Color system for skin color extraction

Naruaki TOMA
Dept. of Information Engineering
Univ. of the Ryukyus
1 senbaru, Nishihara, Okinawa
903-0213,JAPAN

Emi TAMAKI
Dept. of Intelligent Interaction Tech.
Univ. of Tsukuba
1-1-1 Tennodai, Tsukuba-shi Ibaraki-ken
305-8577, JAPAN

Abstract

This paper describes a skin color extraction method, which comes with high-speed and -accuracy. There were many typical systems to achieve the requirements, however, they were useless due to tough limitations such as using two cameras to recognize correct coordinates, running only a specific background, wearing some gears, learning the users detailed information in advance, making personalized database and so on. In contrast, our approach which used YIQ color system with a background estimation method for skin color extraction, achieved well performance more than 30 fps with strong robustness on any-complicated and -illuminated environments.

1 Introduction

This research concerns an intuitively useable and high efficiency interface system, especially be focusing a system based on image processing techniques (vision-based interface).

Recently, vision based interfaces that they were introduced any sensor(s) to use user's behaviors as inputs into interface system, were studied [1, 2, 3, 4]. These type systems give friendly and efficiently interface for specialized situation. However, their systems don't use in general because of strict requirements, for example, they require two cameras to recognize correct coordinates, to set a specific background to recognize correct hand position, to wear some gears (e.g., marker tracking equipments), to learn the users detailed information (e.g., their color skin, finger form) and make personalized database.

This paper describes a skin color extraction method, which comes with high-speed and -accuracy, based on YIQ color system with a background estimation method to adjust the extraction parameters. We preliminarily evaluated the performances compared with other color systems (RGB, HSV and HQV [5]), and our method showed well performance (more than 30 fps) with strong robustness (more than 90% accuracy) on any-complicated and -illuminated environments.

2 Skin color extraction

2.1 Basic idea

Our supposed environment requires only one color movie camera, which must be able to acquire each image with high speed (more than 30 fps). Note that, we don't use any kinds of other sensors, don't care any backgrounds and illuminated environments, don't use any database (don't need to learn detailed information in advance).

Each captured images are processed through 1st stage: translating the color space from RGB space to the target space, then 2nd stage: calculating the thresholds to extract skin color. We use YIQ color system in 1st stage, and, if necessary, we use a background estimation method to adjust the thresholds automatically.

2.2 HSV and HQV

Some color systems were used for the extraction, and HSV and HQV were typical color systems. HSV based approach showed very low (fast) calculation costs and better extraction accuracy than RGB one. HQV (and also modified HQV) were proposed to improve the accuracy and it was achieved some studies (see [5]), but the calculation costs were too large to use for real-time processing like an interface system.

2.3 YIQ color system

YIQ is a color space, formerly used in the NTSC television standard. The YIQ color system is intended to take advantage of human color-response characteristics. The eye is more sensitive to changes in the orange-blue (I) range than in the purple-green range

(Q), therefore less bandwidth is required for Q than for I.

There are two stages to translate from RGB space into YIQ space, gamma correction is performed as equation 1, and then YIQ space was given by equation 2.

$$R' = \begin{cases} 4.5 \times R & R < 0.018 \\ 1.099 \times (R^{0.45}) - 0.099 & R \geq 0.018 \end{cases}$$

$$G' = \begin{cases} 4.5 \times G & G < 0.018 \\ 1.099 \times (G^{0.45}) - 0.099 & G \geq 0.018 \end{cases}$$

$$B' = \begin{cases} 4.5 \times B & B < 0.018 \\ 1.099 \times (B^{0.45}) - 0.099 & B \geq 0.018 \end{cases} \quad (1)$$

$$\begin{aligned} Y &= 0.299 \times R' + 0.587 \times G' + 0.144 \times B' \\ I &= 0.596 \times R' - 0.274 \times G' - 0.322 \times B' \\ Q &= 0.211 \times R' - 0.522 \times G' + 0.311 \times B' \quad (2) \end{aligned}$$

2.4 Filtering with thresholds

Thresholds on each color axis in the color space must be calculated in order to extract skin color. In other words, we assign a color zone on each axis to measure the pixel whether skin color or not.

2.5 Background estimation

When the background image or color is fixed like a blue screen, we can extract our desired objectives easily. But such extraction methods aren't functionable in real world (complex background), because the methods require really tough settings on each unknown environments such as different illuminated ones or really complicated places in advance. In order to decrease user's preliminary settings, we use a background estimation method to adjust the thresholds for extraction automatically.

The background estimation calculates a distance on the color space between current image and a few flames previous image. In Exactly, all distances on each pixel are calculated by formula 3, where $\overrightarrow{\mathrm{D}}$ is 3-dimentional vectors on the color space, $\overrightarrow{\mathrm{M}}(n)$ is movie image at the time n, and the i is a pixel position.

$$\overrightarrow{\mathrm{D}} = \mid \overrightarrow{\mathrm{M}}(n)i - \overrightarrow{\mathrm{M}}(n-1)i \mid \quad (3)$$

If the $\overrightarrow{\mathrm{D}}$ has small values, the estimated background are calculated by formula 4, where $\overrightarrow{\mathrm{P}}$ is the estimated background image and α is adding ratio of current image $\overrightarrow{\mathrm{M}}$.

$$\overrightarrow{\mathrm{P}}(n+1)i = \alpha \times \overrightarrow{\mathrm{M}}(n)i + (1-\alpha) \times \overrightarrow{\mathrm{P}}(n)i \quad (4)$$

2.6 Adjusting thresholds

Our system decides a set of thresholds based on the estimated background image described above section. The thresholds are calculated by formula 5, where $M(R, G, B)$ is a set of average (mean) values, F is translation from RGB to other color system, $D(R, G, B)$ is a dispersion values and F is a constant value.

$$Threshold = F(M(R, G, B)) \pm \beta \times D(R, G, B) \quad (5)$$

3 Experiments

For evaluation of proposed method, we performed one preliminary experiment and four experiments with numerical evaluation, which are compared with traditional color skin extraction methods (based on RGB, HSV and HQV).

We performed some experiments on table 1, and our system was written in Processing language[6] and JMyron library[7] to process movie images.

Table 1: Hardwares.

Video camera	iSight1.0.3
OS	MacOSX 10.4.4
CPU	Dual 2GHz, PowerPC G5
Memory	4GB DDR SDRAM

3.1 Exp1: Background estimation

Left side of figure 1 shows an actual image, which has hidden background parts with fingers, and right side shows the estimated image. It is enough result to decide the set of thresholds, because we use the meanscore only calculated from the image (see also formula 3).

3.2 Exp2: Computational speed

Experiment 2 evaluated computational speeds to translate of RGB to each color system only, here, we

Figure 1: An estimated background result (left: actual image, right:estimated image).

don't care about the accuracy of skin color. In figure 2, *Base* denotes a computational time to get and to display a movie image with Processing. That shows a basic speed without any calculations.

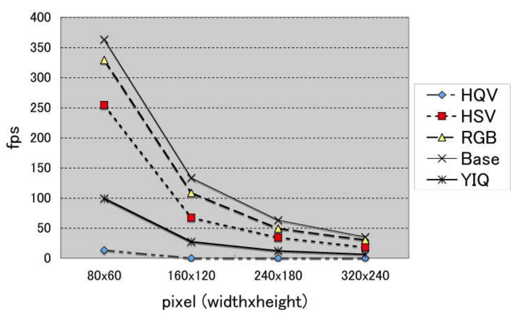

Figure 2: Computational speed for color translation from RGB into the target color space.

As the results, HSV was fastest except for RGB, and YIQ also was fast enough to achieve 30 fps with 160x120 (19,200 pixels) processing. HQV was too slow to use real time processing however it was evaluated as fine color system.

3.3 Exp3: Extraction accuracy 1

Experiment 3 evaluated the extraction accuracy on several backgrounds: *Black* background, *PC* background experimented on keyboard of note-pc, *Table* background experimented on the table with some objects, and *Complex* background experimented on really complicated situation. We measured three kinds of value. However, due to limitations of space, we showed one value 'total recognition-miss ratio' measured by 6, where N_{total} denotes the number of total pixels in one frame, and $\bar{N}_{total}$ denotes the number of total miss-counted pixels.

$$total\_miss\_ratio(\%) = \bar{N}_{total}/N_{total} \times 100 \quad (6)$$

Each recognition-miss ratios are shown in table 3.3, where YIQ_I denotes a result used I value only in YIQ color system. The thresholds on each color system were adjusted in order that recognition-miss ratios become lowest on *Black* background. Note that, the correct skin color pixels are decided by hand, so these results have a margin of error with a few percentages.

Table 2: Total-miss ratio on several backgrounds(%).

color	Black	PC	Table	Complex
RGB	1.69	23.55	56.62	50.41
HSV	2.22	40.08	63.18	67.19
HQV	6.88	6.78	55.42	52.83
YIQ	2.86	1.60	3.41	2.48
YIQ_I	5.12	3.61	5.00	2.88

According these results, the accuracies of RGB and HSV were through 8 to 61 percentages on complicated one. The other, YIQ showed very robust performances, which the accuracies were keeping with through 92 to 96 percentages even though the experiment performed in complicated environments.

3.4 Exp4: Extraction accuracy 2

Experiment 4 evaluated the extraction accuracy on several illuminated environments: *Fluor* illuminated with fluorescent lamp, *Incan* illuminated with incandescent lamp, *Sun(in)* illuminated indoor with sunlight, and *Sun(out)* illuminated outdoor with sunlight. In this experiment, the background was fixed on *PC* environments.

Table 3: Total-miss ratio on several illuminated environments(%).

color	Fluor	Incan	Sun(in)	Sun(out)
RGB	52.99	78.74	62.82	8.22
HSV	54.68	74.43	67.74	29.08
HQV	32.88	90.17	56.37	65.35
YIQ	3.14	94.40	2.17	13.33
YIQ_I	3.15	94.75	2.03	13.33

The results (see table 3.4), showed all kinds of color system didn't have enough accuracy on incandescent lamp. One of the reasons, I axis of YIQ has similar characteristics as the light of incandescent lamp even though I axis is most important axis to extract skin color.

3.5 Exp5: Extraction accuracy 3

To resolve the previous problem (weak recognition rate on incandescent lamp), background estimation method described in section 2.5 was introduced. Here, we don't experiment with RGB, HSV and HQV because they don't have robustness against the difference of illumination.

Tables 3.5 showed the accuracies with background estimation, where *withoutBS* denotes the results without background estimation (same values on experiment 4), and *withBS* denotes with background estimation. Figure 3 showed a extracted result without BS in leftside and with BS in rightside, where black color in hand part shows skin-miss pixels, dark light grey around of hand shows background-miss pixels and dark grey mainly showed in hand part of rightside shows correct recognized pixels.

Table 4: Total-miss ratio with background estimation(%).

color	without BS	with BS
YIQ	94.40	2.40
YIQ_I	94.75	2.17

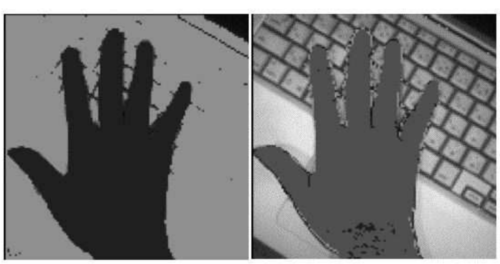

Figure 3: An extracted result (leftside with BS, rightside with BS).

The accuracy of YIQ with proposed threshold adjusting based on background estimation was pretty improved through 97 to 98 percentages. These results shows that our approach has highly robustness not only complicated backgrounds but also several illuminated environments.

4 Conclusion

In case of using only extraction based color system, the accuracy on all color systems were too low to use for practical. The accuracy of YIQ with proposed threshold adjusting based on background estimation was pretty improved through 97 to 98 percentages.

In summary, the proposed method showed well performance with strong robustness for skin color extraction on any-complicated and -illuminated environments. According to this summary, this method so as to apply widely applications because our method requires only one camera and without any personalized database.

References

[1] Nozomu Matsui and Yoshikazu Yamamoto, "Virtual Keyboard with Single Video Camera (in japanese)", in Programming Symposium, 1999.

[2] Nozomu Matsui and Yoshikazu Yamamoto, "Virtual Keyboard: Realization of Robust Real-Time Fingertip Detection from Video-Image (in japanese)", in JSSST SPA2000, 2000.

[3] Yishiyuki Kojima, "Hand Interface for Wearable Mixed Reality - Clip Interface -", Master's Thesis, Department of Information Processing, Graduate School of Information Science, Nara Institute of Science and Technology, NAIST-IS-MT0051037, February 8, 2002.

[4] Takeshi Kurata, Takashi Okuma, Masakatsu Kourogi, and Katsuhiko Sakaue, "The Hand Mouse: GMM Hand Color Classification and Mean Shift Tracking", In Proc. Second International Workshop on Recognition, Analysis and Tracking of Faces and Gestures in Real-time Systems (RATFG-RTS 2001) in conjunction with ICCV 2001, pp.119-124.

[5] Satoshi matsuhashi, Osamu Nakamura and Toshi Minami, "Human-Face Extraction Using Modified HSV Color System and Personal Identification Through Facial Image Based on Isodensity Maps", IEEE CCGEI 1995, Vol 2, No 2, pp.909-912.

[6] Processing.org, http://www.processing.org/

[7] Myron WebCamXtra, http://webcamxtra.sourceforge.net/