國立交通大學

生醫工程研究所

碩 士 論 文

以反透視映射模型與移動補償為基礎的障礙物偵測應用在倒車行為的安全上

Generic Obstacle Detection for Backing-Up Maneuver Safety Based on Inverse Perspective Mapping and Movement Compensation

研 究 生：李嶸健

指導教授：林進燈 教授

中 華 民 國 九 十 九 年 七 月

以反透視映射模型與移動補償為基礎的障礙物偵測應用在倒車行為的安全上

Generic Obstacle Detection for Backing-Up Maneuver Safety Based on Inverse Perspective Mapping and Movement Compensation

研 究 生：李嶸健　　Student：Jung-Chien Lee

指導教授：林進燈　　Advisor：Chin-Teng Lin

國 立 交 通 大 學

生 醫 工 程 研 究 所

碩 士 論 文

A Thesis

Submitted to Institute of Biomedical Engineering

College of Computer Science

National Chiao Tung University

in partial Fulfillment of the Requirements

for the Degree of

Master

in

Computer Science

July 2010

Hsinchu, Taiwan, Republic of China

中華民國九十九年七月

以反透視映射模型與移動補償為基礎的障礙物偵測應用在倒車行為的安全上

學生：李嶸健　　指導教授：林進燈 博士

國立交通大學生醫工程研究所

摘要

近年來，交通事故造成人員受傷或死亡的數量仍然居高不下，因此世界各地提升安全性的高效能車輛與日俱增，其中車輛與其他障礙物碰撞事故發生最為頻繁，很多研究人員提出利用立體視覺的一般障礙物偵測技術，然而，影像式駕駛輔助安全系統中在移動車輛中以單支攝影機的障礙物偵測仍然是一個困難且具有挑戰性的問題。在本篇論文中，我們提出一個以移動補償與反透視映射模型為基礎的單眼影像式障礙物偵測技術，提出一個創新的地面移動量估測技術藉由利用路面偵測取得影像中最有用的地面特徵點，並且分析特徵點的光流主要分布，可以準確的獲得地面移動補償量，此外，利用反透視映射模型對影像中平面與非平面物體的不同特性來偵測障礙物。為了要得到更可靠的偵測結果與最靠近的可能碰撞位置，一個垂直方向的統計將被用來定位物體。我們提出的方法將會估計目標物離本車的距離來做更進一步的安全應用，最後我們測試了很多在日常生活倒車場景中可能會發生的情況，藉由實驗結果可證明我們提出的系統在障礙物偵測與距離估測上可達到一定的穩健性與準確性。

本論文的目標是希望藉由偵測阻礙行進路線上或突出地面具有一定高度的物體來提升倒車行為的安全，此外，藉由標示目標物離本車最近的位置與距離，可有效的避免在倒車上的碰撞。

Generic Obstacle Detection for Backing-Up Maneuver Safety Based on Inverse Perspective Mapping and Movement Compensation

Student: Jung-Chien Lee Advisor: Dr. Chin-Teng Lin

Institute of Biomedical Engineering

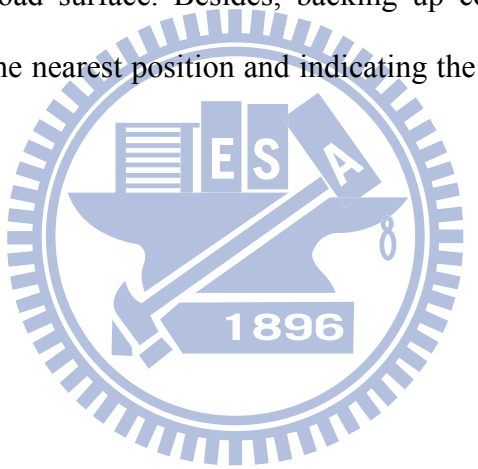
College of Computer Science National Chiao Tung University

ABSTRACT

The number of road traffic accidents which involves the fatality and injury accidents is so much in recent years, so that the number of higher performance vehicles for safety is growing up over the world. Among of all the collision accidents between vehicle and other generalized obstacles occur most frequently. For generic obstacle detection, researchers have proposed many methods which focus on stereo vision. However, automatic obstacle detection with a single camera system mounted on a moving vehicle is still a challenging problem in vision-based safety driving support systems. In this paper, we present a monocular vision-based obstacle detection algorithm which is based on movement compensation and IPM. We proposed a novel ground movement estimation technique which is employing road detection to assist in obtaining most useful ground features in the image, and analysis the principal distribution of optical flow of these feature points, the ground movement for compensation is obtained accurately. Besides, adopting different characteristics

between planar and non-planar object result from IPM to detect obstacle. In order to get more reliable detection and the nearest collision position, a vertical profile is used to locate the obstacle. Furthermore the proposed technique is to measure the distance of target how far away ego-vehicle for safety application. This system has been tested for many conditions which could occur when backing up in real life situations, and the experimental results already demonstrated the effectiveness and accurateness of the proposed obstacle detection and distance measurement technique.

The objective of this paper is to improve the safety for backing up maneuver by detecting objects that can obstruct the vehicle's driving path or anything raise out significantly from the road surface. Besides, backing up collision can be avoided effectively by marking the nearest position and indicating the distance away from our ego-vehicle.



致　　謝

本論文的完成，首先要感謝指導教授林進燈博士這兩年來的悉心指導，不管在學業或研究方法上都提供我許多意見與想法，這些寶貴的知識讓我受益良多。另外也要感謝口試委員們的建議與指教使本論文更爲完整。

其次，感謝超視覺實驗室的蒲鶴章博士與范剛維博士，還有建霆、子貴、肇廷、Linda、東霖、勝智學長姐們，以豐富的知識與過來人的經驗給予我莫大的幫助，同學佳芳、哲男、健豪、聖傑的相互砥礪共同奮鬥，這段日子有苦有笑，我會懷念我們一起爲實驗爲論文奮戰的時光，還有學弟們在研究過程中的幫忙，提供建議、拍攝影片、打氣鼓勵、製造歡樂，因爲有你們在 915 的日子很快樂。其中特別感謝子貴學長，在理論及程式技巧上給予我相當多的幫助與建議，讓我獲益良多。還有感謝超視覺實驗室給我的一切，這兩年來不管是知識的學習還是待人處事上讓我成長很多並且終身受用。

最後要感謝我最愛的父母，對我的教育與栽培，在我求學的路上總是不斷的給予我鼓勵與關懷，一點一滴，永藏於心，並給予我精神及物質上的一切支援，使我能安心地致力於學業。此外也感謝妹妹佳敏對我不斷的關心與支持。

謹以本論文獻給我的家人及所有關心我的師長朋友們。

Contents

List of Tables

List of Figures

Chapter 1.
Introduction

1.1 Background

The history of vehicle could be traced back to the 18th century. At that time, vehicle was a simple steam-powered three wheeler. As time goes on, there are many variations on vehicle improvement, such as the change of power, the change of style, and the change of function.

Nowadays, in addition to the consideration of vehicle style and brand, more and more consumers would take other factors into account. For example: the safety of vehicle, the entertainment equipments, the engine performance of vehicle and so on. Due to the demands of consumers, new generation vehicle which involves in the intelligent transportation system (ITS) has been developed rising and flourishing. The intelligent transportation system sub-items are listed in Table 1-1 [1].

Table 3-1: The sub-items of intelligent transportation system

Items	Some sub-items
Advanced Traffic Management Systems (ATMS)	Changeable Message Sign Weigh-In-Motion Electronic Toll Collection Automatic Vehicle Identification
Advanced Traveler Information Systems (ATIS)	Highway Advisory Radio Global Positioning System Wireless Communications
Advanced Vehicle Control and Safety Systems (AVCSS)	Collision Avoidance Systems Driver Assistance Systems Automatic Parking System
Commercial Vehicle Operations	Automatic Cargo Identification

(CVO)	Automatic Vehicle Location
Advanced Public Transportation Systems (APTS)	Automatic Vehicle Monitoring Electronic Fare Payment

The intelligent transportation system which we emphasize on advanced vehicle can be subdivided into two parts: (1) high performance and pollution-less power (2) intelligent control and advanced safety. For high performance and pollution-less power, many engineers have developed some new fuels or power such as bio-alcohol, bio-diesel fuel, hydrogen-based power, and hybrid power. For intelligent control and advanced safety, the objective of this part is to assist drivers when they are inattentive by using sensors and controllers. There are great deals of researches that have been done in this part. For instance, pre-collision warning system, intelligent airbags, electronic stability program (ESP), adaptive cruise control (ACC), lane departure warning (LDW) and so on.

More and more international vehicle companies or groups have invested in driving safety applications, for example, Toyota, Honda, and Nissan in Japan developed many products continuously. In Europe, the PREVENT project of i2010 will research and promote this project in order to decrease the number of casualty in the traffic accidents.

For driving safety, either vision-based (passive sensor) or radar-based (active sensor) sensing system are used in the intelligent vehicle in order to help driver deal with the situations on the road. The radar-based sensing system can detect the presence of obstacle and its distance, but the drawbacks of this system are low spatial resolution and slow scanning speed. Erroneous judgment will occur when the system is lying in the complicated environment or bad weather. Although vision-based sensing system will also suffer from the environment and weather restriction, it still provide more information and real scene to drivers to see what is happened when the

system is alarming.

The vision-based sensing system used in the intelligent vehicle can be subdivide into four parts by function: (1) obstacle warning system, (2) parking-assist system, (3) lane departure warning system ,(4) interior monitoring systems [2].Within this thesis, we concern about the first and the second part only.

1.2 Motivation

According to the statistics [3] in Table 1-2, the number of higher performance vehicles is growing up over the world in recent years, but the number of road traffic accidents number which involves the fatality and injury traffic accidents is still so much. By researching the factors of them, we can refer to a main reason: improper driving, such as inattentive driving, failing to slow down, drunk driver fatigue and so on.

To analyze the type of accidents, we find out a fact that collision accidents between vehicle and other generalized obstacles occur most frequently. For example, vehicle could collide with pedestrians, collide with other vehicle, or collide with static objects such as tree or pole. And we focus on improving backing-up maneuver safety. According to a recent NHTSA report [4] backing-up crashes are liable for at least 183 fatalities and more than 6700 injuries annually in the US. These numbers show how important it is to give drivers ways to avoid these crashes.

Indeed, many car manufacturers now propose rear obstacle detection systems on their vehicle to prevent backing-up collisions. These systems are mainly based on ultrasonic or radar sensors which have been found [4] to offer poor performance in detecting child pedestrians. Reference [4] reported that rear view cameras ‘have the greatest potential to provide drivers with reliable assistance in identifying people in

the path of the vehicle when backing'. Yet, these conclusions are balanced by concerns on camera performance and driver's ability to use efficiently the additional information provided. Indeed, to show most of the environment behind the vehicle, these cameras have a wide field of view introducing large distortion of images. Such images can prove difficult to interpret (especially for the evaluation of distance and recognition of objects) by untrained drivers as presented in [4] page 39: 'The mirror convexity also caused significant distortion of displayed objects, making them more difficult to recognize. A hurried driver making quick glances prior to initiating a backing maneuver might not allocate sufficient time to allow the driver to recognize an obstacle presented in the mirror.'

Furthermore, as these systems have not yet reached the technical capacities to be used as safety devices they are used more as comfort equipment. Thus drivers do not necessarily look at the screen during all of their backing-up maneuvers.

The objective of our research is to provide a rear obstacle detection system, that is a vision-based system capable of detecting obstacle in the rear view, especially in the path of the vehicle when backing. We hope that utilize this technology to improve the safety when backing the car.

Table 1-4: Case number of traffic accidents over the world from 1997 to 2008

Year	ROC	USA	JP	UK	FR	KOR
1997	2,428	3,347,614	958,925	324,204	169,600	343,159
1998	2,007	3,192,035	990,675	321,791	168,535	340,564
1999	1,636	3,236,238	1,050,397	316,887	167,600	402,967
2000	66,895	3,188,750	1,155,697	316,874	162,117	426,984
2001	80,612	3,032,672	1,180,955	309,859	153,945	386,539
2002	109,594	2,925,758	1,167,855	299,174	137,839	348,149
2003	156,303	2,888,601	1,181,431	287,099	115,900	376,503
2004	179,108	2,788,378	1,183,120	277,619	108,700	346,987

2005	203,087	2,699,000	1,156,633	267,816	108,000	342,233
2006	211,176	2,575,000	1,098,199	255,232	102,100	340,229
2007	216,927	2,491,000	1,034,445	244,834	103,200	335,906
2008	227,423	-	945,504	228,367	-	-

1.3 Objective

The collision accidents between vehicle and other generalized obstacles occur frequently during backing up period. Therefore, developing obstacle detection technique to prevent the crash accident is our main purpose. In our research, the obstacle which we want to detect is any object that can obstruct the vehicle's driving path or anything raising out significantly from the road surface. Therefore, the objective of our research is to develop a monocular vision-based obstacle detection algorithm on rear-view for improving backing-up maneuver safety. Adopting the advantages of vision-based obstacle detector, which have longer and wider detection range relative to radar, and that visual display can help driver to know position of obstacle and to indicate the distance away from ego-vehicle further more. Besides, we expect to utilize single camera to achieve obstacle detection not to adopt stereo vision, because low cost relative to stereo vision is considered. Then we intend to detect obstacle effectively for general environment and estimate the distance between target position and ego-vehicle accurately to alarm the driver. By proposing the obstacle detection for backing up, we want to prevent some backing-up crashes and to improve the safety.

1.4 Thesis Organization

The remainder of this thesis is organized as follows. Chapter 2 describes related

works about IPM and obstacle detection. Chapter 3 introduces system overview and some fundamental techniques such as IPM and optical flow. Chapter 4 shows obstacle detection algorithm including feature point extraction, ground movement estimation, obstacle localization, obstacle verification and distance measurement. Chapter 5 shows experimental results and evaluations. Finally, the conclusions of this system and future work will be presented in chapter 6.

Chapter 2.
Related Works

2.1 Related Works of Inverse Perspective Mapping (IPM)

The objective of inverse perspective mapping method is to remove the perspective effect caused by camera. This effect will cause the far scene to be condensed and always make following processing to be confused.

The main team who researched about the application of inverse perspective mapping topic are Alberto Broggi's team at the University of Parma in Italy. First, they proposed an inverse perspective mapping theory and establish the formulas [5]. Then they use these theories combine with some image processing algorithm, stereo camera vision system, and parallel processor for image checking and analysis (PA-PRICA) system which works in single Instruction multiple data (SIMD) computer architecture to form a complete general obstacle and lane detection system called GOLD system [6]. The GOLD system which was installed on the ARGO experimental vehicle is in order to achieve the goal of automatic driving.

There are other researchers using the inverse perspective mapping method [5] or similar mapping method combining with other image processing algorithm to detect lane or obstacles. For example, W. L. Ji [7] utilized inverse perspective mapping to get the 3D information such as the front vehicle height, distance, and lane curvature. Cerri et al. [8] utilized stabilized sub-pixel precision IPM image and time correlation to estimate the free driving space on highways. Muad et al. [9] used inverse perspective mapping method to implement lane tracking and discussed the factors which influent IPM. Tan et al. [10] combined the inverse perspective mapping and

optic flow to detect the vehicle on the lateral blind spot. Jiang et al. [11] proposed the fast inverse perspective mapping algorithm and used it to detect lanes and obstacles. Nieto et al. [12] proposed the method that stabilized inverse perspective mapping image by using vanish point estimation. Yang [13] adjusted the characteristic of inverse perspective mapping proposed by Broggi [5], which is the property a horizontal straight line in the image will be projected to an arc on the world surface. However, a horizontal straight line in the image will be projected to a straight line on the world surface, and a vertical straight line in the image will also be projected to a straight line and the prolongation will pass the camera vertical projection point on the world surface.

2.2 Related Works of Obstacle Detection

The obstacle detection is the primary task for intelligent vehicle on the road, since the obstacle on the road can be approximately discriminated from pedestrian, vehicle, and other general obstacles such as trees, street lights and so on. The general obstacle could be defined as objects that obstruct the path or anything located on the road surface with significant height.

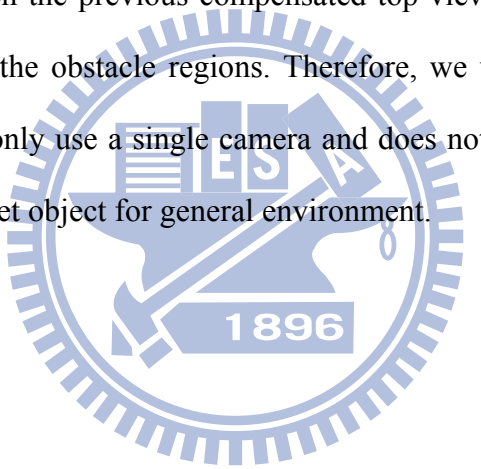
Depending on the number of sensors being used, there are two common approaches to obstacle detection by means of image processing: those that use a single camera for detection (monocular vision-based detection) and those that use two (or more) cameras for detection (stereo vision-based detection).

The stereo vision-based approach utilizes well known techniques for directly obtaining 3D depth information for objects seen by two or more video cameras from different viewpoints. Koller et al. [14] utilized disparities in correspondence to the

obstacles to detect obstacle and used Kalman filter to track obstacles. A method for pedestrian(obstacle) detection is presented in [15] whereby a system containing two stationary cameras. Obstacles are detected by eliminating the ground surface by transformation and matching the ground pixels in the images obtained from both cameras. The stereo vision-based approaches have the advantage of directly estimating the 3D coordinates of an image feature, this feature being anything from a point to a complex structure. The difference in viewpoint position causes a relative displacement, called disparity, of the corresponding features in the stereo images. The search for correspondences is a difficult, time-consuming task that is not free from the possibility of errors. Therefore, the performance of stereo methods depends on the accuracy of identification of correspondences in the two images. In other words, searching the homogeneous points pair in some area is the prime task of stereo methods.

The monocular vision-based approaches utilize techniques such as optical flow. For optical flow based methods which indirectly compute the velocity field and detect obstacle by analyzing the difference between the expected and real velocity fields, Kruger et al. [16] combined optical flow with odometry data to detect obstacles. However, optical flow based methods have drawback of high computational complexity and fail when the relative velocity between obstacles and detector are too small. Inverse perspective mapping (IPM), which is based on the assumption of moving on a flat road, has also been applied to obstacle detection in many literatures. Ma et al. [17][18][19] present an automatic pedestrian detection algorithm based on IPM for self-guided vehicles. To remove the perspective effect by applying the acquisition parameters (camera position, orientation, focal length) on the assumption of a flat road geometry, and predicts new frames assuming that all image points lie on the road and that the distorted zones of the image correspond to obstacles. Bertozzi et

al. [20] develop a temporal IPM approach, by means of inertial sensors to know the speed and yaw of the vision system and the assumption of flat road. A temporal stereo match technique has been developed to detect obstacles in moving situation. Although these methods could utilize the property of IPM to obtain effective results, but all of these should rely on external sensors such as odometer or inertial sensor to acquire ego-vehicle's displacement on the ground that enables them to compensate movement over time for the ground plane. Yang et al. [21] proposed a monocular vision-based approach by compensating for the ground movement between consecutive top-view images using the estimated ground-movement information and computes the difference image between the previous compensated top-view image and the current top-view image to find the obstacle regions. Therefore, we want to propose a pure vision-based algorithm only use a single camera and does not need additional sensor that could detect the target object for general environment.



Chapter 3
System Overview and Fundamental Techniques

3.1 System Overview

The overall system flowchart is shown in Fig. 3-1 and Fig. 3-2. At the beginning, we will utilize the road detection technique to support the feature point extraction. Due to the characteristic of color space, we first transform the RGB images to Lab color space. The road detection algorithm is processed on Lab color space. Then, we analysis the features which are suitable for track in our condition, the road boundary and the features which gradient is satisfied the restriction in the road region is selected, and the detailed contents will be described in Section 4.1.

While we are obtaining the feature points, the optical flow will be calculated among all of the feature points. Due to the perspective effect, the direction and length of the optical flows on the road in the original image are not the same. Therefore, we transform the information include the image coordinate and optical flow of feature points from the image coordinate to the world coordinate by inverse perspective mapping(IPM) [13] and build the bird's view image. When we obtain the information of feature points in world coordinate, the principal distribution of optical flow and the temporal coherence is used to estimate and verify the ground movement respectively. By transforming coordinate between image coordinate and world coordinate and compensating the ground movement, we can build a compensated image which is shifted by ground movement. The ground movement compensation procedure will be shown in Section 4.2 in detail.

As depicted in Fig. 3-2, the obstacle localization procedure is done by image difference. By thresholding the image, we can obtain the obstacle candidate image which the planar object in the image will be eliminated and the non-planar object which is the obstacle regions will be marked. For each obstacle region the closest position to our ego-vehicle is the interesting. The objective of obstacle localization is to look for the closest position and we can warn the driver by showing the target position. The detail of the obstacle localization will be described in Section 4.3. To ensure the detection results whether the objects are obstacles or not, we utilize the road detection result to validate the initial results. When the objects are detected by above flows, the information about the distance between target and our ego-vehicle is estimated by distance measurement procedure. The obstacle validation method and distance measurement method will be described in Section 4.4 and Section 4.5 respectively.

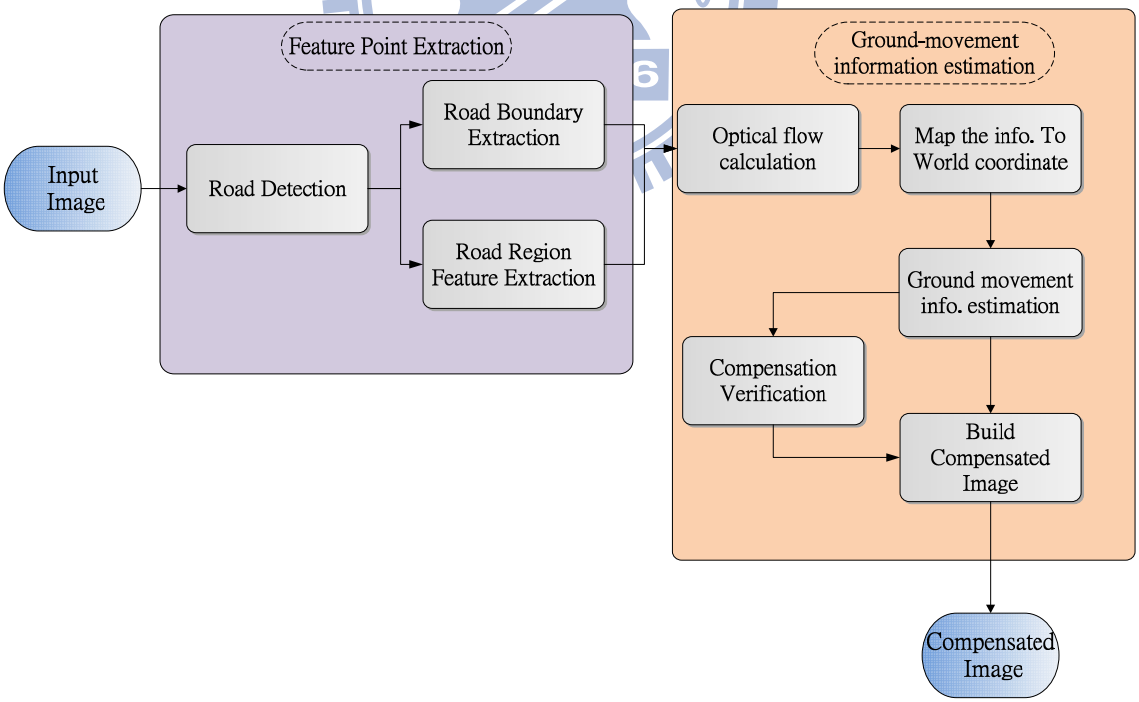


Fig. 3-1 System Flowchart

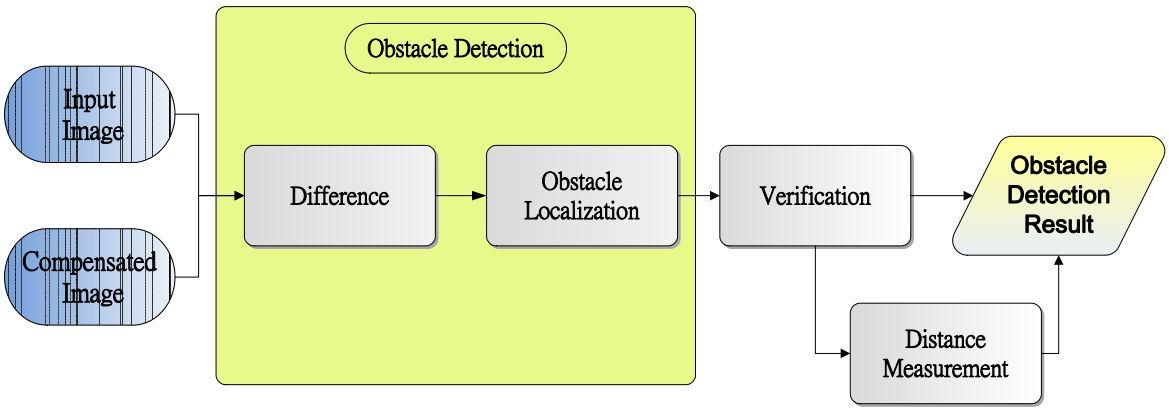


Fig. 3-2 System Flowchart

As mentioned above, ground movement information is estimated from optical flow in the world coordinate system and used to compensation for the difference between consecutive frame images in the world coordinate. Therefore, the inverse perspective mapping (IPM) and optical flow techniques adopted in our system will be introduced in the next section 3.2 and 3.3 respectively.

3.2 Inverse Perspective Mapping (IPM)

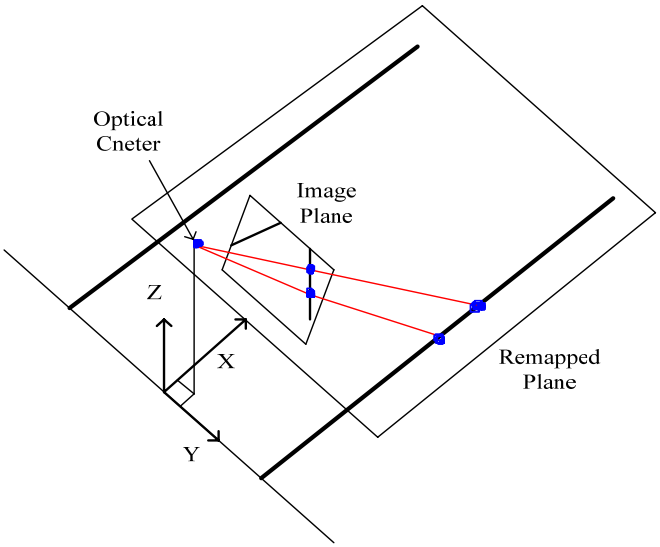
The perspective effect associates different meanings to different image pixels, depending on their position in the image. Conversely, after the removal of the perspective effect, each pixel represents the same portion of the information content, allowing a homogeneous distribution of the information among all image pixels. In other words, a pixel in the lower part of the image represents less information, while a pixel in the middle of the same image involves more information.

To cope with the effect that non-homogenous information content distribution among all pixels, an inverse perspective mapping transformation method will be introduced to remove perspective effect. To remove the perspective effect, it is

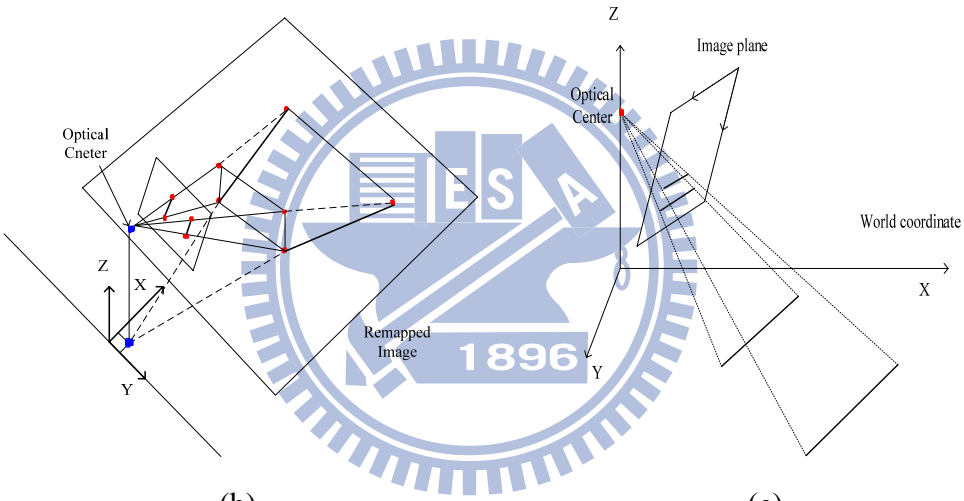
necessary to know the specific acquisition conditions (camera position, orientation, optics, etc.) and the scene represented in the image (the road, which is now assumed to be flat). This constitutes the a priori knowledge. The procedure aimed to remove the perspective effect by resample the incoming image, remapping each pixel on the original camera's image toward a different position and producing a new two-dimensional (2-D) array of pixels. The resulting image represents a top view of the scene, as it was observed from a significant height. Hence we will obtain a new image whose pixels indicate homogeneous information content.

In our research, the inverse perspective mapping [13], which will be able to remove the perspective effect by transforming the image coordinate to world coordinate and then process upon the world coordinate to estimate ground movement.

With some prior knowledge such as the flat road assumption and intrinsic and extrinsic parameters, we will be able to reconstruct a two-dimension image without perspective effect. The expectative results of diagrams are shown in Fig. 3-3. The transformation equation pair [13] with two expectative results: (1) a vertical straight line in the image will still be projected to a straight line whose prolongation will pass the camera vertical projection point on the world surface, (2) a horizontal straight line in the image will be projected to a straight line instead of an arc on the world surface. This result can be verified by similar triangle theorem.



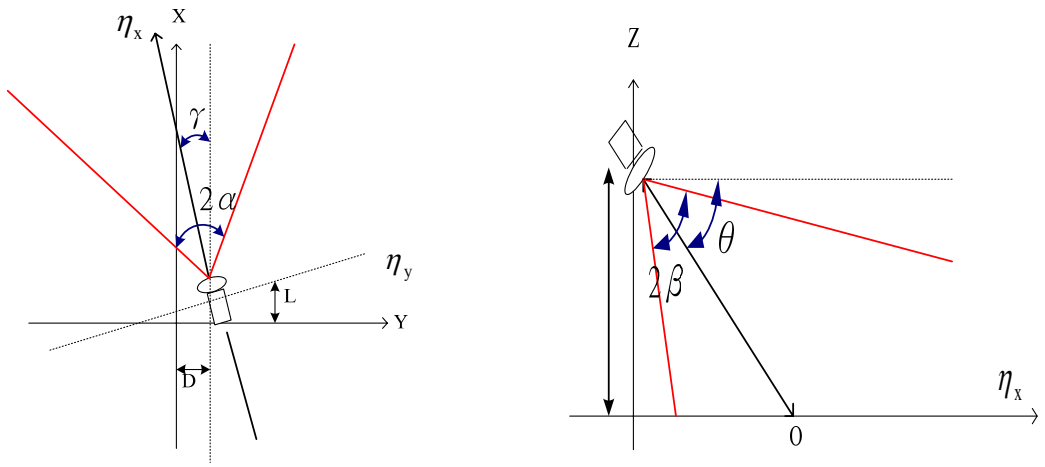
(a)



(b) (c)

Fig. 3-3: The expectative results of diagrams (a) perspective effect removing (b) the property of a vertical straight line (c) the property of a horizontal straight line

The spatial relationship between the world coordinate and image coordinate system is shown in Fig. 3-4, and the illustrations of deriving process shown in Fig. 3-5



(a) (b)

Fig. 3-4: (a) Bird's view and (b) side view of the geometric relation between world coordinate and image coordinate system.

From Fig. 3-5, the following equations will be derived

$$\rightarrow \theta_1 = \mathrm{u}\frac{2\beta}{\mathrm{m}-1} - \beta$$

$$\rightarrow \theta_2 = \mathrm{v}\frac{2\alpha}{\mathrm{n}-1} - \alpha \qquad (3\text{-}1)$$

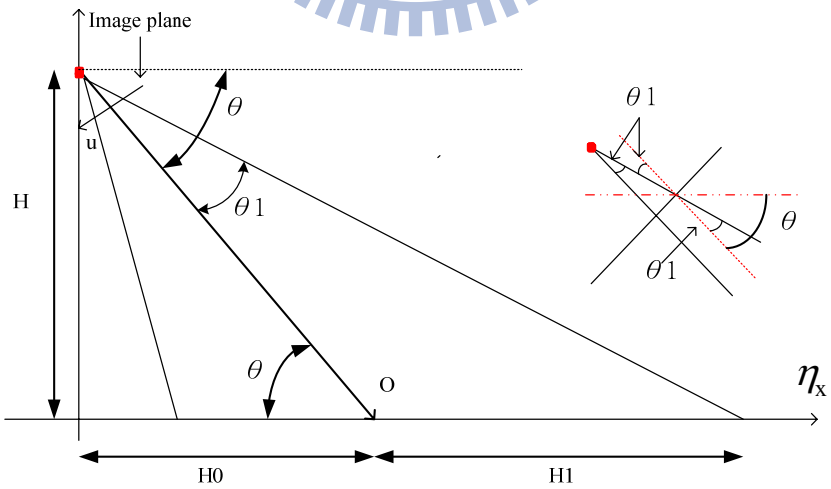


Fig. 3-5: Geometric relation of image coordinate system and world coordinate system

The forward transformation equations will be derived, and the backward transformation equations are easily obtained by some mathematical computation. The

forward transformation and backward transformation equations are shown below:

Forward transformation equations:

$$\Rightarrow X = H*\cot(\theta+\theta_1)*\cos(\gamma)+H*\csc(\theta+\theta_1)*\tan(\theta_2)*\sin(\gamma)$$

$$\Rightarrow Y = -H*\cot(\theta+\theta_1)*\sin(\gamma)+H*\csc(\theta+\theta_1)*\tan(\theta_2)*\cos(\gamma)$$

(3-2)

Backward transformation equations:

$$\Rightarrow u = \frac{m-1}{2\beta}*(\cot^{-1}(\frac{X*\cos(\gamma)-Y*\sin(\gamma)}{H})-\theta+\beta)$$

$$\Rightarrow v = \frac{n-1}{2\alpha}*(\tan^{-1}(\frac{X*\sin(\gamma)+Y*\cos(\gamma)}{H*\csc(\theta+\theta_1)})+\alpha)$$

(3-3)

The notations in the above equations and figures are introduced as follows:

(u,v) : The image coordinate system.

(X,Y,Z) : The world coordinate system where (X,Y,0) represent the road surface.

(L,D,H) : The coordinate of camera in the world coordinate system.

θ: The camera's tilt angle.

γ: The camera's pan angle.

α,β: The horizontal and vertical aperture angle.

m,n : The dimension of image (m by n image).

O: The optic axis vector.

η_x,η_y: The vector which represents the optic axis vector O projected on the road surface and its perpendicular vector.

To implement the inverse perspective mapping, we use the equations (3-3) instead of equations (3-2) and scan row by row on the remapped image to compute

the mapping points on the original image while we do not want an image full of hollows. However, the objective of inverse perspective mapping in our research is not to build the bird's view image, but aiming to use the transformation to remove the perspective effect and to transform coordinate between the original image coordinate and the world coordinate system. By utilizing inverse perspective mapping, the ground movement estimation procedure will process in the world coordinate system.

3.3 Optical Flow

When we are dealing with a video source, as opposed to individual still images, we may often want to assess motion between two frames (or a sequence of frames) without any other prior knowledge about the content of those frames. The optical flow itself is some displacement that represents the distance a pixel has moved between the previous frame and the current frame. Such a construction is usually referred to as a dense optical flow, which associates a velocity with every pixel in an image. The Horn-Schunck method [22] attempts to compute just such a velocity field. One seemingly straightforward method simply attempting to match windows around each pixel from one frame to the next; this is known as block matching. Both of these methods are often used in the dense tracking techniques.

In practice, calculating dense optical flow is not easy. Consider the motion of a white sheet of paper. Many of the white pixels in the previous frame will simply remain white in the next. Only the edges may change, and even then only those perpendicular to the direction of motion. The result is that dense methods must have some method of interpolating between points that are more easily tracked so as to solve for those points that are more ambiguous. These difficulties manifest themselves

most clearly in the high computational costs of dense optical flow.

This leads us to the alternative option, sparse optical flow. Algorithms of this nature rely on some means of specifying beforehand the subset of points that are to be tracked. If these points have certain desirable properties, such as the "corners", then the tracking will be relatively robust and reliable. The computational cost of sparse tracking is so much less than dense tracking that many practical applications are often adopting. Therefore, we consider the most popular sparse optical flow technique, Lucas-Kanade (LK) optical flow [23][24], this method also has an implementation that works with image pyramids, allowing us to track faster motions.

The Lucas-Kanade (LK) algorithm, was applied to a subset of the points in the input image, it has become an important sparse technique. The LK algorithm can be applied in a sparse context because it relies only on local information that is derived from some small window surrounding each of the points of interest. The disadvantage of using small local windows in Lucas-Kanade is that large motions can move points outside of the local window and thus become impossible for the algorithm to find. This problem led to development of the "pyramidal" LK algorithm, which tracks starting from highest level of an image pyramid (lowest detail) and working down to lower levels (finer detail). Tracking over image pyramids allows large motions to be caught by local windows.

The basic idea of the LK algorithm rests on three assumptions:

1. *Brightness constancy*. A pixel from the image of an object in the scene does not change in appearance as it (possibly) moves from frame to frame. For grayscale images, this means we assume that the brightness of a pixel does not change as it is tracked from frame to frame.
2. *Temporal persistence*. The image motion of a surface patch changes slowly in time. In practice, this means the temporal increments are fast enough relative to the scale

of motion in the image that the object does not move much from frame to frame.

3. *Spatial coherence*. Neighboring points in a scene belong to the same surface, have similar motion, and project to nearby points on the image plane.

By using these assumptions, the following equation can be yielded, where *y* component of velocity is *v* and the *x* component of velocity is *u*:

$$I_x u + I_y v + I_t = 0 \tag{3-4}$$

For this single equation there are two unknowns for any given pixel. This means that measurements at the single-pixel level cannot be used to obtain a unique solution for the two-dimensional motion at that point. Instead, we can only solve for the motion component that is perpendicular to the line described by the flow equation. Fig. 3-6 presents the mathematical and geometric details.

$$I_x u + I_y v + I_t = 0 \Rightarrow \nabla I^T u = -I_t$$
$$u = \begin{bmatrix} u \\ v \end{bmatrix} \quad \nabla I = \begin{bmatrix} I_x \\ I_y \end{bmatrix} \tag{3-5}$$

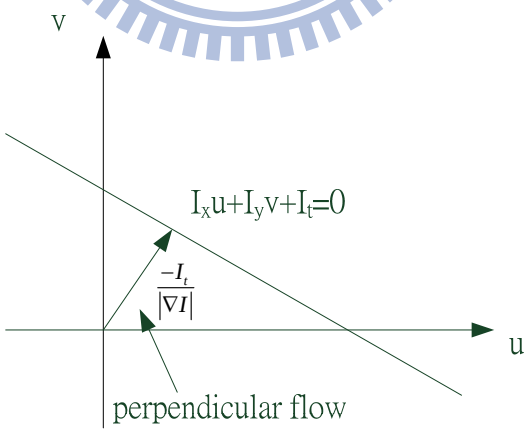


Fig. 3-6: Two-dimensional optical flow at a single pixel: optical flow at one pixel is underdetermined and so can yield at most motion, which is perpendicular to the line described by the flow equation

Normal optical flow results from the aperture problem, which arises when you

have a small aperture or window in which to measure motion. When motion is detected with a small aperture, you often see only an edge, not a corner. But an edge alone is insufficient to determine exactly how (i.e., in what direction) the entire object is moving. To resolve the aperture problem, we turn to the last optical flow assumption for help. If a local patch of pixels moves coherently, then we can easily solve for the motion of the central pixel by using the surrounding pixels to set up a system of equations. For example, if we use a 5-by-5 window of brightness values, round the current pixel to compute its motion, we can then set up 25 equation as follows:

$$\underbrace{\begin{bmatrix} I_x(p_1) & I_y(p_1) \\ I_x(p_2) & I_y(p_2) \\ \vdots & \vdots \\ I_x(p_{25}) & I_y(p_{25}) \end{bmatrix}}_{A} \underbrace{\begin{bmatrix} u \\ v \end{bmatrix}}_{d} = \underbrace{\begin{bmatrix} I_t(p_1) \\ I_t(p_2) \\ \vdots \\ I_t(p_{25}) \end{bmatrix}}_{b} \tag{3-6}$$

To solve for this system, we set up a least-squares minimization of the equation, whereby $min\|Ad-b\|^2$ is solved in standard form as: $(A^TA)d = A^Tb$. From this relation we obtain our *u* and *v* motion components. Writing this out in more detail and the solution to this equation is yielded:

$$\underbrace{\begin{bmatrix} \sum I_x I_x & \sum I_x I_y \\ \sum I_x I_y & \sum I_y I_y \end{bmatrix}}_{A^TA} \begin{bmatrix} u \\ v \end{bmatrix} = -\underbrace{\begin{bmatrix} \sum I_x I_t \\ \sum I_y I_t \end{bmatrix}}_{A^Tb} \Rightarrow \begin{bmatrix} u \\ v \end{bmatrix} = (A^TA)^{-1}A^Tb \tag{3-7}$$

Chapter 4
Obstacle Detection Algorithm

4.1 Feature Point Extraction

4.1.1 Feature Analysis

There are many kinds of local features that one can track. Features which are used to estimate ground movement based on its motion, that is to find these features from one frame in a subsequent frame of the video stream. Obviously, if we pick a point on a large blank wall then it won't be easy to find that same point in the next frame of a video. If all points on the wall are identical or even very similar, then we won't have much luck tracking that point in subsequent frames. On the other hand, if we choose a point that is unique then we have a pretty good chance of finding that point again. In practice, the point or feature we select should be unique, or nearly unique, and should be parameterizable in such a way that it can be compared to other points in another image. Therefore, we might be tempted to look for points that have some significant change within neighboring local area that is the good features which have a strong derivative in spatial domain. Another characteristic of features is about the position of the image. Due to the objective of the following procedure is to estimate the ground movement information, features lie on the ground region is useful for the following ground movement estimation algorithm. According to above analysis, a good feature to track should have two characteristics. First, a feature should have strong derivative which could assist us to track them and obtain a precise motion. Then, the position of feature should be restricted on the road region (non-obstacle region). The features which we will use them to estimate the ground

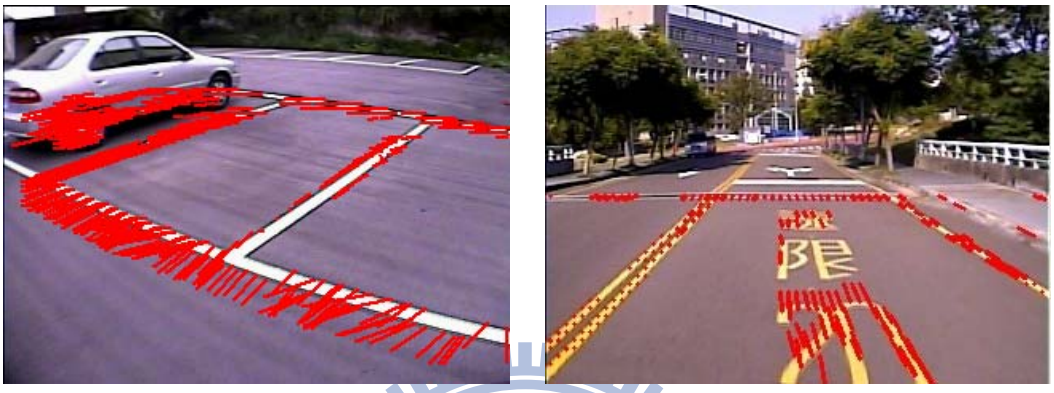
movement information should conform the above two characteristic, these features will be suitable for estimating ground movement information.

To consider the first characteristic – strong derivative, a point to which a strong derivative is associated may be on an edge of some kind. Then considering this property of edge we employ the Sobel operator to find out the edge of image. The points which be extracted by edge detection are used to be feature points, which then calculate optical flow for all of these feature points. These edge points and its optical flow of image is shown in Fig. 4-1. But a problem is arising as depicted in Fig 4-1(b), it could look like all of the other points along the same edge. An ambiguous optical flow will happen when the edge points parallel to the direction of motion. It turns out that strong derivative of a single direction is not enough. However, if strong derivatives are observed in two orthogonal directions then we can hope that this point is more likely to be unique. For this reason, many trackable features in the image that are corners. Intuitively, corners are the points that contain enough information to be picked out from one frame to the next. We examined by the most commonly used definition of a corner was provided by Harris[25]. This definition relies on the matrix of the second-order derivatives ($\partial^2 x$, $\partial^2 y$, $\partial x \partial y$) of the image intensities. We can think of the second-order derivatives of images, taken at all points in the image, as forming new second-derivative images or, when combined together, a new *Hessian* image. This terminology comes from the Hessian matrix around a point, which is defined in two dimensions by:

$$H(p) = \begin{bmatrix} \frac{\partial^2 I}{\partial x^2} & \frac{\partial^2 \mathrm{I}}{\partial x \partial y} \\ \frac{\partial^2 I}{\partial y \partial x} & \frac{\partial^2 I}{\partial y^2} \end{bmatrix} \tag{4-1}$$

By using Harris corner definition, the result of corner detection and its corresponding optical flow is shown Fig. 4-2. It is obvious to observe that motion of corner is

relative accurate than edge point. Although corner have more precise optical flow, another problem is arising that is position of corner almost lie on obstacle region such as vehicle component. The Fig. 4-2(b) is shown the result of corner detection in a common driving condition, we can see that the feature points are nearly located on obstacle regions. These feature points unable let us to estimate the ground movement.



(a) (b)

Fig. 4-1 Results of edge detection and its corresponding optical flow

(a) (b)

Fig. 4-2 Results of corner detection and its corresponding optical flow

By considering above analysis of features, we proposed a feature point extraction method employ road detection procedure to assist in getting ground features. The flowchart of proposed feature point extraction is shown in Fig. 4-3. The objective is to distinguish the major road region and non-road region, and utilize the result of the road detection, that is to extract the boundary of major road and some good features within road region. By integrating road detection, the more useful ground features

could be extracted and could improve results of ground movement effectively. The next chapter 4.1.2 will introduce the detail of road detection and describe what feature point will be selected.

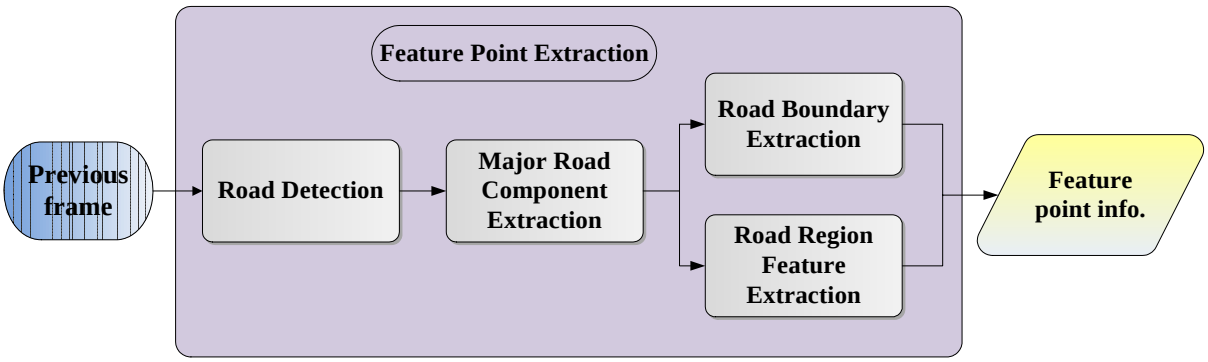


Fig. 4-3 Flowchart of feature point extraction

4.1.2 Road Detection

The proposed feature point extraction technique is integrating a road detection procedure [26] which is used an on-line color model that we can train an adaptive color model to fit road color. The main objective of road detection is to discriminate the road and non-road region roughly, because the result is used to support feature extraction not used to extract obstacle regions. However, we adopt an on-line learning model that allows continuously update during driving, through the training method that can enhance plasticity and ensure the feature is on the road region.

Due to the color appearance in the driving environment, we have to select the color features and using these color features to build a color model of the road. Therefore, we have to choose a color space which has uniform, little correlation, concentrated properties in order to increase the accuracy of the model. In computer color vision, all visible colors are represented by vectors in a three-dimensional color

space. Among all the common color spaces, RGB color space is the most common color feature selected because it is the initial format of the captured image without any distortion. However, the RGB color feature is high correlative, and the similar colors spread extensively in the color space. As a result, it is difficult to evaluate the similarity of two color from their 1-norm or Euclidean distance in the color space.

The other standard color space HSV is supposed to be closer to the way of human color perception. Both HSV and L*a*b* resist to the interference of illumination variation such as the shadow when modeling the road area. However, the performance of HSV model is not as good as L*a*b* model because the road color cause the HSV model not uniform that lead to the HSV color model not as uniform as the L*a*b* color model. There are many reasons attribute this result. Firstly, HSV is very sensitive and unstable when lightness is low. Furthermore, the Hue is computed by dividing (I_{max} - I_{min}) in which I_{max} = max(R,G,B), I_{min} = min(R,G,B), therefore when a pixel has a similar value of Red, Green and Blue components, the Hue of the pixel may be undetermined. Unfortunately, most of the road surface is in similar gray colors with very close R, G, and B values. If using HSV color space to build road color model, the sensitive variation and fluctuation of Hue will generate inconsistent road colors and decrease the accuracy and effectiveness. L*a*b* color space is based on data-driven human perception research that assumes the human visual system owing to its uniform, little correlation, concentrate characteristics are ideally developed for processing natural scenes and is popular for color-processed rendering. L*a*b* color space also possesses these characteristics to satisfy our requirement. It maps similar colors to the reference color with about the same differences by Euclidean distances measure and demonstrates more concentrated color distribution than others. Then considering the advantaged properties of L*a*b* for general road environment, the L*a*b* color space for road detection is adopted.

The RGB-L*a*b* conversion is described as follow equations:

1. RGB-XYZ conversion:

$$X = 0.431 \cdot R + 0.342 \cdot G + 0.178 \cdot B$$

$$Y = 0.222 \cdot \mathrm{R} + 0.707 \cdot \mathrm{G} + 0.071 \cdot \mathrm{B}$$

$$Z = 0.020 \cdot R + 0.130 \cdot G + 0.939 \cdot B$$

2. Cube-root transformation:

$$\begin{cases} L^* = 166 \cdot \left(\frac{Y}{Y_n}\right)^{\frac{1}{3}} - 16 & if \ \frac{Y}{Y_n} > 0.008856 \\ L^* = 903.3 \cdot \left(\frac{Y}{Y_n}\right)^{\frac{1}{3}} & if \ \frac{Y}{Y_n} \le 0.008856 \end{cases}$$

$$a^* = 500 \cdot [f(\frac{X}{X_n}) - f(\frac{Y}{Y_n})]$$

$$b^* = 200 \cdot [f(\frac{Y}{Y_n}) - f(\frac{Z}{Z_n})]$$

where X_n, Y_n, Z_n, *are XYZ tristimulus values of reference white point*

$X_n = 95.05, Y_n = 100, Z_n = 108.88$

$$f(x) = \begin{cases} t^{\frac{1}{3}} & \frac{Y}{Y_n} > 0.008856 \\ 7.787t + 16/116 & \frac{Y}{Y_n} \le 0.008856 \end{cases}$$

By modeling and updating of the L*a*b* color model, the built road color model can be used to extract the road region. The L*a*b* model is constituted of K color balls, and each color ball m_i is formed by a center on $(L_{m_i}, {}^*a_{m_i}, {}^*b_{m_i})$ and a fixed radius $\lambda_{\max} = 5$ as seen in Fig. 4-4. In order to train a color model, we set a fixed area of the lower part of the image and assume pixels in the area are the road samples. For each of these pixels in the beginning 30 frames are used to initialize the color model, and updating the model every ten frames to increase processing speed but still maintain high accurate performance.

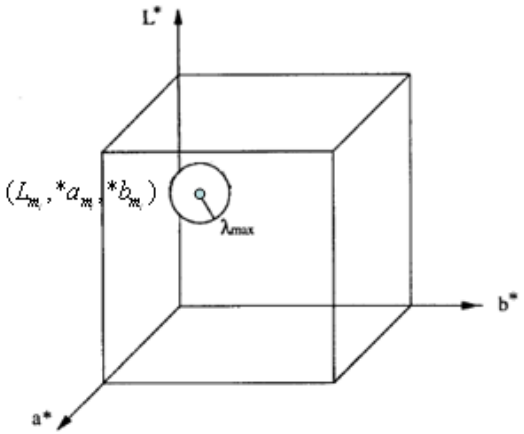


Fig. 4-4 A color ball i in the L*a*b* color model whose center is at (L_m, $*a_m$, $*b_m$) and with radius λ_{max}

The sampling area is used to be modeled by a group of K weighted color balls. We denote the weight and the counter of the $m_{i\,th}$ color ball at a time instant t by $W_{m_i,t}$ and $Counter_{m_i,t}$, and the weight of each color ball represents the stability of the color. The color ball which more on-line samples belonged to over time accumulated a bigger weight value shown in Fig. 4-5. Adopting the weight module increases robustness of the model.

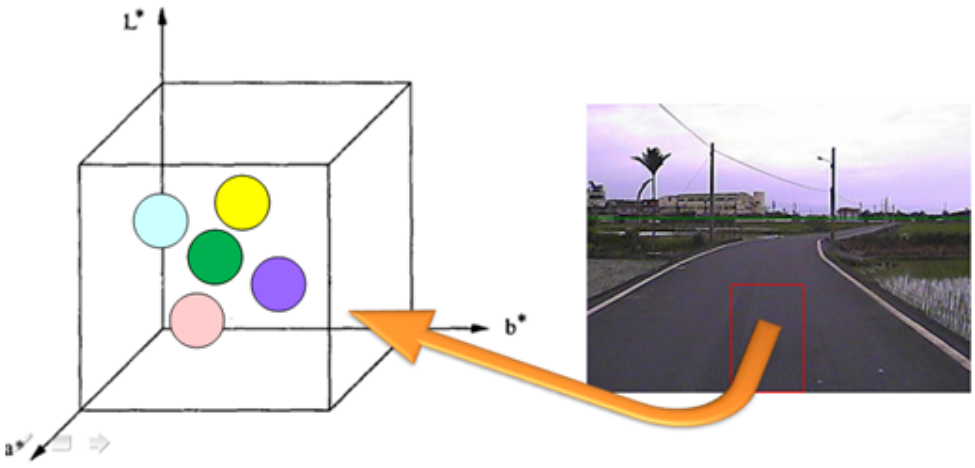


Fig. 4-5 Sampling area and color ball with a weight which represents the similarity to current road color.

The weight of each color ball is updated by its counter when the new sample is coming which is called one iteration. Therefore the counter would be initialized to zero at the beginning of iteration. The counter of each color ball records the number of pixels added from the on-line samples in the iteration. The first thing to do is that which color ball is chosen to be added. We measure the similarity between new pixel x_t and the existing K color balls using a Euclidean distance measure (4-1). The maximum value of K is 50 which represents each on-lined model contains 50 color balls at most.

$$Similarity(x, m_i) = sqrt\left[(L_{m_i} - L_x)^2 + (a_{m_i} - a_x)^2 + (b_{m_i} - b_x)^2\right] <= \lambda_{\max} \quad (4\text{-}1)$$

If a new pixel x_t was covered by any of the color ball in the model, one will be added to the counter of best matching color at this iteration as the equation (4-2). After entire new sample pixels at this iteration undertake the matching procedures mentioned above, the weights of every color ball are updated according to their current counter and their weight at last iteration. The updating method is as follows:

$$m(x_i) = arg_{m_i} min(Similarity(x_i, m_i) <= \lambda_{max}) \quad (4\text{-}2)$$

$$W_{m_i,t+1} = \alpha_w \cdot W_t + (1 - \alpha_w) \cdot counter_{m_i} / \left|N_{sample}\right|$$
$$\alpha_w \in [0,1], \quad \left|N_{sample}\right|$$

, where α_w is user-defined learning rate, N_{sample} is the sampling area

Then using the weight to decide which color ball of the model most adapt and resemble current road. The color balls are sorted in a decreasing order according to their weights. As a result, the most probable road color features are at the top of the list. The first B color balls are selected to be enabled as standard color for road detection, and these color balls with a higher weight has more importance in detection

step. Road detection is achieved via comparison of the new pixel x_t with the existing B standard color balls selected at the previous instant of time shown in Fig. 4-6. If no match is found, the pixel x_t is considered as non-road. On the contrary, the pixel x_t is detected as road.

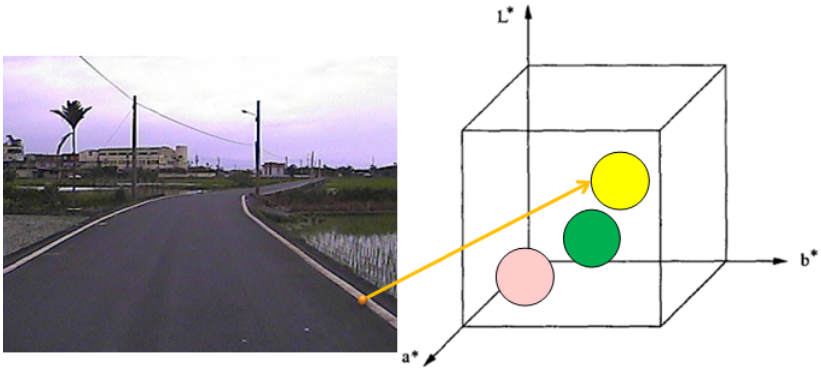


Fig. 4-6 Pixel matched with first B weight color balls which are the most represent standard color.

Fig. 4-7 shows some results of on-line L*a*b* road detection, the green areas are determined as road else are the non-road regions. However, we will not undertake road detection to all of the image because of the following procedure is processed in the world coordinate, that is the range of road detection is restricted to the horizon which is caused by the geometrical characteristic of IPM, and could be the different position in the image due to camera set up environment (camera height, tilt angle and so on).

Fig. 4-7 Results of road detection

4.1.3 Feature Point Extraction

The objective of road detection is to distinguish the major road and non-road region, and the result will be used to extract feature point. As above mentioned, we consider the two characteristics that are strong derivative and ground feature, the road boundary and strong gradient points are selected to be feature points.

Therefore, the first step of feature point extraction is to extract the major road region. When result of road detection is obtained, the dilation and erosion procedure is used to merge the neighboring region and to reduce the fragmentation then the connected component is executed to separate the road regions to several components. After that we will find out the maximum component of all components and assumed that to be the major road region. Then the boundary of major road region is extracted to be feature points, because the border of road and non-road should be the road feature and have strong derivative. Besides, we analyze the gradient distribution of major road area, and the more strong gradient points will be extracted to be feature points because of their strong derivative and position. Then the feature points are collected completely by these road boundary and high gradient features. The Fig. 4-8 is shown some results of feature point extraction. By employing road detection to support feature point extraction, the more useful ground features can be extracted.

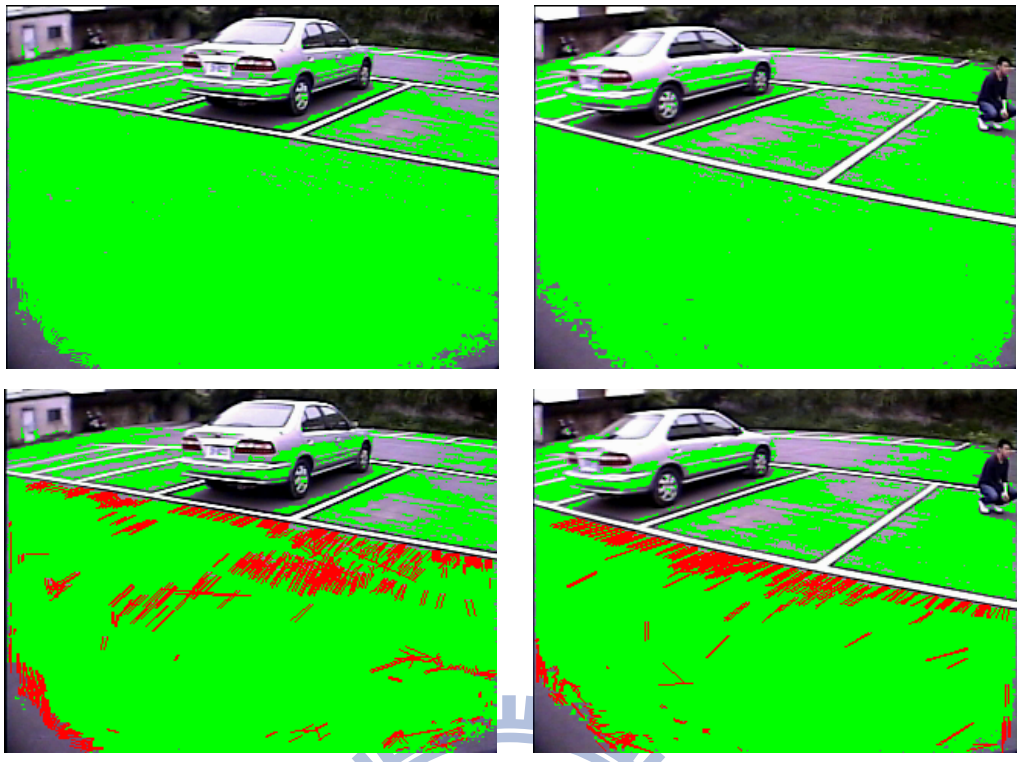


Fig. 4-8 Results of feature point extraction. The upper image is result of road detection, and lower image is position of feature points.

4.2 Ground Movement Estimation

In this section we will introduce the proposed ground movement estimation procedure. Ground movement information is estimated from optical flow in the world coordinate system. By analyzing the principal distribution of optical flow can let us get the most representative ground movement, which will be used to compensate for the previous frame and difference with current frame. In addition, ground movement will be verified via temporal coherence. The flow of ground movement estimation is illustrated in Fig. 4-9.

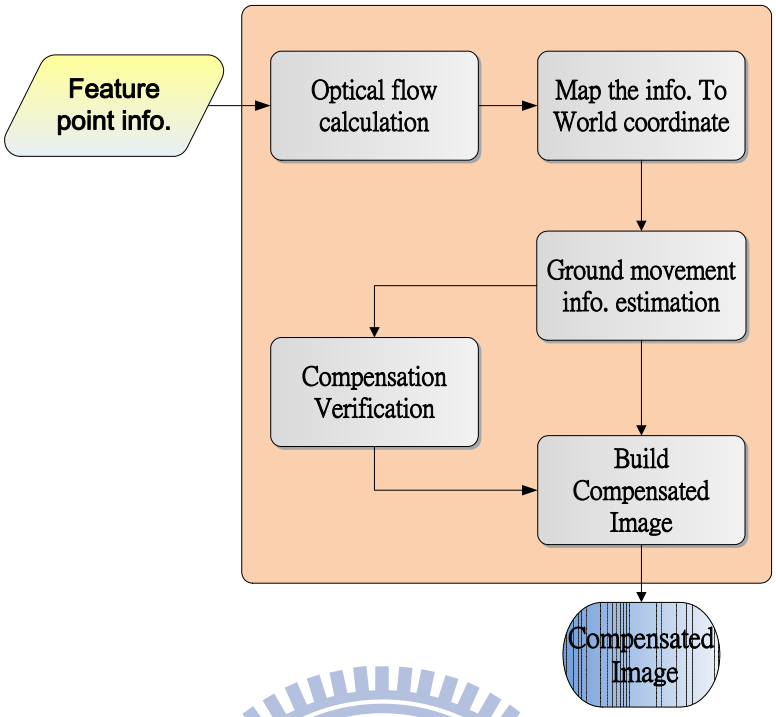


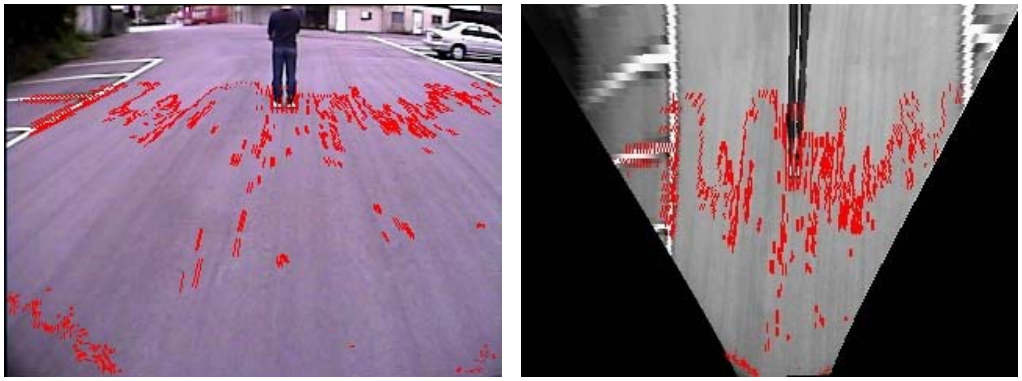
Fig. 4-9 flow of ground movement estimation

4.2.1 Ground Motion Estimation

By feature point extraction procedure as described in section 4.1, the useful features for ground movement estimation is obtained. Then these feature points will be used to estimate ground motion. Therefore, when the feature points are acquired the main tasks of ground movement estimation procedure are high accuracy optical computation and ground movement information estimation. The first step is to calculate the optical flow for all of these feature points. The pyramidal Lucas and Kanade algorithm introduced in section 3.3, which copes efficiently with large movements, is used to calculate the optical flow for these feature points in the original image. As a result, Fig. 4-10(a) and Fig. 4-11(a) shows the feature points and its corresponding optical flow in the original image. Due to the perspective effect, the directions and lengths of the optical flows on the road in the original image are not the same when vehicle is moving straight shown in Fig. 4-10(a). The case in which the

vehicle is turning is shown in Fig. 4-11. In this case the complicated optical flow distribution appeared in the original image. However, the inconsistent optical flow of road in the original image let us encounter a difficulty to estimate the ground movement. Therefore, we will take advantage of the IPM to remove the perspective effect. The optical flow information of an original image is mapped into world coordinate. The objective of inverse perspective mapping (IPM) is to remove the perspective effect by transforming the image coordinate to world coordinate, and scale the world coordinate that can obtain a bird's view image. Therefore, the world coordinate information is same as bird's view image that both of them are perspective removal. In our research, the IPM is used to remove the perspective effect and transform the image coordinate information to world coordinate and the ground movement procedure is processed in the world coordinate the bird's view image is used to display and examine some results, which we will not process on it.

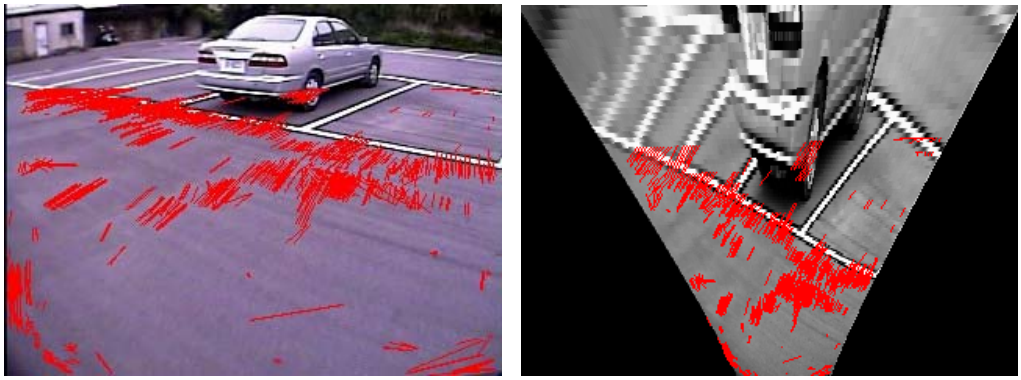
The difference between the optical flow of the original image and that of the bird's view image is shown in Fig. 4-10 and Fig. 4-11. When a vehicle is moving straight, the optical flows in the bird's view image have the same direction and length independent of the locations of the optical flows. Similarly when vehicle is turning, the optical flow distribution in the bird's view image draws concentric circle following the movement of the ground but roughly have a similar magnitude.



(a) optical flows of original image (b) optical flows of bird's view image

Fig. 4-10 Difference between optical flow of original image and those of bird's view

image when vehicle is moving straight.



(a) optical flows of original image (b) optical flows of bird's view image

Fig. 4-11 Difference between optical flow of original image and those of bird's view image when vehicle is turning.

The kernel concept of ground compensation based detection algorithm is adopting the following characteristics. The optical flow distribution of ground region is approximately consistent in the bird's view image. On the contrary, a vertical straight line in the image which represented the vertical edge of obstacle in the world coordinate system is projected to a straight line whose prolongation will pass the camera vertical projection point on the world surface. Therefore, the optical flow distribution of obstacle regions are different drastic to the ground region. Then we can estimate the ground movement and used build a compensated image which assume the image is all planar object (ground). Therefore, the planar region will be eliminated but the obstacle regions will not. Fig. 4-12 shows the difference between optical flow of obstacle and those of planar object.

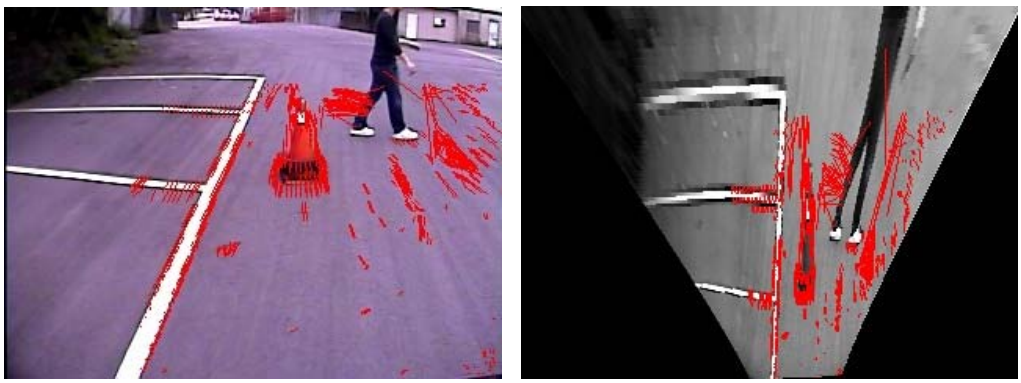


Fig. 4-12 Differences between optical flows of obstacle and those of planar object.

Thanks to the mapping between original image and world coordinate system, ground movement information can be estimated based on the optical flow of feature points in the world coordinate system. The feature points which we obtained via integrating road detection are with a characteristic that most of these features will lie on ground region. Besides, the specific optical flow distribution of ground region in the world coordinate system, that is optical flow of them will approximately consistent. Due to the above characteristics, we would like to find out the principal distribution of optical flow that can let us get the most representative ground motion. We analyzing the principal distribution of optical flow by calculating the histogram of optical flow according to its direction and magnitude, and the peak of the histogram is considered to be ground motion. As shown in Fig. 4-13, that is the histogram distribution of corresponding optical flow. By utilizing the principal motion, that can let us avoid some errors such as ambiguous optical flow or the non-peak value of optical flow is possibly causing by obstacle feature point.

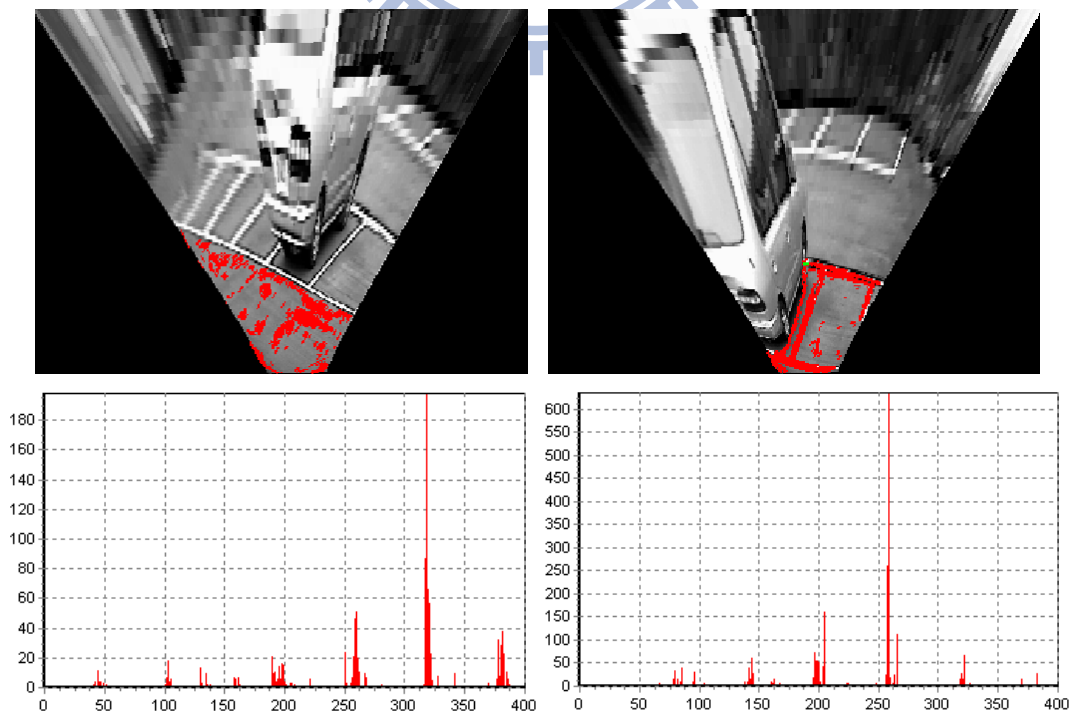


Fig. 4-13 Histogram distribution of optical flow in world coordinate system

4.2.2 Compensated Image Building

The specific optical flow distribution in the world coordinate system shows that the movement of ground in the world coordinate system can be described as a translation or rotation of a two-dimensional coordinate plane. The ground movement information is estimated by using optical flow in world coordinate system, and the optical flow is ground motion which is estimated by analyzing the principal distribution of optical flow of feature points, that is utilized the procedure as described in last section 4.2.1. Then we can acquire ground motion in the world coordinate system which is used to estimate ground movement information.

In the world coordinate system converted from consecutive images captured at difference location O1 and O2 as depicted in Fig. 4-14, if we represent the world coordinates of a ground feature point P as $(x_1, y_1)^T$ and $(x_2, y_2)^T$ before and after vehicle movement respectively, the optical flow of the feature point P can be written as Eq. (4-3).

$$\begin{pmatrix} f_x \\ f_y \end{pmatrix} = \begin{pmatrix} x_2 \\ y_2 \end{pmatrix} - \begin{pmatrix} x_1 \\ y_1 \end{pmatrix} \tag{4-3}$$

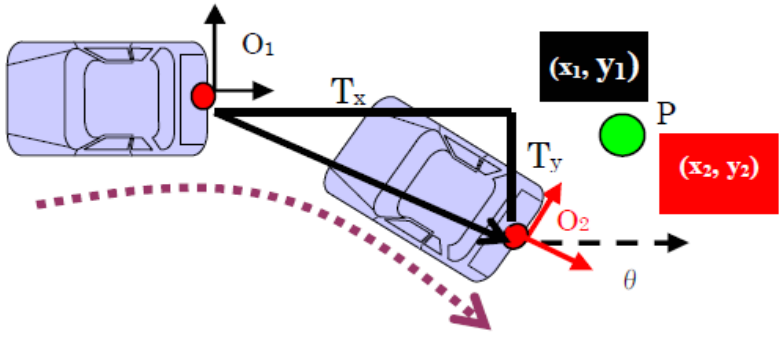


Fig. 4-14 Two-dimensional coordinate plane

If we represent the ground movement information between consecutive world

coordinate system by a two-dimensional coordinate plane as shown in Fig. 4-14, the relationship of corresponding points between consecutive world coordinate system can be described with the Eq. (4-4), where Θ is the rotation component of vehicle movement, and $(T_x, T_y)^T$ are the translation components of ground movement. Then ground movement information including rotation and translation component. That is we can utilize Eq. (4-4) to find out the compensated coordinate of point $(x_1,y_1)^T$ but Θ and $(T_x, T_y)^T$ should be calculated in advance. By substituting Eq. (4-3) into Eq. (4-4), then the relationship between Θ and $(T_x, T_y)^T$ can be derived as Eq. (4-5).

$$\begin{pmatrix} x_2 \\ y_2 \end{pmatrix} = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} x_1 \\ y_1 \end{pmatrix} + \begin{pmatrix} T_x \\ T_y \end{pmatrix} \tag{4-4}$$

$$\theta \begin{pmatrix} y_1 \\ -x_1 \end{pmatrix} - \begin{pmatrix} T_x \\ T_y \end{pmatrix} + \begin{pmatrix} f_x \\ f_y \end{pmatrix} = 0 \tag{4-5}$$

Where Θ and $(T_x, T_y)^T$ are unknown parameters, $(f_x,f_y)^T$ and $(y_1,-x_1)^T$ can be obtained during the process of optical flow calculation. Therefore, when obtaining the ground motion we can use these ground point information including world coordinate and magnitude of optical flow to calculate the ground movement information, that is unknown parameters Θ and $(T_x, T_y)^T$ can be acquired.

By undertaking above procedure, the ground movement information can be obtained and used to interpolate image, then compensated image is generated. Therefore, thanks to the mapping between image coordinate and world coordinate, we will compensate previous frame image to a new image by projecting the image coordinates to the ground plane and compensating the ground movement on the ground plane then back projecting to the image plane again. Fig. 4-15 illustrates the procedure of the compensated image building. For each pixel of the previous frame will be transformed to ground plane which is world coordinate by IPM forward mapping, then for each world coordinate is compensated to new world coordinate by

using ground movement information and Eq. (4-4), the coordinates are back transforming to image coordinate by IPM backward mapping, a new image is interpolated in the image plane. Therefore, the new built image is assuming formed by planar object because of compensated using ground movement. Due to these assume the non-planar object (obstacle) can be extracted by comparing current image and compensated image.

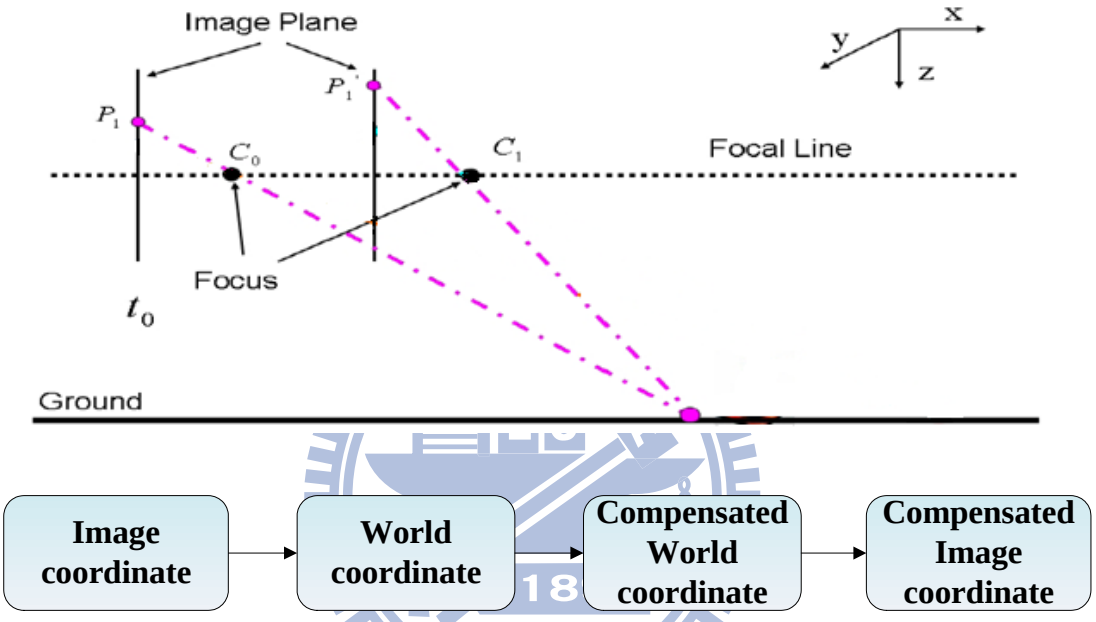


Fig. 4-15 procedure of the compensated image building

4.2.3 Compensation Verification

The ground movement information is used to compensate the image and a new image is built. By considering the temporal coherence, that is to consider the condition that the scene will not change a lot during the few frames. For these reason, when the ground movement information is obtained, the compensated image will be built and used to compare with current frame image then we can acquire an obstacle candidate image which indicate the pixel is obstacle or not. The detail of how to generate obstacle candidate image will be introduced in section 4.3. That is number of

obstacle candidate should be stable because of little varied scene during few frames. Therefore, number of obstacle candidate is anomalous relative to the neighboring frames would indicate the compensation information is not correct.

Observing the recent frames can assist us to verify and ensure the ground movement information is correct. Because the correctness of ground movement is greatly determining the results of following detection procedure, the verification is essential and worth to undertake. As shown in Fig. 4-16, number of obstacle candidate in previous ten frames will be record, and the mean of these is calculated. The equation (4-6) is used to check correctness of ground movement. If the equation is conformed, that is indicating anomalous amount of obstacle candidate and the ground movement information in the current frame possibly erroneous. Then the previous compensation information will be utilized to compensate.

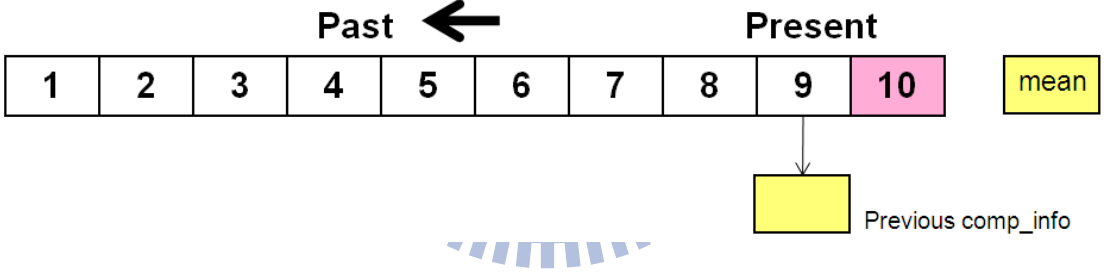


Fig. 4-16 Chart of temporal coherence

Quantitative criterion:

$$\begin{aligned} if\ \ obs\_num &> High\_Thd * mean \\ &< Low\_Thd * mean \end{aligned}$$

$$\Rightarrow comp\_info = previous\ comp\_info \qquad (4\text{-}6)$$

,where High_Thd is user-defined threshold about 1.3, Low_Thd about 0.8

By using temporal coherence, the average amount of obstacle candidate in recent frames is calculated and could be validate the compensation information.

4.3 Obstacle Localization

4.3.1 Obstacle Candidate Image

Due to the property of compensated image as described in section 4.2.2, that is the compensated image is interpolated with ground movement information so that the planar object such as road will be correspond to the same position with current image and the non-planar object will not. Therefore, these characteristic is utilized to detect the obstacle region via image difference as depicted in Fig. 4-17 the image difference block. By image compare between current frame image and compensated image, the planar region will be eliminated and the obstacle region will be marked.

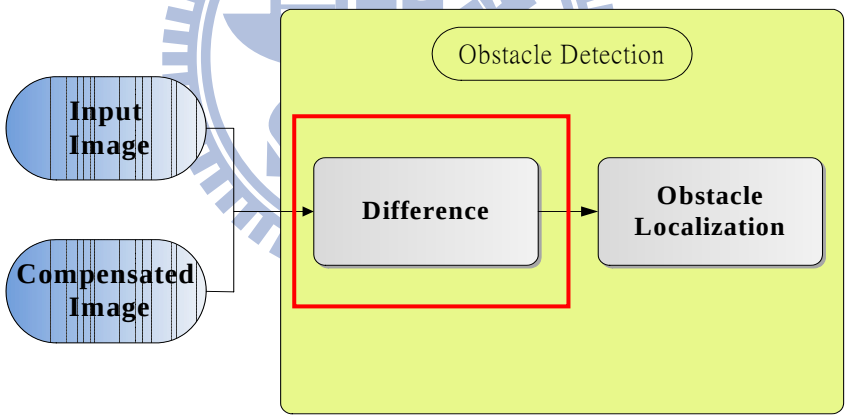


Fig. 4-17 Flow of Obstacle Detection

Here, it can be determined whether a point is on the ground plane or not by comparing the gray values of the corresponding pixels on two image frames (see Eq. 4-7), where Δ is the threshold for the maximum disparity of gray value of two corresponding points. In this way, we can detect obstacles above the ground plane. When image difference is completed, the obstacle candidate image is obtained which is used to indicate every pixel is belonging to obstacle or not. Fig. 4-18 shows some

results of image difference between current image and compensated image, the pixels which are marked as blue are determined as obstacle.

$$\left|cur\_image(P) - comp\_image(P)\right| = \begin{cases} \geq \Delta \quad , \quad \rightarrow \text{obstacle} \\ < \Delta \quad , \quad \rightarrow \text{ground} \end{cases} \tag{4-7}$$

,where cur_image is current frame image, and comp_image is compensated image

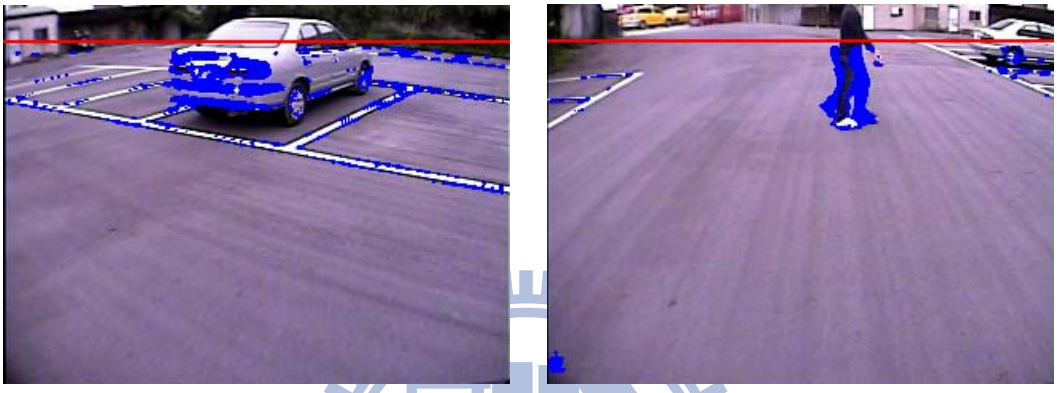


Fig. 4-18 results of image difference between current image and compensated image

4.3.2 Obstacle Localization

In this section, the objective of the obstacle localization procedure is to locate the position of each target objects. The obstacle candidate image is used to extract obstacle region and locate the position of obstacle. Fig. 4-19 illustrates the flow of obstacle localization, obstacle candidate image is integrating vertically and for each obstacle we will locate the closest position to ego-vehicle. Owing to the distance between each obstacle and the ego-vehicle is the most important task for driver when backing up. Therefore, in our research the main purpose is to look for the closest point to the camera not to bind the entire obstacle region.

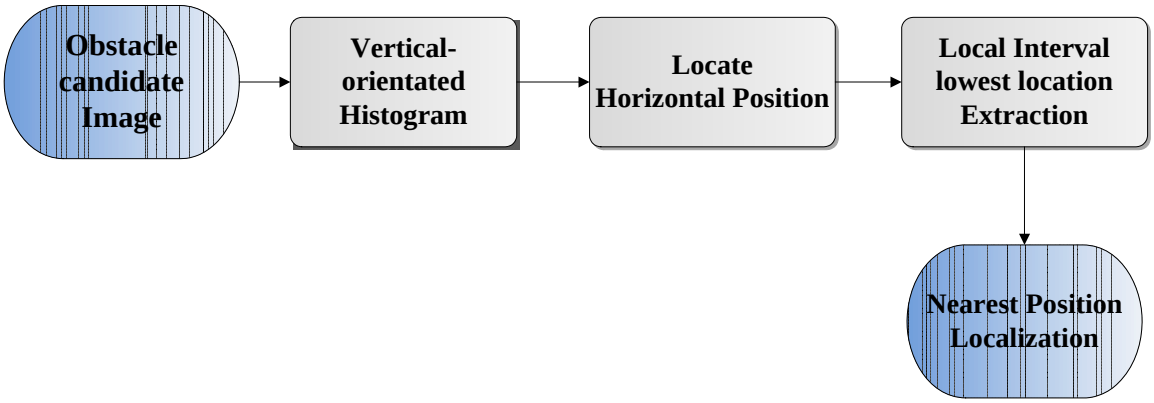


Fig. 4-19 Flowchart of obstacle localization

The first step of the obstacle localization is to calculate the vertical-orientated histogram. The vertical-orientated histogram of binary detection result(obstacle candidate image) is created to ensure horizontal positions of obstacle by extracting the bins which exceeding threshold. The idea is based on the fact that obstacle have more vertically oriented edges compared with their background. By integrating the detection result vertically, the detection on vertically oriented objects is amplified. The vertical-orientated histogram P is created with Eq. (4-8).

$$P_x = \sum_y I_{xy} \quad \forall x \tag{4-8}$$

Here, x is the horizontal coordinate of the image and y is the vertical coordinate of the image. Since the obstacle candidate image is binarized, the pixel intensity I either has a value 0 or 1. Fig. 4-20 illustrates the obstacle candidate image and the corresponding vertical-orientated histogram. Then by using a threshold to indicate which horizontal position have significant accumulation that means obstacle could lie on these x coordinates.

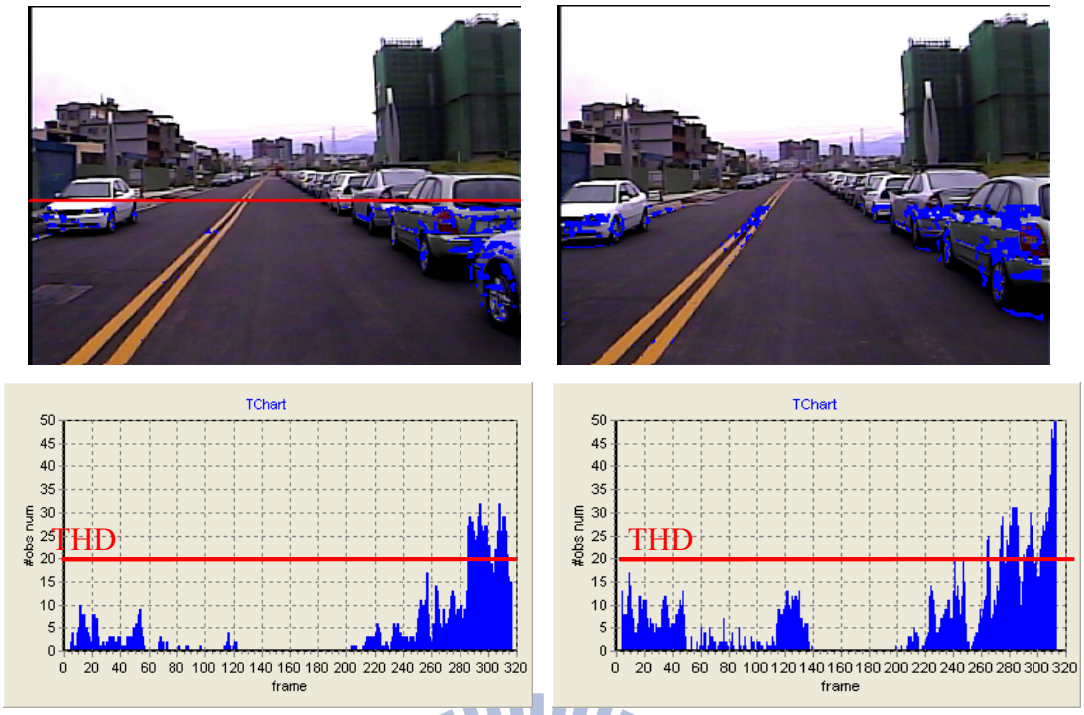


Fig. 4-20 the obstacle candidate image and the corresponding vertical-orientated histogram

The detail of creating vertical-oriented histogram is depicted in Fig. 4-21, first the image will be divided vertically to several bins, for each bin is to scan from bottom to top and to accumulate the number of obstacle candidate. If the interval between two candidates is smaller than a threshold, we consider that these still belong to the same object so that accumulate to the histogram in these bin; otherwise, clean up the accumulation. Besides, creating the vertical-oriented histogram we will save the lowest position of each bin simultaneously. Finally, by extracting lowest position of local interval, the nearest position to the ego-vehicle of each obstacle region can be obtained and the results can be refined. That is for all of bins of histogram which is over the threshold, in every local area that will reserve single position which is the lowest coordinate. For every obstacle region, the closest position to our ego-vehicle will be located and marked to alert the driver.

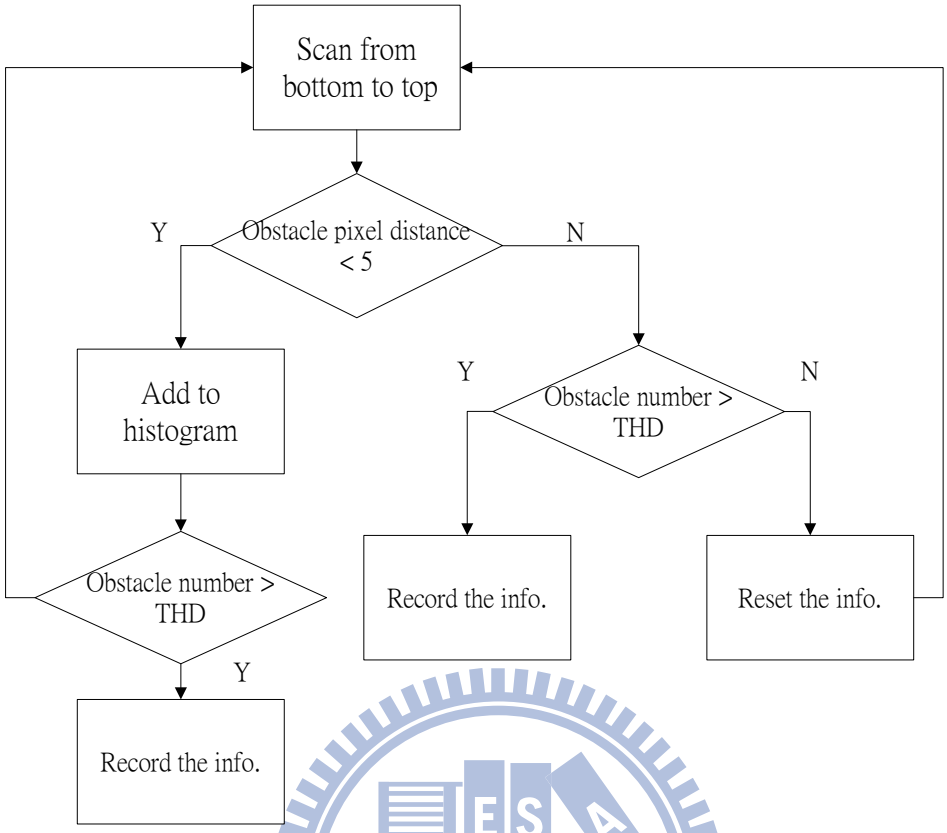


Fig. 4-21 Procedure of creating vertical-orientated histogram

4.4 Obstacle Verification

To decrease amount of false alarming which would cause by erroneous ground movement, the false alarming have occurred commonly on road marking or shadows. Therefore, knowledge of road detection is used to verify results of obstacle detection. In other words, for all of potential target locations are verified using some prior knowledge about the road information. For all of potential target positions, verification is exceeded by checking road ratio of fixed bounding box. If road ratio of a bounding box is over than a threshold, that the position could be the road region, then we will filter out the result. On the contrary, road ratio of a bounding box is

smaller than a threshold then the position is not considered to be a ground region. Therefore, we consider that the region is part of obstacle and reserve the result. Fig. 4-22 is illustrated the idea of obstacle verification. By final checking of road ratio, some error such as road marking would be removed and the result of obstacle detection is more robust.

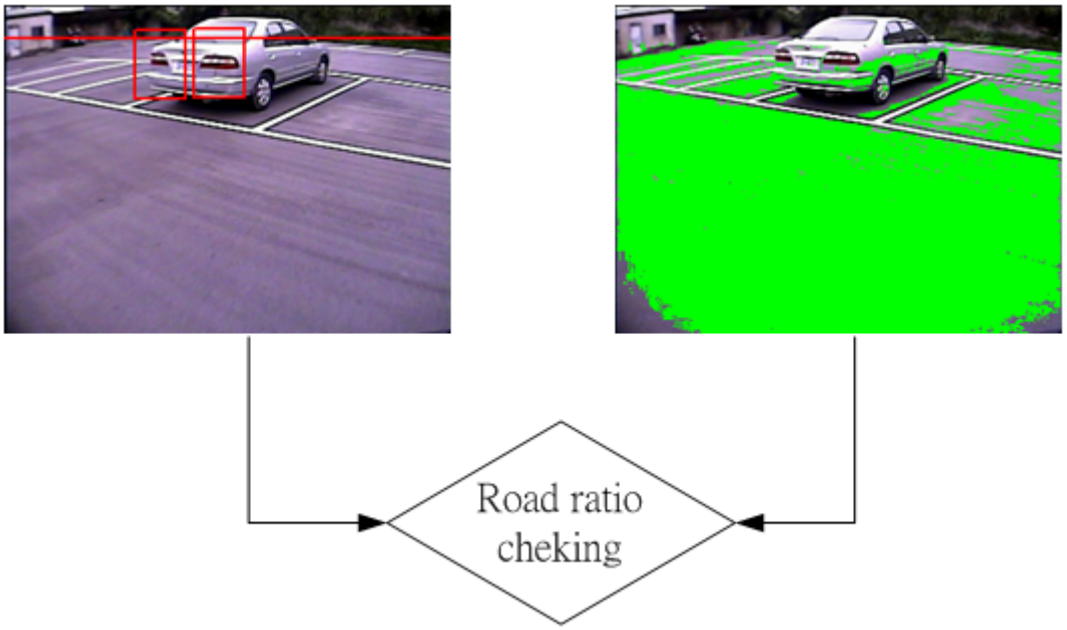


Fig. 4-22 procedure of obstacle verification

4.5 Distance Measurement

For backing up maneuver safety, the position of objects in rear view which are driver concerning with during parking period. Among of all obstacle information the distance between target and ego-vehicle is most essential. By acquiring the distance, we can realize objects are near or far away from our ego-vehicle. Then some backing-up crashes could be avoided. The distance which we want to know is the closest position of obstacle to our vehicle because of the nearest collision would occur

here. Therefore, the distance between the lowest position of each obstacle which is located by the obstacle localization procedure and our vehicle will be estimated by our procedure.

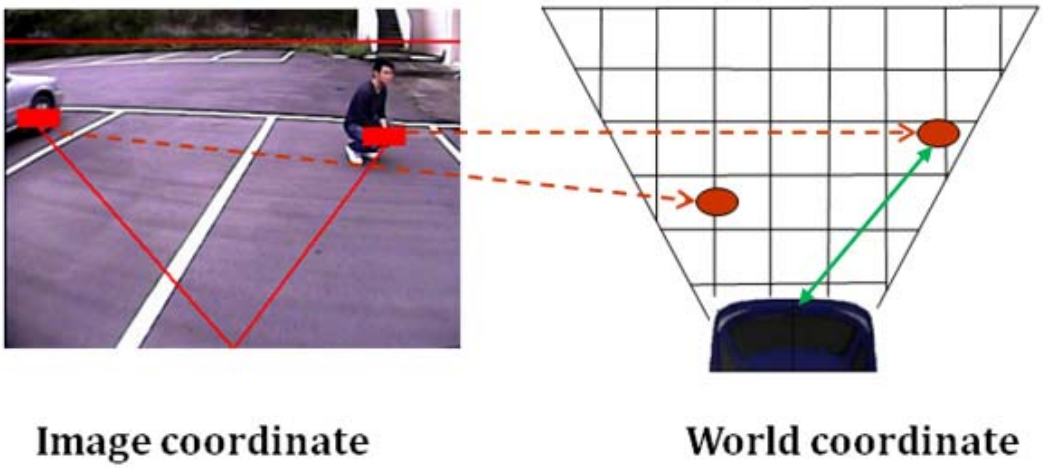


Fig. 4-23 Transformation between image coordinate and world coordinate

As shown in Fig. 4-23, all the obstacle positions can be transformed from the image coordinate to world coordinate by IPM calibration procedure. Therefore, the first step of distance measurement is to obtain the world coordinate of each target. Then the world coordinate will be transformed to real distance by scaling. In order to estimate the scale between real length and world coordinate, the horizontal and vertical reference line on the ground should be measured as depicted in Fig 4-24. By calculating the proportion between reference line in real scene and in the world coordinate, the scale of horizontal and vertical will be acquired to undertake the mapping between world coordinate and real length. The Fig. 4-25 is illustrated the flow of distance measurement, that is transforming the image coordinate of target to world coordinate first, then mapping the world coordinate to real distance. The distance how far away our vehicle is obtained.

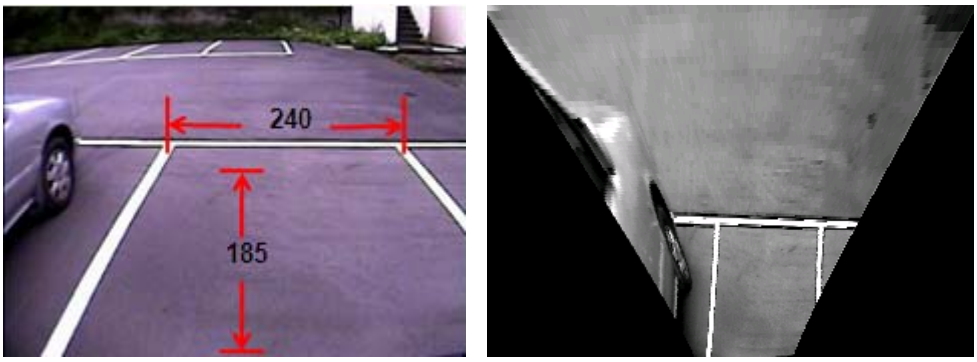


Fig. 4-24 Scale measure between world coordinate and real length

Fig. 4-25 flow of distance measurement

Chapter 5
Experimental Results

5.1 Experimental Environments

Due to our research is to develop a vision-based obstacle detection algorithm, we mounted a CCD camera on back of vehicle with a fixed height and tilt angle. The environment of camera setup is shown in Fig. 5-1. Then our algorithm was implemented on the platform of PC with Intel Core2 Duo 2.2GHz and 2GB RAM. Borland C++ Builder is our developing tool and operated on Windows XP. All of our testing inputs are uncompressed AVI video files. The resolution of video frame is 320*240.

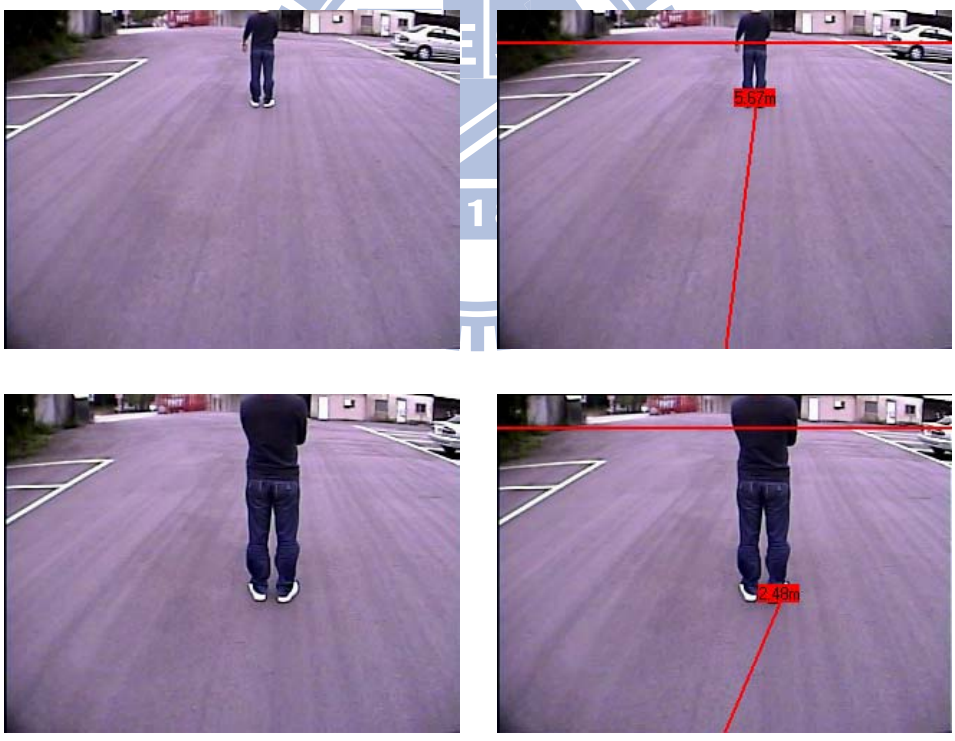
Fig. 5-1 Environment of camera setup

5.2 Experimental Results of Obstacle detection

In this section, several results of obstacle detection will be presented. We will show the experimental results of the proposed algorithm on some conditions which could happen during backing up period such as pedestrian waking or stop on vehicle's

driving path, or some common parking situations. In the following, the columns of left side contain original video sequence and the columns of right side contain detection results of the proposed algorithm. We use “red” block to represent the target position and mark a value to indicate the distance between target and our vehicle. The upper horizontal red line is to indicate the detection range which is below the line in the image. Because of the geometrical characteristic of IPM, the detection range is limited to the region under the horizon, and it will change the position of line with the different camera setup information (such as height or tilt angle of camera).

Fig. 5-2 illustrates a scenario that is a pedestrian stand or squat on the vehicle’s driving path. The results are show that the static pedestrian could be detected no matter standing or squatting by a lower height.



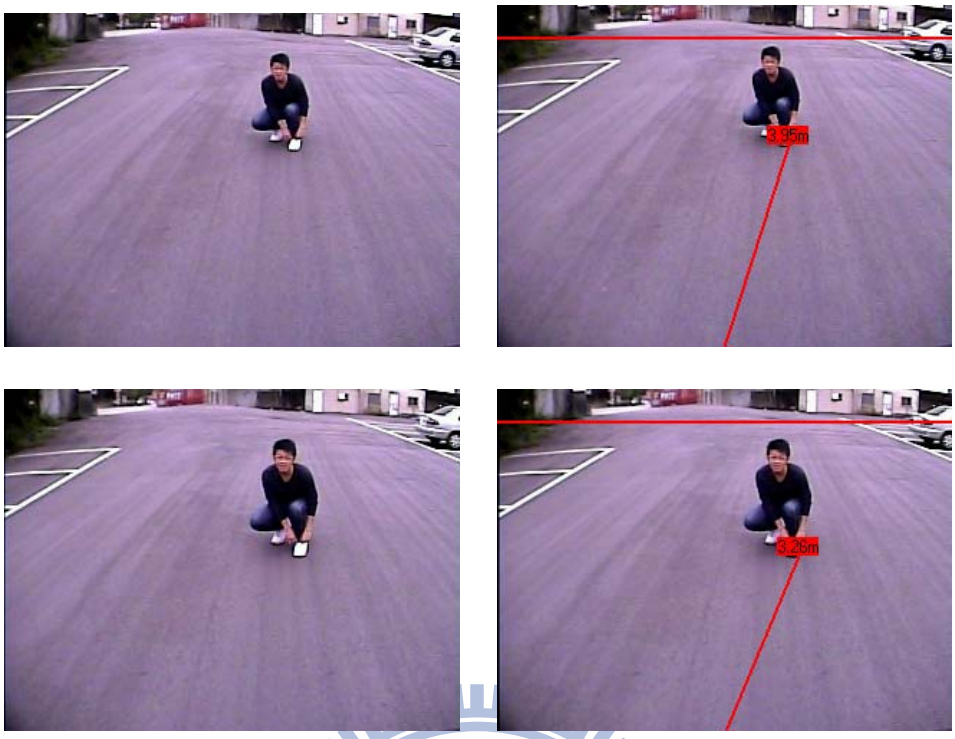


Fig. 5-2 Pedestrian stop on vehicle's driving path

Fig. 5-3 illustrates an example a pedestrian crosses a vehicle's path when vehicle is moving straight. Passing pedestrian can be detected correctly.

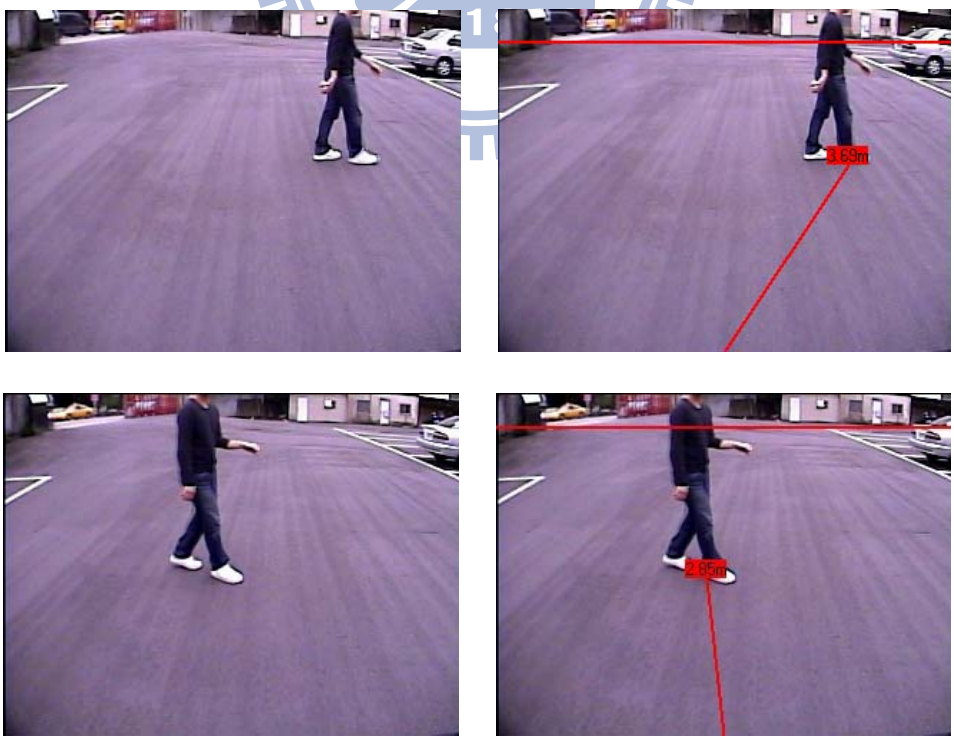


Fig. 5-3 Pedestrian crossing on vehicle's driving path

Fig. 5-4 shows an example of a typical parking situation, there is a vehicle stopping on the side of parking space. Stopping vehicle can be detected correctly when driver is parking.

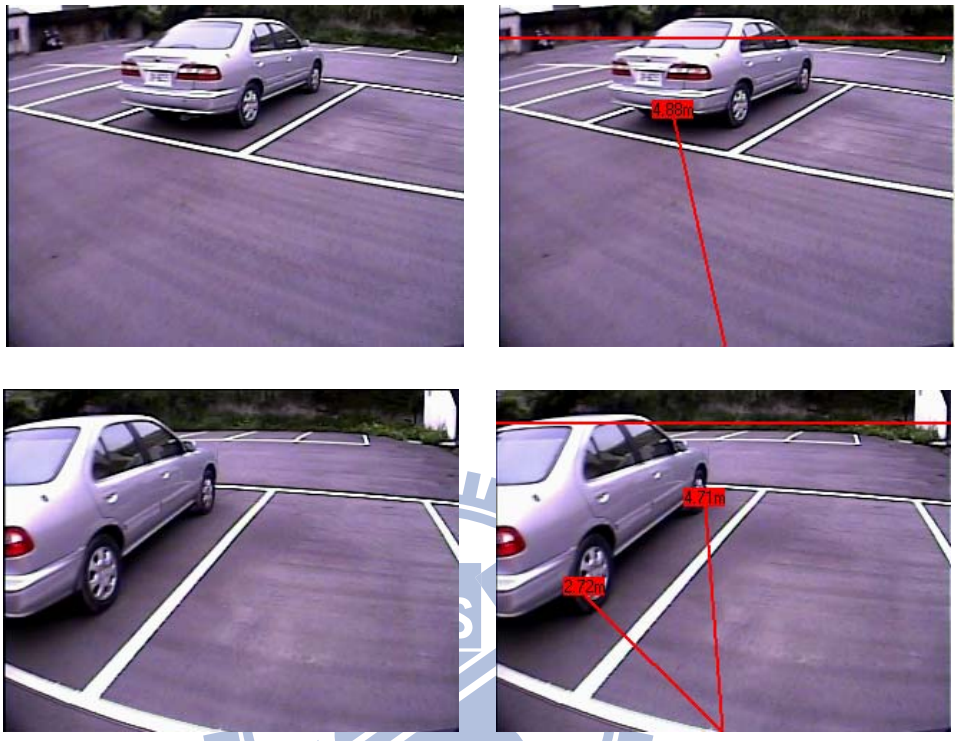


Fig. 5-4 A typical parking situation

Fig. 5-5 illustrates a multiple objects condition when backing up a car, there is a vehicles stopping on the side of parking space and a pedestrian squatting in the parking space for a lower height which driver could ignore commonly. The proposed system can detect them simultaneously.

Fig. 5-5 Multiple objects in parking situation

In Fig. 5-6 we test our system in a low contrast environment which is in the evening, and that is commonly occurring in the parking condition. Fig. 5-7 is also in a low contrast environment and interfered by brake lights furthermore. Our system wouldn't be affected by the external environment and obstacle region can be detected correctly.

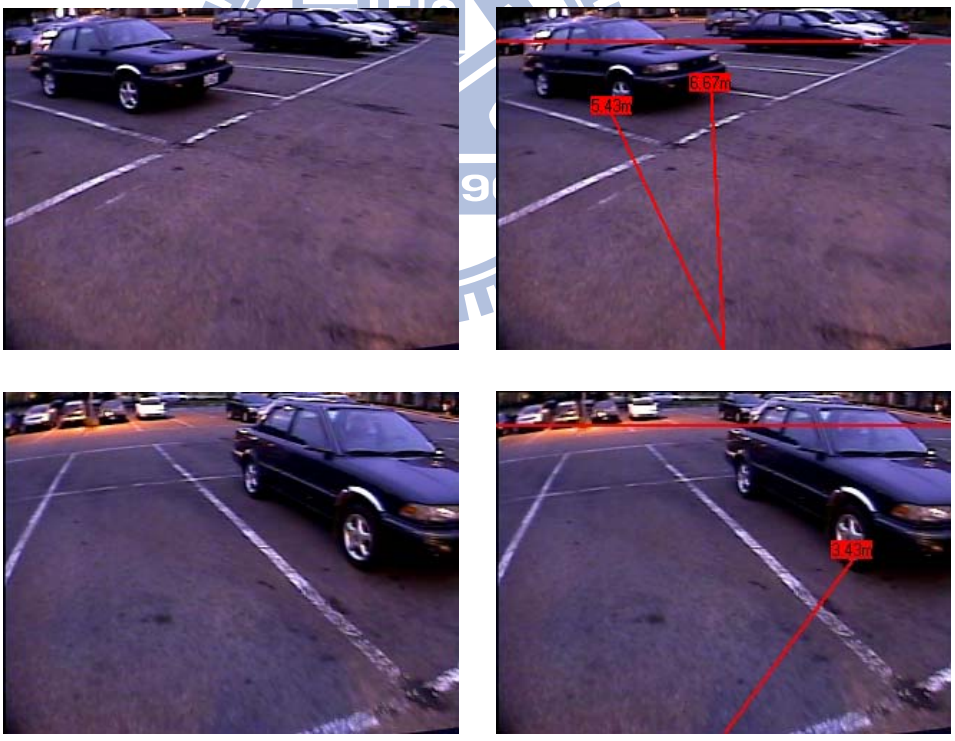


Fig. 5-6 A low contrast environment when backing up

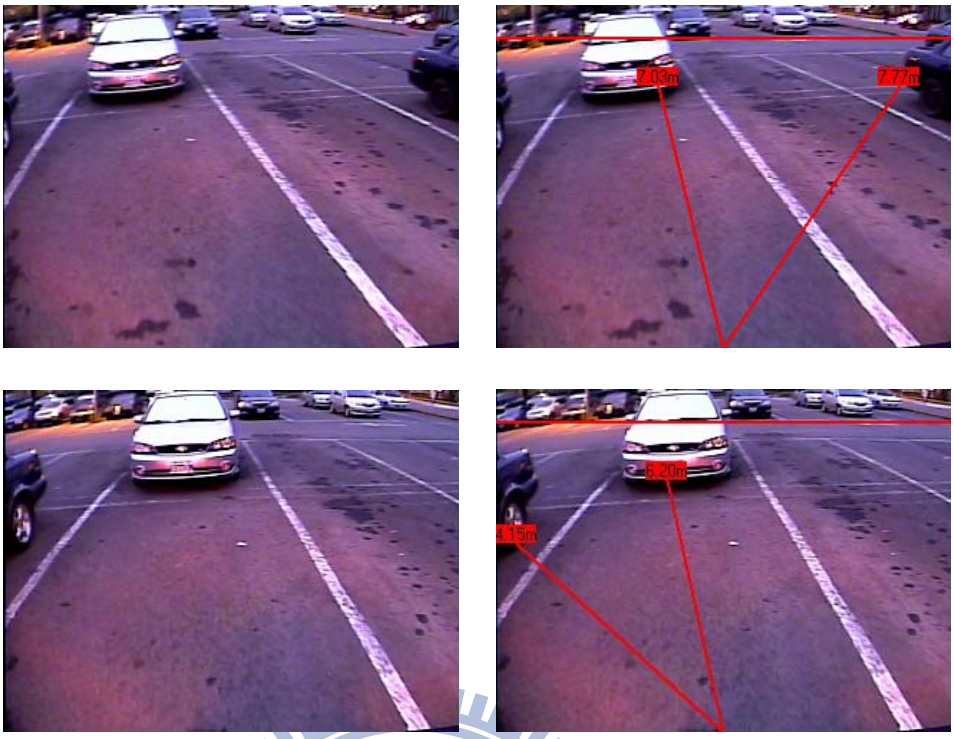
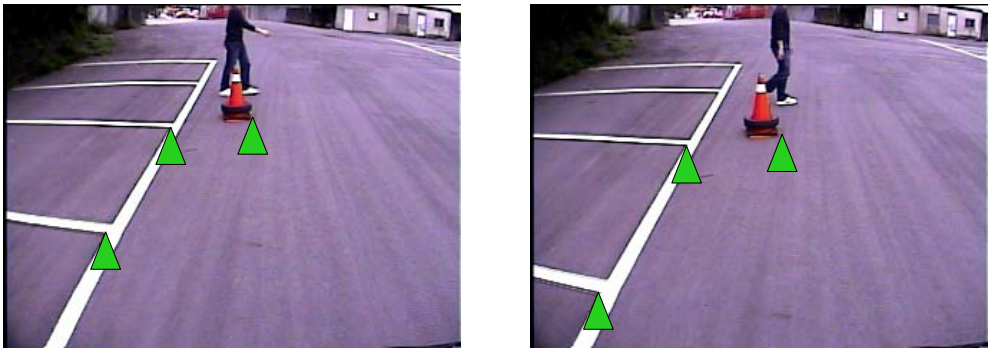


Fig. 5-7 A low contrast environment and interfered by brake lights when backing up

5.3 Accuracy Evaluation

5.3.1 Compensation Evaluation

In order to verify the accuracy of the proposed ground movement estimation technique, an experiment is designed to check compensation of ground movement. The first task we should accomplish is to establish ground truth manually that can assist us to evaluate the results. Due to evaluate the compensation of ground movement, we marked three ground points for each frame and corresponding positions in previous frame. As shown in Fig. 5-8, the three ground points marked between consecutive frame images.



Previous frame　　Current frame

Fig. 5-8 Diagram of ground truth building

When the ground truth is established, for each ground point in previous frame we utilize the proposed algorithm to obtain compensated position. If ground movement information is accuracy, the compensated position should be identical to position in current frame of ground truth. Therefore, utilizing Eq. (5.1) to calculate compensation error, that is to indicate the error distance between actual position and compensated position. Then we implement an approach as introducing in [6] and calculate the experimental results. The comparison results are presented in Table 5-1 by testing 514 ground truth data. In addition, we calculate the compensation error in real distance and bird's view image which is scaled to 320*240 separately. The proposed approach outperforms the method in [6] with average compensation error of 5.5 cm and 11.2 cm in real scene.

$$Compensation\ error = dist(current\_position,\ compensated\_position) \quad (5.1)$$

Compensation error Avg.		
	Proposed method	[6]
Bird's view image (pixel)	**0.9**	1.82
Real distance (cm)	**5.5**	11.2

Table 5-1 Comparison results of compensation error

5.3.2 Accuracy Evaluation of Obstacle detection

In order to assess the performance of the proposed obstacle detection algorithm it is necessary to test the system on many different conditions corresponding as much as possible to situations that one may find when backing up in real life situations. We developed a group of scenarios that are used when evaluating the system, such as a pedestrian crossing or stopped in the vehicle's driving path, or some commonly parking situations with some objects in the parking space when daytime or evening. Then we verify the effectiveness of the proposed system with an event-based method that is to consider the detection rate and false alarm rate by count how many events have to trigger in every frame. The detection rate and the false alarm rate are calculated as follow:

$$\text{Detection Rate} = \frac{N_{\text{detected}}}{N_{\text{obstacle}}} \times 100\% \tag{5.2}$$

$$\text{False Alarm Rate} = \frac{N_{\text{false detected}}}{N_{\text{correct detected}}} \times 100\% \tag{5.3}$$

$$\text{Failed Detection Rate} = \frac{N_{\text{non-detected}}}{N_{\text{obstacle}}} \times 100\% \tag{5.4}$$

In this experiment, a total of 851 frames of images were extracted from seven daytime and evening situations and there are 813 obstacle spots that need to be extracted. The experimental results are shown in Table 5-2, and the detection rate is 86.7% false alarm rate is 2.5%. Thus, the experimental result demonstrates the effectiveness of the proposed technique.

	Correct detection	False detection	Failed detection
Obstacle spots	705	18	108
Rate (%)	86.7%	2.5%	13.3%

Table 5-2 Accuracy evaluation of proposed obstacle detection

Then we implement an approach as introducing in [6] and calculate the experimental results. Table 5-3 shows the experimental results, and the detection rate is 78.5% false alarm rate is 24.8%. The proposed approach outperforms the method in [6] with detection rate especially in false alarm rate. This is because the method in [6] utilized the concept of motion similarity to extract ground feature point and estimate ground movement, but in many cases the ground movement information is erroneous. Therefore, the erroneous ground movement information cause many false alarming, and the detection rate is also decreased because of false position is localized in the lower spot caused by lane marking that obstacle position is missed. As depicted in Fig. 5-9, the erroneous ground movement can result in a large amount false alarm.

	Correct detection	False detection	Failed detection
Obstacle spots	638	158	175
Rate (%)	78.5%	24.8%	21.5%

Table 5-3 Accuracy evaluation of [6]

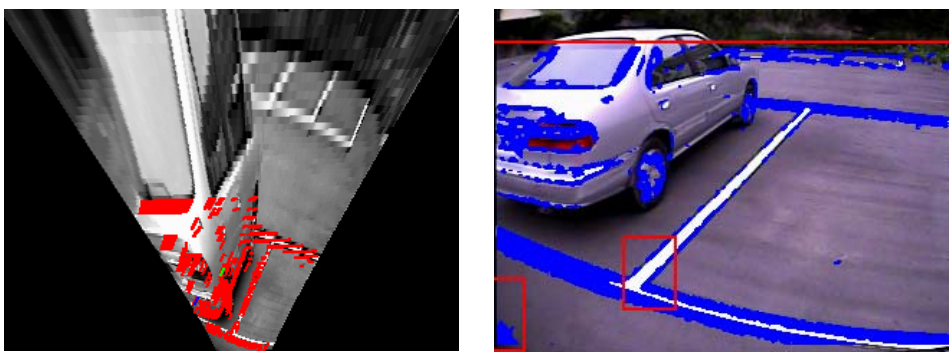


Fig. 5-9 Result of erroneous ground movement

5.3.3 Accuracy Evaluation of Obstacle Distance

For evaluation of distance measurement, we compare length of real line with the length which estimated by the proposed procedure to obtain the distance measurement error. By testing 590 data such as illustrated in Fig. 5-10 the land marking, the error of distance measurement within short and long range respectively is shown in Table 5-4. Due to the geometric characteristic of calibration procedure, the calibration error in the far range will be enlarged. Therefore, the average distance measure error of near range is about 0.16m but of far range is about 0.68m. The experimental results demonstrate that estimation for distance of target position is accurately.

Fig. 5-10 Land marking for distance measurement

	Near range (3~5m)	Far range (5~8m)
Average Distance Error (m)	0.16 m	0.68 m

Table 5-4 Experimental result of distance measurement

Chapter 6

Conclusions and Future Work

For generic obstacle detection, researchers have proposed many methods which focus on stereo vision. Compared to other researches we proposed a system which could automatic detect obstacle with only a single camera mounted on a moving vehicle. Besides, the movement of ego-vehicle which is generally acquired by external sensors such as odometer, we propose a ground movement estimation method that can only adopt image knowledge to obtain these information effectively.

In our research, we intend as obstacle any object that can obstruct the vehicle's driving path or anything raise out significantly from the road surface. Therefore, we propose a ground movement compensation based approach to detect non-planar objects. In addition, adopting different characteristics between planar and non-planar object result from IPM to detect obstacle. The proposed ground movement estimation technique is employing road detection to assist in obtaining most useful ground features in the image, and analysis the principal distribution of optical flow of these feature points, the ground movement for compensation is obtained accurately. The accurately ground movement information can improve the performance of obstacle detection. Thus, the experimental results on many conditions which could occur during backing up period are already used to demonstrate the effectiveness of the proposed obstacle detection algorithm. Finally, we utilize calibration procedure to achieve distance measurement for every detected obstacle. By indicating information of obstacle distance, driver can realize objects are near or far away from our ego-vehicle. Besides, the estimation for distance of target position is verified by

practical measurement, our proposed method can achieve the accurately.

So far, the proposed obstacle detection algorithm can operate well in variant conditions during the backing up maneuver. However, a weak point of the proposed compensation based detection is the detection when vehicle is stationary. In the future, the work should be committed toward utilizing single frame to detect non-planar objects to improve the performance on a stationary scene.

References

[1] www.freeway.gov.tw/Publish.aspx?cnid=590&p=94

[2] www.iek.itri.org.tw

[3] http://www.motc.gov.tw/mocwebGIP/wSite/mp?mp=1

[4] NHTSA: 'Vehicle backover avoidance technology study', Report to Congress, November 2006

[5] M. Bertozzi, A. Broggi, and A. Fascioli, "Stereo inverse perspective mapping: theory and applications," *Image and Vision Computing,* vol. 16, pp. 585-590, Jun 1998.

[6] M. Bertozzi and A. Broggi, "GOLD: A parallel real-time stereo vision system for generic obstacle and lane detection," *IEEE Transactions on Image Processing,* vol. 7, pp. 62-81, Jan 1998.

[7] Wen-Liang Ji , "A CCD-Based Intelligent Driver Assistance System-Based on Lane and Vehicle Tracking, " National Cheng Kung University, PhD degree, 2005.

[8] P. Cerri and P. Grisleri, "Free Space Detection on Highways using Time Correlation between Stabilized Sub-pixel precision IPM Images," in *Robotics and Automation, 2005. ICRA 2005. Proceedings of the 2005 IEEE International Conference on*, 2005, pp. 2223-2228.

[9] A. M. Muad, A. Hussain, S. A. Samad, M. M. Mustaffa, and B. Y. Majlis, "Implementation of inverse perspective mapping algorithm for the development of an automatic lane tracking system," in *TENCON 2004. 2004 IEEE Region 10 Conference*, 2004, pp. 207-210 Vol. 1.

[10] S. Tan, J. Dale, A. Anderson, and A. Johnston, "Inverse perspective mapping and optic flow: A calibration method and a quantitative analysis," *Image and Vision*

Computing, vol. 24, pp. 153-165, Feb 2006.

[11] J. Gang Yi, C. Tae Young, H. Suk Kyo, B. Jae Wook, and S. Byung Suk, "Lane and obstacle detection based on fast inverse perspective mapping algorithm," in *Systems, Man, and Cybernetics, 2000 IEEE International Conference on*, 2000, pp. 2969-2974 vol.4.

[12] M. Nieto, L. Salgado, F. Jaureguizar, and J. Cabrera, "Stabilization of Inverse Perspective Mapping Images based on Robust Vanishing Point Estimation," in *Intelligent Vehicles Symposium, 2007 IEEE*, 2007, pp. 315-320.

[13] Ching-Chiuan Yang, "Construction of Fisheye Lens Inverse Perspective Mapping Model and Its Application of Obstacle Detection", National Chiao Tung University, Master degree, June 2008.

[14] Q. T. Luong, J. Weber, D. Koller, and J. Malik, "An integrated stereo-based approach to automatic vehicle guidance," in *Computer Vision, 1995. Proceedings., Fifth International Conference on*, 1995, pp. 52-57.

[15] ONOGUCHI K.: "Shadow elimination method for moving object detection". Proc. 14th Int. Conf. Pattern Recognition, Brisbane, Proc, Qld, Australia, August 1998, vol. 1, pp. 583–587

[16] W. Kruger, W. Enkelmann, and S. Rossle, "Real-time estimation and tracking of optical flow vectors for obstacle detection," in Intelligent Vehicles '95 Symposium., Proceedings of the, 1995, pp. 304-309.

[17] Guanglin Ma, Su-Birm Park, S. Miiller-Schneiders, A. Ioffe, A. Kummert, "Vision-based Pedestrian Detection - Reliable Pedestrian Candidate Detection by Combining IPM and a 1D Profile," Intelligent Transportation Systems Conference, 2007. ITSC 2007. IEEE

[18] Guanglin Ma, Su-Birm Park, S. Miiller-Schneiders, A. Ioffe, A. Kummert, "A Real Time Object Detection Approach Applied to Reliable Pedestrian Detection,

" Proceedings of the 2007 IEEE Intelligent Vehicles Symposium Istanbul, Turkey, June 13-15, 2007

[19] Guanglin Ma, Su-Birm Park, S. Miiller-Schneiders, A. Ioffe, A. Kummert, "Pedestrian detection using a singlemonochrome camera," Intelligent Transport Systems, IET, March 2009, pp. 42–56

[20] M. Bertozzi, A. Broggi, P. Medici, P. P. Porta and R. Vitulli, "Obstacle detection for start-inhibit and low speed driving", Intelligent Vehicles Symposium, 2005. Proceedings. IEEE

[21] Changhui Yang, Hitoshi Hongo, and Shinichi Tanimoto, "A New Approach for In-Vehicle Camera Obstacle Detection by Ground Movement Compensation", Proceedings of the 11th International IEEE Conference on Intelligent Transportation Systems Beijing, China, October 12-15, 2008

[22] B. K. P. Horn and B. G. Schunck, "Determining optical flow," Artificial Intelligence 17 (1981): 185–203.

[23] J.-Y. Bouguet et. al., "Pyramidal implementation of the Lucas kanade feature tracker description of the algorithm," Intel Corporation, Microprocessor Research Labs, OpenCV document.

[24] Gary Bradski, Adrian Kaehler, Learning OpenCV Computer Vision with the OpenCV Library, O'REILLY

[25] C. Harris and M. Stephens, "A combined corner and edge detector," Proceeding of the 4th Alvey Vision Conference (pp. 147–151), 1988.

[26] Shen-Chi Chen, "A New Method of Efficient Road Boundary Tracking Algorithm Based on Temporal Region Ratio and Edge Constraint", National Chiao Tung University, Master degree, June 2009.