INVESTIGATION OF BLOCK MATCHING ALGORITHM FOR VIDEO CODING

FAIZUL HADI BIN MOHAMAD JAMIL

THESIS SUBMITTED IN FULFILLMENT FOR THE DEGREE OF MASTER OF ENGINEERING (COMPUTER ENGINEERING)

SCHOOL OF ENGINEERING AND INFORMATION TECHNOLOGY
UNIVERSITI MALAYSIA SABAH
2013

PUMS 99:1

UNIVERSITI MALAYSIA SABAH

BORANG PENGESAHAN TESIS

JUDUL : ____
____
____

IJAZAH : ____
____

SAYA : ____________________________________ **SESI PENGAJIAN** : ________________________________
(HURUF BESAR)

Mengaku membenarkan tesis *(LPSM/Sarjana/Doktor Falsafah) ini disimpan di Perpustakaan Universiti Malaysia Sabah dengan syarat-syarat kegunaan seperti berikut:-

1. Tesis adalah hak milik Universiti Malaysia Sabah.
2. Perpustakaan Universiti Malaysia Sabah dibenarkan membuat salinan untuk tujuan pengajian sahaja.
3. Perpustakaan dibenarkan membuat salinan tesis ini sebagai bahan pertukaran antara institusi pengajian tinggi.
4. Sila tandakan (/)

☐ SULIT (Mengandungi maklumat yang berdarjah keselamatan atau kepentingan Malaysia seperti yang termaktub di AKTA RAHSIA RASMI 1972)

☐ TERHAD (Mengandungi maklumat TERHAD yang telah ditentukan oleh organisasi/badan di mana penyelidikan dijalankan)

☐ TIDAK TERHAD

Disahkan oleh:

____________________ ________________________
(TANDATANGAN PENULIS) (TANDATANGAN PUSTAKAWAN)

Alamat Tetap: ________________
____
____
__________________________ ________________________
(NAMA PENYELIA)

TARIKH: ______________ TARIKH: _________________

Catatan:
*Potong yang tidak berkenaan.
*Jika tesis ini SULIT dan TERHAD, sila lampirkan surat daripada pihak berkuasa/organisasi berkenaan dengan menyatakan sekali sebab dan tempoh tesis ini perlu dikelaskan sebagai SULIT dan TERHAD.
*Tesis dimaksudkan sebagai tesis bagi Ijazah Doktor Falsafah dan Sarjana Secara Penyelidikan atau disertai bagi pengajian secara kerja kursus dan Laporan Projek Sarjana Muda (LPSM).

DECLARATION

I hereby declare that the material in this thesis is my own except for the quotations, excerpts, equations, summaries and references, which have been duly acknowledged.

12 September 2012

Faizul Hadi bin Mohamad Jamil

PK2009-8065

CERTIFICATION

NAME	:	**FAIZUL HADI BIN MOHAMAD JAMIL**
MATRIC NO.	:	**PK20098065**
TITLE	:	**INVESTIGATION OF BLOCK MATCHING ALGORITHM FOR VIDEO CODING**
DEGREE	:	**MASTER OF ENGINEERING (COMPUTER ENGINEERING)**
VIVA DATE	:	**31 JULY 2013**

DECLARED BY

1. **SUPERVISOR**

   Prof. Dr. Ali Chekima — Signature

   ____

2. **CO-SUPERVISOR**

   Dr. Rosalyn R. Porle

   ____

ACKNOWLEDGEMENT

I would like to thanks to my main supervisor, Dr. Ali Chekima and my co-supervisor, Dr. Rosalyn R. Porle for some constructive advices they gave me, and also my friends for their valuable help. I sincerely acknowledge both of my supervisors for spending their valuable time for my research. I place my sincere thanks to my friends for their cooperation and assistance. Thanks for their willingness in spending time to have wonderful discussion on my work.

Faizul Hadi Bin Mohamad Jamil

15 September 2012

ABSTRACT

INVESTIGATION OF BLOCK MATCHING ALGORITHM FOR VIDEO CODING

The model of advanced video coding technique can be divided into two main parts that are the spatial model and the temporal model. The Spatial model exploits the redundancy in a single video frame (*I* frame) while the temporal model exploits the redundancy among frames (*P* frame). The temporal model deals with motion estimation (ME) and motion compensation (MC) algorithm with the matching technique called "Block Matching Algorithm" (BMA) to produce the next encoded video frame with motion vector. In this research, seven types of famous BMA techniques that are Exhaustive Search (ES), Three Step Search (TSS), New Three Step Search (NTSS), Simple and Efficient Three Step Search (SETSS), Four Step Search (4SS), Diamond Search (DS) and Adaptive Rood Pattern Search (ARPS) have been applied. These techniques have been used to analyze the video frames quality with different Macroblock size and different sequence of *I* and *P* frame and also with different search ranges of block prediction. Then, the best BMA algorithm technique will be chosen to develop a hybrid method that varies with the motion type of the video. This method will detect whether the video has a high, medium or low motion movement and it will determine the suitable Macroblock size (MB) and search range (p) for it. By using the hybrid method, it shows that the high motion video requires small MB size and larger search range to gain high encoded video quality while for the low motion video, larger MB size and small search range can be used to generate an acceptable video quality. Overall, this hybrid method can produce high quality video for the high motion movement video and it can reduce the computational complexity and yield to an acceptable quality for the low motion video.

ABSTRAK

Model untuk kaedah pengkodan video lanjutan boleh dibahagikan kepada dua bahagian utama iaitu model spatial dan juga model temporal. Model spatial mengeksploitasi pengulangan data pada satu gambar manakala model temporal pula mengeksploitasi pengulangan data di antara gambar-gambar. Model temporal ini bertindak dengan menggunakan algoritma jangkaan pergerakan (motion estimation) dan juga pergerakan pampasan (motion compensation) dengan teknik pengecaman yang dinamakan "Block Matching Algorithm" (BMA). Teknik ini digunakan untuk menghasilkan sebuah gambar video berpandukan pada gambar rujukan dan vektor pergerakan. Seterusnya, di dalam penyelidikan ini, tujuh teknik BMA yang terkenal iaitu Exhaustive Search (ES), Three Step Search (TSS), New Three Step Search (NTSS), Simple and Efficient Three Step Search (SETSS), Four Step Search (4SS), Diamond Search (DS) dan Adaptive Rood Pattern Search (ARPS) telah digunakan. Teknik-teknik ini telah digunakan untuk menganalisa kualiti gambar video dengan olahan pada saiz blok makro, olahan pada saiz carian blok pengecaman dan juga olahan susunan gambar jenis I dan gambar jenis P. Selepas teknik BMA yang terbaik telah dikenalpasti, satu kaedah hibrid dibangunkan untuk mengenalpasti pergerakan corak video. Kaedah ini dapat menentukan sama ada ianya video yang mempunyai banyak pergerakan, atau pergerakan yang sederhana atau pun mempunyai sedikit pergerakan. Kaedah ini dapat menentukan saiz blok makro serta saiz carian pengecaman yang sesuai berdasarkan corak video. Kajian ini mendapati bahawa kualiti video yang banyak pergerakan boleh diperolehi dengan menggunakan saiz blok makro (MB) yang kecil dan juga saiz carian jarak (p) yang besar manakala untuk video yang mempunyai sedikit pergerakan, memadailah dengan menggunakan saiz MB yang besar dan saiz carian jarak (p) yang kecil untuk mendapatkan output video yang baik. Secara keseluruhannya, kaedah hibrid ini dapat bertindak berpandukan kepada corak pergerakan video dimana ia dapat menghasilkan kualiti video yang lebih baik untuk video yang mempunyai pergerakan yang tinggi dan juga dapat mengurangkan pengiraan kompleks serta menghasilkan kualiti video yang baik untuk video yang mempunyai sedikit pergerakan.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

4SS	Four Step Search
ARPS	Adaptive Rood Pattern Search
BMA	Block Matching Algorithm
CIF	Common Intermediate Format
CODEC	Coder-Decoder pair
DCT	Discrete Cosine Transform
DS	Diamond Search
DWT	Discrete Wavelet Transform
ES	Exhaustive Search
FBSME	Fixed Block Size Motion Estimation
IDCT	Inverse Discrete Cosine Transform
LDSP	Large Diamond Search Pattern
MAD	Mean of Absolute Difference
MB	MacroBlock
MC	Motion Compensation
ME	Motion Estimation
MSE	Mean Squared Error
MV	Motion Vector
NTSS	New Three Step Search
PSNR	Peak Signal-to-Noise Ratio

QCIF	Quarter Common Intermediate Format
RGB	Red, Green and Blue Colour space
ROS	Region Of Support
SAD	Sum of Absolute Difference
SDSP	Small Diamond Search Pattern
SETSS	Simple and Efficient Three Step Search
TSS	Three Step Search
URP	Unit Size Rood Pattern
VBSME	Variable Block Size Motion Estimation
YUV	a Luma component (Y) and two Chrominance component (U, V)
YCbCr	a Luma component (Y) and two Chrominance components (Cb, Cr)
ZMP	Zero Motion Prejudgement

LIST OF APPENDIX

CHAPTER 1

INTRODUCTION

1.1 Overview

Nowadays, the application of image and video has become popular around the world to interpret data information. With the massive growth of multimedia and communication, people can view and play multimedia data anywhere on any available display devices such as television, computer, mobile phone, cinema, portable media player, tablet, and others. Not only that, the user can also watch movies, play video songs, view CCTV video on security systems, view their friend's video while interacting with them and many more through the internet, satellite TV, or from a single storage or memory device. It is a fact that most of the video applications are used for entertainment applications. With the growth of internet technology, the images and videos can be accessed easily from any image or video web hosting like 'flickr', 'youtube', and the famous personal social website 'facebook'. Users are free to upload and download their personal images or videos through the website. They also can access any pictures and videos that are uploaded by other users. The famous video web hosting 'youtube' contains a huge number of various videos where the user can view various videos and they are free to leave comment on it (Chatzopoulou *et al.,* 2010).

Almost all of the images or videos are represented in binary number. Each pixel of an image has its own value to represent the image information on that coordinate. An image might have a large number of pixels in vertical and horizontal direction. More number of pixels requires more data storage. A raw image requires a huge data storage size and it is extremely too huge for a raw video. By using the raw video, the 'youtube' website might not be able to store the entire video database due to its huge storage size. The user also might not be able to watch the video smoothly because the current transmission technology cannot transfer a huge file at very high speed. By compressing the raw video, it can reduce the video storage size and yield to faster transmission (Ghanbari, 2003). Because of this fact, the development of video compression technology is growing rapidly.

Many compression and coding techniques have been proposed to improve the image and video compression efficiency. Although the storage capacity of hard disks, flash memories and optical media are now becoming higher and cheaper than before, the use of video compression is still relevant because the uncompressed video requires extremely huge storage size which makes it unsuitable for current storage device usage and also current data transmission (Richardson, 2003).

The video compression/coding standards have been developed with lossy compression based so its storage size can be smaller than the lossless compression type (Jain, 1981). However, the generated compressed video data from lossy compression might not be identical to the original raw data. Some of the data might be lost and yield lower picture quality compared to its original video. As the users nowadays are always demanding to get higher video quality, it becomes necessary to compress video without compromising too much on quality while maintaining an acceptable frame rate. Thus, the field of video compression is rapidly gaining much attention in research and development, and every now and then, new video compression techniques are introduced as the demand for higher quality of video rapidly increases with cheaper hardware cost.

Many image and video compression techniques have been standardized since the last two decades to gain an efficient image and video compression ratio, its quality and its complexity of the algorithm. The International Standards Organisation (ISO) and the International Telecommunications Union (ITU) are the standard bodies which took the responsibility to develop a series of visual communications industry standard format such as JPEG and JPEG2000 for still image, and MPEG-1, MPEG-2, and MPEG-4 for the video codec standard (Richardson, 2002). The video codec consists of two main parts that are the encoder and the decoder. The standardization of the video codec is only formed at the decoder part, not at the encoder part. The encoder part is normally much more complex than the decoder part (Kalva, 2006). Normally, all the images and video standards can be viewed and played on the computer because almost all of the computer operating systems have the media codec standards. The success of digital TV and DVD video is based upon MPEG-2 standard. This technology has proved its efficiency in video application but now, this standard seemed unsuitable for high performance video application like High Definition video such as HD-DVD (Kobota, 2004). The "MPEG-4 part 10" or "H.264/AVC" standard has been proposed to achieve a high artificial intelligence technique to get the higher quality video and higher compression ratio especially for the high definition video for user satisfaction (Chen *et al.*, 2006).

Video coding basically exploits the spatial and temporal redundancy by removing the redundant data within and among the video frames. The data removal within a frame or spatial compression is the same as the image coding technique. The temporal redundancy attempts reduction of the redundant data among the video frames by exploiting the similarities between the neighbouring frames and generates a prediction frame of the current frame (Richardson, 2003). The prediction frame is generated from one or more previous or future frames or combination of it. A frame can be divided into several 4x4, 8x8, or 16x16 blocks for each frame partition to represent the data. Theoretically, the data difference between two frames can be coded as the moving vector (MV) at each block of the reference frame. The MV represents the displacement between the current block and the best matched block and each block will have its own direction for the next frame. Basically, the MV is created from the difference or residual of two successive frames with the constraint of criterion cost function like SAD (Sum of Absolute Difference), MAD (Mean of Absolute Difference) or MSE (Mean Square Error).

One of the popular study and research is on a part of the video coding technique which is to improve the Motion Estimation (ME) algorithm or the Block Matching Algorithm (BMA). This technique predicts the motion of a block of pixels between two or more frames in sequence. It will try to search the best matched block within the determined search area. The simplest searching technique is full search (FS) or sometimes is called Exhaustive Search (ES) (Manikandan *et al.*, 2006). This technique calculates every pixel point with the matching criterion to find the best motion vector and hence, it generates high computational complexity and it is not practical for real-time video applications and also not suitable for low power consumption devices such as mobile phone (Al-Mualla *et al.*, 2002). For this reason, the other techniques of BMA have been proposed to reduce the computational complexity such as Three Step Search (TSS), New Three Step Search (NTSS) (Li *et al.,* 1994), Simple and Efficient Three Step Search (SETSS) (Lu and Liou, 1997), Four Step Search (4SS) (Po and Ma, 1996), Diamond Search (DS) (Zhu and Ma, 2000), Adaptive Rood Pattern Search (ARPS) (Nie and Ma, 2002) and others. There are also other new techniques of BMAs which have been proposed by many researchers to improve the previous techniques regarding to its image quality and simplicity of the algorithm.

A number of video coding techniques have been proposed mainly to improve the coding performance regarding to its compression efficiency, high quality video and simplicity of the algorithm (Liyin *et al.*, 2010). Satyanarayana and co-researchers (2009) proposed an improved algorithm by the prediction of the block of search range and block size by comparing the movement motion of the video. Liu and Chong (2010) introduced a motion estimation algorithm which adaptively chooses different algorithm templates according to movement complexity and module precision. Both techniques have the same objective that is to improve the search speed and give acceptable quality of the reconstructed video. This research proposed a hybrid video encoder and decoder that are based on video motion types that are high, medium and low. The output video will have an appropriate video quality and acceptable computational complexity depending on the video motion types.

1.2 Problem Statement and its Significance

As stated before, the original raw video requires extremely large storage size and it requires high transmission bandwidth during video streaming. For example, to transmit a normal 90 minutes of coloured movie that has thirty frames per second (30 frames/s) with 720 x 480 pixels dimension, it will require around 162000 frames to be processed. Each frame has 720 x 480 pixels and each pixel contain three coloured components that are Red (*R*), Green (*G*), and Blue (*B*) where each of it requires 1 byte. A pixel will require 3 bytes and a frame will require 3 x 720 x

480 which is equal to 1,036,800 byte. Hence, its video storage size will be approximately 162 Gigabytes for a 90 minutes movie. This is not suitable for current storage devices and also for transmission purposes. To reduce the video storage size, the original raw video needs to be compressed by applying several processes to remove the spatial and temporal redundancies which can be achieved in the encoder part. But the compressed or encoded video will have lower picture quality compared to its original raw video. Normally, higher compression ratio will cause lower picture quality and vice versa. The storage size of a raw video can be computed by using Equation (1.1) where (Fr/s) is number of frames per second, M and N is the number of pixels in horizontal and vertical, and t indicates the time taken to play movie in second.

$$\textit{Size of video (byte)} = 3 \times (Fr/s) \times M \times N \times t \qquad (1.1)$$

Nowadays, with the rapid growth of video technology, the video picture quality is becoming one of the important issues. Users always want to watch high quality video smoothly. With the evolution of video coding which is done by many researchers, new technique of video coding is introduced to gain better compression ratio and improve its picture quality at the same time. High compression ratio gives an advantage to the storage size and transmission purpose while high picture quality gives an advantage for user satisfaction and also video identification.

The research in video coding, especially in BMA technique, has been conducted for two to three decades mainly to find the best matched block at the next frame within the search area. The traditional Exhaustive Search (ES) technique will exhaustively evaluate all possible blocks over the determined search window to find the best match. It is the most straight forward strategy and generates better image quality compared to the other search techniques, but its number of computation to find the best matched MB is extremely high which is not practical to use. Normally, most of the data among the video frames have zero motion movement. The other BMA techniques generally have more complex searching strategies and generate lower picture quality compared to the ES technique but it reduces the number of searching candidate points within the search window. The lower number of searching points may yield to an efficient algorithm and faster processing time. Basically, the idea of new technique of BMA is proposed to achieve one or more of the three objectives that are to get higher compression ratio, to reduce the computational complexity and to gain more accurate motion, hence, higher picture quality (Liyin *et al.*, 2010).

1.3 Research Objectives

The primary objective of this research is to investigate and develop an efficient Block Matching Algorithm (BMA) technique for video coding. The broad objective may be expressed in term of a number of aims:

1. To analyze the video coding technique by performing seven different types of BMA techniques that are Exhaustive Search (ES), Three Step Search (TSS), New Three Step Search (NTSS), Simple and Efficient Three Step Search (SETSS), Four Step Search (4SS), Diamond Search (DS), and Adaptive Rood Pattern Search (ARPS). The best BMA is chosen based on the combination of video quality and the computational complexity of the encoded video.

2. To develop an algorithm to examine the most suitable and practical MB size and search area for each video motion type. Hence, the best or highest picture quality and the lowest computational complexity from the BMA technique is examined by applying different macroblock (MB) sizes and different search area blocks to three types of video motions that are high, medium and low motion video.

3. To develop a Hybrid BMA algorithm that varies according to the motion types whether the video has high, medium or low motion video. The encoding process will be based on the motion type so that it can improve its picture quality and reduce the computational complexity.