國立交通大學

電控工程研究所

博士論文

數位影像穩定技術及其應用

Digital Image Stabilization Technique and its Applications

研究生：徐聖哲

指導教授：林進燈　教授

中華民國九十九年六月

數位影像穩定技術及其應用

Digital Image Stabilization Technique and its Applications

研 究 生：徐聖哲　　Student：Sheng-Che Hsu

指導教授：林進燈博士　　Advisor：Dr. Chin-Teng Lin

國 立 交 通 大 學

電控工程研究所

博 士 論 文

A Dissertation

Submitted to Institute of Electrical Control Engineering

College of Electrical and Computer Engineering

National Chiao Tung University

in partial Fulfillment of the Requirements

for the Degree of

Doctor

in

Electrical Control Engineering

June 2010

Hsinchu, Taiwan, Republic of China

中華民國九十九年六月

數 位 影 像 穩 定 技 術 及 其 應 用

學生：徐聖哲指導　　教授：林進燈教授

國立交通大學電控工程研究所 博士班

摘 要

本論文為提出數位影像穩定技術研究成果及其應用，本文中提出如何在手持式、車用或固定式監視攝影機，因人為不可預期的抖動、車輛行駛巔岐、方向盤轉動的效應或風力、外力等的影響而造成影像的抖動，以數位影像處理的方式去除不必要的抖動而保留必要的移動。數位影像穩定技術主要可分為兩部份：(1)如何從影像序列中有效率地估算出準確、可靠的全域移動向量；(2)如何從所取得的移動向量在邊界的限制下補償出一平滑的影像運動軌跡。

本文在**移動向量估測**上提出倒三角方法以尋找區域移動向量、最佳化代表點選擇以降低尋找移動向量的計算量、產生精煉型移動向量以應用在乏特徵的狀況仍可估算全域移動向量、另提出天際線的檢測方法與以背景為基礎的對等演算法以求得較可靠的全域移動向量。

本文在**移動向量補償**上提出移動軌跡的繪製與平滑指標的計算以驗證所提出移動向量補償方法改善數位影像穩定的定量分析，同時在補償方法中提出在回路中加上一內部回授積分器以改善鏡頭在定速移動時造成補償效果不良的問題。最後則以模糊推論的機制，以兩種不同的方法透過模糊推論，選擇較佳的補償方式，以適應數位影像防振在各種情況的應用。

經實驗的結果，本文所提出的方法可適用在不同狀況的影像序列如乏特徵、重複圖樣、大移動物件及大區域低對比的影像狀況，而能估算出準確的全域移動向量。在移動向量補償上則解決在定速移動下所造成補償效果降低的問題，並以模糊推論整合方式有效地改善補償向量的補償效果，從平滑指標與移動軌跡圖上均可驗證本文所提出的方法有效地改善數位影像穩定的效能。

Digital Image Stabilization Technique and its Applications

Student：Sheng-Che Hsu　　Advisors：Dr. Chin-Teng Lin

Institute of Electrical Control Engineering
National Chiao Tung University

Abstract

In this dissertation, a digital image stabilization (DIS) technique and its applications are proposed as a way to remove the unwanted shaking phenomena in the image sequences captured by hand-held, in-car or fixed-type surveillance camcorders, without being affected by moving objects in the image sequence or by the intentional panning motion of the camera. DIS contains two major parts: (1) How to estimate an efficient, precise, and reliable global motion vector. (2) How to use the existing GMV to compensate for a smooth motion trajectory within the window shifting allowance boundary.

For motion estimation (ME), an inverse triangular method is proposed to look for the local motion vector (LMV). An optimization of the representative points is proposed to reduce computation complexity, and a refined motion vector is proposed to apply to any ill-conditions of the GMV estimation. Skyline detection for in-car applications and background based peer to peer evaluation are proposed to enforce the reliability of the GMV estimation as well.

For motion compensation (MC), a plot of the motion trajectory and a smoothness index evaluation are proposed to quantitatively verify the analysis that shows the improvement of the MC. An inner feedback-loop integrator has been applied to the MC to improve image stabilization during the camera's panning motion. Finally, Fuzzy inference digital image stabilization (FIDIS) is proposed to adaptively determine better motion compensation methods through the use of two different MCs.

Experimental results show that the proposed methods of this dissertation adapt to different conditions of image sequencing, such as a lack of features, repeated patterns, large moving objects and large low-contrast areas in the image and that they can also estimate the GMV precisely as well. The degradation of image stabilization during the panning motion is solved by adding an inner feedback-loop integrator. The proposed FIDIS also shows effective

improvements in different conditions of image sequence through the evaluations of the smoothness index and the motion trajectory.

Acknowledgement

首先感謝指導教授林進燈教務長多年來的指導。無論在專業上或生活上的指導與照顧，都使我受益良多。林進燈教授在研究上的精神是我學習最佳的榜樣，其待人誠懇平易近人及樂於助人的特質更是讓我在進修博士班十年的歲月中最大的鼓勵。對於本論文的完成，林教授給與我很大的鼓勵與支持，再次由衷的感謝。

在家人方面，謝謝年邁的父親，從您過去以身作則堅持與毅力的表現，在在影響我對進入博士班深造與進修的持續。謝謝我的妻子巽俐與兩位可愛的女兒蕙心、蕙薰，多年來因為學業與工作的關係使得陪你們的時間相對減少，你們對我的期待與包容成為我可以完成這樣的一個理想，如果沒有你們的支持與鼓勵，那將是一件不可能的任務，謝謝！我愛你們。

最後謝謝論文口試委員郭耀煌教授、蔡文祥講座教授、陳永耀教授、陳永昌教授、張志永教授、陳永平教授、蒲鶴章博士，由於您們熱心的指導與建議使論文更加完整，在此致上最高的敬意。

Table of Contents

List of Tables

List of Figures

1. Introduction

1.1. Motivation

Digital video sequences acquired by image captured devices are usually affected by undesired motions which can be classified into several aspects. The image captured by compact and lightweight video cameras, i.e. hand held devices, are usually affected by unstable camera holding or moving platform. The image captured by in-car video cameras are usually affected by a bumpy ride or by the steering of the driver. The image captured by video surveillance systems are usually affected by wind blowing, bird jumping or earthquake. The unwanted positional fluctuations of the video sequence will affect the visual quality and impede the subsequent processes for various applications such as motion coding, video compression, feature tracking, etc. The challenges of a digital image stabilization system (DIS) are: how to compensate for the unwanted shaking of the devices without being affected by large moving objects, ill conditions in the image, and the panning motion of the camera. In this dissertation, the related techniques of DIS will be proposed to tackle these challenges.

1.2. Related Works Review

The image stabilization systems can be classified into three major types: (1) the electronic image stabilizers (EIS) [1]; (2) the optical image stabilizers (OIS) [2]; (3) the digital image stabilizers (DIS) [5]. The EIS stabilizes the image sequence by employing motion sensors to detect the camera movement for compensation. The OIS employs a prism assembly that moves opposite the shaking of the camera for stabilization. Because both EIS and OIS are hardware dependant, the applications are restricted to device built-in on-line process. The (DIS) is the process of removing the undesired motion effects to generate a compensated image sequence by using digital image processing techniques without any mechanical devices such as gyro sensors or fluid prism [4]. The major advantages of DIS are: (1) no restriction of on/off-line applications, (2) suitable for miniature hardware

implementation (since the mechanical device is not required for compensation) [5]. The DIS can be performed either as post-processing after the video sequence was acquired, or in real-time during the acquisition process, depending on the applications. Archive films with undesired shaking effects require post-processing for the video sequences, while camcorders require a real-time compensation process.

The DIS system is generally composed of two processing units. One is the motion estimation unit and the other is the motion compensation unit. The purpose of motion estimation unit is to estimate the reliable camera global movement through three processing steps on the acquired image sequence. (1) evaluation of local motion vectors (LMVs) is the first step of the process; (2) detection of unreliable motion vector components is the next step. (3) determination of the global motion vector (GMV). Following the motion estimation, the motion compensation unit generates the compensating motion vector and shifts the current picking window according to the compensating motion vector to obtain a smoother image sequence. Fig. 1.1 shows the motion compensation schematics. The window of frame$(t-1)$ is the previous compensated image. The compensating motion vector v is generated by the DIS according to the global motion vector between two consecutive images. The window of frame(t) is the picking window according to the compensating motion vector v to minimize the shaking effect.

Various algorithms had been developed to estimate the local motion vectors in DIS applications such as representative point matching (RPM) [3][5][6], edge pattern matching (EPM) [7][8], bit-plane matching (BPM) [4][9] and others [10][11][15][16][17]. It had also been demonstrated that the DIS could reduce the bit rate for video communication [18]. The major objective of these algorithms is to reduce the computational complexity, in comparison to the large area full-search block-matching method, without losing too much accuracy. In general, the RPM can greatly reduce the complexity of the computation in comparison with the other methods. However, it is sensitive to irregular conditions such as moving objects and intentional panning, etc. [9]. Therefore, the reliability evaluation is necessary to screen the undesired motion vectors for the RPM method. In [6], a fuzzy-logic-based approach was proposed to discriminate the reliable motion vector from the local motion vectors. This method produced two discriminating signals based on some image information such as contrast, moving object, and scene changing to determine the global motion vector. However, these two signals cannot widely cover various irregular conditions such as the lack of features or containing large moving objects in the images, and it is also hard to determine an optimum

threshold for discrimination with these various conditions. Some researchers estimate local motion vector using feature based techniques, which track a small number of image features (points, lines, and contours or certain object, etc.) to evaluate the motion vector. This makes it efficient and available for real-time implementation. But the difficulty is that, especially for outdoor applications, it can not stably and accurately find available features in the image [19]. Based on the optical flow technique, a fundamental approach in computer vision, many methods have been proposed in literature to solve different types of problems. The estimation of optical flow is based on the assumption that the intensity of the object (or specified pixel) in the image sequence is constant. The difficulty is that most consumer video camcorders have an auto-shutter function that adjusts itself to average intensity dynamically, so that maintaining constant intensity of the object becomes impossible in real applications. In this paper, a reliable local motion vector extraction method is proposed to determine the global motion vectors for practical applications.

In the motion compensation of DIS, accumulated motion vector estimation [7] and frame position smoothing (FPS) [20][21][22] are the two most popular approaches. The accumulated motion vector estimation needs to compromise stabilization and intentional panning (constant motion) preservation since the panning condition causes a steady-state lag in the motion trajectory [20]. The FPS accomplished the smooth reconstruction of an actual long-term camera motion by filtering out jitter components based on the concept of designing the filter with appropriated cut-off frequency. The disadvantage of FPS is that it does not guarantee the availability of the determined compensating motion vector when the specified-bound is restricted to preserve the effective image area in the DIS applications.

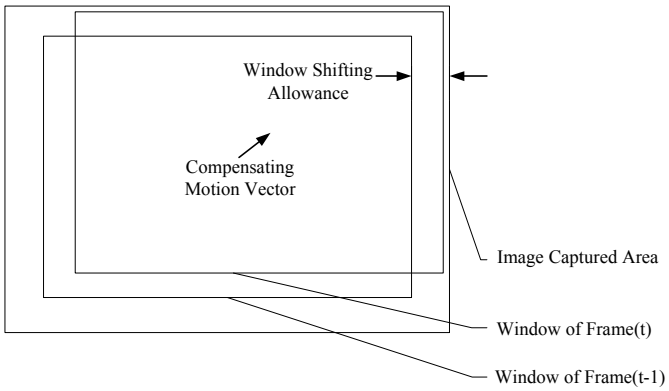


Fig. 1.1. Motion compensation schematics.

1.3. Overview of This Dissertation

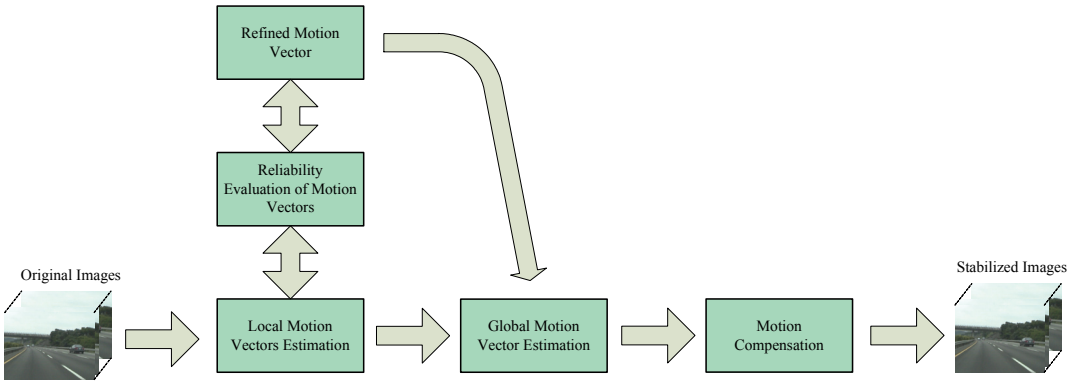


Fig. 1.2. Schematic overview of the digital image stabilization developing techniques

Fig. 1.2 describes a schematic overview of the digital image stabilization developing techniques. It illustrates the three main issues of DIS developments. These issues are local motion estimation, global motion estimation and motion compensation. The issues presented in this dissertation ,such as motion estimation, which include local motion vectors and global motion vector, and motion compensation will be addressed in separate chapters, each with its own introduction, literature overview and algorithm development. This will make it possible to read each chapter without having to cross-reference to the other chapters.

In chapter 2, we start with the issues of local motion vector estimation and global motion vector estimation. Due to these two issues being mutually related in their information process and development technique, it is hard to separate them. Based on the difference in dynamic and real video characteristics, we propose several techniques to tackle irregular images that contain large moving objects, low-contrast area and lack of features to improve robustness and accuracy. The advantages of our proposed methods in the different functions are:

- Use representative point matching to dramatically reduce the computation complexity. This makes it possible to build in the regular processors for realization.
- Propose an inverse triangle method to discriminate the reliability of motion estimation. Based on the inverse triangle method, the related development techniques are:
  - Discriminate the reliability of each local motion vector with respect to each axis.
  - Form a refined motion vector.

- Determine the background-based evaluation area by coarsely detecting the skyline.

- Determine the optimum representative points for stationary video surveillance system. This will improve the computation efficiency.
- Use background based evaluation to determine the global motion vector.

In chapter 3, we address the issue of motion compensation. The objective of motion compensation is to reduce the jiggled image phenomenon by generating a compensating motion vector to stabilize the image sequence. The major work in this part is to improve the existing compensation method within the specified requirements such as limited window shifting allowance, panning condition etc. Therefore, an inner feedback-loop integrator and fuzzy inference have been applied to approach the problem. Quantitative evaluation terms, such as motion trajectory and smoothness index, have been developed for final results comparison as well.

In chapter 4, the experimental results of the algorithms developed in chapter 2 & 3 will be interpreted. First, the accuracy of global motion vector estimation will be compared by the root mean square error (RMSE) method. In this part of the experiments, video sequences captured by different applications of hand-held camcorders and in-car video capture devices will be compared with the RPM_FUZZY method. Secondly, motion compensation with the inner feedback-loop integrator will be demonstrated to show the improvements of the motion trajectory and the smoothness index in panning conditions. Furthermore, we describe the results of applying the fuzzy inference algorithm to stabilize the image sequences in various conditions, which can reduce the drawbacks and keep the merits of each motion compensation method.

2. Motion Estimation

2.1. Introduction

Motion estimation is the process of finding optimal or near-optimal motion vectors. Most of these techniques are developed for video compression to increase compression ratios by making better use of redundant information between successive frames. But the motion estimation techniques used in DIS have some differences. The purpose of the motion estimation in DIS is to estimate the reliable global motion vector through three processing steps on the acquired image sequence: (1) evaluation of local motion vectors (LMVs); (2) detection of unreliable motion vector components; (3) determination of the global motion vector (GMV).

The local motion vectors estimation techniques presently used can be divided into three main categories [22], [24], [25]:

- The intensity gradient techniques [15], [24], [26], [35], [36].
- The frequency-domain techniques [31], [32], [33], [38].
- The block matching techniques [5], [10], [27], [29], [30].

The initial hypothesis in calculating image motion with the intensity gradient techniques is that the intensity structures of the local regions are approximately constant for a short duration. In other words, the image luminance is invariant along motion trajectories. If $I(t,\mathbf{x})$ is the intensity function of image with respect to t and $\mathbf{x}$.

$$I(t,\mathbf{x}) \approx I(t+\delta t,\mathbf{x}+\delta\mathbf{x}), \tag{2.1}$$

where $\delta\mathbf{x}$ is the displacement of the local image region at $(t,\mathbf{x})$ after time δt. The Taylor series expansion of the right hand in (2.1) yields

$$I(t+\delta t,\mathbf{x}+\delta\mathbf{x}) = I(t,\mathbf{x}) + \nabla I(t,\mathbf{x})\cdot\delta\mathbf{x} + \delta t\frac{\partial I(t,\mathbf{x})}{\partial t} + O^2 \tag{2.2}$$

where $\nabla = [(\partial/\partial x),(\partial/\partial y)]$ is the gradient operator, and O^2 includes the 2^{nd} and higher

order terms, which are assumed negligible. According to (2.1), ignoring O^2 and dividing by δt yields

$$\nabla I(t,\mathbf{x})\cdot\mathbf{v}+\frac{\partial I(t,\mathbf{x})}{\partial t}=0, \tag{2.3}$$

where $\nabla I(t,\mathbf{x})$ is the spatial intensity gradient and $\mathbf{v}$ is the image velocity. Equation (2.3) is known as the optical flow constraint equation or spatio-temporal constraint equation. As the image intensity changes at a point due to motion, the constraint equation is not sufficient to compute both components of $\mathbf{v}$. That is to say, only the projection of $\mathbf{v}$ on ∇I can be determined from (2.3). This problem is known as the aperture problem. Only at image locations where there is sufficient intensity structure, such as the Gaussian curvature, can the motion be fully estimated with the use of the optical flow constraint equation. Therefore, an additional constraint must be introduced to regularize the ill-posed problem and to solve the optical flow. To estimate the exact image motion with optical flow, several conditions have to be satisfied. These are:[24]

- Uniform illumination
- Lambertian surface reflectance
- Pure translation parallel to the image plane

The drawback is that these conditions are not entirely satisfied in scenery images. It can only be assumed that these conditions are held partially in the scene. The degree to which these conditions are partly satisfied determines the accuracy with which optical flow approximates image motion. These techniques also suffer from two serious drawbacks [25]: (1) the smoothness constraint leads to an increased prediction error, especially on moving objects boundaries. (2) the dense motion field (i.e., a motion vector per pixel) requires much overhead information.

A second category of motion vector estimation techniques are based on the use of velocity-tuned filters. It is found that motion-sensitive mechanisms operating on Fourier transform techniques can estimate the motion vector of an image for which other matching approaches would fail. The Fourier transform of a translational intensity of a specified image in (2.1) is

$$\mathrm{I}(\omega,\mathbf{k})=\mathrm{I}_0(\mathbf{k})\delta(\mathbf{v}^T\mathbf{k}+\omega), \tag{2.4}$$

where $\mathrm{I}_0(\mathbf{k})$ is the Fourier transform of $I(0,\mathbf{x})$ and $\mathbf{x}$ denotes spatial position. δ is a Dirac delta function and $\mathbf{k},\omega$ denote spatiotemporal frequency. This yields the optical flow constraint equation in frequency space

$$\mathbf{v}^T\mathbf{k}+\omega=0 \tag{2.5}$$

which means $\mathbf{v}$ is a function of $\mathbf{k}$ and ω and forms a plane through the origin of the Fourier space. It was shown that velocity-tuned filters could be tuned to ranges of component image velocity. In general, noise robustness is enhanced because the filters can be designed to attenuate noise through time as well as space [39]. If the images are corrupted by frequency-dependent noise, then the Fourier methods are preferred rather than other methods. The phase correlation method is based on the Fourier Shift Theorem and it was originally proposed for the registration of translated images. The method shows strong robustness against the correlated and frequency dependent noise and non-uniform, time varying illumination disturbances [28]. The drawbacks of frequency domain techniques are: (1) They do not provide a means of assigning confidence to the computed velocities [31]; (2)There's a requirement for a large number of filters to cover frequency space [37]; (3) Numerical differentiation is sometimes impractical because of small temporal support [23].

The block matching techniques are based on the matching of blocks between two images and aim to minimize a disparity measure [25]. These techniques are sometimes called correlation-based techniques or template matching techniques. The basic idea of block matching is defined as a block $B(x_p,y_p)$ around the point p, in which we try to find the best similar block $B(x_p+i,y_p+j)$ shifted by the integer values in pixels in a search space S composed by the i,j such as $-\triangle< i <\triangle$ _and $-\triangle< j <\triangle$. The best similar block is defined as a minimum measurement of a distance or a maximum measurement of the correlation between the intensity of an image in the two corresponding blocks. The block matching techniques are the most intuitive and also the most widely applied techniques to compute the motion vector from an image sequence. Different kinds of algorithms use different criteria for the comparison of blocks. Originally, the algorithm to be used for block matching is referred to as the Full Search or Exhaustive Search. In this, each block within a pre-selected search window is evaluated to the current block by disparity measure criteria. The lowest disparity measure value is the best match for the current block. It shows the excellence of quality and simplicity, but it has a really high computation complexity. Therefore, there are several approaches for reducing the computation complexity, such as the three-step search [41], the .new three-step search [42], [43], the four-step search [48], the one-dimensional full search [47] and the

diamond search [44] have been developed to reduce the search positions. Another method is to simplify the matching operations, such as pixel decimation [49], mini-max criterion [51], boundary match [52], pixel truncation [53], representative point matching [5], [6], edge pattern matching [7], [8] and bit-plane matching [9]. In general, these fast algorithms suffer from considerable peak signal-to-noise ratio (PSNR) degradation compared to full-search block matching algorithm (FSBMA), especially when the motion field is large and complex [30]. One of the motion estimation problems is the reliability measure. The confidence measures can indicate the reliability of estimated motion vectors which may suffer from transparency, occlusion, lacking features, repeated patterns or low contrast. The confidence measures can be used to screen the unreliable motion vectors during the estimation process. In the digital image stabilization applications, some of the motion estimation properties should be emphasized such as (1) low computation complexity; (2) large motion estimation areas that are larger than the macroblocks with a size of 8x8 or 16x16, which are used in MPEG video compression; (3) full usefulness of motion estimation information; (4) high reliability of motion vectors; (5) background based motion vectors.

In this chapter we present the details of motion estimation for digital image stabilization. Firstly, we proposed the architecture of the DIS system, the block diagram of motion estimation and then addressed the evaluation of local motion vectors by the representative point matching method. Secondly, we discussed the detection of irregular conditions in the image sequence. In this section, the inverse triangular method is introduced to measure the reliability of local motion vectors. Based on this method, the selection of the representative point amount is discussed. Thirdly, the next section addressed the generation of refined motion vector (RMV) to tackle the ill-conditions of image sequences.

2.2. Motion Estimation

2.2.1. Motion Estimation Architecture

The architecture of the proposed DIS system, shown in Fig.2.1, includes two processing units, the motion estimation unit and the motion compensation unit. The motion estimation unit consists of three estimators: the local motion vectors (LMV), the refined motion vector (RMV), and the global motion vector (GMV) estimators. The motion compensation unit consists of the compensating motion vector (CMV) estimation and image compensation. In the first part of motion estimation, as shown in Fig. 2.2, the LMVs and RMV estimation is to generate the LMVs and RMV for global motion vector estimation. The LMVs can be obtained from the correlation between two consecutive images by the representative point matching (RPM) algorithm. The RMV can be obtained from LMVs by evaluating the corresponding confidence indices through the irregular condition detection and the proposed RMV generation algorithm.

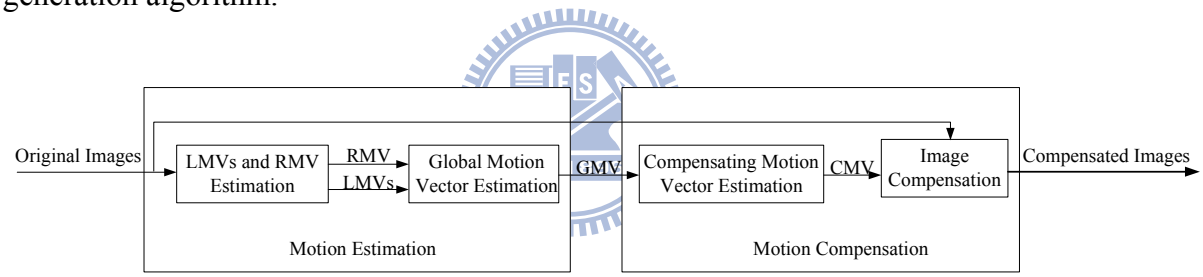


Fig. 2.1. System architecture of the proposed digital image stabilization technique

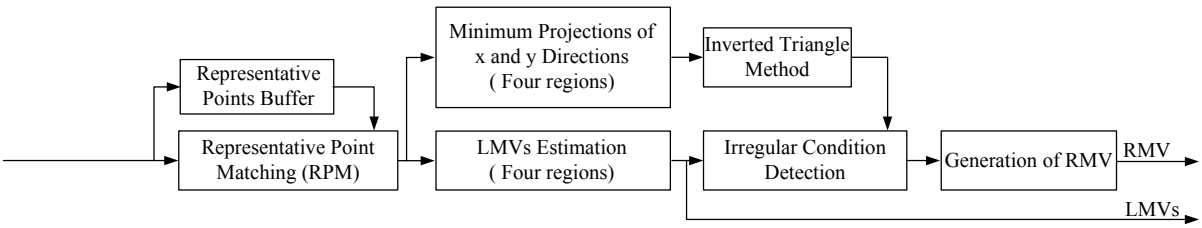
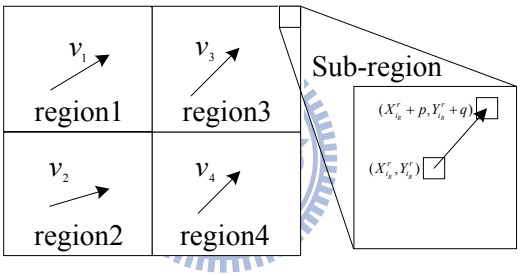


Fig. 2.2. Block diagram of LMVs and RMV estimation.

2.2.2. RPM and Local Motion Estimation

It has been demonstrated that a local approach that uses a regional matching process is more robust and stable than a direct global matching process [50]. This means that using the local motion vectors estimated by the divided regions to determine the global motion vector is more robust and stable than a direct approach. There is also a tradeoff for the size of the divided region. Reducing the size of the divided region increases the robustness, but the size of the divided region should be sufficiently large enough to hold the average distribution. If we want to divide the image so that the horizontal and vertical components have the same partitions, it should be divided into n^2 regions. More divided regions will increase the computational cost to estimate the local motion vector for each region. Therefore, we only divided the image into four regions, as shown in Fig. 2.3, for the RPM method and it can cover various situations in the DIS applications by combining the proposed inverse triangle method and the adaptive background evaluation model.



V_i : local motion vector(i=1,2,3,4)

Fig. 2.3. Division of image for local motion vector estimation

Each region is further divided into 30 sub-regions (with each side of 5 rows×6 columns) and the central pixel of each sub-region is selected as the representative point to represent the pattern of this sub-region. This layout is based on the size of images captured by the regular imaging devices, such as 640×480 or 320×240. In order to make the representative points equally distributed in spatial, the ratio of row and column should be maintained by as close to 3:4 as possible.

Then the correlation calculation of RPM with respect to representative point (X_r, Y_r) is performed as

$$R_i(p,q) = \sum_{r=1}^{N} \left| I(t-1, X_r, Y_r) - I(t, X_{r+p}, Y_{r+q}) \right|, \tag{2.6}$$

where N is the number of representative points in one region, $I(t-1, X_r, Y_r)$ is the intensity

of the representative point (X_r, Y_r) at frame$(t-1)$, and $R_i(p,q)$ is the correlation measure for a shift (p,q) between the representative points in region i at frame$(t-1)$ and the relative shifting points at frame(t). Assuming R_{iMin} is the minimum correlation value in region i, i.e. $R_{iMin} = \underset{p,q}{Min}(R_i(p,q))$, the shift vector v_i that produces the minimum correlation value for region *i* represents the local motion vector of this region, i.e.,

$$v_i = (p,q), \quad \text{for } R_i(p,q) = R_{iMin}. \tag{2.7}$$

2.2.3. Irregular Condition Detection

By analyzing the curves of correlation values corresponding to image sequences with various conditions, it is found that the curve of correlation values is related to the reliability of motion detection. Figs. 2.4 and 2.5 show the various correlation curves corresponding to different video images with different conditions. Fig. 2.4 is a video image captured by a hand-held camcorder of a daily life scene. Fig. 2.5 is a video image captured by an in-car video camera of an outside scene. It is found that the irregular conditions detected in daily life scenes can be detected in in-car applications as well. Fig. 2.4 (a) and 2.5 (a) show a normal condition where the peak of correlation is obvious in each region. In Fig. 2.4 (b) and 2.5 (b), the curves look like a valley. This means only one dimension of correlation data is reliable and it lacks of feature in the other direction. Fig 2.4 (b) is short of features in the x direction. On the contrary, Fig 2.5 (b) is short of features in the y direction. Fig. 2.4 (c) and Fig. 2.5 (c) show examples with repeated patterns. Fig. 2.4 (c) is an image of an office partition filled with grid-hole patterns, especially on region 2 and 4, which causes multiple peaks on these regions. Fig. 2.5 (c) is an image of a brick wall with a fence in the bottom area, and it also causes multiple peaks in the correlation curves, especially within region 1 due to the pure bricks that are repeated in this area. Fig. 2.4 (d) represents the condition of a moving object; a bear moves from the left side to the right side of the image sequence. It causes double peaks in the curve of region 2 and the value of R_{iMin} is larger than the value of a normal image, such as Fig. 2.4 (a). Fig. 2.5 (d) represents an image sequence of a motorcycle moving from the right side to the left. It causes double peaks within region 1 of the curve and the value of R_{iMin} is larger than that of a normal image as well. The example shown in Fig. 2.4 (e)

contains a white wall on the right side of the image. Obviously, it is very hard to distinguish the peak from the correlation curve in region 4. Fig. 2.5 (e) contains a large low-contrast area (blue sky) on the upper left corner of the image. It is also hard to distinguish the peak in region 1.

Although the curve of correlation values is related to the reliability of motion detection, it is still too complex to directly use these curves to evaluate the reliability of motion detection. Therefore, we propose a strategy that combines the minimum projections of correlation curve in *x* and *y* directions (minimum projections) and the inverse triangle method to detect the irregular conditions from each region. The mathematical expression of minimum projections can be written as:

$$\begin{aligned} x_i\_\min(p) &= \min_q R_i(p,q) \\ y_i\_\min(q) &= \min_p R_i(p,q) \end{aligned}, \tag{2.8}$$

where $x_i\_\min(p)$ and $y_i\_\min(p)$ are the minimum projections of correlation curve in *x* and *y* directions in region i, respectively. Fig. 2.6 shows the examples of minimum projections of the correlation curve in *x* and *y* directions. Fig. 2.6(a) is the minimum projection of Fig. 2.5(a), that is regular and the determination of motion vector in each region is clear and consistent. Fig. 2.6(b) is the minimum projection of Fig. 2.5(b), which shows a lack of features in *y* direction (horizontal). The values of the minimum projection of correlation curve in y direction are within a small range and erratic with multiple peaks so that the determination of the minimum value is very hard.

In order to determine the reliability of motion vector easily, the feature extraction of reliability is performed by the proposed inverse triangle method through the minimum projections in *x* and *y* directions to obtain the reliability indices. Fig. 2.7 shows the illustration of the inverse triangle method. In the first step, we find $T_i\_\min$ that represents the global minimum of the minimum projection curve in region i and can be calculated by Eq. (2.9).

In the second step, we calculate S_{xi} and S_{yi} by Eq. (2.10), where *offset* is the altitude of the inverse triangle. n_{xi} and n_{yi} are defined as the numbers of S_{xi} and S_{yi}, respectively (see Eq. (2.11)). d_{xi} and d_{yi} are defined as the distances of two vertexes of the base of the inverse triangle obtained by Eq. (2.12). The cost level of *x* and *y* directions are calculated by Eq. (2.13). The higher cost level means a lower confidence level. Since the condition of multiple peaks seriously degrades and affects the determination of reliability, the penalty of multiple peaks is taken into account by Eq. (2.13) to improve the discrimination of

reliability. The example shown in Fig. 2.7 is a curve with twin peaks which will get the penalty of $d_{xi} - n_{xi}$. In the third step, we determine the confidence indices of x_i and y_i in region i through a threshold denoted as TH. The lower cost level represents a higher reliability. In the final step, summing up the counts of reliable motion components of x and y in the four regions as Eq. (10), we get $Num(x_i)$ and $Num(y_i)$, $i = 1 \sim 4$.

Step 1. Find global minimum $T_i\_\min$ from $x_i\_\min(p)$ or $y_i\_\min(q)$

$$T_i\_\min = \min_p(x_i\_\min(p)) \text{ or } \min_q(y_i\_\min(q)) \tag{2.9}$$

Step 2. Calculate the cost level, x_i_cost and y_i_cost

$$\begin{cases} S_{xi} = \{p \mid x_i\_\min(p) < T_i\_\min + offset\} \\ S_{yi} = \{q \mid y_i\_\min(q) < T_i\_\min + offset\} \end{cases} \tag{2.10}$$

$$\begin{cases} n_{xi} = \text{number of } S_{xi} \\ n_{yi} = \text{number of } S_{yi} \end{cases} \tag{2.11}$$

$$\begin{cases} d_{xi} = \max_P S_{xi} - \min_P S_{xi} \\ d_{yi} = \max_q S_{yi} - \min_q S_{yi} \end{cases} \tag{2.12}$$

$$\begin{cases} x_i\_\text{cost} = 2d_{xi} - n_{xi} \\ y_i\_\text{cost} = 2d_{yi} - n_{yi} \end{cases} \tag{2.13}$$

Step 3. Set the threshold, TH , for determining the reliability indices

If $x_i\_\text{cost} < TH$ Then (2.14)

x_i is reliable

Else

x_i is unreliable

End if

. $y_i\_\text{cost} < TH$ Then

y_i is reliable

Else

y_i is unreliable

End if

Step 4. Calculate the numbers of x_i and y_i in four regions

$$\begin{cases} Num(x_i) = \text{sum of } (x_i \text{ is reliable}) \\ Num(y_i) = \text{sum of } (\text{y}_i \text{ is reliable}) \end{cases} \tag{2.15}$$

$i = 1 \sim 4$

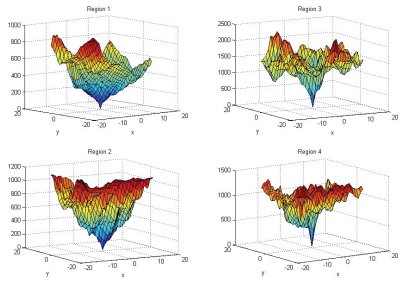
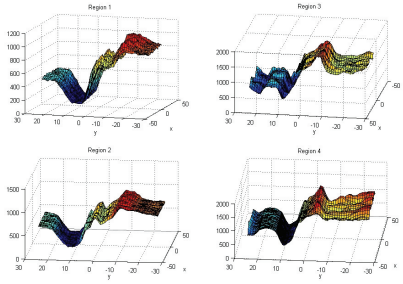
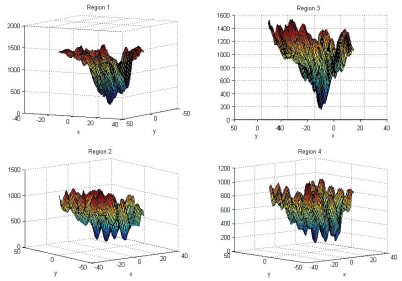
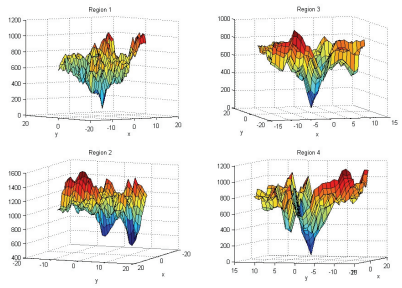
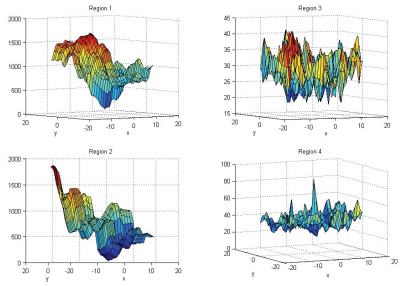
(a)		
(b)		
(c)		
(d)		
(e)		

Fig.2.4. Various correlation curves that correspond to image sequences with different conditions.(I) (a) A normal condition (b) Features lacking in the vertical direction (book) (c) Repeated patterns (office) (d) Moving object (bear) (e) Large low-contrast area (white wall) Video images captured by a camcorder in daily life scene.

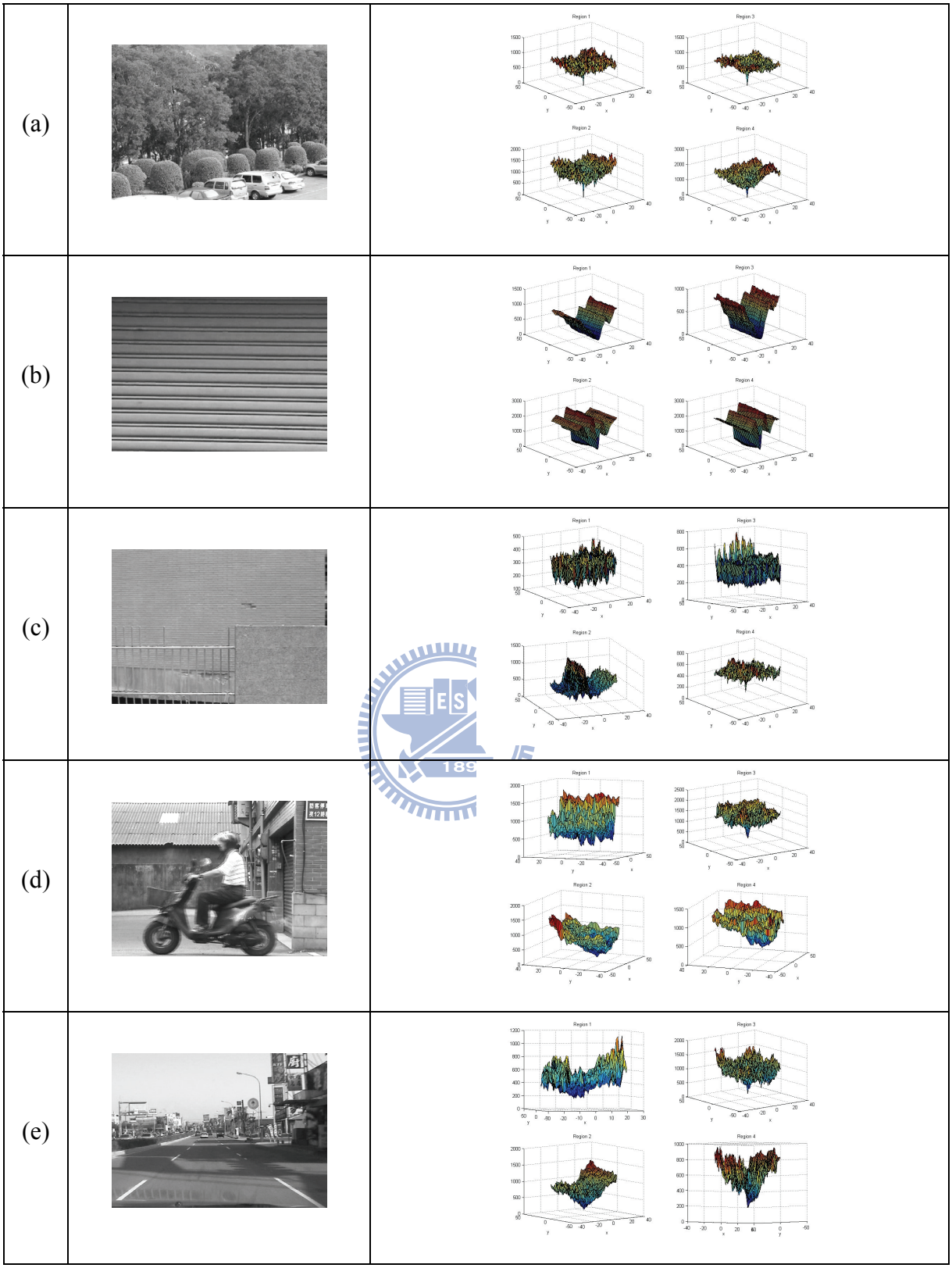
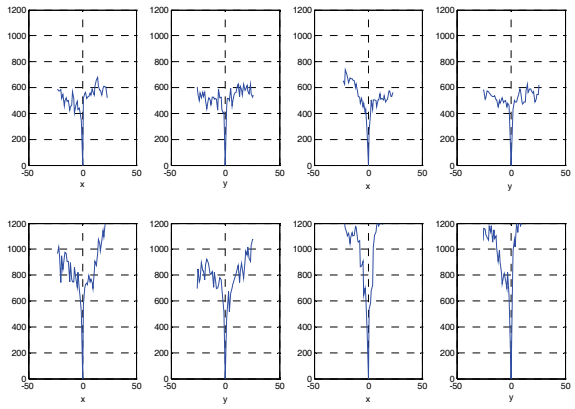
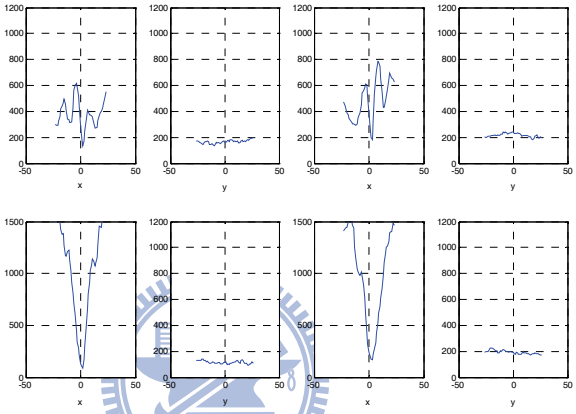


Fig.2.5. Various correlation curves that correspond to image sequences with different conditions (III). (a) A normal condition (b) Features lacking in the horizontal direction (gate) (c) Repeated patterns (brick) (e) Moving object (motorcycle) (f) Large low-contrast area (sky) Video images captured by an in-car video camera during outside scene.



(a)



(b)

Fig. 2.6. Examples of minimum projections of correlation curve from x and y directions in four regions. (a) Regular image sequence. (b) Ill-conditioned image sequence.

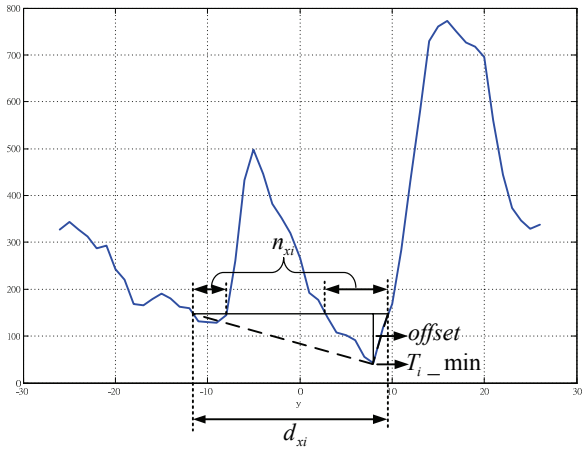


Fig. 2.7. Illustration of the proposed inverse triangle method

2.2.4. Determination of Representative Point Amount

Determining the representative point amount is quite an issue when using the RPM method. How many points are necessary to determine the local motion vector? Trial and error is one approach to get a better number, but it is neither efficient nor scientific. The reliability measure by inverse triangle method can be used for this issue. The cost value evaluated by Eq. (2.13) is an index of reliability. The higher cost level indicates a lower reliability. According to our experimental results, a threshold of cost value set at 18 can get the reliable motion vectors from our experimental video sequences. Fig. 2.8 shows the experimental result of calculating the cost level by using a different number of representative points. This is the averaged testing result for the experimental video sequences used in chapter 4. It can be found that if the number of the representative points is larger than 30, the cost level will get down to the threshold and almost all the motion vectors calculated by the RPM method are reliable. In other words, in this case, the cost level will be good enough at the lower value in regards to higher reliability. In order to keep low computation time complexity, 30 representative points are used in our system.

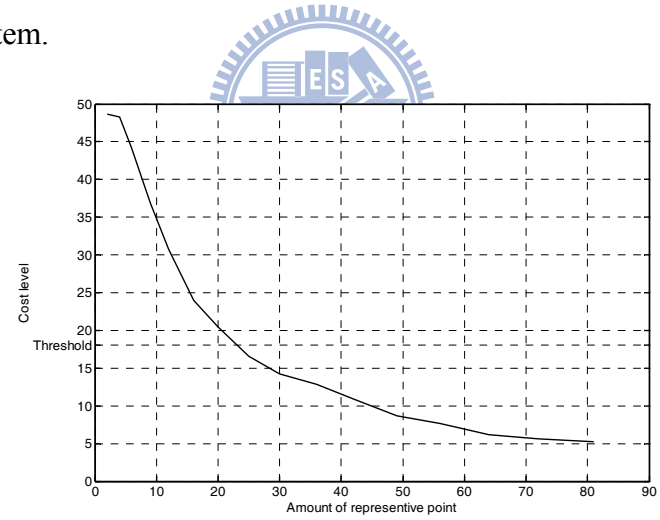


Fig. 2.8. The experimental result of calculating the cost level (that is an index of reliability defined in Eq. (2.13)) by using a different number of representative points.

2.2.5. Optimization of Representative Points

In the fixed-type video surveillance system, the DIS process can be divided into two sub processes: the initialization process and the compensation process. The main characteristics are: (1) the background is stationary; (2) the fluctuation of an image is caused by wind disturbance or object vibration such as a bird jumping or an earthquake; (3) the car moving or

people walking can cause partial local motion in the image, which will disturb the global motion detection. The initialization process can optimally select the representative points from the background image by the difference measurement. It can increase the reliability of motion vector detection.

The optimization process consists of difference measurement and optimization of representative points. First, the difference measurement derives a difference matrix for each selected representative point. Second, the optimization process chooses the optimal representative point among those points in the corresponding sub-region.

The difference measurement is based on the concept of the aforementioned RPM method. Each representative point is selected from a point of the corresponding sub-region, which can be expressed as (X_l^r, Y_l^r), where l is the index of representative points in each sub-region, r is the index of the sub-region. The difference matrix of each representative point can be written as

$$\begin{aligned} &DIF_l^r(p,q) = \left| I(X_l^r, Y_l^r) - I(X_l^r + p, Y_l^r + q) \right|, \\ &-\frac{1}{2}\text{total rows} \le p \le \frac{1}{2}\text{total rows of sub-region}, \\ &-\frac{1}{2}\text{total columns} \le q \le \frac{1}{2}\text{total columns of sub-region}. \end{aligned} \tag{2.16}$$

where $I(t, X_l^r, Y_l^r)$ is the intensity of the representative point $(X_{l_R}^r, Y_{l_R}^r)$, and $DIF_l^r(p,q)$ is the difference matrix corresponding to representative point $(X_{l_R}^r, Y_{l_R}^r)$.

The difference matrix expresses the degree of similarity between the representative point and the neighborhood. The mean value of difference matrix can globally represent the similarity between the representative point and the neighborhood, which is

$$M_l^r = mean(DIF_l^r(p,q)), \tag{2.17}$$

where l is the index of the representative point in each sub-region, r is the index of the sub-region. The largest M_l^r means the representative point is less similar to its neighborhood, i.e. the best candidate for the representative point, which can be expressed as

$$i_R^r = \underset{l}{Max}(M_l^r), \tag{2.18}$$

where l_R^r is the index of the optimal representative point in the sub-region r. If the values of M_l^r in sub-region r are smaller than a pre-defined threshold, the representative point will be neglected due to the lesser difference between itself and the neighborhood. Fig. 2.9(a) shows the background of the video surveillance image captured from the road scene. Fig. 2.9(b) shows the locations of traditional representative points, which are usually located within the

centers of the corresponding sub-regions respectively. The horizontal and vertical lines are formed to separate the sub-regions for illustration. Fig. 2.9(c) shows the locations of the optimal representative points. Each point is selected by the most significant point (highest M_l^r) in the sub-region. The optimal representative points can increase the confidence index through the cost level calculation by Eq. (2.13). After neglecting the insignificant points by that M_l^r are smaller than pre-defined threshold, the remaining representative points are shown in Fig. 2.9(d). The neglecting insignificant points can reduce the computation complexity without affecting the confidence index of motion estimation.

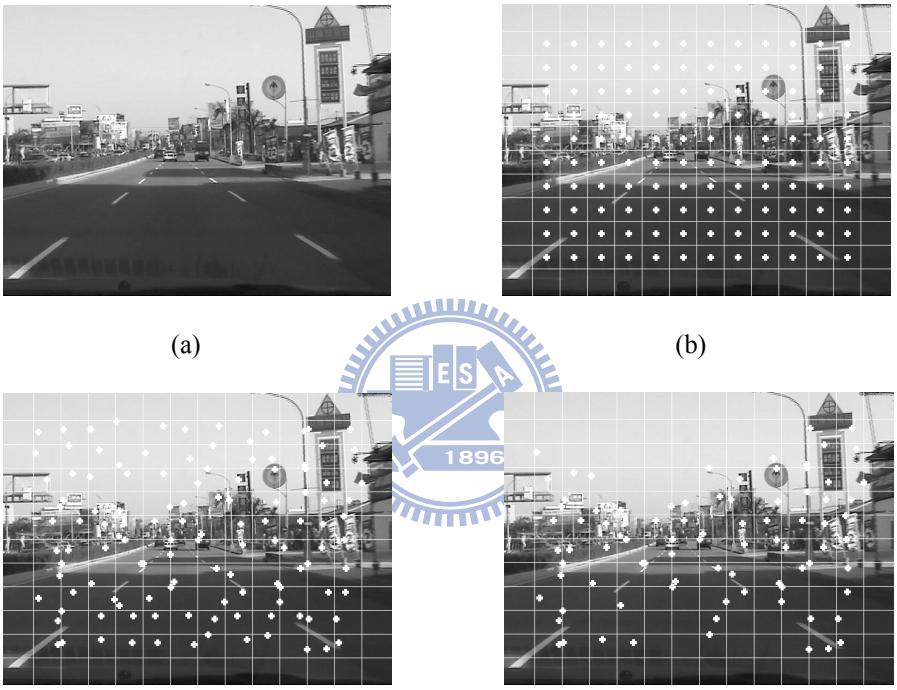


Fig. 2.9. The illustration of optimal representative point selection method. (a) The original image. (b) The traditional representative points. (c) The optimal representative points. (d) The optimal representative points after neglecting insignificant points.

2.2.6. Generation of Refined Motion Vector

Irregular motion vectors can be detected and excluded by using the minimum projection and inverse triangle method; however, image sequences with ill conditions, such as a lack of features, a large low-contrast area, a moving object or a repeated pattern, may contain fewer available MVs (most of the MVs are irregular) in the four regions. Therefore, a recombination of these available components of regular MVs is necessary to form a refined motion vector

(RMV). To solve this problem, a median function is used to extract a motion vector with respect to each direction for the ill condition. The calculation to determine the RMV is described as follows in detail.

Case 1. If $Num(x_i(t)) = 4$ then

$$V_{refined\_x}(t) = Med(V_{a\_x}(t), V_{b\_x}(t), V_{c\_x}(t), V_{d\_x}(t), GMV_x(t-1))$$

Case 2. If $Num(x_i(t)) = 3$ then

$$V_{refined\_x}(t) = Med(V_{a\_x}(t), V_{b\_x}(t), V_{c\_x}(t))$$

Case 3. If $Num(x_i(t)) = 2$ then

$$V_{refined\_x}(t) = Med(V_{a\_x}(t), V_{b\_x}(t), GMV_x(t-1)) \tag{2.19}$$

Case 4. If $Num(x_i(t)) = 1$ then

$$V_{refined\_x}(t) = V_{a\_x}(t)$$

Case 5. If $Num(x_i(t)) = 0$ then

$$V_{refined\_x}(t) = \gamma \times GMV_{avgx}(t-1)$$

where $Num(x_i(t))$ is the number of x components of reliable LMVs, $V_{refined\_x}(t)$ is the x component of RMV, $V_{a\_x}(t)$, $V_{b\_x}(t)$, $V_{c\_x}(t)$, and $V_{d\_x}(t)$ represent x components of reliable LMVs in the different regions, respectively, $Med(\,)$ is the function of median operation, $GMV_x(t-1)$ is the x component of the previous GMV, t is frame number, γ is attenuation coefficient, $0 < \gamma < 1$. The $GMV_{avgx}(t)$ can be calculated by

$$GMV_{avgx}(t) = \zeta GMV_{avgx}(t-1) + (1-\zeta)GMV_x(t), \quad 0<\zeta<1. \tag{2.20}$$

Then we apply the similar process to obtain $V_{refined\_y}(t)$. The resultant RMV is represented by

$$V_{refined}(t) = \begin{bmatrix} V_{refined\_x}(t) \\ V_{refined\_y}(t) \end{bmatrix}. \tag{2.21}$$

2.2.7. Global Motion Estimation

The objective of global motion vector estimation is to determine a motion vector from the existing data that has been evaluated from the motion estimation process. In a practical video sequence, it always suffers from moving objects, repeated patterns etc. The LMV in

each region may represent the global motion vector, the moving object motion vector, or even the error vector, respectively. The error vector may be caused by an ill condition, a repeated pattern, or the mixture of global and moving object motion. A reliable global motion vector is essentially selected from the LMVs and RMV. However, in the worst case, i.e. if the LMVs and RMV are all faulty, this will induce a worse result after compensation compared with the original images. Therefore, if the evaluation includes the zero motion vector (ZMV), it can prevent the occurrence of this case. Similarly, for an image sequence with constant motion in the scene, it will induce a worse result if it is compensated by ZMV or error motion vector rather than by the average motion vector (AMV). In the proposed DIS technique, the seven motion vectors, which include the four LMVs, the RMV, the ZMV, and the AMV, which are referred to as pre-selected motion vectors (pre_MV), are employed to estimate the GMV of the current frame. In general, one of LMVs is the highly probable GMV for the regular image. The RMV is the highly probable GMV for the ill-conditioned image. The ZMV can prevent a worse compensation result caused by the unreliable MVs, and the AMV is useful for the constant motion of the car. In addition, if the image sequence contains a large moving object, the determination of global motion is troublesome because the determined motion vector probably switches between the background and the large moving object or it is totally dominated by the large moving object. In this case, it will lead to artificial shaking and cause an important challenge in DIS.

2.2.7.1. Background Based Peer to Peer Evaluation

In this dissertation, a background-based evaluation function is proposed to overcome the large moving object problem. Fig. 2.10 shows the areas for the background-based evaluation. Five regions are selected to evaluate the result, which are located on the surrounding areas of the image. The reason is that, in most cases, the foreground object is located in the center of the image, so the surrounding areas of the image are the best candidates for background detection. The estimation of the GMV is calculated by the summation of absolute difference (SAD),

$$SAD_{B_i,c} = \sum_{X,Y \in B_i} \left| I(t-1,X,Y) - I(t,X+X_c,Y+Y_c) \right|, \quad 1 \le i \le 5, \quad 1 \le c \le 7, \tag{2.22}$$

where $I(t-1,X,Y)$ is the intensity of the point (X,Y) at frame *t-1*, B_i is the *i*-th background region in the image, X_c, Y_c are the components of the seven pre-select motion vectors (pre_MV_c) in *x* and *y* directions.

Each pre_MV_c can obtain it's $SAD_{B_i,c}$ in each region. The smaller $SAD_{B_i,c}$ represents the higher probability of the desired motion vector among theses pre-selected motion vectors. The score for each pre_MV_c in region i is denoted as $S_{i,c}$, which is the order of the $SAD_{B_i,c}$ value, and the higher $SAD_{B_i,c}$ indicates the higher score. The total score for each pre_MV_c can be obtained by

$$S_c = \sum_{i=1}^{5} S_{i,c} \ . \tag{2.23}$$

Five-region peer-to-peer evaluation can prevent the situation that some partial high-contrast image regions dominate the evaluation result. In this algorithm, each region has an equal priority to determine the result. In (14), S_c is the index to determine the GMV. The pre_MV_c with the smallest S_c is the desired GMV and it can be expressed as

$$\text{GMV}= pre\_MV_i \text{, for } i = \arg(\min_c S_c) \ . \tag{2.24}$$

According to these sophisticated evaluation areas, the evaluation function can detect the attributed background motion vector precisely in most circumstances.

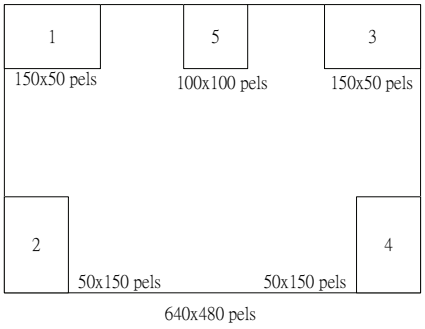


Fig. 2.10. Areas for background detection and evaluation.

2.2.7.2. Skyline Detection

To improve the robustness of the global motion vector estimation, the adaptive background-based evaluation function is proposed. Firstly, the skyline detection will be

performed. Then, five regions, based on the estimated skyline, are selected to evaluate the result. In most outdoor applications, such as in-car video capture, the pixels of the area above the skyline are low contrast. The skyline detection can prevent an invalid result due to some of the five regions located on the low-contrast area. Selecting the regions surrounding the boundary of the image to evaluate the obtained motion vector can avoid the disturbance of moving-object effects for global motion vector estimation. Fig. 2.11 shows the adopted areas for the adaptive background-based evaluation according to detecting the skyline. The proposed skyline detection algorithm combines RPM correlation evaluation, minimum projection, and the inverse triangle method. First, we calculate the absolute differences between the representative point at frame$(t-1)$ and the corresponding neighborhood pixels in the same sub-region at frame(t) by Eq. (2.25) that is regarded as the intermedium of Eq. (2.6),

$$C_{i,j}(p,q)=\left|I(t-1,X_i,Y_j)-I(t,X_{i+p},Y_{j+q})\right|, \tag{2.25}$$

where (i,j) denotes the position of one sub-region with respect to the row and column as shown in Fig. 2.12. There are 120 sub-regions (10 rows x 12 columns in this paper). (X_i,Y_j) is the coordinate of the representative point in the (i,j) sub-region, $I(t-1,X_i,Y_j)$ is the intensity of the representative point (X_i,Y_j) at frame$(t-1)$, and (p,q) is a shifting vector within the sub-region. Then we can derive the correlation curve for detecting the skyline by calculating

$$C_l(p,q)=\sum_{i=1}^{l}\sum_{j=1}^{M}C_{i,j}(p,q), \tag{2.26}$$

where M is the total number of sub-regions in the horizontal axis ($M=12$), l represents the l-th row of the sub-regions. Initially, $l=1$ and the minimum projection and inverse triangle method presented in Eq. (2.9) ~ (2.13) are applied to $C_l(p,q)$ to get the confidence index in the horizontal direction. The cost level is relatively high when the corresponding area is a low-contrast area such as the sky. If the level is lower than the presetting threshold then we stop the evaluation process and the horizontal position of the representative points of the sub-regions located in the last row of $C_l(p,q)$ is defined as the coarse skyline. Otherwise, we set $l=l+1$ and continue the evaluation of $C_l(p,q)$ till the level is lower than the presetting threshold. Fig. 2.13 shows the results of skyline detection in the video sequence taken from the highway. The coarse skyline is used to adaptively layout the background-based

evaluation blocks located on the higher contrast area. It improves the robustness of global motion vector estimation in image stabilization applications.

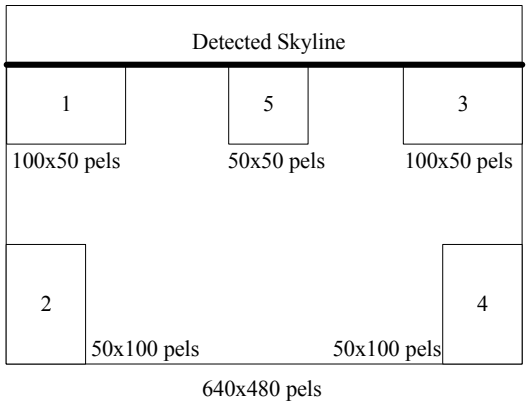


Fig. 2.11. Areas for the background-based evaluation adapted by the detected skyline.

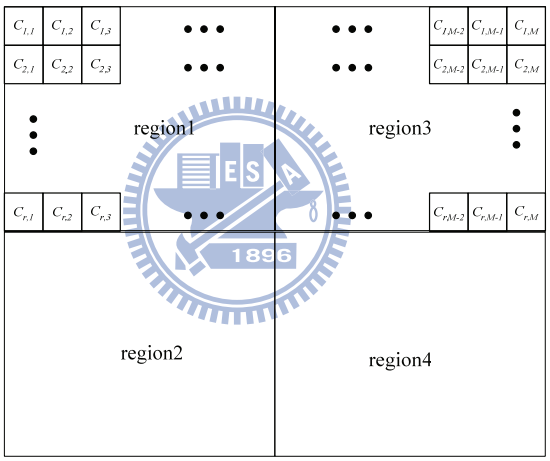


Fig. 2.12. Skyline detection algorithm is to combine RPM correlation evaluation, minimum projection and inverse triangle method.

(a) (b)

Fig. 2.13. Skyline detection applies on the in-car video sequence taken from the highway.

2.3. Summary

In this chapter we address the related research works of motion estimation and the advantages/disadvantages of various algorithms in more detail. To detect irregular conditions, we proposed an inverse triangle method to measure the reliability of evaluated motion vectors according to the correlation curve. The inverse triangle method evaluates the cost value in x and y directions for each motion vector, respectively, and then uses the cost value of each component of the motion vector to generate the refined motion vector. It can positively contribute to determinate the global motion vector in ill-conditioned images that have a lack of features, repeated pattern etc.

The determination of the representative point amount is also addressed by using the inverse triangle method. Based on this result, the representative point amount is set on 30 for each estimation region. In the fixed-type video surveillance system, the optimization of selecting the representative points from the background image by the difference measurement can increase the reliability of motion vector detection. The optimization process consists of the difference measurement and the optimization of the representative points. First, the difference measurement derives a difference matrix for each selected representative point. Second, the optimization process chooses the optimal representative point among those points in the corresponding sub-region.

Skyline detection and background based peer to peer evaluation have been proposed to improve the robustness of the global motion estimation. The background based peer to peer evaluation solves the large moving object problem and reduces artificial shaking in DIS. The skyline detection uses the coarse skyline to adjust the background-based evaluation area. It can prevent an invalid result due to some of the five regions located in the low-contrast area.

The DIS system combines with the aforementioned motion estimation approaches and can adapt to various conditions such as daily, in-car and surveillance applications.

3. Motion Compensation

3.1. Introduction

Motion compensation used in video compression is a technique that describes a picture in terms of translated copies of portions of a reference picture, often 8x8 or 16x16-pixel blocks. The objective is to increase compression ratios by making better use of redundant information between successive frames [54]. But motion compensation used in DIS is a technique for generating the compensating motion vector. This will compensate and shift the current picking window, accordingly, to obtain a smoother image sequence. Most of the research in the field of digital images stabilization is concentrated in the motion estimation part of the system. But motion compensation still plays a decisive role in the DIS system, especially in practical applications. Accumulated motion vector estimation and frame position smoothing (FPS) are the two most popular approaches to generate the compensating motion vector. J. K. Paik proposed the accumulated motion vector method in 1992. The computation procedure is given by [7]:

$$CMV(t) = k(CMV(t-1)) + (\alpha GMV(t) + (1-\alpha)GMV(t-1)) \quad (3.1)$$

where t represents the frame number, $0 < k < 1$ and $0 \le \alpha \le 1$. The constant k is used for smooth panning. The constant α is used to filter out the unexpected noise effect. The increase in k causes the decrease in unwanted shaking effect but the increase in the value of CMV leads more allowance to compensate. It means that, in the definite allowance, it may not fully stabilize the image sequence by the CMV, especially in a high panning condition.

The accumulated motion vector estimation needs to compromise stabilization and intentional panning (constant motion) preservation since the panning condition causes a steady-state lag in the motion trajectory [20]. The FPS accomplished the smooth reconstruction of an actual long-term camera motion by filtering out jitter components based on the concept of designing the filter with appropriated cut-off frequency. The disadvantage of FPS is that it does not guarantee the availability of the determined compensating motion vector when the specified-bound is restricted for preserving the effective image area in the DIS applications.

3.2. Compensating Motion Estimation

3.2.1. Motion Trajectory

To easily illustrate CMV estimation afterwards, the motion trajectories will be introduced first. These can then be calculated to analyze the problem that will be addressed in 3.2.2. The motion trajectories can be obtained by

$$MTraj_o(t) = \sum_{i=1}^{t} GMV(i), \tag{3.2}$$

$$MTraj_c(t) = \left(\sum_{i=1}^{t} GMV(i) \right) - CMV(t), \tag{3.3}$$

where $MTraj_o(t)$ and $MTraj_c(t)$ are the original and the compensated motion trajectories of the image sequence at frame(t).

Fig. 3.1 shows trajectories of motion generated by original global motion vectors and compensated global motion vectors. There are two trajectories in the figure. One is the original trajectory calculated by Eq. (3.2) and the other one is the compensated trajectory calculated by Eq. (3.3). The explanation of trajectories in the figure will be addressed in 3.2.2.

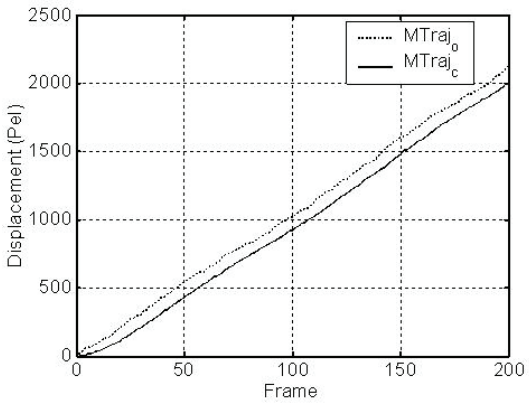


Fig. 3.1. Trajectories of motion generated by original global motion vectors and compensated global motion vectors. The compensated motion trajectory generation method in (3.1)

3.2.2. Quantitative Evaluation

The shaking effect of images can be evaluated by the summation of absolute differences of momentums within every two consecutive frames. The mass of an image can be set as a constant, such as one, for simplicity, or a value from zero to one, according to the degree of shaking in the images measured by human visual perception. The smoothness index (SI) is proposed to quantitatively evaluate the performance of different DIS algorithms and it is defined as

$$SI = \frac{1}{N-1}\sum_{t=2}^{N}\Delta m(t) = \frac{1}{N-1}\sum_{t=2}^{N} m \times \left| GMV(t) - GMV(t-1) \right|, \tag{3.4}$$

where t is the frame number, N is the number of total frames, m is the mass of the image, and $\Delta m(t)$ is the change rate of the absolute value of momentum. The lower SI means less shaking components in the image sequence and it represents a smoother effect.

3.2.3. Compensation with Inner Feedback-Loop Integrator

The compensating motion vector (CMV) estimation is used to generate CMVs for removing the undesired shaking motion, but still keep the steady motion in the image sequence. The estimation follows the conventional compensating motion vector estimation by (3.1) as $CMV(t) = k(CMV(t-1)) + (\alpha GMV(t) + (1-\alpha)GMV(t-1))$.

The CMVs in Fig. 3.1 are generated by the conventional method shown in Eq. (3.1). Obviously, $MTraj_c(t)$ has tremendous lag compared to $MTraj_o(t)$ due to the constant motion effect. Fig. 3.1 and 3.2 show the performance comparison of three different CMV generation methods applied to a video sequence with constant motion and jitter in the image. There are two trajectories in each subfigure; one is the original trajectory calculated by Eq. (3.2) and the other one is the compensated trajectory calculated by Eq. (3.3). The CMVs in Fig. 3.1 are generated by the conventional method shown in Eq. (3.1). Obviously, $MTraj_c(t)$ has tremendous lag compared to $MTraj_o(t)$ due to the constant motion effect. The CMV probably exceeds the window shifting allowance such that the available effective image area during the compensation process is reduced. The CMVs in Fig. 3.2(a) are generated by Eq. (3.1) with clipper function as

$$CMV(t) = clipper(CMV(t)) = \frac{1}{2}\left(\left| CMV(t) + l \right| - \left| CMV(t) - l \right| \right), \tag{3.5}$$

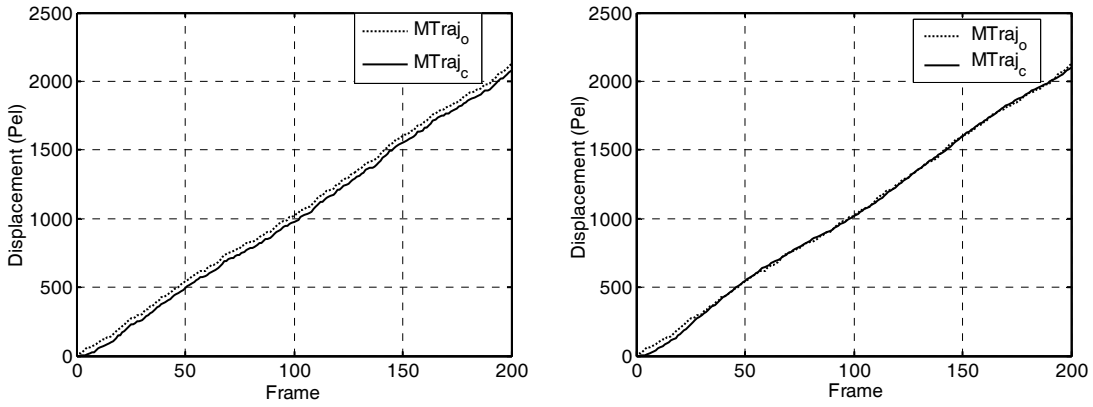
where l is boundary limitation, i.e. the maximum window shift allowance. In this case, the lag can be reduced to a certain range. However, it will also decrease the performance of shaking compensation due to the picking window operating near the boundary area.

In order to deal with the above problem, Vella, et al. used the passive method of ceasing to correct in this condition [9]. That implied that the undesired shaking effect could not be eliminated in the constant motion condition. To attack this drawback, we combine the inner feedback-loop integrator with a clipper function to reduce the steady-state lag for steady motion. This also keeps the CMV operating in the appropriate range. Fig. 3.3 shows the block diagram of the proposed CMV generation method. There is an integrator in the inner feedback loop, which can eliminate the steady-state lag of the CMV in the constant motion condition. That means, by employing the integrator, the shaking components of the images with constant motion effect as well as those in regular images can be stabilized. It is noted that the CMV computation procedure is applied to x and y components separately. That is, parameters corresponding to the x and y directions can be set as different values. In general, the constant motion condition usually occurs in the horizontal direction such that the shaking patterns are different in both directions. The proposed CMV computation procedure is presented by

$$\begin{aligned} CMV(t) &= \mathbf{k} \bullet CMV(t-1) + GMV(t) - \boldsymbol{\beta} \bullet CMV\_I(t-1) \\ CMV\_I(t) &= CMV\_I(t-1) + CMV(t) \\ CMV(t) &= clipper(CMV(t)) \end{aligned} \quad , \tag{3.6}$$

where $\begin{bmatrix} 0 \\ 0 \end{bmatrix} \le \mathbf{k}, \boldsymbol{\beta} \le \begin{bmatrix} 1 \\ 1 \end{bmatrix}$, $\bullet$ denotes array multiplication, and $clipper(\ )$ is defined in Eq. (3.5).

Fig. 3.2(b) shows the compensated motion trajectory generated by the proposed method. Compared with Figs. 3.1 and 3.2(a), the proposed method can reduce the steady-state lag of the compensated motion trajectory in the constant motion condition and keep the CMVs in an appropriate range.



(a) (b)

Fig. 3.2. The original and the compensated motion trajectories. (a) CMV generation method in (3.1) with clipper in (3.5). (b)Proposed method in (3.6).

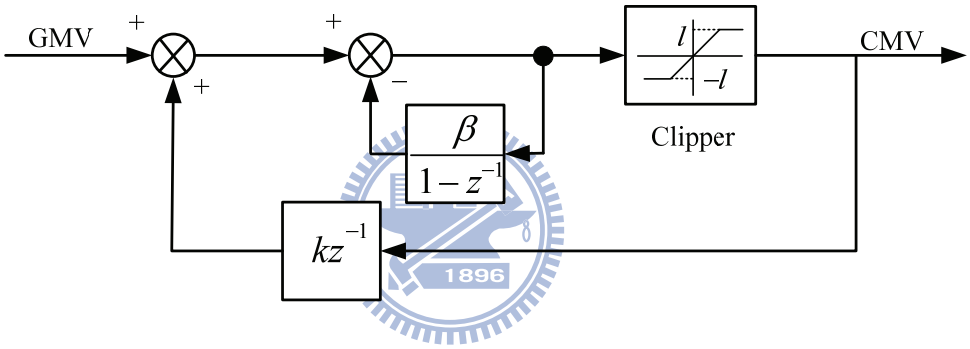


Fig. 3.3. Block diagram of the proposed CMV generation method.

3.2.4. Fuzzy Inference Digital Image Stabilization (FIDIS)

The objective of motion compensation is to achieve the optimal stabilization of shaking images within a specified compensation range (window shifting allowance). Eq. 3.1 can get a good result in a general case. But in the constant motion case, the undesired shaking effect cannot be eliminated by this method (mentioned in 3.2.3). The method proposed in 3.2.3 can overcome this drawback. The inner feedback-loop integrator was combined with a clipper function to reduce the steady-state lag for steady motion, as well as to keep the CMV operating in the appropriate range. Due to the characteristics of the integral, the overshoot will occur in certain back and forth low frequency image panning. Fig. 3.4 shows overshoot in low frequency back and forth image panning. The motion trajectory has obvious overshoot. This overshoot will reduce the effect of stabilization.

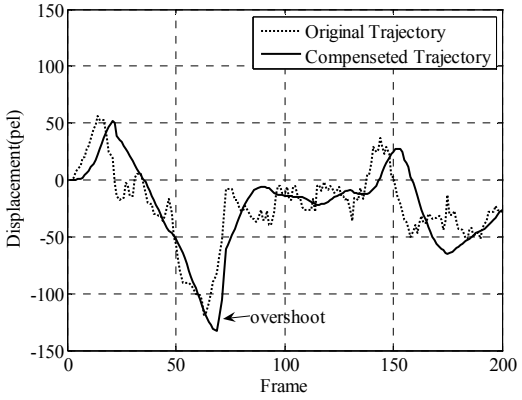


Fig. 3.4. Overshoot in low frequency back and forth image panning.

In this section, the architecture of fuzzy inference applied to DIS will be proposed. The idea is to keep the merits of each motion compensation method, i.e. reduce compensated motion trajectory lag and overshoot, to improve the DIS in different conditions. Fig. 3.5 shows the architecture of fuzzy inference applied to digital image stabilization. The system architecture includes two processing units, the motion estimation unit and the motion compensation unit. The motion estimation unit is the same as 2.2.1. The motion compensation unit includes two motion compensation methods, fuzzy inference system and motion compensation. Due to the differences of environments, conditions and shaking patterns of the captured images, the different motion compensation methods have individual merits separately. Fig. 3.6 shows the architecture of the fuzzy inference system. The pre-process block inputs GMV, CMV#1, CMV#2 are to transform into delta of short term smoothness index($\Delta SI(n)$) and delta of deviation(Δd) which are inputs of fuzzy inference. The fuzzy inference determines the weighted value to calculate the compensation vector. And then the block of image compensation uses the compensation vector to stabilize the image sequence.

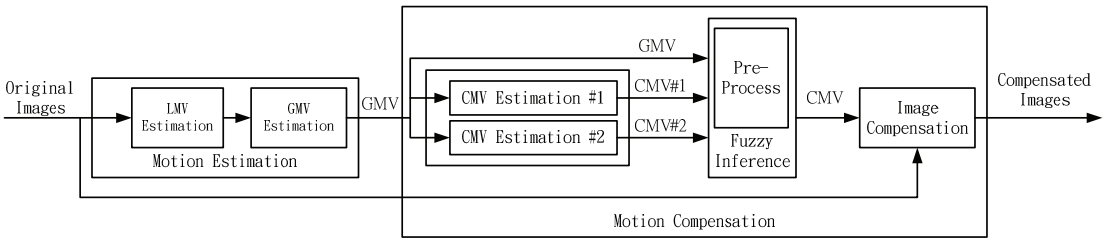


Fig. 3.5. Architecture of fuzzy inference applied to digital image stabilization.

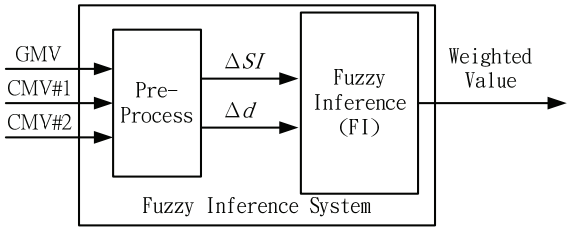


Fig. 3.6. Architecture of fuzzy inference system.

The motion compensation unit uses two motion compensation methods. These are the conventional CMV estimation mentioned in Eq. 3.1 and the proposed CMV estimation mentioned in Eq. 3.3. The results are denoted as CMV#1 and CMV#2, respectively.

The main function of pre-process is to transform external signals into distinguished signals of fuzzy inference input. The major consideration is the smoothness index. To realize the effect of the compensation method in a short term period, the short term smoothness index $SI(n)$ is defined as

$$SI(n) = \frac{1}{p+1}\sum_{i=n-p}^{n}\Delta m(i) = \frac{1}{p+1}\sum_{i=n-p}^{n} m \times \left|GMV(i) - GMV(i-1)\right| \tag{3.7}$$

where n is the frame number, p is the short term frame amount.

The delta of short term smoothness index($\Delta SI(n)$) can be expressed as

$$\Delta SI(n) = SI\#1(n) - SI\#2(n) \tag{3.8}$$

where $SI\#1(n)$ and $SI\#2(n)$ are short term smoothness indexes generated by Eq. 3.1 and 3.3 , respectively.

When $\Delta SI(n) < 0$, the motion trajectory compensated by CMV#1 is smoother than what was compensated by CMV#2 and vice versa. The secondary consideration is the deviation between the original and compensated motion trajectories. The deviation ($d(n)$) is defined as

$$d(n) = \left|MTraj_o(n) - MTraj_c(n)\right| \tag{3.9}$$

where $MTraj_o(n)$ is the original motion trajectory and $MTraj_c(n)$ is the compensated motion trajectory.

The large deviation between original and compensated motion trajectories will affect the image stabilization result. Therefore, the delta of deviation (Δd) is defined as a distinguished signal which is expressed as

$$\Delta d(n) = d\#1(n) - d\#2(n) \tag{3.10}$$

where $d\#1(n)$ and $d\#2(n)$ are deviations of motion trajectory compensated by CMV#1

and CMV#2. When $\Delta d(n) < 0$ means the deviations of motion trajectory compensated by CMV#1 is better than compensated by CMV#2 and vice versa.

Fig. 3.7 shows the Architecture of internal fuzzy inference. It contains a fuzzifier, a fuzzy rule base, an inference engine and a defuzzifier. The function of the fuzzifier is to transform crisp measured data $\Delta SI(n)$ and $\Delta d(n)$ into suitable linguistic values. The membership functions of these two signals are shown in Fig. 3.8. The linguistic descriptors are: large negative (FN), small negative (SN), zero (Z), small positive (SP), large positive (LP). Fuzzy rule base is characterized by a collection of fuzzy IF-THEN rules in which the preconditions and consequents involve linguistic variables. The collection of fuzzy inference rules characterizes the simple input-output relation of the system. The Eq. 3.8~3.18 are the statements of fuzzy rules:

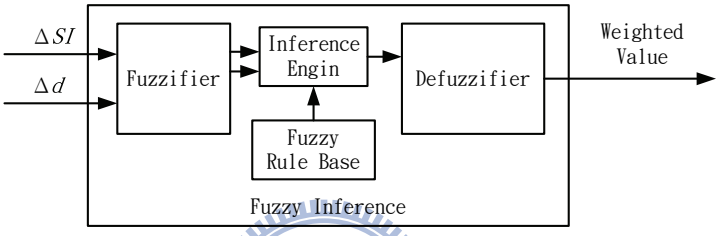


Fig. 3.7. Architecture of internal fuzzy inference.

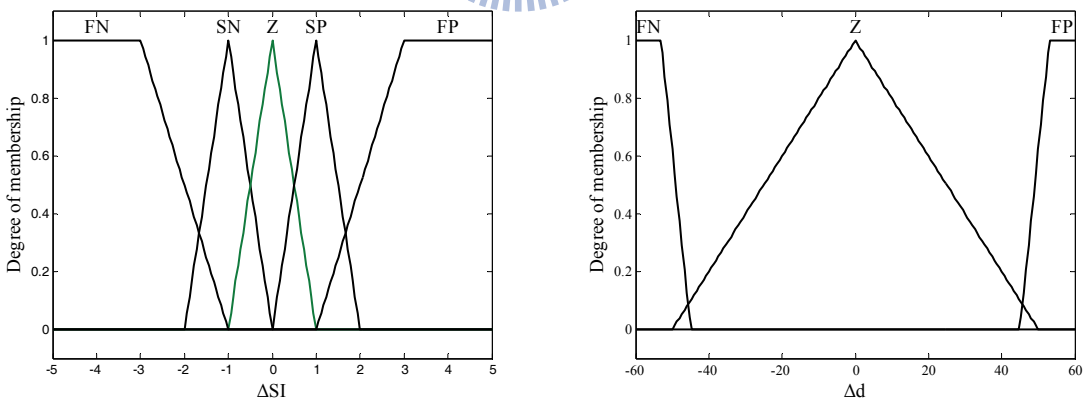


Fig. 3.8. Membership functions of inputs (a) $\Delta SI(n)$; (b) $\Delta d(n)$.

IF ΔSI is SN and Δd is FN THEN Output is L (3.11)

IF ΔSI is SN and Δd is Z THEN Output is L (3.12)

IF ΔSI is SN and Δd is FP THEN Output is M (3.13)

IF ΔSI is Z and Δd is FN THEN Output is L (3.14)

IF ΔSI is Z and Δd is Z THEN Output is M (3.15)

IF ΔSI is Z and Δd is FP THEN Output is S (3.16)

IF ΔSI is SP and Δd is FN THEN Output is M (3.17)

IF ΔSI is SP and Δd is Z THEN Output is S (3.18)

IF ΔSI is SP and Δd is FP THEN Output is S (3.19)

IF ΔSI is FN THEN Output is L (3.20)

IF ΔSI is FP THEN Output is S (3.21)

Mamdani's fuzzy implication method R_c, associated with the min operation, is used in the proposed inference system. The membership function of output signal is shown in Fig. 3.9. The linguistic descriptors are: large (L), Medium (M), small (S). The output represents the weighted value of motion compensation method#1. Because two methods are included in the system, the final motion compensation vector is calculated by:

$$\begin{aligned} CMV(n) &= k \times CMV(n-1) \\ &+ y(n) \times (CMV\#1(n) - CMV\#1(n-1)) \\ &+ (1-y(n)) \times (CMV\#2(n) - CMV\#2(n-1)) \end{aligned} \quad (3.22)$$

where y(n) is the weighted value, $0 < k < 1$, $CMV\#1(n)$ and $CMV\#2(n)$ are the compensated vectors estimated by method #1 and #2.

Defuzzifier is a mapping from a space of fuzzy set defined over an output universe of discourse into a space of nonfuzzy set. The method of defuzzification used in the fuzzy inference system is the center of area (COA).

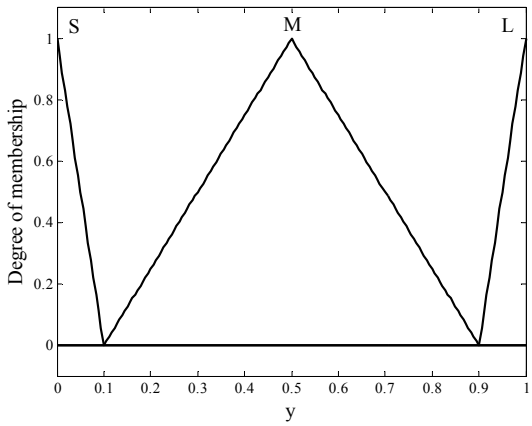


Fig. 3.9. Membership function of output.

3.3. Summary

In this chapter, we addressed the related research works of motion compensation. First, the motion trajectory of an image sequence was introduced to analyze the problem of motion compensation. From the motion trajectory, it is easy to view the smoothness of processing digital image stabilization and the deviation between the original image sequence and the compensated image sequence. Secondly, the quantitative evaluation method was introduced to evaluate the degree of image stabilization quantitatively. The smoothness index (SI) was proposed to quantitatively evaluate the performance of different DIS algorithms. Then two methods of motion compensation, the CMV estimation with inner feedback-loop integrator and fuzzy inference applied to motion compensation, were proposed. The CMV estimation with inner feedback-loop integrator solved the lag problem in the constant motion condition. The fuzzy inference DIS (FIDIS) was proposed to adaptively determine better motion compensation methods through two different MCs for various conditions of image sequences.

4. Experimental Results

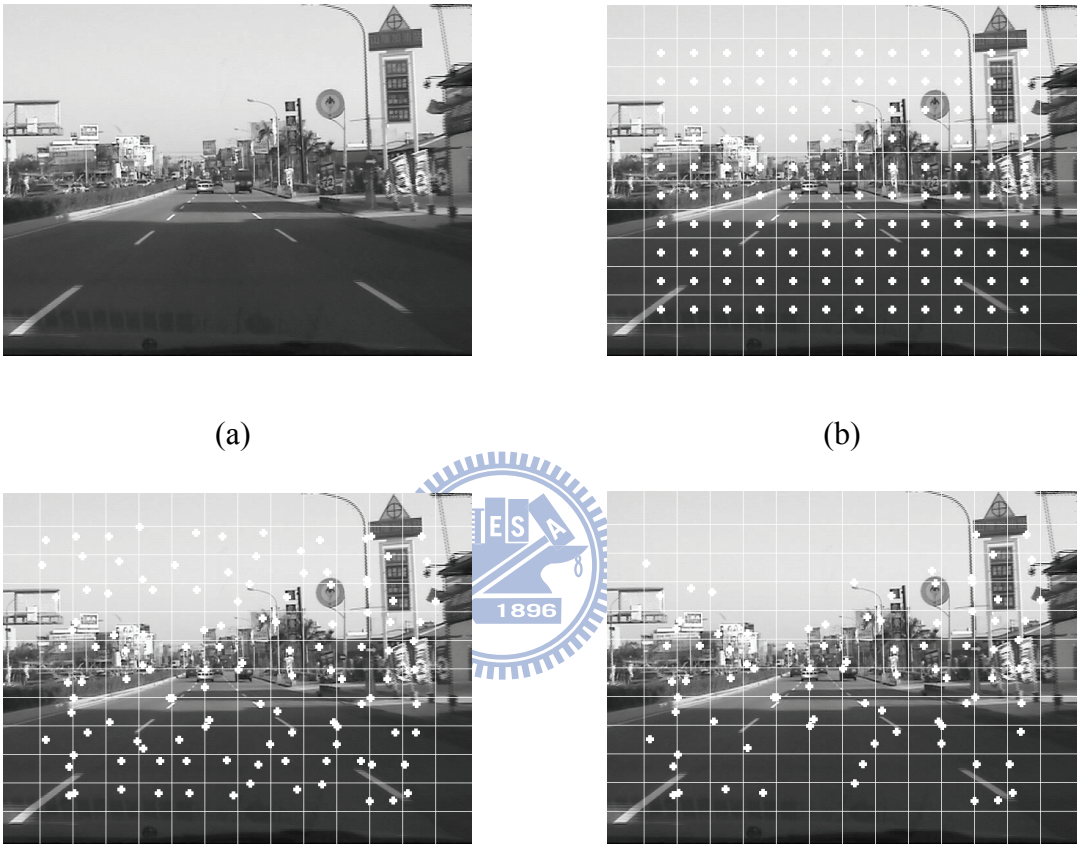
In this chapter, the experimental results of method with optimization of representative points (ORP) will be addressed first. Secondly, the performance of the proposed DIS technique is evaluated and compared to other existing DIS methods based on the performance indexes of motion estimation and motion smoothing, respectively. Finally, the experimental results of fuzzy inference applied to DIS (FIDIS) will be addressed.

4.1. Experiment Results of ORP

The experimental results of method with optimization of representative points will contain three parts for demonstration. An image with resolution of 640x480 is used for the first part of the testing. Fig. 4.1(a) shows the background of the video surveillance image captured from a road scene. Fig. 4.1(b) shows the locations of the traditional representative points, which are usually located within the centers of the corresponding sub-regions, respectively. The horizontal and vertical lines are formed to separate the sub-regions for illustration. Fig. 4.1(c) shows the locations of the optimal representative points. Each point is selected by the most significant point (highest M_l^r) in the sub-region. After neglecting the insignificant points, M_l^r, that are smaller than the pre-defined threshold, the remaining representative points are shown in Fig. 4.1(d).

The second part illustrates the generation of the local motion vectors. The local motion vectors are generated from the four divided regions of the image. Each local motion vector represents the corresponding local motion of the region. Fig. 4.2 shows the local motion vector in each region of the image. The bear is moving from the right to the left side and the background is stationary. Thus, only the motion vector appears in the region 4. Fig. 4.3 shows the combination of local and global motions. The bear is moving from the left to the right side and the background is moving from the bottom-left to the upper-right. Thus, regions 1, 2 and 3 contain only the global motion, but region 4 contains the combination of local and global motion.

The third part illustrates the experimented result of a real surveillance video sequence. Fig. 4.4 shows the image sequence captured when focusing the camera on the campus entrance gate. The passing in and out of cars or people are the moving objects disturbing the selection of the global motion vector, which plots the original and compensated motion trajectories of x and y axes. It can be found that the fluctuation of the image sequence can be mostly removed as the motion deviation is within the specified range.



(a) (b)

(c) (d)

Fig. 4.1. The illustration of optimal representative point selection method. (a) The original image. (b) The traditional representative points. (c) The optimal representative points. (d) The optimal representative points after neglecting insignificant points.

Fig. 4.2. The local motion vector of each region. Only region 4 appears as local motion due to the bear moving from the right to the left side.

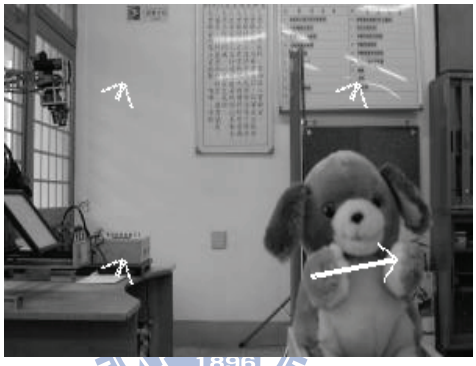
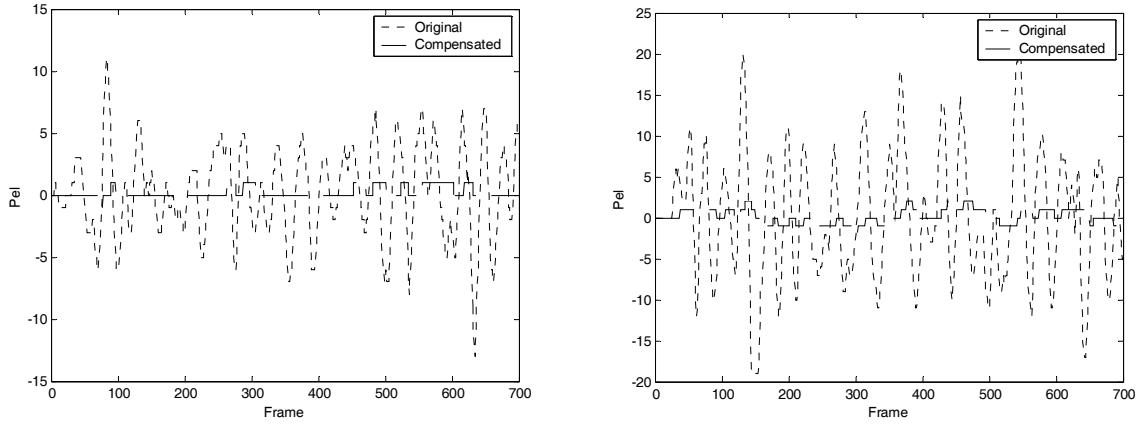


Fig. 4.3. The bear is moving from the left to the right side and the background is moved from the left-bottom to the right-up. The motion vector in regions 1, 2 and 3 contain the global motion. Region 4 contains the combination of the local and global motions.



(a) (b)

Fig. 4.4. The plots of original and compensated motion trajectories. (a) In x axis (b) In y axis.

4.2. Performance of Compensation with Inner Feedback-Loop Integrator

In this section, the performance of the proposed DIS technique is evaluated and compared to other existing DIS methods based on the performance indices of motion estimation and motion smoothing, respectively. To do this, eight real video sequences, captured by hand-held and in-car cameras with various irregular conditions, are used for testing. Each video sequence has a resolution of 640x480. VS#1 is a video taken of books on a bookshelf that has constant and intermittent panning in the horizontal direction. Obviously, there is a lack of features in the vertical direction. VS#2 is a video taken of a forest with constant panning, so it has the effects of hand shaking in both the horizontal and vertical directions. VS#3 is a video taken of a child, which contains a large moving object and displays the effects of hand shaking as well. VS#4 is a video taken of a car that has poor image quality and tremendous fluctuation. VS#5 is a video taken of a gate that has constant camera motion and jitter. In this sequence, there is a lack of features in the horizontal direction. VS#6 is a video taken of a community road during a bumpy condition. VS#7 is a video taken of a highway with jitter. VS#8 is a video taken of a car turning in a parking lot. The motion estimation performance is evaluated based on: the root mean square error (RMSE) between the algorithmically estimated motion vectors and the desired motion vectors evaluated by human visual perception as well as considering the background factor frame by frame. The RMSE is given by

$$RMSE = \sqrt{\frac{1}{N}\sum_{n=1}^{N}\left[(x_n - x_{dn})^2 + (y_n - y_{dn})^2\right]}, \tag{4.1}$$

where (x_{dn}, y_{dn}) is the desired motion vector and (x_n, y_n) is the motion vector generated from the evaluated DIS algorithms.

The proposed method is compared to a RPM approach with fuzzy set theory (RPM_FUZZY). The motion estimation results of these two methods are summarized in Table 4.1. The result with respect to VS#1 shows that the proposed method is superior to the RPM_FUZZY method since the proposed technique applies the minimum projection approach and the inverse triangle method to detect the irregular components of LMVs and then recombines available MVs to form an IMV. The result with respect to VS#3 shows that the GMV evaluated by the proposed background evaluation scheme can avoid the influence of large objects in motion. In VS#4, the higher RMSE indicates that some frames with

tremendous fluctuation in the image sequence are out of the MV detection range and include more rotation components as well. However, the proposed technique still performs better than the RPM_FUZZY method on this video sequence. VS#5 lacks for feature in the horizontal direction such that only one component of motion vector is reliable (see Fig. 2.5 (b)). The proposed method applies the minimum projection approach and the inverse triangle method to detect the irregular components of LMVs and then recombines available MVs to form an RMV. This approach can sufficiently use the existing information to estimate the global motion vector. The testing result with respect to VS#5 shows that the RMSE reduces from 5.8348 to 2.5269 by using our method since the RPM_FUZZY method did not consider the condition of lack of feature. The results with respect to VS#6~8 also show that RMSEs of our method are superior to the RPM_FUZZY method since the resultant GMV through the adaptive background-based evaluation can avoid the influence of large moving objects. The irregular components of motion vectors are also considered here as well. According to these experiments, the proposed technique is more robust than the RPM_FUZZY method is in dealing with video sequences with irregular conditions, such as a lack of features, large moving objects, and poor image quality.

The motion smoothing performance is evaluated by the smoothness index (SI) proposed in Section 3.2.2. Fig. 4.5(c) shows the original motion trajectory vs. the compensated motion trajectory generated by the proposed method. Compared with Figs. 4.5(a) and (b), the proposed method can reduce the steady-state lag of the compensated motion trajectory in constant motion condition and keep the CMVs in an appropriate range.

Table 4.2 shows the SI comparisons of three CMV generation methods presented in Fig. 4.5. The generation of CMV without clipper is impractical since it loses too much effective image area, i.e., the maximum of the CMVs does not guarantee to fit the practical compensation range. The proposed CMV generation method dramatically reduces the SI value from 5.6482 to 0.9346 compared with the CMV generation without the integrator. The reason is that the effect of the inner feedback-loop integrator greatly reduces the steady-state lag in the image sequence with constant motion.

Table 4.1.

RMSE comparisons of RPM_FUZZY and the proposed method with respect to eight real video sequences.

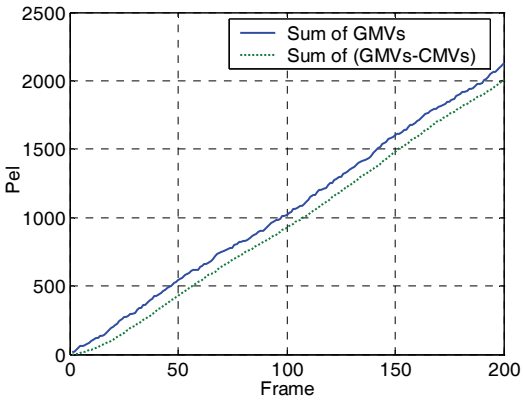
Method	Real video sequences							
	VS#1	VS#2	VS#3	VS#4	VS#5	VS#6	VS#7	VS#8
RPM_FUZZY	2.5729	2.4166	2.3958	6.0469	5.8348	0.8031	2.6618	2.2749
The proposed method	0.2449	0.7280	1.2369	2.5632	2.5269	0.3536	1.6837	0.5701

Table 4.2.

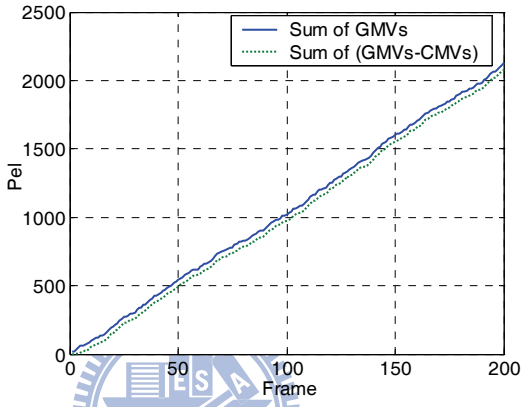
SI comparisons of three CMV generation methods.

Methods	SI	Max. CMV value (pels)
Eq. (3.1)	0.7990	134
Eq. (3.1) with clipper	5.6482	47
The proposed CMV generation method (Eq. (3.6))	0.9346	47

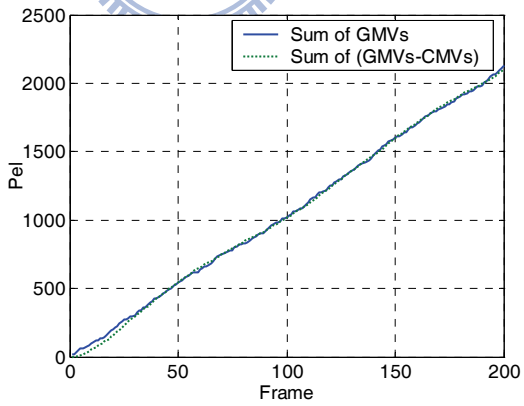
Note: The original SI is 7.4372. The clipper is bounded within ±47 pels.



(a)



(b)

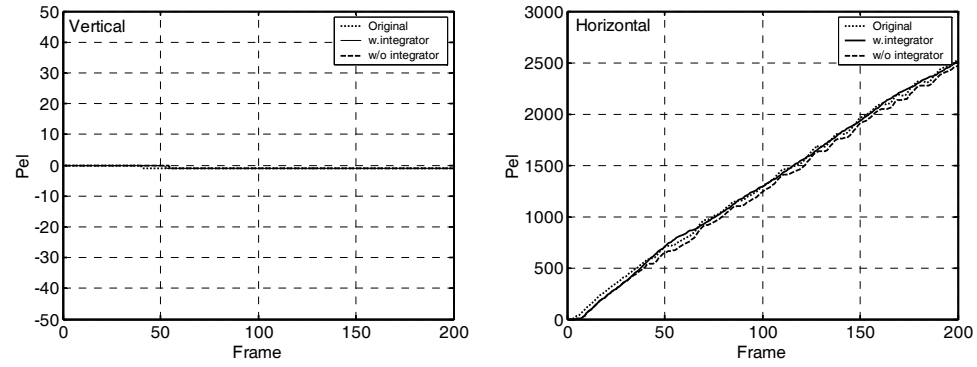


(c)

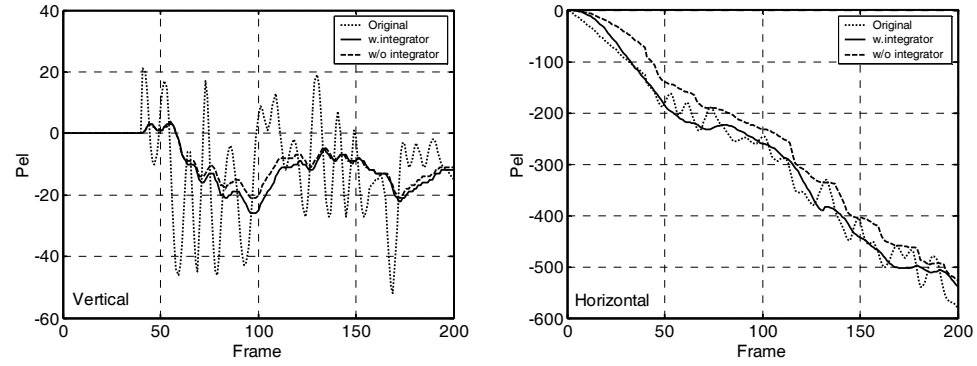
Fig. 4.5. Performance comparison of three different CMV generation methods applied to a video sequence with panning and hand shaking. (a) CMV generation method in Eq. (3.1). (b) CMV generation method in Eq. (3.1) with clipper in Eq. (3.5). (c) The proposed method in Eq. (3.6).

We also evaluate the CMV generation methods by four GMV sets generated from real video sequences (GMV sets #1~4). Fig. 4.6 shows the comparison of original and compensated motion trajectories by using two different CMV generation methods, Eq. (3.1) with clipper and Eq. (3.6), with respect to these four GMV sets. The parameter settings in Eq. (3.1) with clipper and Eq. (3.6) are listed in Table 4.3. The k is set as the same value for both horizontal and vertical directions. This means they have the same shaking absorption effects. The parameter β is the inner feedback-loop integral gain and it can determine the speed of the steady lag elimination during constant motion. The gain should not be too high to avoid resonance. In the in-car DIS applications, the constant motion occurs more frequently in the horizontal direction than in the vertical direction. Therefore, we set a higher gain for β in the horizontal direction to get a better visual quality.

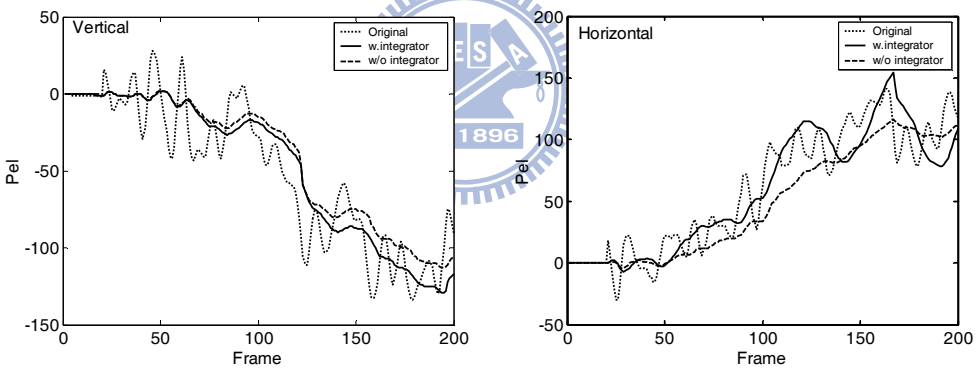
In each subfigure, the dotted line, solid line, and dashed line indicate the original trajectory and the compensated CMV trajectories are indicated by Eq. (3.6) and Eq. (3.1) with clipper, respectively. The GMV sets #1 and #2 (Fig. 4.6(a) and (b)) are estimated from video sequences with constant motion in the images. The GMV set#3 (Fig. 4.6(c)) is estimated from VS#3. The GMV set#4 (Fig. 4.6(d)) is estimated from VS#4. According to the results, the compensated horizontal motion trajectories of GMV set#1, set#2 and set#4, all of which have more constant motion in the images and were generated by the proposed CMV generation method, are, when compared to others, closer to the original horizontal motion trajectories. This means that the proposed method can reduce the steady-state lag and provides more space to absorb the shaking effect of image sequences without violating the physical range limitation. The GMV set #3 is estimated from the video captured in the highway. The result of the method with the integrator is that there is a slight overshooting phenomenon. This is evident when it is compared to the method without the integrator. This is the intrinsic property of adding an integrator in the process loop. But it is a good trade-off since it can greatly reduce the steady lag of motion trajectory. Table 4.4 shows the SI comparisons corresponding to Fig. 4.6. The original SIs can be regarded as the smoothness index of the original sequences with constant motion and undesired shaking components. In general, the proposed CMV generation method has better motion smoothing performance than the approach without the integrator on the compensation of most real video sequences with constant motion. The experimental results show that the proposed method can deal with various circumstances and has better performance in quantitative evaluations (such as RMSE and SI) and in human visual evaluation, too.



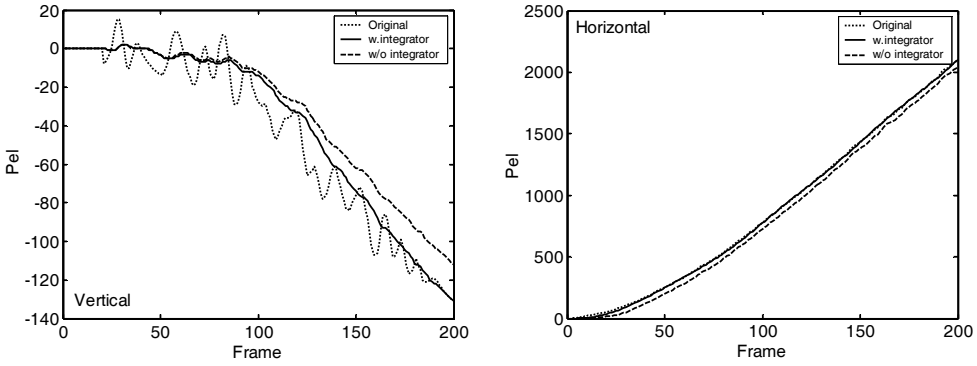
(a)



(b)



(c)



(d)

Fig. 4.6. Comparisons of original and compensated motion trajectories by two different CMV generation methods (with and without integrator) with respect to (a) GMV set #1, (b) GMV set #2, (c) GMV set #3, (d) GMV set #4.

Table 4.3.

The parameters applied to CMV generation with different equations

Method (Equation)	Parameters		
	k	β	Clipper Limit
18	$\begin{bmatrix}0.95\\0.95\end{bmatrix}$	$\begin{bmatrix}0\\0\end{bmatrix}$	±47pels (vertical) ±53pels(horizontal)
22	$\begin{bmatrix}0.95\\0.95\end{bmatrix}$	$\begin{bmatrix}0.001\\0.01\end{bmatrix}$	

Table 4.4.

SI comparisons of two different CMV generation methods with respect to four different GMV sets.

Video sequences	SI		Original SI
	Eq. (3.1) with clipper	Eq.(3.6)	
GMV set #1	2.8794	0.7237	3.2111
GMV set #2	1.3352	0.8153	4.2129
GMV set #3	0.8343	1.005	4.666
GMV set #4	1.3207	0.6891	2.1375

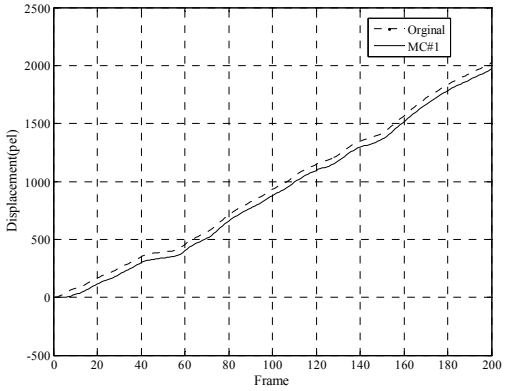
4.3. Experimental Results of FIDIS

In this section, the performance of FIDIS in various conditions is demonstrated. To do this, GMVs estimated by three real video sequences that contain 200 frames are named GMV#1~3. To simplify the problem, one horizontal dimension is discussed and illustrated. The characteristics of GMVs are: GMV#1 contains constant speed movement and small fluctuation, GMV#2 contains constant speed movement and large fluctuation, GMV#3 contains only fluctuation without constant speed movement. The three sets of GMV observe the effects of image stabilization. To easily illustrate this, the three motion trajectories are plotted respectively to compare the improvement of the three methods, MC#1, MC#2, and FIDIS.

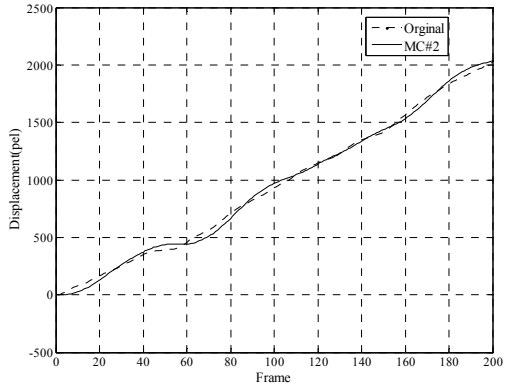
The image size of the experiment is 640x480 pels. The clipper is bounded within ± 53 pels. Parameters of MC#1 are set $k = 0.95$ and $\alpha = 1$, MC#2 are set $k = 0.95$ and $\beta = 0.015$, FIDIS are set $k = 0.98$ and membership functions are shown in Fig. 3.8 and 3.9.

Fig. 4.7 shows the comparison of motion trajectories of GMV#1 which are compensated by three different methods, MC#1, MC#2 and FIDIS, respectively. From (a), due to the GMV#1 having a constant speed movement attribute, MC#1 has the drawback to deal with it. Small fluctuation affects the smoothness. This is minor, but the lag of motion trajectory is obvious. From (b), compensated by MC#2, the lag of motion trajectory has improved. But due to the addition of an integrator in MC#2, there is a little oscillation across the original motion trajectory. From (c), compensated by FIDIS, obvious improvements in both the lag and oscillation problems have been realized.

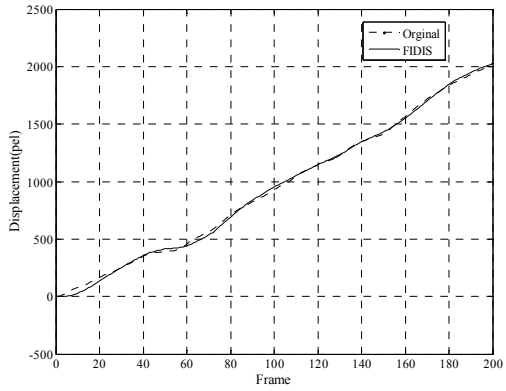
Fig. 4.8 shows the comparison of motion trajectories of GMV#2 among three compensated methods. From (a), due to the GMV#2 having a constant speed movement and large fluctuation attribute, MC#1 appears to have the drawback that is unable to compensate the large fluctuation. The reason is that the compensating vector exceeds the window shift allowance. Fig. 4.8 (b) and (c) almost have the same results, i.e. the FIDIS almost chose MC#2 as the motion compensating vector.



(a)

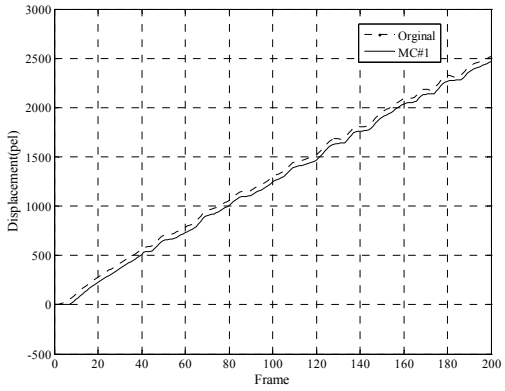


(b)

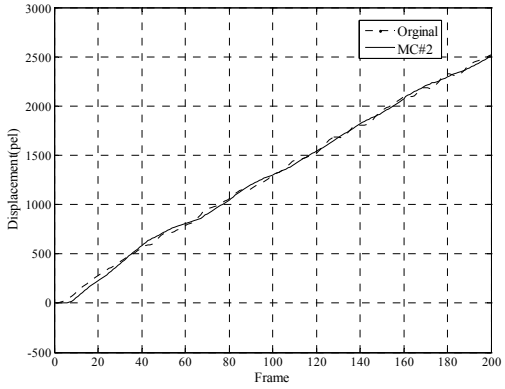


(c)

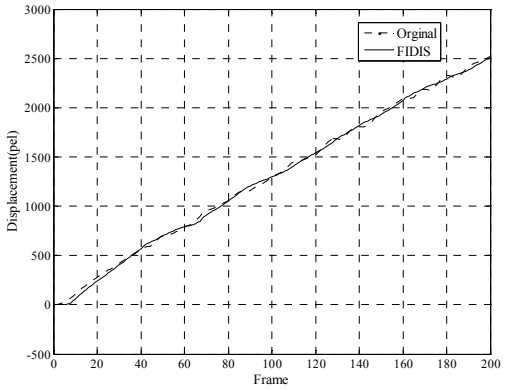
Fig. 4.7. Comparisons of original and compensated motion trajectories by three methods, MC#1, MC#2 and FIDIS with respect to GMV#1.



(a)



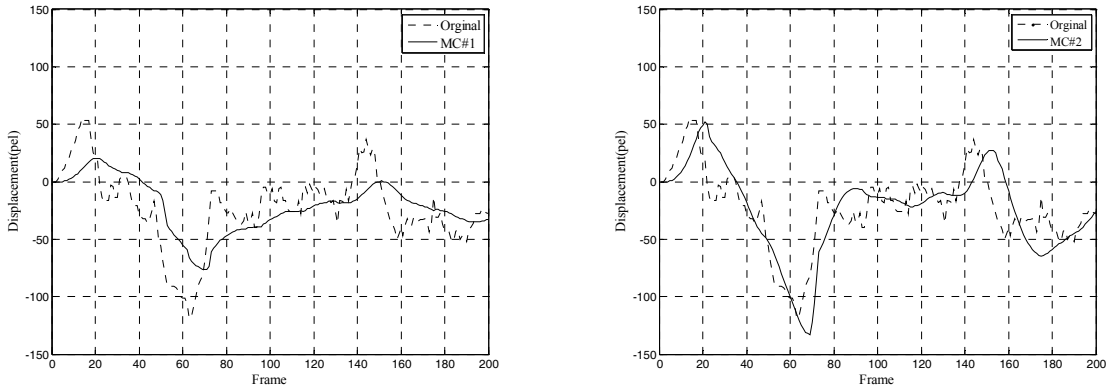
(b)



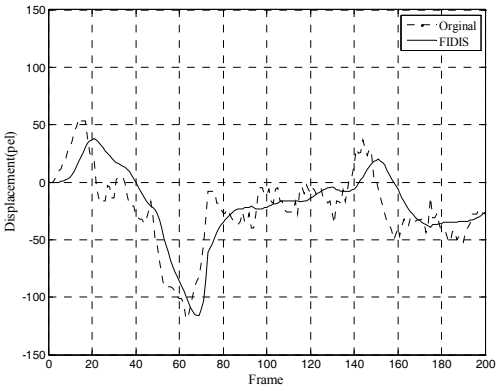
(c)

Fig. 4.8. Comparisons of original and compensated motion trajectories by three methods, MC#1, MC#2 and FIDIS with respect to GMV#2.

Fig. 4.9 shows the comparison of motion trajectories of GMV#3. From (a), it observes the original trajectory. It only has back and forth fluctuation. Therefore, there is a good smooth compensated trajectory by MC#1. From (b), MC#2 also derives a good smooth compensated trajectory, but it overshoots somewhere such as frame=60~80. From (c), FIDIS is observed that the compensated trajectory still keeps the smoothness and eliminates the overshoot problem as well.



(a) (b)



(c)

Fig. 4.9. Comparisons of original and compensated motion trajectories by three methods, MC#1, MC#2 and FIDIS with respect to GMV#3.

Table 4.5 shows the SI comparison of MC#1, MC#2 and FIDIS in GMV#1~4. The GMV#4 cascades GMV#1~3. That means GMV#4 contains various conditions in the global motion sequence. The SI of the original motion vector can look as intrinsic mobility plus external fluctuation. Three compensation methods can mostly improve the stability of the image. From the table, both MC#1 and MC#2 have equally good merits. Except for panning, MC#1 is suitable for general conditions. MC#2 is suitable for panning with constant speed movement. FIDIS choose the better $SI(n)$ from MC#1 and MC#2 to generate the compensation vector. This shows that FIDIS has superior performance in GMV#4.

Table 4.5.

SI comparisons of three methods with respective to four GMV sets

Video sequences	SI(smoothness index)			
	Original	MC#1	MC#2	FIDIS
GMV#1	0.7940	0.7839	0.7035	0.5678
GMV#2	3.2111	2.8794	0.7889	0.9699
GMV#3	7.3719	0.8040	0.9347	0.8292
GMV#4	3.8214	1.5593	0.9048	0.8264

Fig. 4.10. shows the result of FIDIS which uses a short term smoothness index to determine the better compensation method. The sequence n=1~200 corresponds to GMV#1, n=201~400 corresponds to GMV#2, and so on. In the first segment (n=1~200), the CMVs processed by MC#1 and MC#2 have almost the same $SI(n)$. Therefore, the fuzzy output is about 0.5. In the segment of n=250~400, the $SI(n)$ of MC#1 is higher; then the fuzzy output is close to 0. In other words, MC#2 has a better $SI(n)$ in this segment. In the segment of n=422~438, the $SI(n)$ of MC#2 has a spiked shape. Therefore, the fuzzy output approaches 1 and MC#1 is selected to compensate. Obviously, FIDIS is superior to MC#1, MC#2, respectively, for evaluating motion trajectory, the smoothness index or the short term smoothness index.

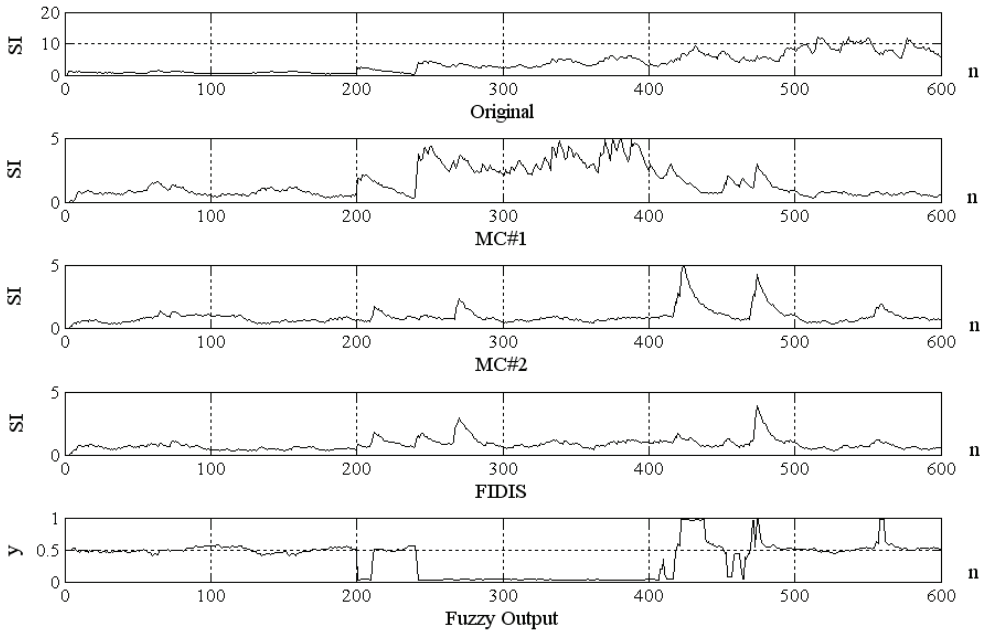


Fig. 4.10. FIDIS operation illustrated by short term smoothness $SI(n)$ and fuzzy output.

5. Conclusions

Digital image sequences that have been acquired by hand-held, in-car or fixed-type surveillance video cameras are usually affected by the undesired motions produced by hand-shaking , bumpy rides, and many other external forces. The unwanted positional fluctuations of the image sequence will affect the visual quality and impede the subsequent processes for various applications. Although undesired motions are usually irregular and uneven compared to intentional global motions, the challenge of image stabilization systems is how to compensate for the unwanted shaking of the camera, without being influenced by the object moving in the image or by an intentional motion. In this dissertation research works, the related problems have been partially solved. The major contributions are summarized as:

(1) The local motion estimation algorithm is proposed by using the inverse triangle method to evaluate the reliability of a calculated local motion vector, and then generate a refined motion vector with a reliability index to form a set of selected local motion vectors which can determine the global motion vector robustly. (This research work has been published in "A robust digital image stabilization technique based on inverse triangle method and background detection," IEEE Trans. on Consumer Electronics, vol. 51, no. 2, pp. 335 - 345, May 2005 (SCI) and granted in Taiwan invention patent 泛用型數位影像防振系統與方法。中華民國發明專利 I307443 號 2009/03/11.).

(2) To improve the robustness of global motion estimation, skyline detection and background based peer to peer evaluation are proposed. The background based peer to peer evaluation solves the large moving object problem and reduces artificial shaking in DIS. The skyline detection uses the coarse skyline to adjust the background-based evaluation area. It can prevent the invalid result due to some of the five regions located on the low-contrast area. (This research work has been published in "A Robust In-Car Digital Image Stabilization Technique," IEEE Trans. on Systems, Man, and Cybernetics, Part C: Applications and Reviews, vol. 37, no. 2, pp. 234-247, Mar. 2007 (SCI)).

(3) The optimization of selecting the representative points from the background image in the fixed-type video surveillance application is proposed to reduce the computation complexity. (This research work has been published in "Digital Image Stabilization Technique and its Application to Video Surveillance System," 2006 MIICS Mechatronic and Industry Interact Cross Strait Conference pp. 309-314, Hsinchu Taiwan, Nov. 23 2005).

(4) The objective of motion compensation is to achieve the optimal stabilization of shaking images within a specified compensation range. The proposed FIDIS combines two motion compensation methods to adaptively determine a better method for motion compensation. It solves the problem of lag and overshoot of motion trajectory. The proposed FIDIS shows effective improvements in different conditions of image sequence through the evaluations of a smoothness index and motion trajectory. (This research work has been published in "Fuzzy Inference Applied to Digital Image Stabilization Techniques," Journal of Image Processing and Pattern Recognition, vol. 13, no. 3, pp. 55-66, 2007.(Domestic Journal)).

.

Bibliography

[1] M. Oshima, et al., “VHS camcorder with electronic image stabilizer,” *IEEE Trans. on Consumer Electronics*, vol. 35, no. 4, pp. 749-758, Nov. 1989.

[2] K. Sato, et al., “Control techniques for optical image stabilizing system,” *IEEE Trans. on Consumer Electronics*, vol. 39, no. 3, pp. 461-466, Aug. 1993.

[3] S. C. Hsu, S. F. Liang, C. T. Lin, "A robust digital image stabilization technique based on inverse triangle method and background detection," *IEEE Transactions on Consumer Electronics*, vol. 51, no. 2, pp. 335 - 345, May 2005.

[4] S. J. Ko, S. H. Lee, and K. H. Lee, “Digital image stabilizing algorithms based on bit-plane matching,” *IEEE Tran. on Consumer Electronics*, vol. 44, no. 3, pp. 617-622, Aug. 1998.

[5] K. Uomori, et al., “Automatic image stabilizing system by full-digital signal processing,” *IEEE Trans. on Consumer Electronics*, vol. 36, no. 3, pp. 510-519, Aug. 1990.

[6] Y. Egusa, et al., “An application of fuzzy set theory for an electronic video camera image stabilizer,” *IEEE Transactions on Fuzzy Systems*, vol. 3, no. 3, pp. 351-356, Aug. 1995.

[7] J. K. Paik, Y. C. Park, and D. W. Kim, “An adaptive motion decision system for digital image stabilizer based on edge pattern matching,” *IEEE Trans. on Consumer Electronics*, vol. 38, no. 3, pp. 607-616, Aug. 1992.

[8] J. K. Paik, Y. C. Park, and S. W. Park, “An edge detection approach to digital image stabilization based on tri-state adaptive linear neurons,” *IEEE Tran. on Consumer Electronics*, vol. 37, no. 3, pp. 521-530, Aug 1991.

[9] S. W. Jeon, et al., “Fast digital image stabilizer based on Gray-coded bit-plane matching,” *IEEE Trans. Consumer Electronics*, vol. 45, no. 3, pp. 598-603, Aug. 1999.

[10] F. Vella, et al., “Digital image stabilization by adaptive block motion vectors filtering,” *IEEE Trans. on Consumer Electronics*, vol. 48, no. 3, pp. 796-801, Aug. 2002.

[11] S. Erturk, “Digital image stabilization with sub-image phase correlation based global motion estimation,” *IEEE Trans. on Consumer Electronics*, vol. 49, no. 4, pp. 1320-1325, Nov. 2003.

[12] S. C. Hsu, S. F. Liang, and C. T. Lin, "Digital Image Stabilization Technique and its Application to Video Surveillance System," *2005 MIICS Mechatronic and Industry Interact Cross Strait Conference*, pp. 309-314, Hsinchu Taiwan, Nov. 23 2005.

[13] S. C. Hsu, S. F. Liang, K. W. Fan and C. T. Lin, "A Robust In-Car Digital Image Stabilization Technique," *IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews*, vol. 37, no. 2, pp. 234-247, Mar. 2007.

[14] L. Xu, and X. Lin, "Digital image stabilization based on circular block matching," *IEEE Transactions on Consumer Electronics*, vol. 52, no. 2, pp. 566 - 574, May 2006.

[15] J. Y. Chang, et al., “Digital image translational and rotational motion stabilization using

optical flow technique," *IEEE Trans. on Consumer Electronics*, vol. 48, no. 1, pp. 108-115, Feb. 2002.
[16] J. S. Jin, Z. Zhu, and G. Xu, "A stable vision system for moving vehicles," *IEEE Trans. on Intelligent Transportation Systems*, vol. 1, no. 1, pp. 32-39, Mar. 2000.
[17] G. R. Chen, et al., "A novel structure for digital image stabilizer," *Proc. of 2000 IEEE Asia-Pacific Conference on Circuits and Systems*, pp. 101-104, 2000.
[18] Engelsberg, G. Schmidt, "A comparative review of digital image stabilising algorithms for mobile video communications," *IEEE Transactions on Consumer Electronics*, Vol. 45, no 3, pp. 591-597 Aug. 1999
[19] M.B. van Leeuwen. "*Motion estimation and interpretation for in-car systems,*" PhD thesis, Informatics Institute, University of Amsterdam, The Netherlands, May 2002.
[20] S. Erturk, "Image sequence stabilisation: motion vector integration (MVI) versus frame position smoothing (FPS)," *Proc. of the 2nd International Symposium on Image and Signal Processing and Analysis*, pp. 266-271, 2001.
[21] M. K. Gullu, S. Erturk, "Membership function adaptive fuzzy filter for image sequence stabilization," *IEEE Trans. on Consumer Electronics*, vol. 50, no. 1, pp. 1-7, Feb. 2004.
[22] M. K. Kim, E. Kim, D. Shim et al., "An efficient global motion characterization method for image processing applications," *IEEE Trans. on Consumer Electronics*, vol. 43, no. 4, pp. 1010-1018, Nov. 1997.
[23] J. L. Barron, D. J. Fleet, and S. S. Beauchemin, "Performance of optical flow techniques," *International Journal of Computer Vision*, vol. 12, no. 1, pp. 43-77, 1994.
[24] S. BEAUCHEMIN S., and L. BARRON J., "The Computation of Optical Flow," *ACM Computing Surveys*, vol. l 27, no. 3, pp. 433-467, Sep. 1995.
[25] F. Dufaux, and F. Moscheni, "Motion estimation techniques for digital TV: a review and a newcontribution," *Proceedings of the IEEE*, vol. 83, no. 6, pp. 858-876, Jun. 1995.
[26] A. Bainbridge-Smith, and R.G. Lane, "Determining optical flow using a differential method," *Image and Vision Computing*, vol. 15, no. 1, pp. 11-22, Jan. 1997.
[27] A. Giachetti, "Matching techniques to compute image motion," *Image and Vision Computing*, vol. 18, no. 3, pp. 247-260, Feb. 2000.
[28] Barbara Zitová, Jan Flusser, and, "Image registration methods: a survey," *Image and Vision Computing*, vol. 21, no. 11, pp. 977-1000, Oct. 2003.
[29] T. Ha, S. Lee, and J. Kim, "Motion compensated frame interpolation by new block-based motion estimation algorithm," *IEEE Trans. on Consumer Electronics*, vol. 50, no. 2, pp. 752 - 759, May 2004.
[30] Y. W. Huang, S. Y. Chien, B. Y. Hsieh et al., "Global elimination algorithm and architecture design for fast block matching motion estimation," *IEEE Trans. on Circuits and Systems for Video Technology*, vol. 14, no. 6, pp. 898 - 907, Jun. 2004.
[31] J. Fleet David, and D. Jepson Allan, "Computation of component image velocity from local phase information," *International Journal of Computer Vision*, vol. 5, no. 1, pp. 1573-1405, Aug. 1990.
[32] D. J. Fleet, and A. D. Jepson, "Stability of phase information," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 15, no. 12, pp. 1253-1268, 1993
[33] David Vernon, "Computation of instantaneous optical flow using the phase of Fourier components," *Image and Vision Computing*, vol. 17, no. 3-4, pp. 189-199, Mar. 1999.
[34] S. Erturk, "Digital image stabilization with sub-image phase correlation based global motion estimation," *IEEE Transactions on Consumer Electronics*, vol. 49, no. 4, pp.

1320-1325, Nov. 2003.
[35] Pierre-Yves Burgi, and, "Motion estimation based on the direction of intensity gradient," *Image and Vision Computing*, vol. 22, no. 8, pp. 637-653, 2004.
[36] D. J. Fleet, and K. Langley, "Recursive filters for optical flow," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 17, no. 1, pp. 61-67, 1995.
[37] S. BEAUCHEMIN S., and L. BARRON J., "The Computation of Optical Flow," *ACM Computing Surveys*, vol. l 27, no. 3, pp. 433-467, Sep. 1995.
[38] S. Erturk, T. J. Dennis, and, "Image sequence stabilisation based on DFT filtering," *IEE Proceedings of Vision, Image and Signal Processing,*, vol. 147, no. 2, pp. 95-102, Apr. 2000.
[39] H. Adelson Edward, and R. Bergen James, "Spatiotemporal energy models for the perception of motion," *JOSA A*, vol. 2, no. 2, pp. 284-299, 1985.
[40] J. R. Jain, and A. K. Jain, "Displacement measurement and its application in interframe image coding," *IEEE Tran. on Communications*, vol. 29, no. 12, pp. 1799-1808, Dec. 1981
[41] T. Koga, K. Iinuma, A. Hirano et al., "Motion compensated interframe coding for video conferencing," *Proceedings of National Telecommunications Conference*, pp. C9.6.1-C9.6.5, 1981.
[42] R. Li, B. Zeng, and M. L. Liou, "A new three-step search algorithm for block motion estimation," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 4, no. 4, pp. 438 - 442, Aug. 1994.
[43] K. Seth, P. Rangarajan, S. Srinivasan et al., "A parallel architectural implementation of the New Three-Step Search algorithm for block motion estimation," *Proc. 17th International Conference on VLSI Design, 2004.*, pp. 1071 - 1076, 2004.
[44] S. Zhu, K. K. Ma, "A new diamond search algorithm for fast block matching motion estimation," *Proceedings of 1997 International Conference on Information, Communications and Signal Processing, 1997. ICICS.*, pp. 292 - 296, 9-12 Sep. 1997.
[45] J. B. Xu, L. M. Po, and C. K. Cheung, "Adaptive motion tracking block matching algorithms for video coding," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 9, no. 7, pp. 1025 - 1029, Oct. 1999.
[46] J. L. Chen, and P. Y. Chen, "An efficient gray search algorithm for the estimation of motion vectors," *IEEE Transactions on Systems, Man and Cybernetics, Part C*, vol. 31, no. 2, pp. 242-248, May 2001.
[47] M. J. Chen, L. G. Chen, and T. D. Chiueh, "One-dimensional full search motion estimation algorithm for video coding," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 4, no. 5, pp. 504 - 509, Oct. 1994.
[48] L. M. Po, and W. C. Ma, "A novel four-step search algorithm for fast block motion estimation," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 6, no. 3, pp. 313 - 317, Jun. 1996.
[49] B. Liu, A. Zaccarin, "New fast algorithms for the estimation of block motion vectors," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 3, no. 2, pp. 148 - 157, Apr. 1993.
[50] L. Chen, and N. Tokuda, "A general stability analysis on regional and national voting schemes against noise-why is an electoral college more stable than a direct popular election?," *Artificial Intelligence*, vol. 163, no. 1, pp. 47-66, March, 2005.
[51] M. J. Chen, L. G. Chen, T. D. Chiueh et al., "A new block-matching criterion for motion

estimation and its implementation," *IEEE Transactions on Circuits and Systems for Video Technology,*, vol. 5, no. 3, pp. 231 - 236, Jun. 1995.

[52] S. Garg, and S. N. Merchant, "Interpolated Candidate Motion Vectors for Boundary Matching Error Concealment Technique in Video," *IEEE Transactions on Circuits and Systems II: Express Briefs*, vol. 53, no. 10, pp. 1039 - 1043.

[53] Z. L. He, C. Y. Tsui, K. K. Chan et al., "Low-power VLSI design for motion estimation using adaptive pixel truncation," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 10, no. 5, pp. 669 - 678, Aug. 2000.

[54] *Motion compensation - Wikipedia, the free encyclopedia*. http://en.wikipedia.org/wiki/-Motion_compensation:2007/12/14.

[55] Y. C. Peng, M. T. Lu, and H. Chen, "DSP implementation of digital image stabilizer," *Proc. ICME, IEEE Int'l Conference on Multimedia and Expo*, pp. 4, 6-8 July 2005.

[56] Y. C. Peng, H. A. Chang, H. H. Chen et al., "Digital image stabilization and its integration with video encoder," *IEEE Consumer Communications and Networking Conference*, pp. 544 - 549, 3-6 Jan. 2005.

Author's Publication List

I. Referred Papers

[1] S. C. Hsu, S. F. Liang, KW Fan and C. T. Lin, "A Robust In-Car Digital Image Stabilization Technique," IEEE Trans. on Systems, Man, and Cybernetics, Part C: Applications and Reviews, vol. 37, no. 2, pp. 234-247, Mar. 2007.(SCI)

[2] S. C. Hsu, S. F. Liang, and C. T. Lin, "A robust digital image stabilization technique based on inverse triangle method and background detection," IEEE Trans. on Consumer Electronics, vol. 51, no. 2, pp. 335 - 345, May 2005.(SCI)

II. Conference Papers

[1] S. C. Hsu, and T. S. Chen, "The Research of Image Motion Blurring Effect and Restoration," 2008 MIICS Mechatronic and Industry Interact Cross Strait Conference (in Chinese, pp. 329-333, Hsinchu Taiwan, Nov. 12 2008.

[2] T. S. Chen, and S. C. Hsu, "Image Features Extration by Automatic Multilevel Thresholding," 2007 MIICS Mechatronic and Industry Interact Cross Strait Conference (in Chinese), pp. 18-21, Hsinchu Taiwan, Nov. 7 2007.

[3] T. S. Chen, and S. C. Hsu, "Automatic Threshold Selection from Gray Level Histogram," 2006 MIICS Mechatronic and Industry Interact Cross Strait Conference (in Chinese), pp. 493-497, Hsinchu Taiwan, Nov. 1 2006.

[4] S. C. Hsu, T. S. Chen, and H. L. Liao, "Dual Mode Design of Temperature and Humidity Control," 2006 MIICS Mechatronic and Industry Interact Cross Strait Conference (in Chinese), pp. 409-414, Hsinchu Taiwan, Nov. 1 2006.

[5] Y. C. Cheng, J. F. Chung, C. T. Lin, S. C. Hsu, "Local motion estimation based on cellular neural network technology for image stabilization processing," The 2005 9th International Workshop on Cellular Neural Networks and Their Applications, pp. 286 - 289, 28-30 May 2005.

[6] S. C. Hsu, S. F. Liang, and C. T. Lin, "Digital Image Stabilization Technique and its Application to Video Surveillance System," 2005 MIICS Mechatronic and Industry Interact Cross Strait Conference (in Chinese), pp. 309-314, Hsinchu Taiwan, Nov. 23 2005.

[7] S. C. Hsu, "The LabVIEW Based Battery Capacity Measure System," The 2004 Seminar and Symposiumon of Harmonic and Grounding Engineering Technology (in Chinese), pp. 43-48, Hsinchu Taiwan, Oct. 1 2004.

[8] S. C. Hsu, et al., "The e-lab Management and Control," The 2003 Seminar and Symposiumon of Photo and Semiconductor Technology (in Chinese), pp. 18-22, Hsinchu Taiwan, Nov. 28 2003.

[9] T. M. Chen, and S. C. Hsu, "The Design of PLC SCADA Teaching System," The 2001

Seminar and Symposiumon of Applied Power electronics Technology (in Chinese), pp. 91-94, Hsinchu Taiwan, Nov. 23 2001.

[10] S. C. Hsu, "The Technology of Fingerprint Verification," The 2001 Seminar and Symposiumon of Applied Power electronics Technology (in Chinese), pp. 48-50, Hsinchu Taiwan, Nov. 23 2001.

III. Patents

[1] 林進燈、徐聖哲、梁勝富。2009/03/11。泛用型數位影像防振系統與方法。中華民國發明專利 I307443 號.

Vita

徐聖哲

學經歷：

1. 62 年 9 月～65 年 6 月　台灣省立台南第一高級中學
2. 65 年 9 月～69 年 6 月　中原理工學院電機工程系
3. 71 年 9 月～76 年 9 月　工業技術研究院機械工業研究所助理/副研究員
4. 76 年 9 月～78 年 6 月　美國紐澤西理工學院電機工程研究所碩士進修
5. 78 年 8 月～79 年 1 月　工業技術研究院機械工程研究所副研究員
6. 79 年 2 月～81 年 7 月　福祿遠東股份有限公司工程部資深工程師
7. 81 年 8 月～95 年 1 月　大華技術學院電機工程系專任講師
8. 89 年 9 月～Now　國立交通大學電機與控制工程學系博士進修
9. 96 年 2 月～97 年 2 月　大華技術學院電機工程系專任副教授
10. 97 年 3 月～Now　大華技術學院電機工程系專任副教授兼系主任