MOVING OBJECT DETECTION IN A SEQUENCE OF IMAGES TAKEN FROM NON-STATIONARY CAMERA

NORAN AZIZAN BIN CHOLAN

A thesis submitted in fulfilment
of the requirements for the award
of the Degree of Master of Engineering (Electrical)

Faculty of Electrical Engineering
Universiti Teknologi Malaysia

SEPTEMBER 2004

DECLARATION

" I hereby declare that the materials presented in this thesis are the result of my own work except as cited as reference. The thesis has not been accepted for any degree and is not concurrently submitted in candidature of any degree."

Signature : ……………………………..........
Name : Noran Azizan Bin Cholan
Date : 21 September 2004

To My Loving and Caring Family ...

ACKNOWLEDGEMENTS

My first thanks is for my supervisor, Associate Prof. Dr. Syed Abdul Rahman, whose constant support, patience and unbounded enthuasiasm were of invaluable help. His devotion to the needs of the students and the encouragements have made working with him a true delight. Thanks for helping me to kickstart this research by providing insights and his work as reference.

My sincere appreciation to my fellow collegues in the Computer Vision, Video and Image Processing (CVVIP) Research Group, sharing the similar research interests. I value the camaraderie we share as well as the time they spent to share with me enriching ideas, as well as their concern. My gratitude especially goes to Sani, Phaik Yong, Nuha and Nansah for many hours of discussions, as well as assistance with the programming stuff and image acquisation.

My sincerest thanks to all those who have helped to make this thesis possible. Warmest regards to my parents, sisters and brother for their seamless caring encouragement and moral support that has made this journey possible.

Without exception, a special thanks to my fiance Nor Hafizah Bt. Ngajikin for her consistent encouragement and concern over time that made the journey enjoyable and meaningful. Without her unwavering support, love and devotion, this achievement would not have progressed as far as it did.

ABSTRACT

Moving object detection is a vital aspect of motion analysis. It has drawn an increasing attention in the recent years due to its applications such as in communication, traffic monitoring, security surveillance, robot navigation and servoing. Despite the fact that much research efforts have been devoted to this area, detecting moving object using non-stationary moving camera remains a great challenge. The research undertaken in this thesis is mainly concentrated on developing a reliable and robust detection system which incorporates some operation on images such as thresholding, blob labelling, blob matching, filtering and blob analysis. The basic idea behind this system is that the motion of the moving object is different with the motion of background object. Path transversed within a certain period of observation of the moving object is usually longer than background object. By using blob labelling and blob matching operation, this system would be able to track binary blobs over an arbitarily long image sequence. The criteria for matching binary blobs from two adjacent frames are position, height, width, area, colour and aspect ratio. If the the path transversed of a binary blob within a certain period of observation is sufficiently long, then the tracked blob is considered as moving object.

ABSTRAK

Mengesan objek bergerak merupakan satu aspek yang penting dalam analisis pergerakan. Ia semakin mendapat perhatian dalam beberapa tahun kebelakangan ini disebabkan aplikasinya dalam komunikasi (persidangan video), pengawasan lalulintas, pengawasan keselamatan, pengemudian dan pengawalan robot dan sebagainya. Walaupun banyak usaha penyelidikan telah ditumpukan dalam bidang ini, pelaksanaan satu sistem mengesan objek bergerak dengan menggunakan kamera bergerak masih menjadi satu cabaran. Penyelidikan yang dijalankan ini memberi penekanan kepada penghasilan satu sistem mengesan objek bergerak yang tepat, jitu dan efisien, dengan meggunakan kaedah-kaedah dalam Pemerosesan Imej seperti *thresholding, blob labeling, blob matching, filtering* and *blob analysis.* Konsep atau perkara asas dalam membangunkan sistem ini ialah pergerakan objek bergerak adalah berbeza dengan pergerakan objek statik. Jarak perjalanan bagi objek bergerak dalam satu jangka masa tertentu adalah lebih jauh berbanding jarak perjalanan bagi objek statik. Dengan meggunakan *blob labeling* dan *blob matching,* sistem ini berupaya menjejak *binary blobs* dalam satu jujukan imej. Kriteria yang digunakan untuk proses *blob matching* ialah posisi, ketinggian, keluasan, saiz dan warna blob. Jika jarak perjalanan dalam jangka masa tertentu bagi sesuatu blob itu telah melebihi jarak minima yang telah ditetapkan, maka blob itu dikira sebagai blob bergerak.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF APPENDICES

LIST OF ABBREVIATIONS

OFC	-	Optical Flow Constraint
RGB	-	Red Green Blue
PC	-	Personal Computer
RAM	-	Random Access Memory
CCD	-	Charged Coupled Device
CPU	-	Central Processing Unit
GUI	-	Graphical User Interface

CHAPTER I

INTRODUCTION

In everyday life, humans visually keep detecting and tracking a multitude of objects with certain objectives in mind. Examples are orientation in the environment, recognizing persons in the surroundings, locating, recognizing, monitoring and handling of objects.

Although much has been learned from biological and human vision for image processing and analysis, the markedly different tasks and capabilities of the function modules have to be kept in mind. It is not possible to simply transfer or copy the human biological vision to machine vision system. However, the basic functionality of the human vision system can guide us in the task of how these principles can be transferred to technical vision system such as motion analysis. The analysis of motion gives access to the dynamics of processes. Motion analysis is generally a very complex problem. The true motion of objects can only be inferred from their motion after the 3-D reconstruction of the scene.

The increasing use of video sensors, with Pan-Tilt and Zoom capabilities or mounted on moving platforms in surveillance and autonomous driving vehicle applications, have focused the attention of researchers on the detection of moving objects in a video streams acquired by a moving platform.

1.1 Objectives

The main goal of this project is to develop a system which is capable to detect and track moving object from the frames captured by a non-stationary camera. It attempts to make use of the blob labeling, blob analysis and blob matching technique in order to accurately detect and track a moving object

1.2 Project Scopes

In general, there are 2 main components in any motion detection system. One is the scene, more often called the world, and the other is the observer, which, in most cases, is represented by the camera. For analyzing motion detection, it is imperative to understand the configuration between the camera and the world, as each one of them is treated in a slightly different manner. There are 3 possible configurations

1. stationary camera, moving objects (SCMO)
2. moving camera, stationary objects (MCSO)
3. moving camera, moving objects (MCMO)

In this project, the research is based on moving camera, moving objects (MCMO) system.

The proposed detection and tracking algorithm is mainly concentrated on robustly detecting the movement of a single moving object from an image sequence obtained from the moving camera. Nevertheless, the algorithm has a capability to detect multiple moving objects but with some constraints.

Each frame of the captured image sequence is fixed to a size of 256 x 256 pixels. The processing of the acquired image sequence is performed in the binary format, which has 2 value; '0' and '1'.

The system that will be developed is an off-line system and not real time or simulation. The entire moving object detection and tracking program is developed using Microsoft Visual C++ version 6.0.

Camera motion is strictly translation and not rotation or pan-tilt.

1.3 Problem Statements

Problems concerning about this system can be classified into two categories; motion detection and motion tracking.

Motion detection involves verifying the presence of a moving object in image sequences based on the object's temporal change and possibly locating it precisely for tracking or recognition purpose.

Whereas motion tracking is an iterative process of determining the trajectory of a moving object during a video sequence, by monitoring the the object's spatial and temporal changes, including its presence, position, size, shape, etc. This is done by solving the temporal correspondence problem, i.e the problem of matching the target region in successive frames of a sequence of images taken at closedly-spaced time intervals.

These two processes are closely related because motion tracking usually starts with detecting moving objects, while detecting an object repeatedly in subsequent image sequence is often necessary to help and verify tracking.

1.4 Application of Motion Detection and Tracking System

Detecting and tracking a moving object in a dynamic video sequence has been a vital aspect of motion analysis. This detecting and tracking system has become increasingly important due to its application in various areas, including communication (video conferencing), transportation (traffic monitoring and autonomous driving vehicle), security (premise surveillance) and industries (dynamic robot vision and navigation).

Specifically, the main application targeted by the proposed detecting and tracking algorithm is for implementing an autonomous driving system. The system can be used to automatically detect and track any moving object exist in the traffiic scenes such as moving cars in highway within the view of the moving camera. Obstacle detection is one of the key functions in an autonomous driving vehicle.

1.5 Thesis Outline

Chapter 1 provides readers a first glimpse at the basic aspects of the research undertaken, such as objectives, scopes, problem formulation and the application targeted by the developed moving object detection and tracking system.

Chapter 2 gives an insight to the existing vision-based moving object detection and tracking algorithms developed by the various researchers, and subjectively classify them into four categories.

Chapter 3 elaborates on the methodology of the proposed detection and tracking algorithm. This chapter gives an explaination for each main stages in the developed detection system such as blob labeling, blob analysis, blob matching and a testing for moving blob criteria.

Chapter 4 is mainly devoted for demonstrating the experimental results and performance of the proposed detection and tracking algorithm on some off-line image sequences.

Chapter 5 deals with the summary and conclusions of the research. A number of research findings obtained from the empirical results of the implemented detection and tracking system are also discussed. Lastly, some realistic extensions as well as possible enhancements for the research are provided.

CHAPTER 2

LITERATURE REVIEW

In general, the existing approaches and algorithms formulated to deal with objects detection and tracking from the images taken by non-stationary moving camera, can be subjectively classified into a few categories, i.e. feature-based region matching technique, correlation-based matching technique, gradient-based technique and frame differencing technique with certain frame adjustment. The basic idea behind all these techniques is that the motion of the moving object is different with the motion of background object.

The choice of the algorithm that will perform well depends upon many considerations. Of primary concern is the selection of an efficient model that well suited the requirements of the particular desired application.

2.1 Correlation-Based Matching Technique

The idea of this approach is to partition an image I_t into segmented region and to search for each of the segmented regions the corresponding segmented regions in the successive image I_{t+1} at time t+1

The detection and tracking algorithm proposed by Volker Rermann in [14] utilizes both correlation and features as matching technique to detect and track moving object in the scene intended for robotic and vehicle guidance applications. In order to provide more robust approach, color regions that could not be matched on the feature level are matched on the pixel level by the integration of a correlation-based mechanism.

From his finding, with feature-based matching technique, nearly 90% of the image area can be matched in an efficient and accurate way. The correlation-based matching technique, which requires more elaborate processing, is thus only applied to a small fraction of the image area. It yields very accurate displacement vectors and most of the time finds the corresponding color regions.

2.2 Gradient-Based Technique

Gradient-based method, especially optical-flow method, has been shown to be powerful tool for analysis of scene in the case of static camera [10][11]. The main principle in the optical-flow method is to solve the so-called optical flow-constraint (OFC) equation. The equation states that the temporal gradient of the moving object is equal to the spatial gradient, multiplied by the speed of the moving object in both x and y directions. Both of the gradients can be computed easily, leaving the speed in both directions to be unknown.

2.3 Frame Differencing Technique (with certain frame adjustment)

In the case when the camera moves, the background is no longer stationary. Therefore we should use a camera motion modelling to recover the motion of the camera By applying this motion to previous frame and comparing the result with the current frame we can estimate the areas for which the camera motion model does not hold. These areas correspond to the actual moving objects in the scene.

In order to model and compute motion, an understanding is needed as to how images (and therefore image motion) are formed [13]. Motion in image sequences acquired by a video camera is induced by movements of objects in a 3-D scene and by camera motion. Thus, camera parameters, such as its 3-D motion (rotation, translation) or focal length, play an important role in image motion modelling. If these parameters are known precisely, only object motion needs to be recovered. The 3-D motion of objects and camera induces 2-D motion on the image plane *via* a suitable projection system. To compute motion trajectories, underlying models must be selected, e.g., motion model (representation, region of support) and occlusion model. The choice of models and their parameters is application-dependent; for example the occlusion model may or may not be relevant for an image analysis.

Consider a point on an object moving in a 3-D space. Let its position at time t be $\mathbf{X} = \mathbf{X}(t) = (X(t); Y(t); Z(t))^T \in IR^3$ expressed in camera coordinates. (X(t); t) defines a curve in 3-D space over time which we refer to as the *world motion trajectory*. For any two time instants t and τ , the world motion trajectory identifies a 3-D displacement in position can be represented by Equation 2.1.

$$D_{t,\tau}(X) = X(\tau) - X(t) \qquad (2.1)$$

For a review of 3-D motion and its relationship to the apparent 2-D motion of interest here, the information is referred to [13].

An image acquisition system projects the 3-D world onto a 2-D image plane with image coordinates $x = (x; y)^T \in \Lambda$ where Λ is a sampling grid, usually an orthogonal lattice. Upon this projection, world motion trajectories result in (2-D) *motion trajectories* (x(t), t). Definition of a 2-D motion is adopted from [13]: a trajectory is defined only in the time interval in which the associated point is visible in the image. Thus, assuming non-transparent objects, each spatio-temporal position (x; t) belongs to a motion trajectory of one visible point only. As depicted in Fig. 2.1, the 2-D displacement can be expressed, similarly to the 3-D displacement, as follows

$$d_{t,\tau}(x) = x(\tau) - x(t) \tag{2.2}$$

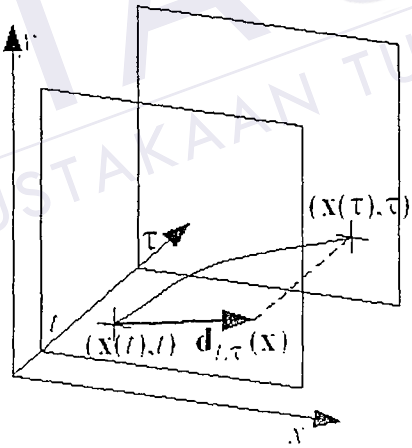

Fig. 2.1. Motion trajectory x(t) and associated displacement vector $d_{t,}(x)$.

In general, a *motion field* is a vector-valued function of continuous spatial coordinates. In practical applications, this function is often described in a parametric form using a finite, usually small, number of parameters. Since 2-D motion results from the projection of moving 3-D objects onto the image plane, a model for 2-D motion fields can be derived from models describing 3-D motion, 3-D surface function and camera projection geometry. If these models are parametric, the resulting 2-D motion model will be parametric as well. As a simple example, consider a 3-D planar patch undergoing 3-D affine motion under orthographic projection. The 3-D affine motion can be written as follows

$$D(X) = (R - I)X + s \qquad (2.3)$$

In general, the 3x3 matrix ***R*** = (rij) has 9 degrees of freedom, and the translational motion vector $s = (s1; s2; s3)^T$ has another 3 degrees of freedom. Equation (2.3) includes rigid motion as a special case. Then, ***R*** is a rotation matrix, i.e., its columns (rows) are orthonormal, thus allowing only 3 degrees of freedom corresponding to the three rotation axes. Let the planar patch be specified by three parameters α, β and γ as follows,

$$\alpha X + \beta Y + \gamma Z = 1 \qquad (2.4)$$

The camera model provides two additional scalar equations mapping 3-D world coordinates onto 2-D coordinates of the image plane. For an orthographic projection, the following relationship holds:

$$x = cX,\ \ y = cY\ ;\ \ c \in IR^3 \qquad (2.5)$$

Substituting equations for the camera model (2.5) and for the 3-D surface (2.4) into (2.6), one readily obtains a model for 2-D motion which, for the given example, becomes the 2-D affine model

$$d(x) = (A - I)x + b; \qquad (2.6)$$

with

$$A=\begin{pmatrix} r_{11}-\frac{\alpha}{\gamma}r_{13} & r_{12}-\frac{\alpha}{\gamma}r_{13} \\ r_{21}-\frac{\alpha}{\gamma}r_{23} & r_{22}-\frac{\alpha}{\gamma}r_{23} \end{pmatrix} \qquad b=\begin{pmatrix} \frac{\varepsilon}{\gamma}r_{13}+\varepsilon s_1 \\ \frac{\varepsilon}{\gamma}r_{23}+\varepsilon s_2 \end{pmatrix}$$

Table 2.1 : Motion Model

2-D model			3-D model		Camera model
	Number of parameters	Motion field	3-D surface function	3-D motion	
Translational	2	$d(x)=(a_1,b_1)^T$	arbitrary	rigid 3-D translation	orthographic
Affine	6	$d(x)=\begin{pmatrix} a_1 & a_2 \\ b_1 & b_2 \end{pmatrix}x+\begin{pmatrix} a_3 \\ b_3 \end{pmatrix}$	planar	3-D affine	orthographic
Projective linear	8	$d(x)=\begin{pmatrix} \frac{a_1+a_2x+a_3y}{1+a_4x+b_4y} \\ \frac{b_1+b_2x+b_3y}{1+a_4x+b_4y} \end{pmatrix}-x$	planar	3-D affine	perspective
Quadratic	12	$d(x)=\begin{pmatrix} a_1+a_2x+a_3y+a_4x^2+a_5xy+a_6y^2 \\ b_1+b_2x+b_3y+b_4x^2+b_5xy+b_6y^2 \end{pmatrix}$	parabolic	3-D affine	orthographic
Sampled	2 per Δ^2 pixel	$d(x)=\sum_{i,j}\begin{pmatrix} a_{ij} \\ b_{ij} \end{pmatrix}H(x-\Delta\cdot i,y-\Delta\cdot j)$	'smooth' as specified by interpolation kernel H		arbitrary
Polynomial	2|K| motion-adaptive	$d(x)=\sum_{(i,j)\in K}\begin{pmatrix} a_{ij} \\ b_{ij} \end{pmatrix}x^i y^j$	'smooth' as specified by K		arbitrary

This model has been extensively used in the literature for 2-D motion representation [13]. Clearly, a 2-D motion model does not uniquely correspond to one 3-D model; identical 2-D motion models may result from different assumptions about 3-D motion, surface and camera projection models. Table 2.1 summarizes some parametric models for 2-D motion and provides possible underlying assumptions.

REFERENCES

[1] Syed Abdul Rahman, Phd Thesis: "Moving Object Feature Detector & Extractor Using a Novel Hybrid Technique", University of Bradford, Bradford, UK.1997.

[2] Wan Ayub Bin Wan Ahmad, Master Thesis: "Menjejaki Objek Yang Bergerak Dalam Satu Jujukan Imej", Universiti Teknologi Malaysia, Skudai, Malaysia.2002.

[3] Yeoh Phaik Yong, Master Draft Thesis: "Integration of Projection Histograms For Real Time Tracking Of Moving Object", Universiti Teknologi Malaysia, Skudai, Malaysia.2003.

[4] Bernd Heisele, W.Ritter, "Obstacle Detection Based on Color Blob Flow". Proc. Intelligent Vehicles Symposium, pages 282-286, Detroit, 1995.

[5] Bernd Heisele, W.Ritter and U.Krebel, "Tracking Non-Rigid, Moving Objects Based on Color Cluster Flow". Proc. Computer Vision and Pattern Recognition, pages 253-257, San Juan, 1997

[6] Bernd Heisele, "Motion-Based Object Detection and Tracking in Color Image Sequences". Fourth Asian Conference on Computer Vision, pages 1028-1033, Taipei, 2000

[7] B. Heisele, T. Serre, S. Mukherjee and T. Poggio. "Feature Reduction and Hierarchy of Classifiers for Fast Object Detection in Video Images". Proceedings of 2001

IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2001), Kauai, Hawaii, Vol. 2, 18-24, December 2001.

[8] Yoav Rosenberg and Michael Werman, "Real-Time Object Tracking from a Moving Video Camera: A Software Approach on a PC". 4th IEEE Workshop on Applications of Computer Vision (WACV'98), New Jersey, 1998

[9] S.Jehan-Besson, M. Barlaud and G. Aubert, "Region-Based Active Contours for Video Object Segmentation with Camera Compensation". International Conference on Image Processing, Thessaloniki, Greece, 2001.

[10] Horn and Schunck, "Determining Optical Flow". Artificial Intelligence, 1981

[11] Enkelmen, Gengenbach, Kruger, Rossle and Tolle, "Obstacle Detection by Real-Time Optical Flow Evaluation". IEEE Transactions on Pattern Analysis and Machine Intelligence, 1996.

[12] Alireza Behrad, Ali Shahrokni, Seyed Ahmad Motamedi and Kurosh Madani, "A Robust Vision-based Moving Target detection and Tracking System". Proceeding of Image and Vision Computing Conference (IVCNZ2001), New Zealand, 2001.

[13] Christoph Stiller, Janusz Konrad "Estimating Motion in Image Sequences: A tutorial on modeling and computation of 2D motion". IEEE 1999.

[14]Volker Rehrmann, "Object Oriented Motion Estimation in Color Image Sequences". 5th European Conference on Computer Vision (ECCV '98), Freiburg, 2-6th June 1998.

[15] Scott e Umbaugh, "Computer Vision and Image Processing",Prentice Hall PTR, NJ, USA, 1999

[16] K. Suzuki, I. Horiba, N. Sugie, “Linear time connected-component labeling based on sequential local operations”, Proc. 15th Int. Conf. Pattern Recognition, Barcelona, Spain, vol. 2, 2000, pp. 434–437.