Tracking and Surveillance
Lim Siong Boon
Supervisor, Dr Chua Chin Seng

**School of EEE, Nanyang Technological University,
Nanyang Avenue, Singapore 639798**

Abstract

A technique is introduced to track human movement. The proposed method uses the color of human's clothing to track their movement.

Human clothing can uniquely determine the identity of the person. The surface area available for detection is also larger than skin region. Therefore the chances of detecting rotational, deformed and occluded target also increase.

The color of the clothing can be determined with initial help from the skin color. The system learns the clothing color by monitoring the colors near the skin color over a few frames. After which skin color can be discard and tracking can be based mainly on clothing color.

The method is robust with regard to occlusion, non-stationary background, and random human movement.

1 Introduction

There are a lot of techniques used in object detection. The common detection techniques include Motion, Color, Texture, Shape, etc…. Each technique has their strength and weaknesses.

In this project, color is examined for its advantages in tracking application. From its properties, color has the ability of differentiating objects. Therefore it provides very useful information especially during the case of occlusion.

Color detection technique can be robust against non-stationary background. Therefore it is applicable in the case of mobile robot. Camera can be fixed on the robot to do tracking and navigation.

In Human tracking application, skin color is often use to signify the presence of human target. However, tracking skin color is unreliable due to the limited skin region available. The skin color can be easily block by objects or when the target is rotating. In this paper, the method to extracting the color of human clothing is discussed.

2 Color System

Color information is commonly represented in the widely used RGB coordinate system. This representation is hardware oriented and is suitable for acquisition or display devices but not particularly applicable in describing the perception of colors. On the other hand, the HSV (hue, saturation, value) color model corresponds more closely to the human perception of color. The hex-cone model shown in Fig 2-a conveniently represented HSV color space.

HSV domain allows color and brightness isolation. This is useful as brightness information is not useful in color object detection. The extra brightness information could be cause by external factor (external light source), therefore holds no meaning.
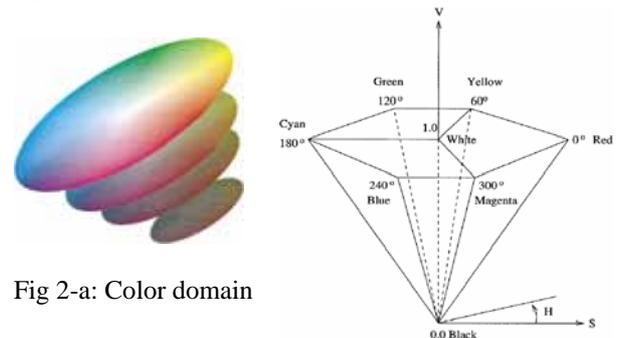

Fig 2-a: Color domain

2.1 Advantage using Color
- Color can be unique to an object
  -Object identity can be determine easily even after occlusion
- Color representation of object is robust against
  -Complex, deformed & changeable shape (human)
  -Orientation/rotational
  -Size (zoom in, out)
  -Partially blocked view (object)
- Color representation independent on resolution
  -Low resolution
  -Fast to process image
  -Low cost for the equipment
- Background can be dynamic
  -Camera can be mobile (fix on mobile robot)

2.2 Disadvantage using Color
- Object is not unique to the color
  -Affected by other object with same color as object targeted
- Color of object cannot be obtain precisely
  -Color is affected by illumination/brightness & low saturation (too bright looks like white, too dark looks like black)

-Environmental factor. Source of poor lightings can be from highlight; shadow, light spectrum emitted by different/ numerous lightings, daily variation due to sunlight and temperature
-Color balance in every camera is different
- Object color may not be distinctive
-Wide range of skin color (reddish, yellow, brown, white, black)

Depending on the application of color detection, some parameters can be controlled. Camera can be calibrated (white balanced) before operating and lightings source can be adjusted. Human skin color can be predefined.

There are also techniques to minimize the disadvantages listed. Threshold adjustment/adaptive filtering [03] [04], statistic estimation [06], neural network [08], white/black calibration.

3 Skin Color

Human skin is composed of several layers of tissue, which consist essentially of blood cells, and a yellow pigment called melanin. The appearance of the skin is affected by a number of factors, which include the degree of pigmentation (varies amongst individuals and different races), the concentration of blood, and the incident light source. The combination of these entire factors give rise to a variation in skin color which spans over the range of red, yellow and brownish-black. Nevertheless, this corresponds to a restricted range of hue values.

The three common skin hue distributions are conveniently summarized in Table 3-a [01].

Race	Mean, m	Standard Deviation, σ
Asian	28.9°	5.1°
African-American	8.6°	8.2°
Caucasian	25.3°	6.8°

Table 3-a: Statistic of Hue distribution

The tabulated values indicate that the Asian test samples have the highest mean value of the three distributions, m =28.9° (i.e. greater shift towards Yellow) with the lowest standard deviation, σ. The Caucasian sample set has similar statistics with a slightly smaller mean value, m =25.3° and a slightly larger value of σ. The African-American distribution has the smallest mean value of the three, m =8.6° (shift towards red) and the largest standard deviation. The large value in σ can be attributed to the variation in skin colors within the African-American sample set.

4 Methodology

4.1 Skin Color Extraction

The hue component is the most significant feature in defining the desired polyhedron (skin color). The hue values can be represented by a limited range 340-360°

(magenta-red) and 0-50° (red-yellow) for all skin types. This range is very effective in extracting skin colored regions under higher levels of illumination and sufficiently saturated colors. However, the hue can be unreliable when the following two conditions arise:

(1) When the level of brightness (i.e. value) in the scene is low or (2) when the regions under consideration have low saturation values.

The first condition can occur in areas of the image where there are shadows or, generally, under low lighting levels. In the second case, low values of saturation correspond to achromatic regions. Many objects, by nature, are achromatic (i.e. white clouds, gray asphalt roads, etc.), however, shadows or conditions of non-uniform illumination (i.e. speculate reflection) can cause chromatic regions such as skin areas to appear achromatic. Thus, we must define thresholds for the value and saturation components where the hue attribute is reliable.

For experimental reason, the following range is used to extract (mask out) skin color.
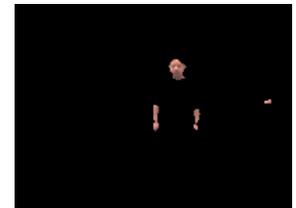
The success of object segmentation will greatly depend on the quality of the extracted Skin Color.

	Actual Range	Normalise 0-255
Hue:	1.4°-11.2°	1-8
Saturation:	27.3%-58.6%	70-150
Value:	40.2%-100%	103-256

Table 4-a: Experiment skin color threshold

| (a) Original | (b) HSV masked |

Fig 4-b: Masking

There will be a lot of false skin like color detected, due to the wide range of skin color to detect. By readjusting the threshold of skin color with reference to the detected skin, this error can be reduced [03] [04]. Therefore a more accurate skin color model could be obtained and extracted.
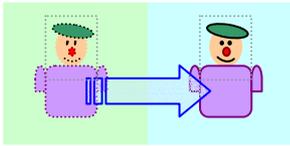
Blob process helps to identify object after successful skin color extraction. A blob consists of a group of adjacent pixels. Since they have similar color and are closely packed to each other, they are assume to be pixels of the same object.

Before a blob can be accurately identified, a dilution process can helps to eliminate possible false detection or undetected pixel. The process estimates the possible existence of a pixel from its surrounding pixels.

After obtaining the true skin target, identity will be assigned to each blob. The blob will correspond to a detected skin target.

4.2 Object Representation

After a target is found, a HSV (color) histogram data will be form to represent the target. The histogram data is collected within the box-up target. It defines the pixels that are associated with the blob (skin color pixels). The histogram relates the colors associated to the blob. It is use as an ID to differential between different targets.

Any color changes during movement implies that they belong to the background. The stabled color will be the color associated to the target.

Fig 4-c: Histogram ID updating process

The histogram obtained, does not represent target accurately due to the presents of background color. To extract the target's histogram (colors associated to target), a learning process is introduced. The idea is to monitor and update the colors that are always present within the boxed target. The color that is always present should be the blob (skin color) and it's clothing color. This is under the assumption that background color changes as the target moves.

With the target's histogram extracted it can then be use as ID to differential itself from other targets.

A thing to note is that the first target histogram (reference) is captured only when the whole target entered the scene. If only partial histogram is obtained, it will be treated as the initial histogram and makes the updating inaccurate. Two-target width is a possible distance to extraction the ID's color.

During typical scenario, a target is detected with 3 skin color parts, two hands and a head. The 3 ID comes with the same surrounding color of the clothing. Using this cue, the 3 ID can be merge together because they belongs to the same target. Distance information between IDs can be used to further confirm identity of the object. This algorithm is possible with assumption of low histogram resolution used (discuss in next section). Slight variation in hue value is still consider the same hue as they are rounded to the nearest hue.

Face detection algorithm could also be use instead since each target has only a face [02] [05].
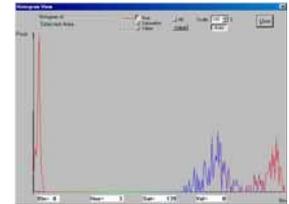
4.3 Analyzing Target

The target ID is determined from the peaks of the histogram. In the experiment, a sample skin color human head is found. The color associate around the head is form in the histogram shown in Fig 4-d (b). From the histogram, it can be clearly seen that the Hue histogram consist of mainly 3 colors only (red, blue/purple, pink). However the peak algorithm is unable to pick up the 3 main cluster of color. This is because the generated noise forms peaking, which cause false detection. To counter this, the histogram is formed using lower resolution. This will force the

neighboring weightage to their nearest bin (color representation). Therefore a noiseless histogram can be obtained Fig 4-d (c).
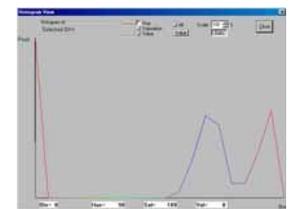
The disadvantage of using the reduce resolution is that the peak detected may slightly off from its actual hue. A better method would be pulling smaller neighbouring weightage to it's own bin [07]. The hue position can be maintained and noise is also removed.

(a) Target image

(b) Hue histogram of target (256 bins)

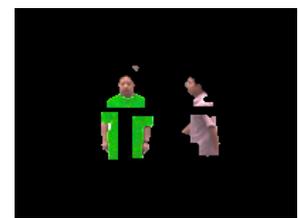(c) Hue histogram of target (20 bins)

Fig 4-d: Target histogram

4.4 Extracting Target's ID

Having the color associated with the target obtained, the target can now be extracted. The target's histogram is form by a few Gaussian distribution at various peaks Fig 4-d (c). This is so because environment color is form by gradual change in color. Therefore to extract the object related to the clustered color range, Gaussian distribution is assumed. In the experiment, the five hue's bin nearest from the peak will be use to determine the hue range use for extracting the colors.
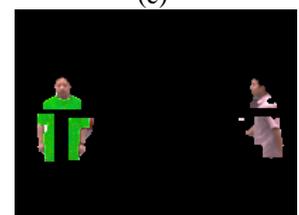
A more accurate approach would be to use mean and standard deviation to obtain the range of the cluster. This can produce a higher accuracy covering up to 90 over percent of the area.

(a)

(b)

Fig 4-e: Target extraction Histogram Updating Experiment (18th Dec 2001) Target Bin resolution: 20 Threshold: 20 pixel above, to consider Color as ID Extract 5 dominant colors for display.

(c)

Although the pink shirt is close to skin color, two colors have been detected due to two detected peaks.

4.5 Object Tracking

With the clothing information extracted, human tracking will be mainly base on clothing's color. It is easier to track clothing color than skin because of the large area available for detection. The chances of detecting rotational, deform, occluded target also increase due to the large area Fig 4-f.

Under occlusion, no more updating is to be done to object occlusion. They will be closely monitored until they are out of occlusion. Target's ID is use for comparing and differential between the occlusion targets when they are out from occlusion.

Due to the large tracking area (clothing color), it is possible to track the object under occlusion. This is under the assumption that the object cannot be totally blocked. A portion of its clothing color is expose throughout the occlusion. Texture can be also be looked into as another extension to color detection
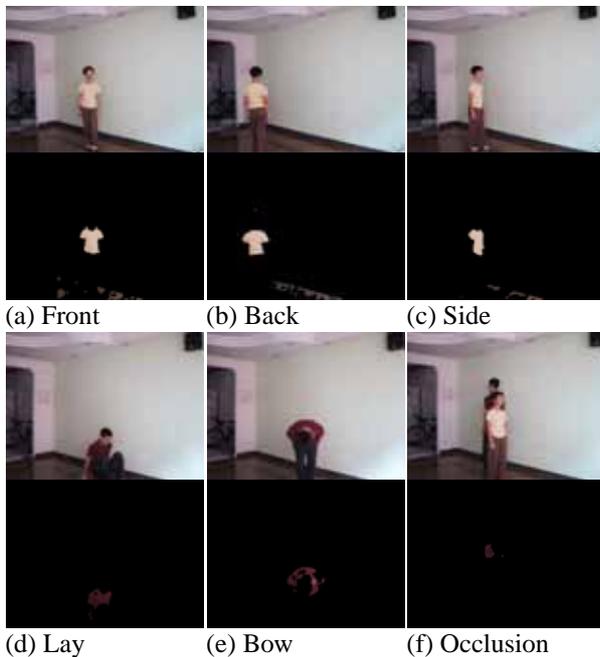

| (a) Front | (b) Back | (c) Side |
| (d) Lay | (e) Bow | (f) Occlusion |

Fig 4-f: Clothing Tracking

5 ASSUMPTION & PROBLEM

5.1 Assumption

- Object color is distinctive and not dynamic.
- Skin color Hue (Chinese race).
- Skin color is unique from background color.
- Clothing is of uniform color (can have different color but must be able to see them 360°).
- Humans present in frame have distinct color for identification (example clothing, things human carries).
- Assume human movement is small distance between adjacent frames captured.
- Color histogram obtain is in terms of guassian distribution

5.2 Problem Encountered

- Performance poor with homogenous background.
- Pants color cannot be always extracted.
- Color under tracking changes under different lighting condition. (Saturation value shifted up in darker environment)
- Not ready for white and black color detection

6 Conclusion

In this paper, the method to extracting the color of human clothing is discussed. The clothing color is use for tracking the human because of the high probability of detecting the target. The clothing colors are obtained by monitoring from the surrounding of skin region.

Simple experiments were conducted using Visual C++. The image processed is done under real time processing using sequential images in BMP file format.

The experiment results show that the proposed method is robust to rotational, deformed and occluded human targets. It is also robust under non-stationary background (moving camera).

7 Acknowledgement

I would like to thank my supervisor Dr Chua Chin Seng and Li Jiang for their assistance and suggestions.

8 References

[1] N. Herodotou, K.N. Plataniotis, A.N. Venetsanopoulos, Automatic location and tracking of the facial region in color video sequences, Signal Processing: Image Communication 14 (1999), pp.359-388.

[2] Yanjiang Wang, Baozong Yuan, A noval approach for human face detection from color images under complex background, Pattern Recognition 34 (2001), pp.1983-1992.

[3] Kyung-Min Cho, Jeong-Hun Jang, Ki-Sang Hong, Adaptive skin-color filter, Pattern Recognition 34 (2001), pp.1067-1073.

[4] Maricor Soriano, Birgitta Martinkuappi, Sami Huovinen, Skin detection in video under changing illumination conditions, Machine Vision & Media Processing Unit University of Oulu FINLAND, pp.839-842.

[5] Jeonghee Park, Jungwon Seo, Dongun An, Seongjong Chung, Detection of human faces using skin color and eyes, Computer Engineering Chonbuk National University South Korea, pp.133-136.

[6] Xiaojin Zhu Jie Yang Alex Waibel, Segmenting Hands of Arbitrary Color, School of Computer Science Carnegie Mellon University 2000 IEEE, pp.446-453.

[7] F. Kurugollu, B. Sankur, A.E. Harmanci, Color image segmentation using histogram multithresholding and fusion, Image and Vision Computing 17 (2001), pp.915-928.

[8] Son Lam Phung, Douglas Chai and Abdesselam Bouzerdoum, A universal and robust human skin color model using neural networks, Visual Information Processing Group Edith Cowan University 2001 IEEE, pp.2844-2849.