2014 International Conference on Advances in ICT for Emerging Regions (ICTer): 029 - 034

Frame Feature Tracking for Speed Estimation

Y.M.P.C.Y. Bandara[#1], U. A. A. Niroshika[#2] and T. U. Abeygunawardhane[#3]

[#1,2,3]*Department of Information Technology, Faculty of Computing, Sri Lanka Institute of Information Technology Malabe, Sri Lanka*

[1]IT11051794@my.sliit.lk, [2]aruni.n@sliit.lk and [3]IT11051862@my.sliit.lk

Abstract—**Frame feature tracking, relative pixel distance, relative pixel direction measurement and conversion of pixel distance to real world distance are the techniques used to estimate the speed of the vehicle. The balance between the performance and accuracy are the critical tasks in many existing speed detection techniques. In this paper, a novel speed estimation technique using feature tracking is presented. The main goal of this research is to increase the accuracy of the speed estimation while having an acceptable performance level. The proposed technique use feature tracking of subsequent image frames and estimate the speed using relative distance of the features. Experimental results with several traffic situations revealed that new technique has the desired accuracy with a satisfactory level of performance.**

***Keywords*— Feature Tracking, Region of Interest, Good Features**

I. INTRODUCTION

Measuring the speed with high accuracy is a requirement of many automated vehicle systems. Consider a vehicle active safety system developed by a vendor, which is different from the car manufacturer and the system, depends on the speed of the vehicle. To obtain the speed, gain access to the dashboard of the vehicle or to measure the interval between each wheel revolutions are possible options. In either of the above-mentioned possibilities, retrieval of the vehicle speed will depend on the technology used in the vehicle. The inbuilt vehicle system will need to be exposed or changed to accommodate the external safety system. There are certain situations where a vehicle owner will deny exposing or changing the car system [1], and it would have an adverse effect on the market of the safety system.

The objective of this research is to develop a speed estimation method which works independent from the inbuilt vehicle features (dashboard, rotating shafts) and which neither expose nor change the vehicle inbuilt features to obtain the current speed of the vehicle. The proposed method is intended to have a high accuracy level of ± 6%.

Lin and Wang [2] have introduced a new model based on frame feature detection and Hidden Markov Model. In this approach, a vehicular digital video recorder (VDVR) is used to record the path of the vehicle. These recorded images have been used to estimate the vehicle speed, and it compares the estimated speed with the actual speed. In this method, scale invariant feature transform (SIFT) [3] has been adopted for the feature matching while photogrammetry techniques [2] have been used to minimize the errors caused by the camera. The system uses Hidden Markov Model (HMM), which is a statistical model in which the system is assumed to be a Markov model with unobserved states. The HMM decides whether the vehicle is in a speed up state, speed down state or a constant state and the speed of the vehicle is calculated based on the current state of the vehicle and the speed of the vehicle in the previous frame.

The proposed technique uses the concept of image features (shapes, colours) of subsequent video frames to estimate the speed of the vehicle. There are two types of methods used in feature-based object detection: shape-based approach and colour-based approach. Even though, solely, colour is not always adequate to track features, the low computational cost in colour based feature matching is an advantage in high-performance systems [4].

Feature filtering process is responsible for removing unnecessary image features from the image frame using the properties of good features since the accuracy of the feature detection depends on the quality of the features [5].A good feature have two main properties: The feature should be able to track in subsequent image frames, and the feature should drift towards the midpoint of the frame [6].Good features to track has been adopted in this proposed method. It is important to reduce errors propagating by the camera to match and track good features. Photogrammetry techniques [2] can be used to reduce errors propagating by the camera.

In the process of feature matching and tracking between subsequent frames, issues regarding changes in image scales have to be addressed [2]. Scale invariant feature transform [3] and speed up robust features [7] techniques can be used to overcome these issues. Speed up robust feature technique has a low computational cost than scale invariant feature transform. Although many researches have been carried out in the area of feature matching and tracking, there is ample opportunity for high accuracy and precision research on feature matching and tracking [4].

In this study, pyramidal implementation of the Lucas-Kanade feature tracker description of the algorithm [8] is used for feature tracking. Feature matching between subsequent frames over adjacent frames are chosen to add flexibility in computational efficiency in this study. The selection of the subsequent image frames, (distance between the feature matching frames) are based on the camera frame capturing speed (frames per second) and the processing power of the processor used in this proposed study.

To avoid the errors propagating from different traffic situations, region of interest concept (ROI) [2] has been adopt in this method and this concept allows figuring out the directions in which the calculations needs to be done in order to estimate the speed of the vehicle.

The images retrieved by the video sequence are 2D images. The retrieved 2D image coordinate used to match and track the features. These results are then mapped into a 3D image coordinate system that will yield the world coordinates of the matched and tracked features [9]. The conversion is done with high accuracy since a small error could result in a huge variation in the final analysis.

A camera in which the video sequences are captured is a dependent factor in the above-mentioned conversion [9]. Different cameras have different intrinsic characteristics (focal length, skew, distortion, pixel error). Initially, the camera will be calibrated to obtain the camera specific characteristics [9].

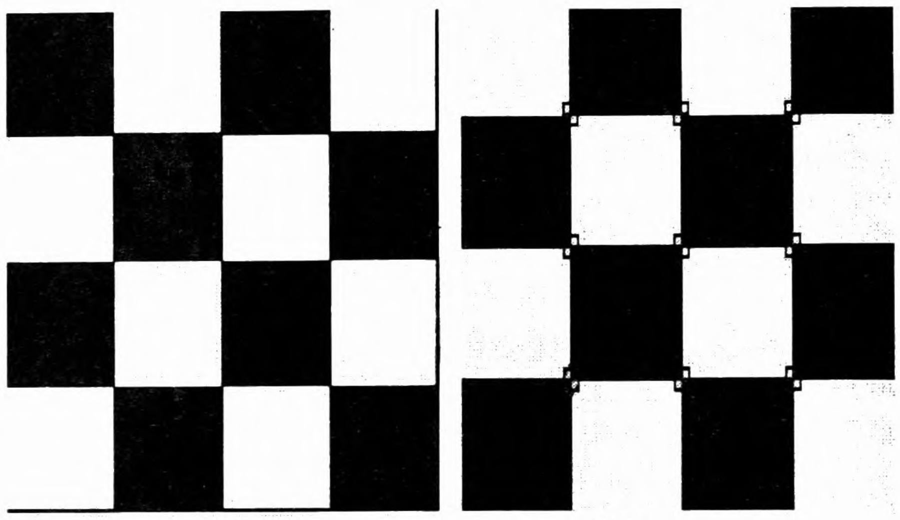
The proposed method initially identifies the good features to track in the subsequent image frames. The tracked features will be categorized into the regions and the real world distance travelled in each region will be computed. These distances will be an input to the voting algorithm which decides the best matching distance travelled in the subsequent frames. Inputs to the speed estimation calculations will be the output of the voting algorithm. Experimental results show an improved accuracy in speeds detected while having an accepted level of performance.

II. Methodology

The proposed technique introduces a novel method of speed detection that improves the accuracy of the speed estimation in computer vision while having the required performance level. The proposed technique can be described in four main steps as follows,

A. Camera Calibration.

Camera calibration toolbox [9]-[10] is used to calibrate the camera and intrinsic parameters such as focal length, skew, distortion, pixel error will be stored for computational purposes.



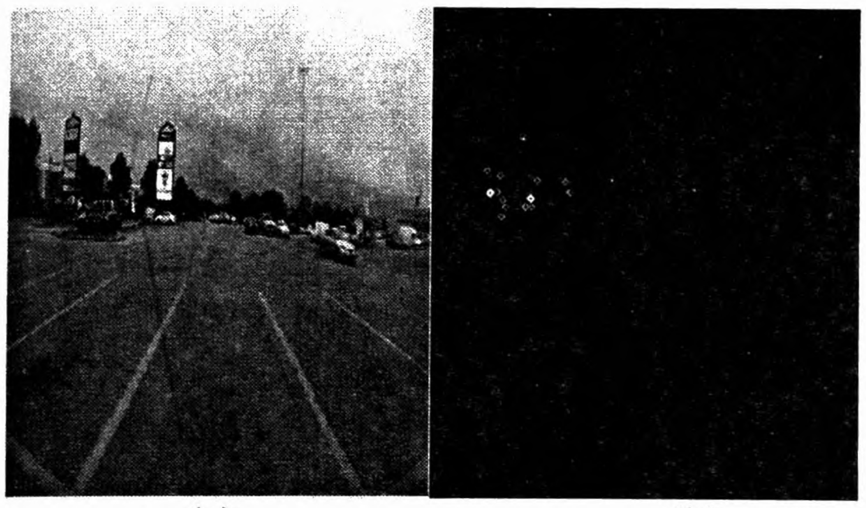
(a) (b)

Fig. 1 Images of camera calibration (a) Input image for calibration (b) output image of camera calibration.

Initially, chess board image as shown in Fig. 1 (a) is captured using the camera dedicated for speed detection. This image will be used as the input image for the camera calibration. Camera calibration method in [10] is adopted to calibrate the camera in this study. The projection of the world coordinates in the camera will depend on the camera intrinsic parameters and hence camera calibration has a direct effect on the final result [11].

B. Feature Matching and Tracking in Subsequent Frames.

Initially, the camera captures the video sequence and keeps the images of the required frames in the memory. The capturing of the video sequence will be done by a pinhole camera. Three subsequent image frames are stored in the memory, and the selection of the subsequent frames can be adjacent or distant than one frame from each other. Limitations in choosing subsequent image frames are that the distance between all the image frames has to be the same.



(a) (b)

Fig. 1 Feature tracking of subsequent frames (a) frame captured from the video sequence (b) features matched and tracked from the captured subsequent frames.

The features of the three subsequent image frames can be tracked and matched using Lucas-Kanade feature tracker [13].However, Lucas-Kanade algorithm works only for objects with slow-speed (displacement of the image contents of two subsequent image frames are small and approximately constant within a neighbourhood of a point) motions [13].The proposed method deals with slow speed motions and high-speed motions. The inability to track features of high-speed motion is a drawback of Lucas-Kanade algorithm. The pyramids provide a solution to the Lucas-kanade algorithm to track features of fast moving objects [14].

Pyramidal implementation of the Lucas Kanade feature tracker description of the algorithm [5] will be used to track the features of the subsequent image frames. This implementation enables to track both the slow moving features and fast moving features. Lucas kanade Feature tracker has a limitation on choosing subsequent image frames. The constant flow for all the pixels in a larger window will not be reasonable if the tracking is done at a longer period of time. The selection of the subsequent images should be small enough to hold the assumptions of the Lucas-Kanade algorithm.

As an example, suppose the current frame of the camera capture is F_t. Let the subsequent image frames are chosen 'n' frames apart from each other (assuming that the assumptions of Lucas Kanade algorithm holds for the subsequent image frames of distance 'n'). Then the subsequent image frames will be F_t, F_{t+i} and F_{t+2i}. Then, the frame features between F_t, F_{t+n} and F_{t+n}, F_{t+2n} will be matched and tracked respectively. Fig. 1 (a) is an image captured by a video sequence and the feature point will matched and tracked as shown in Fig. 1 (b).

To measure the distance travelled by the matched and tracked features, information about the relative depth changes are also important. 2D to 3D conversion based on motion, and colour mergence [12] has been adopted to obtain the 3D coordinate system in this study.

Let S_1, S_2,S_3 denotes the sets of good feature points matched and tracked using pyramidal implementation of the Lucas-Kanade feature tracker [8] between subsequent frames F_t, F_{t+n} and F_{t+2n} respectively. Let $Diff_i$ be the distance travelled by the tracked points in the sets S_i and S_{i+1}.Then the distance of the tracked features can be measured using the following formulas.

$$Diff_1 = \|a_i - b_i\| \; ; \; a_i \in S_1, b_i \in S_2 \quad (1)$$

$$Diff_2 = \|a_i - b_i\| \; ; \; a_i \in S_2, b_i \in S_3 \quad (2)$$

C. Region of Interest and Matched Features Categorisation

Images captured from the video sequence may include feature points that have adverse effects on the final analysis. To avoid these adverse effects, features that are matched and tracked have to go through a filtering mechanism. A good understanding of the pixel movements is a necessary when identifying the feature points which are needed to be filtered.

Consider a situation of a vehicle travelling forward. Then the matched and tracked feature points can be categorised into five regions as shown in Fig. 2 based on the relative movement of the features [2].

Let the regions be numbered as shown in Fig. 2. The basic characteristics of pixel movement directions differs in regions 1,2,3,4 and 5. Features tracked in region 1 have a tendency to move vertically upwards while features tracked in region 3 have a tendency to move vertically downwards. In a similar manner, Features tracked in region 2 have a tendency to move horizontally leftwards while features tracked in region 4 have a tendency to move horizontally rightwards. All the measurements are taken with respect to the vehicle moving forward.

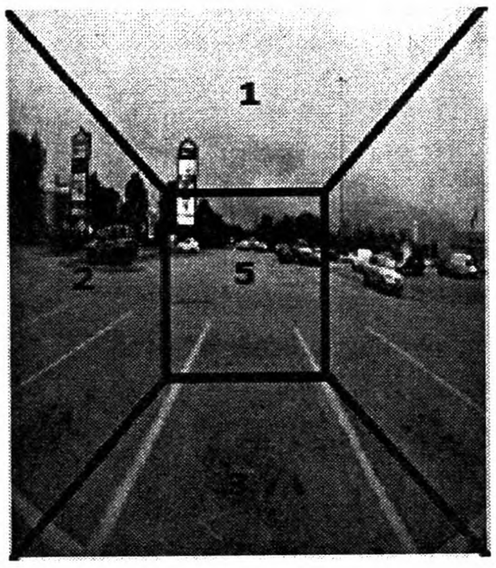
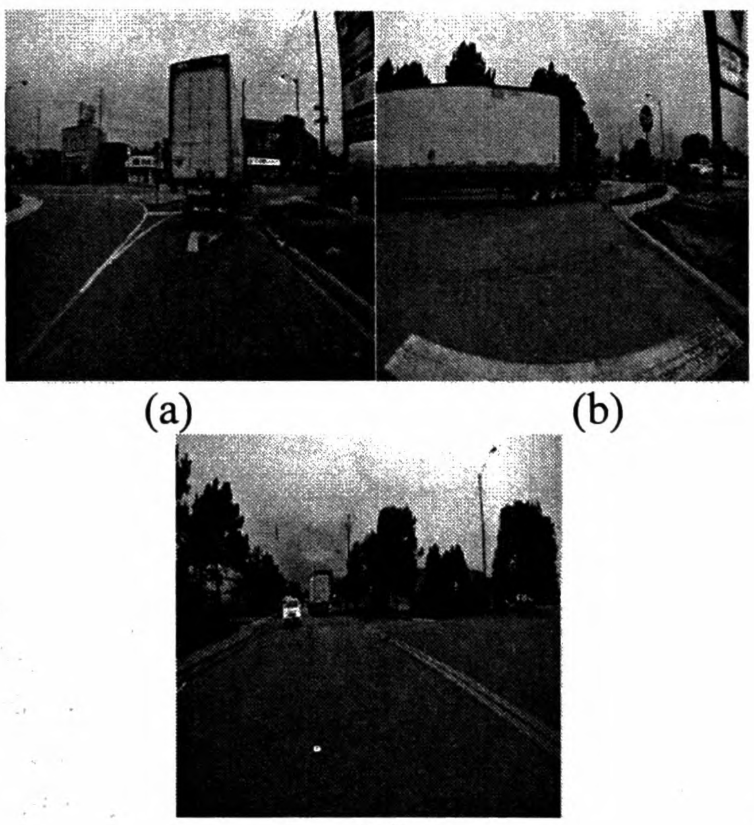


Fig. 2 Regions of interest in image frame

1) *Traffic Status Analysis:* There are three major traffic status to be countered in order to obtain accurate measurements of feature distance in subsequent frames [2].The first state of the traffic is the vehicle is stoped, but the vehicle in front is moving as shown in Fig. 3 (a). Since the new method only considers the feature movements of a particular direction, this can be countered using the above-mentioned theory. Fig.3 (b) depicts the second state of traffic analysis, a vehicle moving horizontally with respect to the vehicle. This movement will be captured in the regions two or four and with a threshold value for feature matching, and the tendency to move in those regions will nullify the effect of this traffic status. The third traffic status analysis is an overtaking vehicle or an oncoming vehicle as shown in Fig. 3 (c). Combination of the methods used to nullify the above mentioned two traffic states is implemented here to minimize the effect of this traffic status.



(a) (b)

(c)

Fig. 3 Traffic state analysis (a) vehicle in front is moving (b) horizontal movement of a vehicle (c) a vehicle is moving towards the camera in the oncoming lane.

Considering these facts, formulas mentioned in (1), (2) can be re-formulated as follows,

$$Diff_{1,j} = \|a_i - b_i\| \; ; \; a_i \in S_1, b_i \in S_2, (b_i - a_i) \in t_j \quad (3)$$

$$Diff_{2,j} = \|a_i - b_i\| \; ; \; a_i \in S_2, b_i \in S_3, (b_i - a_i) \in t_j \quad (4)$$

Where $Diff_{i,j}$ is the $Diff_i$ in j^{th} region and t_j is tendency to move in the j^{th} region.

D. Speed Estimation

1) *Average Distance of Features*: The matched and tracked points undergo calculations related to the regions in which the features are tracked and obtain the average distance of the tracked features of all the regions.

$$D_{1,J} = \frac{1}{i}\sum_{k=1}^{i} Diff_{1,j,k} \quad (5)$$

$$D_{2,J} = \frac{1}{i}\sum_{k=1}^{i} Diff_{2,j,k} \quad (6)$$

Where $D_{I,j,k}$ is the distance of the tracked point k has travelled in j^{th} region and $D_{i,j}$ is the average distance of all the tracked points in j^{th} region. The above average differences will be the input to the voting algorithm which is used with a predefined threshold value to obtain the average feature difference of the subsequent frames.

2) *Speed Estimation Procedure*:

Video capture
↓
Subsequent frames
↓
Feature matching and tracking
↓
Feature conversion to the real world coordinates
↓
Average distance travelled in each region
↓
Voting algorithm
↓
Speed Estimation

Fig. 4 flow of the speed estimation methodology

III. EXPERIMENTAL RESULTS

This section presents the results obtained from the proposed speed estimation method on the Mathlab version R2012a executed on an Intel Core i5-4200 CPU @ 1.60 GHZ in standard laptop version. For the understanding purpose of the effects of road situation on the proposed method, two types of road situations (traffic and non-traffic) are experimented under the proposed method. Experiments were carried out using a web camera having a frame rate of 30 frames per second (fps).

The experiments were carried out focusing on subsequent image frame capturing and traffic analysis to test the effect of each on the speed estimation. Subsequent image frames were captured one frame apart, and three frames apart, to test the effect of image capturing rate on speed estimation and experiments were carried out in two road situations (traffic and non-traffic) to verify the effect of road situation on speed estimation.

In table I and table III, experimental results of a vehicle travelling in a traffic situation are shown. In table II and table IV, experimental results of a vehicle moving in a non-traffic situation are shown. Experimental results of table I and table II corresponds to subsequent image frames were captured three frames apart and table III, and table IV corresponds to subsequent images were captured one frame apart.

TABLE I
SPEED ESTIMATED IN A TRAFFIC SITUATION WITH SUBSEQUENT IMAGE FRAMES TAKEN 3 FRAMES APART

Actual speed $V(Kmh^{-1})$	Estimated speed $V^{l}(Kmh^{-1})$	Error $\|V-V^{l}\|$	Percentage error (%)
20	20.82	0.82	4.12
40	41.9	1.9	4.75
60	63.04	3.21	5.35
80	84.28	4.73	5.92

TABLE III
SPEED ESTIMATED IN A NON-TRAFFIC SITUATION WITH SUBSEQUENT IMAGE FRAMES TAKEN 3 FRAMES APART

Actual speed $V(Kmh^{-1})$	Estimated speed $V^{l}(Kmh^{-1})$	Error $\|V-V^{l}\|$	Percentage error (%)
20	20.82	0.82	4.12
40	41.74	1.74	4.34
60	62.81	2.81	4.68
80	83.98	3.98	4.97

TABLE IIII
SPEED ESTIMATED IN A TRAFFIC SITUATION WITH SUBSEQUENT IMAGE FRAMES TAKEN 1 FRAME APART

Actual speed $V(Kmh^{-1})$	Estimated speed $V^{l}(Kmh^{-1})$	Error $\|V-V^{l}\|$	Percentage error (%)
20	21.10	1.10	5.05
40	42.22	2.22	5.55
60	63.67	3.67	6.12
80	85.08	5.08	6.35

TABLE IVV
SPEED ESTIMATED IN A NON-TRAFFIC SITUATION WITH SUBSEQUENT IMAGE FRAMES TAKEN 1 FRAME APART

Actual speed $V(Kmh^{-1})$	Estimated speed $V^{l}(Kmh^{-1})$	Error $\|V-V^{l}\|$	Percentage error (%)
20	20.94	0.94	4.71
40	41.96	1.96	4.92
60	63.12	3.12	5.21
80	84.38	4.38	5.48

A. Experiments on traffic situations.

The speed estimation method was tested in many traffic situations (relative movements of vehicles in front, horizontal movement of vehicles' relative to the camera capture, vehicles' moving to the direction of the camera) to identify the proposed method's ability to work in different traffic situations.

Table I and table III shows that the percentage error increases as the speed increases. Though the percentage error increases, the variation of the percentage is low.

B. Experiments on Non-Traffic situations.

The speed estimation method was tested in a non-traffic situation to identify the proposed method's functionality when the noise is limited and to compare the results with the results obtained in traffic situations.

Analysing the values on Table II and table IV shows that the percentage error increases as the speed increases.

The results shown in both Table I and Table II are having comparatively a lower error rate in low speeds while having a comparatively a higher error rate in high speeds. The same trend is visible in comparison of table III with table IV. One of the reasons for this change is the effect of blurring in image frames. When the vehicle speed is increased, the stationary features which are closer to vehicle have the same blurring effect as an object which moves fast relatively distant from the camera [15].

The accuracy of the non-traffic situation is high compared to the results obtained from the traffic situation. This shows that the noise features have been influenced to the calculations despite of the filtering process. As the speed increase, the amounts of noise features propagate to the calculations have been increased.

In comparison with the results of table I with table III, it is apparent that the distance between the subsequent image frames has an effect on the final results. One of the reasons for this is that every measuring instrument has a built in error. The error is significant when the measured distance is relatively low. The distance calculated with subsequent image frames taken one frame apart is less than the value of the distance taken three frames apart. It should also be noted that there is a maximum distance between subsequent images since Lucas-Kanade algorithm has its own limitations.

The comparison of percentage errors in traffic and non-traffic situations shows the important of traffic analysis. Robustness in the presence of noise in Lucas-kanade algorithm [16] is one of the reasons to have a lower error rate in both the situation while having the required performance in real time.

IV. Conclusions and Future Works

The problem of developing a vehicle independent speed estimation technique can be addressed using technologies in computer vision. In this paper, a new technique is introduced for speed estimation using the concept of frame features in computer vision. The proposed technique uses ROI concept to minimize the effects of traffic conditions and to improve the quality of the feature tracking. The extreme weather conditions and road conditions are the main limitation in this method. The proposed technique has an accuracy level of ± 6% and the maximum processing rate adopted is 30 frames per second (fps).

References

[1] Aa.co.nz, 'Vehicle Modification Rules & Regulations - aa.co.nz', 2014. [Online]. Available: http://www.aa.co.nz/cars/licensing-safety-fees/warrant-of-fitness-wof/vehicle-modification-rules-and-regulations/. [Accessed: 06- Jun- 2014].

[2] C. Lin and M. Wang, 'Vehicle speeding early warning model using frame feature detection and HMM', pp. 241–244, 2011.

[3] D. Lowe, 'Object recognition from local scale-invariant features', vol. 2, pp. 1150–1157, 1999.

[4] T. Ling, L. Meng, L. Kuan, Z. Kadim and A. Al-Deen, 'Colour-based object tracking in surveillance application', vol. 1, 2009.

[5] J. Shi and C. Tomasi, 'Good features to track', pp. 593–600, 1994.

[6] L. Yang, Y. Cai, A. Hanjalic, X. Hua and S. Li, 'Video-based image retrieval', pp. 1001–1004, 2011.

[7] H. Bay, T. Tuytelaars and L. Van Gool, 'Surf: Speeded up robust features', *Springer*, pp. 404–417, 2006.

[8] J. Bouguet, 'Pyramidal implementation of the affine lucas kanade feature tracker description of the algorithm', *Intel Corporation*, vol. 5, 2001.

[9] S. Indu, M. Gupta and A. Bhattacharyya, 'Vehicle tracking and speed estimation using optical flow method', *Int. J. Engineering Science and Technology*, vol. 3, no. 1, pp. 429–434, 2011.

[10] V. Douskos, I. Kalisperakis and G. Karras, 'Automatic calibration of digital cameras using planar chess-board patterns', vol. 1, pp. 132–140, 2007.

[11] S. Chhaniyara, P. Bunnun, L. Seneviratne and K. Althoefer, 'Optical flow algorithm for velocity estimation of ground vehicles: A feasibility study', *International Journal on smart sensing and intelligent systems*, vol. 1, no. 1, pp. 246–268, 2008.

[12] F. Xu, G. Er, X. Xie and Q. Dai, '2D-to-3D conversion based on motion and color mergence', pp. 205–208, 2008.

[13] S. Baker and I. Matthews, 'Lucas-kanade 20 years on: A unifying framework', *International journal of computer vision*, vol. 56, no. 3, pp. 221–255, 2004.

[14] L. Li and Y. Yang, 'Optical flow estimation for a periodic image sequence', *Image Processing, IEEE Transactions on*, vol. 19, no. 1, pp. 1–10, 2010.

[15] Ee.iitm.ac.in, 'Research', 2014. [Online]. Available: http://www.ee.iitm.ac.in/ipcvlab/research. [Accessed: 05- Jun- 2014].

[16] A. Bruhn and J. Weickert, 'Towards ultimate motion estimation: Combining highest accuracy with real-time performance', vol. 1, pp. 749–755, 2005.