximately about 1~2 second and still need to improve. Although the train time is quite long but actually it's no impact to our system because the train process can be finished in advanced. The reason about high take false rate is almost resulted from the seaside resort scene pictures; means that the model need to enhance skin detection. But this system presents a practice ability which can be applied in embedded platform. In our paper we only give the method to detect women breast characteristic, but the method can be easily used to detect other obscene feature just add a time spare sub module, which can control the system to detect different feature every several seconds.

4. Conclusion

These and the Reference headings are in bold but have no numbers. Text below continues as normal.

In this paper a pornographic picture detection algorithm is presented based on wavelet transform and SVM. The algorithm combines the coarse template matching, building pyramid layer tile , illumination compensation and histogram equalization, after that a Haar wavelet transform of pictures is performed, and then the LL data is input to SVM and classify according to the eigenvector trained and built by given samples. The test results show that the system has practice ability which can be applied in embedded platform.

5. Acknowledgements

This work is financially supported by National 973 Program (contract No. 2011CB302303) and National 863 Project (contract No. 2009AA01A402).

6. References

[1] Hung-Chih Lai, Marios Savvides, Tsuhan Chen, "Proposed FPGA hardware architecture for high frame rate (>100 fps) face detection using feature cascade classifiers," Biometrics: Theory, Applications,and Systems, pp.1-6, 2007.

[2] Chun He et al., Alexandros Papakonstantinou, “A Novel SoC Architecture on FPGA for Ultra Fast Face Detection”, Computer Design, 2009. ICCD 2009.IEEE International Conference on Issue, page: 412 – 418

[3] Junguk Cho, Ryan Kastner, Jason Oberg, and Ryan Kastner, “FPGAbased face fetection system using Haar classifiers,” International symposium on Field programmable gate arrays, 2009.

[4] ZHANG X. Y., ZHAO X.Y. and LI X., “Face Detection Based on Wavelet Trans form and Support Vector Machine”, Microcomputer Information, Vol.34, 2007

[5] JI Hu, SUN Ji-xiang, YAO Wei , “Wavelet moment for images”, Journal of Circuits and Systems, Vol.6, 2005