GAME THEORETICAL MODELS AND ALGORITHMS FOR RATE CONTROL IN VIDEO COMPRESSION

by

JIANCONG LUO

Presented to the Faculty of the Graduate School of

The University of Texas at Arlington in Partial Fulfillment

of the Requirements

for the Degree of

DOCTOR OF PHILOSOPHY

THE UNIVERSITY OF TEXAS AT ARLINGTON

August 2005

ACKNOWLEDGEMENTS

I am deeply indebted to my supervisor Prof. Dr. Ishfaq Ahmad whose helps, stimulating suggestions and encouragement helped me in all the time of research for writing of this dissertation. Without his invaluable guidance through these years in HKUST and UTA, my doctorial studies would not have been completed.

Dr. Alp Aslandogan, Dr. Hao Che, Dr. Hua-mei Chen and Dr. Jean Gao have spent their precious time in reviewing this dissertation and provided me many useful suggestions. I would like to express my gratitude to them for all their helps and supports. I would like to thank Dr. Weiguo Zheng, Dr. Yu Sun for their kind helps and useful discussions. I also would like to thank my colleagues in CSE Multimedia Lab for their advices, discussions and warm helps all these years.

I dedicate this dissertation to my parents, who have provided me the greatest help and constant encouragement. Without their support, I would not have gone through all the difficult times. Especially, I would like to give my special thanks to my wife Yamin, whose patient love enabled me to complete this work.

July 14, 2005

ABSTRACT

GAME THEORETICAL MODELS AND ALGORITHMS FOR RATE CONTROL IN VIDEO COMPRESSION

Publication No. ______

Jiancong Luo, PhD.

The University of Texas at Arlington, 2005

Supervising Professor: Ishfaq Ahmad

This thesis investigates game theory based rate control algorithms for optimizing the bit allocation in video compression. The first algorithm utilizes the cooperative bargaining game in a MB level rate control algorithm to optimize the perceptual quality while guaranteeing "fairness" in bit allocation among macroblocks. The algorithm first allocates the target bits to frames based on their coding complexity; a method to estimate the coding complexity of the remaining frames is proposed. Next, macroblocks of a frame play cooperative games such that each macroblock competes for a share of resources (bits) to optimize its quantization scale while considering the

human visual system (HVS) perceptual property. Since the whole frame is an entity perceived by viewers, macroblocks compete cooperatively under a global objective of achieving the best quality with the given bit constraint. The major advantage of the proposed approach is that the cooperative game leads to an optimal and fair bit allocation strategy based on the *Nash Bargaining Solution*. Another advantage is that it allows multi-objective optimization with multiple decision makers (e.g., macroblocks). The algorithm achieves accurate bit rate with good perceptual quality, and to maintain a stable buffer level. The second algorithm based on a non-cooperative strategic game is aimed for video object level bit allocation. We formulate a two-player bi-matrix game, in which the utilities of the players are pre-determined by a set of available strategies (i.e., the possible quantization parameters). The game is non-deterministic in which the players' strategies are bound to a probability distribution over the set of available actions. The outcome of the game is a mixed strategy Nash equilibrium. The proposed algorithm achieves accurate bit rate regulation and smooth buffer occupancy.

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS

LIST OF TABLES

CHAPTER 1

INTRODUCTION

With the increasing demands of multimedia applications, such as HDTV system, video conferencing, video on demand and other kinds of video based interactive multimedia services, various digital videos need to be stored, processed and transmitted. Problem raises that the storage and bandwidth is not satisfied the requirement of storing and transmitting the raw video data.

For instance, the Federal Commission (FCC) requires that the HDTV transmitted signal fits inside the same 6 MHz channel spacing as is currently used for today's NTSC transmissions. This requirement implies that 1.5Gbits per second of video data need to be transmitted via a 6 MHz channel, which can support only about 20Mbits per second.

Another example is video phone applications. The common modem available today for PSTN can transmit at a bit rate of 56kbps, if the line condition is very good. A common format for such an application is the quarter common intermediate format (QCIF), which has a dimension of 176×144 pixels. For a 24-bit color sequence, which uses 24 bits for one pixel (8 bits for each red, green and blue channel respectively), and assuming that the frame rate is 30 frames per second, the size of one second video is $176\times144\times24\times30=18.3$Mbits. That means that it requires transmitting video data of bite rate at 18.3Mbps via a channel with a bandwidth of 56kbps. It is a 330 times larger bandwidth than what is available with PSTN.

The storage and bandwidth requirements of the raw video data are nearly incomprehensible. For the sake of efficient bandwidth utilization, compression is inevitable. With the help of compression technology, the requirements can be reduced by a long way. Alternatively, instead of sending the raw video sequence at a given spatial and temporal resolution, one can send a higher resolution compressed video sequence in the same channel.

Due to the high demands for video application, several international video compression standards have been or are being developed during the last two decades, such as MPEG-1 [15], MPEG-2 [16] , MPEG-4 [17], H.261 [20], H.263 [21] and H.264 [23]. A series of different disciplines of digital signal processing principles, such as coding theory, rate-distortion theory, prediction techniques and control theory, have been applied cooperatively to improve the compression ratio, coding efficiency, picture quality and complexity, etc. In this work, we will focus on the bit rate control problem and aim to design rate-distortion efficient rate control schemes for frame-based and object-based video coding.

The exchange of video information between remote sites requires that the digital video be encoded and transmitted through specified network connections. Due to the variable amount of redundancy and irrelevancy of video contents, the amount of compressed video data varies in a rather unpredictable manner. Therefore, the compressed video data rate is inherently variable and inconsistent with the channel bandwidth. It may be lower than the available bandwidth, which leads to unnecessary degrading of visual quality and waste of bandwidth; or higher than the bandwidth constraints, resulting in traffic congestion and loss of data. Rate control is a mechanism that regulates the compressed video data rate to meet the channel requirement. It is critical in determining the encoding efficiency and video quality. This chapter will give an overview of video compression schemes with an emphasis on the rate control techniques.

1.1 Video Coding Architecture

Although many different video compression standards exist, they adopt a similar architecture: Hybrid DCT/DPCM coding scheme. This scheme employs Discrete Cosine Transform (DCT) and quantization to exploit the spatial correlation while utilizes inter-frame *Differential Pulse Coding Modulation* (DPCM) coding techniques to reduce the temporal redundancy. The common video processing functions in Hybrid DCT/DPCM coding are: DCT and Inverse DCT (DCT/ DCT^{-1}), Quantization and De-quantization (Q/Q^{-1}), Entropy Coding and Decoding (EC/ED), Motion Estimation (ME) and Motion Compensation (MC), and Rate Control (RC). Figure 1.1 shows a generic block diagram for hybrid DCT/DPCM coding scheme. “FS” in Figure 1.1 symbolizes the frame storage.

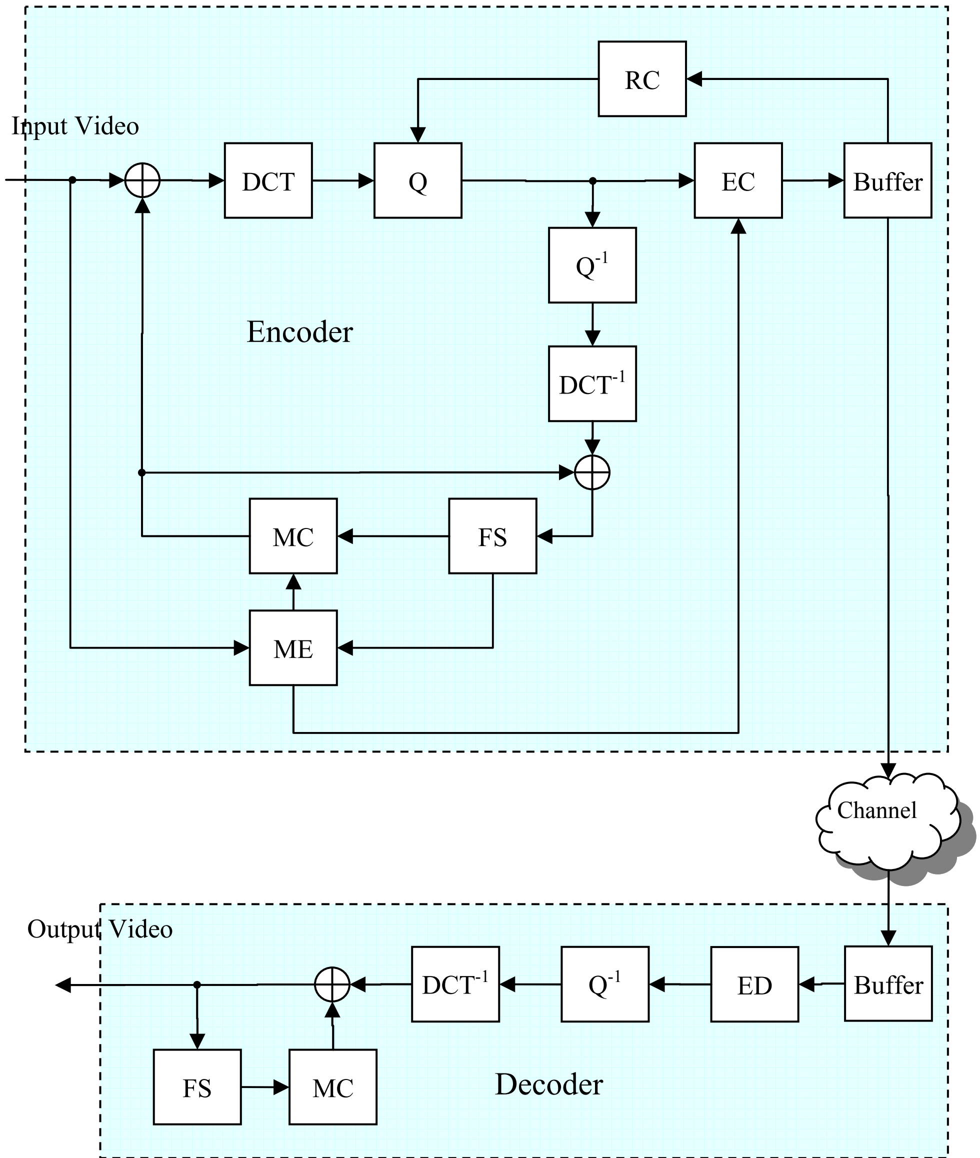

Figure 1.1 Block diagram of Hybrid DCT/DPCM coding scheme

Discrete Cosine Transform (DCT) is a technique converting signals to frequency components. For 2D DCT transform, the lower-right DCT coefficients relate to low spatial frequencies within the image block and the upper-left DCT coefficients relate to the high spatial frequencies. Based on the human visual systems criteria, the human eyes are more sensitive to the components with low spatial frequencies than those with high spatial frequencies. This property is used to remove the subjective redundancies. Figure 1.2 depicts the variance distribution of 8x8-block DCT coefficients. Coefficients with small value are less significant for the reconstruction of the image blocks than coefficients with large variances. We can observe that the energy is concentrated around the upper-left corner, which corresponds to the low spatial frequencies. Most of the DCT coefficients are small enough to vanish in the quantization step and need not to be encoded and transmitted.

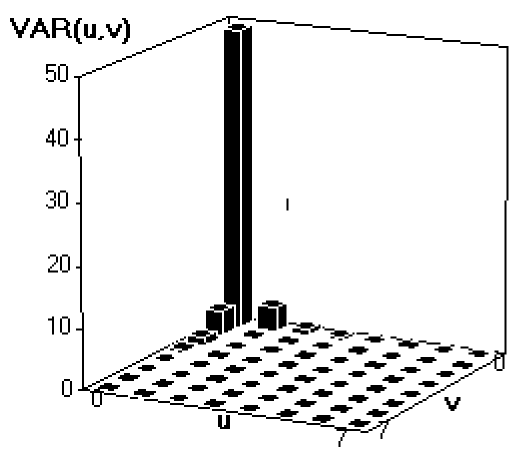

Figure 1.2 Distribution of DCT coefficients

Quantization is the major source of data loss in DCT based compression algorithms. By converting amplitudes that fall in certain ranges to one in a set of pre-determined levels, quantization reduces the amount of information required to represent the amplitudes. Although different types of quantization are proposed, for simplicity, all the standard image compression algorithms use linear quantization where the step size quantization levels are constant. Quantization in the frequency domain has many advantages over quantization in pixel domain. Quantization in pixel domain leads to "contour" artifact where small amplitude changes in a smoothly gradient area cause step-sized changes in the reconstructed amplitude. Some other quantization schemes have been reported in [55] [59].

Entropy coding is a lossless process based on statistics of the image or the motion picture sequence to be compressed. Although there is a variety of different implementation of entropy coding in the existing video compression standards, the underlying theory of entropy coding is to encode the most frequently occurring patterns with the shortest code word and encode the less frequently appearing patterns with longer code words. In this manner, data can be compressed by a factor of 3 or 4. Entropy coding for video compression applications is a two step process: Zero Run-Length Coding (RLC) and Huffman coding. RLC data is an intermediate symbolic representation of the quantized bins which utilizes a pair of numbers. The first element is the number of consecutive zeros while the second element represents the non-zero value immediately following the previous run of zeros. For example, the RLC code (3, 6) represents the sequence (0, 0, 0, 6) of numbers. Huffman coding assigns a variable length code to the RLC data, producing variable length code word. The Huffman coding is a table-lookup process. A Huffman tables is pre-computed based on statistical properties of the image (as it is in JPEG) or pre-determined if a default table is to be used (as it is in H.261 and MPEG). The same table as used in encoding is used to decode the bitstream data. Since entropy coding produces variable length code word, the digital compressed stream has no specific boundaries or fixed length.

In general, consecutive frames in a motion video tend to be highly correlated, i.e., the frames only change slightly over a small period of time, which implies that the difference between these frames is small. For this reason, compression ratios for motion video sequences may be increased by encoding the difference between two or more successive frames.

In contrast, motions of objects increase the difference between frames. This implies that more bits are required to encode the sequence. To address this issue, motion estimation is utilized to determine the displacement of an object.

Motion estimation (ME) is the process of finding motion vectors during the encoding process. In the block matching motion estimation, a frame is divided into 16x16 blocks, called macro-block. Motion vectors encapsulate the displacements between the macro-blocks in the current frame and the best correlated macro-blocks in the past and/or future reference frames containing previously decoded pixels that are used to form the prediction and the error difference signal. Forward motion vectors refer to correlation with previous pictures. Backward motion vectors refer to correlation with future pictures. Full search algorithm exhaustively evaluates all possible displacements within the predetermined or adaptive search range to find out the best matching 16x16 block in the reference frame for a macro-block of the current frame. The advantage of full search block matching algorithm is accuracy. It always leads to the global minimum by looking at everywhere in the search area. The defect is also obvious. The computational complexity for a full search block matching algorithm is very heavy. The computational complexity of full search algorithm can be 60~80% of total computations in the encoding process involves in the purpose of full search algorithm [32] [55]. In order to reduce the computational complexity of the encoding algorithm, a number of fast ME (FME) algorithms are proposed. Extensive survey of FME has been included in [31]. The representative FME algorithms include [14] [24] [29] [36] [38] [40] [56] [70] [71] [79] [80].

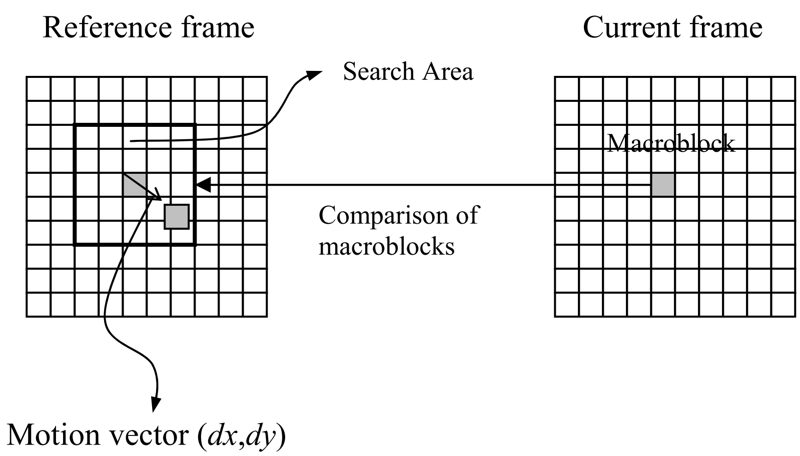

Figure 1.3 Motion vector

As mentioned above, video compression produces variable bit rate output streams due to the variable amount of redundancy and irrelevancy in the input sequences. This characteristic of video encoders inherently mismatches the constrained media channels. For example, the constant bit rate (CBR) channels require that a constant number of bits per time unit be sent to the channel, and the variable bit rate (VBR) channels require that a set of specified traffic parameters be met since otherwise data losses may occur. In both cases, bit rate regulation is necessary. The objective of rate control algorithms is to regulate the output bit rate to meet the channel constrains. Meanwhile rate control algorithms optimize the visual quality subject to the available transmission data rate and certain quality of service (QoS) requirements, such as delay, quality fluctuation, number of frame skipped, etc.

Although the video coding standards designate the syntax of encoded presentation, the methods of generating the standard conformed bitstream is not defined. Rate control is critical to the visual quality and rate-distortion performance. It is not standardized and is an open issue for research. Our research focuses on the rate control problem.

1.2 Digital Video Coding Standards

Standardization of video compression facilitates manipulation and storage of motion video as a form of computer data, and its transmission over existing and future networks or over terrestrial broadcast channels. Furthermore, interchange and interoperation of video contents become possible only when the coded representations of video contents conform to the standards. To this end, ISO/IEC has published MPEG-

1/2/4 targeting on different applications and ITU also proposed H.261/263/264, mainly aiming to real time video conferencing applications.

Moving Picture Experts Group (MPEG) is an ISO/IEC working group, established in 1988 to develop standards for digital audio and video formats. Five MPEG standards are being used or in development. Each standard was designed for a specific application and bit rate, although MPEG compression scales well with increased bit rates. Telecommunication Standardization Sector of the International Telecommunications Union (ITU-T) also defined its own video compression standards. These standards are mainly focus on real time video applications, such as video phone and video conference. In the sequel, we will introduce the major video coding standards in the chronicle manner.

The first video coding standard is H.261, which was ratified in 1990 by ITU. The increasingly obsolete H.261 was specifically designed for transmission over ISDN lines in that data rates are multiples of 64kbit/s. The standard supports CIF and QCIF video frames at resolutions of 352x288 and 176x144 respectively. The coding algorithm is a hybrid of inter-picture prediction, transform coding, and motion compensation. The data rate of the coding algorithm was designed to be able to be set to between 40Kbits/s and 2Mbits/s.

MPEG-1 was completed in 1992 as the first MPEG standard for the compression of moving pictures and audio. It was designed for CD-ROM video applications and a bit rate of up to 1.5 Mbit/sec. MPEG-1 is also a popular standard for video on the Internet, transmitted as .mpg files. MPEG-1 is the standard of compression for Video CD, the most popular video distribution format throughout much of Asia. Comparing with H.261, MPEG-1 employed some advanced techniques, such as bi-directional prediction frame and half-pixel motion compensation.

MPEG-2 is a standard on which Digital Television set top boxes and DVD compression is based. Ratified in 1994, MPEG-2 was based on and was a superset of MPEG-1 but specifically designed for the compression and transmission of digital broadcast television which applied higher bit rate (between 1.5 and 15 Mbit/sec). MPEG-2 supports some significant enhancement from MPEG-1, including efficient compression of interlaced video, data partitioning, SNR scalability, spatial and temporal scalability, etc. MPEG-2 supports flexible picture size and bit rate. It scales well to HDTV resolution and bit rates, obviating the need for an MPEG-3.

H.263 is a video compression standard designed by the ITU-T as a low bit-rate encoding solution for videoconferencing. It was first designed to be utilized in H.323 based systems, but now is finding use in RTSP (streaming media) and SIP (Internet conferencing) solutions as well. H.263 provides a suitable replacement for H.261 at all bit-rates. It has been superseded by H.263v2 (a.k.a. H.263+ or H.263 1998). H.263 provides advanced coding modes, such as half-pixel motion compensation, optimized run length coding (VLC) table, unrestricted motion vector, PB frame mode, advanced motion vector prediction, arithmetic coding and enhanced error resilience, etc. H.263 has a substantial improvement in video quality and coding efficiency compared with H.261.

MPEG-4 was initiated for multimedia and Web compression. The motivation of

MPEG-4 is the convergence various multimedia technologies, such as computer graphic, TV and film industry, 3D image and video and telecommunications. MPEG-4 supports a wide range of bit rates (5kbps-5Mbps) and applications such as video conferencing, mobile videophone, multimedia cooperative work, tele-teaching and, last but not least, games. What distinguishes MPEG-4 from other video coding standards is the concept of object-based video coding. The new or improved functionalities of MPEG4 have been clustered in three sets - content-based interactivity, compression and universal accessibility [53]:

Content-based interactivity

- Content-based multimedia data access tools

- Content-based manipulation and bitstream editing

- Hybrid natural and synthetic data coding

- Improved temporal random access

Compression

- Improved coding efficiency

- Coding of multiple concurrent data streams

Universal access

- Robustness in error-prone environments

- Content-based scalability

The MPEG-4 video compression scheme is based on the block-based concept as used in the MPEG-1, MPEG-2 and H.261/263, but has been extended to support arbitrarily-shaped video objects. Arbitrarily-shaped video objects are split up into macroblocks within a bounding box. An alpha-plane represents the shape of a video object. Alpha-plane can be gray-scale or binary image. Visual objects can be natural video object or synthetic objects. Every object is encoded and decoded by a difference encoder and decoder instance and may use different coding options. The video object at a time instance is called video object plane (VOP). Individual objects within a scene are encoded and decoded separately by different encoder and decoder instance and multiplexed together to create an MPEG4 file. This results in very efficient and very scalable compression. It also allows developers to control objects independently in a scene, and therefore introduce interactivity.

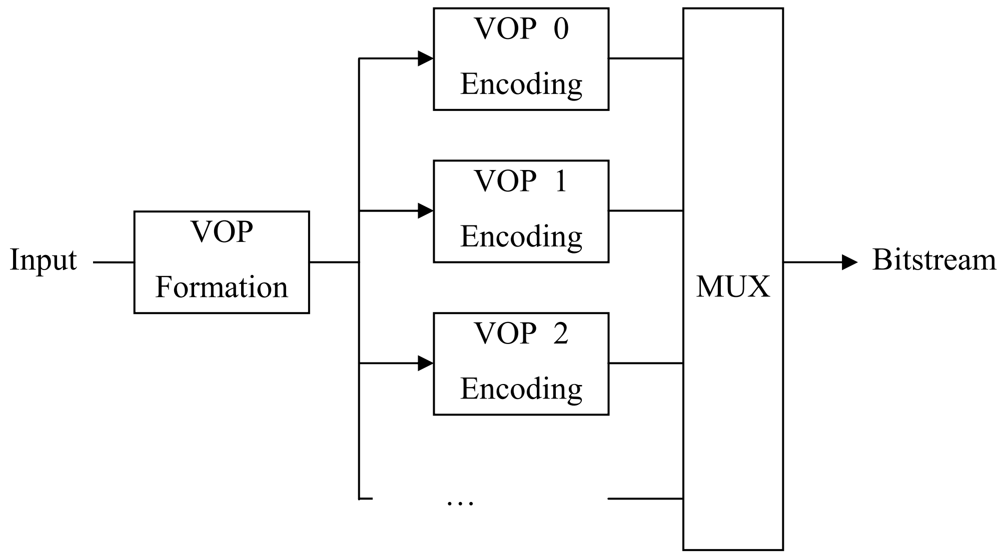

Figure 1.4 MPEG-4 Encoder Architecture

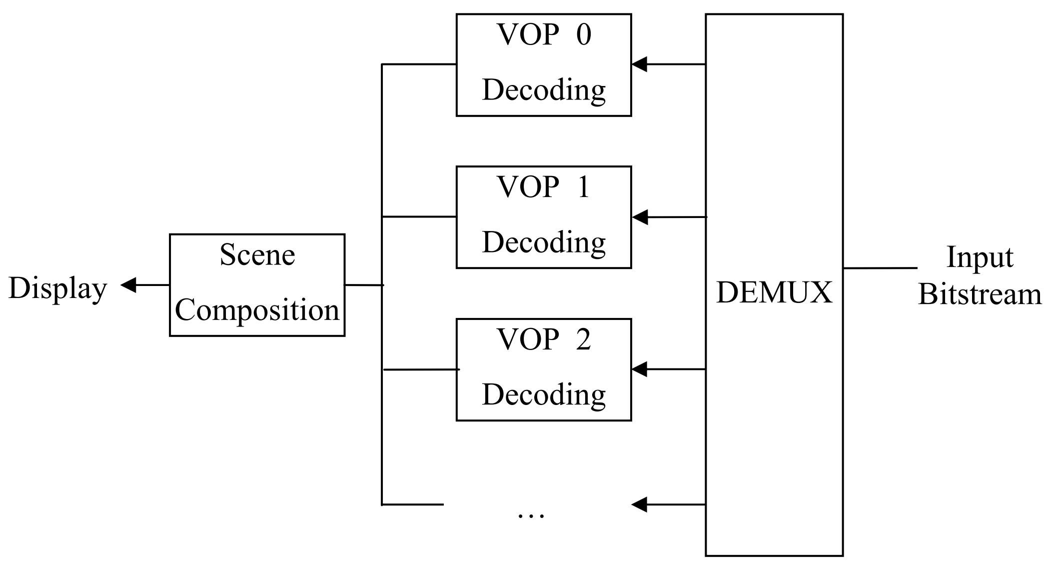

Figure 1.5 MPEG-4 Decoder Architecture

Advanced Video Coding (AVC) is a high compression digital video compression standard written by the ITU-T Video Coding Experts Group (VCEG) together with the ISO/IEC Moving Picture Experts Group (MPEG) as the product of a collective effort known as the Joint Video Team (JVT). This standard is also known as H.264 and identical to ISO MPEG-4 part 10. As the latest video compression standard, AVC is already widely used for videoconferencing. It has also been preliminarily adopted as a mandatory part of the future DVD specification known as HD-DVD, now under development by the DVD Forum. A number of broadcasters in Japan and Korea have announced future support for the codec, and it is under consideration for other broadcast use -- for example, it is under consideration in the United States' Advanced Television Systems Committee (ATSC) and in Europe's Digital Video Broadcast (DVB) standards bodies. In the wireless world, it is under consideration for adoption by the 3rd-Generation Partnership Project (3GPP).

AVC employs a series of new technologies. The major enhancements include: multiple reference frame motion compensation prediction in inter mode, Intra prediction, multiple block size from 4x4 to 16x16, sub-pixel motion compensation (down to 1/8 pixel), rate-distortion optimized (RDO) mode decision and motion estimation, 4x4 integer DCT/IDCT, support both Context-based Adaptive Binary Arithmetic Coding and Variable Length Coding, advanced in-loop filter, etc. AVC offers significantly better coding efficiency than the previous ITU and MPEG standards.

1.3 Rate Control Overview

1.3.1 Principles

Limited bandwidth network usually is used as the infrastructure of the digital video communication and transmission. Though the network is capable to provide a wide rage service from narrow bandwidth to wide bandwidth, the video data generally are transmitted over a constant bit rate (CBR) channels. In CBR channels, a transmission buffer is used to accommodate the variable data rate nature of the compressed video. Therefore, the possibility of buffer overflow or underflow exists. In this channel mode, the target of the rate control is to regulate the output bit rate to avoid buffer overflow and underflow. Some network service, such as B-ISDN [9], also supports variable bit rate (VBR) channel. The buffer is not necessary in VBR channel since the network can accommodate the variable rate video data. But network congestion still can happen since the burst of peak rate of the video data may exceed the network capacity. In VBR channel, rate control is still required to avoid network congestion. Overall, the objective of rate control is to regulate the output bit rate of a video encoder to the bandwidth requirement, meanwhile, depending on the required data rate to maintain a satisfactory reconstructed quality.

In lossy video coding, rate-distortion function can well embody the behavior of video encoder. The output bit rate can be increased by increasing the distortion level, and vice versa. Figure 1.6 illustrates the rate-distortion relation of a generic video encoder. Therefore, by choosing the proper operation point on the rate-distortion curve, the output bit rate of and encoder can be controlled.

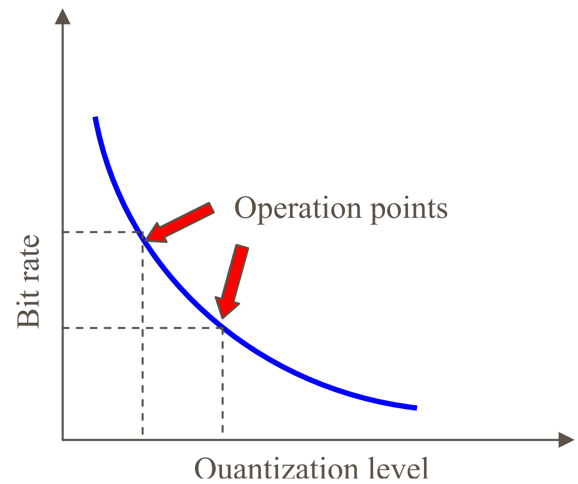

Figure 1.6 Rate-distortion relation of DCT-based video encoder

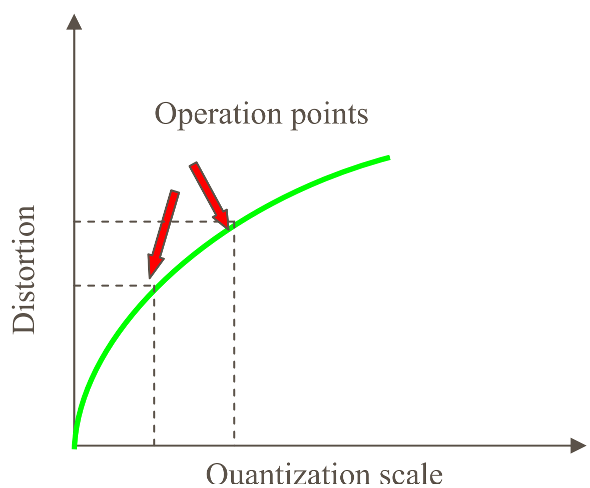

Figure 1.7 Distortion-quantization relation of DCT-based video encoder

Quantization is the major source of data loss in DCT based video compression thus is closely related to the distortion level of the reconstructed image. The output bit rate of a DCT-based video encoder is typically controlled by choosing a proper quantization scale. A larger quantization scale masks more DCT coefficients, therefore, saves more bits, but at the expense of picture quality. A finer quantization scale produces more bits but results in better perceptual quality.

Depend on the control granularity, rate control algorithms can be classified into the following categories:

(1) Frame-level rate control: a uniform quantization parameter (QP) is assigned to all the MBs in a frame. The control unit is a frame.

(2) Macroblock-level rate control: QPs for macroblocks can be different. The control unit is a macroblock.

For the multiple-video-object (MVO) coding, VOP-level rate control can be employed where different QPs can be assigned to video object plane (VOP), but a uniform QP is used for all MBs within a VOP.

Compare with frame-level rate control, Macroblock-level rate control has the following advantages: 1) macroblock-level rate control is able to achieve more accurate bit rate, since macroblock-level rate control use a smaller basic control unit, which generally leads to better responds to the change in video characteristics and model parameters. 2) With macroblock-level rate control, it is possible to compute QP for each macroblock to optimize the visual quality of a frame, whereas using a uniform QP may not be optimal in the rate-distortion (R-D) sense. The defects of macroblock-level rate control include: 1) it requires extra bits to encode the change of QP. In the other word, in the same bit rate, macroblock-level rate control assigns less bits to encode the DCT coefficients compare with frame layer rate control, which may lead to the lower visual quality, especially in low bit rate coding environment. 2) A macroblock-level rate control usually is more complicate that a frame-level rate control.

The rate control algorithms can also be categorized frame-based rate control and object-based rate control, depending on the encoding infrastructure. Frame-based rate control algorithms are employed in the frame-based video encoding such as MPEG-1 and MPEG-2, whilst the object-based rate control algorithms are utilized in the object-based video coding framework which supports the encoding the arbitrary shape object, such as MPEG-4. Compared with frame-based algorithms, the object-based rate control algorithms deals with additional issues including data for shape coding (coding of alpha plane), coding of scene description information and bit allocation for multiple objects.

For the object-based video coding such as MPEG-4, different rate control strategies can be employed. In the independent rate control, each object encoder has its own rate controller. The objects are encoded separately at independent constant bit rates or quality. In the joint multiple-video-object (JMVO) rate control, data rate of object encoders are jointly considered. Multiple object encoders share a total data rate. Efficient resource allocation among objects is required.

1.3.2 Rate Control Architecture

The following rate control architecture is commonly used in most of the current video applications: A buffer is placed between the encoder and the channel in order to absorb the variation of the output bit rate and stabilize the output bit rate.

Generally, the rate control algorithm consists of the following steps:

1. Target bit allocation

The target bit constraint allocated prior to the encoding of a frame, a VOP or a macroblock. The allocation can be based on the available bits, remaining frames, buffer status and the statistic of previous coded units.

2. Quantization parameter determination

A proper quantization parameter for the coding unit is obtained in this step. The quantization parameter is estimated such that the output bit rate meets the target bit constraints determined in step 1. Depending on the control granularity, the quantization parameter can be for a frame, an object or a macroblock. The popular approach of obtaining the quantization parameter is the utilization of R-D models.

3. Post-Encoding Processing

After encoding a frame, an object or a macroblock, the encoder status is updated in this step. These values include available bits, remaining coding unit, buffer occupancy, R-D model parameters, etc. The updated values are used for the encoding of the next coding unit.

Several aspects strongly affect the performance of a rate control algorithm.

– Quality fluctuations. Fluctuations in image quality between consecutive time instances can be visually annoying and thus should be reduced as much as possible.

– Buffer occupancy. Smaller buffer occupancy leads to lower delay.

– Buffer overflow. Buffer overflow occurs when the encoder output bit rate is too high such that the accumulate bits is more than the buffer capacity. In this case, the excess bits will be dropped. Buffer overflow always causes loss of data, therefore must be avoided.

– Buffer underflow. Buffer underflow usually occurs when the encoder output bit rate is lower than the dripping bit rate of the buffer. In this case, not enough bits are dripped to the channel, which results in sub-optimal use of the channel. It is not as critical as buffer overflow but should also be avoided.

– Traffic contract. In VBR networks, like ATM networks, packets transmitted outside the bounds established by the negotiated contract may be discarded if traffic congestion occurs resulting in loss of data which must be avoided.

1.3.3 Related Works

Several rate control algorithms have been proposed and utilized in video compression standards and applications. In the early rate control algorithms, such as H.263 Test Model Near-term version 5 (TMN5) [22], a fixed bit allocation for each frame is employed. The target bit budget for each frame is obtained by dividing the target bit rate by the frame rate. MPEG Test Model 5 (TM5) [18] rate control employs a hierarchical bit allocation scheme, the target number of bits is decided for the GOP layer first. Constrained by the GOP bit budget, target number of bits for a frame is then calculated. Target bits for I, P, B frame are allocated according to the complexity of the previous frame. With respect to the model-based quantization scale decision, early rate control algorithms, such as TM5, adopt a linear rate-distortion model. Thereafter, various rate-distortion models have been utilized to improve the accuracy of the quantization scale estimation. Chiang and Zhang proposed a quadratic rate-distortion model that can be applied to both DCT and wavelet-based coders [4]. An improvement of [4] has been proposed by Pan et al. to fine tune the bit allocation according to the position of a frame in a GOP [52]. Based on [4], Ngan, Meier and Chen have introduced a new constraint for the least-mean-square estimation of the model parameter of the rate-distortion function [49]. Cheng and Huang have proposed an adaptive piecewise linear model [2]. He and Mitra have modeled the rate and distortion as the functions of ρ, which is the percentage of zeros among the quantized DCT coefficients, and an optimized bit allocation has been proposed based on this model [10] [11] [12]. Chiang, Lee and Zhang have proposed a scheme for MB level bit allocation and two distinct models for high bit rate and low bit rate situations respectively [3]. The model used in high bit rate adapts the quantization scale with the energy of the block by using finer quantization for MB of flatter image regions. The model used in low bit rate maintains a near-constant quantization scales, in order to minimize the overhead bits for DQUANT, which is to define the change of quantization scale. Another MB level rat-distortion model has been addressed by Oehler and Webb [50]. Ribas-Corbera and Lei have proposed a rate-distortion model, which was adopted in the H.263+ testing model TMN8 [62]. Based on [62], Tsai and Hsieh have proposed to modify the encoding order of MBs to favor the more complex MBs [70]. For object-based video coding, Lee, Chiang and Zhang have proposed an algorithm that is scalable for various bit rates, spatial and temporal resolution [33] [34]. Vetro and Sun also developed a scheme for multiple video objects in [73] [74]. Lee et al. has proposed models for coded frames and objects as well as skipped frame and objects [35]. Other rate control algorithms have been proposed in [27] [38] [68] [78], etc. The above rate control algorithms reportedly achieve efficient bit rate regulation. What distinguishes our approach from these works is that we define a new criterion of optimizing the rate-distortion efficiency under the bit rate and "fairness" constraints, and propose an efficient solution.

Since the quantization scale can refer the distortion of the decoded frame, some rate control algorithms seek optimal quantization to maximize the aggregate perceptual quality or minimize the aggregate distortion subject to the bit rate constraints, using Lagrange multiplier or dynamic programming [58] [60] [62] [65] [77]. Although these algorithms address the rate-distortion optimization problem, the perceptual redundancy of human vision system has not been efficiently exploited. Moreover, the fairness of bit allocation among the macroblocks has not been discussed.

1.4 Research Methodology

Although rate control is a well studied topic in video compression and several efficient schemes have been widely used in standards, there still are different methods that have not been explored and improvement can be achieved by using new approaches. E.g. a few algorithms have considered the rate-distortion optimization problem; the perceptual redundancy of human vision system has not been efficiently exploited; the "fairness" of bit allocation has not been considered.

The objective of our research is to define a new criterion of optimizing the rate-distortion efficiency under the bit rate and "fairness" constraints, and propose an efficient solution. Towards this goal, our study applied cooperative and non-cooperative game theory approaches for rate control.

Game theory deals with the decision making problems. A game is a mathematical model of an interactive decision problem – a situation in which the outcome is determined by the choices of two or more parties, who are called decision makers or players. Game can be found in everyday life. E.g., when you are driving, you need to make decisions depending on the behaviors of the other drivers on the road, thus you are playing a game with other drivers; when you are bidding in an auction, your decision is based on how you value the item and the money in your pocket, sometimes estimation of other bidders' bids, thus you are playing with other bidders. Each player may of may not be fully informed about the other players' choices, and is assumed to react only in his/her own interests. Players can act simultaneously or sequentially, once or repetitively. Players may or may not communicate and may make binding agreements. In every game we look for a "solution", which is the decisions each player makes and their corresponding outcomes.

Two classes of games are distinguished, called *cooperative* and *non-cooperative*. In *cooperative* games, players can communicate and make binding commitments. The units of analysis (primitives) are groups. They are supposed to be able to reach cooperatively some desirable gains (for the group and each player). The question in cooperative games is how the players share the gains from their cooperation. In *non-cooperative* games, no communications and binding commitments are allowed. Each player is treated as a "selfish" individual who cares about maximizing its own utility in presence of strategic interdependences.

The history of game theory is not long. It is widely thought to have originated in the early twentieth century, when von Neumann laid the groundwork by proving the min-max theorem in a paper in 1928 [47]. However, it was not until 1944 when Von Neumann and Morgenstern's classic book *The Theory of Game and Economic Behaviors* [48] appeared, that the world realized the power of game theory for studying the behaviors in Economy, Political Science, Social Science and Biology. Others had anticipated some of the ideas. Later, many economists and mathematicians made innovative contributions. Some of the most pioneering results were reported within a year, when Nobel Laureate John Nash made seminal contributions to both *cooperative* and *non-cooperative* games. In [45], Nash proved the existence of a strategic equilibrium for non-cooperative games (Nash Equilibrium). He also proposed that cooperative games were reducible to non-cooperative games. He accomplished that by pioneering the axiomatic bargaining theory and proved the existence of the Nash Bargaining Solution (NBS) for cooperative games (a notion similar to the Nash Equilibrium) [43]. The remarkable property of Game Theory is its abstractly defined mathematics and notions of optimality. In no other branch of Sciences do we find so many understandable definitions and levels of optimality [51]. Game Theory has been used as a powerful method to solve and analyze problems that contain natural competition in several areas of social sciences [29], Biology [13], Political Science [42] Economics [41], etc. In computer science, a few applications of Game Theory applied to job scheduling and networking problems have been documented [61]. Recently, auction theory is being recognized as the emerging solution for problems in microeconomics [41], and agent-based systems [64].

Cooperative and non-cooperative games are applied in our study of game theoretical rate control. We formulate the bit allocation problems in rate control to different types of games and derive the bit allocation strategies from the solutions. Moreover, we incorporate the human visual property into the game theoretical frameworks. The perpetual visual quality is optimized under the bit rate and “fairness” constraint. To our knowledge, this work is the first attempt to apply game theory in video compression.

1.5 Organization of the Dissertation

The dissertation organized as follows:

In chapter 2, a macroblock level rate control schemes using cooperative game theory is presented.

In chapter 3, a JMVO rate control schemes using non-cooperative game theory in object level bit allocation is proposed.

Chapter 4 summarizes the studies and research presented in this dissertation, and discusses the possible future research.

CHAPTER 2

PERCEPTUALLY TUNED RATE CONTROL USING COOPERATIVE GAME THEORY

2.1 Introduction

Most rate-distortion optimization algorithms optimize a unique objective function, which is typically the perceptual quality or distortion of an entire frame. While such an approach allows an overall good visual quality, the same objective function may not yield the best results for the whole video frame or a sequence. This work presents a game theory based rate allocation strategy that allows multi-objective optimization with multiple decision makers (e.g., blocks), while working under an overall constraint (e.g., a given bit budget for a frame). The proposed approach is based on cooperative game theory, under which each decision maker has its own objective function, which is its own perceptual quality measure. The solution for the cooperative game yields the optimal bit allocation that is fair to each macroblock (MB) under the give constraints. The proposed rate control algorithm has two stages. In the first stage, target bits are allocated at the frame level. In the second stage, the quantization scale for each MB is determined by using a game theoretical approach.

At the frame level, the algorithm allocates target bits to the current frame based on the coding complexity of the frame, which is the mean absolute prediction error. Since the remaining frames are unavailable, we propose a method to estimate the coding complexity of the remaining frames from the encoded frames. The target number of bits of the current frame is optimized by using the current frame coding complexity as well as the estimated coding complexity of the remaining frames.

At the MB level, the algorithm utilizes game theory for bit allocation. We formulate the bit allocation as a bargaining problem [44]. Each MB competes for a share of resources, which are the target bits for a frame. Based on *Nash Bargaining Solution* [46], we derive a cooperative optimal quantization scale for each MB. Furthermore, we incorporate the human visual system (HVS) perceptual property in the game theory framework. Initial visual quality of the game setting, which is guaranteed for each MB, is determined proportional to the perceptible distortion, which is the distortion that exceeds the Just-Noticeable-Distortion (JND) threshold [24]. Figure 2.1 shows the block diagram of the proposed rate control algorithm that can be embedded in a DCT-based video encoder. The gray box in the diagram shows the modules of the proposed algorithm.

The rest of this chapter is organized as follows: Section 2 presents the related work on rate control algorithms reported in the literature. Section 3 describes the proposed frame level bit allocation scheme and the proposed game theoretical formulation for MB layer quantizer optimization. Section 4 presents simulation results of the proposed scheme including a comparison to the quadratic rate control proposed by [4] and recommended by MPEG-4 VM8 [19]. Finally, Section 5 concludes the work with some final remarks.

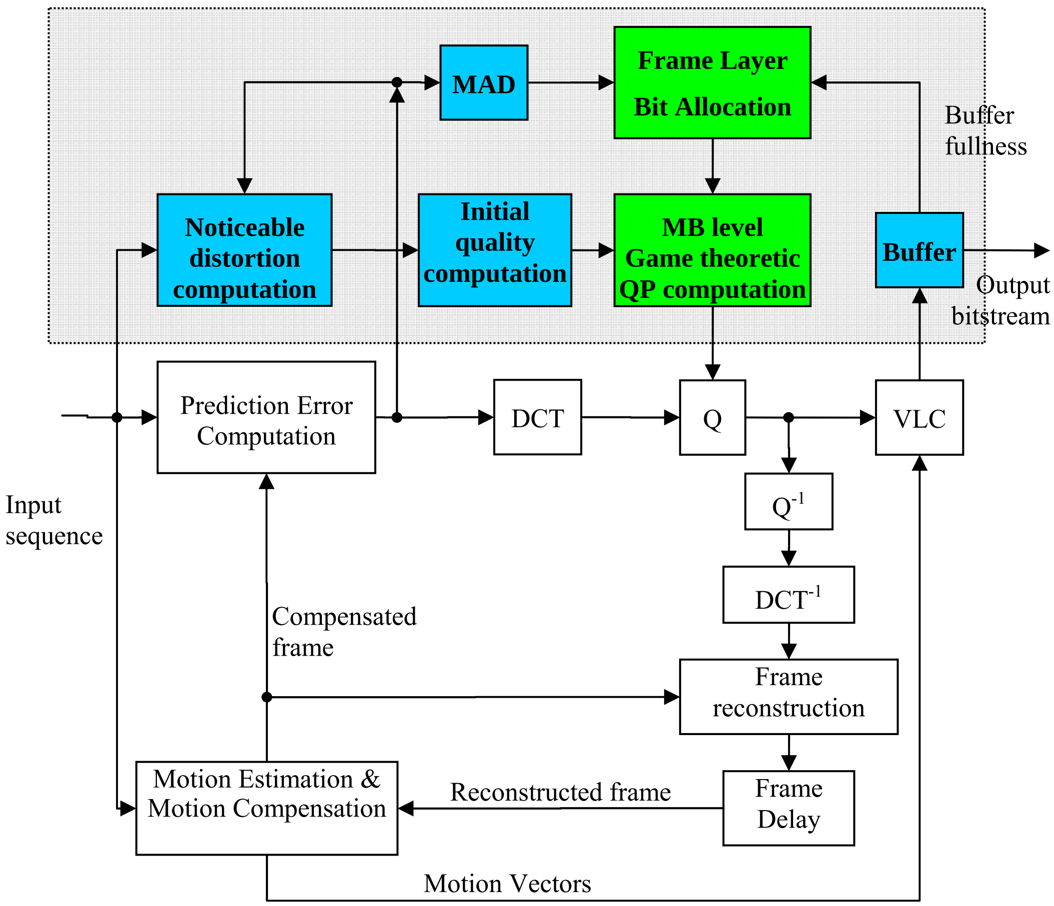

Figure 2.1 The proposed algorithm embedded in a DCT-based video encoder

2.2 Frame-Level Bit Allocation

The visual quality of an encoded frame is related to the bits it consumes. To maintain stable quality throughout the video sequence, one needs to consider the distribution of the bit budget for each frame. Since it takes more bits to encode a complex frame than a simple frame to obtain the same visual quality, we propose an approach to tune the frame level target bit budget according to the estimated coding complexity of the frame. We measure the coding complexity of a frame using the mean absolute prediction error, denoted by *mad*:

$$mad = \frac{1}{W \cdot H} \sum_{x=0}^{W} \sum_{y=0}^{H} \left| p(x,y) - \hat{p}(x,y) \right| \tag{2.1}$$

where *(x,y)* is the pixel coordinate, W and H are the width and height of the frame in pixel. The target bit budget is allocated proportional to the coding complexity of the current coding frame. Therefore, we need to know the coding complexity of the remaining frames. Since the encoder only contain the current frame and the reference frame due to the memory and delay constraints, we estimate the coding complexity of the remaining frame by computing the weighted mean of the coding complexity of the previous coded frames. Suppose we are encoding a frame at time t. The estimated coding complexity of the remaining frames is denoted by mad_r_t. and given by:

$$
\begin{aligned}
mad\_r_t &= \frac{\sum_{i=1}^{t}(i \cdot mad_i)}{\sum_{i=1}^{t} i} \\
&= \frac{2 \cdot \sum_{i=1}^{t}(i \cdot mad_i)}{(t+1) \cdot t} \\
&= \frac{2 \cdot \sum_{i=1}^{t-1}(i \cdot mad_i) + 2t \cdot mad_t}{(t+1) \cdot t} \\
&= \frac{(t-1)}{(t+1)} \cdot \frac{2 \cdot \sum_{i=1}^{t-1}(i \cdot mad_i)}{t \cdot (t-1)} + \frac{2}{t+1} mad_t \\
&= \frac{(t-1)}{(t+1)} \cdot mad\_r_{t-1} + \frac{2}{t+1} mad_t
\end{aligned}
\tag{2.2}
$$

where mad_i is the coding complexity of frame at time i. The number of bits allocated to a frame at time t is given by:

$$
T_t' = \frac{R_{t-1}}{N_{t-1}} \cdot \times \frac{mad_t}{mad\_r_t} \tag{2.3}
$$

where E_{t-1} and N_{t-1} are the numbers of remaining bits and remaining P frames after encoding the frame at time $t-1$.

Initially, I frame at time 0 is encoded using an external input quantization scale. The number of the remaining bits after encoding the first I frame is:

$$
R_0 = Br \cdot L - R_f \tag{2.4}
$$

where Br is the target bit rate of the sequence, L is the total length of the sequence in unit of second. R_f is the number of bits used to encode the first I frame.

We need to further adjust the target bits T_t' to the current buffer fullness with the following equation:

$$T_t'' = \frac{B_t + 2\times(B - B_t)}{2\times B_t + (B - B_t)} \times T_t' \tag{2.5}$$

where B is the buffer size and B_t is the buffer fullness at time t. (2.5) intends to maintain the buffer fullness in the middle of the buffer. If the buffer fullness is lower than the middle level, more bits will be allocated to the current frame. Otherwise, fewer bits will be allocated.

Finally, the target number of bits is bounded by:

$$T_t = \min\left(\frac{2Br}{F}, \max\left(\frac{Br}{4F}, T_t''\right)\right) \tag{2.6}$$

where F is the frame rate. Br/F is the average bits for a frame (2.6) intends to avoid allocating extremely large or extremely small number of target bits to a frame, in order to prevent the buffer overflow or under flow. The upper bound of target number of bits for a frame is two times of average bits for a frame and the lower bound is a quarter of the average bits for a frame.

2.3 MB Level Bit Allocation

The problem to identify the optimal quantization scale is equivalent to find an optimal allocation of the frame target bits to maximize the perceptual quality, which is a resource optimization problem. Each MB competes for a share of resources to optimize its own performance. Since the whole frame is an entity perceived by the viewers, MBs need to work cooperatively. We solve the problem by playing a multiple-player cooperative game. An optimal and fair bit allocation strategy is derived based on the Nash bargaining solution.

2.3.1 Quadratic Rate-Distortion Model

We propose a quadratic rate-distortion model to formulate the relation of the consumed bits and the quantization scale used to encode a MB. The proposed model is plugged in to the game theoretical framework presented in section 0, to determine the optimal quantization scale. The model is given by:

$$\frac{r_i}{m_i^{\alpha}} = \frac{K}{Q_i^{\;2}} \tag{2.7}$$

where r_i is the number of bits spend to encode the *i*th MB, m_i is the standard deviation of prediction errors of the *i*th MB, Q_i is the quantization step size for the *i*th MB, K are model parameters. α is a constant ($\alpha=0.8$).

The rate-distortion model is updated after the encoding of each MB. The rate-distortion model parameter K for the $(i+1)$th MB will be updated as the following:

$$K_{i+1} = \begin{cases} \dfrac{(i-n_s-1)}{i-n_s} K_i^{'} + \dfrac{A_i Q_i^2}{m_i^{\alpha}(i-n_s)} & A_i > 0 \\ K_i \times 0.9 & A_i = 0 \end{cases} \tag{2.8}$$

$$\text{and } K_{i+1}^{'} = \begin{cases} K_i & A_i > 0 \\ K_i^{'} & A_i = 0 \end{cases},$$

where A_i is the actual bits used to encode the *i*th MB, n_s is the number of skipped MBs. The initial value K_0 is set to 3000.

2.3.2 Solving Quantizer Optimization with Game Theory

In the proposed MB level bit allocation, the bargaining game is configured as following:

Players: Each frame contains N uncoded MBs. Each MB is regarded as a player in the game. N players compete for the use of a fixed resource, which is the target bit budget R_t for the frame at time t.

Strategies: The strategy of a player is the number of bits it requests for, denoted by r_i. Since the target number of bits for a frame is constrained, the sum of the bits requested by the N remaining MBs should be no more than the remaining bits for a frame, i.e.:

$$\sum_{i=1}^{N} r_i \leq R_c \tag{2.9}$$

where R_c is the remaining bits for the N MBs.

Preference: A utility function u_i for each player i reflects its preference. We use the visual quality as a measure of utility. Higher visual quality is more preferable. Given a combination of strategies carried out by all the MBs $r = (r_1 ,r_2 ,\ldots,r_N)$, $u = (u_1(r), u_2(r), \ldots, u_N(r))$ is the utility of the game. Since in the DCT video coding, the visual quality of a MB is related only to the number of bits it obtained, the utility can be represented by $u=(u_1(r_1), u_2(r_2), \ldots, u_N(r_N))$.

Initial utility: The initial utility of the ith MB, denoted by d_i, is the initial perceptual quality that required to be guaranteed. The initial quality is determined according to the perceptible distortion of each MB, which is defined in section 0. Denote the initial quality of the game $d=(d_1, d_2, \ldots, d_N)$, we have $u>d$. Define the number of bits achieving d_i as r_i^0. Since $u>d$, we have $r > r^0$, where $r^0 = (r_1^0 , r_2^0 ,\ldots, r_N^0)$, which means:

$$r_i > r_i^0 \tag{2.10}$$

Define U the set of achievable utilities. Tuple $<U, d>$ represents the game setting.

Nash bargaining solution (NBS) is a unique solution that satisfies a set of axioms for fair bargain [46]. Let strategy r^* be the NBS of the game $<U, d>$, the corresponding utility $u^*(<U,d>)$ is called the Nash bargaining point. The following axioms define a unique NBS:

Efficiency: *The Nash bargaining solution is Pareto optimal, i.e., there is no other solution produces better utility for one player without hurting another player.*

Linearity: *Given a monotonous increasing linear transform function F, Nash bargaining point u* satisfies that:* $u^*(<F(U),F(d)>) = F(u^*(<U,d>)$

Independence of irrelevant alternatives: *Let X, Y are sets of attainable utilities, and* $X \subseteq Y$. *If* $u^*(<Y,d>) \in X$, *then* $u^*(<Y,d>) = u^*(<X,d>)$.

Symmetry: *If U is symmetric with respect to any two players in the game, and if their initial utility is equally preferable, then exchange the two players will not affect the solution.*

The efficiency axiom states that the Nash bargaining solution is cooperatively optimal. The linearity axiom implies that the NBS will not change if the player's objective is linearly transformed. The irrelevant alternatives axiom expressed that if we remove the irrelevant subset (the subset does not contain the Nash bargaining point), the NBS of the game will not change. The symmetry axiom says that the solution is only depended on the players' initial utilities and their utility functions. The last three axioms are the axioms of fairness.

NBS for a multiple players bargaining game is characterized by the following property [66]:

The solution r* is a NBS if and only if $\prod_i (u_i(r^*) - d_i) \geq \prod_i (u_i(r) - d_i)$ for all $r \in H$, where H is the set of all feasible combination of strategies.

Therefore, to find the Nash bargaining solution, we need to solve the following maximization problem:

$$\max \prod_{i=1}^{N} \left(u_i(r_i) - d_i\right) \qquad \text{s.t. } r_i > r_i^0, \ \sum_{i=1}^{N} r_i \leq R_c \tag{2.11}$$

The conditions in (2.11) are constraints from (2.9) and (2.10).

The approximate mean square error (MSE) distortion of the *i*th MB is $D_i = Q^2/12$ [8]. Therefore, we defined the visual quality of the ith MB by:

$$u_i = \frac{1}{D_i} = \frac{12}{Q_i^2} = \frac{12 r_i}{K m_i^{\alpha}} \tag{2.12}$$

Since u_i is concave and injective, ln(u_i) is strictly concave. The above problem is equivalent with the following problem:

$$\max \sum_{i=1}^{N} \ln\left(u_i(r_i) - d_i\right) \qquad \text{s.t. } r_i > r_i^0, \ \sum_{i=1}^{N} r_i \leq R_c \tag{2.13}$$

The above inequality constrained optimization problem can be solved by maximizing the following Lagrangean, using the theorem of Kuhn and Tucker [67].

$$J = \sum_{i=1}^{N} \ln(u_i - d_i) + \lambda(R_c - \sum_{i=1}^{N} r_i) + \sum_{i=1}^{N} \theta_i (r_i - r_i^0) \tag{2.14}$$

where λ and θ_i are the Lagrange multiplier.

$$
\begin{cases}
\dfrac{\partial J}{\partial r_i} = \dfrac{\partial \ln(u_i - d_i)}{\partial r_i} - \lambda + \theta_i = 0 \\
\dfrac{\partial J}{\partial \lambda} = R_c - \sum_{i=1}^{N} r_i \geq 0 \\
\lambda \dfrac{\partial J}{\partial \lambda} = \lambda (R_c - \sum_{i=1}^{N} r_i) = 0 \\
\dfrac{\partial J}{\partial \theta_i} = r_i - r_i^0 \geq 0 \\
\theta_i \dfrac{\partial J}{\partial \theta_i} = \theta_i (r_i - r_i^0) = 0
\end{cases}
\tag{2.15}
$$

$$, \text{where } i \in (1 \ldots N)$$

Since $\sum_{i=1}^{N} r_i^0 < R_c$, there must be a solution *r* that strictly superior to r^0, so that $r_i - r_i^0 > 0$. Therefore, $\theta_i = 0$. From (2.15), we have

$$
\begin{aligned}
\frac{\partial \ln(u_i - d_i)}{\partial r_i} &= \lambda \\
\Rightarrow r_i &= \frac{1}{\lambda} + \frac{K}{12} m_i^{\alpha} d_i
\end{aligned}
\tag{2.16}
$$

and

$$
\begin{aligned}
R_c - \sum_{i=1}^{N} r_i &= 0 \\
\Rightarrow R_c - \frac{N}{\lambda} - \frac{K}{12} \sum_{i=1}^{N} m_i^{\alpha} d_i &= 0
\end{aligned}
\tag{2.17}
$$

Therefore, the NBS for the game is given by:

$$r_i = \frac{R_c}{N} - \frac{K}{12N} \sum_{j=1}^{N} m_j^{\alpha} d_j + \frac{K}{12} m_i^{\alpha} d_i \tag{2.18}$$

And the optimal quantization step size is given by:

$$Q_i = \sqrt{Km_i^{\alpha} \Big/ \left(\frac{R_c}{N} - \frac{K}{12N} \sum_{j=1}^{N} m_j^{\alpha} d_j + \frac{Km_i^{\alpha} d_i}{12} \right)} \tag{2.19}$$

Proof of fairness:

An assignment of resource $s = (s_1, s_2, \ldots, s_N)$ is said to be *proportionally fair* with respect to a utility function *f*, if for any other feasible allocation *s'*, the aggregate proportional changes is zero or negative [26].

$$\sum_{i=1}^{N} \frac{f(s') - f(s_i)}{f(s_i)} \le 0 \tag{2.20}$$

Denote r^* the allocation from NBS. Consider a small changes of r^* to r', where $r_i' = r_i^* + \Delta r_i^*$. In the context of our game setting, the proportional fairness criteria can be rewritten as $\sum_{i=1}^{N} \frac{[u_i(r_i') - d_i] - [u_i(r_i^*) - d_i]}{u_i(r_i^*) - d_i} \le 0$

Replace *u* with (2.12), we have

$$\sum_{i=1}^{N} \frac{[u_i(r_i') - d_i] - [u_i(r_i^*) - d_i]}{u_i(r_i^*) - d_i} \le 0$$

$$\Leftrightarrow \sum_{i=1}^{N} \frac{\left[\frac{12r_i}{Km_i^{\alpha}} - d_i\right] - \left[\frac{12r_i^*}{Km_i^{\alpha}} - d_i\right]}{\frac{12r_i^*}{Km_i^{\alpha}} - d_i} \le 0$$

$$\Leftrightarrow \sum_{i=1}^{N} \frac{\frac{12r_i}{Km_i^{\alpha}} - \frac{12r_i^*}{Km_i^{\alpha}}}{\frac{12r_i^*}{Km_i^{\alpha}} - \frac{12r_i^0}{Km_i^{\alpha}}} \le 0$$

$$\Leftrightarrow \sum_{i=1}^{N} \frac{r_i' - r_i^*}{r_i^* - r_i^0} \le 0 \tag{2.21}$$

Let $v_i(r_i) = \ln(u_i(r_i) - d_i)$. Since r^* is the optimal solution for (2.13), any change in r^* will make a zero or negative change in $\sum_{i=1}^{N} v_i(r_i)$. Therefore, we have

$$\sum_{i=1}^{N} \left[v_i(r_i') - v_i(r_i^*) \right] \le 0 \tag{2.22}$$

$$\sum_{i=1}^{N} \left[\left. \frac{\partial v_i}{\partial r_i} \right|_{r_i = r_i^*} \Delta r_i^* \right] \le 0 \tag{2.23}$$

$$\sum_{i=1}^{N} \frac{\left. u_i'(r_i) \right|_{r_i = r_i^*}}{u_i(r_i^*) - d_i} \Delta r_i^* \le 0 \tag{2.24}$$

Plug (2.12) into (2.24), we have

$$\sum_{i=1}^{N} \frac{\Delta r_i^*}{r_i^* - r_i^0} \le 0 \Rightarrow \sum_{i=1}^{N} \frac{r_i' - r_i^*}{r_i^* - r_i^0} \le 0 \tag{2.25}$$

2.3.3 Perceptually Tuned Quantizer

The perceptual redundancy is inherent in video signals. It is found that the human visual system is insensitive to the signals in some spatial frequencies. Moreover, the human vision is much easier to detect the luminance difference rather than the absolute intensity. And the sensitivity to the luminance contrast is depended on the average background intensity. Due to the above observation, a metric of just-noticeable-distortion (JND) is proposed in [24] to measure the perceptual lower bound of the signal distortion. JND is a threshold below which the distortion is imperceptible. The computation of JND of a pixel at (x,y) is given by [5]:

$$JND(x, y) = \max\{f_1(bg(x, y), mg(x, y)), f_2(bg(x, y))\} \tag{2.26}$$

$$f_1(bg(x, y), mg(x, y)) = mg(x, y) \cdot \alpha(bg(x, y)) + \beta(bg(x, y)) \tag{2.27}$$

$$f_2(bg(x, y)) = \begin{cases} T_0 \cdot (1 - (bg(x, y)/127)^{1/2}) + 3 & bg(x, y) \le 127 \\ \gamma \cdot (bg(x, y) - 127) + 3 & bg(x, y) > 127 \end{cases} \tag{2.28}$$

$$\alpha(bg(x, y)) = bg(x, y) \times 0.0001 + 0.115 \tag{2.29}$$

$$\beta(bg(x, y)) = \lambda - bg(x, y) \times 0.01 \tag{2.30}$$

where T_0, γ and λ are 17, 3/128 and 1/2, respectively.

bg(x,y) and *mg(x,y)* represent the average background luminance and the maximum weighted average of luminance differences around the pixel at *(x,y)*. The calculations of *bg(x,y)* and *mg(x,y)* are given below:

$$bg(x, y) = \frac{1}{32} \sum_{i=1}^{5} \sum_{j=1}^{5} p(x - 3 + i, y - 3 + j) \cdot B(i, j) \quad \text{for} \quad 0 \le x < H, 0 \le y < W, \tag{2.31}$$

where *p(x,y)* is the luminance of pixel at *(x,y)*. *B(i,j),* (*i,j=1,..,5)*, is a matrix of weights, which is shown in Figure 2.2. mg is the maximum gradient among different directions. $grad_k$ represents the gradient in direction *k*, where *k*={1,2,3,4} is one of the following four directions, 1: vertical, 2: diagonal (upper-left to lower-right), 3: horizontal, 4: diagonal (upper-right to lower-left). $grad_k$ is computed with (2.33).

1	1	1	1	1
1	2	2	2	1
1	2	0	2	1
1	2	2	2	1
1	1	1	1	1

Figure 2.2 Matrix B

0	0	0	0	0
1	3	8	3	1
0	0	0	0	0
-1	-3	-8	-3	-1
0	0	0	0	0

(a)

0	0	1	0	0
0	8	3	0	0
1	3	0	-3	-1
0	0	-3	-8	0
0	0	-1	0	0

(b)

0	0	1	0	0
0	0	3	8	0
-1	-3	0	3	1
0	-8	-3	0	0
0	0	-1	0	0

(c)

0	1	0	-1	0
0	3	0	-3	0
0	8	0	-8	0
0	3	0	-3	0
0	1	0	-1	0

(d)

Figure 2.3 Matrix G_k (a) G_1 , (b) G_2 , (c) G_3 , (d) G_4

$$mg(x, y) = \max\{|grad_k(x, y)|\} \tag{2.32}$$

$$grad_k(x, y) = \frac{1}{16}\sum_{i=1}^{5}\sum_{j=1}^{5} p(x-3+i, y-3+j)\cdot G_k(i, j) \tag{2.33}$$

where G_k is the matrices of weights in four directions, which is given in Figure 2.3.

In (2.19), a factor that affects the optimal quantization step size Q_i is the initial quality d_i. In the proposed algorithm, the initial quality of a MB is proportional to its noticeable distortion e. The noticeable distortion in the ith MB is given by:

$$e_i = \frac{1}{256}\sum_{m=1}^{16}\sum_{n=1}^{16}\left\{ \left[h(m+u, n+v) - JND(m+u, n+v)\right]\cdot\delta(m+u, n+v) \right\} \tag{2.34}$$

where (u, v) is the top-left corner of the ith MB, and

$$h(x, y) = \left|p(x, y) - \hat{p}(x, y)\right| \tag{2.35}$$

$$\delta(x, y) = \begin{cases} 1, & h(x, y) > JND(x, y) \\ 0, & h(x, y) \le JND(x, y) \end{cases} \tag{2.36}$$

Due to the bit rate constraint, the total bits corresponding to the initial quality of the N uncoded MBs in a frame cannot exceed the remaining bits, i.e., $\sum_{i=1}^{N} r_i^0 < R_c$. Therefore, we bound the initial quality by a scale factor C, such that the total bits corresponding to the initial quality are bounded by a half of the remaining bits for the N uncoded MBs ($0.5R_c$). C is computed as follows:

$$C = \begin{cases} 1 & R_c \geq \dfrac{K}{6}\sum_{j=1}^{N} m_j^{\alpha} e_j \\ \dfrac{6R_c}{K \cdot \sum_{j=1}^{N} m_j^{\alpha} e_j} & R_c < \dfrac{K}{6}\sum_{j=1}^{N} m_j^{\alpha} e_j \end{cases} \tag{2.37}$$

The initial quality for the ith MB is given below:

$$d_i = C \cdot e_i \tag{2.38}$$

2.4 Algorithm Summary

The proposed algorithm, name Game Theoretic (GT) algorithm, is embedded in a DCT-based video encoder (e.g. H.263 or MPEG-4), as shown in Figure 2.1. The functionality of each block in the diagram is introduced in the previous sections. We summarize the GT algorithm in the following steps:

(1) Frame layer bit allocation will be performed prior to the encoding of a frame. The mean absolute prediction error of the current frame is computed with (2.2) and (2.3), in order to determine the initial estimation of the target number of bits T_t'. The Buffer fullness feedback further tunes T_t' with (2.5). Then T_t'' is further adjusted with (4.6) to get T_t.

(2) Before encoding a frame, R_c is initially set to T_t. The noticeable distortion for each MB in the frame is calculated with (2.34).

(3) Based on the noticeable distortion and remaining bits R_c, the initial qualities for the remaining MBs are computed with (2.37) and (2.38).

(4) The quantization step size of current MB is computed based on the derived game theoretical formulation, given by (2.19).

(5) After encoding each MB, the quadratic rate-distortion model is updated with (2.8). R_c is updated as $R_c=R_c-A_i$. Go to step 6 if all the MBs in the frame is finished, otherwise go to step 3.

(6) Update the buffer status after encoding each frame. Stop if all the frames are encoded, otherwise go to step 1.

2.5 Experiments and Results

First, we compare the GT algorithm to the quadratic rate control algorithm suggested by MPEG-4 VM8 [19]. Both comparing algorithms are implemented in Momusys encoder for MPEG-4 Verification Model. To be consistent with the VM8 rate control algorithm, we set the buffer size to *Br/8*, which means the maximum delay is 125ms. The initial buffer fullness is *Br/16*.

The performance of a rate control algorithm is evaluated by the following metrics.

I. The bit rate accuracy

II. Percentage of bits saved

III. The number of skipped frames

IV. The peak-signal-to-noise ratio (PSNR)

V. The peak-signal-to-perceptible-noise ratio (PSPNR)

A good rate control algorithm should be able to control the actual bit rate as close as possible to the target bit rate. We measure the bit rate accuracy with the following equation:

$$bit\_rate\_accuracy = 1 - \frac{\left|R_{\text{actual}} - R_{\text{target}}\right|}{R_{\text{target}}} \tag{2.39}$$

where R_{actual} and R_{target} are the actual bit rate and the target bit rate respectively. Besides the bit rate accuracy, we also measure the percentage of bits saved of the GT algorithm compared to the VM8 algorithm. For the same visual quality, lower bit consumption means the higher rate-distortion efficiency.

Frame skip technique is employed to avoid the buffer overflow. Once the buffer fullness is over a threshold, the encoding of the next frame will be skipped and will not be buffer in order to cut down the buffer fullness level. In the decoder side, the skipped frame will be replaced by a duplication of the previous frame in order to maintain the continuity of the video decoding. However, a frame skip will degrade the signal quality. A good rate control algorithm should be able to avoid the buffer overflow and minimize the number of skipped frames.

PSNR is a widely adopted metric to measure visual quality. PSNR averages the noise of all pixels in a frame, regardless if it is perceptible to human visual system or not. PSPNR proposed by Chou and Li [5] measures the perceptible visual quality incorporating the human perceptual property. PSPNR only take account of the perceptible noise and is defined as:

$$PSPNR = 20\log_{10} \frac{255}{\sqrt{\frac{1}{HW}\sum_{x=0}^{H}\sum_{y=0}^{W}(h(x,y) - JND(x,y))^2 \cdot \delta(x,y)}} \tag{2.40}$$

where *h(x,y)* and *δ(x,y)* and is defined in (2.35) and (2.36) respectively. According to the MPEG-4 core experiment, the PSNR (and PSPNR) of the skipped frame is computed by considering the skipped frame as the duplication of the previous decoded frame in the decoded sequence [62] [63].

We encode QCIF and CIF format test sequences at various target bit rates. The frame rate in the experiments is 30 frames per second. For each sequence, 100 frames are encoded. The temporal prediction structure used in the experiment is IPPP…, i.e., only the first frame is encoded as I-frame and the remaining frames are encoded as P-frame (IPP…). It is to be noted that the GT algorithm is not limited to P frames, and easily extended to the frame level bit allocation for I frame and B frame. The reason of using this temporal prediction structure in the experiment is that the comparative algorithm (VM8) supports only this structure. To be consistent with VM8 in the comparison, we adopt IPPP… structure in the experiment. The detailed simulation results and the comparisons are shown in Table 2.1 and 2.2. Table 2.1 shows the results of QCIF format and Table 2.2 shows the results of CIF format.

2.5.1 Bit Rate Accuracy and Percentage of Bits Saved

From Table 2.1, we observe that the GT rate control algorithm produces fewer bits and achieves more accurate bit rate than the VM8 algorithm in most test cases. The average bit rate accuracy of VM8 is 96.45% for CIF format and 96.61 for QCIF format while the GT algorithm achieves 99.95% for CIF format and 99.93% for QCIF format. Compare to VM8, the GT algorithm saves 3.57% and 3.39% bits for CIF format and QCIF format respectively.

One can note that the GT algorithm uses fewer bits in the situation that without frame skipped. That means, comparing with VM8 that skips a certain number of frames to achieve the target bit rate, the proposed algorithm averagely use even less bits to encode a frame.

2.5.2 Frame Skipping

On the number of skipped frames comparison, one can notice that various numbers of frames are skipped when using VM8 algorithm. On the contrast, there is no frame skips in the experiments when using the GT rate control algorithm. The advantage of the GT algorithm is especially prominent in the low bit rate tests. For example, in the “Table” QCIF 64Kbps test and “Stefan” QCIF 96Kbps test, VM8 skips 3 frames and 5 frames respectively, while the GT algorithm does not skip any frame.

The reason of improvement of frame skipping is because the GT can maintain a stable buffer level, which thanks to the accurate estimation of frame level target bit budge and the high accuracy of MB level rate control.

Figure 2.4, Figure 2.6 and Figure 2.8 illustrate the frame-to-frame comparison on buffer fullness. One can observe that the proposed algorithm has less fluctuation on buffer level and is able to control the buffer fullness around the middle of the buffer size. The buffer fullness level is maintained within a safe margin to avoid frame skipping.

2.5.3 PSNR and PSPNR

The overall visual quality is measure by the average PSNR and PSPNR. Different rate control algorithm may lead to different number of frame to be skipped. If a rate control algorithm skips more frames than the other one, more bits are used to encode the non-skipped frame, which will lead to a higher PSNR and PSPNR value on the non-skipped frame. Therefore, it is unfair to compare the average PSNR and PSPNR only taking into account the encoded frame. The distortion of the skipped frame should be considered to for a fair comparison. In the MPEG-4 decoder, a skipped frame will be repeated by the previous encoded frame. Hence, in the MPEG-4 rate control test the previous frame is used in the PSNR and PSPNR computation when a frame is skipped [62].

Table 2.1 shows the PSNR and PSPNR comparison between VM8 and the proposed algorithm. It can be observed that the GT algorithm achieves higher PSNR than its counterpart in the PSNR comparison, although the GT algorithm is optimized for the PSPNR.

With regard to the perceptible distortion comparison, the GT algorithm successfully masks the imperceptible distortion. This is due to the bit allocation based on the noticeable distortion. Therefore, the algorithm produces substantial improvement in perceptual quality, compared to the VM8 algorithm. For example, the PSPNR improves by 1.68dB in the "Container" qcif 384kbps test, and by 2.23dB in the "Akiyo" qcif 384kbps test. The GT algorithm outperforms the VM8 algorithm in terms of both the bit rate and the visual quality.

Figure 2.4 to Figure 2.9 show detailed results of buffer fullness level and PSPNR values of each frame achieved by the two algorithms. One can observe that the GT algorithm achieves higher PSPNR and maintains the buffer fullness level within a safe margin to avoid frame skipping.

Table 2.1 Bit rate accuracy, PSNR, PSPNR and number of skipped frame comparison between VM8 and the GT algorithm (QCIF sequences)

video sequence	Algorithm	Bit rate		bit rate accuracy	bits saved	Average PSNR_Y	Average PSPNR_Y	Frame skipped	PSNR gain (dB)	PSPNR gain (dB)
		Target	Actual							
Table	VM8	64	66.64	95.88%		28.88	32.38	3		
	GT	64	64.1	99.84%	3.96%	29.04	32.6	0	0.15	0.22
	VM8	96	99.24	96.63%		30.26	34.36	3		
	GT	96	96	100.00%	3.37%	30.34	34.48	0	0.08	0.12
	VM8	128	131.54	97.23%		31.2	35.76	2		
	GT	128	128	100.00%	2.77%	31.34	35.98	0	0.15	0.22
Stefan	VM8	96	101.54	94.23%		24.45	26.87	5		
	GT	96	96.12	99.88%	5.64%	24.51	26.94	0	0.06	0.07
	VM8	112	119.81	93.03%		24.86	27.39	6		
	GT	112	111.98	99.98%	6.99%	25.03	27.61	0	0.17	0.22
	VM8	128	135.9	93.83%		25.34	28.03	5		
	GT	128	128.05	99.96%	6.13%	25.46	28.18	0	0.11	0.14
Foreman	VM8	64	69.7	91.09%		29.73	33.56	8		
	GT	64	64.04	99.94%	8.84%	29.84	33.7	0	0.11	0.14
	VM8	128	131.99	96.88%		32.53	37.93	2		
	GT	128	127.94	99.95%	3.17%	32.62	38.23	0	0.09	0.31
	VM8	192	195.27	98.30%		34.27	40.98	1		
	GT	192	191.79	99.89%	1.81%	34.32	41.29	0	0.05	0.31
container	VM8	192	199.06	96.32%		38.15	47.46	3		
	GT	192	192.09	99.95%	3.63%	38.48	47.81	0	0.33	0.36
	VM8	384	386.87	99.25%		41.15	54.62	0		
	GT	384	384.12	99.97%	0.72%	41.95	56.3	0	0.79	1.68
	VM8	512	512.01	100.00%		41.96	56.84	0		
	GT	512	512.2	99.96%	-0.04%	42.99	58.78	0	1.03	1.93

Table 2.1 – *Continued*

Akiyo	VM8	192	198.22	96.76%		42.31	55.97	2		
	GT	192	191.88	99.94%	3.30%	42.42	58.05	0	0.11	2.07
	VM8	384	384.91	99.76%		44.16	60.62	0		
	GT	384	383.41	99.85%	0.39%	44.72	62.85	0	0.56	2.23
	VM8	512	511.99	100.00%		45.39	64.18	0		
	GT	512	511.32	99.87%	0.13%	46.25	67.71	0	0.86	3.53

Table 2.2 Bit rate accuracy, PSNR, PSPNR and number of skipped frame comparison between VM8 and the GT algorithm (CIF sequences)

video sequence	Algorithm	Bit rate		bit rate accuracy	bits saved	Average PSNR_Y	Average PSPNR_Y	Frame skipped	PSNR gain (dB)	PSPNR gain (dB)
		Target	Actual							
container	VM8	192	193.32	99.31%		32.54	37.33	0		
	GT	192	191.82	99.91%	0.78%	32.57	37.42	0	0.03	0.09
	VM8	256	263.35	97.13%		33.37	38.64	2		
	GT	256	256.12	99.95%	2.82%	33.44	38.80	0	0.06	0.16
	VM8	384	400.84	95.61%		34.86	41.26	4		
	GT	384	384.32	99.92%	4.30%	34.90	41.32	0	0.04	0.05
Coastguard	VM8	256	270.36	94.39%		27.40	30.86	5		
	GT	256	256.04	99.98%	5.59%	27.43	30.88	0	0.03	0.02
	VM8	384	395.19	97.09%		28.69	32.67	2		
	GT	384	383.95	99.99%	2.93%	28.79	32.77	0	0.10	0.10
	VM8	512	521.55	98.13%		29.73	34.10	1		
	GT	512	511.88	99.98%	1.89%	29.78	34.15	0	0.05	0.05
foreman	VM8	256	271.10	94.10%		32.05	36.42	5		
	GT	256	255.82	99.93%	5.97%	32.35	36.97	0	0.30	0.55
	VM8	384	398.17	96.31%		33.57	38.98	3		
	GT	384	383.48	99.86%	3.83%	33.69	39.18	0	0.12	0.21
	VM8	512	522.32	97.98%		34.59	40.79	1		
	GT	512	511.41	99.89%	2.13%	34.64	40.81	0	0.05	0.02
table	VM8	384	397.00	96.61%		30.80	34.89	3		
	GT	384	383.77	99.94%	3.45%	31.08	35.29	0	0.27	0.40
	VM8	512	522.09	98.03%		31.93	36.56	1		
	GT	512	511.99	100.00%	1.97%	32.07	36.80	0	0.14	0.24
	VM8	640	648.14	98.73%		32.75	37.85	1		
	GT	640	640.03	100.00%	1.27%	32.88	38.04	0	0.12	0.20

Table 2.2 – *Continued*

Mobile	VM8	512	548.51	92.87%		24.10	27.01	6		
	GT	512	512.33	99.94%	7.06%	24.17	27.10	0	0.07	0.09
	VM8	640	674.20	94.66%		24.88	28.07	4		
	GT	640	640.19	99.97%	5.31%	24.95	28.16	0	0.07	0.09
	VM8	768	800.36	95.79%		25.56	29.01	3		
	GT	768	768.16	99.98%	4.19%	25.62	29.09	0	0.06	0.08

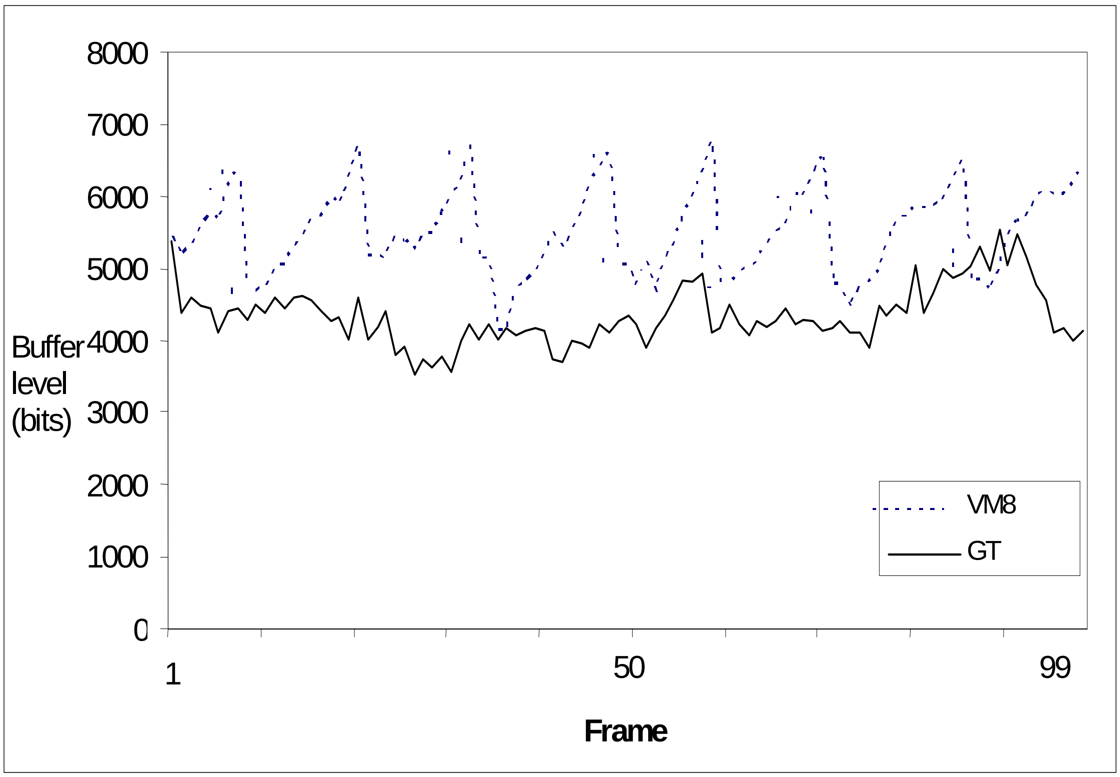

Figure 2.4 Buffer occupancy in the coding of “Foreman” QCIF at 64kbps and 30Hz with buffer size = 8kbits

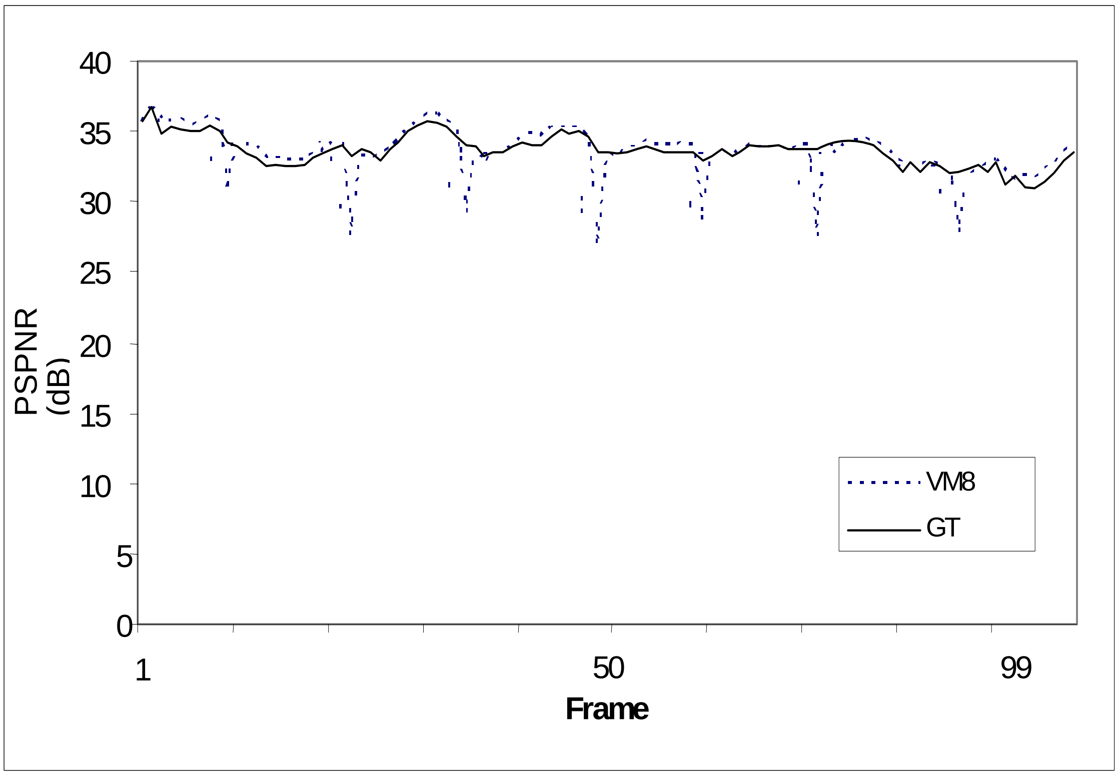

Figure 2.5 PSPNR in the coding of “Foreman” QCIF QCIF at 64kbps and 30Hz with buffer size = 8kbits

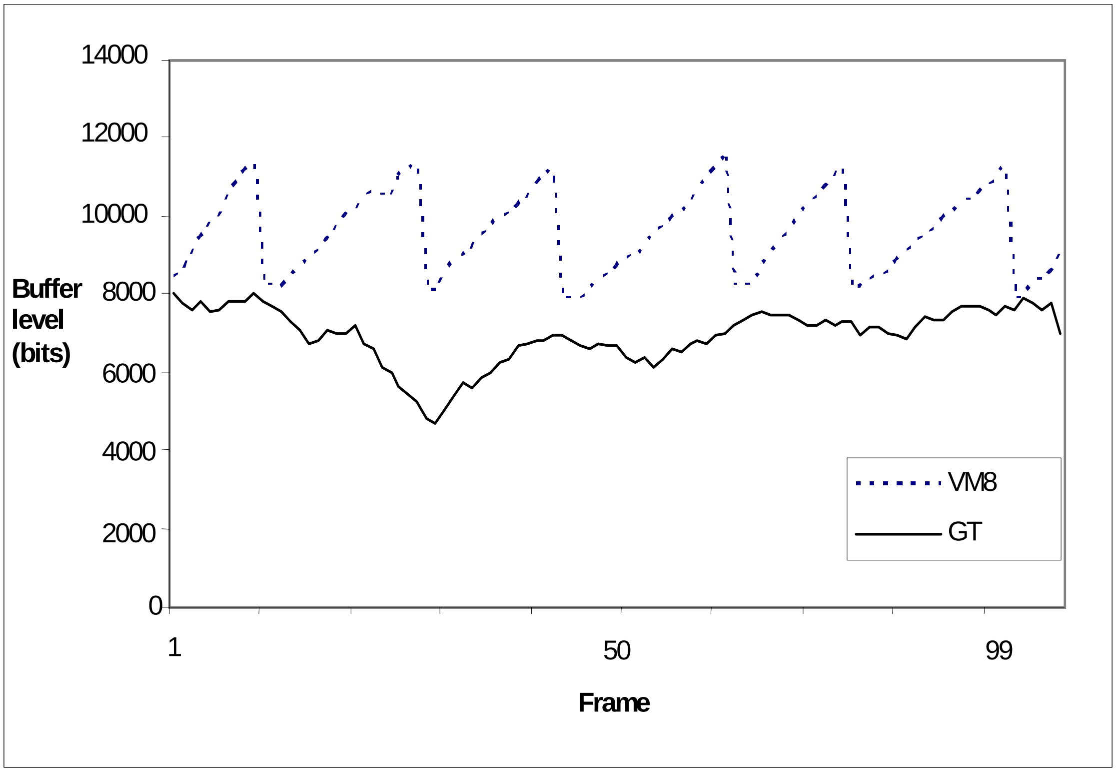

Figure 2.6 Buffer occupancy in the coding of "Stefan" QCIF at 12kbps and 30Hz with buffer size = 14kbits

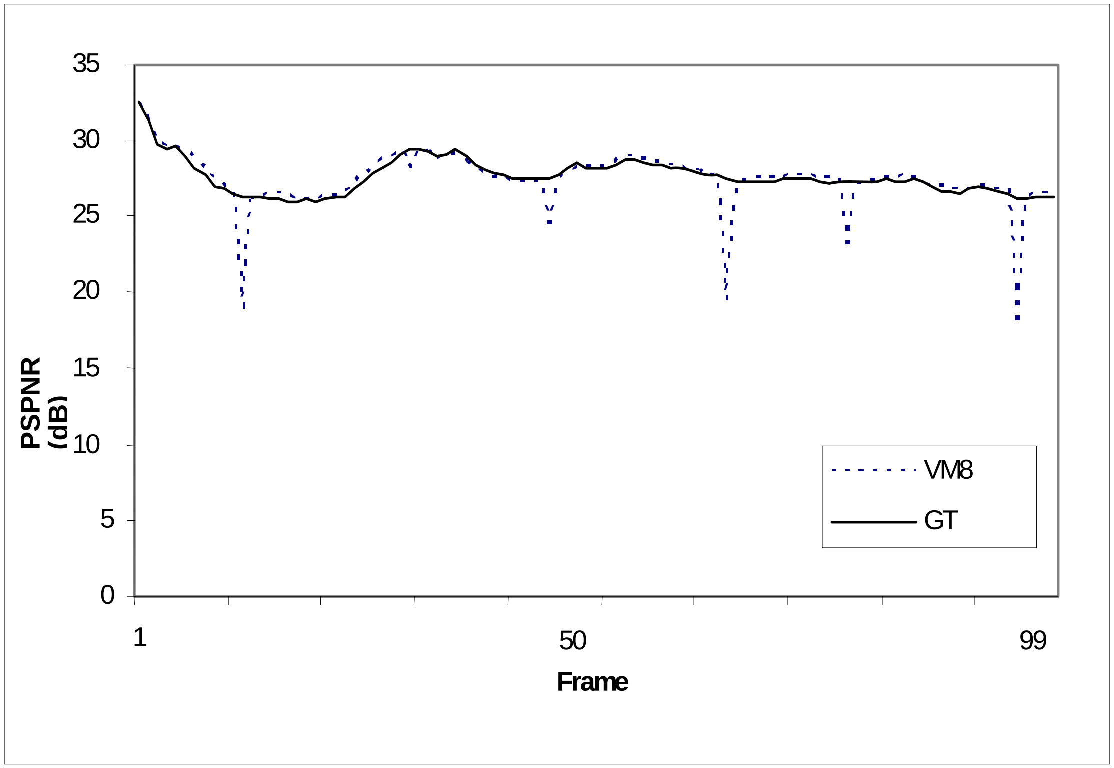

Figure 2.7 PSPNR in the coding of “Stefan” QCIF at 12kbps and 30Hz
with buffer size = 14kbits

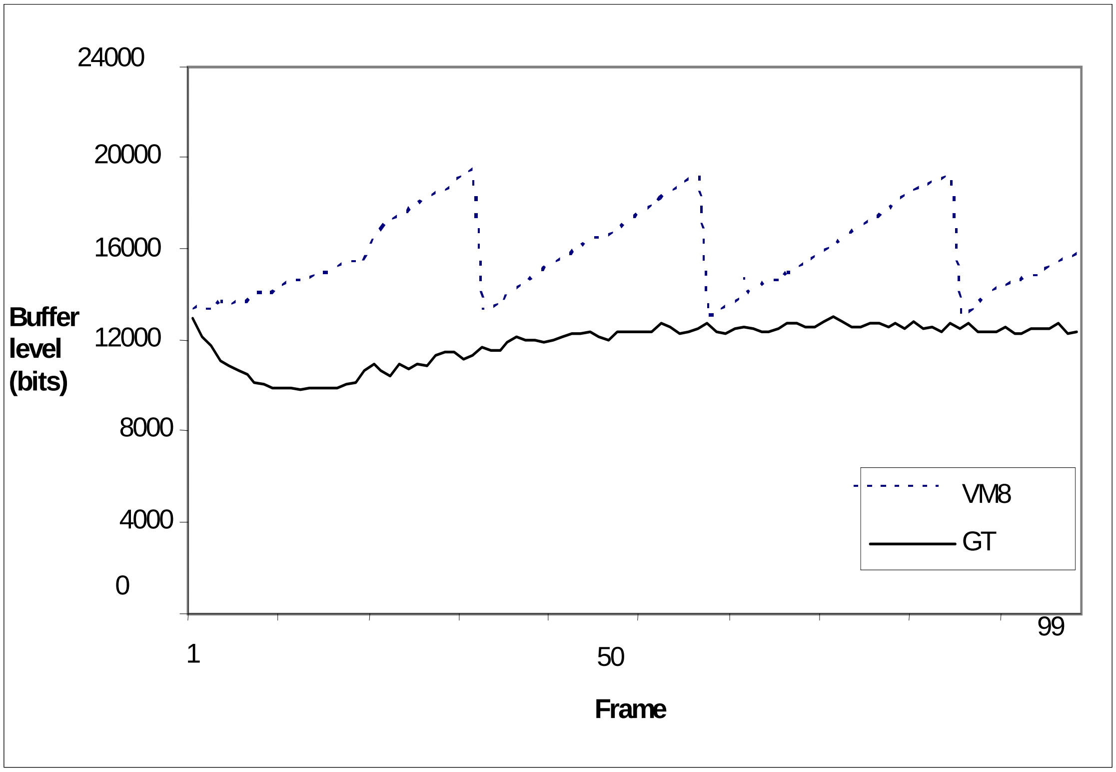

Figure 2.8 Buffer occupancy in the coding of "Container" QCIF at 192kbps and 30Hz with buffer size = 24kbits

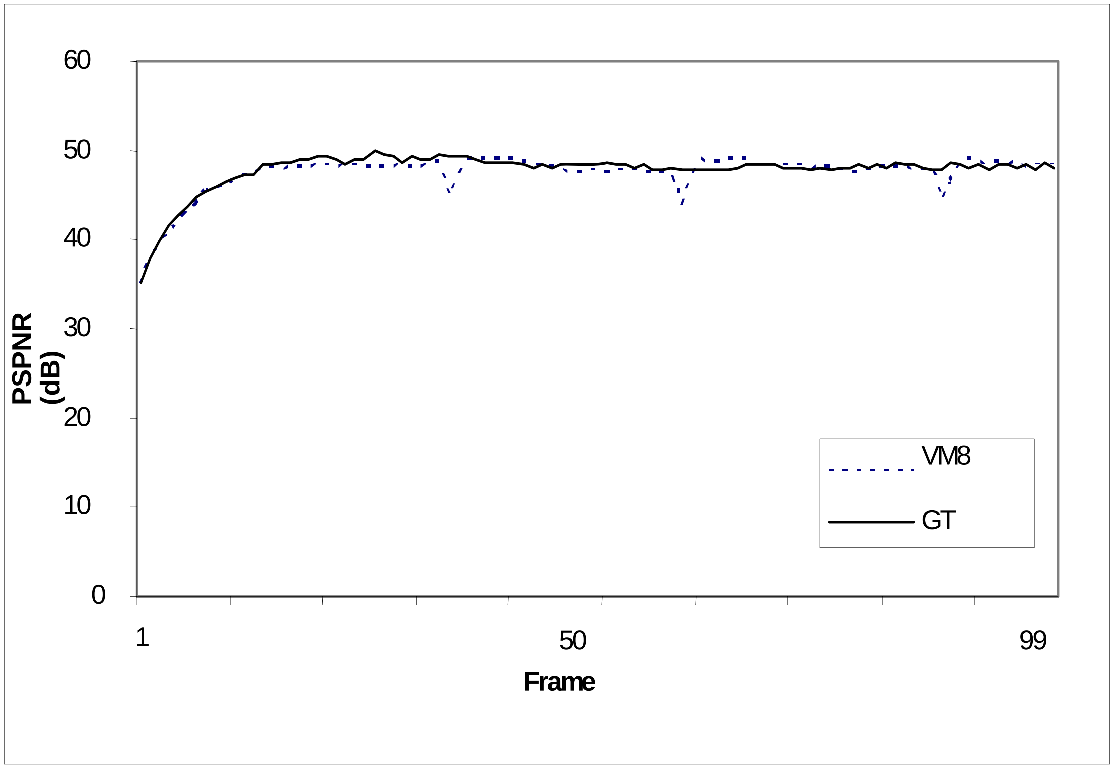

Figure 2.9 PSPNR in the coding of “Container” QCIF at 192kbps and 30Hz
with buffer size = 24kbits

2.6 Summary

In this proposal, we propose adaptive and perceptually tuned motion estimation and rate control schemes for a new paradigm of video compression called *content adaptive video compression*. By adapting to the video characteristics extracted by a online video analysis process, the proposed motion estimation scheme outscores the other predictive motion estimation algorithms in terms of computational cost and visual quality, while showing the adaptability to various types of scenes and abrupt scene change occasions. This scheme has the best overall performance among the compared algorithms after considering the overhead introduced by the video analysis process. This proposed rate control algorithm utilizes a game theoretical approach. The algorithm models the bit allocation problem on the MB level as a bargaining problem. Bit allocation and quantization scale of each MB are decided based on the Nash Bargaining Solution. The proposed algorithm masks the imperceptible distortions by adjusting the initial quality of each MB based on noticeable distortion. The algorithm includes an efficient frame level bit allocation according to the frame coding complexity. The proposed rate control algorithm outperforms the VM8 rate control algorithm in terms of several aspects, including bit rate accuracy, PSNR, PSPNR and the buffer stability.

CHAPTER 3

JOINT MULTIPLE VIDEO OBJECT RATE CONTROL USING NON-COOPERATIVE GAME THEORY

3.1 Introduction

As a standard for multimedia and Web compression, MPEG-4 supports object-based interactivity, high compression, and universal accessibility [28]. In MPEG-4 standard, a scene is composed of one or more audio-video objects, which are separately tracked and compressed but multiplexed into a single MPEG-4 bit stream for transmission. This scheme leads to efficient and scalable compression, and makes MPEG-4 able to support interactive functionalities, such as the manipulation and control of individual objects. Because of the difference with the conventional frame-based standards, such as MPEG-2 and H.263, the rate control algorithms for MPEG-4 need to be redesigned to accommodate to the MVO coding functionality, in order to provide the maximum efficiency. In [3] and [4], Chiang and Zhang have reported a scalable rate control algorithm for both DCT and wavelet-based coders. Vetro and Sun extended this framework with a scheme for JMVO rate control [73], [74]. Their scheme distributes the target bits to video objects proportional to the normalized factors of object size, motion and variance. Ronda and Eckert focused on rate control in real-time applications [63]. The goal was to minimize the average distortion of the objects, to guarantee desired qualities to the most relevant ones, and to keep constant ratios among the object qualities.

The core of JMVO rate control is the bit rate allocation among video objects subject to constraint data rate. The allocation scheme strongly impacts the coding efficiency, the quality of individual audio-visual objects and the overall bit rate accuracy.

In this chapter, we propose a JMVO rate control algorithm that distributes bits to VOPs based on the mixed strategy Nash equilibrium of a non-cooperative bi-matrix game.

For the reason of simplicity, most of the current rate control algorithm adopts a hierarchical bit allocation scheme, where the target bits are first distributed in coarser level then in finer level [3] [34] [35] [52] [73] [74]. In our algorithm, hierarchical structure is employed. The total available bits for a video sequence, which are computed based on the bit rate requirement, are firstly distributed for each time instance; then the bits for a time instance are further allocated to VOPs. The quantization parameter for each VOP is calculated from a rate-distortion model. In each level, we utilize efficient bit allocation schemes to regulate the bits in order to meet the bit rate requirement while optimizing the visual quality. In the time instance level, the number of bits allocated is depended on the number of remaining bits for the remaining VOPs as well as the number of bits used for the previous time instance. In the object level, a non-cooperative matrix game is employed, where each VOP is regarded as a rational player. The bi-matrix strategic game simulates the behaviors of non-cooperative players competing for available bits to optimize their own performance. We look for the "solution" of the game, the mixed strategy Nash Equilibrium, which is the probability distribution of the actions carried by players that optimizes their expected utility, the number of bits. The game is played iteratively. The expected utility of each play will be accumulated. The game will terminate when all available bits for the specified time instance are distributed to VOPs.

3.2 Target Bit Estimation for a Time Instance

The bit allocation scheme in the proposed rate control algorithm is hierarchical. In this section, we describe an approach that estimate the target number of bits for all VOPs in a time instance. This simple yet efficient approach is first appeared in [33], [34].

First, the estimation of target number of bits for each VOP is initially estimated from the remaining bits by:

$$T_i = \frac{R}{N \times n_r} \tag{3.1}$$

where N is the number of video objects and n_r is the number of remaining VOPs.

Taking account of the actual bits consumption for the previous VOP, the initial estimation is further adjusted by Eq. (3.2) in order to maintain a smooth change in target bit budget:

$$T_i' = 0.8 \times T_i + 0.2 S_i \tag{3.2}$$

where S_i is the actual bits for the previous VOP of the ith video object.

Finally, the number of bits for a specific time instance is the sum of estimated bits for all VOPs:

$$T' = \sum_{i=1}^{N} T_i' \tag{3.3}$$

3.3 Buffer Policy

A buffer is placed between encoder and the transmission channel to accommodate bit rate variation in the CBR channel mode. In order to maintain an accurate regulation and avoid buffer overflow and underflow, the buffer control tries to keep buffer fullness in the middle level. The buffer control system can be regarded as a tracking system where the actual buffer fullness level traces the target buffer level, which is a half of the buffer size. Close-loop control is widely used in tracking systems. The feedback of the buffer status is used to adjust the target bit estimation. Proportional-Integral-Derivative (PID) control is widely used in the feedback control systems [7] [54]. It combines to the proportional, integral and derivative feedback compensation components to improve the steady state response and transient response. In this work, we adopt the PID buffer feedback control in our related works [75]. A block diagram of the PID buffer control system is shown in figure 3.1.

The PID buffer feedback adjustment can be formulated as following:

$$\begin{gathered} err_t = 1 - 2B_t / B_s \\ T = T' + T' \times K_p (err_t + K_i \cdot \sum_{i=0}^{t} err_t + K_d \cdot (err_t - err_{t-1})) \end{gathered} \tag{3.4}$$

where B_s is the buffer size, B_t is the buffer fullness at time t, K_p, K_i and K_d are the proportional, integral and derivative control parameters. They are set to 1, 0.01 and 0.4 respectively. The proportional feedback has the intension to reduce the error between the actual buffer level and the target buffer level. The integral feedback intends to eliminate the steady-state error. The negative feedback will increase along the increasing of the accumulated error. Generally, the integral feedback improves the steady-state response but degrades the system's transient response, i.e. the system will overshoot to the accumulated error, such that the feedback keeps increasing while the error is decreasing, which makes the system slowly responding to the current error. To compensate it, a derivative compensator is used to add more stability to the system by reducing the overshoot. Combining the proportional, integral and derivative compensator, and choosing proper control parameters, the buffer level can be accurately control to the desired level.

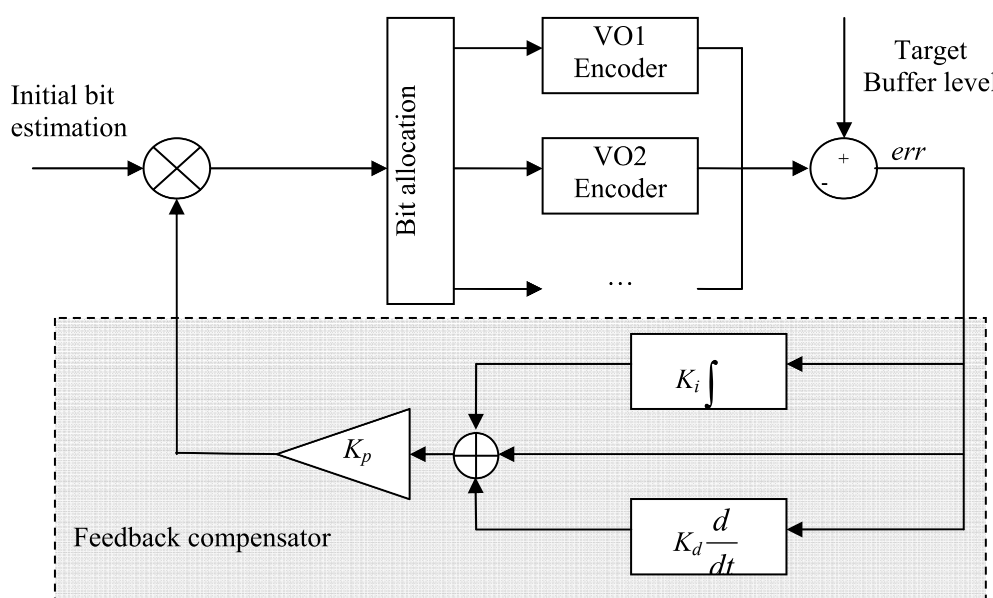

Figure 3.1 PID buffer control diagram

After the PID buffer feedback adjustment, the number of target bits is bounded by [*0.25Br/Fr, 2Br/Fr*], where *Br* is the target bit rate and *Fr* is the frame rate. *Br/Fr* is the average number of bits for a time instance. The objective of this adjustment is to avoid the buffer overflow and guarantee a minimum quality.

3.4 VOP Level Bit Allocation

Given the bit target for a specified time instance, bits are further distributed to VOPs at this time instance. Most current existing rate control algorithms for MVO coding distribute bits proportional to the linear combination of several factors, such as video object size, residual data complexity, motion vector length, etc. This approach is based on the assumption that these factors are linear related to the bit consumption of encoding an object. The target is to obtain a balance visual quality over all objects.

From another point of view to look at the bit allocation problem, we proposed a non-cooperative game approach for bit allocation among VOPs. In the MVO coding, each VOP is encoded with a quantization parameter, which is related to the encoded bits. According to the rate-quantization relation, using a smaller quantization parameter can achieve higher visual quality but produce more bits. Assume the VOPs can choose their quantization parameters independently. They need to consider the other VOPs' choices. If a VOP chooses a small quantization parameter in order to obtain a high quality, it may happen that the other VOPs also choose small quantization parameter such that the total number of bits consume will exceed the target bit budget. On the other hand, if a VOP chooses a large quantization parameter while the other VOPs choose small ones, then it will lose most of the bits. Assuming each VOP is a rational player, the behaviors of the VOPs can be modeled by a bi-matrix strategic game.

In all game theory models the basic entity is a player. A player is an individual or a group of individual making decision. A strategic game is a game in which each player chooses his plan of action once and for all, and all players' decisions are made simultaneously. When choosing an action, each player is not informed the actions chosen by other players. Each player is "rational" in the sense that he is aware of his available strategies and the preference on these strategies, thus he chooses to carry out an action that will maximize his utility. The utility of a player is a function of the combination of actions that carry out by all players.

A strategic game can be described conveniently by a matrix. Figure 3.2 shows an example of two player strategic game. One player's actions are identified with the row and the other player's actions with the column. In each entry of the matrix are the utilities of player A and player B given they choose the corresponding actions.

	H	L
H	a_1,b_1	a_2,b_2
L	a_3,b_3	a_4,b_4

Figure 3.2 A sample of bi-matrix game

An interpretation of game in Figure 3.2 is that: Both player (player A in column and player B in row) have a strategy space {H, L}. When player A chooses H and player B choose H, player A's utility is a_1 and player B's is b_1. When player A chooses H and player B chose L, player A's utility is a_2 while player B's is b_2.

A strategic game can be utilized to model the behaviors of VOPs. We define a strategic game as following. For the ease of description, we limit the number of objects to two. More objects can be applied in the same framework, but the computational complexity will increase accordingly. In this game, the objects are regard as players. The strategy space of the player is defined on the set of quantization parameters. In the MPEG-4 environment, the strategy space is {1,..,31}. Each entry in the matrix contains the utilities of players given corresponding combination of actions taken by the two players. The utility is defined as follow:

$$u_1(q_1,q_2)=\begin{cases} R_1(q_1) & \text{if } R_1(q_1)+R_2(q_2)\le s \\ 0 & \text{if } R_1(q_1)+R_2(q_2)>s \end{cases} \tag{3.5}$$

$$u_2(q_1,q_2)=\begin{cases} R_2(q_2) & \text{if } R_1(q_1)+R_2(q_2)\le s \\ 0 & \text{if } R_1(q_1)+R_2(q_2)>s \end{cases} \tag{3.6}$$

where u_1 and u_2 are utilities of player 1 and player 2 respectively. R_1 and R_2 are the rate quantization functions of player 1 and player 2. q_1 and q_2 represent the strategies chosen by player 1 and player 2. s is the number of remaining bits for texture coding.

$$s=T-\sum_i T_{hdr,i} \tag{3.7}$$

where $T_{hdr,i}$ is the total number of bits used for coding the shape, motion and header of the previous VOP of the ith video object.

Different rate-quantization functions have been proposed in the literatures to model the rate quantization relation [6] [27] [69] [76]. In our work, we employ a quadratic model which first appeared in [4], due to its wide usage and accuracy. The rate quantization function is shown in (3.8)

$$R_i = \frac{X_{1,i} \times MAD_i}{q_i} + \frac{X_{2,i} \times MAD_i}{q_i^{\;2}} \tag{3.8}$$

where $X_{1,i}$ and $X_{2,i}$ are the model parameters for video object i.

		Player 2			
		1	2	…	31
Player 1	1	$u_1(1,1)$	$u_1(1,2)$	…	$u_1(1,31)$
	2	$u_1(2,1)$	$u_1(2,2)$	…	$u_1(2,31)$
	…	…	…	…	…
	31	$u_1(31,1)$	$u_1(31,2)$	…	$u_1(31,31)$

(a)

		Player 2			
		1	2	…	31
Player 1	1	$u_2(1,1)$	$u_2(1,2)$	…	$u_2(1,31)$
	2	$u_2(2,1)$	$u_2(2,2)$	…	$u_2(2,31)$
	…	…	…	…	…
	31	$u_2(31,1)$	$u_2(31,2)$	…	$u_2(31,31)$

(b)

Figure 3.3 The utility matrices (a) matrix of player 1 (b) matrix of player 2

The utilities defined (3.5) and (3.6) can be interpreted as follows: the utility of a player is the number of bits required to encode it with a chosen quantization level; if the total number of bits requested by both players is exceeding the total available bits for both players, both players get nothing.

The strategic game for VOP level bit allocation can be described as in figure 3.3. Each entry of the matrix contains a tuple <u_1, u_2>. Therefore, it is also called a bi-matrix game. We can also use two matrices: one represents the utilities of player 1 and the other represents player 2.

Assume the players' choices are nondeterministic, i.e., the players have some probability distributions over the set of strategies. The probability distribution is called a mixed strategy of a player. For example, if x ($x = \begin{pmatrix} x_1 & x_2 & \cdots & x_{31} \end{pmatrix}$, $x_i \geq 0$ and $\sum_i x_i = 1$) is the mixed strategy of player 1 in the game of Figure 3.3, x_1 is the probability of player 1 choosing 1 and x_2 is the probability of player 1 choosing 2. Obviously, the expect utility of each player is depended on both players' mixed strategies. Assume A is the matrix of player 1's utilities and B is player 2's. The expected utilities of player 1 and player 2 are xAy^T and $x^T By$, where x is the mixed strategy of player 1 and y is the mixed strategy of player 2. Assuming all players are rational, the game is expected to converge to equilibrium. A pair of mixed strategies (x^*, y^*) for a bi-matrix game (A, B) is said to be in equilibrium if, for any other mixed strategies, x and y:

$$x^* A y^{*T} \geq x A y^{*T} \text{ and } x^* B y^{*T} \geq x^* B y^T \tag{3.9}$$

(x^*, y^*) is called mixed strategy Nash equilibrium. In other words, a mixed strategy is said to be Nash equilibrium if no player has any positive reason for changing his mixed strategy, assuming none of the other players will change their mixed strategy. $x^* Ay^{*T}$ and $x^* By^{*T}$ are the expected utilities.

In each play of the game, the players will choose the mixed strategy Nash equilibrium, since it is the optimal strategy for a player in the "Nash" sense. The expected utilities of the equilibrium are allocated to player 1 and player 2 respectively. The game is played iteratively. After each play, the remaining number of bits is updated by taking out the number of bits allocated and the bi-matrix is updated accordingly. The game will terminate when all the remaining bits are allocated.

(3.9) can be converted to the form of linear complementary problem (LCP). LCP can be solved by a complementary pivot algorithm due to Lemke [37]. According to the characteristics of this game, we derive a simple process that simplifies the computation of expected utility.

Assume we have constructed matrix A and B based on the rate-quantization function and the available bits. In matrix A, if there exists some k, such that $u_1(i,j) \leq u_1(k,j)$ for all j, then action i of player 1 is said to be dominated by action k. i is the dominated strategy and k is the dominant strategy, i.e., row i is the dominated by row k. Similarly, in matrix B, if there exists if there exists some k, such that $u_2(i,j) \leq u_2(i,k)$ for all i, then strategy j of player 2 is said to be dominated by strategy k, i.e., column j is the dominated by column k. From the definition of Nash equilibrium, we know that if a strategy of a player is dominated by other strategies of the same player, there is no positive reason the player will play this strategy since there is always a better choice. Thus, the probability of the player carrying out this strategy is zero. Therefore, we can eliminate the dominated row and column in the matrices without interfere the solution of game.

If the total bits in an entry do not exceed the available bits, the following relations are held:

1. $u_1(i,j) \geq u_1(m,j)$ for $i \leq m$, and $u_2(i,j) \geq u_2(n,j)$ for $j \leq n$, since the rate-quantization function is monotonously decreasing;

2. $u_1(i,j) = u_1(i,m)$ for any j and m, and $u_2(i,j) = u_2(n,j)$ for any i and n, since a object's bits is only related to its own quantization level.

Therefore, after eliminating the dominated row and column, matrix A and B will be simplified to a form as shown in Figure 3.4.

A pair of mixed strategy (x,y) is a Nash equilibrium, if $Ay^T = ke_n^T$ and $xB = re_n$.
Proof: $xAy^T = x'Ay^T = k$ for any x' and $xBy^T = xBy'^T = r$ for any y'; according to the definition of mixed strategy Nash equilibrium in (3.9), mixed strategy (x,y) is a Nash equilibrium.

With matrices in the above form, (x,y) with such property can be find by solving equations:

$$\begin{gathered}
a_1 y_n = a_2 (y_{n-1} + y_n) \\
a_2 (y_{n-1} + y_n) = a_3 (y_{n-2} + y_{n-1} + y_n) \\
\cdots \\
a_{n-1} (y_2 + y_3 + \cdots + y_n) = a_n (y_1 + y_2 + \cdots + y_n)
\end{gathered}$$

$$
\begin{aligned}
b_1 x_n &= b_2(x_{n-1}+x_n) \\
b_2(x_{n-1}+x_n) &= b_3(x_{n-2}+x_{n-1}+x_n) \\
&\cdots \\
b_{n-1}(x_2+x_3+\cdots+x_n) &= b_n(x_1+x_2+\cdots+x_n)
\end{aligned}
$$

Player 2

Player 1		q_1	q_2	…	q_{n-1}	q_n
	p_1	0	0	…	0	a_1
	p_2	0	0	…	a_2	a_2
	…	…	…	…	…	…
	p_{n-1}	0	a_{n-1}	…	a_{n-1}	a_{n-1}
	p_n	a_n	a_n	…	a_n	a_n

Player 2

Player 1		q_1	q_2	…	q_{n-1}	q_n
	p_1	0	0	…	0	b_n
	p_2	0	0	…	b_{n-1}	b_n
	…	…	…	…	…	…
	p_{n-1}	0	b_2	…	b_{n-1}	b_n
	p_n	b_1	b_2	…	b_{n-1}	b_n

Figure 3.4 The simplified matrices by eliminating the dominated rows and columns, where $a_i > a_j > 0$ and $b_i > b_j > 0$, for $i<j$

Solving these equations, we have:

$$x_1 = \frac{b_{n-1} - b_n}{b_{n-1}},\ x_n = \frac{b_n}{b_1} \text{ and } x_i = \frac{b_{n-i} - b_{n-i+1}}{b_{n-i}} \cdot \frac{b_n}{b_{n-i}} \text{ for } i \neq 1 \text{ and } i \neq n \qquad (3.10)$$

$$y_1 = \frac{a_{n-1} - a_n}{a_{n-1}},\ y_n = \frac{a_n}{a_1} \text{ and } y_j = \frac{a_{n-j} - a_{n-j+1}}{a_{n-j}} \cdot \frac{a_n}{a_{n-j}} \text{ for } j \neq 1 \text{ and } j \neq n \qquad (3.11)$$

The expected utility of (x,y) is (a_n, b_n).

The following procedure distributes the bits to the VOPs:

1. Construct matrix A and B, based on the objects' rate-quantization functions and the remaining bits for texture coding.

2. Eliminate the dominated rows and columns.

3. Distribute the bits according to the expected utility of the mixed strategy Nash equilibrium to the VOPs.

4. Subtract the distributed bits from the remaining bits. If there are bits available, go to step 1.

3.5 Quantization Level Calculation and Frame-skipping Control

Given the model parameter $X_{1,i}$ and $X_{2,i}$ and MAD_i, with the estimated number of bits for VOP_i the quantization parameter *QP* for the VOP_i can be calculated using (3.8). To maintain a smooth change of visual quality, the change of *QP* is restricted to 25% of the VOP in the previous time instance. Conforming to the standard the *QP* is limited to vary between 1 and 31.

To avoid buffer overflow, the encoder will skip encoding and transmitting one or more frames when the buffer occupancy is too high. The same method as VM8 [74] is employed for frame-skipping control. Frame-skipping controls are performed in both pre-encoding stage and post-encoding stage. In pre-encoding stage, if the VOP header coding consumes too many bits such that there are no bits for the texture coding, frames will be skipped. In the post-encoding stage, if the buffer occupancy is higher than 80% of the buffer size, frames will be skipped to avoid buffer overflow.

3.6 Experiments and Results

We implement the proposed algorithm on the Momusys MPEG-4 software verification model (VM 8.0). Five common test sequences are used in the experiments. They are the first two objects of "news", "bream", "coastguard", "Children" and "container". For all sequences, we encode the first 150 frames and the frame rate is 30Hz. Different target bit rate are tested, from 128Kbps to 384Kbps with a step of 64Kbps.

First, the bit rate accuracy is evaluated. As shown in Table 2.1, the results verified that our algorithms can accurately control the bit rate to the desired value. The bit rate accuracy maintains above 98%. Compared with VM8 JMVO rate control algorithm, our algorithm achieves a similar accuracy level in bit rate control.

Rate control has strong impacts on the visual quality of the reconstructed video signal. Peak-signal-to-noise-ratio (PSNR) is widely used to evaluate the reconstructed signal quality. We measure the PSNR of each VOP and average the PSNR values of all the VOP in a video object to obtain the reconstructed PSNR of a video object. As reported in Table 2.1, with the bit allocation among video objects based on the non-cooperative strategic game, our rate control algorithm can properly distribute the bits to video object. The distribution is closely related to the utility function of each video object. For example, in the coding of “news” sequence at 30Hz and 128kbps, the foreground object (video object 1) obtains a larger portion of bit budget through the competition with the background object (video object 2). The coding complexity of video object 1 is higher than video object 2. By taking the same action; video object 1 consumes more bits than video object 2. Thus, video object 1 is more aggressive to obtain bits than the video object 2, which is embodied by their utility functions. Therefore, video object 1 wins more bits than video object 2.

The buffer occupancy is examined and plotted in Figure 3.5-3.9. The detail frame by frame buffer occupancy curve shows the fluctuation of buffer fullness. If the fluctuation level is large, the algorithm is much easier to overflow or underflow the buffer. From the buffer occupancy figure, it can be observed that our rate control algorithm efficiently maintains the buffer fullness around the middle level of the buffer. The buffer occupancy is maintained around 40%-60% of the buffer size. Comparing with VM8 JMVO rate control algorithm, the buffer occupancy of proposed algorithm is smoother and has fewer fluctuations. For example, in the coding of “news” sequence at 30Hz and 256kbps, the buffer occupancy with VM8 JMVO rate control algorithm fluctuates around 30%-70%, while the buffer occupancy with our rate control algorithm is around 40%-65%. Another example is the coding of “container” sequence at 30Hz and 256kbps. With VM8 rate control algorithm, the buffer occupancy varies form 10%-60%, and our algorithm produces a smoother change of the buffer fullness level and the variation is about 15%-20% around the middle of the buffer size. Thanks to the PID control, our rate control algorithm produces a quick response to the buffer level change. For instance, in the coding of "coastguard" at 30Hz and 256kbps as shown in figure 3.7, our algorithm much quickly responds to the buffer error than the VM8 MVO rate control algorithm.

Besides the average PSNR of each video object, the frame to frame PSNR fluctuation is also a measure of video quality. Large PSNR fluctuation between consecutive encoding frames will degrade the visual quality. Figure 3.10-3.19 shows the frame to frame PSNR curves. In this comparison, the proposed algorithm and VM8 algorithm have similar level of fluctuation.

Table 2.1 also shows the number of skipped frames. Due to the smooth buffer control and proper bit allocation, fewer frames are skipped with our algorithm compared with VM8 JMVO rate control. For example, in the coding of "children" sequence, our algorithm skipped few frames compared with VM8 JMVO rate control.

Table 3.1 The performance comparison between game theory algorithm and the VM8 reference algorithm

Sequence	Alg.	Target Bits Rate	Actual Bit Rate				skipped frames	PSNR	
			VO1	VO2	Total	Accuracy		VO1	VO2
news	VM8	384	205.1	177.2	382.2	99.5%	2	41.18	41.44
	GT	384	242.7	139.7	382.4	99.6%	0	42.10	40.07
	VM8	320	172.0	145.1	317.0	99.1%	0	39.91	40.20
	GT	320	202.6	114.2	316.8	99.0%	0	40.90	38.88
	VM8	256	137.9	117.4	255.4	99.8%	0	38.42	39.14
	GT	256	165.1	88.7	253.8	99.1%	0	39.56	37.12
	VM8	192	98.2	92.6	190.8	99.4%	0	36.16	37.33
	GT	192	117.9	72.8	190.7	99.3%	0	37.22	35.78
	VM8	128	63.8	63.4	127.1	99.3%	0	33.59	34.92
	GT	128	75.2	51.7	126.8	99.1%	0	34.51	33.70
bream	VM8	384	89.1	297.0	386.1	99.5%	0	44.93	33.69
	GT	384	97.5	284.2	381.7	99.4%	0	45.33	33.38
	VM8	320	56.0	263.0	319.0	99.7%	0	43.93	32.89
	GT	320	61.1	258.2	319.2	99.8%	0	44.30	32.74
	VM8	256	52.4	203.1	255.6	99.8%	0	43.71	31.22
	GT	256	54.9	200.9	255.8	99.9%	0	44.05	31.10
	VM8	192	40.9	150.4	191.3	99.6%	0	42.98	29.34
	GT	192	39.8	151.8	191.5	99.8%	0	43.04	29.43
	VM8	128	28.8	99.0	127.8	99.9%	0	42.26	27.02
	GT	128	28.6	98.3	126.8	99.1%	0	42.22	26.91

Table 3.1 - *Continued*

coastguard	VM8	384	192.9	188.2	381.1	99.3%	0	39.05	40.95
	GT	384	215.8	165.0	380.8	99.2%	0	39.86	39.87
	VM8	320	163.2	153.7	316.9	99.0%	0	37.68	39.22
	GT	320	177.8	139.2	317.0	99.1%	0	38.33	38.42
	VM8	256	135.1	118.9	254.0	99.2%	0	36.21	37.23
	GT	256	147.2	106.4	253.6	99.1%	0	36.79	36.16
	VM8	192	102.6	87.4	190.0	98.9%	0	34.25	34.62
	GT	192	112.0	78.3	190.3	99.1%	0	34.76	33.79
	VM8	128	68.7	58.1	126.8	99.0%	0	31.36	31.79
	GT	128	75.6	51.4	127.0	99.2%	0	31.91	30.89
children	VM8	384	207.2	171.7	379.0	98.7%	10	29.94	34.51
	GT	384	239.7	138.3	378.0	98.4%	5	30.97	30.85
	VM8	320	183.6	131.8	315.5	98.6%	6	28.90	30.46
	GT	320	190.3	125.8	316.1	98.8%	4	29.08	29.71
	VM8	256	145.8	107.0	252.8	98.7%	7	27.40	28.88
	GT	256	149.7	102.8	252.5	98.6%	5	27.47	28.27
	VM8	192	105.8	83.9	189.7	98.8%	9	25.50	27.74
	GT	192	110.4	79.5	190.0	99.0%	6	25.70	25.65
	VM8	128	72.1	54.6	126.6	98.9%	6	23.14	23.32
	GT	128	75.6	50.9	126.6	98.9%	3	23.40	21.69

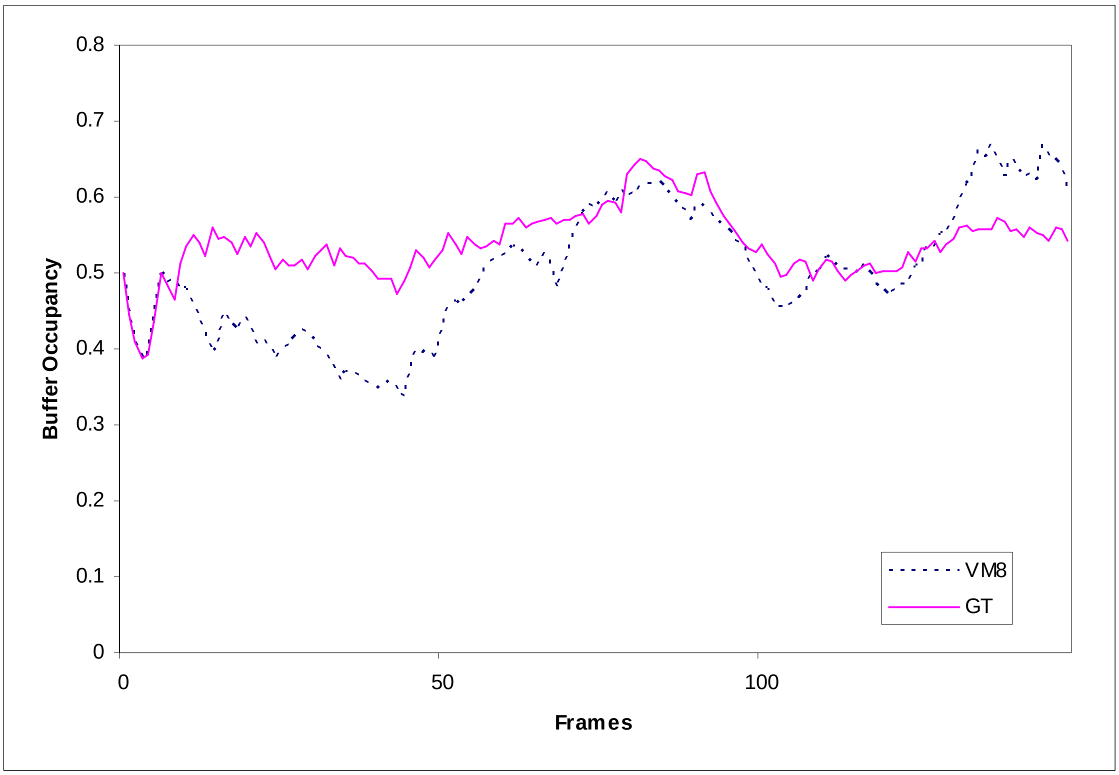

Figure 3.5 Buffer occupancy in the coding of “news” at 30Hz and 256kbps

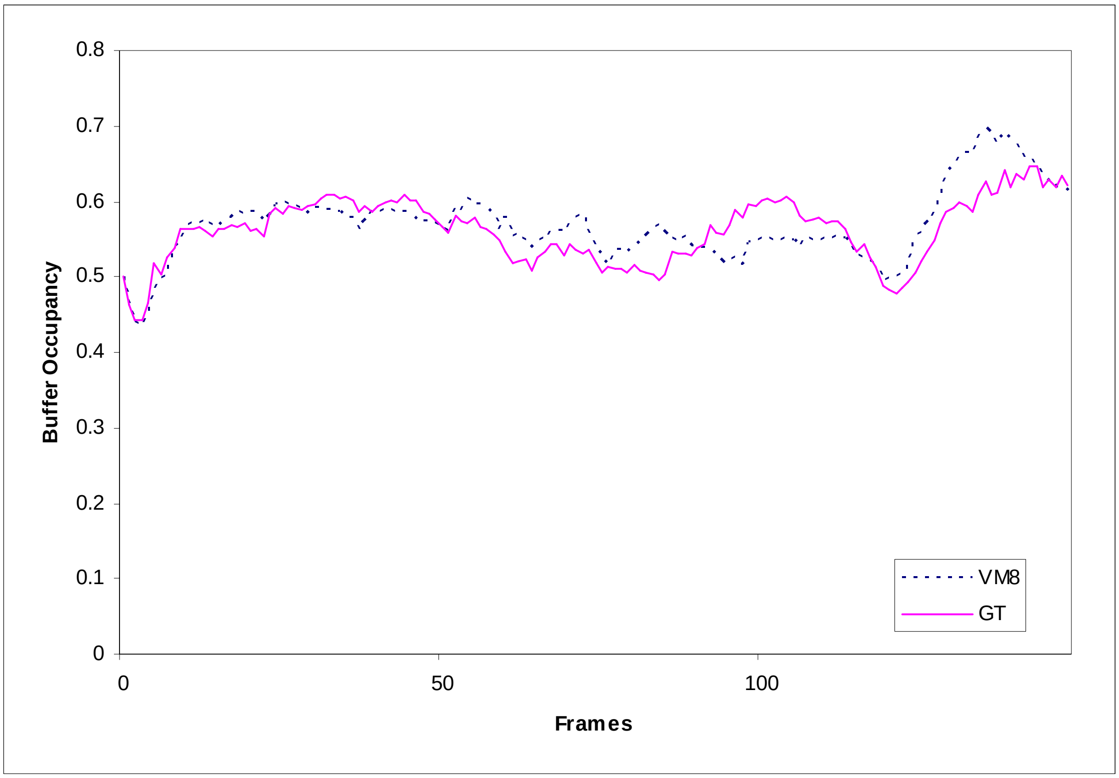

Figure 3.6 Buffer occupancy in the coding of “bream” at 30Hz and 256kbps

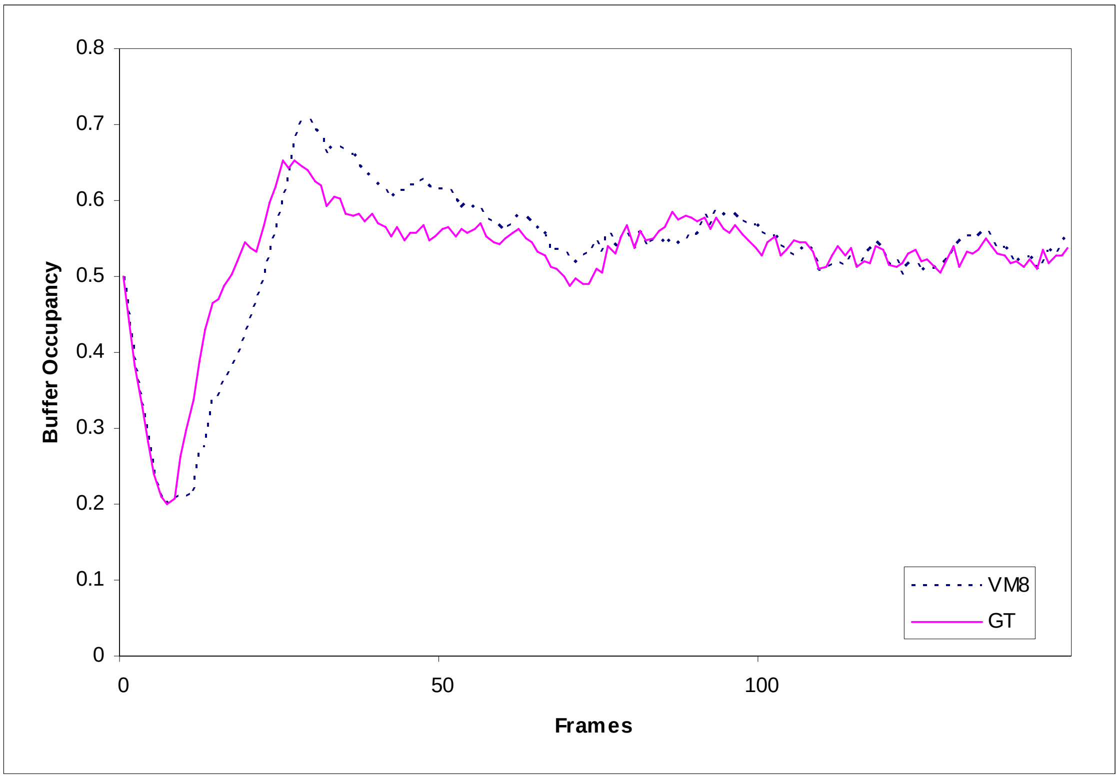

Figure 3.7 Buffer occupancy in the coding of “coastguard” at 30Hz and 256kbps

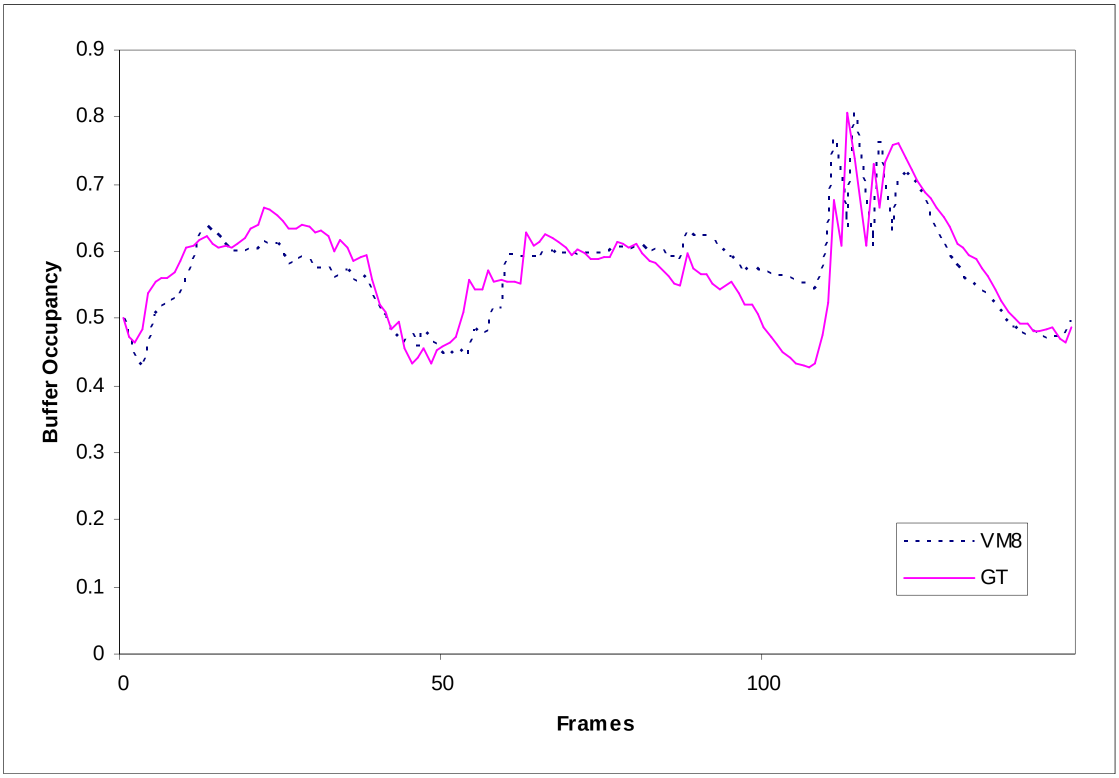

Figure 3.8 Buffer occupancy in the coding of “children” at 30Hz and 256kbps

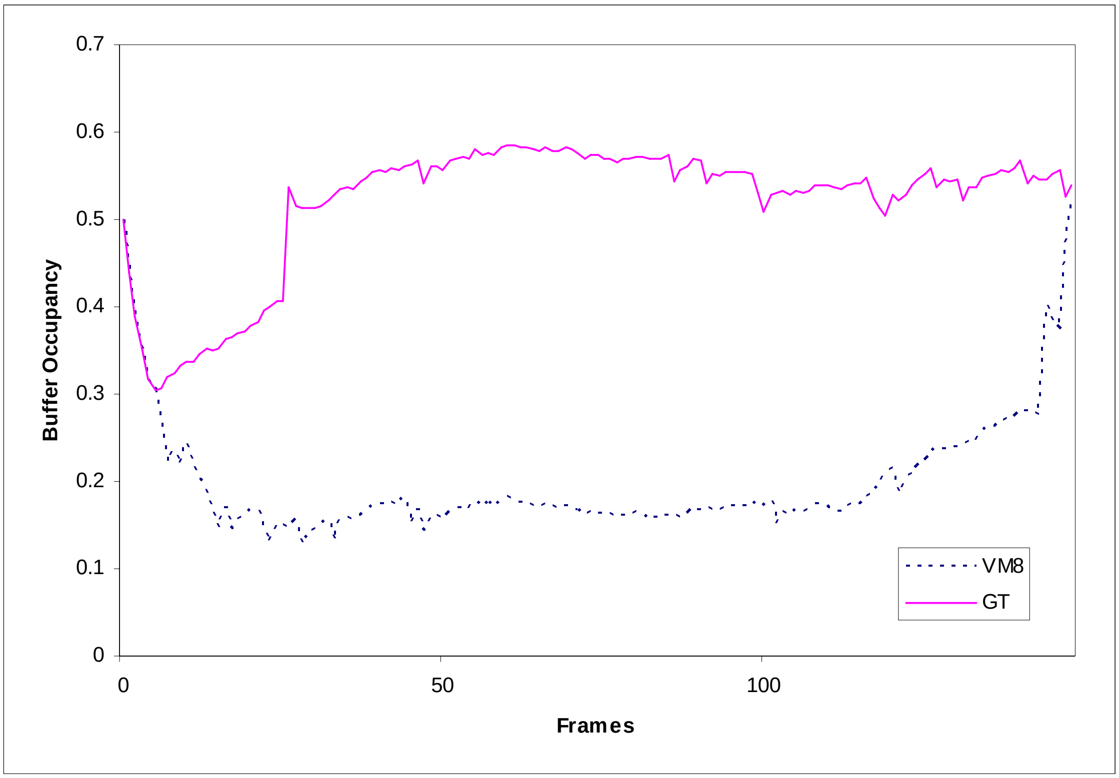

Figure 3.9 Buffer occupancy in the coding of “container” at 30Hz and 256kbps

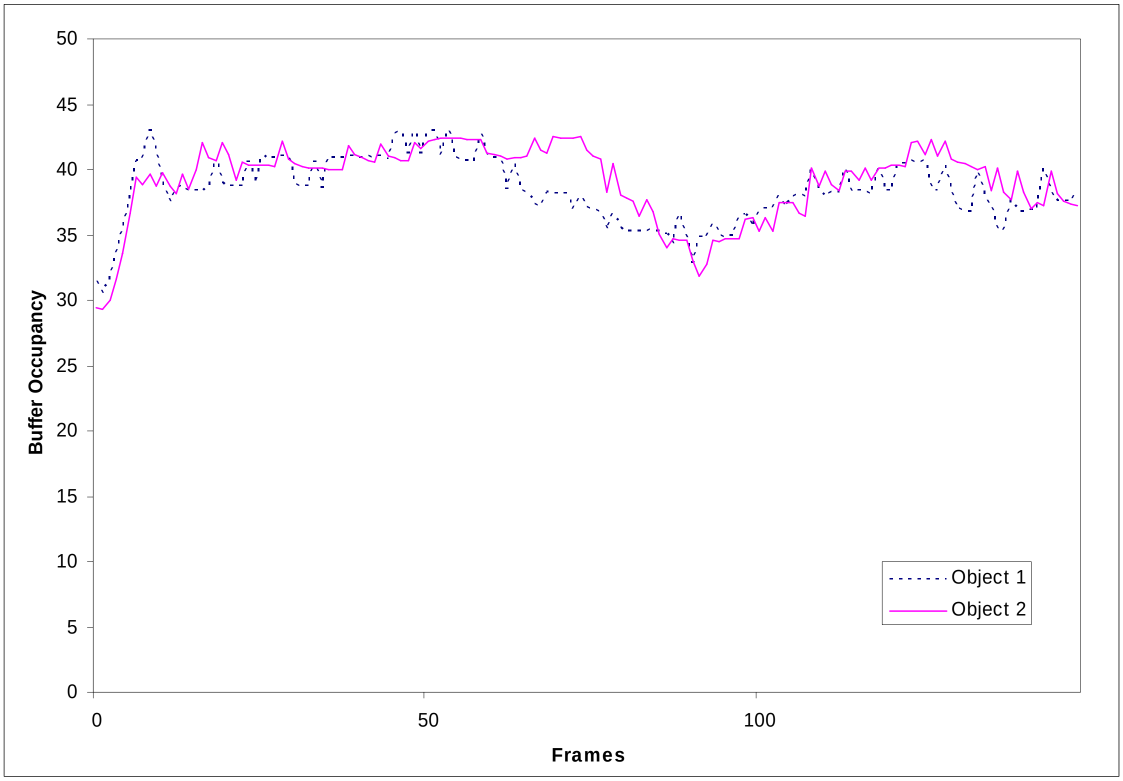

Figure 3.10 PSNR of “news” using VM8 at 256kbps

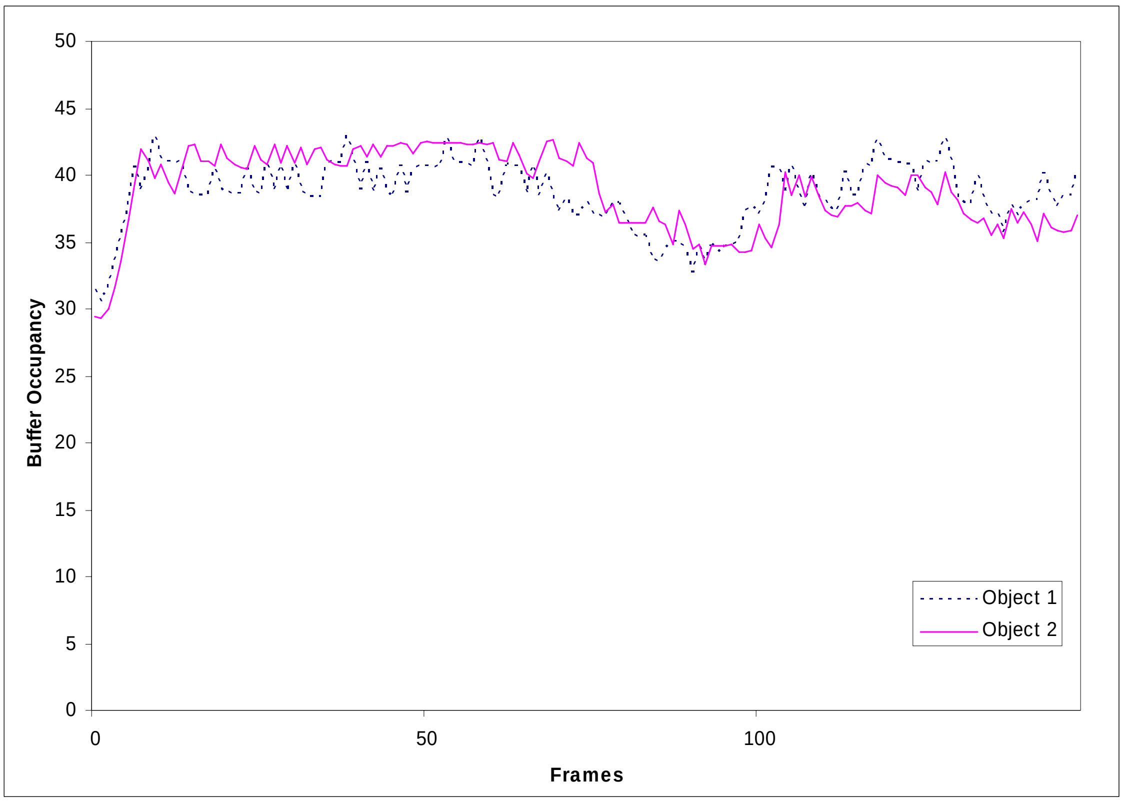

Figure 3.11 PSNR of “news” using the proposed rate control at 256kbps

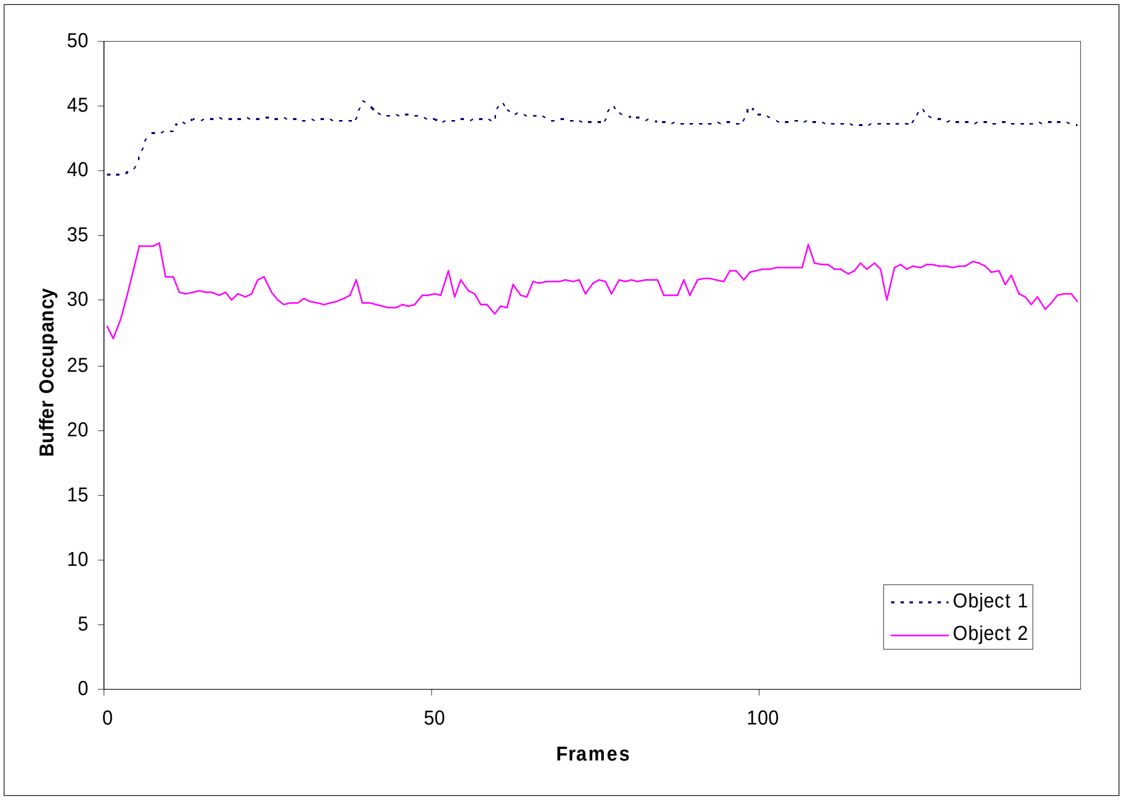

Figure 3.12 PSNR of "bream" using VM8 rate control at 256kbps

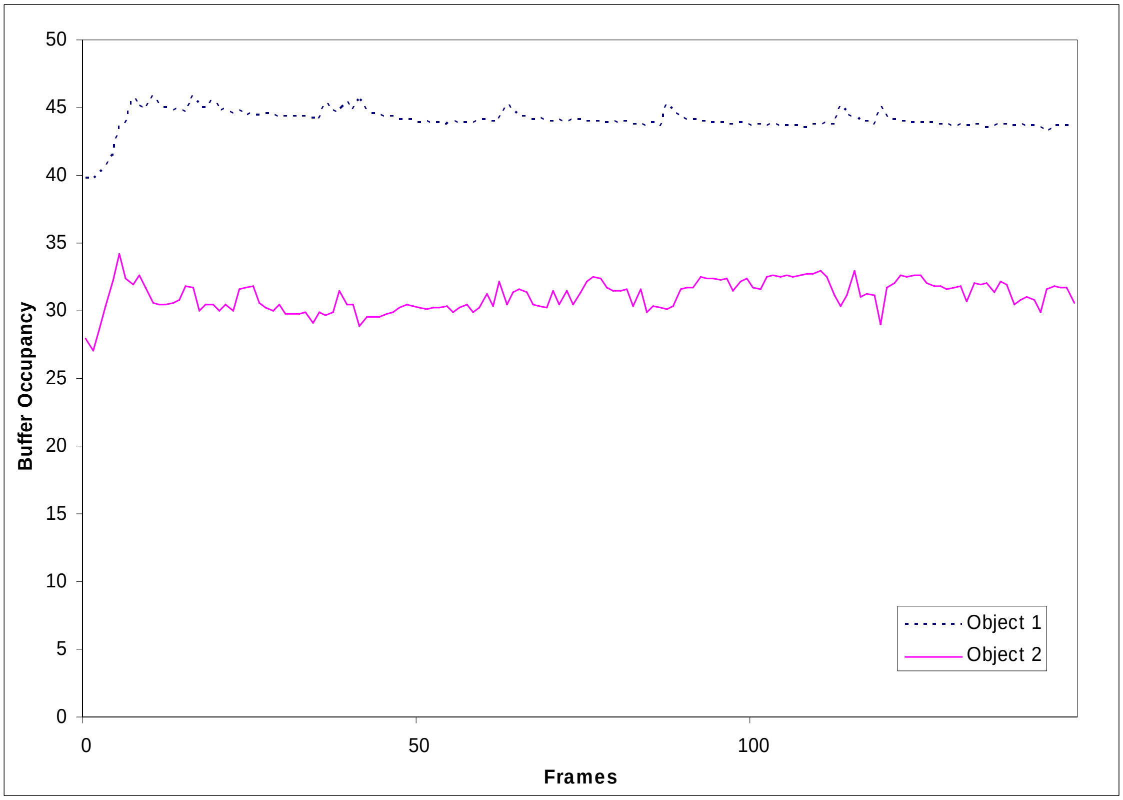

Figure 3.13 PSNR of “bream” using the proposed rate control at 256kbps

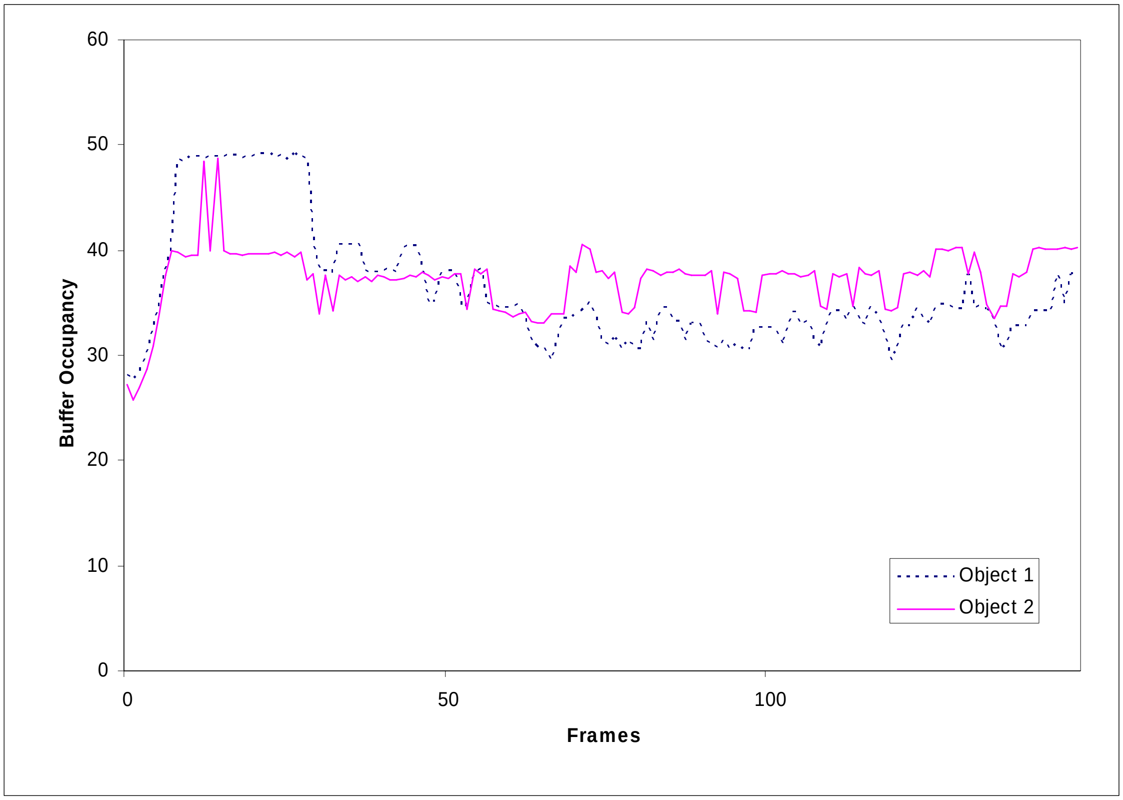

Figure 3.14 PSNR of “coastguard” using VM8 rate control at 256kbps

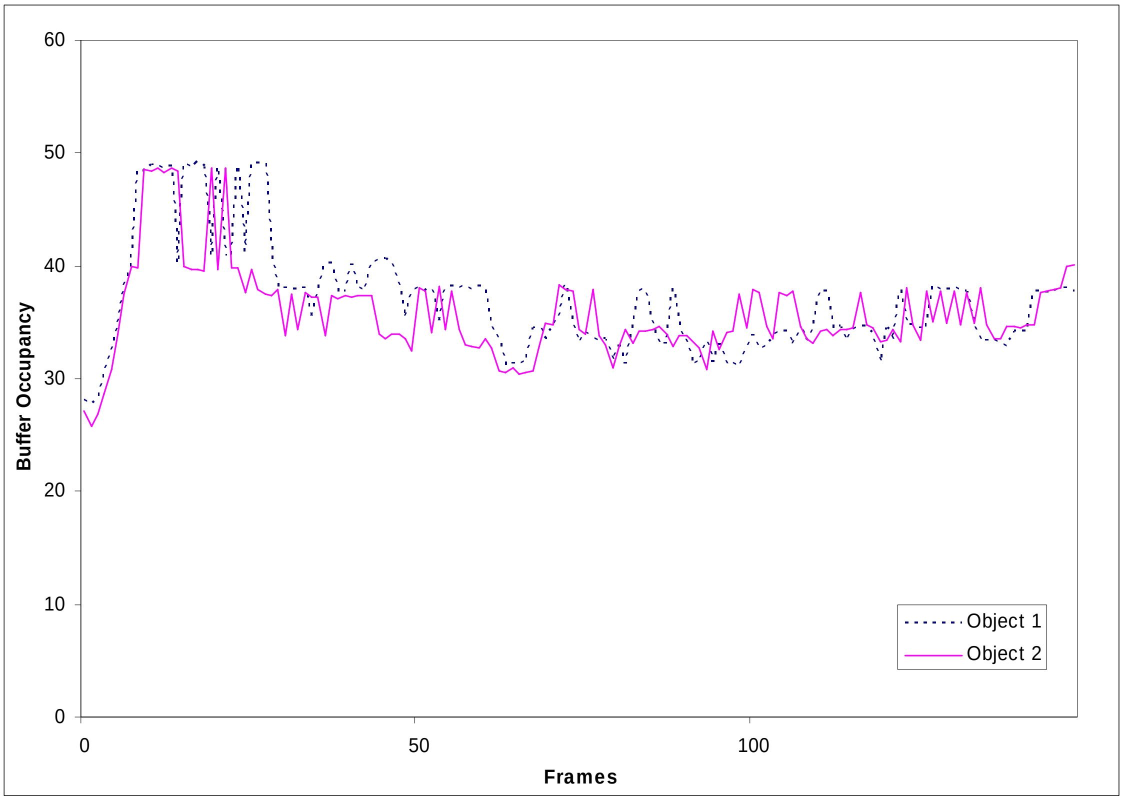

Figure 3.15 PSNR of “coastguard” using the proposed rate control at 256kbps

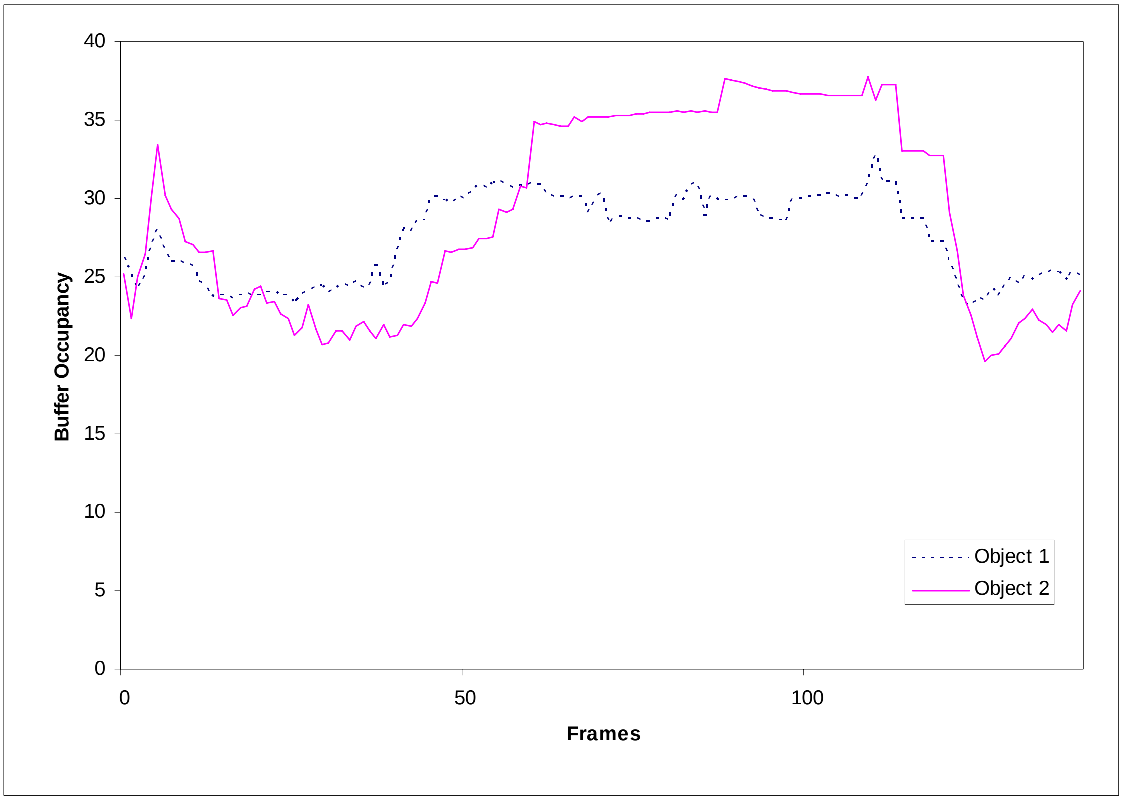

Figure 3.16 PSNR of “children” using VM8 rate control at 256kbps

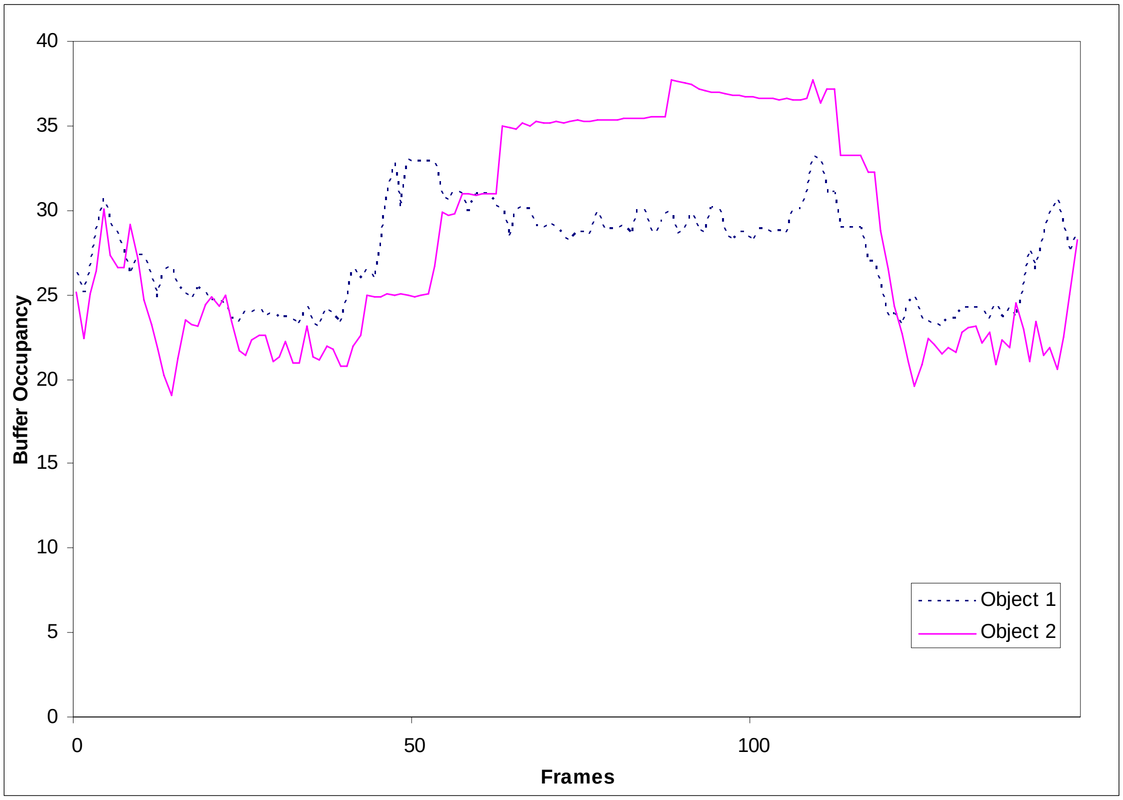

Figure 3.17 PSNR of “children” using the proposed rate control at 256kbps

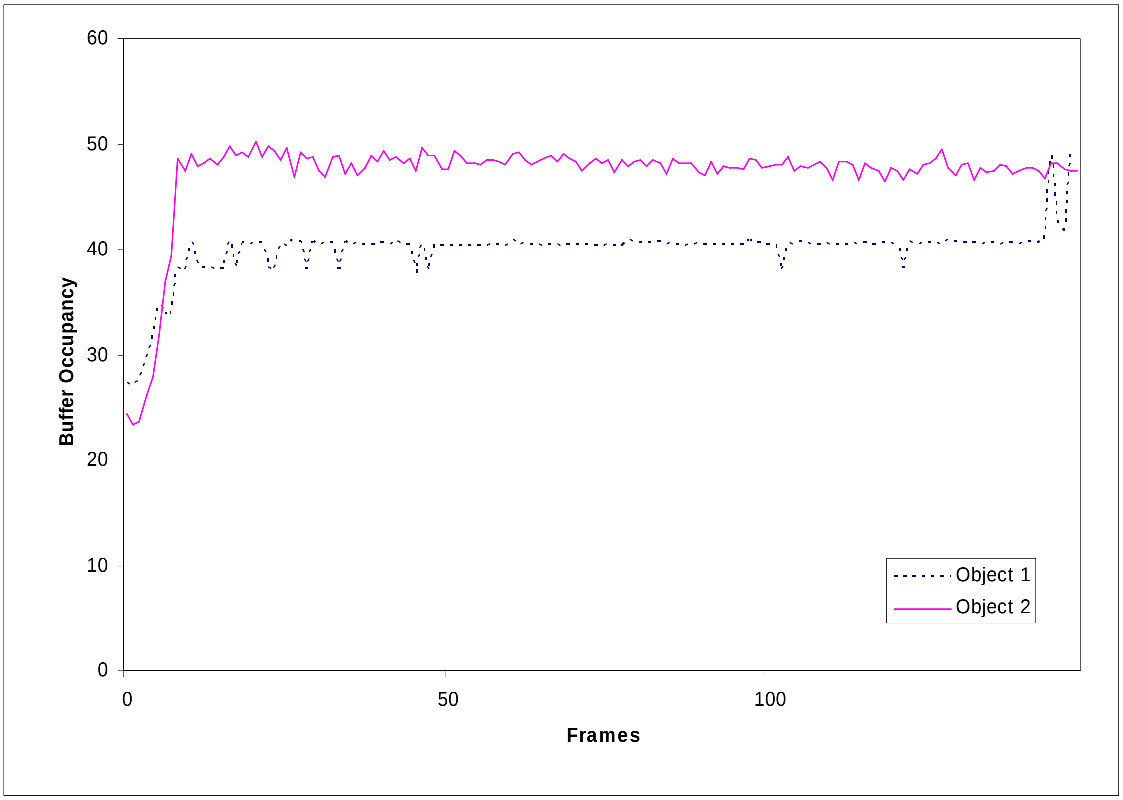

Figure 3.18 PSNR of “container” using VM8 rate control at 256kbps

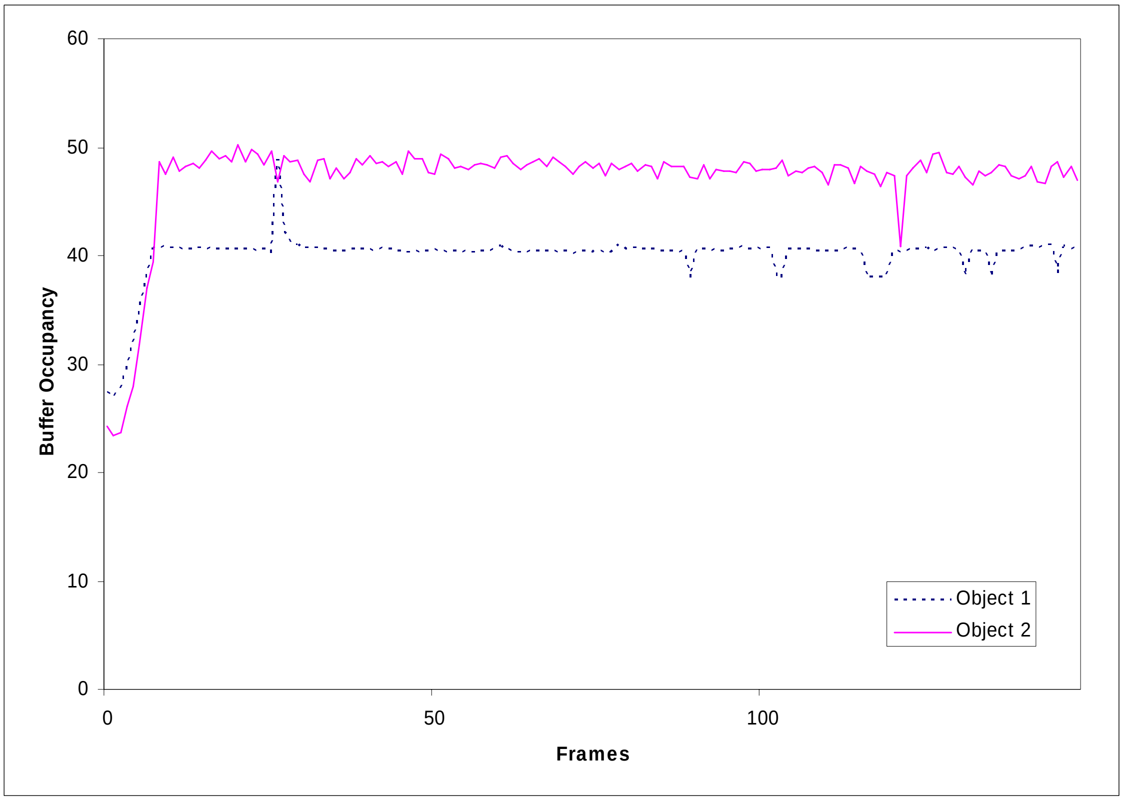

Figure 3.19 PSNR of “container” using the proposed rate control at 256kbps

3.7 Summary

In this chapter, a JMVO rate control algorithm based on non-cooperative game theory is proposed. A bi-matrix strategic game is employed to model the behavior of VOPs. The utility functions are defined based on the rate-quantization relations of video objects. The game is non-deterministic. The mixed strategy Nash equilibrium solution is used to compute the bit allocation. Based on the characteristics of the game, a simple approach computing the mixed strategy Nash equilibrium is derived.

The experiment results show that the proposed game theory algorithm achieves accurate bit rate control and smooth buffer occupancy to avoid buffer overflow and underflow. The proposed algorithm can automatically and properly allocate bits among video objects based on the non-cooperative game.

CHAPTER 4

SUMMARY AND FUTURE WORKS

Rate control plays an important role in the video compression and communication. It has strong impact on the visual quality and the bandwidth utility. Towards the goal of optimizing the current existing video compression algorithms, we have presented and discussed several motion estimation and rate control algorithms. As a conclusion, we will summarize the major contributions followed by the discussion of future research.

4.1 Summary of Contributions

Realizing the power of game theory in solving multiple objective optimization problems with multiple decision makers, we designed cooperative and non-cooperative games for rate control in video coding. To our knowledge, our work is the first attempt to game theory to solve the video compression problems. Two algorithms have been proposed in our study, based on cooperative and non-cooperative game theory respectively.

In the first algorithm, we formulated the MB level rate control problem as a multiple objective optimization problem. We converted the problem to a bargaining game and derived an optimal and fair bit allocation strategy from the Nash Bargaining Solution. Due to the non-uniform distortion visibility of human eyes, the perceptual redundancy can be further explored by incorporating the perceptual sensitivity of human vision system into the cooperative game theoretical rate control framework. The proposed algorithm successively achieves accurate bit rate control, smooth buffer fullness and promising visual quality.

Object-based video coding, such as MPEG-4, support the coding and transmission of independent video objects. JMVO rate control is a efficient rate control scheme for MVO coding. We extended our work to cope with the MVO coding. The attempt of using non-cooperative matrix game for object level bit allocation is made. Each VOP is regarded as a player. Players play the game by non-cooperative choosing a proper quantization strategy to optimize its expected utility, which is the expected number of bit obtained. The game repeats until all the bits are distributed. It leads to a bit allocation corresponding to the mixed strategy Nash equilibrium, a steady status that both players will converge to. Our JMVO rate control algorithm achieves accurate bit rate control and smooth buffer occupancy to avoid buffer overflow and underflow.

4.2 Future Works

4.2.1 Joint Multiple Video Object Rate Control using Cooperative Game

In chapter 3, we have presented an object level rate control algorithm that employs non-cooperative game theory to model the behaviors of the VOPs and solves the bit allocation problem using a bi-matrix strategic game. In a non-cooperative game, there is no communication and binding commitment allowed. Individual players play selfishly to maximizing their own utilities. On the contrary, cooperative game allows communications and binding commitments among players. The players play cooperatively. The utilities are optimized for a group of players. In chapter 2, we proposed a cooperative game for MB level rate control, which optimizes the perceptual quality under the constraint of fair bit allocation among MBs in a frame. In the future research, we plan to explore cooperative game theory models for the object level bit allocation in MVO coding. In MVO coding, such as MPEG-4, video objects are encoded separately but composite together into a scene. The human vision system perceives a scene composition instead the individual objects. Therefore, cooperation in the video object coding may benefit the visual quality. Refer to the cooperative game theory framework for MB level rate control in chapter 3, we can apply cooperative game in object level rate control. Due to the disparity of the object characteristics and the user's recognition, video objects may have different importance to the observer. The bit allocation problem among multiple objects can be mapped into a generalized bargaining problem with asymmetric players. The axiom of symmetry in the Nash bargaining solution can be relaxed in a generalized bargaining problem such that the players can have different powers in the bargaining. By translating the importance into the bargaining power, Nash bargaining solution can be used to solve the object level bit allocation problem in MVO coding.

4.2.2 Improved Utility Function in Game Theoretical Models

A virtue of game theoretical models and algorithms for rate control is the joint optimization of multiple utility functions by multiple decision makers. By taking advantage of this property, in chapter 2 we have proposed the utility functions that are associated with MB's perceptual redundancies of human vision system. But the current criterion for quantitative perceptual redundancy only takes into account the spatial redundancy which is originally developed for the image instead of video. While images have only the spatial dimension, the video contents have both spatial and temporal dimension. The perceptual property in temporal dimension has not been exploited in the utility functions in the game theory framework. As one of our potential researches, we will explore proper criteria for subjective quality of video contents. The new criteria should account for the human visual sensitivity and perceptual property for video contents, in spatial dimension (for instance, size, texture complexity, luminance and position in the scene composition, etc.) as well as temporal dimension (for instance, motions and changes in luminance, size and position, etc.). In the game theoretical models, different utility functions in can be defined for different players, based on their individual subjective quality.

4.2.3 Game Theory Applications on Joint Source-Channel Rate Control

The video communication over wired network presumes that there is zero channel error during the transmission. Most of the existing rate control algorithms are designed based on this channel model. However, this assumption is no longer true when the data is transmitting through the error prone wireless channels. The video data transmission is strongly interfered by the condition of the wireless channel, which the current rate-distortion is not apt for. Efficient mechanism is required to adapt the rate control algorithm to the channel condition. Channel coding is commonly used in error prone environment. Redundancies are added to the data streams before transmission such that channel error can be detected and corrected to some extent. The more bits are devoted to the channel coding, the lower error rate occurs in the receiver end, thus better reconstruction visual quality can be achieved. But due to the constrained data rate, this implies a lower bit rate for the video encoder, which leads to the higher compression ratio and lower visual quality. To optimize the over all quality, proper bandwidth allocation between source coding and channel coding is necessary. Potential research can be done on using the game theory models for the bandwidth allocation among source coding and channel coding. Each video encoder or channel encoder can be regarded as a player, who bids for bandwidth in an auction. The objective is to optimize the overall quality at the receiver end under the constraint of channel condition, by jointly controlling the bit rate consumption of source encoders and channel encoders.

REFERENCES

[1] E. Abdou, "Image Restoration of Multiple Frame Sequences," *Applications of Digital Image Processing XIX*, *Proceedings of SPIE*, 2847, pp. 192-202, 1996.

[2] J. Cheng, and H. Hang, "Adaptive Piecewise Linear Bits Estimation Model for MPEG Based Video Coding," in *Proceeding of 1995 International Conference on Image Processing*, vol. 2, pp. 551-554, 1995.

[3] T. Chiang, H. J. Lee, and Y. Q. Zhang, "Macroblock Layer Rate Control," ISO/IEC JTC/SC29/WG11 MPEG97/M2555, July 1997.

[4] T. Chiang, and Y. Q. Zhang, "A New Rate Control Scheme using Quadratic Rate Distortion Model," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 7, no. 1, pp. 246-250, Feb. 1997.

[5] C. H. Chou, and Y. C. Li, "A Perceptually Tuned Subband Image Codec Based on the Measure of Just-Noticeable-Distortion Profile," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 5, no. 6, pp. 467-476, Dec. 1995.

[6] W. Ding and B. Liu, "Rate Control of MPEG Video Coding and Recording by Rate-Quantization Modeling," *IEEE Transactions on Circuits and Systems for Video Technology,* vol. 6, pp. 12-20, Feb. 1996.

[7] A. F. D'Souza, *Design of Control System*, Englewood Cliffs: Prentice-Hall, 1988.

[8] Gersho and R. Gray, *Vector Quantization and Signal Compression*. Kluwer Academic, 1992.

[9] R. Handel, M.N. Huber, and S. Schroder, *ATM nectworks: Concepts, Protocols, Applications.* Addison-Wesley.

[10] Z. He, Y. K. Kin, and S. K. Mitra, “Low-Delay Rate Control for DCT Video Coding via ρ-domain Source Modeling,” *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 11, no. 8, pp. 928-940, Aug. 2001.

[11] Z. He, and S. K. Mitra, “Optimum Bit Allocation and Accurate Rate Control for Video Coding via ρ-domain Source Modeling,” *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 12, no. 10, pp. 840-849, Oct. 2002.

[12] Z. He, and S. K. Mitra, “A Linear Source Model and a Unified Rate Control Algorithm for DCT Video Coding,” *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 12, no. 11, pp. 970-982, Nov. 2002.

[13] J. Hofbauer and K. Sigmund, *Evolutionary Games and Population Dynamics*, Cambridge University Press, U.K., 1998.

[14] P.I. Hosur and K.K. Ma, “Motion Vector Field Adaptive Fast Motion Estimation,” *Second International Conference on Information, Communications and Signal Processing (ICICS ’99)*, Singapore, 7-10 Dec. 1999

[15] ISO/IEC CD 11172/2 (MPEG-1 Video), Information technology----Coding of Moving Pictures and Associated Audio for Digital Storage Media at Up to About 1.5 Mbits/s: Video, 1993.

[16] ISO/IEC CD 13818/2 (MPEG-2 Video), Information technology----Generic Coding of Moving Pictures and Associated Audio Information: Video, 1995.

[17] ISO/IEC CD 14496/2 (MPEG-4 Video), Information technology----Coding of Natural/Visual Objects, Part 2: Visual, 1999.

[18] ISO/IEC JTC1/SC29/WG11, Test Model 5 (TM5), Document N400, 1993.

[19] ISO/IEC JTC1/SC29/WG11, Coding of Moving Picture and Associated Audio, MPEG97/W1796, ISO/IEC 14496-2, MPEG-4 Video VM Version 8.0, Stockholm, Sweden, Jul., 1997.

[20] ITU-T, Video Codec for Audiovisual Services at p x 64 kbits/s, Recommendation H.261, Geneva, 1990.

[21] ITU-T, Video Codec for Low Bit-Rate Communication, Recommendation H.263, Dec. 1995.

[22] ITU-T SG 15 Experts Group for Very Low Bitrate Visual Telephony, *Video Codec Test Model for the Near-term 5 (TMN5)*, 1995.

[23] ITU-T Rec. H.264 | ISO/IEC 14496-10 AVC, *Advance Video Coding, Draft*, Oct. 2002.

[24] J. R. Jain and A. K. Jain, “Displacement Measurement and Its Application in Interframe Image Coding,” *IEEE Transactions on Communication*, vol. COM-29, pp. 1799-1808, Dec. 1981.

[25] N. Jayant, J. Johnston, R. Safranek, “Signal Compression Based on Models of Human Perception,” *Proceedings of the IEEE*, vol. 81, no. 10, pp. 1385-1422, Oct. 1993.

[26] F. Kelly, “Charging and Rate Control for Elastic Traffic,” *Eur. Trans. Telecomm. -Focus on Elastic Service over ATM Networks*, vol. 8, no. 1, pp. 33-37, 1997.

[27] H. M. Kim, “Adaptive Rate Control using Nonlinear Regression,” *IEEE Transactions on Circuits and System for Video Technology*, vol. 13, no. 5, pp. 432-439, Jan. 2003.

[28] R. Koenen, Ed., “Overview of the MPEG-4 standard,” Doc. ISO/IEC JTC1/SC29/WG11 N2725 Seoul, Korea, Mar.1999.

[29] T. Koga, K. Ilinuma, A. Hirano, Y. Iijima, and T. Ishiguro, “Motion Compensated Interframe Coding for Video Conferencing,” in *Proceeding of 1981 National Telecommunication Conference (NTC81)*, New Orleans, pp. C9.6.1-C9.6.5, Nov. 1981.

[30] V. Krishna. *Auction Theory*, Academic Press, San Diego, U.S.A., 2002.

[31] P. Kuhn, Algorithms, *Complexity Analysis and VLSI Architectures for MPEG-4 Motion Estimation*, Kluwer Academic Publishers, Boston, 1999.

[32] P. Kuhn, W. Stechele, “Complexity Analysis of the Emerging MPEG-4 Standard as a Basis for VLSI Implementation,” in *Proceeding of SPIE Visual Communications and Image Processing*, San Jose, pp. 498-509. Jan. 1998.

[33] H. J. Lee, T. Chiang, and Y. Q. Zhang, "Scalable Rate Control for Very Low Bit Rate (VLBR) Video," in *Proceeding of 1997 International Conference on Image Processing*, vol. 2, pp. 768-771, Oct. 1997.

[34] H. J. Lee, T. Chiang, and Y. Q. Zhang, "Scalable Rate Control for MPEG-4 Video," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 10, no. 6, pp. 878-894, Sept. 2000.

[35] J. W. Lee, A. Vetro, Y. Wang, and Y. S. Ho, "Bit Allocation for MPEG-4 Video Coding With Spatial-Temporal Tradeoffs," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 13, no. 6, pp. 488-502, Jun., 2003.

[36] L. W. Lee, J. F. Wang, J. Y. Lee, J. D. Shie, "Dynamic Search-Window Adjustment and Interlaced Search Block-matching Algorithm," *IEEE Transaction on Circuits and Systems for Video Technology*, vol. 3, no. 1, pp. 85-87, Feb. 1993.

[37] C.E. Lemke and J.T. Howson, "Equilibrium Points of Bimatrix Games," *SIAM Journal of Applied Mathematics* 12, pp. 413-423, 1964

[38] R. Li, B. Zeng, and M. Liou, "A New Three-Step Search Algorithm for Block Motion Estimation," *IEEE Transactions on Circuit and Systems for Video Technology*, vol. 4, no. 4, pp. 438-442, Aug. 1994.

[39] L. Lin, A. Ortega, "Bit-rate Control using Piecewise Approximated Rate-distortion Characteristics," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 8, no. 4, pp. 446-459, Aug. 1998.

[40] L. K. Liu, E. Feig, “A Block-based Gradient Descent Search Algorithm for Block-based Motion Estimation in Video Coding,” *IEEE Transactions on Circuits and Systems on Video Technology*, vol. 6, no. 4, pp. 419-422, Aug. 1996.

[41] Montet and D. Serra, *Game Theory and Economics*, Palgrave, 2003.

[42] J. Morris, M. Satyanarayanan, M. Conner, J. Howard, D. Rosenthal and F. Smith, “Andrew: A Distributed Personal Computing Environment,” *Communications of the ACM*, vol. 29, no. 3, pp. 184-201, Mar. 1986.

[43] J. Nash, "Equilibrium Points in N-person Games," in *Proceeding of the National Academy of Sciences*, vol. 36, pp. 48-49, Jan. 1950.

[44] J. Nash, “The Bargaining problem,” *Econometrica*, vol. 18, pp. 155-162, Apr. 1950.

[45] J. Nash, “Non-Cooperative Games,” *Annals of Mathematics*, vol. 54, pp. 286-295, Sept. 1951.

[46] J. Nash, “Two-Person Cooperative Games,” *Econometrica*, vol. 21, pp. 128-140, Jan., 1953.

[47] J. von Neumann, “Zur Theorie der Gesellschaftsspiele,” *Mathematische Annalen*, vol. 100, pp. 295-320, 1928.

[48] J. Von Neumann, and O. Morganstern, *Theory of Games and Economic Behaviors*. Princeton (Princeton University Press), 1944, 1947.

[49] K. N. Ngan, T. Meier, and Z. Chen, "Improved Single Video Object Rate Control MPEG-4," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 13, no. 5, pp. 385-393, May 2003.

[50] K. Oehler and J. L. Webb, "Macroblock Quantizer Selection for H.263 Video Coding," in *Proceeding of IEEE International Conference on Image Processing*, vol. 1, pp.365–368, Oct. 1997.

[51] M. J. Osborne and A. Rubistein, *A Course in Game Theory*, MIT Press, Massachusetts, U.S.A., 1994.

[52] F. Pan, Z. Li, K. Lim, and G. Feng, "A Study of MPEG-4 Rate Control Scheme and Its Improvements," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 13, no. 5, pp. 440-446, May 2003.

[53] Fernando Pereira, "MPEG-4: a New Challenge for the Representation of Audio-Visual Information," *Keynote speech at Picture Coding Symposium' 96*, 1996

[54] C.L. Phillips and R.D. Harbor, *Basic Feedback Control Systems*, 2nd Edition, Englewood Cliffs: Prentice-Hall, 1991.

[55] P. Pirsch, N. Demassieux, W. Gehrke, "VLSI Architectures for Video Compression- A Survey," *Proceedings of the IEEE*, vol. 83, no. 2, pp. 220-246, Feb. 1995.

[56] L. M. Po and W. C. Ma, "A Novel Four-Step Search Algorithm for Fast Block matching," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 6, no. 3, pp. 313-317, Jun. 1996.

[57] Puri and R. Aravind, "Motion-Compensated Video Coding with Adaptive Perceptual Quantization," *IEEE Transactions on Circuits and Systems for Video Technology,* vol. 1, no. 4, pp. 351-361, Dec. 1991.

[58] K. Ramchandran and M. Vetterli, "Best Wavelet Packet Bases in a Rate-Distortion Sense," *IEEE Transactions on Image Processing*, vol. 2, pp. 160-175, Apr. 1993.

[59] K. Ramchandran, A. Ortega, and M. Vetterli, "Bit Allocation for Dependent Quantization with Applications to Multiresolution and MPEG Video Coders," *IEEE Transactions on Image Processing*, vol. 3, no. 5, pp. 533-545, Sept. 1994.

[60] Eric C. Reed and Frederic Dufux, "Constrained Bit-Rate Control for Very Low Bit-Rate Streaming-Video Applications," *IEEE Transactions on Circuits and Systems for Video Technology,* Vol. 11, No. 7, pp. 882-889, Jul. 2001.

[61] T. Roughgarden, "Stackelberg Scheduling Strategies," in *Proc. of ACM STOC*, pp. 104-113, 1991.

[62] J. Ribas-Corbera and S. Lei, "Rate Control in DCT Video Coding for Low-Delay Communication," *IEEE Transactions on Circuit and System for Video Technology*, vol. 9, no. 1, pp. 172-185, Feb. 1999.

[63] J. I. Ronda, M. Eckert, F. Jaureguizar, and N. Garcia, "Rate Control and Bit Allocation for MPEG-4," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 9, no. 6, pp. 1243-1258, Dec. 1999.

[64] T. Sandholm, “Limitations of Vickery Auction in Computational Multiagent Systems,” in *Proceeding of the 2nd International Conference on Multiagent Systems*, pp. 299-306, 1996.

[65] Y. Shoham and A. Gersho, “Efficient Bit Allocation for an Arbitrary Set of Quantizers,” *IEEE Transactions on Acoust. Speech, Signal Processing*, vol. 36, no. 9, pp. 1445-1453, Sept. 1988.

[66] Stefanescu and M. W. Stefanescu, “The Arbitrated Solution for Multiobjective Convex Programming,” *Rev. Roum. Math. Pure Applicat*, vol. 29, pp. 593-598, 1984.

[67] R. K. Sundaram, *A First Course in Optimization Theory*, Cambridge University Press, 1996.

[68] Tao, H. A. Perterson, and B.W. Dickinson, “A Rate-Quantization Model for MPEG Encoders,” in *Proceeding of 1997 International Conference on Image Processing*, pp. 338-341, Oct. 1997.

[69] B. Tao, B. W. Dickinson, and H. A. Peterson, “Adaptive Model-Driven Bit Allocation for MPEG Video Coding,” *IEEE Transactions on Circuits and Systems for Video Technology,* vol. 10, pp. 147-157, Feb. 2000.

[70] A.M. Tourapis, O.C. Au, and M.L. Liou, "Fast Block-Matching Motion Estimation using Predictive Motion Vector Field Adaptive Search Technique (PMVFAST)", *ISO/IEC JTC1/SC29/WG11 MPEG99/m5866, Noordwijkerhout, the Netherland*, Mar. 2000.

[71] A.M. Tourapis, O.C. Au, and M.L. Liou, "Highly efficient predictive zonal algorithms for fast block-matching motion estimation," *IEEE Transactions on Circuits and Systems for Video Technology*, Volume: 12, Issue: 10, Oct. 2002, pp 934 – 947.

[72] J. C. Tsai, and C. H. Hsieh, “Modified TMN8 Rate Control for Low-Delay Video Communication,” *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 14, no. 6, pp. 864-868, Jun. 2004.

[73] A. Vetro and H. Sun, “Joint Rate Control for Multiple Video Objects Based on Quadratic Rate-Distortion Model,” *ISO/IEC JTC1/SC29/WG11 Coding of Moving Pictures and Associated Audio MPEG 97/M1631*, Sevilla, Spain, Feb. 1997.

[74] A. Vetro, H. Sun, and Y. Wang, “MPEG-4 Rate Control for Multiple Video Objects,” *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 9, no. 1, pp. 186-199, Feb. 1999.

[75] Sun Yu and Ishfaq Ahmad, “A New Rate Control Algorithm for MPEG-4 Video Coding,” in *Visual Communications and Image Processing*, *Proc. SPIE* 4671, pp. 698-709, San Jose, CA. Jan. 2002.

[76] Viscito and C. Gonzales, “A Video Compression Algorithm with Adaptive Bit Allocation and Quantization,” *in SPIE Visual Commun. Image Processing*, vol. 1605-1, pp. 58 –72, Nov. 1991.

[77] S. W. Wu and A. Gersho, “Rate-constrained Optimal Block-Adaptive Coding for Digital Tape Recording of HDTV,” *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 1, no. 1, pp. 100-112, Mar. 1991.

[78] K. H. Yang, A. Jacquin, and N. S. Jayant, “A Normalized Rate-Distortion Model for H.263-compatibal Codecs and Its Application to Quantizer Selection,” in *Proceeding of 1997 International Conference on Image Processing*, pp. 41-44, Oct. 1997.

[79] W. Zheng, I. Ahmad and M.L. Liou, "Adaptive Motion Search with Elastic Diamond for Mpeg-4 Video Coding," in *Proceeding of 2001 International Conference on Image Processing (ICIP2001)*, vol. 1 pp.377-380, 2001

[80] S. Zhu and K. K. Ma, “A New Diamond Search Algorithm for Fast Block-Matching Motion Estimation,” *IEEE Transactions on Image Processing*, vol. 9, no. 2, pp. 287-290, Feb. 2000.

BIOGRAPHICAL INFORMATION

Jiancong Luo received the B.Eng and M.S. degrees from the Department Electrical Engineering at Zhongshan University, China, in 1997 and 2000 respectively. From 2000 to 2001, he was a Ph.D. student at the Computer Science Department of Hong Kong University of Science and Technology. He transferred to the University of Texas at Arlington in spring 2002. He received his doctorate in Computer Science and Engineering in August 2005. His research interests include multimedia systems, video compression, video processing, 3D video, video over wireless device, pervasive audio/video collaboration, video surveillance, game theory application in video compression, 2D/3D image for remote sensing and medical application.