Journal of Fundamental and Applied Sciences

ISSN 1112-9867

Available online at http://www.jfas.info

Research Article

Special Issue

THE EFFECT OF ILLUMINATION COMPENSATION METHODS WITH HISTOGRAM BACK PROJECTION FOR CAMSHIFT APPLICATION

I. Saleh* and W. Rahiman

School of Electric and Electronic Engineering, Universiti Sains Malaysia, Engineering Campus, 14300 Seberang Perai Selatan, Pulau Pinang, Malaysia

Published online: 10 November 2017

ABSTRACT

This paper presents the results of a factorial experiment performed to determine the effect of illumination compensation methods with histogram back projection to be used for object tracking algorithm continuous adaptive mean-shift (Camshift). Since Camshift tracking can be used for distance approximation of an object, a precise tracker algorithm is required. This study compared two types of illumination compensation methods using Design of Experiment (DOE) in the presence of illumination inconsistency. Based on the results, it was found that selecting two channels as reference in histogram back projection weakens the accuracy of Camshift tracking whereas the combination of both methods produces results that are more desirable.

Keywords: object tracking; camshift; vision system; DOE.

Author Correspondence, e-mail: nims14_eee041@student.usm.my

doi: http://dx.doi.org/10.4314/jfas.v9i6s.47

1. INTRODUCTION

One of the parameters that could be manipulated from object tracking is distance estimation. However, one of the drawbacks of using vision system [14] in object tracking is that there are bound to have disturbances like occlusions and illumination inconsistency. Three methods which arethe estimator (Kalman filter), perspective transform (Inverse Perspective Mapping) and descriptor called Scale Invariant Feature Transform (SIFT) were reviewed in relation to their memory requirement and their efficiency in various environmental conditions.

A study adopting Kalman filter was conducted for optimal estimation of the state of a dynamic system, on the basis of fluctuating data peaks and an uncertainty in the system model dynamic [1-4]. A Kalman filter was implemented in the tracking to guess the locations of objects in the$(n + 1)^{th}$video frames using stereovision camera as an image input. The study found that Kalman filter was able to estimate the coordinate of a moving attribute point in the next frame and assessed the credibility of the estimation [1]. The processing time was also optimized as Kalman filter reduced the search area for redetecting an object which also led to lesser false positives.

Another method evaluated was Inverse Perspective Mapping. This method utilized inverse perspective mapping for distance estimation [5-6]. To predict the distance of the tracked object under various conditions, regression analysis was employed to derive the equation needed [6]. Linear regression models are straightforward and require minimum memory to be implemented, so they perform well on embedded controllers that have limited memory space [7]. Although linear regression is faster in comparison to neural networks, the inverse perspective matching method consumes a lot of memory space. Therefore, this combination was not an ideal combination. The system had been tested in various situations and was found to be efficient in distance estimation and object detection. However, it is extremely perceptive to changes in the environmental conditions [6]. It was also found that its tolerance to changes in light intensity can be improved by observing its effect on predicted distance value. The last method would be SIFT. SIFT is a descriptor for extracting distinctive invariant features from images that can be used to perform reliable matching between different views of an object or scene [8-9]. Further study however is needed to improve the accuracy of the method. Plus, this method requires a lot of processing resources to be implemented [9].

2. CAMSHIFT APPLICATION AS OBJECT TRACKER

The mean-shift algorithm is a robust method of finding local extrema in the density distribution of a data set. This is an affluent process for continuous distributions, which is essentially a hill climbing method applied to a density histogram of data. Fig. 1 shows how mean-shift tracker works. Assume B1 is the initial bounding box. Its center is marked with triangle. However, the true centroid point inside bounding box B1 is located differently (circle). Therefore, the mean-shift tracker will reiterate the center of the bounding box so that it can match with the previous centroid. As it iterates it will finally obtain bounding box with maximum pixel distribution which is B2 with its desired center marked with square. Supposedly the centroid and the desired center match or only differ with slightest error.

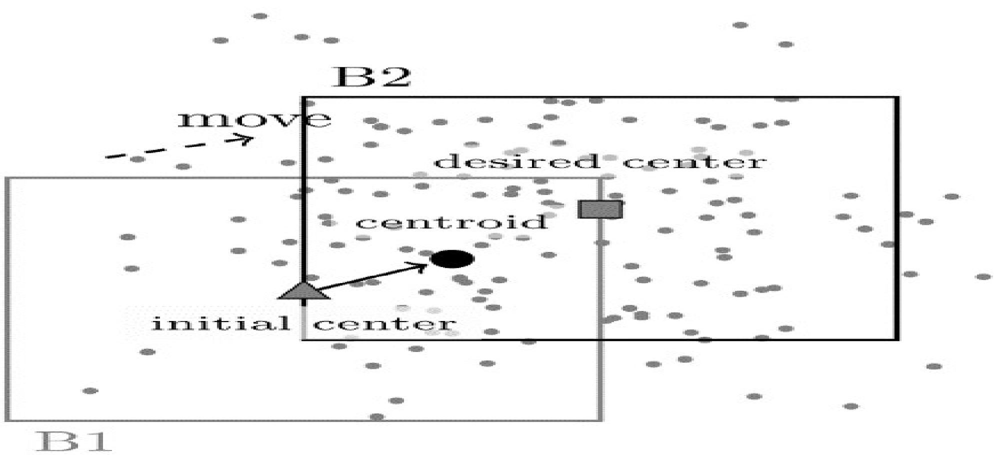

Fig.1. Mean-shift mechanism

Camshift (continuous adaptive mean-shift) differs from mean-shift in a way that the search window adapts itself in size. Suppose there are features whereby the pixels distributed evenly and stay compact, then this algorithm will automatically adjust itself for the size of the object as the object moves closer to and further from the camera. Fig. 2 shows the flowchart of Camshift. The highlighted box is an additional step provided by Camshift. Should there be well-segmented distributions (say object features that stay compact), then this algorithm will automatically adjust itself for the size of the object as the object moves closer to and further from the camera [10].

However, object tracking using vision system is susceptible to disturbances such as occlusions and illumination inconsistency. Therefore, light compensation method(s) must also be utilized in addition to tracking method, to get good estimation. A combination of Kalman filter and Camshift were previously implemented to address occlusion problems but not illumination inconsistency [2]. While there are intelligent descriptors that could be employed without light compensator to address these problems, the methods of interest should not consume too much memory and should work efficiently even at lower image resolution.

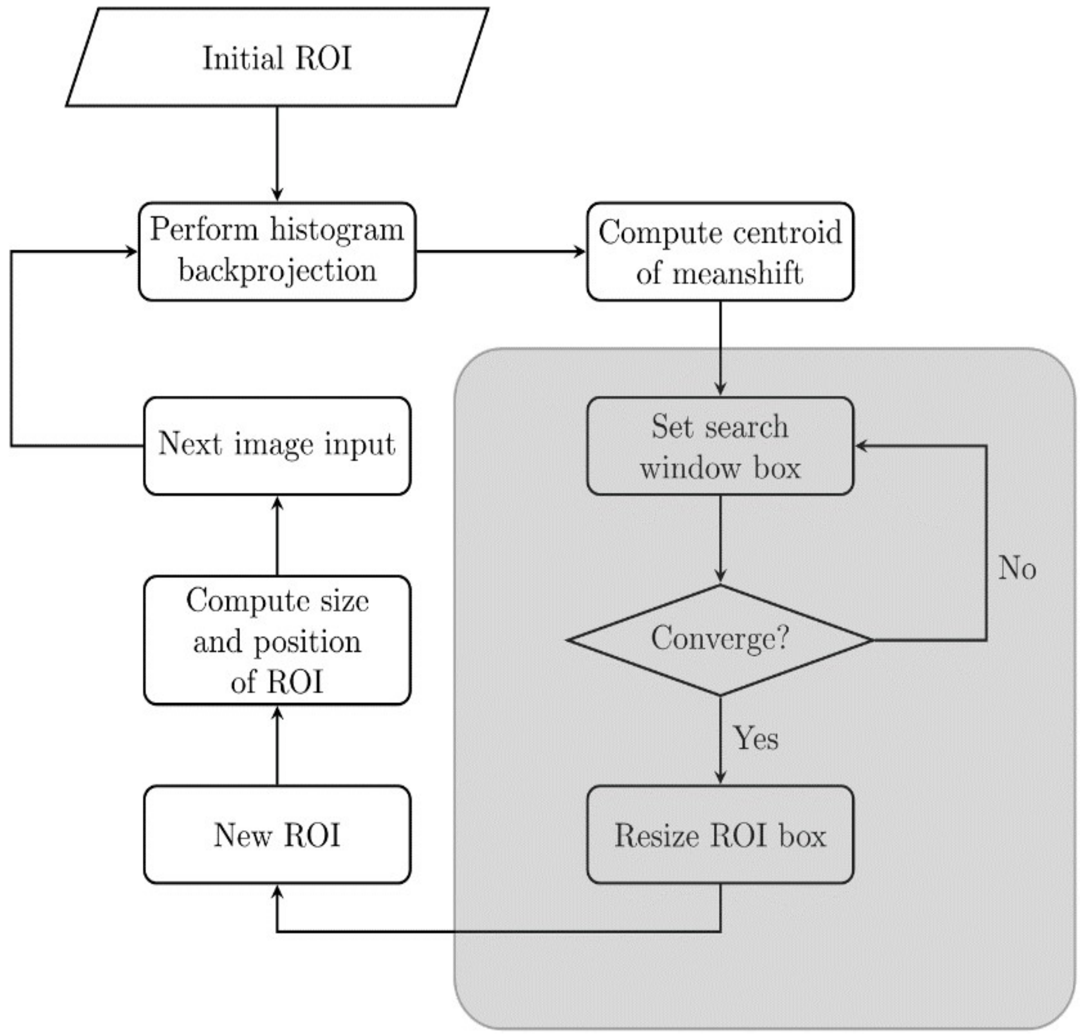

Fig.2. Camshifttracker flowchart

2.1. Distance Estimation with Camshift

There are several works that discusses the application of monocular vision camera. These works had proposed their own geometric model to calculate the object distance [4, 11-12] .

The geometric model in Fig. 3 shows that P is the target object, d is the distance between the monocular vision sensor and the target object, γ is the camera angle and h is the elevation height of the monocular vision sensor from ground level. Based on the geometry relationship, the distance d could be calculated.

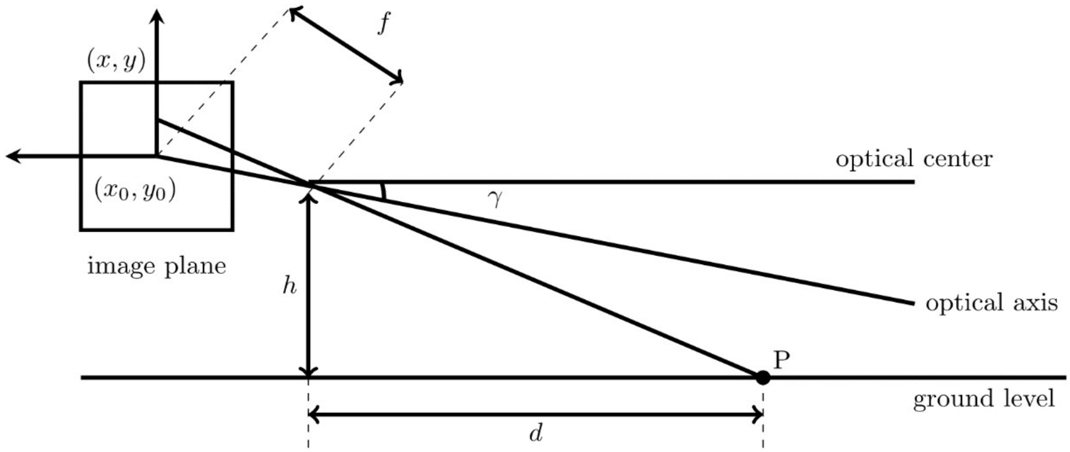

Fig.3. Distance estimation geometry model

Suppose that f is the focal length of the monocular vision sensor, (u_0, v_0) is the camera coordinate intersection point of optical axis and image plane. The physical dimension of a pixel corresponding to the x-axis and y-axis on the image plane are δ_x and $\delta_y$respectively, while the projection of P on the image plane (image coordinate) (x, y) can be represented as in Equation (1)-(5):

$$d = h/\tan\left(\gamma + \arctan\left((y - y_0)/f\right)\right) \tag{1}$$

$$\begin{cases} x = (u - u_0)\delta_x \\ y = (v - v_0)\delta_y \end{cases} \tag{2}$$

The optical point u and v or the apparent coordinate in image space can be represented by:

$$u = \frac{x}{\delta_x} + u_0, \ v = \frac{y}{\delta_y} + v_0 \tag{3}$$

By assuming (x_0, y_0) as origin:

$$x_0 = y_0 = 0 \tag{4}$$

The final equation distance d is:

$$d = h/\tan\left(\gamma + \arctan\left(\frac{(v - v_0)}{a_y}\right)\right) \tag{5}$$

whereγ and h are constants and v_0 and a_y can be obtained through camera calibration matrix and$a_y = f/\delta_y$. From the equation, it could be said that v influences the calculation of distance$d$.

3. METHODOLOGY

The objective of the experiment was to determine the effects of different types of illumination compensation methods on histogram back projection to be used for object tracking algorithm continuous adaptive mean-shift (Camshift). To achieve the objective the algorithms were coded in Python with the aid of OpenCV library. OpenCV (Open Source Computer Vision Library) is a computer vision and machine learning software library. It was developed as an open source to provide foundation for computer vision purposes and to provide function aid to the use of machine perception in the commercial products [10]. The algorithms were run on a single board computer (SBC) which is Raspberry Pi.

3.1. Image Segmentation Method

Fig. 4 shows the pseudo code performed for image segmentation. Firstly, the camera video input was divided into multiple independent frames. These individual frames are still digital images. The algorithm would process each n frame until the program was terminated by the user. Firstly, the program would prompt the user to select the ROI region. Each frame was then converted from its original colour space RGB to HSV colour space in real time. HSV colour space was chosen as it performed better than other colour space (for example L*A*B) [13]. With the sampled ROI the histogram back projection could be calculated and the resultant image was converted to white. Then, median filter was performed using filter of 3 x 3 kernel size and consequently morphological operation was applied. In addition to this, two steps were employed for this experiment.

Algorithm 1 Image Segmentation

i=320, j=240
while (camera=True) **do**
    read nth video frame
    $Hist_{ROI} \rightarrow [H_{min}, S_{min}, V_{min}], [H_{max}, S_{max}, V_{max}]$
    Convert nth frame colour space $BGR \rightarrow HSV$
    for $pixel = (0,0), pixel \leq (i, j)$
      increment pixel
      if $[H_{min}, S_{min}, V_{min}] \leq [H, S, V] \leq [H_{max}, S_{max}, V_{max}]$ **then**
        $pixel \rightarrow 0$ (white)
      else $pixel \rightarrow 1$ (black)
      end if
    end for
    Median filter
    Morphological operation
end while

Fig.4. Image segmentation algorithm

3.2. Illumination Compensation by Normalizing Y Channel

The first step was to normalize the intensity component of the image. This method was employed before image segmentation. The importance of equalizing intensity component is that even if the image was a darker image (which varied from the brighter image reference used), after equalization the output would be similar to the reference used. This will render the image to mimic similar lighting condition.

The process on equalizing Y channel is shown in Fig.5. The input image obtained from vision sensor was converted from its default colour space $(RGB)$to (YCC). This colour space was then split into its respective channels(Y, C_b, C_r). Out of these three channels, Y represents the luminance component. Therefore, Y channel was equalized using histogram equalization. The equalized$Y$ channel Y_n was then merged back with the other two channels (C_b) and (C_r).The resulting colour space was then converted back to RGB colour space.

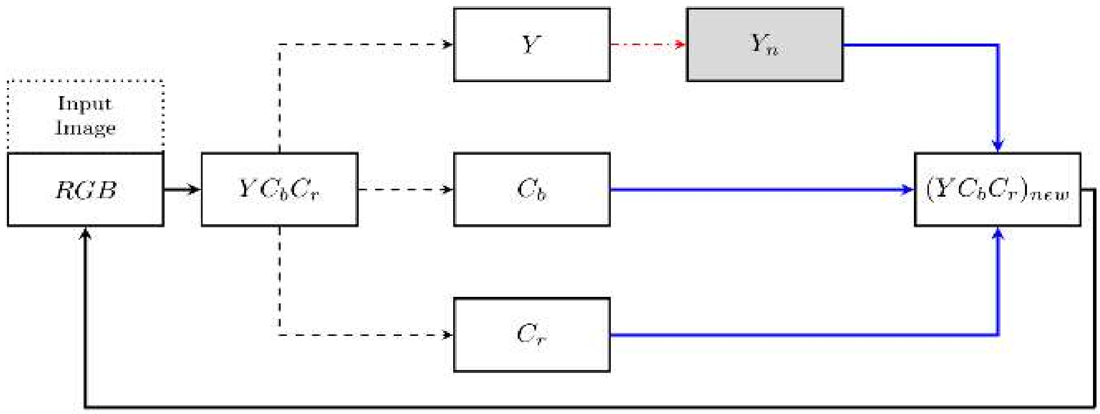

Fig.5. The process on equalizing Y channel

3.3. Isolation of V Channel in Histogram Back Projection

The second step was the selection of histogram properties to be used in histogram back projection. For tracking purposes the object of interest's colour was used as reference feature for histogram back projection. In HSV colour space,H channel was derived based on human perception of colour,S channel indicates the saturation of the hue and$V$ channel indicates the darkness/lightness of the image. Since V channel is influenced by the illumination, therefore this channel is not used as an image property in histogram back projection. Since V channel had been isolated, the remaining $H$and Schannels became the reference features for the histogram back projection.

3.4. Determining the Effect using Design of Experiment

In order to see the effects of these methods the design of experiment was performed. Factorial experiment was devised in order to test the effects of these methods. There were three factors which were used for this experiment. The experiment utilized a 2-level factorial method which means that each of the factors has two levels.

Table 1 shows the factors that had been selected with their respective levels. Factor A is the normalization of Y channel. The normalization of Y channel was used only for high level (+) and not for low level (-). Factor B is the reference properties for the histogram back projection. For low level (-) only H channel was used. For high level (+) two channels, H and Swere used. It would be irrelevant to set only S channel as one of the levels because the hue properties belong to H channel. Factor C is the light disturbance. During the course of the experiment, when the trial uses the combination high (+) level, the system will be tested at ideal condition.Whereas, when it is at low (-) level shadow is purposely introduced.

Table 1. Factor selections and its levels

Term	**Factor**	**Level**	
		-	+
A	Normalize Y	No	Yes
B	Channels	H	H+S
C	Light Disturbance	Yes	No

As discussed in section 2.1, v is the key component for distance estimation. Equation (6) shows the response selectedv_{err}. The response v_{err} from cycle i to n is the average difference between real time optical point v_{out} and the calculated value v_{calc} which was kept constant.

$$v_{err} = \sum_{i=1}^{n} \frac{|v_{out}-v_{calc}|_i+\cdots+|v_{out}-v_{calc}|_n}{n} \tag{6}$$

Since 3 factors were used there are altogether a total of 8 trials. The experiment was performed thrice (3 replicates) in order to verify the statistical significance of the effects. Therefore, the total number of trials was 24. The trials were performed in a totally randomized order to ensure that each trial has an equal chance of being affected by extraneous factors. Extraneous factors may affect the responses of the trials between replicates and within replicates.

3.5. Execution of the Experiment

Table 2 lists down the trial combinations of low (-) level or the high (+) level of each factors A, B and C. The flow of the execution of the experiment and its combination for each trial for each i^{th} image frame is then demonstrated in Fig. 6. The outputv_{err} will then be calculated.

Table 2. Factor selections and its levels

Trial	Factor A		Factor B		Factor C	
	Level	Notation	Level	Notation	Level	Notation
1	-	No Y_n	-	H	-	Shadow
2	+	Y_n	-	H	-	Shadow
3	-	No Y_n	+	$H+S$	-	Shadow
4	+	Y_n	+	$H+S$	-	Shadow
5	-	No Y_n	-	H	+	No Shadow
6	+	Y_n	-	H	+	No Shadow
7	-	No Y_n	+	$H+S$	+	No Shadow
8	+	Y_n	+	$H+S$	+	No Shadow

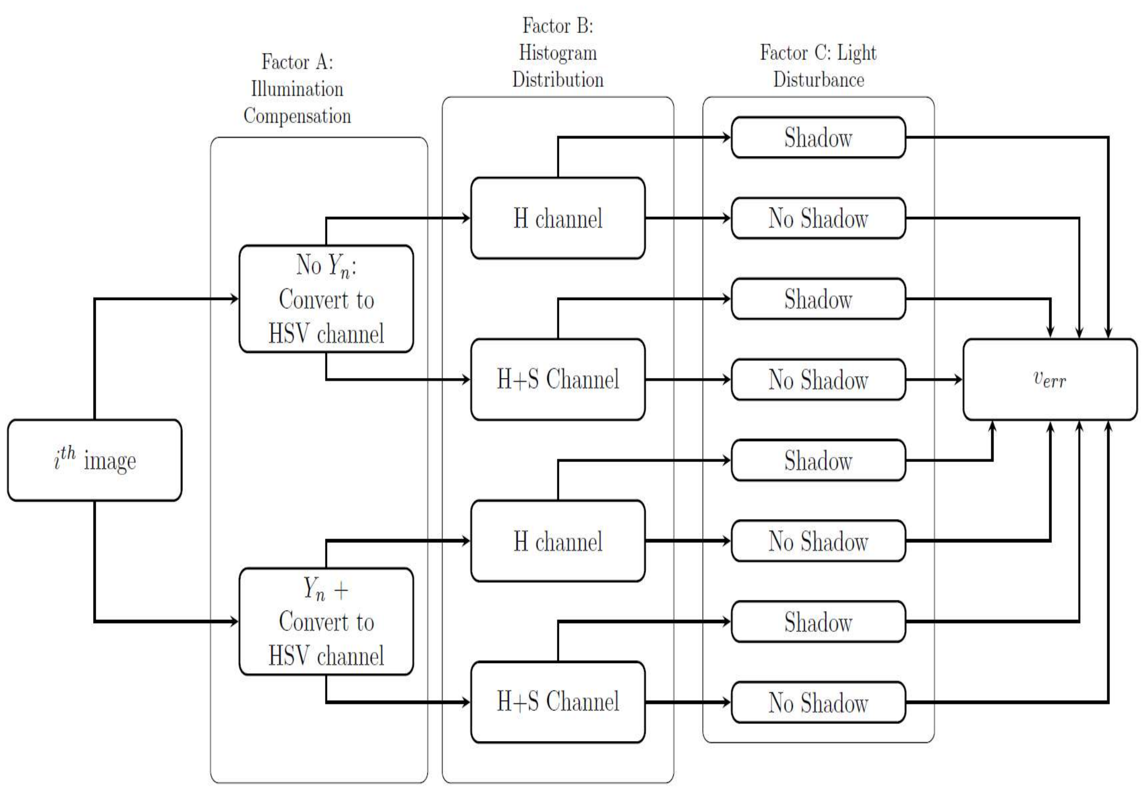

Fig.6. DOE sequence

4. RESULTS AND DISCUSSION

Table 3 shows the calculated effects from DOE performed and its consequent interaction. From the main effect table, it could be observed that the calculated effect and its consequent p-value show that factor B had produced the most significant result.

Table 3. Calculated effects and its interaction

Term	Effect	P	Term	Effect	P
Constant	-	0.00	AB	-0.22	0.71
A	-0.31	0.60	AC	0.00	1.00
B	2.61	0.00	BC	0.04	0.95
C	-0.53	0.37	ABC	-0.59	0.33

In order to study the interaction between the factors, interaction plot was mapped where factor A was kept constant. In Fig. 7, factor A was kept at high level which means that the luminance component was normalized. From the plot, it could be observed that the usage of H+S channel method produces high error regardless of the presences of light disturbance/shadow. On the contrary, the error was kept low and stable with the usage of only H channel.

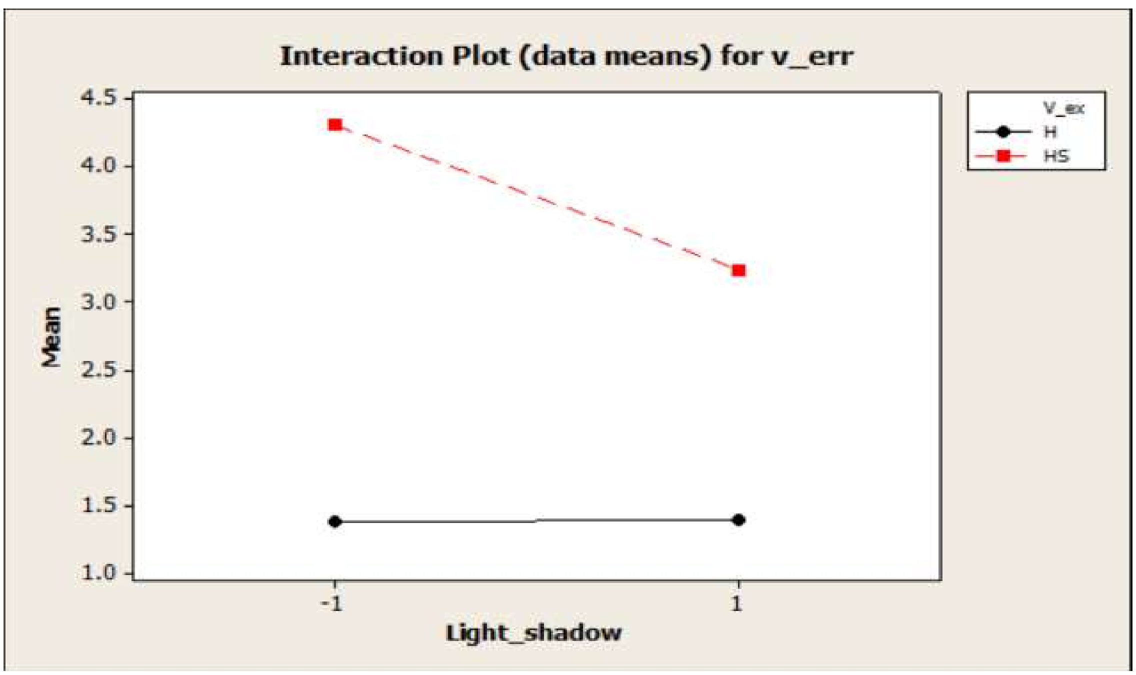

Fig.7. Illumination compensation method is used

In Fig. 8, Factor A is kept low which means that the luminance component is not normalized. From the plot it could be observed that the usage of $H + S$ channel also yields high error. With the usage of H channel the error is the least at ideal lighting condition (1). However, the error value escalates quite drastically in the presence of disturbance/shadow (-1).

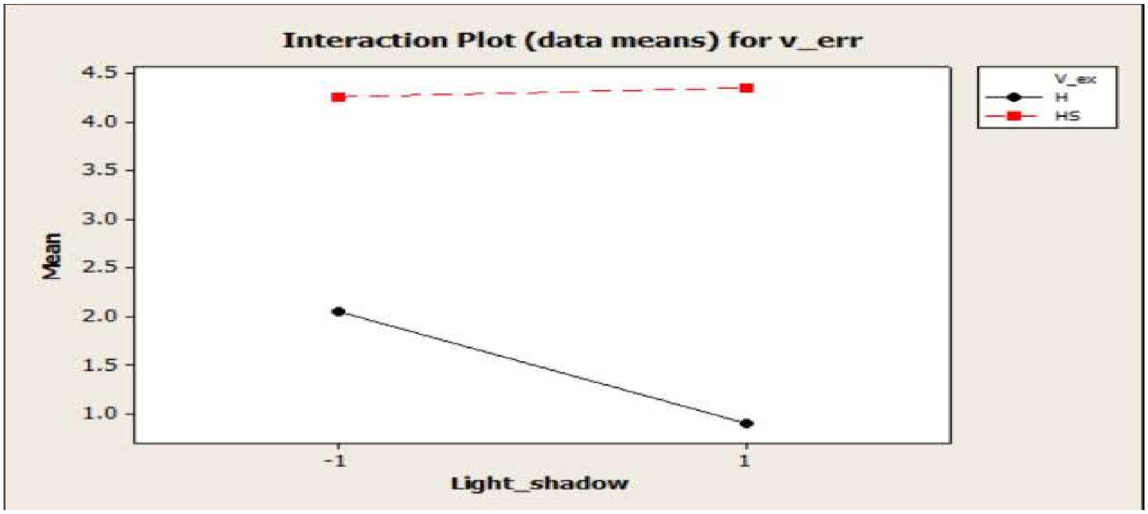

Fig.8. No illumination compensation method is used

5. CONCLUSION

It can be concluded from the experiment that the combination of Factor A at high level where the normalization of Y channel was used and Factor B at low level where only H channel was used, was able to keep the error low and stable (even before the implementation of filter to remove fluctuating peaks). These traits are desirable when approximating distance. Therefore, it could be said that the combination of both methods yield satisfactory result. Contrary to the assumption that the Camshift tracking algorithm accuracy will increase by introducing S channel as the reference for histogram back projection, it was found that it only deteriorates the performance of the tracking algorithm.

6. ACKNOWLEDGEMENTS

This research was supported by Multi-Agent System for Data Link Interaction in SPALAN Grant (grant number: 1001/PELECT/814265).

7. REFERENCES

[1] Khalid Z, Abdenbi M. Stereo vision-based road obstacles detection. In 8th IEEE International Conference on Intelligent Systems: Theories and Applications, 2013, pp. 1-6

[2] Le T, Combs M, Yang Q. Vehicle tracking based on Kalmanfilter algorithm. Technical reports, MSU Department of Computer Science, 2013,

https://www.cs.montana.edu/techreports/1314/Le.pdf

[3] Jiangwei C, Lisheng J, Lie G, Rongben W. Study on method of detecting preceding vehicle based on monocular camera. In IEEE Intelligent Vehicles Symposium, 2004, pp. 750-755

[4] Lien S F, Hsia K H, Su J P. Moving target tracking based on CamShift approach and Kalman filter. Applied Mathematics and Information Sciences, 2015, 9(1):395-401

[5] Tuohy S, O'cualain D, Jones E, Glavin M. Distance determination for an automobile environment using inverse perspective mapping in OpenCV. In Irish Signals and Systems Conference, 2010, pp. 1-6

[6] Bharade A, Gaopande S, Keskar AG. Statistical approach for distance estimation using inverse perspective mapping on embedded platform. In Annual IEEE India Conference, 2014, pp. 1-5

[7] Montgomery D. C. Design and analysis of experiments.New Jersey: John Wiley and Sons, 2013

[8] Lowe D G. Distinctive image features from scale-invariant keypoints. International Journal of Computer Vision, 2004, 60(2):91-110

[9] Zhu C. Video object tracking using SIFT and mean shift.Gothenburg: Chalmers University of Technology, 2011

[10]Kaehler A., BradskiG. Learning OpenCV 3: Computer vision in C++ with the OpenCV library. California: O'Reilly Media, Inc., 2016

[11]Lu W, Wang T T, Chu J H. The method of real-time distance measurement based on monocular vision. Advanced Materials Research, 2012, 403:1451-1454

[12]Li X, Wang L. A monocular distance estimation method used in video sequence. In IEEE International Conference on Information and Automation, 2012, pp. 390-394

[13]Bora D J, Gupta A K, Khan F A. Comparing the performance of L* A* B* and HSV color spaces with respect to color image segmentation. International Journal of Emerging Technology and Advanced Engineering, 2015, 5(2):192-203

[14] Zaman FH, Ali MH, Shafie AA, Rizman ZI. Efficient human motion detection with adaptive background for vision-based security system. International Journal on Advanced Science, Engineering and Information Technology, 2017, 7(3):1026-1031

How to cite this article:
Saleh I and Rahiman W. The Effect of Illumination Compensation Methods with Histogram Back Projection for Camshift Application. J. Fundam. Appl. Sci., 2017, *9(6S), 618-631*.