International Journal of Information, Control and Computer Sciences
ISSN: 2517-9942
Vol:2, No:12, 2008

Motions of multiple objects detection based on video frames

Ms. Khin Thandar Lwin , Dr. Than Htike, Dr. Zaw Min Naing
Department of Electronic Engineering
Mandalay Technological University
Mandalay, Myanmar
lwinlwinth@gmail.com

***Abstract*—:** This paper introduces an intelligent system, which can be applied in the monitoring of vehicle speed using a single camera. The ability of motion tracking is extremely useful in many automation problems and the solution to this problem will open up many future applications. One of the most common problems in our daily life is the speed detection of vehicles on a highway. In this paper, a novel technique is developed to track multiple moving objects with their speeds being estimated using a sequence of video frames. Field test has been conducted to capture real-life data and the processed results were presented. Multiple object problems and noisy in data are also considered. Implementing this system in real-time is straightforward. The proposal can accurately evaluate the position and the orientation of moving objects in real-time. The transformations and calibration between the 2D image and the actual road are also considered.

***Keywords*—** Motion Estimation, Image Analyses, Speed Detection

I. INTRODUCTION

REAL -time motion analysis is becoming mandatory for many applications, which attracted many research works in this area. On this basis, many opportunities arise for new applications, and many others could be fully exploited. Some system that was infeasible with traditional techniques can now be implemented. Several examples can be observed, such as in the areas of cinema, virtual reality, medical imaging, and automated inspection. In most of these applications, real time synthesis of actors animated by the estimated motions and deformations is needed. For this reason, the integration among motion analysis and synthesis techniques assumes a high relevance on the quality of the results. This project makes use of real-time imaging algorithm for speed detection of moving objects. Survey study identified the drawbacks of typical applications, such as those used by the police speed trap. Then based on the weakness of the current speed detection radar, image-processing techniques were applied to tackle the problem. The feasibility of applying the proposed image processing technique with applications in this area will be discussed and analysed.

II. 3D SPACE RECONSTRUCTION

In the motion analysis of image sequences, Optical flow computation is a fundamental problem. It has been proved that optical flow is generally different from the motion field, the projections of the motions of scene points relative to the observer. Lai and Vemuri [6] present two algorithms for computing optical flow. They are the Modified Gradient-based Regularization (MGR) method and the SSD-based regularization method. The gradient-based method is used to amend the errors in the discrete image flow equation caused by numerical differentiation as well as temporal and spatial aliasing in the brightness function. The image flow constraint and a contour-based flow constraint are selectively combined into the data constraint by using a reliability measure. Each data constraint is appropriately normalised to obtain an approximate minimum distance constraint instead of the conventional linear flow constraint. Such modifications result in robust and accurate optical flow estimation.

An extension to the inverse perspective mapping geometrical transform, which has demonstrated to be robust with respect to vehicle movements, was proposed by Bertozzi et. al. [3]. Obstacles can also be detected with a high confidence even in case the camera orientation and camera height have drifts up to ±1° and ±10cm. The advantage of employing the technique is that only a simple match of a model encoding with a prior knowledge is needed, instead of an exhaustive search for homologous points is required.

Instead of model-free approach, Hagg and Nagel [5] incorporated optical flow measurements and explicitly model occlusions for the tracking of partially occluded vehicles in image sequences adopt Model-based tracking approach. The consequent exploitation of model knowledge result in robust tracking results even for complex driving manoeuvres significant occlusions, low contrast and small vehicle images.

Algorithm for local 3D reconstruction of a road from its image plane boundaries was proposed by Guiducci [1], which assumed that a stationary camera records the complex driving manoeuvres, so the field of view is suitably large to monitor the road. However, a large monitoring area may leads to too small vehicle images, which will affect the resolution of the object. Here the proposed algorithm is based on a model-free approach instead of a model-based tracking approaches which extract image features in separate frames to track a moving object in any instant. Koller et. al. [2], proposed an algorithm to extract the edge segments and associates these with model segments obtained by projecting a vehicle model into the image plane. The segmentation of an optical flow field achieves the initial pose estimation. Lin and Nevatia [4] proposed static object detection by a single monocular aerial image, in which buildings are detected and 3-D shape descriptions of buildings can then be constructed.

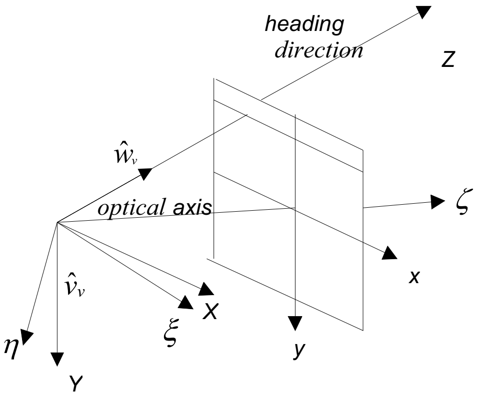

Figure 1. The camera reference frame

III. Notation and Calibration

In Figure 1, the camera axes are indicated by ξ, η, and ζ, with ζ in the direction of the optical axis, ξ in the direction of the scan lines, and η downward to form a right-handed reference frame. Considering the image plane being at unit distance from the projection centre along the optical axis ζ and the image plane coordinates are (x,y) with the x and y axes parallel to ξ and η, respectively. The camera frame and the world frame are connected by rotation matrix R^T, where

The calibration consists of the determination of the mapping between pixel plane of the camera and the 3D world directions. The intrinsic calibration maps each pixel to a unit

$$R^T = [\hat{u}, \hat{v}, \hat{w}].$$

vector in the camera reference and assumed known by any of the standard procedure. The extrinsic calibration gives the position and orientation of the camera reference with respect to a suitably chosen world reference frame as Figure 2. For this paper the most convenient choice of the world reference is a reference oriented as the road in correspondence of the vehicle. Then the extrinsic calibration consists of the determination of the components of the unit vectors of the world reference in the camera reference and of the instantaneous position of the camera with respect to the road is employed. The determination of the vehicle triad and the on-road calibration were performed according to Guiducci [1].

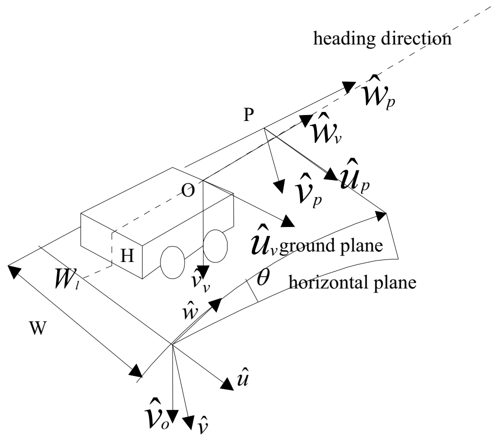

Figure 2. The vehicle, world, and reference frames

IV. Computation Errors

There are various sources of error and they include (1) Error due to the Image Noise, (2) Errors due to the Algorithm and (3) Error due to Transformation. Detailed analysis of these errors is not detailed here as they can easily be corrected mathematically. When the video frames are being captured, it is inevitable for quantisation error to occur. Also, the system requires the source, which must be in YUV format. Undoubtedly, truncation error and rounding error occur in the process converting the frame format from RGB to YUV. In the following, the error estimation is shown. The exact RGB to YUV transformation, defined by the CCIR 601 standard, is given by the following transformation: $Y = 0.299R + 0.587G + 0.114B$, $U=0.564(B-Y)$ and $V=0.713(B-Y)$, where Y is the luminance component, and (U, V) are two chrominance components. For computer display representation, all values are required to be integer. It results in truncation error or rounding error. With resolution of and *256* gray levels, other errors are negligible. The percentage of quantisation error for each pixel is equal to $\varepsilon = (1/256)\ X\ 100\% = 0.3906$. For the entire image of *256X256* pixels, the maximum probable noise power, *(N)*, is given by

$$N = \frac{1}{L_1 L_2}\left\{\sum_{y=0}^{L_2-1}\sum_{x=0}^{L_1-1}\left[f(x,y) - r(x,y)\right]^2\right\} = 1$$

Assuming that the image is well equalised and the average value per pixel is equal to 128. The maximum probable signal power, *(S)*, is given by

$$S = \frac{1}{L_1 L_2}\left\{\sum_{y=0}^{L_2-1}\sum_{x=0}^{L_1-1}\left[f(x,y)\right]^2\right\} = 128^2 = 16384$$

and the corrresponding signal-to-noise ratio (S/N) can be computed by 10 $\log_{10}$ (16384) and the peak signal to noise ratio by 10 $\log_{10}$ (256^2) respectively. The corresponding values are (S/N)= 42.144 dB and Peak (S/N) = 48.165 dB.

V. System Implementation

1) Hardware Description

In this project, the complete real-time image processing system was not implemented, instead an off-line approach was used for evaluation purposes. The off-line system comprised of two parts; namely, the image captures hardware and the image processing system. However integrating these two parts is straightforward. For the convenience of off-line analysis, an image database was set-up first. A video camera was used to capture several series of video frames in different locations. The database was built up using these real-life captured video frames. To facilitate the video capture process, the miroVideo DC20 capture interface was employed. The system comprises of PC hardware to digitise the video information and software to process the digitised images. The complete system configuration is as shown in Figure 3. After the video is converted into AVI format in the laboratory, a sequence of frames can be extracted from the AVI file. All the frames being processed are in RAW data format. With the captured data files, image-processing techniques were used to extract the multiple objects from individual video frame. Subsequent to that, the location of each moving object was identified and the corresponding speed were abstracted.

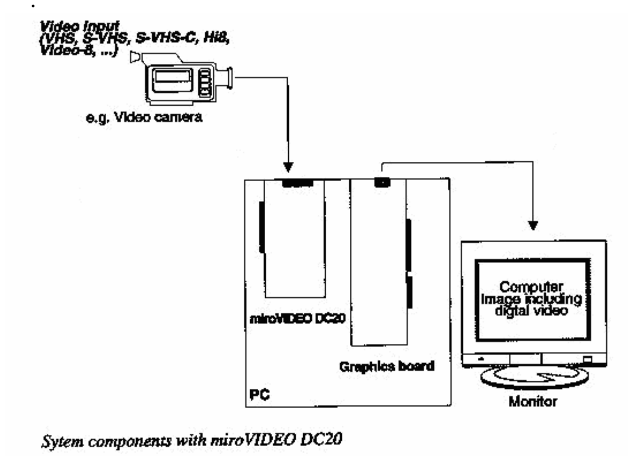

Figure 3 Complete Processing Systems

2) Software Description and Algorithm Outline

The software is developed using gcc running under Microsoft Window environment. The reason to select the development platform and the compiler is based on their code efficient. Samplings of every 30 frames per second will provide sufficient accuracy, during the image processing only odd number of frames were being used for computation. By using the Intel PIV- 1.60GHz CPU, the software package can process images in real-time.

Foreground and background separation is first performed follows by filtering out the noise on the image by Morphology and other image filters. Locating each moving objects by clustering and connectivity follows by estimation of the distance traveled for each moving object in a succession of images.

Pseudo Code

```
#include<stdio.h>
#include<time.h>
# ...
…
void main()
{…
  unsigned foreground[256][256];
  unsigned foreground[256][256];
  …
  FILE *fp;
     fp=fopen("0.raw","rb");
//get background reference image frame
…
fclose(fp); }
…
//read other frames with moving objects
//separation of background & foreground
//find out the location of each object
//calculate distance traveled by objects in
successive frames
```

VI. Sample Results

Figure 4(a) Background of Location

Figure 4(b) Video frame with occurrence of multiple moving objects

Figure 4(a) - (c) shows samples of video frame as taken on a popular piece of road that police speed trap was set. Figure 4(a) is the background of the road, which is used as a reference for the subsequent image processing. Figure 4(b) is a typical

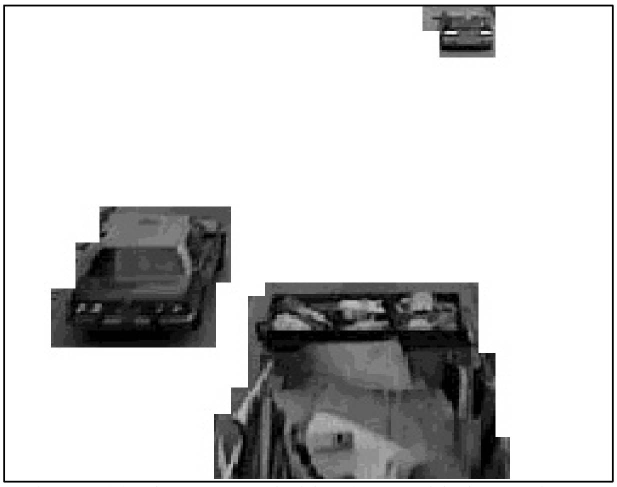

Figure 4(c) The processed video frame with the moving objects

video frame with multiple moving objects. Figure 4(c) indicated the processed video frame with the moving objects being extracted from the background. With consecutive video frames and the corresponding extract frames like Figure 4(c), Figure 4(c) Image showing multiple objects being extracted a succession of time sequence frame with multiple object being extracted can be obtained. Following the 3D-reconstruction algorithm and with know timing between consecutive frame, speed of individual objects can be evaluated accordingly. Samples of two partial outputs listing of the program execution are as follows;

1) Sample 1 - location of multiple objects

Partial program output with the snap shot sequence including Figure 4(c)

```
Moving object at
x=166 y=53 total 630 pixels
Moving object at
x=207 y=152 total 910 pixels
Moving object at
x=16 y=193 total 68 pixels
```

1) Sample 2 - speed evaluation of multiple objects

```
Reading background image file
<1>    background.raw
Reading consecutive image files
<2>    1.raw
<3>    3.raw
<4>    5.raw
***
***
***
<169>  335.raw
<170>  337.raw
***
***
Object 4: velocity = 63.26 km/hr
Date/time Sun May 02 18:12:23 1999
Object 5: velocity = 111.68 km/hr
Date/time Sun May 02 18:12:32 1999
```

Sample 1 identified 3 moving objects with their location and relative size reported. In sample 2 there are multiple objects being detected and the speeds of the moving object 4 and object 5 are calculated to be approximately 63 km/hr and 111 km/hr respectively. With the legal speed limit in Hong Kong for that piece of road being 50 km/hr, object 4 had exceeded 13 km/hr, which is liable to a fix penalty. However, object 5 is exceeded by 61 km/hr, which is subjected to prosecution and possible suspension of license.

VII. Conclusion

This paper demonstrated the application of image-processing techniques for speed detection. In the past, the traditional speed detectors make use of microwave, infrared or even the laser. Such previous systems have many drawbacks, as these previous detectors are easily jammed and defeated by the fast-changing technology. Also, there are numerous reflection errors for current detector. In view of this, this paper focuses on tackling the faults of current speed traps. An image processing approach has been employed. In order to obtain the speed of moving object, the foreground and background are separated. During the processing, noise is inevitable therefore, image-processing filters are employed to filter out noise. Since the foreground (moving object) is found out, its location is easily taken. Accompanying with the Road's calibration, the exact location and the time between successive frames can be used to evaluate the speed of multiple moving objects.

Acknowledgment

The kindness and help of Dr. Zaw Min Aung , Dr. Zaw Min Naing, Dr. Hnin Si and Dr. Yin Mon Myint, Department of Electronic Engineering, Mandalay Technological University are gratefully acknowledged. Especially, I would like to express my special thank to my parents for their noble support and encouragement.

VIII. References

[1] A. Guiducci, “3D Road Reconstruction from a Single View”, Computer Vision and Image Understanding, vol. 70, no. 2, 1998, pp.212-226.

[2[D. Koller, K. Daniilidis, H.-H. Nagel, "Model-based object tracking in monocular image sequences of road traffic scenes", Int. Journal of Computer Vision vol. 10, 1993, pp.257-281.

[3] M. Bertozzi, A. Broggi and A. Fascioli, "Stereo Inverse Perspective Mapping: Theory and Applications", Image and Vision Computing, vol.16, 1998, pp.585-590.

[4] C. Lin and R. Nevatia, "Building Detection and Description from a Single Intensity Image", Computer Vision and Image Understanding vol. 72, no. 2, 1998, pp.101-121.

[5] M. Haag and H.-H. Nagel, "Tracking of Complex Driving Maneuvers in Traffic Image Sequences", Image and Vision Computing, vol. 16, 1998, pp.517-527.

[6] S. Lai and B. C. Vemuri, "Reliable and Efficient Computation of Optical Flow", Int. Journal of Computer Vision, vol. 29, no. 2, 1998, pp.87-105.