MEAN SHIFT BASED OBJECT TRACKING: THE EFFECT OF COLOR SPACE

S. MEDOUAKH, M. BOUMEHRAZ

Energie Systems Modeling Laboratory, University of Biskra, Algeria
medsaadia@yahoo.fr
medboumehraz@mselab.org

ABSTRACT

The mean shift algorithm is widely used in object tracking because of its speed, efficiency and simplicity. This algorithm is used to track the location of a non rigid object in image sequences using the object's color histogram. Mean shift tracker maximizes iteratively the appearance similarity by comparing the color histogram of the target model and the target candidate. In this paper we focus on studying the effect of color spaces (HSV, YCrCb, YIQ, YUV, I1I2I3 Lab) on object tracking by mean shift algorithm since this algorithm is based on color information (color histogram) and traditional color histogram uses only the RGB color space. The experimental results show that each color space can influence the tracking robustness for different color targets having the same background.

KEYWORDS: Mean Shift algorithm, Color Histogram, Color Spaces, Object Tracking.

1 INTRODUCTION

Object tracking is one of the most important tasks in the field of computer vision [1]-[3]. It's widely applied in military and civil field such as surveillance, visual navigation [3] and human computer interaction [1], etc. Object tracking purpose is to detect, extract, recognize, and track a moving object, get its state parameters, and understand the behavior of the object. Although object tracking [2] has been studied for several decades, and much progress has been made in recent years, it remains a very challenging problem. Numerous factors affect the performance of a tracking algorithm, such as the presence of noise, illumination variation, occlusion, background clutters, and scale appearance change of the objects variation.

Among various tracking algorithms, the Mean-Shift tracking is one of the most efficient tracking algorithms for real-time applications, due to its simplicity and robustness [4]. Mean shift [5], which was proposed by Fukunaga and Hostetler in 1975, is an unsupervised clustering method. Cheng [2], developed, in 1995, a more general formulation of mean shift and forecasted its potential applications in clustering and global optimization. Since Comaniciu [1],[5] studied the application of mean shift in 1999, mean shift is widely used for object tracking [6], image segmentation, etc.

The mean shift algorithm [5],[6],[11] is a simple and fast adaptive tracking procedure that finds the maximum of the Bhattacharyya coefficient given a target model and a starting region. This algorithm based on the color histogram, is used to describe the target region [9]. Color has been used widely in many areas of computer vision such as image segmentation and object tracking. The objective of this work is to study the effect of the color spaces, used to calculate the color histogram of the target region, on the tracking efficiency and quality using mean shift algorithm.

This paper is organized as follows. Section 2, is a short description of the mean shift tracking algorithm. The different color spaces are described, in Section 3. The experimental results of object tracking in video sequences, and comparisons for different color spaces are given in Section 4. Finally, section 5 gives the conclusion of this work.

2 MEAN SHIFT ALGORITHM

Mean shift is a nonparametric estimator of density gradient. This algorithm finds local maxima in any probability distribution. It is used for tasks such as clustering, mode seeking, probability density estimations, and tracking [3],[4]. In Mean shift, the object is represented by a kernel-weighted histogram.

2.1 Object representation

2.1.1 Target model

The first step is the initialization of the target, we select a rectangular region which contains the object in the first frame. Suppose the coordinate of the target's centre point is x0, the kernel function is k(x), and the bandwidth is h, then for all the pixel points $\{x_i^*\},\ i = 1,2,...,\ n$ within the target region, the probabilities of each feature in feature space can be calculated. The target probability model [2] corresponds to each feature (bin) u=1, 2,…, m can be expressed as :

$$\hat{q}_u = C\sum_{i=1}^{n} k\left(\left\|\frac{x_0 - x_i}{h}\right\|^2\right)\delta[b(x_i) - u] \tag{1}$$

Where C is a normalizing term so the sum of $\hat{q}_u$ equals to one, δ is Dirac delta function so each pixel contributes in only one feature, b(x_i) is the color of pixel x_i. The role of functions b and δ is to determine whether the color value of x_i is in the feature value u. Mean shift [2] uses the Epanechnikov kernel written as:

$$k = \begin{cases} \frac{1}{2} c_d^{-1}(d+2)(1-\|x\|^2) & if\,\|x\| < 1 \\ 0 & otherwise \end{cases} \tag{2}$$

Where d is the number of dimensions, and c_d is the volume of dimension.

2.1.2 Target candidate

Candidate Target area is the area that may contains the Target to track following the first frame. Suppose pixel points in the candidate target area are, where the center point is y, the kernel function is k(x), and the bandwidth is h, then the target probability candidate [2] corresponds to each feature, u=1,2,…,m, can be expressed as:

$$\hat{p}_u = C_h\sum_{i=1}^{n_h} k\left(\left\|\frac{y - x_i}{h}\right\|^2\right)\delta[b(x_i) - u] \tag{3}$$

2.2 Similarity function

Similarity function is used to describe similarity measure between the target model and the target candidate. The Bhattacharyya coefficient is used as the similarity function; it is defined as [2]:

$$\rho(\mathrm{y}) = \rho(\hat{p}_u(y), \hat{q}_u) = \sum_{i=1}^{m}\sqrt{\hat{p}_u(y)\hat{q}_u} \tag{4}$$

The value of ρ(y) is much larger, the target model and candidate are more similarity.

The distance between the target model and the target candidate is defined as:

$$d(\mathrm{y}) = \sqrt{1 - \rho\left[\hat{p}_u(y), \hat{q}_u\right]} \tag{5}$$

2.3 Target localization

The most probable location y of the target in the current frame is obtained by maximizing the Bhattacharyya coefficient $\rho(\mathrm{y})$. The search for the new target location in the current frame starts at the estimated location y_0 of the target in previous frame. The target's feature probability of the target candidate at location y_0 in the current frame is $\hat{p}_u(y_0),\ u = 1,2,...,\ m$, have to be computed first. Using Taylor expansion around the values $\hat{p}_u(y_0)$ the Bhattacharyya coefficient is approximated as [2]:

$$\rho(\hat{p}_u(y), \hat{q}_u) \approx \frac{1}{2}\sum_{u=1}^{m}\sqrt{\hat{p}_u(y)\hat{q}_u} + \frac{C_h}{2}\sum_{u=1}^{n_h} w_i k\left(\left\|\frac{y - x_i}{h}\right\|^2\right) \tag{6}$$

Where

$$w_i = \sum_{u=1}^{m}\sqrt{\frac{\hat{q}_u}{\hat{p}_u(y_0)}}\,\delta[b(x_i) - u] \tag{7}$$

As the first term of (6) is independent of y, the second term of (6) must be maximized. Mean shift vectors moving from y0 to y can be calculated iteratively by the second term, and the center point of Mean shift in iterative computation is [2]:

$$y_1 = \frac{\sum_{i=1}^{n_h} x_i w_i g\left(\left\|\frac{y_0 - x_i}{h}\right\|^2\right)}{\sum_{i=1}^{n_h} w_i g\left(\left\|\frac{y_0 - x_i}{h}\right\|^2\right)} \tag{8}$$

Where g(x)= - k'(x), target's best position y1 in current frame can be through iterative calculation by (8). The iterative process stops when ‖ y_1 - y_0 ‖<threshold, usually threshold=1 pixel.

The mean shift iteration is described in following algorithm:

Algorithme mean shift tracker iteratio

Input: The target model $\left\{ \hat{q}_u \right\}, u = 1,2,..., m$ and its location y_0 in the previous frame.

1. Initialize the center of the rectangular in the current frame at y_0, compute $\hat{p}_u(y_0), u = 1,2,..., m$ using (3) and evaluate $\rho(\hat{p}_u(y_0), \hat{q}_u)$ using (4).
2. Compute the weights $w_i, i = 1,2,..., n$ according to (7).
3. Compute the next location of the target candidate according to (8).
4. Compute $\hat{p}_u(y_1), u = 1,2,..., m$ using (3) and evaluate $\rho(\hat{p}_u(y_1), \hat{q}_u)$ using (4).
5. While $\rho(\hat{p}_u(y_1), \hat{q}_u) < \rho(\hat{p}_u(y_0), \hat{q}_u)$,
   Do $y_1 \leftarrow \frac{1}{2}(y_0 + y_1)$,
   and evaluate $\rho(\hat{p}_u(y_1), \hat{q}_u)$ using (4).
6. If $\|y_1 - y_0\| < \varepsilon$ stop.
   Otherwise, set $y_0 \leftarrow y_1$ and go to Step 2.

3 COLOR SPACE REPRESENTATION

The study of color is important in design and development of color vision systems. The use of color in image of display is not only more pleasing, but it also enables us to grasp quickly more information.

A color space is a method of describing and representing colors in a standard way. The choice of the color space can be a very important decision which can dramatically influence the results of the processing. There are many color spaces are used such as RGB, YCbCr, YUV, YIQ, YCoCg, HSV, HSL, HSI, CIE XYZ, L*a*b*, L*u*v*, and I1I2I3, etc.

The color models selected for the study, chosen between the most usual in applications of computer vision are the following [8(11), 9(12)] :

RGB, rgb : the most commonly employed color space in computer technology is the RGB color space, which is based on the additive mixture of three primary colors R(red), G(green), B(blue). From this model defines the variant normalized rgb, that consists in dividing the values of RGB by (R+G+B). The importance of the RGB color space is that it relates very closely to the way that the human eye perceives color.

YUV, YIQ, YCbCr, YCgCo: (Luminance-Chrominance)

YUV and YIQ are standard color spaces used analogue television transmission. YUV is used in European TVs and YIQ in North American TVs (NTSC).These color spaces are “derived” from the RGB space. Where the Y is component of luminance, *and* U, V and I, Q are components of chrominance. The YCbCr color space is used for component digital video. This color space similar to YUV and YIQ that separate RGB into luminance Y and chrominance Cb, Cr information and is useful in compression applications. The YCgCo color model was developed to increase the effectiveness of the image compression. This color model comprises the luminance (Y) and two color difference components (Cg - offset green, Co - offset orange).

CIEXYZ: The XYZ color space is an international standard developed by the CIE (Commission Internationale de l’Eclairage) in 1931. XYZ are known as tristimulus values, This space is based on three colors primaries, XYZ, and all visible colors can be represented by using only positive values of X, Y, and Z. The Y primary is luminance, while X and Z primaries give color information. The main advantage of the CIEXYZ space is that this space is completely device-independent.

L*a*b*, L*u*v*: in 1976 the CIE proposed two color spaces CIELuv and CIELab which are perceptually uniform. It is derived based on the standard CIE XYZ space and white reference point. The Euclidean distance between two color points in the CIELuv/CIELab color spaces corresponds to the perceptual difference between the two colors by the human vision system. The Lab separates the color information into lightness (L) and color information (a, b), (a) correlates with the red-green components and (b) correlates with the yellow-blue components, Luv evolved from the Lab space.

HSV, HSL: The HSV (hue, saturation, value), HSL (hue, saturation, lightness), color spaces based on the idea of human visual system, were developed to be more intuitive in manipulating with color and were designed to approximate the way humans perceive and interpret color. HSL and HSV are used widely in computer graphics. These color spaces uses cylindrical coordinates for the representation of RGB points. The luminance information is place in the V(HSV), L(HSL) and varies with color saturation, the hue H represents color and saturation S indicates the range of grey in the color space are the chrominance components.

I1I2I3: Some models have been designed specifically for some applications. For example, the color space I1I2I3 proposed for the segmentation of color was obtained through the decorrelation of the components RGB using the transformed Karhunen-Loeve.

4 EXPERIMENTAL RESULTS

In this section, we present the experimental results obtained by applying Mean Shift tracking algorithm with different color spaces, and illustrate the effect of color spaces in object tracking. The approach was tested using two images sequences: “Football” and “Woman”. The first one is 56 frames sequence and the size of each frame is 756x720, and the second is 597 frames sequence of size 288x352. In all experiments, the number of histogram is 16 bins. The tracking target is manually selected in first frame. The tracking results of experiment are presented in the following Figures. The results can verify the effectiveness of different color spaces; we used 3 components of each color space. We tested all the color spaces RGB, rgb, HSV, HSI, HSL, YCrCb, YIQ, YUV, I1I2I3, XYZ, Lab and Luv. In this paper, we compare the tracking results of color spaces which gave us the robust tracking (HSV, Lab, YCrCb, YIQ, YUV, I1I2I3).

Figure1 shows tracking results of player in orange outfit (target 1) in “Football” sequence, for frames 15, 27, 28, and 50. We can see that the color spaces HSV, YUV and I1I2I3 can well recapture the target when it reappears after occluded (frame 28) better than Lab, YCrCb, and YIQ. We compare between the direction X and Y of real (or accurate) center and of tracking center through different color spaces are shown in figure 2. We can see that in direction Y for all color spaces the center of target did not move except when the target is occluded (frames 27 and 28). But in direction X, the center of target moved when the target reappears after occluded with HSV, YUV and I1I2I3.

Figure 3 shows tracking results of player in white outfit (target 2) in frames 8, 21, 40, and 51. We can see that all color spaces can track the target, except the color space HSV failed. From figure 4, we can observe that in direction X and Y for all color spaces the center of target moved but the center by HSV largely moved especially for direction X.

Figure 5 shows tracking results for Woman sequence in frames 22, 92, 124, and 555. We can see that for all the color spaces the target can be tracked accurately in the frame 22, except for Lab color space. However, it failed in the frame 92 since the background has changed, and the color difference between this background and the target is indistinctive. For the color space HSV, in frame 124, the target can be tracked even under the circumstance of the variation of illumination and partial occlusion better than the other colors spaces. The target is lost in YUV and I1I2I3 after 93 and 162 frame, respectively.

To evaluate the tracking results for each color space, comparisons with the manually labeled ground truth have been performed. The success rate is evaluated through the overlap rate (bounding box overlap). We compute [13] the bounding box overlap S of r_t and r_g in each frame, where r_t is the bounding box outputted by a tracker and r_g is the ground truth bounding box. The bounding box overlap is defined as:

$$S = \frac{|r_t \cap r_g|}{|r_t \cup r_g|} \tag{9}$$

Where ∩ and ∪ denote the intersection and union of two regions, respectively.

To measure the performance on a sequence of frames, we compute the average overlap over all the frames in this sequence as:

$$overlap\,rate = \sum_{i=1}^{N} S / N$$

Where N is the total number of frames.

Table 1 summarizes the average overlap between tracked bounding box and the ground truth bounding box for tracking of each color space, We observe that the HSV has better performance of success rate than the other color spaces for the target 1(player in orange outfit) in “football sequence”. But has failed for the target 2 (player in white outfit); for this target the I1I2I3 and YIQ tracking performance are better than Lab, YUV, YCrCb. We can say that the HSV color HSV is very sensitive to target and background colors. As it can be seen, the target in the "woman sequence" is lost for the YUV and I1I2I3 color spaces. However, using HSV the target is tracked successfully than Lab, YCrCb, YIQ color spaces.

Table 01: Overlap rate the ratios of successful

Sequence	HSV	Lab	YCrCb	YIQ	YUV	I1I2I3
“football” Target 1	**0.7725**	0.7248	0.7187	0.7461	0.7434	0.7411
“football” Target 2	0.4900	0.7375	0.7096	**0.7569**	0.7069	**0.7681**
“woman”	**0.7035**	0.5891	0.6226	0.6311	lost	lost

From the comparison of tracking results between different color spaces for different color targets in the same sequence (same background) and for targets under variation of illumination and partial occlusion, we can say that the color space can influence the tracking robustness, because each color space has characteristic according to the target color and the same background in football sequence. The disparity results demonstrate that the color space choice can be a very important step in Mean shift tracking algorithm because this algorithm is based only on color information. However, we can say that the HSV color space performance is better than that of Lab, YCrCb, YUV, YIQ and I1I2I3 color spaces in the case of partial occlusion and illumination variation.

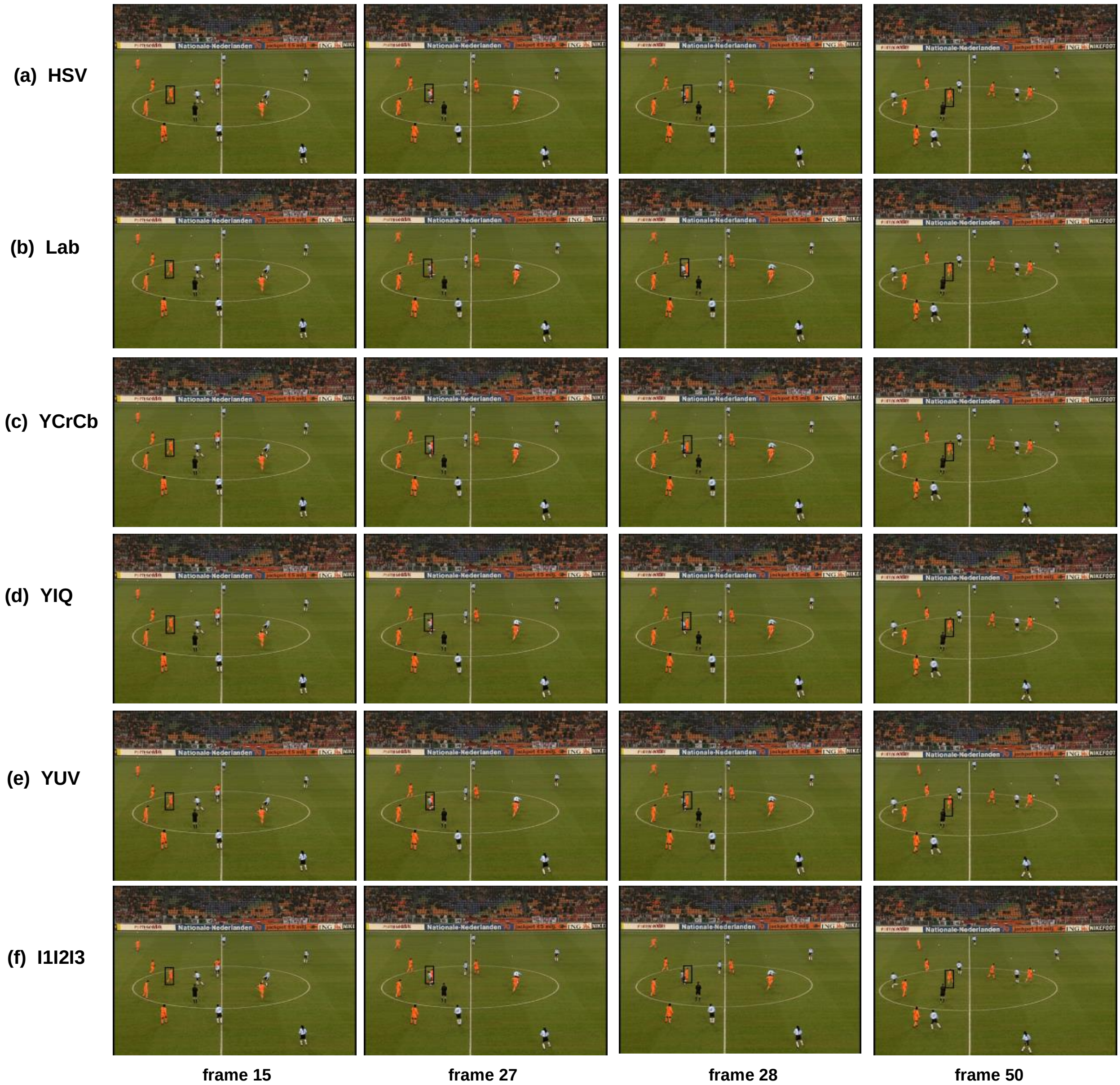


Figure 01: Tracking results of different color spaces of player in orange outfit in frames 15, 27, 28, and 50

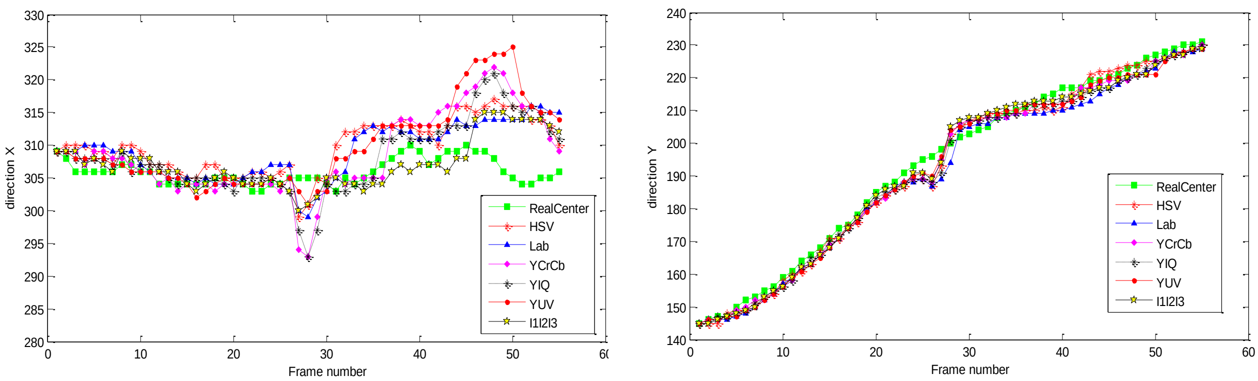
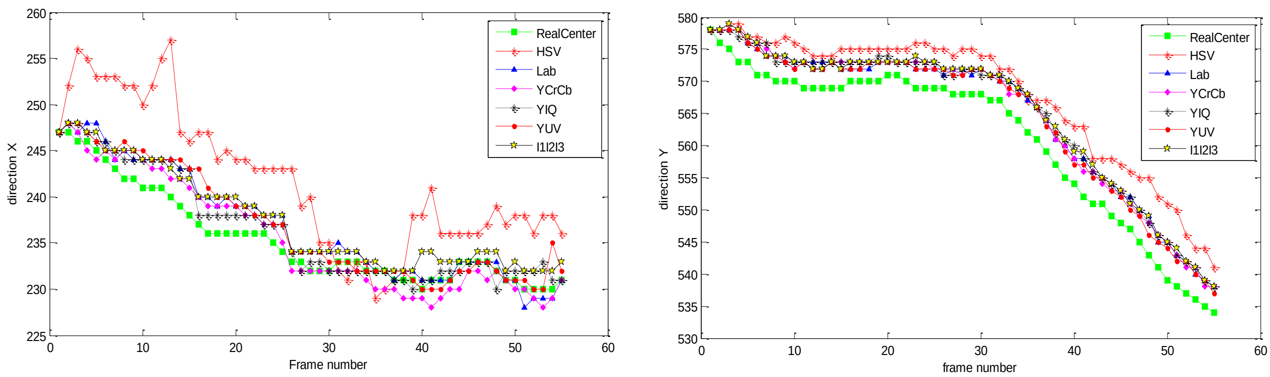


Figure 02: Direction of center moving object (player in orange outfit), (a) Direction X, (b) Direction Y

(a) HSV
(b) Lab
(c) YCrCb
(d) YIQ
(e) YUV
(f) I1I2I3

frame 8 frame 21 frame 40 frame 51

Figure 03: Tracking results of different color spaces of player in white outfit in frames 8, 21, 40, and 51



(a) (b)

Figure 04: Direction of center moving object (player in white outfit), (a) Direction X, (b) Direction Y

Figure 05: Tracking results of different color spaces of Woman frames 29, 92, 124, and 555

5 CONCLUSION

In this paper, we have presented the effect of color space on object tracking using Mean Shift algorithm. Comparative tracking results showed that the quality of tracking depends not only on the color space but also on the object and background color. The choice of the color space can be very important and very difficult, because the appropriate color space depends on the actual situation and may vary between the frames. The selected color space must have the maximum ability to distinguish the object from its background. The choice of the space color can be done automatically as a pre-step before applying the mean shift.

REFERENCES

[1] A. Yilmaz, O. Javed, and M. Shah, "Object tracking: A survey," ACM Computing Surveys, vol. 38, no. 4, pp. 13-57, 2006.

[2] D. Comaniciu, V. Ramesh, and P. Meer, "Kernel-based object tracking," IEEE Trans. Pattern Anal. Mach. Intell, vol. 25, pp. 564-577, May 2003.

[3] Q. Zhao, W. Yao, J. Xie, "Research on object tracking with occlusion", WIT Transaction on information and Communication Technologies, vol. 51, pp. 537-545, 2014.

[4] J. Zhao, W. Qiao, G-Z. Men, "An approach based on mean shift and kalman filter for target tracking under occlusion", International Conference on Machine

Learning and Cybernetics, vol. 4, pp. 2058-2062, 2009.

[5] Z. Wen, Z. Cai, “ A robust tracking approach using mean shift ”, Proceedings of the Third International Conference on Natural Computation, Vol.2, pp.170-174, Aug 2007.

[6] K. Fukunaga and L. Hostetler, “The Estimation of the Gradient of a Density Function, with Applications in Pattern Recognition”, IEEE Tran. IT, Vol.21, 1975, pp 32-40.

[7] K. Bai, W. Liu, “Improved object tracking with particle filter and mean shift”, Proceedings of the IEEE International Conference on Automation and Logistics. pp. 431-435, 2007.

[8] X. Yang, H. Yin, Y. Chai, “An improved moving target tracking algorithm based on mean shift”, Journal of Computational Information Systems 8: pp. 39-46, 2012.

[9] K.H. An, M.J. Chung, “Mean shift based object tracking with new feature representation”, the 3rd International Conference on Ubiquitous Robots and Ambient Intelligence, pp. 365-370, 2006.

[10] I. Leichter, M. Lindenbaum, E. Rivlin, “Mean shift tracking with multiple reference color histograms”, Computer Vision and Image Understanding 114, 400-408, 2010.

[11] N.A. Ibraheem, et al, “Understanding color models: A review”, ARPN Journal of Science and Technology, vol. 2, no. 3, pp. 265-275, April 2012.

[12] P. Sebastia, Y. Vooi Voon, R. Comley, “Color space effect on tracking in video surveillance”, International Journal on Electrical Engineering and Informatics, vol. 2, no. 4, pp. 298-312, 2010.

[13] Y. Wu, J. Lim, and M. H. Yang, “Online object tracking: a benchmark,” in Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR '13), pp. 2411–2418, 2013