Applied Math. & Information Sciences Letters
*An International Journal*



# Driver's Face Tracking Based On Improved CAMShift for Drowsiness Detection

**Jie Ma[1] , Rui Xiao[1] and Kamarul Hawari Bin Ghazali[1]**

[1] Faculty of Electrical & Electronics Engineering, UNIVERSITI MALAYSIA PAHANG, 26600 Pekan, Malaysia

*Email: ma_jie@zoho.com*

**Abstract:** The statistic shows that the number of casualty increase in every year due to road accident related to driver drowsiness. After long journey or sleepless night, vehicle driver will perform some bio-features with regard to drowsiness on them face. It is self-evident that getting location information of face in continuous monitoring and surveillance system rapidly and accurately can help prevent many accidents, and consequently save money and reduce personal suffering. In this paper, according the real situation in vehicle, an improved CAMShift approach is proposed to tracking motion of driver's head. Results from experiment show the significant performance of proposed approach in driver's head tracking.

**Keywords:** Color Space, Face Tracking, CAMShift, Mean Shift, Probability Distribution Function

## 1 Introduction

With the increasing popularity of automobiles, the traffic accidents have become severe social problems. Recent statistic shows that the number of fatalities due to road accidents in Malaysia increased 1.9% to 6,872 deaths in 2010 from 6,745 deaths in 2009. In a separate online survey by the community project Malaysians Unite for Road Safety (MUFORS), 61.6% of the respondents believed that human error is the biggest cause for road carnages, while 15.6% of the respondents blamed road conditions [1]. Malaysia is paying a heavy price due to road accidents, and the cost to the economy.

Successfully addressing the issue of driver drowsiness in the commercial motor vehicle industry is a formidable and multi-faceted challenge. A lot of research and experiments have been done in the last decade. Driver drowsiness detection techniques can be broadly classified into three categories: physiological measurement (such as brain waves (EEG), eye movements (EOG) and heart rate (ECG)), visual cues (such as eyelid movement, facial orientation, as well as yawning) and driving performance (monitor how the driver handles the vehicle) [2]. The first two categories monitor the driver's bio-signals directly, whereas the third category monitors the driver indirectly. The first category have to be attached to driver, making this approach both intrusive and impractical, the last two categories have non-intrusive nature. By reason of its non-intrusive nature, considerable accuracy and fully portability, the visual cues are the most widely used technique to detect the driver fatigue.

The National Sleep Foundation suggests a list of signs that can be used for determining when a driver is no longer in conditions for being driving. These signs are the following:

- Difficulty focusing and daydreaming.
- Frequent blinking and heavy eyelids
- Trouble remembering the last few miles driven
- Frequent yawning or rubbing eyes
- Drifting from lane or tailgating
- Trouble keeping head up

From this list, we know eye is the key organ for driver fatigue detection. In many applications, such as face recognition, face analysis, eye detection, and eye tracking, the first step is usually face detection. The face detection techniques can be divided into two major categories. The first class is based on face features which utilize some well-known knowledge on human faces, for example the shape of face, the relative locations of the eyes, the nose, and the mouth in face. The second class is based on face colors which build on specific color models to locate face based on skin colors. The skin color has quite stable distribution in some color models.

Object tracking is a fundamental and an important problem of dynamically extracting two-dimensional (2D) information in most visual applications including image processing, computer vision, video surveillance, human-computer interaction (HCI). With the proliferation of high-powered personal computers and portable and low-cost video cameras has brought forth a large numbers of algorithms in object tracking. There are many different approaches for tracking an object proposed to overcome the difficulties arising from noise, similar color distribution, and complex background environment. The CAMShift tracking algorithms stands out as its simplicity and efficiency. CAMShift stands for Continuously Adaptive Mean Shift and it was first proposed by G. Bradski et al [3], aiming at efficient head and face tracking from a stationary camera in a perceptual user interface but has since been modified for a variety of other tracking situations[4][5].

The CAMShift algorithm was derived from the earlier Mean Shift algorithm [6] and is a simple, yet very effective, color-based tracking technique. It is proved by lot of proposed work. Bogdan kwolek [6] track human head and body respectively in order to tracking people. In the paper [7], color-histogram was presented for tracking; it is insensitive to small object pose change due to discard spatial information. Zhaowen Wang et al [8] used particle filter and CAMShift to Increase the robustness, but the calculation time of particle filter reduce the real-time performance. In the paper [9], David Exner, Erich Bruns et al had accomplished CAMShift efficiently on the GPU; all of the excellent performance makes full use of a GPU's high parallelization capabilities. That means need special hardware support.

This paper proposed an improved CamShift algorithm based on segmentation of skin color in H channel of VSH space to increase the robustness performance. Local back projection was proposed to saving the calculation times and increased the weight of the skin color pixel。 Due to captures a richer description of target, the improved CamShift algorithm increase robustness in complex background circumstance. The chart flow of this system can be described as following figure.

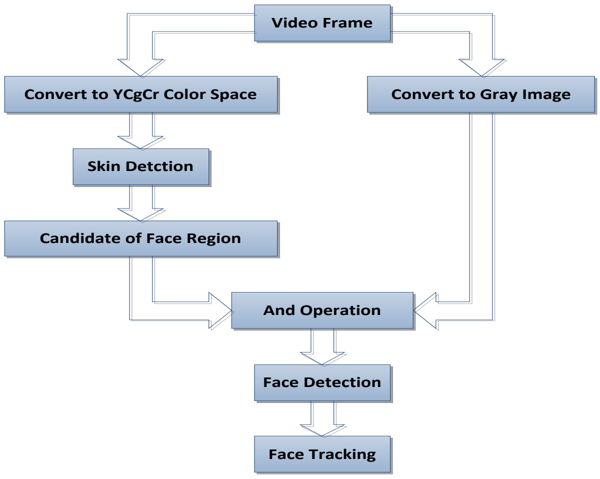


Figure 1: The flow chart of system

## 2 Algorithm Analysis

From the side of defects in CAMShift algorithm, it can help us to know how to make use of this algorithm properly. The defects of CAMShift algorithm are described as follows:

- The CAMShift algorithm is semiautomatic tracking algorithm. It needs be given the target object position information by other object detection algorithm in the first frame to do object tracking.
- The CAMShift algorithm is sensitive to color, so if the similar color distribution exist in background of target object. In the end, the CAMShift will lost the target object or locate wrong object.
- The CAMShift algorithm is suitable for uniform-colored object tracking, so it will fail in high probability when using to track multi-colored object.
- Although the CAMShift algorithm uses HSV color space, it has resistance against changing illumination. But when illumination changes intensively, this algorithm wills underperformance

When detecting driver's fatigue in the vehicle, it is usually to install a camera in front of the driver. In the captured video frame, the vehicle driver's hand or face is the biggest object comparing other object in the video frame. In other words there have seldom other objects to disturb the CAMSfit algorithm. Although driving car on the road, the background behind vehicle driver is almost stationary. Human being's face is uniform-colored distribution object. Light changing in vehicle can affect the performance of CAMShift, but the result can acceptable. Using only hue from the HSV color space gives CAMSHIFT wide lighting tolerance [16]. All the analysis shows that CAMShift is suitable to tracking deriver's face in vehicle.When detecting driver's fatigue in the vehicle, it is usually to install a camera in front of the driver. In the captured video frame, the vehicle driver's hand or face is the biggest object comparing other object in the video frame. In other words there have seldom other objects to disturb the CAMSfit algorithm. Although driving car on the road, the background behind vehicle driver is almost stationary. Human being's face is uniform-colored distribution object. Light changing in vehicle can affect the performance of CAMShift, but the result can acceptable. Using only hue from the HSV color space gives CAMSHIFT wide lighting tolerance [16]. All the analysis shows that CAMShift is suitable to tracking deriver's face in vehicle.

## 3 Methodology

The first step is face detection. After that, CAMShift was employed to tracking face. When using CAMShift algorithm to track object, the first step is that convert color video frame to HSV. After this step, the other steps can named as back projection, Mean Shift and CAMShift Procedure. All the steps will be present in detail in this paper.

### 3.1 Face Detection

#### A. Skin model in YCgCr Color Space

YCgCr color space was derived from YCbCr color space. In the YCbCr color space, Y is the luminance component and Cb and Cr are the blue-difference and red-difference chroma components. Considering the YCbCr color space, a human skin color model can be considered practically independent on the luminance and concentrate in a small region of the Cb-Cr plane. YCgCr color space also has the above advantage that is non-sensitive to the luminance [17].

The YCgCr color space is a variation of YCbCr color space that uses the color difference (G-Y) instead of (B-Y). YCgCr components can be derived from the following matrix expression:

$$\begin{bmatrix} Y \\ Cg \\ Cr \end{bmatrix} = \begin{bmatrix} 16 \\ 128 \\ 128 \end{bmatrix} + \begin{bmatrix} 65.481 & 128.553 & 24.996 \\ -81.095 & 112 & -30.915 \\ 112 & -93.786 & -18.214 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} \quad (1)$$

We found that a 2D projection of YCgCr skin color into Cg-Cr subspace, in which we find that the skin color clustering effect of Cg-Cr color space is better than that of Cb-Cr color space in Figure 2

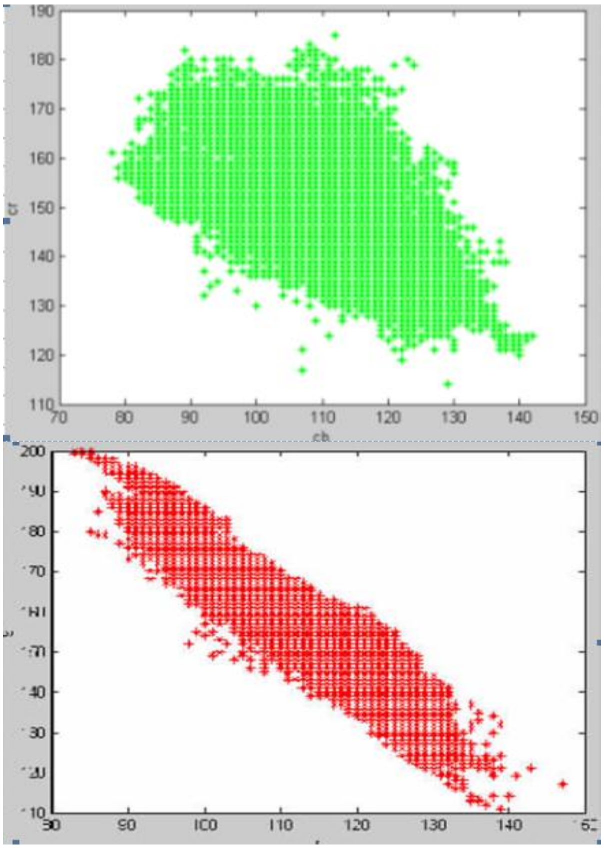


Figure 2: Skin color distribution in Cb-Cr (above) color space and Cg-Cr (below) color space.

In order to segment human skin regions from non-skin region based on color, we need a reliable skin color model that is adaptable to people of different skin colors and to different lighting conditions. The color distribution of skin colors of different people was found to be clustered in a small area of the chromatic color space. Although skin colors of different people appear to vary over a wide range, they differ much less in color than in brightness. In other words, skin colors of different people are very close, but they differ mainly in intensities. The color histogram revealed that the distribution of skin-color of different people is clustered in the chromatic color space and a skin color distribution can be represented by a Gaussian model. In our work, each colorful image is concerted into Cg-Cb color space, we can get $X = (Cg, Cb)^T$ ; $m_x = E(x)$ ; $C_x = E\left[(X - m_x)(X - m_x)^T\right]$ . Where $m_x$ is it's mean, and $c_x$ is the covariance matrix of x.

The video frame was converted into YCgCr color space. In the YCgCr color space, we use the following formula to get a black-white image.

$$\begin{cases} 100 < Cg < 130 \\ 135 < Cr < 175 \end{cases} \quad (2)$$

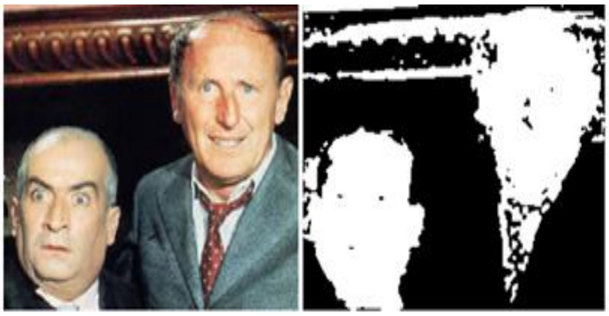


Figure 3: The original image and after Gaussian detection

### B. Viola-Jones Detector

In this paper, the face detection was accomplished based on the Viola-Jones objection frame proposed by viola et al [10]. In this appeared-based method, the algorithm is insensitive to color. So the system can get good performance even with lot of skin-color like decorations surrounding. The Viola-Jones framework is based on the idea of a boosted cascade of weak classifiers. Each classifier uses a set of Harr-like (figure 4) features, acting as a filter chain. Only those image regions that manage to pass through all the stages of the detector are considered as containing the face. For each stage in the cascade, a separate sub classifier is trained to detect almost all target objects while rejecting a certain fraction of those non-object patterns that have been incorrectly accepted by previous stage classifiers [11]. The procedure of processing can be sketched in follow picture 4. More details about the framework, please review relative literatures.

Figure 4: The feature prototype of simple haar-like

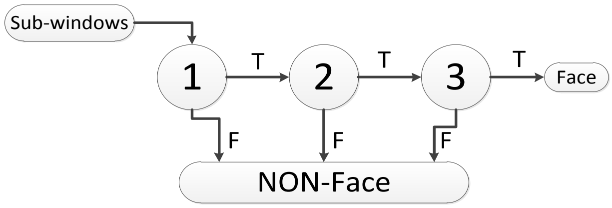



Figure 5: Cascade classifier

### 3.2 HSV Color Space

HSV are the two most common cylindrical-coordinate representations of points in an RGB color model, which rearrange the geometry of RGB in an attempt to be more perceptually relevant than the cartesian representation. HSV stands for hue, saturation, and value, and is also often called HSB (B for brightness) [13]. Hue is the attribute of a visual sensation according to which an area appears to similar to one of the perceived colors: red, yellow, green and blue, or to a combination of two of them. It can be described in terms of an angle on the above circle. Although a circle contains 360 degrees of rotation, the hue value is normalized to a range from 0 to 255, starting from red. Saturation measures the departure of a hue from achromatic, i.e., from white or gray. In other words, saturation represents the vibrancy of the color. Its value ranges from 0 to 255. The lower the saturation value, the more gray is present in the color, causing it to appear faded. Value measures the departure of a hue from black, the color of zero energy. It ranges from 0 to 255, with 0 being completely dark and 255 being fully bright. From figure 1, it can be explained more clearly.

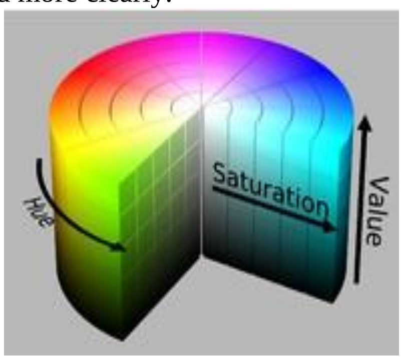



Figure 6: HSV color cylinder [13]

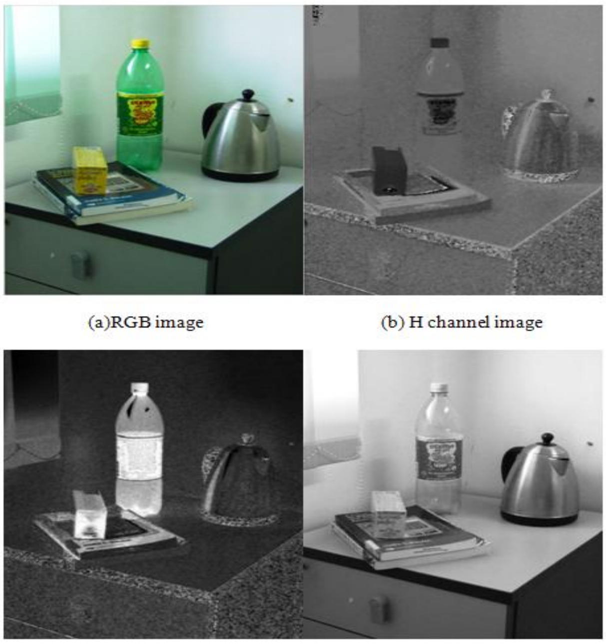
**RGB to HSV Algorithm** [14]

Given: R, G and B, each on domain [0, 1]

Desired: The equivalent H, S, and V, each on range [0, 1]

1. V: = max(R,G,B);
2. Let X: = min(R,G,B);
3. $S := \frac{V - X}{V}$; if S = 0 return;
4. Let $r := \frac{V - R}{V - X}$; $g := \frac{V - G}{V - X}$; $b := \frac{V - B}{V - X}$;
5. If R=V then H: =(if G=X then 5+b else 1-g);
   If G=V then H :=( if B=X then 1+r else 3-b);
   Else H = (if R=X then 3+g else 5-r);
6. $H := \frac{H}{6}$;

The image of H, S and V channels is as follows：



(a)RGB image (b) H channel image

(c) S channel image (d) V channel image

Figure 7: The image of H, S and V channels

### 3.3 Back Projections

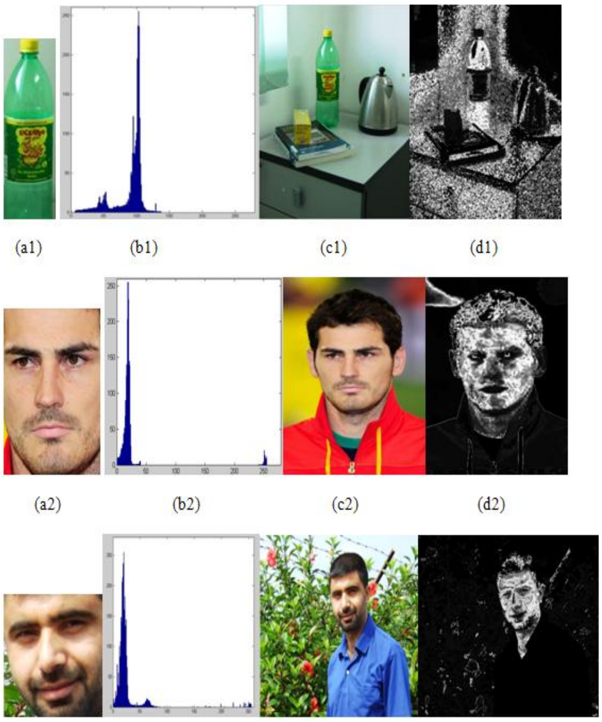
Back projection is a primitive operation that associates the probability of being part of the tracked object in the image of each pixel with the value of the calculated color histogram [10]. The histogram is calculated as：

$$q_u = \sum_{i=1}^{n} \delta[p(x_i) - u], \ u = 1 \ldots m. \quad (3)$$

Here δ is unit impulse function. The $p(x_i)$ associating the pixel at $x_i$ with the index of its color bin in the quantized feature space is the image histogram value of the elements in H channel. The formula of histogram normalization is as:

$$\bar{q}_u = \min\{\frac{255}{\max(q)} * q_u, 255\}, \text{ for } u = 1 \ldots m. \quad (4)$$

The range of values in [0, max(q)] can be normalized to the interval of [0, 255] based on formula (4) . Here the function max(q) denotes the maximum value of the histogram. Then the image element value of image corresponding to the value of the histogram for each color can be linked up by back-projection. That is the greatest probability of Neurosurgery which is corresponding to the histogram of the pixel, which can be showed with high brightness. This probability distribution of the color can be showed in Fig. 8. Figures (a) are the RGB picture of target object. Figures (b) are the histogram of target object in V channel of HSV color space. Figures(c) are original image. Figures (d) are back projection of whole image according target histogram in H channel.



(a1) (b1) (c1) (d1)

(a2) (b2) (c2) (d2)

(a3) (b3) (c3) (d3)

Figure 8: The image of H, S and V channels

### 3.4 Mean Shift

The Mean Shift algorithm is used as a way to converge from an initial area for location and scale to the best match based on the color-histogram similarity. Mean shift procedure is applied to find the mass center of the current tracked region, it described by Figure 4.. And the calculated mass center will be used as the initial center of the next frame.

Let a kernel function is given as following:

$$K(x_i - x) = e^{c\|x_i - x\|^2} \quad (5)$$

This Gaussian kernel function determines the weight of nearby points for re-estimation of the mean. The weighted mean of the density in the window determined by K is

$$m(x) = \frac{\sum_{x_i \in N(x)} x_i * K(x_i - x)}{\sum_{x_i \in N(x)} K(x_i - x)} \quad (6)$$

where N(x) is the neighborhood of x, a set of points for which $K(x) \neq 0$ . The mean shift algorithm sets $x \leftarrow m(x)$ , repeats the estimation until m(x) converges.

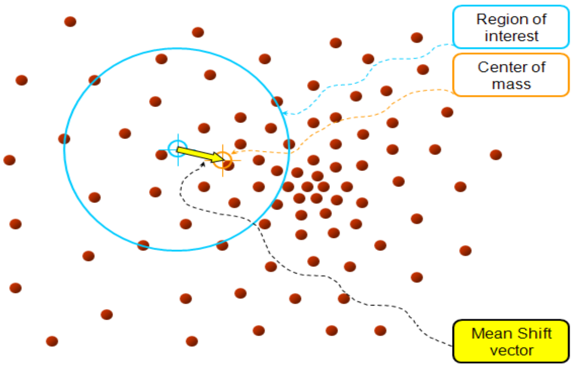



Figure 9: Find the densest regio

### 3.5 Summary of Classic CAMShift

The CAMShift algorithm has been used to adaptively meet the size and location of the search window. Current frame's localization results of the tracked object are reported and used to set the size and location of the search window in the next frame image. Firstly, it is usually converts video frame to HSV colorful space. The CAMShift algorithm is shown as follows [10, 11, 12].

(a) Set the calculation region of the probability distribution to the whole image.

(b) Choose the initial location of the 2D mean shift search window.

(c) Calculate the color probability distribution in the 2D region centered at the search window location in an ROI slightly larger than the mean shift window size

(d) Run Mean Shift algorithm, to get a new location and size of the search window. The centroid of search window can be obtained by calculating the moment of search window as follows :

The zero-order moment:

$$M_{00} = \sum_x \sum_y I_c(x,y) \quad (7)$$

The first-order moments of x and y:

$$\begin{cases} M_{10} = \sum_x \sum_y x I_c(x,y) \\ M_{01} = \sum_x \sum_y y I_c(x,y) \end{cases} \quad (8)$$

The centroid of search window:

$$x_0 = \frac{M_{10}}{M_{00}}; \quad y_0 = \frac{M_{01}}{M_{00}} \quad (9)$$

The size of search window;

$$s = 2 * \sqrt{\frac{M_{00}}{256}} \quad (10)$$

(e) In the next frame of video images, use the value obtained by Step c to re-initialize the search window's location and size, then jump to Step C to continue to run until convergence.

### 3.6 Improved CAMShift

In the back projection step of classic CAMShift, looking for exist color pixel in H channel histogram and changing value of each pixel are time-consuming works. Unfortunately, the back projection step will map object H channel histogram into whole image through the whole period of tracking. In paper [18], skin color segmentation using H threshold is presented, if pixels belong to range of skin color then the pixel's value in V channel equal one, otherwise the value is zero in H channel. But the value of threshold in V channel is affected by illumination. In different light condition, the central point of the skin distribution is different. As shown in the figure 7.b, the uniform-colored object has single peak in its histogram chart in regular light condition. Therefore, a self-adaptive threshold method was presented in there. In the frame of video, the histogram of target object was calculated in H channel of HSV. Then, the peak of histogram will be selected; the value of bin (P) in peak is the center-point of skin-color distribution. In the paper, the values of H channel are divided into 256 levels (bins) from 0 to 255.after that According experiments, the range [P-13, P+13] can cover the color of skin, this threshold will be used to segment sequent frames. The segmentation is shown in the Figure 8. In general, they have the same function for find target object in video frame. So, this step will be substituted for the back projections step. By using proposed method, there are 2 aspects contribute to tracking algorithm. Firstly, driver's face will be found very quickly it will improve real-time performance. Secondly, self-adaptive threshold can enhance the robust performance of tracking algorithm.

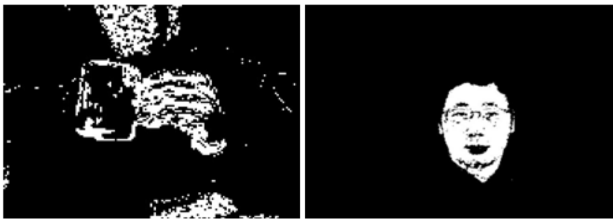


Figure 10: Self-adaptive threshold segmentation

When human being drives a vehicle, driver's head move very slow and has limited active area. A sub-window of video frame can be set up; all the process of CAMShift will be limited in the sub-window. In other words, searching driver's face, doing morphological operation and running mean shift operation have relationship just with the sub-windows. It will improve real-time performance of tracking algorithm by reason of saving calculation consumption.The tracking process of improved CAMShift can be denoted on figure 11.

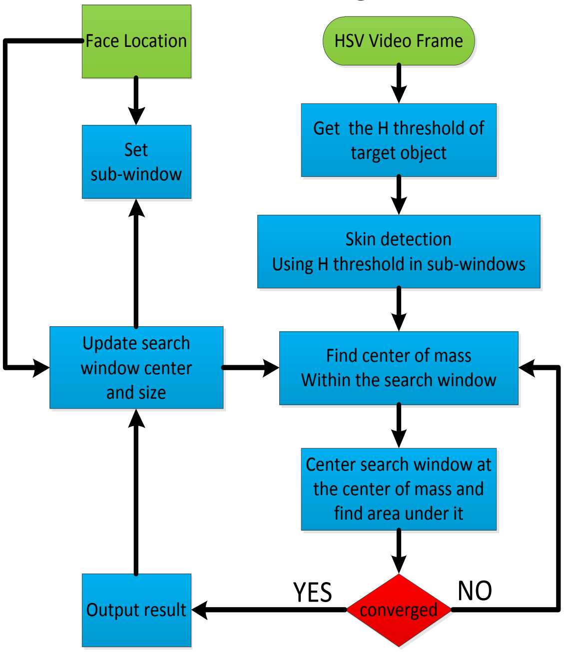



Figure 11: Flow chart of improved CAMShift

## 4 Experiments

In this phase, two video will be used to verify the validity the algorithm proposed in this paper. The hand tracking video (red cup) is downloaded from internet. The moving head video was record by using web-camera in 640x480 resolutions.

In first experiment, the improved CAMShift algorithm compared with classic CAMShift algorithm in hand tracking video. The comparing items include result of tracking and movement trajectory.

Figure 12 is the result of improved CAMShift algorithm and classic CAMShift algorithm.to track hand. The above images are result of the improved CAMShift algorithm; the below ones are the classic CAMShift algorithm result. Figure 11 is the motion trajectory of tracked hand. The left image is outcome of the improved CAMShift algorithm, the right one is outcome of classic CAMShift algorithm.

A. Using improved CAMShift

B.Using Classic CAMShift

Figure 13: Motion Trajectory Motion Trajectory

In the second experiment, the performance of system will be verified by using moving head vide. In the moving head video, the tested person moved his head left and right direction and shifted head backwards and forwards. The example results are shown as following figure 12, where shown the experiment results in high accurate rate about 98%.

Figure 14: Results of face tracking

## 5. Conclusions future work

In this paper, an improved CAMShift algorithm according the driver's head motion was proposed. The relative techniques were discussed in detail. Experimental results indicate that the overall performance of the proposed algorithm is more rapid and more accurate compared with the classic algorithms. The performance of improved algorithm has considerable robustness when illumination changing. However this algorithm still has space to improve such as if an object's color similar to skin color in sub-window, the algorithm maybe miss of target object after some time. These will be addressed in future works

## References

[1] Malaysia Sees Increased Road Fatalities. Retrieved June 18, 2011, from www.roadtrafic-technology.com/news/news108439.html.

[2] Reinier Coetzer, Driver fatigue detection based on eye tracking, in Southern Africa Telecommunication Networks and Applications Conference (SATNAC), 2010.

[3] Bradski G R. Computer V ision F ace T racking F or Use i n a Perceptual User In terface[M].Intel T echnology Journal,1998

[4] J. G. Allen, R. Y. D. Xu and J. S. Jin, Object tracking using camshift algorithm and multiple quantized feature spaces, in Proceedings of the Pan-Sydney area workshop on Visual information processing, ACM International Conference Proceeding Series Vol. 100

[5] N. Liu and B. C. Lovell, Mmx-accelerated real-time hand tracking system, in IVCNZ 2001, (Dunedin, New Zealand, 2001)

[6] Bogdan Kwolek. CamShift - based tracking in joint color-spatial spaces [J]. Computer Analysis of Images and Pat2 terns, 2005 (3691): 693 – 700

[7] Zivkovic and B. Krose, An em-like algorithm for color-histogram based object tracking, in IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR), 2004

[8] Zhaowen Wang, XiaoKang Yang , Yi Xu, CamShift guided particle filter for visual tracking, Signal Processing Systems. China: Shanghai: IEEE Workshop, 2007: 301- 306

[9] David Exner, Erich Bruns, Daniel Kurz, and

Anselm Grundhőfer, Fast and Robust CAMShift Tracking, in IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR), 2010

[10]Xia Liu, Hongxia Chu, Pingjun Li. Research of the Improved Camshift Tracking Algorithm. IEEE International Conference on Mechatronics and Automation, PP 5-8, August, 2007

[11]Liu Xue, Chang Faliang, Wang Huajle. An object tracking method based on improved camshift algorithm. Control & Automation.2007,v23, n7-3, p297-298, 305

[12]Openvc Source Computer Vision Library Reference Manual, version 0.001,2000

[13]HSL and HSV, Retrieved June 18, 2011, from http://en.wikipedia.org/wiki/ HSL_and_HSV# Hue_and_chroma

[14] A.R. Smith, "Color Gamut Transform Pairs,"SIGGRAPH 78, pp. 12-19, 1978

[15]Paul Viola, Michael J. Jones, Robust Real-Time Face Detection, International Journal of Computer Vision 57(2), 2004

[16]Gary R. Bradski, Computer Vision Face Tracking For Use in a Perceptual User Interface, Intel Technology Journal Q2 '98

[17]Kamarul Hawari Bin Ghazali, Jie Ma, Rui Xiao, An Innovative Face Detection Based On Skin Color Segmentation, International Journal of Computer Applications, Volume 34-No.2,November 2011

[18]Tian Wei, Zhuang zhenquan, Self-adaptive Skin Color Detection Based on HSV Color Space.[j], Computer Engineering And Applications,2004,14(40):82-85

**Jie Ma** received the Master degree in Signal Processing from Communication University of China, Beijing, China, in 2006. Now is working toward the PhD degree at University Malaysia Pahang. His research interests include image processing, computer vision and intelligent system.

**Rui Xiao** is currently pursuing her Master at UNIVERSITI MALAYSIA PAHANG, Pekan, Malaysia. She received the B.S. degree from XI'AN International University, PR China in 2001. Her research interests are in pattern recognition and image processing

**Kamarul Hawari Ghazali** is a senior lecturer at Universiti Malaysia Pahang since 2010. He received his first degree in 1995 from Universiti Teknologi Malaysia in Electrical Engineering. He then continues his study in Master of Electrical Engineering and was graduated in 2003. He has PhD in Electronics System Engineering specialist in Computer Vision System. He was graduated in PhD from Universiti Kebangsaan Malaysia. His research area is in image processing, computer vision and intelligent system.