IMPLEMENTATION OF CROSS DIAMOND SEARCH (CDS) ALGORITHM FOR MOTION ESTIMATION USING MATLAB

NOOR IRWAN ZILAH BINTI YAHYA

This report is submitted in partial fulfillment of the requirements for the award of Bachelor of Electronic Engineering (Telecommunication Electronics) With Honours

Faculty of Electronic and Computer Engineering
Universiti Teknikal Malaysia Melaka

April 2009

UNIVERSTI TEKNIKAL MALAYSIA MELAKA
FAKULTI KEJURUTERAAN ELEKTRONIK DAN KEJURUTERAAN KOMPUTER

BORANG PENGESAHAN STATUS LAPORAN
PROJEK SARJANA MUDA II

Tajuk Projek : Implementation of Cross Diamond Search (CDS) Algorithm For Motion Estimation Using MATLAB

Sesi Pengajian : 2005/2009

Saya NOOR IRWAN ZILAH BINTI YAHYA

mengaku membenarkan Laporan Projek Sarjana Muda ini disimpan di Perpustakaan dengan syarat-syarat kegunaan seperti berikut:

1. Laporan adalah hakmilik Universiti Teknikal Malaysia Melaka.
2. Perpustakaan dibenarkan membuat salinan untuk tujuan pengajian sahaja.
3. Perpustakaan dibenarkan membuat salinan laporan ini sebagai bahan pertukaran antara institusi pengajian tinggi.
4. Sila tandakan (√) :

[ ]	**SULIT***	(Mengandungi maklumat yang berdarjah keselamatan atau kepentingan Malaysia seperti yang termaktub di dalam AKTA RAHSIA RASMI 1972)
[ ]	**TERHAD***	(Mengandungi maklumat terhad yang telah ditentukan oleh organisasi/badan di mana penyelidikan dijalankan)
[/]	**TIDAK TERHAD**	

Disahkan oleh:

(TANDATANGAN PENULIS)

Alamat Tetap: PT 3030, Jalan Wakaf Utama,
Kg. Tun Razak, Bkt. Katil, Melaka

Tarikh : 30 APRIL 2009

(COP DAN TANDATANGAN PENYELIA)

REDZUAN BIN ABDUL MANAP
Ketua Jabatan (Kej. Telekomunikasi)
Fakulti Kejuruteraan Elektronik Dan Kejuruteraan Komputer
Universiti Teknikal Malaysia Melaka (UTeM)
Karung Berkunci No 1752
Pejabat Pos Durian Tunggal
76109 Durian Tunggal, Melaka

Tarikh : 30 APRIL 2009

"I hereby declare that this report is the result of my own work except for quotes as cited in the references."

Signature :

Author : NOOR IRWAN ZILAH BINTI YAHYA

Date : 5 APRIL 2009

"I hereby declare that I have read this report and in my opinion this report is sufficient in terms of the scope and quality for the award of Bachelor of Electronic Engineering (Telecommunication Electronics) With Honours."

Signature :

Supervisor's Name : EN. REDZUAN BIN ABDUL MANAP

30/4/2009

ACKNOWLEDGEMENT

All Praise be to Allah, the Almighty God, the Holy Lord. He is the Exalted Sustainer, the Worthy Cherisher. There is no God but Allah, the Great, the Tolerant, there is no God but Allah, the Lord of the Magnificent Throne. There is no God but Allah, the Lord of the heaven and the Earth, the Lord of the Edifying Throne.

First and foremost, I would like to express my sincere gratitude to my supervisor, En. Redzuan bin Abdul Manap, lecturer in the Telecommunication Engineering Department, Universiti Teknikal Malaysia Melaka for the guidance throughout this project. Thank you for the help, support, motivation and advice.

To dear friends; Siti Hajar binti Ahmad, Nor Hashela binti Jamadin, Elya Irwana binti Zullkifly, and Raudzatul Adawiah binti Yunos, thank you all for lending hands during the progress of the project, especially for your assistance in developing the program codes. Your contributions are highly appreciated.

To the family, thank you for your prayers and moral support. Without your supports, I might not be able to complete this project. Last but not least, million of thanks to lecturers, classmates and friends who contributed in this project directly or indirectly.

ABSTRACT

In block motion estimation, search patterns with different shapes or sizes and the center-biased characteristics of motion-vector distribution have a large impact on the searching speed and quality of performance. In this project, an algorithm using a cross-search pattern as the initial step and large/small diamond search (DS) patterns as the subsequent steps for fast block motion estimation is proposed. The initial cross-search pattern is designed to fit the cross-center-biased motion vector distribution characteristics of the real-world sequences by evaluating the nine relatively higher probable candidates located horizontally and vertically at the center of the search grid. The implemented cross-diamond search (CDS) algorithm employs the halfway-stop technique and finds small motion vectors with fewer search points than the most algorithms while maintaining similar or even better search quality. Experimental results show that the CDS is much more robust, and provides faster searching speed and smaller distortions than other popular fast block-matching algorithms.

ABSTRAK

Di dalam anggaran pergerakan blok, paten pencarian dengan pelbagai bentuk atau saiz dan ciri-ciri kecenderungan di bahagian pusat oleh taburan vektor pergerakan mempunyai impak yang besar di dalam kelajuan pencarian dan kualiti pelaksanaan. Dalam projek ini, satu algoritma yang mengaplikasikan paten pencarian silang sebagai langkah permulaan dan paten pencarian berlian besar/kecil sebagai langkah selanjutnya bagi mempercepatkan anggaran pergerakan blok telah diusulkan. Permulaan paten pencarian silang direka sebagai padanan kepada ciri-ciri taburan vektor pergerakan bagi kecenderungan carian silang di bahagian pusat untuk turutan sebenar dengan menilai sembilan titik kemungkinan tertinggi yang ditempatkan secara melintang dan menegak di bahagian pusat grid pencarian. Algoritma pencarian silang-berlian (CDS) menggunakan teknik berhenti separuh jalan dan mencari vektor pergerakan terkecil dengan titik pencarian terendah berbanding kebanyakan algoritma di samping mengekalkan atau memperbaiki kualiti pencarian. Keputusan kajian menunjukkan bahawa paten pencarian silang-berlian adalah lebih kukuh di samping menyediakan kaedah pencarian yang lebih laju serta herotan yang lebih kecil berbanding algoritma padanan blok terkenal yang lain.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURE

LIST OF ABBREVIATIONS

ATM		Asynchronous Time Multiplexing
BMA		Block Matching Algorithm
CDS	-	Cross-Diamond Search
CCB	-	Cross-center Biased
CIF	-	Common Intermediate Format
CS	-	Cross Search
CSP	-	Cross-shaped Pattern
DCB	-	Diamond-center Biased
DCT	-	Discrete Cosine Transform
DPCM	-	Differential Pulse Code Modulator
DS	-	Diamond Search
DV	-	Digital Video
DVD	-	Digital Versatile Disk
FS	-	Full Search
FSS	-	Four-Step Search
HDTV	-	High-definition Television
ISDN	-	Integrated Services Digital Network
ITU	-	International Telecommunication Union
JPEG	-	Joint Photographic Experts Group
LAN	-	Local Area Network
LDSP	-	Large Diamond Search Pattern
MAD	-	Mean Absolute Difference
ME	-	Motion Estimation
MPEG	-	Moving Picture Experts Group
MSE	-	Mean Squared Error
MV	-	Motion Vector

NTSS	-	New Three-Step Search
PSNR	-	Peak-to-Noise-Ratio
SAD	-	Sum of Absolute Difference
SDSP	-	Small Diamond Search Pattern
TSS	-	Three-step Search
WAN	-	Wide Area Network

CHAPTER 1

INTRODUCTION

1.1 Project Background

To achieve high compression ratio in video coding, a technique known as Block Matching Motion Estimation has been widely adopted in various coding standards. This technique is implemented conventionally by exhaustively testing all the candidate blocks within the search window. This type of implementation, called Full Search (FS) Algorithm, gives the optimum solution. However, substantial amount of computational workload is required in this algorithm. To overcome this drawback, many fast Block Matching Algorithms (BMA's) have been proposed and developed. Different search patterns and strategies are exploited in these algorithms in order to find the optimum Motion Vector (MV) with minimal number of required search points.

One of these fast BMA's, which is proposed to be implemented in this project, is called Cross Diamond Search (CDS) Algorithm. The student is required to implement the algorithm in MATLAB and then compare its performance to FS algorithm as well as to other fast BMA's in terms of the peak signal-to-noise ratio (PSNR) produced, number of search points required and computational complexity.

1.2 Problem Statement

Substantial amount of computational workload is required during the execution of Full Search algorithm; however this drawback can be overcome by many types of fast BMA's which have been proposed and developed. Different search patterns and strategies are exploited in these fast BMA algorithms in order to find the optimum MV with minimal number of required search point. One of these fast BMA's is called Cross Diamond Search (CDS) Algorithm.

1.3 Scope

This project will focus on 3 main areas which include :

i. Literature review on video coding, Video Compression, Motion Estimation, and BMAs including CDS.
ii. The development and implementation of CDS algorithm using MATLAB platform.
iii. The performance analysis of CDS to FS algorithm and CDS to other BMAs'.

1.4 Objectives

The aims of this project are :

i. To implement the CDS algorithm in MATLAB.
ii. To compare its performance to FS algorithm as well as to other fast Block Matching Algorithms' such as Diamond Search (DS), Four Step Search (FSS), Cross Search (CS) and New Three Step Search (NTSS).

CHAPTER 2

LITERATURE REVIEW

2.1 Video Compression And Coding Technique

Video takes a lot of space. Uncompressed footage from a camcorder takes about 17MB per second of video. Because it takes so much space, video must be compressed before it is put on the web [1]. "Compressed" just means that the information is packed into a smaller space. There are two kinds of compression which is lossy and lossless.

Lossy compression means that the compressed file has less data in it than the original file. In some cases this translates to lower quality files, because information has been lost. However, at the time we start to notice the difference, we have lose a relatively large amount of data. Lossy compression makes up for the loss in quality by producing comparatively small files. For example, DVDs are compressed using the MPEG-2 format, which can make files 15 to 30 times smaller, but we still tend to perceive DVDs as having high-quality picture.

Lossless compression is exactly what it sounds like, compression where none of the information is lost. This is not nearly as useful because files often end up being the same size as they were before compression. This may seem pointless, as reducing the file size is the primary goal of compression. However, if file size is not an issue, using lossless compression will result in a perfect-quality picture. For example, a video editor transferring files from one computer to another using a hard drive might choose to use lossless compression to preserve quality while he or she is working.

2.1.1 The Needs For Compression

A single digital television signal in CCIR 601 format requires a transmission rate of 216 Mbps. This bit rate is too high for most existing practical communication networks. For example, most local area networks (LANs) offer data transmission at rates on the order of 10 Mbps, and most wide area networks (WANs) support much lower data rates than this. The emerging Asynchronous Time Multiplexing (ATM) networks are capable of transmitting higher bit rates. However, distributing an uncompressed CCIR 601 bitstream over these networks is still prohibitively expensive [2].

This means that the digital video information must be compressed or encoded prior to transmission to accommodate for different transmission media's capabilities. At the receiver's end, the compressed bitstream received is first decompressed or decoded and then displayed. A number of video coding techniques and standards have been developed within the last few years that exploit the inherent redundancy in still images and moving video sequences to provide significant data compression.

Some compression can be achieved by exploiting the statistical redundancy within the data. For example, video data is often highly correlated both spatially and temporally. This redundancy can be removed by coding the data with entropy encoders. Compression of this nature does not sacrifice any visual information carried by the original data and is hence a reversible process. This kind of compression is called lossless compression. The degree of compression achievable by lossless compression is quite limited. To achieve higher compression, remove the subjectively redundant information, which is information that is not visually obvious to the viewer and can be removed without severely reducing the subjective quality of the decoded video signal. This type of compression destroys some of the original image information, which cannot later be recovered. This compression is called lossy compression [3].

2.1.2 Compression Standards

MPEG stands for Moving Picture Experts Group. It is an ISO/IEC working group, established in 1988 to develop standards for digital audio and video formats. There are five MPEG standards being used or in development. Each compression standard was designed with a specific application and bit rate, although MPEG compression scales well with increased bit rates. They include [4]:

- MPEG-1

  Designed for up to 1.5 Mbit/sec Standard for the compression of moving pictures and audio. This was based on CD-ROM video applications, and is a popular standard for video on the Internet, transmitted as .mpg files. In addition, level 3 of MPEG-1 is the most popular standard for digital compression of audio known as MP3. MPEG-1 is the standard of compression for VideoCD, the most popular video distribution format thoughout much of Asia.

- MPEG-2

  Designed for between 1.5 and 15 Mbit/sec Standard on which Digital Television set top boxes and Digital Versatile Disk (DVD) compression is based. It is based on MPEG-1, but designed for the compression and transmission of digital broadcast television. The most significant enhancement from MPEG-1 is its ability to efficiently compress interlaced video. MPEG-2 scales well to High-Definition Television (HDTV) resolution and bit rates, obviating the need for an MPEG-3.

- MPEG-4

  Standard for multimedia and Web compression. MPEG-4 is based on object-based compression, similar in nature to the Virtual Reality Modeling Language. Individual objects within a scene are tracked separately and compressed together to create an MPEG4 file. This results in very efficient compression that is very scalable, from low bit rates to very high. It also allows developers to control objects independently in a scene, and therefore introduce interactivity.

- MPEG-7

  This standard is currently under development, is also called the Multimedia Content Description Interface. When released, the group hopes the standard will provide a framework for multimedia content that will include information on content manipulation, filtering and personalization, as well as the integrity and security of the content. Contrary to the previous MPEG standards, which described actual content, MPEG-7 will represent information about the content.

- MPEG-21

  This standard, also called the Multimedia Framework, has just begun. MPEG-21 will attempt to describe the elements needed to build an infrastructure for the delivery and consumption of multimedia content, and how they will relate to each other.

- JPEG stands for Joint Photographic Experts Group. It is also an ISO/IEC working group, but works to build standards for continuous tone image coding. JPEG is a lossy compression technique used for full-color or gray-scale images, by exploiting the fact that the human eye will not notice small color changes.

- JPEG 2000 is an initiative that will provide an image coding system using compression techniques based on the use of wavelet technology.

- DV is a high-resolution digital video format used with video cameras and camcorders. The standard uses Discrete Cosine Transform (DCT) to compress the pixel data and is a form of lossy compression. The resulting video stream is transferred from the recording device via FireWire, a high-speed serial bus capable of transferring data up to 50 MB/sec.

- H.261 is an International Communications Union (ITU) standard designed for two-way communication over Integrated Services Digital Network (ISDN) lines (video conferencing) and supports data rates which are multiples of 64Kbit/s. The algorithm is based on DCT and can be implemented in hardware or software and uses intraframe and interframe compression. H.261 supports Common Intermdiate Format (CIF) and Quarter Common International Format (QCIF) resolutions.

- H.263 is based on H.261 with enhancements that improve video quality over modems. It supports CIF, QCIF, SQCIF, 4CIF and 16CIF resolutions.

- DivX Compression is a software application that uses the MPEG-4 standard to compress digital video, so it can be downloaded over a DSL/cable modem connection in a relatively short time with no reduced visual quality.

2.1.3 MPEG Video Compression Technique

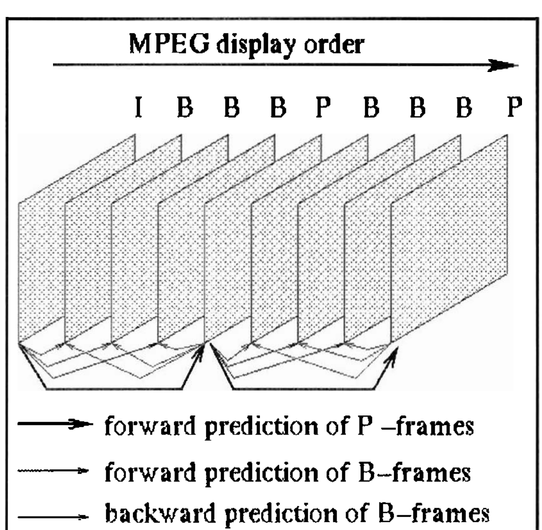

Figure 2.1 Video sequence [4]

The I-frames are intra coded, which is can be reconstructed without any reference to other frames. The P-frames are forward predicted from the last I-frame or P-frame, it is impossible to reconstruct them without the data of another frame (I or P). The B-frames are both, forward predicted and backward predicted from the last or next I-frame or P-frame, which there are two other frames necessary to reconstruct them. P-frames and B-frames are referred to as inter coded frames which mean a Java program must buffer at least three frames, one for forward prediction and one for backward prediction. The third buffer contains the frame coming into being. As the figure shows the frame for backward prediction follows the predicted frame. That would require to suspend the decoding of B-frames till the next P- or B- frame appears. But fortunately the display order is not the coding order. The frames appear on MPEG data stream in such an order that the referred frames precede the referring frames.

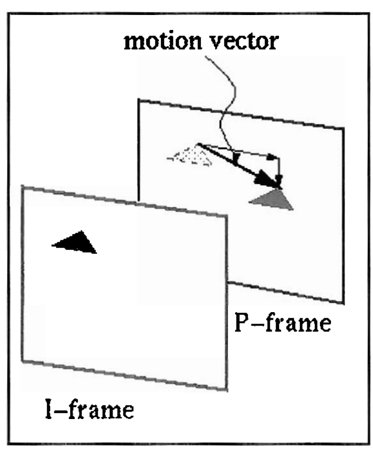

Figure 2.2 Reference frame compare to current frame [4]

Imagine an I-frame showing a triangle on white background. A following P-frame shows the same triangle but at another position. Prediction means to supply a MV which declares how to move the triangle on I-frame to obtain the triangle in P-frame. This MV is part of the MPEG stream and it is divided in a horizontal and a vertical part. These parts can be positive or negative. A positive value means motion to the right or motion downwards, respectively. A negative value means motion to the left or motion upwards, respectively.

But this model assumes that every change between frames can be expressed as a simple displacement of pixels. But the figure to the right shows this is not true.

The red rectangle is shifted and rotated by 5° to the right. So a simple displacement of the red rectangle will cause a prediction error. Therefore the MPEG stream contains a matrix for compensating this prediction error.

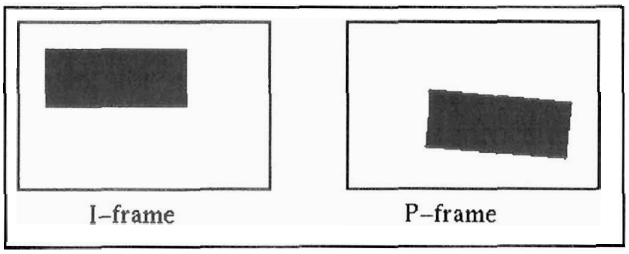

Figure 2.3 Prediction error [4]

Thus, the reconstruction of inter coded frames goes ahead in two steps:

1. Application of the MV to the referred frame
2. Adding the prediction error compensation to the result

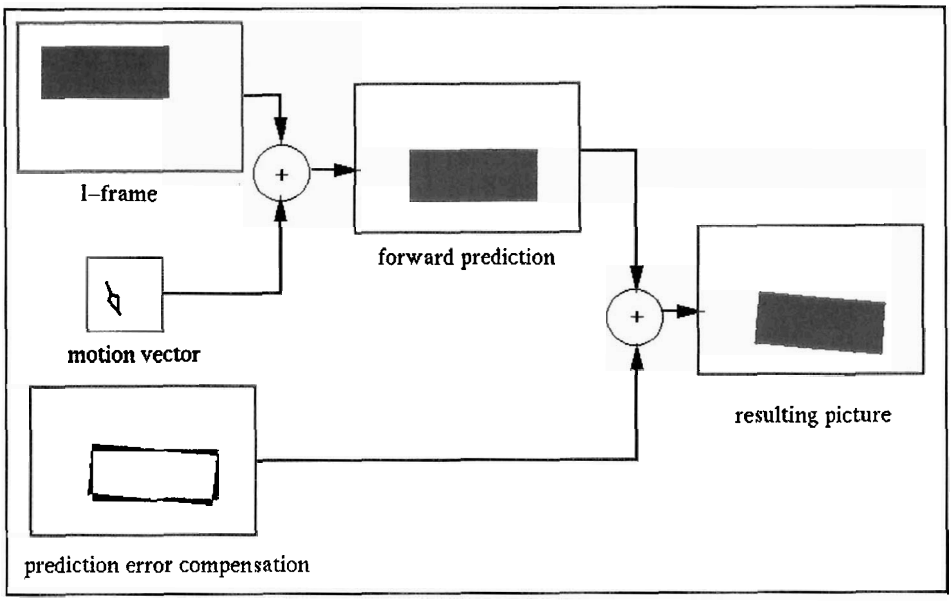

Figure 2.4 Reconstructed of inter coded frames [4]

2.1.4 Coding Of Moving Images

The concept of differential prediction discussed in Section 2.1.3 can be extended to coding moving video sequences. Most video sequences are highly redundant in the temporal domain. Within one video sequence, successive frames are usually very similar. The differences between successive frames are usually due to