Real Time Face Detection Based on skin tone detector

Yuefeng Huang†, Xinyu Ao† and Yongping Li†

† Shanghai Institute of Applied Physics BioMetrics Laboratory
2019 Jialuo Road, Shanghai, China 201800

Summary

Face detection is critical to the application based on human face. Accurate and quick face detection has a determining effect on its following process, like face verification and face recognition etc. Adaboost algorithm [4] has been regarded as a very effective machine learning algorithm that can be applied to the face detection. However, with the increasing of image size determined by the advanced video or image capture device, the search space for finding face is incremented dramatically, to get these images processed in given time is becoming a challenge. In this paper, an algorithm combining skin tone detector and adaboost learning structure is put forward to push the state-of-art of the algorithm, it could reduce the number of window to be scanned to 20% of the conventional approach, thus can satisfy the real-time requirement in our related application where the facial images are captured from video stream of D1 format.

Key words:
Face Detection, Skin Tone, Adaboost, Machine Learning

1. Introduction

Face is the primary media for human interaction. It is the basic symbol of personal identity. It is natural for the face to be the primary candidate for human machine interface. Research attention around face pattern is focused on the development of *reliable face recognition algorithm* in recent decades. The National Institute of Standards and Technology (NIST) conducted Face Recognition Vendor Test (FRVT) [3] from 2000 to 2006, and from 2006, it started Face Recognition Grand Challenge (FRGC) [2] to search a robust face recognition algorithm. BioAPI [10] group pushes the effort to standardize the biometrics verification application programming interface to facilitate related industrial business opportunity.

As the necessary step for face recognition application, face detection algorithm has been a hotpot in the research of pattern recognition. Skin tone has been employed to dig out face area in the given image, plus some post processing to filter out the non-face skin-tone area. This approach is not satisfying for the outcome usually contains some false alarm result. Kanade and Schneiderman put forward an approach based on neural network [5], however the detection speed is very slow and the result is not good enough. In 2000, Viola and Jones put forward a real-time face detection scheme; it gave a satisfying performance for some mid-size image. However, in the up-to-date application, like video surveillance, the data to be processed for face detection becomes bigger and bigger. Currently, the size of the output video format is D1. Viola algorithm usually requires to scan the whole image in different size window, and to test them if these areas could pass their layered classifiers built by adaboost during training phase. Their speed cannot be fulfilled the real time requirement of related application for the size above, especially in the embedded system, where computing resource is scare.

To solve the speed problem in the scenario mentioned above, an obvious approach to attack the problem is to combine skin-tone filter in given image usually occupies small percentage of whole area. Thus, if non-skin tone area could be filtered out before being processed by Viola's Haar face detector, the amount of related computation will be greatly reduced, and hopefully, the face detector could reach the real time processing requirement. However, to determine a fitting threshold for given image is not simple. With the change of content, the percentage of skin-tone in the image is also changing. Therefore a global threshold for all images is not feasible whilst a step to determine the threshold is necessary. Based on this idea, skin tone percentage index is put forward in this paper to reach the goal; it has been integrated to the layered adaboost classifier. Experimental result is good enough for related application.

In following section, a theory based on color space will be illustrated, the related experiments and results are

Manuscript received July 5, 2009
Manuscript revised July 20, 2009

described in section 3, and finally, the discussion and summary will be drawn in the last section based on the experimental results.

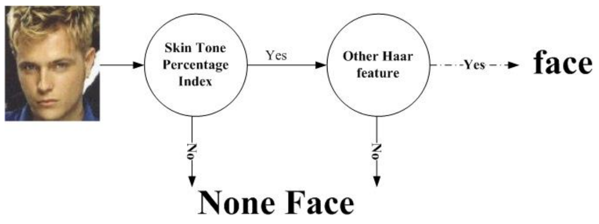

Fig 1. Skin tone percentage index integrated face detecting scheme

2 Theory and related Algorithm

2.1 Color Space

There are lots of color spaces in usage today, like YUV, RGB, HSV, Normalized RGB, YIQ, YES, CIE-Lab and CIE-Luv etc. For the simplicity of computation, YUV space is chosen, also, the output of raw data from the industrial CCD and video encoder is of YUV format, no extra transferring computation is needed anymore. At the same time, YUV space separates the luminance space from chroma space.

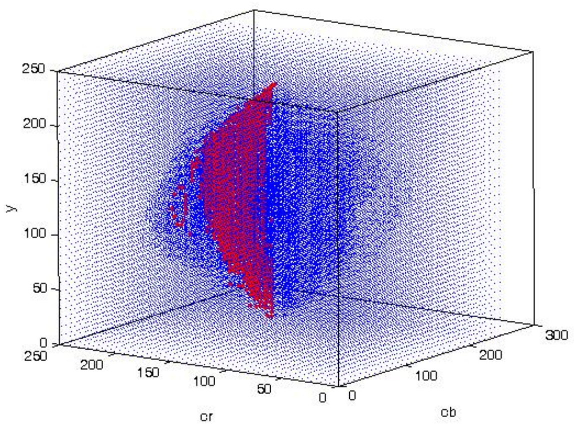

Fig 2. Skin-tone pixels' distribution in YUV space

2.2 skin tone

Human, that lives together, usually shares the same skin tone. And, lots of algorithms take this advantage to determine if there is a person in given image. Although people from different ethnicities many have different skin colors in appearance, experiments have shown that skin colors of individuals cluster closely in the color space, i.e. color appearances in human faces differ more in intensity rather than in chrominance. Therefore, discarding intensity value Y of the original color space and working in the chromatic color space (Cb,Cr) introduces some invariance against lighting conditions. In Fig2, it is illustrated how the skin-tone distributed in YUV space. There are several algorithms proposed for skin color pixel classification. They include piecewise linear classifiers, the Bayesian classifier with the histogram technique, Gaussian classifiers, and the Multi-layer perception [11]. The decision boundaries of these classifiers range from simple shapes (e.g., rectangle and ellipse) to complex parametric and nonparametric forms. Because the skin detection strategy is used only as a preprocessing step to face detection, computational requirement is extremely tense .As the same as reference [1], we simply uses rectangular boundary Classifier with the Cb value of range [77, 127] and Cr value of range [133, 173].

2.3 Skin tone percentage index

2.3.1 Quality factor of skin tone segmentation

In a simple model of image formation the image color C(x) is regarded as the product of object reflectance R(x) and the illuminance L(x) at each point x = (x, y). Additionally the camera influence can be modeled by a gain factor g and a bias term b, which are assumed to be constant on the image plane. Thus a simple image formation model is

$$C(x) = gL(x)R(x) + b \qquad (1)$$

In the same way, skin tone detection can also express as equation 2, B(x) belongs to {0,1},and Fs is skin classifier function.

$$B(x) = Fs(C(x)) = Fs(gL(x)R(x) + b) \qquad (2)$$

Robust skin detection systems should only be based on the object color which is conveyed by its reflectance properties. But without any knowledge or assumptions on the illumination field L(x) determining the object's reflection field R(x) is an ill posed problem. A popular assumption on L(x) and R(x) is that they vary only smoothly in x. Based on the assumption we consider the lighting parameters and object reflectance properties (color) to be almost same in a small 3×3 neighborhood, L(x) = L, R(x) = R. when pixel i(x) is classified as a skin pixel, R(x) may be a true or similar skin point, but also a pseudo skin point (noise). If R(x) belongs to the former, B(x)'s 3×3 neighborhood pixels are almost true or similar skin points, otherwise not. The other way round, as to a specifically skin classifier function (Fs), the better form the image, the compacter are skin pixels distributed, or else the discreter. So at point X = (x, y), the compactness degree of B(x) can basically represent the Quality of skin tone segmentation in equation 2, and we define it as $Q_f(x)$. Obviously, when pixel x belongs to a pseudo or noise skin point, $Q_f(x)$ should nearly equal 1/9. whereas, Q_f of a similar skin point is more than 1/9 and less than 1.0, Q_f of a true skin point is almost 1.0.therefor,the global Q_f of whole image is a weighted sum of all three kinds of skin points, it vary from 1/9 to 1.0 with all effect factors in the process from image forming to skin tone segmenting. At last, the global Q_f effectively reflects the whole Quality of skin tone segmentation and considered as a threshold estimate of skin tone percentage.

2.3.2 Calculation and use of skin tone percentage threshold

Based on the YUV space, skin tone percentage index is put forward to filter out the non skin tone face area. To integrate with the adaboost algorithm, it could be treated as the first weak classifier in the adaboost classifier chain, as shown in Figure 1. The skin tone filter is not used to locate the human face in image; instead it is used to filter out the non-skin area in the image. Through that, the searching space is reduced, this is similar with the work of Peng [7].

For the construction of first filter, the image is filtered by ordinary skin tone filter, after that a binary image will be gotten. By applying equation 3, the first filter will compute the sum matrix for the binary image. Suppose the binary image I(x, y), its value belongs to {0, 1}, 1 <= x <= N ,1<= Y <= M , the 3 × 3 neighboring area, including the point itself, The way to compute the sum matrix is to count the neighboring area's binary image value. so, the possible value for $S_n(x,y)$ would be in the set of

0	0	0	0	0	0	0
0	0	1	1	1	0	0
0	0	1	1	0	1	0
0	1	1	0	1	0	0
0	0	1	0	0	0	0
0	0	0	0	0	0	0

Fig 3. Binary image filtered by skin tone

0	1	2	3	2	1	0
0	2	4	5	4	2	1
1	4	6	7	5	3	1
1	4	5	5	3	2	1
1	2	3	2	1	1	0
0	1	1	1	0	0	0

Fig 4. Sum image of 3x3 neighboring area of fig 3

{0,1, 2, 3, 4, 5, 6, 7, 8, 9}, Suppose the binary image is shown as figure 3, the $S_n(x,y)$ image should be the image shown in figure 4. According to percentage of 1,2 ,3 respectively in $S_n(x,y)$, τ may be 1,2 or 3,for example, if percentage of 1 is much, τ must be 1, or else 2 or 3.

$$S_n(x,y) = \sum_{i,j=-1}^{1} I(x+i, y+j) \qquad (3)$$

$$S(x,y) = \begin{cases} 1 & if\ S_n(x,y) \ge \tau \\ 0 & \end{cases} \qquad (4)$$

$$Q_f = \frac{\sum_{x,y=1}^{x=N,y=M} I(x,y)}{\sum_{x,y=1}^{x=N,y=M} S(x,y)} \qquad (5)$$

From the above equation, the Q_f could serve as an indicator of the distribution of skin tone in the image, if it is scattered, like noise, its value will be relatively low.

In order to skip over the noise and background regions, we use an integral image introduced by Viola and Jones for real-time object detection [6] to rapidly compute sum of skin color pixels within any rectangle, and only perform face detection on the candidate skin color regions where the ratios of skin color to area of the region are above a threshold. The integral image I_s is an intermediate representation for an image I, and it contains the sum of pixels of the upright rectangle ranging from the top left corner at (0, 0) to the bottom right corner at (x, y) :

$$I_s(x,y)=\sum_{i\le x,\, j\le y} I(i,j) \tag{6}$$

The integral image can be recursively computed by

$$I_s(x,y)=I_s(x,y-1)+I_s(x,y)+I_s(x-1,y)+I_s(x-1,y-1) \tag{7}$$

$$where \quad I_s(-1,y)=I_s(x,-1)=0.$$

From this the pixel sum of any upright rectangle r= (x,y,w,h) can be determined by four table lookups:

$$S(r)=I_s(x,y)+I_s(x+w,y+h)-I_s(x+w,y)-I_s(x,y+h) \tag{8}$$

In our application, skin color image is a binary format. That is, I(i, j) = 1 if a pixel at (i, j) belongs to skin color regions; otherwise it is equal to zero. Suppose the current scanned window Wi, its dimension is of W × H, let Si be the Wi skin tone pixel summation. It could be computed according to equation 9.

$$S_i=I_s(x,y)-I_s(x+W,y)-I_s(x,y+H)+I_s(x+W,y+H) \tag{9}$$

The Wi skin tone ratio Pi could be computed according to equation 10. Figure 5 to Figure 6 illustrate the phase of computing skin tone percentage index. Figure 5.a and 6.a are origin images ; 5.b and 6.b are binary images after skin tone filter;5.c and 6.c are S image after using equation 4 on $S_n(x,y)$.

a b c

Fig. 5 better image with good lighting, Qf=0.73:a)origin image; b)binary image after skin segmentation; c)binary S image

a b c

Fig 6 image with shady lighting, Qf=0.41 a)origin image; b)binary image after skin segmentation; c)binary S image

During the face detection process, only the scanned window whose index Pi is larger than the Q_f, can enter the second stage. Generally this filter should reduce the windows passed the adaboost classifier shown in figure 1 to 10 ~ 20%.

$$P_i=\frac{S_i}{W\times H} \tag{10}$$

2.4 Algorithm computation analysis

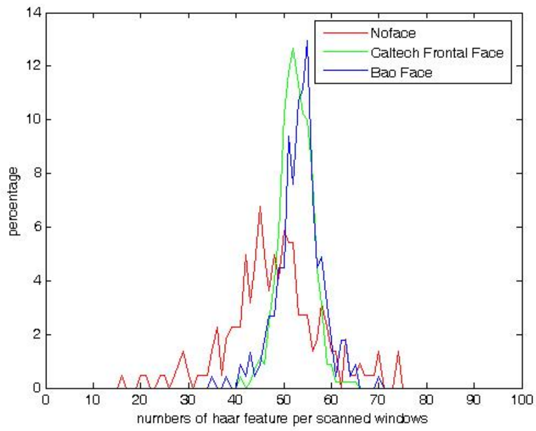

Fig 7 statistical distribution about numbers of haar feature evaluated per scanned windows

From figure 1, we add a skin filter ahead of the general adaboost, it seems to make the algorithm much more complex. Experimental results prove it not true. We make a statistical test about numbers of haar feature evaluated per scanned windows on three static image databases. The first database contains 220 nature landscape images, namely Non-face database; the second is Caltech Frontal Face Database [9] include only one face; the last one is BAO FACE Database [8] containing multiple frontal and non-frontal faces. Figure7 show the correlative experiment result. over 80% windows must evaluate 40~60 haar features, and at least 15 haar features. Evaluating a haar feature needs to calculate two or multi- rectangle pixel summations according to equation 9. In practice, in order to extract features robust to lighting variations, lighting correction are generally used on the candidate image windows before feature extraction. Lighting correction methods based on variance normalization, become time-consuming because of their application to each candidate window in the input image. However, our first skin filter only calculate one rectangle's pixel sum, Moreover, the first cascade can filter 80~90% scanned windows. so its computation much less than general adaboost cascade, and hugely accelerate face detection speed.

3 Experiments and Outcome

The testing face collection comes from two face detection database, BAODATA including 99 face images and 200 face images from Caltech Frontal Face Database database. This database contains various face image, these images contains different number of person in the same image, different pose, lighting, and different expression face image.

These face images are passed to the improved face detector as shown in Figure 1. Comparing with the Viola's implementation, our method reduced the false detection region, and speed the detection rate by 4-5 times. As shown in Figure 8 and Figure 9. Figure 8 is detected with ordinary adaboost detector, and the improved detector is applied in figure 9. The improved face detector reduces the false detection from 2 to 1; and the overall summary is shown in Table 1, the purposed row represent the skin tone percentage. The adaboost row represents the general adaboost of Viola. From the table, the detection rate is almost unchanged but the false alarm is reduced by 7.3%.

Table 1. results of two face detector

Method	Skin process time	Adaboost time	False alarm	rate
Puroposed	7.25ms	37.88ms	1.22%	92.58%
Adaboost	Not applicable	204.80ms	8.53%	93.92%

To further illustrate the effect of floating skin-tone threshold, an experiment is done, the control group consists of choosing the threshold from 0.1 to 1.0, with interval 0.1; also the condition without skin-tone filter is also included to conduct a full review of all the condition. To show both the detection correct percentage and the speed of detection, a variable CPSI is defined as equation 11.

$$CPSI = \left(\frac{C_A - C_w}{2C_A} + \frac{T_s}{2T_n}\right) \qquad (11)$$

C_A represents the number of human face in the image, and C_w represents the correct detected face number; T_s represents the time required to detect the face with skin tone, and T_n represents the time without skin tone. The value of CPSI should be the lower, the better.

The results of the above experiment is shown in figure 10, as we could see, the line of floating threshold stays lowest in the figure 10, indicating the combination factor of speed and correct detection achieved a better performance than the other 10 conditions with our 17 tests. However, it is still not clear on how the speed and effective detection behave. A 2-dimension vector combining the speed of detection and effectiveness is used to represent the mean value of the two components.

Fig 8. Face detection with only adaboost

Fig 9. Face detection with combination of adaboost and skin tone filter

To illustrate it clearly, vector combining of the speed of detection and the correct detection rate is used for all the conditions, the results is shown in the figure 11. The value is the mean value of all instances on 17 example images. The right and upper corner indicates the speed and effectiveness is of the ideal case, since its speed is faster and the effectiveness is better. As indicated in the figure 11, the floating threshold condition stays rightmost, however, its speed is not highest comparing with condition of static threshold 0.9 to 0.5, however, all the condition's detection rate is 20% worse than the floating condition.

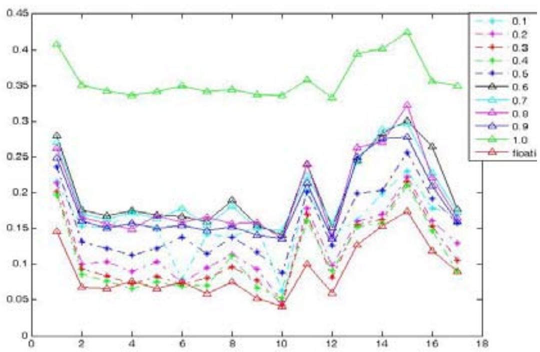

Fig 10. CPSI for different fixed skin-tone filtering threshold from 0.1 to 1.0 and floating threshold

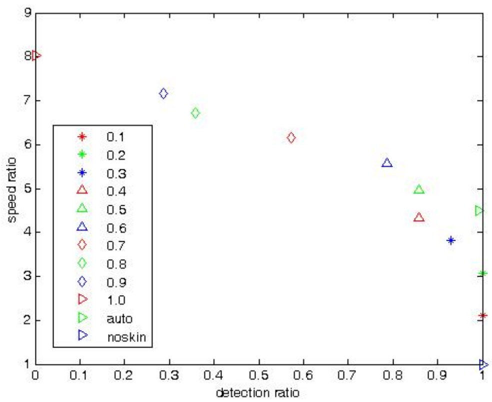

Fig 11. Average performance of fixed skin-tone filtering threshold from 0.1 to 1.0 and floating threshold

4 Conclusion

In this paper, a novel skin tone filter based on percentage index is proposed. It is implemented as the first weak classifier in the adaboost classifier chain. The algorithm first filters the image using general skin tone, then, after getting the binary image, the index is computed as described above. The algorithm designed in the above manner is to consider the implementation environmental constrains and the non-stop increment of the image size. Through experiments we can see, the speed are increased by 4-5 times and detecting performance no impact. And, it has been implemented in TI DM642 based platform, a scarce environment.

Acknowledgment

The authors would like to express their cordial thanks the support of Plan of Hundreds, Chinese Academy of Sciences

References

[1] D. Chai and K. N. Ngan. Face segmentation using skin-color map in videophone applications. IEEE Transaction on Circuits System and Video Technology, 9:551–564, 1999.

[2] T. S. P. Phillips, P. Flynn2 and K. B. etc. Overview of the face recognition grand challenge. In IEEE Conference on Computer Vision and Pattern Recognition, 2005.

[3] R. M. P.J. Grother and P. J. Phillips. Face recognition vendor test 2002 performance metrics. In Proceedings 4th International Conference on Audio Visual Based Person Authentication, 2003.

[4] R. E. Schapire, Y. Freund, P. Bartlett, and W. S. Lee. Boosting the margin: a new explanation for the effectiveness of voting methods. The Annals of Statistics, 26:322–330, 1998.

[5] H. Schneiderman and T. Kanade. Probabilistic modeling of local appearance and spatial relationships for object recognition. In Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 1998.

[6] P. Viola and M. Jones. Robust real-time object detection. In International Journal of Computer Vision, 2001. C. M. Harris, Handbook of Noise Control, McGraw-Hill, New York, 1978.

[7] Y. Peng, Y. Jin, etc Color Model Based Real-Time Face Detection with adaboost in color image, Issue , 28-29 Dec. International Conference on Machine Vision, 2007

[8] R. Frischholz. Bao face database at the face detection homepage.Available: http://www.facedetection.com.

[9] Weber, M., Frontal Face Dataset, California Institute of Technology,http://www.vision.caltech.edu/html-files/archive.html, 1999.

[10] ISO/IEC 19784-1, Information Technology – BioAPI–Biometric Application Programming Interface – Part 1: BioAPI Specification,2004

[11]P. Kakumanu ,etc. A survey of skin-color modeling and detection methods, Pattern Recognition 40 (2007) 1106 – 1122

Yuefeng Huang is a on-job doctorate of Graduate school of Chinese Academy of Sciences while working as electronic engineer. Her research interests are digital signal processing, biometric application and embedded system design.

Xinyu Ao received MS degree in the Electrical Engineering in 2003 from Vanderbilt University. As assistant supervisor at Shanghai Institute of Applied Physics, Chinese Academy of Sciences. His current research interests include biometric algorithms for face and fingerprint recognition, biometrics system performance evaluation, testing, and standardization.

Yongping Li received MS degree in the Electrical Engineering in 1989 from the Graduate University of Chinese Academy of Sciences. He worked as research fellow at Shanghai Institute of Applied Physics for various research projects on computerized instrumentations from 1989 to 1995. He joined the Center for Vision Speech and Signal Processing, Univ. of Surrey, UK in 1996 with a Royal Society fellowship, and received PhD in Pattern Recognition, with emphasis on Biometrics, in 2000 under the supervision of Professor Josef Kittler. He worked on speech signal processing at a startup company in Silicon Valley, California from 2001 to 2002. From 2003 to present, he is a research professor and PhD student supervisor at Shanghai Institute of Applied Physics, and a part-time PhD student supervisor at the Center for Biometrics and Security Research (CBSR), Institute of Automation, Chinese Academy of Sciences. His current research interests include biometric algorithms for face and fingerprint recognition, biometric solution integration, biometrics system performance evaluation, testing, and standardization. Dr. Li is a member of the IEEE, IEEE computer society and IEEE instrumentation and measurement society.