THE EFFECT OF COLOR SPACE ON MEAN SHIFT
OBJECT TRACKING: A COMPARATIVE STUDY

Ahmed Nabil Mohamed [1], Mohamed Moness Ali [2]

[1]*Assistant Lecturer, Department of Computer and Information Systems, Sadat Academy*
[2]*Professor, Department of Computers and Systems, Engineering Faculty, Minia University*

ABSTRACT

The mean shift algorithm is a widely used tool for robust and real-time object tracking that can use any feature space such as the color space. Therefore, in this article, we have studied and analyzed the various effects of using different color space configurations with the mean shift algorithm on the tracking process in surveillance videos. We have done so by investigating the potency of the gray space and four other color spaces (RGB, HSV, YIQ, YCbCr [each color space was studied under all the available seven configurations]). In the end, we have reached a conclusion that deciding which color space (or which configuration of the color space channels) to be used not only can be a very critical decision which may affect greatly the results of the tracking process, but also there is no color space configuration that can succeed all the time in all situations. Therefore, it is better to select the appropriate color space configuration according to the situation at hand.

Keywords: Tracking, Color Spaces, Mean Shift.

1. Introduction

The importance of object tracking in computer vision comes from the urge need of it in applications like automated surveillance, motion-based recognition, augmented reality, video indexing, video compression, perceptual user interface, smart rooms, vehicle navigation, traffic monitoring, driver assistance, etc.

Color is a very popular feature that is used extensively in tracking algorithms. It gains its popularity because it is invariant to translation, rotation, partial occlusion, viewpoint change, pose variations, zooming, resolution, etc. Moreover, it is also easily extracted.

The mean shift algorithm is a widely used tool for robust and fast object tracking that can use any feature space such as the color space. It was originally proposed in [1] by Fukunaga and Hostetler as a robust non parametric technique for gradient estimation of a density function using a generalized kernel approach. It was first applied for the pattern recognition problems such as data clustering and data filtering [1]. Later, it was revived,

---

* Corresponding author.
 E- mail address: anabil.mohamed@live.com

after two decades, by Cheng [2] where he provided a general analysis of the algorithm calling to benefit from its many interesting and useful properties and stressing on its ability of seeking modes of real functions and its applications in clustering analysis and global optimization. The first use of the mean shift algorithm as a real-time tracker (face tracker) was provided by Bradski [3]. The popularity of the algorithm became more evident and well founded in the computer vision society after its successful application to image segmentation and tracking by Comaniciu et al. [4, 5, 6], where they indicated its utility as an important component of any computer vision system.

Although, it has been more than a decade since the mean shift algorithm was introduced as a real-time tracker for the first time, but its use in the literature is still increasing [7] and many modifications have been developed to enhance its performance. However, we have noticed that there is little effort spent in the literature for studying the effects of different color spaces on the tracking process when using the mean shift algorithm. Therefore, our goal in this article is to examine carefully these effects.

The organization of the paper is as follows: In section 2, we briefly discuss the RGB, HSV, YIQ and YCbCr color spaces plus the grayscale space. The mean shift algorithm is discussed and analyzed in section 3. The experimental setup is described in section 4. Experimental comparisons are shown and discussed in section 5. Conclusion and future work are provided in section 6.

2. Color spaces

There are many color spaces such as RGB, HSV, HSL, YUV, YIQ, YCbCr, YCC, CMYK, CIEXYZ, CIELUV, CIELAB, etc. Many of them have been applied successfully in computer vision applications. Each color space has its own characteristics that make it more suitable than others for specific applications. Some color spaces are perceptually linear (i.e. changing the color value produces the same change in perception) while others are not [8]. Some color spaces are intuitive to use while others are not [8]. Some are device dependent while others are device independent [8]. For a short overview of color spaces used in image processing, one can refer to [9] where Tkalcic and Tasic have provided historical, perceptual and applicational background that led to the introduction of different color spaces.

Here, in this section, we only discuss, in brief, the RGB, HSV, YIQ and YCbCr color spaces plus the grayscale space, since they were used in our experiments.

The grayscale space is a space that carries only the intensity information, i.e., the only colors is shades of gray. These shades of gray are bounded by black (the darkest shade) and white (the lightest shade). In other words, when a pixel in a digital image have the same intensity value of its red, green and blue components in the RGB color space, then it can be represented by a shade of gray [10].

The RGB color space is one of the most used color spaces for image and video processing. It has three channels: red, green and blue which correspond to the cones sensors in the human eye. Therefore, these colors are called primary colors. However, a main problem in this color space is that the channels are very correlated (i.e., every channel contains luminance data). This problem affects the performance of the RGB color space in color analysis and color-based recognition algorithms [11]. Therefore, it is desirable to separate the luminance data from the chrominance data, and this can be

achieved by converting the RGB color space to other color spaces such as the HSV and YCbCr color spaces.
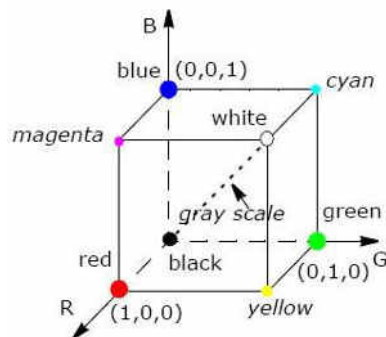

Fig. 1. RGB color space [12]

HSV is a perceptual color space that describes color intuitively (i.e. similar to human perception of color). The HSV color space can be seen as a cone. The hue values can range from 0 to 360 degrees starting from red at the 0° line passing with green at 120° and blue at 240°. The saturation value is represented as the distance from the vertical axis of the cone and measures the purity of the color. Highly saturated colors are located on the outer edge of the cone, whereas gray tones are located at the very center [13]. The brightness is determined by the colors vertical position in the cone [13]. The brightest colors are located at the circular end of the cone, while at the tip end; there exist the darkest colors where there is no brightness [13]. The HSV color space can be calculated from the RGB color space as follows:

$$H = cos^{-1} \frac{\frac{1}{2}((R-G)+(R-B))}{\sqrt{((R-G)^2+(R-B)(G-B))}} \tag{1}$$

$$S = \frac{\max(R,G,B) - \min(R,G,B)}{\max(R,G,B)} \tag{2}$$
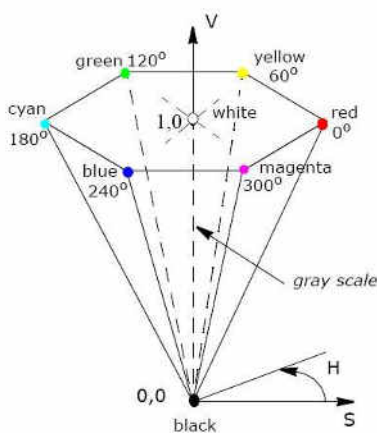
$$V = \max(R, G, B) \tag{3}$$

Fig. 2. HSV color space [12]

YIQ is a color space that is used for analog TV broadcasting (NTSC). It is a rotation of the RGB color space so that the Y axis represents the luminance data. The chrominance

data is contained in the I (orange-blue) and Q (purple-green) axes, which are roughly orthogonal. This arrangement is due to that the human visual system is much more sensitive to changes in the I axis than in the Q axis, thus, allowing the Q axis to be transmitted with less fidelity, conserving bandwidth. The YIQ color space is different from the YUV color space (a closely related color space used for The PAL TV broadcasting color system) in that there is a 33 degree rotation of the chrominance axes [14]. The YIQ color space can be acquired from the RGB color space by using the following equations:
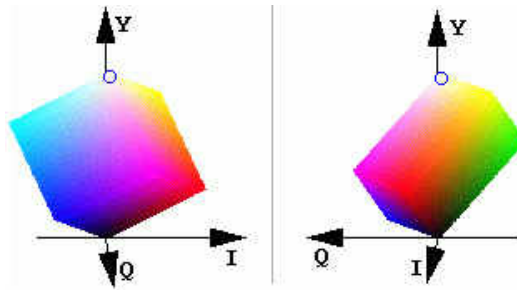
Fig. 3. YIQ color space containing the RGB color cube

$$Y = 0.299\,R + 0.587\,G + 0.114\,B \tag{4}$$
$$I = 0.596\,R - 0.274\,G - 0.312\,B \tag{5}$$
$$Q = 0.211\,R - 0.523\,G + 0.311\,B \tag{6}$$

YCbCr is a color space that is used for digital TV transmission and image compression [8, 11]. It is characterized with explicit separation of luminance and chrominance data [11]. It can be calculated from the RGB color space in the range [0 – 255] according to its definition in the ITU.BT-601 standard as follows:

$$Y = 16 + \frac{1}{255}(65.481\,R + 128.553\,G + 24.966\,B) \tag{7}$$

$$Cb = 128 + \frac{1}{255}(-37.797\,R - 74.203\,G + 112\,B) \tag{8}$$

$$Cr = 128 + \frac{1}{255}(112\,R - 93.786\,G - 18.214\,B) \tag{9}$$
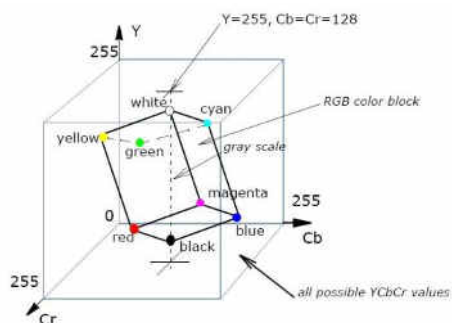
Fig. 4. YCbCr color space containing the RGB color cube [12]

3. The Mean shift algorithm

Let us now review the analysis of the mean shift algorithm [1, 2, 4] for the sake of completeness. Given a set of *n* points in the *d*-dimensional space R^d; $\{x_i\}_{i=1...n}$, the multivariate kernel density estimate with kernel *K(x)* and a bandwidth *h* (also known as window width or the smoothing parameter) is given at the point *x* by the following formula:

$$\hat{f}(x) = \frac{1}{nh^d} \sum_{i=1}^{n} K\left(\frac{x-x_i}{h}\right) \tag{10}$$

$K(x)$ is said to be a kernel if there exists a profile. This profile, which is a radially symmetric kernel, is defined as a function $k : [0,\infty) \to R$ such that

$$K(x) = k(\|x\|^2) \tag{11}$$

and the following conditions apply:

1- k is nonnegative.

2- k is non-increasing, i.e., $k(a) \geq k(b) \, if \, a < b$.

3- k is piecewise continuous.

4- $\int_0^\infty k(r)dr < \infty$

Now, the density estimate in eq. (10) can be written in terms of the kernel profile notation as follows:

$$\hat{f}_K(x) = \frac{1}{nh^d} \sum_{i=1}^{n} k\left(\left\|\frac{x-x_i}{h}\right\|^2\right) \tag{12}$$

by denoting that

$$g(x) = -\acute{k}(x) \tag{13}$$

and assuming that the derivative of k exists for all $x \in [0,\infty)$. A kernel G can be defined as:

$$G(x) = Cg(\|x\|^2) \tag{14}$$

where C represents a normalization constant.

Then, by taking the estimate of the density gradient as the gradient of the density estimate, we get the following:

$$
\begin{aligned}
\nabla f_K(x) \equiv \nabla \hat{f}_K(x) &= \frac{2}{nh^{d+2}} \sum_{i=1}^{n} (x - x_i) \, \acute{k}\left(\left\|\frac{x - x_i}{h}\right\|^2\right) \\
&= \frac{2}{nh^{d+2}} \sum_{i=1}^{n}(x_i - x)g\left(\left\|\frac{x-x_i}{h}\right\|^2\right) \\
&= \frac{2}{nh^{d+2}}\left[\sum_{i=1}^{n} g\left(\left\|\frac{x-x_i}{h}\right\|^2\right)\right]\left[\frac{\sum_{i=1}^{n} x_i g\left(\left\|\frac{x-x_i}{h}\right\|^2\right)}{\sum_{i=1}^{n} g\left(\left\|\frac{x-x_i}{h}\right\|^2\right)} - x\right]
\end{aligned} \tag{15}
$$

where $\sum_{i=1}^{n} g\left(\left\|\frac{x-x_i}{h}\right\|^2\right)$ is assumed to be nonzero.

The last bracket of eq. (15) represents the sample mean shift vector, i.e.

$$M_{h,G}(x) \equiv \frac{\sum_{i=1}^n x_i g\left(\left\|\frac{x-x_i}{h}\right\|^2\right)}{\sum_{i=1}^n g\left(\left\|\frac{x-x_i}{h}\right\|^2\right)} - x \tag{16}$$

and the density estimate, computed with kernel G, at x is

$$\hat{f}_G(x) \equiv \frac{c}{nh^d} \sum_{i=1}^n g\left(\left\|\frac{x-x_i}{h}\right\|^2\right) \tag{17}$$

Now, from equations (15, 16, 17), the estimate of the density gradient becomes

$$\widehat{\nabla} f_K(x) = \frac{2/c}{h^2} \hat{f}_G(x) M_{h,G}(x) \tag{18}$$

Then, the sample mean shift becomes

$$M_{h,G}(x) = \frac{Ch^2}{2} \frac{\widehat{\nabla} f_K(x)}{\hat{f}_G(x)} \tag{19}$$

The last equation shows that the sample mean shift vector with kernel G is an estimate of the normalized density gradient obtained with kernel K.

The mean shift algorithm is then specified as to recursively compute the mean shift vector $M_{h,G}(x)$ and translates the center of kernel G by this vector. So, the successive locations of kernel G, denoted as $\{y_j\}_{j=1,2,...}$, are computed as follows:

$$y_{j+1} = \frac{\sum_{i=1}^n x_i g\left(\left\|\frac{y_j-x_i}{h}\right\|^2\right)}{\sum_{i=1}^n g\left(\left\|\frac{y_j-x_i}{h}\right\|^2\right)}, \quad j = 1,2, ... \tag{20}$$

Then, the density estimates computed with kernel K in the points given by eq. (20) are:

$$\hat{f}_K = \{\hat{f}_K(j)\}_{j=1,2...} \equiv \{\hat{f}_K(y_j)\}_{j=1,2...} \tag{21}$$

If the kernel K has a convex and monotonic decreasing profile, then the sequences in equations (20 and 21) are convergent.

3.1. Mean shift as a tracker

We will consider here, the Bradski implementation [3] which can be summarized as follows:
1. Choose the size and the initial location of the region of interest (ROI).
2. Calculate the histogram of the ROI
3. Calculate the color probability distribution image through histogram back-projection
4. Compute the mean location in the selected ROI (x_c, y_c) as follows:

$$M_{00} = \sum_x \sum_y I(x,y) \tag{22}$$

$$M_{10} = \sum_x \sum_y x I(x,y) \tag{23}$$

$$M_{01} = \sum_x \sum_y y I(x,y) \tag{24}$$

where M_{00}, M_{10}, M_{01} are the zeroth and first spatial moments, and $I(x, y)$ is the probability value of the pixel at position (x, y) in the search window (i.e. the ROI)

$$x_c = \frac{M_{10}}{M_{00}}, \quad y_c = \frac{M_{01}}{M_{00}} \tag{25}$$

5. Center the selected ROI at the mean location computed in step 4
6. Repeat steps 4 and 5 until convergence or until the mean location moves less than a predefined threshold.

4. Experimental setup

The international series of PETS "*Performance Evaluation of Tracking and Surveillance*" workshops motivate many researchers in the community of computer vision to use their standard datasets in order to facilitate comparison and evaluation of different algorithms in the domain of surveillance and tracking systems. Therefore, we have used five sequences from the PETS datasets (PETS 2000 and PETS-ECCV 2004) to compare the effects of different color space configurations on the tracking process. These sequences can be downloaded from [15]. Table 1 shows some information about the used sequences.

Table 1.

Used Sequences

Seq. No.	Public Dataset Name	Clip Name	Resolution	Compression
1	PETS 2004 "CAVIAR"	Rest in Chair	384*288	DIV3
2	PETS 2004 "CAVIAR"	Meet Crowd	384*288	DIV3
3	PETS 2004 "CAVIAR"	Browse 2	384*288	XVID
4	PETS 2000	Test	320*240	DIV3
5	PETS 2000	Test	320*240	DIV3

In most of the following experiments, we have tracked distant and very distant objects under very noisy conditions. In every experiment, we have used the same region of interest to deal with all examined color channels, and thus we have ensured a fair evaluation. Each chosen area was represented by 1D, 2D and 3D histograms in different color spaces using different configurations (the number of histogram bins is taken to be 16 for every dimension). Then, each histogram is back-projected to form the probability image, where we can apply the mean shift algorithm to find the selected region of interest. Note that we applied the cam shift tracker in the sixth scenario, where it works by applying the mean shift algorithm and then setting the search window size to a function of the zeroth moment, and then in the next frame, the calculation region is set slightly larger than the search window. Table 2 shows the video length and the tracked object in each scenario.

Table 2.

Constructed Scenarios

| Scn. No.	Used Sequence	Tracked Object	Number of frames
1	Seq. 1	Human	114
2	Seq. 1	Human	130
3	Seq. 2	Human	141
4	Seq. 3	Human	600
5	Seq. 4	Human	250
6	Seq. 5	Truck	125

5. Results and discussion

For more clarity, we represent the successful tracking results in graphical charts. Consequently, it is very clear from [Figure-5] how using different color spaces with different configurations can be very critical and affects greatly the results of the tracking process. We can see that the only configuration that has succeeded in all experiments is the 2D SV from the HSV color space. The second most successful configurations are the 3D HSV, the 2D YI, the 2D YCb, and the 2D YCr representations. The third most successful configurations are the 3D YIQ, the 3D YCbCr, the 2D YQ, the 2D HV, and the 1-D luminance channel from the YIQ color space. It should be noted that the 3D RGB color space only succeeded in one scenario, the same goes true for the grayscale channel. But for the configurations that do not succeed in any sequence, there are five configurations which are the blue channel, the quadrature channel, the blue-difference chroma channel, the red-difference chroma channel, and the 2D CbCr configuration. Figure-6 declares the effect of using different configurations of the RGB color space on the tracking process. Each configuration has succeeded once except for the blue channel. Figure-7 shows the effect of using different configurations of the HSV color space on the tracking process. The most successful configuration is the 2D SV, while the least successful one is the 1-D hue channel. Figure-8 shows the effect of using different configurations of the YIQ color space on the tracking process. Here, the most successful configuration is the 2D YI representation. Figure-9 shows the effect of using different configurations of the YCbCr color space on the tracking process. The 2D YCb and the 2D YCr are the most successful configuration in the YCbCr color space. From Figure-10, it is very clear that the HSV color space with all of its seven configurations has the largest success percentage, then comes second the YIQ color space, and then comes the YCbCr color space in the third place, while the least successful color space is the RGB color space. Figure-11 shows that the 3D HSV has the largest successful percentage and the 3D RGB has the lowest one. Figure-12 shows that the 2D configurations in the HSV color space have the largest successful percentage and the 2D configurations in the RGB color space have the lowest one. Figure-13 shows that the single channels in the HSV color space are the most successful among other color spaces, while the single channels in the YCbCr color space are the least successful. It should be noted that the highest successful single channel among different color spaces is the luminance channel of the YIQ color space, and then comes the value channel from the HSV color space in the second place. Figure-14 shows that the 3D representations in different color spaces have the largest success percentage and the 2D representations have the second most success, while the 1-D representations have the lowest percentage. Generally, these results are logic since every added dimension adds new information which makes the representation more precise. But this should not contradict with the fact that we may find a single channel in a certain color space that is more appropriate than other 3D representations to represent an object for the tracking purpose, because the most issue that matters here is to represent the tracked object in a unique way that it can be distinguished from its background.
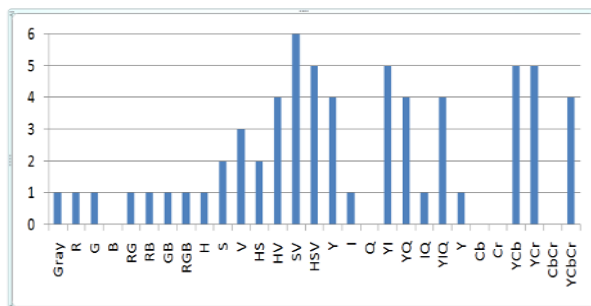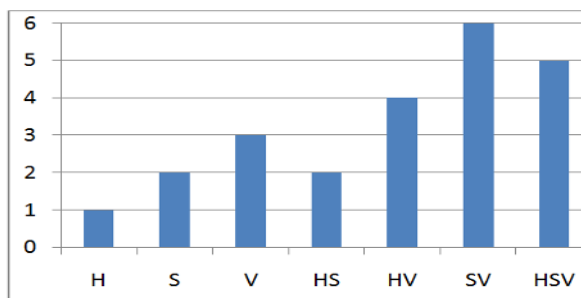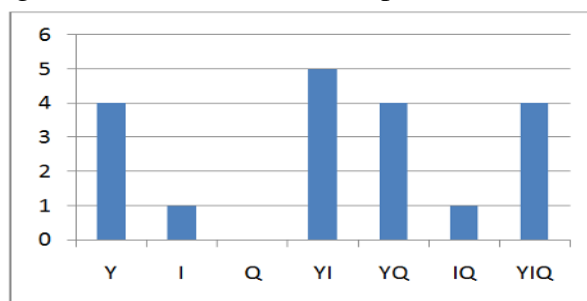
Fig. 5. ROI tracking results using different colors spaces with all available configurations
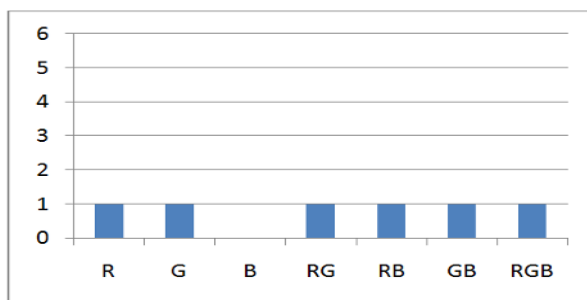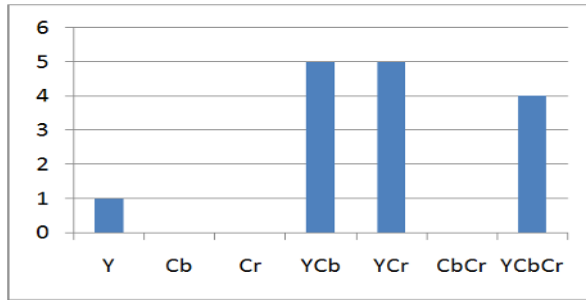

Fig. 6. ROI tracking results in the constructed six scenarios when using the seven available configurations in the RGB color space.


Fig. 7. ROI tracking results in the constructed six scenarios when using the seven available configurations in the HSV color space


Fig. 8. ROI tracking results in the constructed six scenarios when using the seven available configurations in the YIQ color space
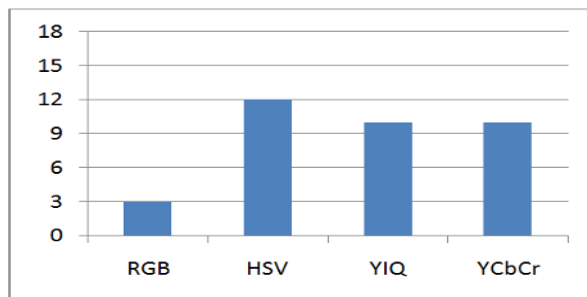
Fig. 9. ROI tracking results in the constructed six scenarios when using the seven available configurations in the YCbCr color space
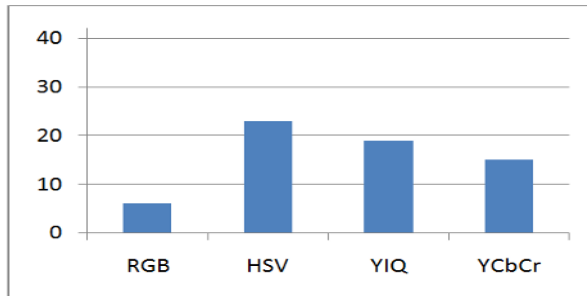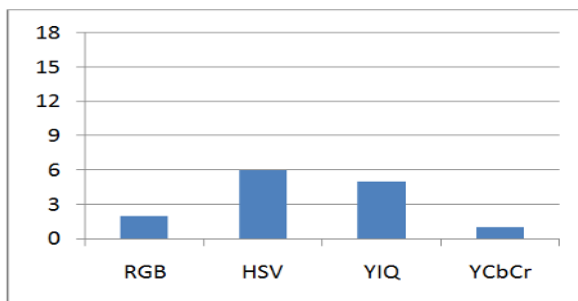

Fig. 10. ROI tracking results using all the seven configurations in RGB, HSV, YIQ, and YCbCr color spaces
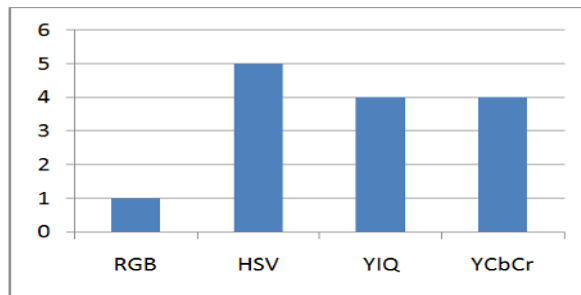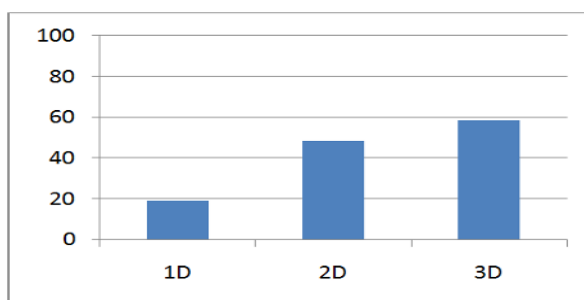

Fig. 11. ROI tracking results using 3D configurations of RGB, HSV, YIQ, and YCbCr color spaces

Fig. 12. ROI tracking results using 2D configurations of RGB, HSV, YIQ, and YCbCr color spaces

Fig. 13. ROI tracking results using 1D configurations of RGB, HSV, YIQ, and YCbCr color spaces

Fig. 14. ROI tracking results using different histogram dimensionalities

6. Conclusion and future work

We have used the mean shift algorithm as a tracker because it is considered a robust, fast, and widely used tool for object tracking that can use any feature space such as the color space. Again, realizing the powerful information provided by color spaces as a very important factor that should be considered when using the color feature for a particular task such as object tracking. Therefore, we have studied and analyzed the various effects of using different color space configurations with the mean shift algorithm on the tracking process in surveillance videos. We have used the gray space and four other color spaces (RGB, HSV, YIQ, YCbCr [each color space was studied under all the available seven configurations]). In the end, we have reached a conclusion that deciding which color space (or which configuration of the color space channels) to be used not only can be a very critical decision which may affect greatly the results of the tracking process, but also there is no color space configuration that can succeed all the time in all situations. Therefore, it is better to select the appropriate color space configuration according to the situation at hand. In other words, the selected color space configuration must have the maximum ability to distinguish the object from its background. This should be done automatically as a pre-step before applying the mean shift.

REFERENCES

[1] Fukunaga, Keinosuke, and Larry Hostetler. "The estimation of the gradient of a density function, with applications in pattern recognition." Information Theory, IEEE Transactions on 21.1 (1975): 32-40.
[2] Cheng, Yizong. "Mean shift, mode seeking, and clustering." Pattern Analysis and Machine Intelligence, IEEE Transactions on 17.8 (1995): 790-799.

[3] Bradski, Gary R. "Computer vision face tracking for use in a perceptual user interface.", Intel Technology Journal, 1 (1998).

[4] Comaniciu, Dorin, Visvanathan Ramesh, and Peter Meer. "Real-time tracking of non-rigid objects using mean shift." Computer Vision and Pattern Recognition, 2000. Proceedings. IEEE Conference on. Vol. 2. IEEE, 2000.

[5] Comaniciu, Dorin, and Peter Meer. "Mean shift: A robust approach toward feature space analysis." Pattern Analysis and Machine Intelligence, IEEE Transactions on 24.5 (2002): 603-619.

[6] Comaniciu, Dorin, Visvanathan Ramesh, and Peter Meer. "Kernel-based object tracking." Pattern Analysis and Machine Intelligence, IEEE Transactions on 25.5 (2003): 564-577.

[7] Smeulders, A.; Chu, D.; Cucchiara, R.; Calderara, S.; Dehghan, A.; Shah, M., "Visual Tracking: An Experimental Survey," accepted for publication in a future issue of the Pattern Analysis and Machine Intelligence, IEEE Transactions on, 2014.

[8] Ford, Adrian, and Alan Roberts. "Colour space conversions." Westminster University, London 1998 (1998): 1-31.

[9] Tkalcic, Marko, and Jurij F. Tasic. "Colour spaces: perceptual, historical and applicational background." Eurocon. 2003.

[10] http://homepages.inf.ed.ac.uk/rbf/HIPR2/gryimage.htm

[11] Vezhnevets, Vladimir, Vassili Sazonov, and Alla Andreeva. "A survey on pixel-based skin color detection techniques." Proc. Graphicon. Vol. 3. 2003.

[12] http://software.intel.com/sites/products/documentation/hpc/ipp/ippi/ippi_ch6/ch6_color_models.html

[13] Cardani, Darrin. "Adventures in hsv space." Laboratorio de Robótica, Instituto Tecnológico Autónomo de México (2001).

[14] N.A. Thacker and P.A. Bromiley. "Tina 5.0 User's Guide", University of Manchester, England, 2005.

[15] http://www.cvg.rdg.ac.uk/datasets/index.html

تأثير الفضاء اللوني على خوارزم متوسط الازاحة في عملية التتبع

الملخص العربي

يعد خوارزم متوسط الازاحة من أشهر وأكثر الطرق المستخدمة في التتبع السريع والمتين، كما أنه يستطيع أن يستخدم أي خاصية كخاصية اللون على وجه المثال. وبناءا على ذلك، فأننا نقدم دراسة تحليلية ومقارنة لتأثير مركبات النماذج اللونية أو الأفضية اللونية المختلفة على خوارزم متوسط الازاحة في عملية التتبع. ولقد استخدمنا في هذة الدراسة الفضاء الرمادي المتدرج بالأضافة إلى أربع أفضية لونية مع ملاحظة أن كل فضاء لوني تم دراسته باستخدام التكوينات الصبغية السبعة لمركباته. ولقد توصلنا إلى أن قرار تحديد الفضاء اللوني (أو التكوين الصبغي) المناسب ليس فقط قرارا حرجا جدا تعتمد عليه نتيجة عملية التتبع، ولكنه أيضا لا يوجد تكوين صبغي يمكنه النجاح طوال الوقت في كل الحالات. ولذا فأنه من الأفضل اختيار التكوين الصبغي المناسب طبقا للظروف الموجودة.