A NOVEL CONSTANT QUALITY RATE CONTROL SCHEME FOR OBJECT-BASED ENCODING

ANG YING HANN

NATIONAL UNIVERSITY OF SINGAPORE

2005

A NOVEL CONSTANT QUALITY
RATE CONTROL SCHEME
FOR OBJECT-BASED ENCODING

ANG YING HANN
(B.ENG)

A THESIS SUBMITTED

FOR THE DEGREE OF MASTER OF ENGINEERING

DEPARTMENT OF ELECTRICAL AND COMPUTER ENGINEERING

NATIONAL UNIVERSITY OF SINGAPORE
2005

ACKNOWLEDGEMENTS

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

CODEC	Coder decoder
CQRC	Constant Quality Rate Control
VO	Video Object
GOP	Group of Picture
HVS	Human Visual System
Q	Quantization Parameter
SRC	Scalable Rate Control
D	Distortion
DCT	Discrete Cosine Transform
MAD	Mean Absolute Difference
PSNR	Peak Signal to Noise Ratio
MB	Macroblock
CDBA	Constant-Distortion Bit Allocation
FIFO	First-In-First-Out
CBR	Constant Bitrate
VBR	Variable Bitrate
QCIF	Quarter Common Intermediate Format
CIF	Common Intermediate Format
VOP	Video Object Plane
SVO	Single Video Object
MVO	Multiple Video Objects
MSE	Mean Squared Error

SUMMARY

This research proposes a novel Constant Quality Rate Control (*CQRC*) algorithm for object-based (MPEG-4) video *CODEC*s. Instead of minimizing distortion of every single frame or minimizing average frame distortion, this controller aims to minimize the *variation* of the frame distortion to achieve *consistent* good quality for *whole* video sequences. The first part of the thesis shows the inadequacy of related conventional rate control schemes in providing constant quality video output while the second part describes the formulation and implementation of the proposed *CQRC* algorithm. The final part compares the performance of this algorithm with MPEG-4 Annex L Scalable Rate Control (*SRC*) algorithm for real-time object-based surveillance applications under numerous real-life constraints.

In most implementations, conventional rate control methods (including the *SRC* algorithm) allocate bit budget to frames based on a measure of buffer fullness so that target bitrate can be obtained and buffer can be stabilized. Due to the non-stationary nature of natural video signals, video output from these rate control schemes will have high visual quality fluctuations. Although there are a few rate control algorithms proposed to minimize quality fluctuations of encoded video, to the knowledge of the author, not much research has been done on rate control schemes based on *object-based CODEC*s.

The proposed *CQRC* scheme is a combination of the state-of-the-art rate control algorithm for MPEG-4's (the *SRC* technique) and the ρ-domain model proposed by He and Mitra**.** The ρ-domain rate control model specifies a relationship between frame bitrate (R), distortion due to quantization (D) and ρ (percentage of zeros among the quantized *DCT* coefficients). Using this *R-D* relationship, one can allocate bitrate based on a target distortion value hence realizing the research objective of maintaining visual quality (distortion) for encoded video sequences. By incorporating the ρ-domain model, the SRC scheme will be able to allocate bit budget based on maintaining a target frame-level distortion. The algorithm is then extended to include multiple video objects by means of target bitrate distribution algorithm based on each individual video object's variance, motion and size.

Experiments to verify the performance of the *CQRC* scheme is conducted using various video sequences and under numerous real-life constraints (including the use of imperfect segmentation masks for encoding). Experimental results and subjective assessments show that in terms of quality fluctuations and temporal quality, the proposed *CQRC* algorithm outperforms the *SRC* algorithm. The *CQRC* algorithm is also able to perform as well as the *SRC* scheme in providing optimal average video quality given buffer and bandwidth constraints.

CHAPTER 1: INTRODUCTION

1.1 Overview

Over the years, rate control for video encoding has been at the centre of interest for researchers, and has been extensively studied under the framework of Rate-Distortion (*R-D*) theory by Claude Shannon **[1]**. Rate control involves modifying the encoding parameters in order to maintain a target output bitrate. At the same time, it aims to minimize distortion in the decoded sequence under real-world constraints (bandwidth, delay and computational complexity). The *R-D* theory gives theoretical bounds on how much compression can be achieved using lossy data compression methods.

More recently, the emergence of the MPEG-4 standard (with its ability to independently code Video Objects (*VO*)) has changed the video encoding paradigm, bringing new challenges to the definition of bitrate control mechanisms. With MPEG-4, a video scene is represented as a composition of arbitrarily shaped video objects characterized by their shapes, motion and texture. Similar to the case of H.263, MPEG-4 rate control must consider both spatial and temporal coding parameters. However, since MPEG-4 also allows the coding of arbitrarily shaped objects, the encoder must consider the significant amount of bits that are used to code the shape information. The independent representation of each *VO* provides coding efficiency and allows the option of prioritizing the subjectively more important objects. This aspect of the encoder makes the rate control problem in MPEG-4 or any other object-oriented encoder unique.

1.2   General Theory on Rate-Distortion Optimization

The *R-D* theory addresses the problem of determining the minimal amount of entropy *R* that should be communicated over a channel such that the coded signal can be reconstructed at the receiver side given distortion *D*. Rate is measured as the total number of bytes transmitted and the notion of distortion is measured in terms of *PSNR* (Peak Signal to Noise Ratio) value of each transmitted frame. However, since we know that most lossy compression techniques operate on data that will be perceived by humans, the distortion measure should include some aspects of human perception (this will be discussed in detail in chapter 3.4). There are many different equations describing the relationship of *R* and *D* and this will be discussed in chapter 2.2.5.

The *R-D* performance of a video *CODEC* describes the tradeoff between video quality and total bitrate. Rate-distortion optimization for rate control is the choosing of a set of encoding parameters that finds the optimal balance between bitrate and quality. It is vital for many applications as variations in the output bitrate may cause problems for various practical delivery and storage mechanisms.

1.3   Major Drawbacks of Existing Rate Control Methods

In general, rate control consists of two main components: *bit allocation* and *quantization parameter determination*. In bit allocation, the target bit budget is estimated (usually based on bits availability and buffer constraints) before being allocated among different coding units such as *GOP*s (Group of Pictures), frames or video objects. For quantization parameter determination, an optimal quantization parameter is estimated to achieve the target allocated bit budget. The research emphasis for this thesis is mainly on improving bit allocation.

A major drawback with conventional bit allocation is that it emphasizes mainly on the stability of the buffer [2-7]. When the buffer is nearly full, the coder will allocate fewer bits to the next frame, resulting in decreased frame quality. When the buffer is nearly empty, more bits will be allocated to the next frame, resulting in increased frame quality. This fluctuation of spatial image quality between adjacent frames will lead to temporal visual degradation in the form of flickering (or blinking) artifacts. It is a known fact that the *HVS* (Human Visual System) is more sensitive to variations in quality compared to the actual frame quality itself [8]. i.e. if a spatially imprecise video sequence (a sequence with high spatial error) is stable over time, it is perceived to be less annoying than a more (spatially) precise video sequence presenting abrupt changes over time [9]. This temporal artifact, caused by a variation of spatial error between frames, has been observed to be one of the most annoying artifacts and has a great impact on the visual quality of video sequences. Many algorithms [3-8] adopt such an approach or similar, as bit allocation based on buffer state is easier to implement than that based on visual quality. Consequently, these rate control schemes will produce

3

video sequences with significant quality fluctuations and ultimately poor subjective video quality.

More recently, bit allocation based on maintaining constant video quality has been investigated and has given some encouraging results [10,11]. The focus of these rate control schemes are very different from that of others in that it does not emphasize on minimizing distortion of every single frame or minimizing average frame distortion. Instead, it seeks to minimize the *variation* of the frame distortion while maintaining buffer stability. We propose here a novel Constant Quality Rate Control (*CQRC*) algorithm and apply it to object-based video encoding within the MPEG-4 framework.

1.4   Organization of Thesis

The organization of this thesis is as follows: Chapter 2 gives a short review on related rate control schemes and their inadequacy in providing constant quality video output. Chapter 3 presents the proposed algorithm that is capable of reducing quality variations for real-time object-based surveillance camera applications. The implementation of the *CQRC* scheme is then presented in Chapter 4 while results and analysis for three different sets of experiments are presented in Chapter 5. Finally, in Chapter 6, conclusion for this research project is drawn and recommendations for further works are discussed.

CHAPTER 2:   LITERATURE REVIEW OF
CONVENTIONAL RATE CONTROL METHODS

2.1   Introduction

Rate control techniques have been studied very intensively for various video-encoding standards such as H.261, H.263, MPEG-1, MPEG-2 and the recent video object coding with MPEG-4. For different *CODEC*s, different coding parameters are employed and different constraints are imposed. The most influential coding parameter with regards to picture quality is the quantization parameter (Q) used for texture coding. In this chapter, several different conventional rate control algorithms are studied, placing emphasis on the bit allocation methods employed. Their performance, advantages and drawbacks are then reviewed and presented.

2.2   Review of Conventional Bit Allocation Schemes

2.2.1 Global Optimal Solution for Bit Allocation

The globally optimal solution for universally constant quality video entails a multi-pass optimization process. By searching *all* possible quantization values for *each* video frame, an optimal solution is guaranteed. This is however impractical as it is both time-consuming and expensive, especially with regards to real time video coding

applications which have a very strict delay constraint and requires a single-pass approach.

2.2.2 Bit Allocation based on Buffer Occupancy

In most implementations, bit allocation is determined based on a measure of buffer fullness so that target bitrate can be obtained and buffer can be stabilized **[9]**. The frame-level rate-control algorithm of MPEG-4 Q2 **[2,3]**, Scalable Rate Control (*SRC*) scheme, is one such algorithm. Its overall purpose is to maintain buffer occupancy of 50% after encoding each frame. Using this approach (i.e. based on rate policy), the bitrate is directly used for rate control. Due to the non-stationary nature of natural video signals, video output from these rate control schemes will have high visual quality fluctuations. Although the buffer can be successfully stabilized, it is difficult to achieve a near constant video quality.

2.2.3 Constant Bit Allocation Scheme

The constant bit allocation scheme is another popular bit allocation scheme used by many standard reference video *CODEC* software such as TM5 for MPEG2 **[12]**, MPEG4 Annex L **[5]** and TMN8 for H.263+ **[13]**. Each *GOP*, slice or frame is allocated the same amount of bits equal to the target bitrate divided by the target frame rate. These works are however based on the assumption that video content is stationary across scenes and frames. This assumption often does not hold in *natural* video sequences and the bitrate must be varied for uniform output video quality. Hence, the

constant bit allocation method needs to be improved or dropped to cater to non-stationary video signals.

2.2.4 *GOP*-based Bit Allocation

One commonly used benchmark for bit allocation is based on MPEG-2 TM5 **[12]**. It uses a *GOP*-based bit allocation mechanism in which size of the *GOP* is fixed and a constant bit budget is allocated for each *GOP*. Using *GOP*-based bit allocation, it is highly unlikely to achieve constant quality for a *whole* video sequence because of the inconsistent video quality between GOPs. To solve this problem, several schemes **[6,10]** adopt the use of a single large *GOP* for each video sequence (i.e. only the first frame is an I-frame whereas the rest of the video sequences are P-frames and/or B-frames). This effectively solves the problem of inconsistent video quality among GOPs, as there is only one *GOP* used for the whole video sequence. However, the absence of I-frames will lead to ineffective video coding, which comes in the form of error propagation. One of the roles of a purely Intra-coded frame is that it serves as an anchor, which will help the sequence recover in case of transmission errors. A transmission error (bit error or packet loss) will cause the decoder to decode incorrectly some of the information. Since each P-frame is used as the reference frames for the following P-frames, quality degradation will propagate to subsequent P-frames.

2.2.5 Bit Allocation based on Probabilistic Models

To solve the problem of variations in output video quality, the use of probabilistic models for bit allocation in coding rate prediction was proposed **[11,14],** where the desired amounts of bits are allocated to each frame based on its complexity value. A probabilistic constant quality model is a specified relationship between the number of bits of a frame and the magnitude of its complexity, when the entire sequence is coded with consistent quality. The *R-D* relationship of the sequence is approximated with some close-form mathematical functions such as logarithmic, polynomial (linear and quadratic), spline and other more complicated models. This estimation of the *R-D* curve is then simplified to a parameter estimation problem using characteristics of the current frame and previously encoded information. If the model has a good match with actual results, the right amount of bits will be allocated to each frame according to its complexity value, and the whole sequence will have small quality variations. Care must be taken to choose a good model as an inaccurate model will cause inaccurate bit budget allocation, hence causing the actual distortion to be quite different from the predicted distortion. An example of an inaccurate model is the Distortion-Quantization (*D-Q*) model adopted by TMN8 ($D = \dfrac{Q^2}{12}$) **[13,15]**. It is based on the assumption that the Discrete Cosine Transform (*DCT*) coefficients of a video frame are uniformly distributed over each interval of length Q. This assumption is however not true as *DCT* coefficients are more likely to be Laplacian distributed **[16]**, especially for natural video sequences.

Since previous data is used to estimate the model parameters, all model-based rate-control methods suffer from model inaccuracy, as estimated model parameters may not truly reflect the characteristics of the current frame. This is especially true for non-stationary video content. In spite of its non-optimality, for real-time single-pass rate control tasks, a constant quality model based bit allocation scheme is inherently more capable of obtaining consistent quality throughout the whole video sequence. Nevertheless, to the best of the author's knowledge, little effort has been spent on the development of this type of method.

2.2.6 "Distortion Policy of Buffer-Constrained Rate Control for Real-Time Variable Bitrate" by Jinho Choi

To cater to non-stationary signals and to minimize variation of video quality, several constant-quality rate control schemes were proposed. Choi [17] proposed a distortion policy based both on the buffer state and on frame distortion. This policy adopts a multi-objective cost function that balances buffer stability and maintaining constant distortion. The weighting factor for these 2 objectives is adaptive; when the state of the buffer is close to the desired state (50%), greater emphasis will be made on reducing quality fluctuations. When the buffer occupancy approaches zero (danger of underflow), more emphasis will be made to bring the buffer occupancy back to 50%. A logarithmic model is used to estimate the R-D relationship and experimental results shows an improvement in distortion variation. The improvement in quality variation is however insignificant as this bit allocation is still based on the objective of maintaining a buffer occupancy of 50%. Moreover, this scheme is based on a JPEG-like image coder, which is seldom used in the video-encoding industry. More efficient video

coding schemes (like MPEG-4 object-based coding) are now available and research on rate control techniques should be carried out with respect to these modern schemes.

2.2.7 "A Sequence-based Rate Control Framework for Constant Quality Video" by B. Xie and W. Zeng

Xie and Zeng proposed a sequence-based bit allocation method **[14]** that is capable of tracking the non-stationary characteristics in a video sequence. The scheme is based on their observation that the *R-MAD* (Bitrate-Mean Absolute Difference) model under constant Q (which is a fairly good solution for constant quality) is quadratic in nature. *MAD* is chosen as the video content complexity measure to calculate bit budget for each frame. It employs a multi-pass quantization approach, which guarantees that the actual bit count will be as close as possible to the target bit count. Firstly the video sequence will undergo an initial Q determination where the bit budget is allocated based on the square-rooted ratio between current *MAD* and the average *MAD* of previously encoded frames.

$$R_{TARGET} = \frac{C}{F} \times \sqrt{\frac{MAD_K}{MAD_{K-1}}} \qquad (2.1)$$

where R_{TARGET} is the target bitrate calculated by the bit allocation process

   C is the target (channel) bitrate.

   F is the target frame rate.

The model is then appraised for its accuracy and if the model mismatch error is larger than a certain threshold, the frame is re-encoded using a new Q. A Q-readjustment module is used to calculate a new Q based on the mismatch error.

$$Q_{new} = Q_{prev} \times \sqrt{\frac{R_{prev\_actual}}{R_{t\arg et}}} \qquad (2.2)$$

where R_{prev_actual} is the actual bitrate of the frame using the previous Q.

This process is repeated until all selected Qs are optimal, and the whole video sequence is encoded. Their proposed algorithm achieves a slight improvement in overall *PSNR* gain (around 0.5 dB) and delivers a more consistent (subjectively) video quality across the whole sequence. However, this multi-pass quantization determination is not acceptable for hard real-time video transmission because of the additional delay induced. About 25% of the frames need extra re-quantization leading to a higher cost both in time and in complexity. The objective of this scheme is not explicitly defined and the effectiveness of their proposed algorithm in containing quality variations is not presented in a concise manner.

2.2.8 "Operational Distortion-Quantization Curve Based Bit Allocation for Smooth Video Quality" by Junqiang Lan, Wenjun Zeng, Xinhua Zhuang

A single-pass frame-level Constant-Distortion Bit Allocation (*CDBA*) scheme was proposed recently by Lan, Zhuang and Zeng **[11]** for smooth video quality throughout

a video sequence. It is based on the ρ-domain rate control model (also known as the linear rate control model) proposed by He and Mitra [18,19] which states that the estimated bit count for texture (R) is linearly related to ρ (percentage of *DCT* zeros in the frame after quantization), and distortion (D) due to quantization is exponentially related to ρ. By modeling this R-ρ relation and D-ρ relation, the R-D relationship can be approximated and bit allocation based on a target distortion value can be achieved. TMN8 [13], a macroblock-based quantization determination algorithm, is then used to calculate the desired Q for each macroblock (MB) based on the estimated target bit budget for each frame. Experimental results show that the *CDBA* scheme provides much smoother video quality on all test sequences compared to Xie and Zeng's *MAD*-based bit allocation scheme [14] and the Constant Bit Allocation Scheme [12]. However the framework for this research is based on H263 *CODEC*s, and has not yet been applied to object-based coding (MPEG-4).

2.3   Conclusion

The main motivation of this research is to implement a rate control scheme that is capable of producing a video output sequence with consistent quality, specifically for object-based video encoders (MPEG-4). The scheme must be able to perform under numerous real-life constraints such as real-time encoding constraints, channel buffer of practical size and uses of imperfect segmentation masks. To the knowledge of the author, little research has been done on rate control schemes based on maintaining consistent video quality, especially for object-based *CODEC*s.

CHAPTER 3: FORMULATION OF THE CONSTANT

DISTORTION RATE CONTROL ALGORITHM

3.1 Introduction

To minimize variation of frame distortion (i.e. keeping video quality constant) in an object-based encoder while maintaining buffer stability, a novel rate control scheme that seeks was proposed. The proposed *CQRC* scheme combines the state of the art rate control algorithm for MPEG-4's (the Scalable Rate Control technique) with the ρ-domain model proposed by He and Mitra **[18,19]**. A frame level bit allocation scheme, based on the ρ-domain rate control model, is used to estimate target bitrate for each arbitrary-shaped *VO* with the purpose of smooth video quality throughout the whole video sequence. This target bitrate will then be used to calculate the quantization parameter for the current *VOP*, based on the quadratic formulation of the Rate-Quantization model adopted by MPEG-4. The scheme is then further extended to encompass multiple (arbitrary shaped) video objects by means of a bitrate distribution algorithm. The proposed single-pass rate control scheme aims to achieve *consistent* good visual quality for *whole* video sequences (under real-world constraints) for an object-based video coding scheme (notably for the MPEG-4 scheme).

This chapter will first give a detailed review on the *SRC* algorithm, followed by an introduction of some of the preliminary concepts of the proposed *CQRC* scheme. Fundamental issues that need to be addressed in order to adapt the ρ-domain rate

control model to the existing *SRC* algorithm are identified and the extension of the proposed scheme to handle multiple *VO*s is presented. Problems associated with rate control under various real-world constraints (buffer constraints, use of imperfect segmentation masks for encoding) will also be addressed.

3.2   The Scalable Rate Control Scheme

3.2.1 Introduction

The objective of this research is to incorporate the ρ-domain rate control model to object-based video coding to achieve the purpose of smooth video quality. Before one can incorporate the ρ-domain model into the *SRC* scheme, an in-depth understanding of the various mechanisms of the *SRC* scheme is required. The VM 8 *SRC* scheme is designed to meet both *VBR* (Variable Bitrate) without delay constraints and *CBR* (Constant Bitrate) with low-latency and buffer constraints. It is scalable for various bitrates (e.g. 10 kbps to 1 Mbps), spatial resolutions (e.g., Quarter Common Intermediate Format (*QCIF*) to Common Intermediate Format (*CIF*)), temporal resolutions (e.g., 7.5 fps to 30 fps), coder (e.g., *DCT* and wavelet), and granularities of video objects (e.g., single *VO* to multiple-*VO*s, frame-layer to macroblock-layer) **[3]**.

The SRC scheme assumes a quadratic formulation of the Rate-Quantization (*R-Q*) model **[20]**, which can adapt to changes in picture activity. This model was adopted by MPEG in November 1996 for single video object simulations forming the benchmark for rate control of object-based video. The *SRC* scheme was designed for P-type Video Object Planes (*VOP*s). The choice of the quantization parameter is dependent on the

available channel bandwidth, output buffer fullness, and picture & motion complexity. There are typically 5 major steps in the *SRC* scheme. *Fig. 1* below illustrates the flow of the algorithm.

```
                    ┌─────────────────────┐
                    │  1) Initialization  │◄──────────────────┐
                    └─────────────────────┘                   │
                              │                                │
                              ▼                                │
                         ╱─────────╲                           │
                        ╱ First PVOP ╲     Y    ┌────────────────────┐
                        ╲  of GOP?   ╱─────────►│   Q = initial Q    │
                         ╲─────────╱            └────────────────────┘
                              │ N                         │
                              ▼                           │
         ┌───────────►┌──────────────────────────┐        │
         │            │  2) Target bit estimation │        │
         │            └──────────────────────────┘        │
         │                      │                          │
         │                      ▼                          │
         │    ┌──────────────────────────────────────┐     │
         │    │ 3) Quantization parameter calculation │     │
         │    └──────────────────────────────────────┘     │
         │                      │                          │
         │                      ▼                          │
         │            ┌──────────────────┐                 │
         │            │     MPEG-4       │◄────────────────┘
         │            │     Encoding     │
         │            └──────────────────┘
         │                      │
         │                      ▼
         │                 ╱─────────╲
         │                ╱ End of GOP? ╲      Y
         │                ╲           ╱──────────────────────┐
         │                 ╲─────────╱                       │
         │                      │ N                          │
         │                      ▼                            │
         │         ┌────────────────────────────┐            │
         │         │ 4) Update Model Parameters │            │
         │         └────────────────────────────┘            │
         │                      │                            │
         │                      ▼                            │
         │         ┌────────────────────────────┐            │
         └─────────│ 5) Post Frame-skipping Control │        │
                   └────────────────────────────┘
```
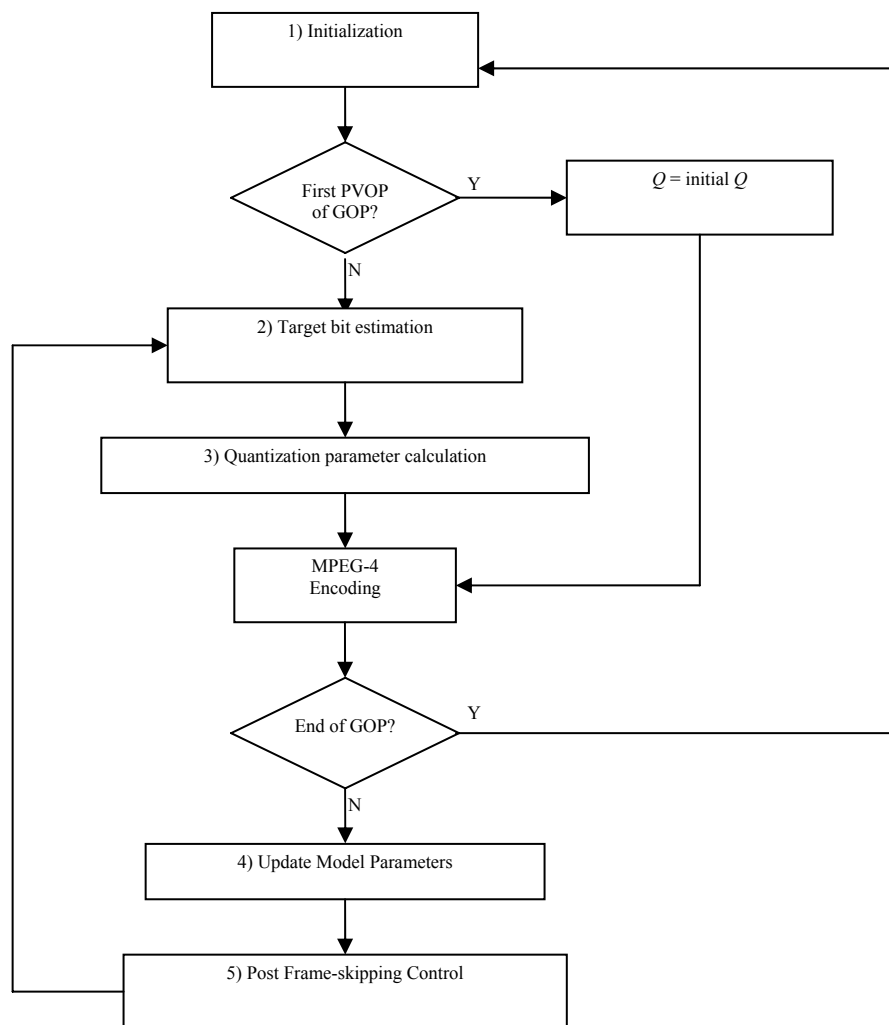
Fig. 1: Flowchart of MPEG-4 Q2 rate control: SRC scheme

3.2.2 Initialization

During this stage, buffer-related quantities are defined and encoding parameters are initialized for use in the algorithm. The two parameters of the quadratic model (X_1 and X_2), initial quantization values for *IVOP*s and *PVOP*s, buffer size and initial buffer occupancy are also initialized in this stage. A user-defined initial Q will be used to encode the 1st *PVOP* of each *GOP*. Rate control will only commence after the 1st *PVOP* of every *GOP* is encoded. The *SRC* assumes that the buffer occupancy is 50% of total buffer size after *every IVOP* is encoded. It is noted that the buffer used by the *SRC* scheme is a virtual buffer that will be reset to 50% of total buffer size after each new *GOP*.

3.2.3 Computation of the Target Bitrate before Encoding

Bit allocation is done in this stage, where the target bit budget is allocated among the different *VO*s to address the rate control problem. *SRC* employs the constant bit allocation scheme at a *GOP*-level, where a pre-determined amount of bits (R_s, based on channel bandwidth and buffer size) is allocated for each *GOP*. At the frame-level, the target bitrate (T) for each *PVOP* is determined in the following 3 stages:

3.2.3.1   Initial Bit Estimation

The estimated initial target bitrate ($T_{initial}$) for the current *VOP* is determined from the bitrate used by the previous *VOP* (R_{prev}) and remaining available bitrate (R_r, allocated

bitrate of each *GOP* less the bits used by the *IVOP* and other previous *PVOP*s; i.e.

$R_r = R_s - R_{IVOP} - R_{previous\_PVOPs}$). A lower bound of $\dfrac{R_S}{30}$ is imposed to ensure a

minimum quality for every *PVOP*.

$$T_{initial} = \max(\frac{R_S}{30}, 0.95 \times \frac{R_r}{N_r} + 0.05 \times R_{prev}) \qquad (3.1)$$

where N_r is the remaining number of *PVOP*s in the *GOP*

3.2.3.2   Joint Buffer Control

After $T_{initial}$ has been determined, it will be scaled based on the current buffer occupancy (B_c) and the buffer size (B_{max}).

$$T = T_{initial} \times \frac{B_c + 2 \times (B_{max} - B_c)}{2B_c + (B_{max} - B_c)} \qquad (3.2)$$

This scaling aims to maintain current buffer occupancy at the target buffer level of 50%. It is based on the difference between the target buffer level and actual buffer fullness. If B_c is more than half of B_{max}, the estimated target bitrate (T) will be reduced. If B_c is less than half of B_{max}, the estimated target bitrate will be increased. If B_c is at 50% of B_{max}, estimated target bitrate will not be amended as the *SRC* scheme's objective of maintaining 50% buffer occupancy is achieved.

3.2.3.3   Buffer Management Strategy

The estimated target bitrate after joint buffer control is further shaped by buffer status, to prevent buffer overflow and underflow, to yield the final estimated target bitrate (T_{final}) for the current frame. By imposing a lower bound of 10% of B_{max} and a higher bound of 90% of B_{max}, the buffer can be successfully stabilized by allocating bits up to this threshold value. i.e. if the buffer is in danger of overflowing (sum of estimated bit budget and current buffer occupancy will result in buffer being more than 90% full), bit budget allocated for the current *VOP* will be reduced to the amount of bits remaining in the buffer just before it is 90% full. The lower bound of $\dfrac{R_S}{30}$ also applies here. A similar policy is used if the buffer is in danger of underflow.

$$
\begin{aligned}
T_{final} \quad &= max \left[\frac{R_S}{30}, (TH_{over} \times B_{max}) - B_c \right] && if \ (B_c + T) \qquad > TH_{over} \times B_{max} \\
&= (TH_{under} \times B_{max}) - B_c + R_p && if \ (B_c + T - R_p) < TH_{under} \times B_{max} \\
&= T && otherwise \qquad (3.3)
\end{aligned}
$$

where the average number of bits to be removed from the buffer (R_p) is:

$$
R_p = (R_s - R_{IVOP})/ N \qquad\qquad (3.4)
$$

N is the number of *PVOP*s in the *GOP*

TH_{over} and TH_{under} are the thresholds for overflow and underflow respectively (set to 90% and 10% respectively)

3.2.4 Quantization Parameter Calculation

After the target allocated bit budget (T_{final}) is estimated for the current *PVOP*, *Q* is computed by using the quadratic *R-Q* model proposed by Chiang and Zhang **[20]**. Since quantization affects only texture, bits for motion, shape and header are subtracted from the target allocated bit budget (i.e. $T_{texture} = T_{final} - T_{header}$, where $T_{texture}$ is the target bitrate allocated for texture and T_{header} is the amount of header bits used for the current *PVOP*. This is approximated to be the same as the header bits used for the previous *PVOP*). The quadratic Rate-Quantization model is as follows:

$$T_{texture} = \frac{X_1 \times MAD_c}{Q_c} + \frac{X_2 \times MAD_c}{Q_c^2} \qquad (3.5)$$

where Q_c is the value of the current quantization parameter,

MAD_c is the mean absolute difference of the current frame after motion compensation (a measure of frame complexity),

X_1 and X_2 are the coefficients of Taylor's expansion of $T_{texture}$ over Q_c.

To ensure frame quality will not vary too much, Q_c is limited to vary between 1 and 31, and is only allowed to change by 25% from the previous quantization parameter (Q_p).

$$Q_c = Min \ \left[\left\lceil 1.25 \times Q_p \right\rceil, Q_c, 31\right] \qquad (3.6)$$

$$Q_c = Max \ \left[\left\lceil 0.75 \times Q_p \right\rceil, Q_c, 1\right] \qquad (3.7)$$

3.2.5 Updating the Model Parameters

The parameters for the Rate-Quantizer model are continuously updated by the encoding results of the current *VOP* as well as a specified number of past *VOP*s. This update is conducted in 3 stages:

3.2.5.1   Sliding window data-point selection

The purpose of this stage is to select relevant data points that will be used to estimate the values of X_1 and X_2. These data points are chosen using a window whose size is dependent on the change in complexity. This sliding window mechanism is used to adaptively smooth the impact of a scene change in updating the *R-Q* model. When the complexity (measured by *MAD*) changes significantly (i.e. high motion scenes), a smaller window with more recent data points will be used. If frame complexity is about the same, a maximum number of 20 data points will be used to calculate X_1 and X_2. The number of past data points (*w*) selected is based on the following:

$$w \quad = \left\lceil [\frac{MAD_c}{MAD_p} \times 20] \right\rceil \quad \quad \textit{if } MAD_p > MAD_c$$

$$= \left\lceil [\frac{MAD_p}{MAD_c} \times 20] \right\rceil \quad \quad \textit{otherwise} \quad \quad \quad (3.8)$$

where MAD_c and MAD_p is *MAD* of the current and previous frame respectively.

3.2.5.2   Statistical removal of data outliers

An outlier is a data point that is located far from the majority of the other data points. To improve the accuracy of the model, data outliers are rejected from the calculation process of the model's parameters. The rejection criterion for this calibration is when the model mismatch error (δ, difference between the predicted number of bits and the actual number of bits) is more than one standard deviation (σ) among the w frames.

$$\sigma = \sqrt{\frac{1}{w}\sum_{i=0}^{w-1}[X_1 \cdot \frac{MAD_c}{Q(i)} + X_2 \cdot \frac{MAD_c}{Q(i)} - R(i) \times MAD_c]^2} \qquad (3.9)$$

$$\delta(i) = X_1 \cdot \frac{MAD_c}{Q(i)} + X_2 \cdot \frac{MAD_c}{Q(i)} - R(i) \times MAD_c \qquad (3.10)$$

where $R(i)$ is the actual number of bits used by the ith frame.

Using least-square estimation, an initial model-parameter estimation is conducted by utilizing the w past data points calculated in the previous stage. Using these initial values of X_1 and X_2, δ of the past w data points and σ can be calculated. The data outliers can then be detected and these will be excluded in the final calculation of X_1 & X_2. Note that only texture bits are considered in this calculation.

3.2.5.3   Updating model parameters

The model parameters (X_1 and X_2) are re-calculated using the least-square approximation method based on the w data points after outlier removal. These values will be used to calculate Q for the next *PVOP*.

3.2.6 Post Frame-Skipping Control

After encoding each *VOP*, the buffer status is updated.

$$B_C = B_p + R_C - R_p \qquad (3.11)$$

where B_p is the previous channel buffer occupancy

R_C is the total bits used for the current frame.

Once B_C reaches 80% of B_{max} (the threshold value chosen for this research), the encoder will skip the next frame. This frame-skipping mechanism will repeat itself until B_C is less than 80% of B_{max}, thus allowing buffer occupancy to reduce to safer levels.

3.2.7 Reset Model at the End of Each *GOP*

At the end of each *GOP*, the *SRC* model will reset itself, clearing all the past data point entries and setting all parameters to their initial value (i.e. resetting the buffer occupancy back to 50%, resets the *Q* for *PVOP* and the model parameters). The algorithm will then repeat itself for the next *GOP* until the whole video sequence is encoded.

3.2.8 Drawbacks of the *SRC* Scheme

The overall objective of the *SRC* scheme is to enforce buffer stability while at the same time optimizing the available bandwidth in order to obtain optimal video quality. However bit allocation based on buffer occupancy will inevitably result in quality variations leading to poor subjective quality. i.e. poorer frame quality at the beginning of the *GOP* (where bits availability is limited as bit budget is divided equally to all *VOP*s in the *GOP*) and better frame quality at the end of the *GOP* (when all remaining bits are allocated to the last *PVOP*). This will be demonstrated in the experimental results shown later. In this research, the ρ-domain rate control model will be implemented into the existing *SRC* scheme to solve this problem of quality fluctuation by allocating frame-level bit budget based on a target distortion measure. The performance of the original *SRC* scheme will be used as a comparison for the performance of the *CQRC* scheme.

3.3 ρ-domain Rate Control Model

To solve the problem of variations in frame quality, the use of probabilistic models in coding rate prediction was proposed. The ρ-domain rate control model was recently proposed by He and Mitra [18,19], in which ρ is the percentage of zeros among the quantized *DCT* coefficients in the current frame. They observed that the bitrate (*R*) has a linear relationship with ρ and the distortion *D* has an exponential relationship with ρ.

$$R = \mu \times N \cdot (1 - \rho) \qquad\qquad (3.12)$$

$$D = \sigma^2 \cdot e^{-\alpha(1-\rho)} \qquad\qquad (3.13)$$

where μ, σ^2 and α are the model parameters

   N is the number of pixels in the frame

If we denote the number of non-zero *DCT* coefficients (after both INTRA and INTER quantization) to be $N_{NZ}(Q)$, the above equations become:

$$R(Q) = \theta \cdot (N_{NZ}(Q)) \qquad\qquad (3.14)$$

$$D(Q) = \sigma^2 \cdot e^{-\beta(N_{NZ}(Q))} \qquad\qquad (3.15)$$

After encoding the first frame, values for $N_{NZ}(Q)$ and *R(Q)* can be calculated and θ can be approximated using *equation (3.14)*. The rest of the model parameters (β and σ^2) can then be calculated by using two pairs of N_{NZ} and *D* values using *equation (3.15)*. In order to minimize model inaccuracy, the 2nd {N_{NZ}, *D*} pair chosen for this research is the one that is nearest to the original {$N_{NZ}(Q)$, *D(Q)*}, namely {$N_{NZ}(Q+1)$, *D(Q+1)*}.

This approximation may be inaccurate when Q is relatively small, where a small change in Q will bring about a relatively large change in D due to the non-linearity of the R-D curve for low Q.

Using these 2 pairs of $\{N_{NZ}, D\}$ as the model references points, β and σ^2 can be calculated. Desired N_{NZ} for the subsequent frame can then be estimated from *equation (3.15)* when given a target distortion. Using *equation (3.14)*, the target bitrate R can be determined. R is then further shaped by buffer constraints and this final R will be used to determine the final Q of the current frame using any R-Q model (e.g. the quadratic model adopted by *SRC* for quantization parameter calculation). To improve the robustness of the model, only the 1st set of model parameters is calculated using $\{N_{NZ}(Q), D(Q)\}$ and $\{N_{NZ} Q+1), D(Q+1)\}$, since this calculation only takes into account the characteristics of the current frame. For the rest of the sequence, the $\{N_{NZ}(Q), D(Q)\}$ of the previous *VOP* and the current *VOP* will be used to update the model parameters. Experiments have proven that this modification improves the rate control scheme in terms of quality variation and number of frames dropped.

Due to the non-stationary nature of natural video sequences, model inaccuracy will still exist in the ρ-domain rate control model. An inaccurate model will cause inaccurate bit budget allocation and will lead to sub-optimal performance. Hence an iterative algorithm is needed to make the actual distortion converge to the desired distortion. The ρ-domain R-D optimization has been successfully applied to the optimal bit allocation within a video frame [18], and within different groups of macroblocks within a frame [11]. It has however, never been applied in an MPEG-4 (object-based) framework before. This is the main motivation for this research: incorporate the ρ-

domain rate control model to object-based *CODEC*s with the purpose of reducing video quality variations and ultimately improve overall visual quality.

3.4 Selection of Distortion Index

A critical point in deriving a good constant quality model is to select a good complexity measure, which is generally a quantity to indicate the difficulty of rendering a certain frame with the same (or similar) quality as the others. For this research, the distortion index *D(Q)* represents the average *DCT* reconstruction error of all luminance pixels of the current *VO* due to quantization.

$$D(Q) = \frac{1}{N} \sum_{u=0}^{N-1} [DCT_{ORIG}(u) - DCT_{RECON}(u)]^2 \qquad (3.16)$$

where N is the number of shape (object) pixels in the current *VOP*,

$DCT_{ORIG}(u)$ is the value of the *DCT* coefficient of pixel u before quantization,

$DCT_{RECON}(u)$ is the reconstructed *DCT* value of pixel u defined by MPEG-4 quantization scheme.

The target distortion index for the subsequent frames is defined as the average distortion index of all previous encoded frames.

$$D_{TARGET,k} = \frac{1}{k-1} \sum_{i=1}^{k-1} D_i \qquad (3.17)$$

where k is the current *VOP* number,

$D_{TARGET,k}$ is the target distortion of the kth frame,

D_i is the actual distortion of the ith frame.

$D_{TARGET,k}$ will undergo large changes in value at the beginning of the sequence when k is relatively small. However as k gets larger and larger, $D_{TARGET,k}$ will converge to a certain value and this value is taken to be the target distortion value for all subsequent *VOP*s. The *D-Q* model adopted for this research will adjust Q in order to obtain a D_k as close to $D_{TARGET,k}$ as possible.

3.5   Buffer Management Strategy

To achieve the best-perceived video quality in real-world scenarios, bandwidth constraints should be considered. The bandwidth constraints on bitrate control are realized by introducing circular *FIFO* (First-In-First-Out) buffers with bits being written in and read out concurrently. The buffer is used at the output side of the *VBR* encoder to absorb variation of the bitrate before sending it to the client via *CBR* links (see *Fig. 2*). In real-time video communications, the buffer usually has very limited size due to the strict end-to-end delay requirement.
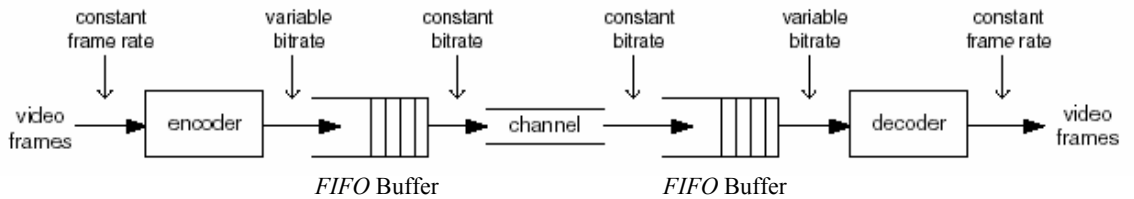

Fig. 2: Encoder output and decoder input buffers [13]

Whenever the encoder generates more bits than the buffer can hold, a buffer overflow will occur. The encoder will either have to re-encode the current frame with a coarser quantizer, or simply skip the frame to avoid the overflow. Similarly, whenever there are no bits available in the encoder buffer, a buffer underflow will occur, wasting some of the available bandwidth. Both encoder underflow and overflow will result in a similar state at the decoder buffer. Hence in order to meet the channel bandwidth constraints, the bitrate control at the encoder side needs to ensure that there is neither overflow nor underflow at any time.

Overflow or underflow of the buffer cannot be prevented without rate control. A good buffer-constrained rate control algorithm should not only prevent buffer overflow/underflow, but also to minimize the distortion. Hence to avoid buffer underflow and overflow, the allocated bit budget calculated from the ρ-domain model needs to be further shaped under buffer constraints.

The buffer management strategy used in the proposed *CQRC* scheme is similar to that of the *SRC* algorithm. Whenever the sum of the current buffer occupancy and the allocated bit budget is more than 90% of buffer size, bit budget allocated for current *VOP* will be the amount of bits that will fill the buffer to exactly 90% of buffer size. This will ensure that B_c will not exceed 90% of buffer size, hence preventing the occurrence of buffer overflow. To prevent buffer underflow, a similar approach is adopted to ensure B_c will not be lower than 10% of buffer size. Details for the buffer management strategy will be presented in Chapter 4 under implementation of the *CQRC* algorithm.

3.6　Video Segmentation for Object-based Video Encoding

Sophisticated forms of interactivity have been developed for direct interaction with audiovisual content, resulting in an evolution towards more semantically meaningful representations, such as those allowed by the ISO MPEG-4 object-based coding standard. Traditional video standards such as MPEG-1, MPEG-2, H.261, or H.263 are considered low-level techniques as they represent video as arrays of pixels. Although they can achieve high compression ratios, these representations are inappropriate if one wants to be able to interact with objects in the image. New video coding schemes are necessary due to this increase in popularity of multimedia applications and content-based interactivity. The performance of algorithms for video processing, compression and indexing often depends on prior efficient segmentation.

Object segmentation in computer vision consists of the extraction of the shape of physical objects projected onto the image plane, ignoring edges due to texture inside the object borders. While output from a general segmentation algorithm (based only on intensity similarity) can be a large number of irregular segments, object segmentation aims to recognize the shapes of complete physical objects present in the scene. However, it is generally known that the state of the art is still inadequate for robust object segmentation algorithms, which are able to deal with generic images and video sequences. Occlusion with the background, lighting and colour changes, shadows, non-static backgrounds, camera motion and other problems will lead to erroneous segmentation results with inaccurate boundaries and artifacts.

Most research on object-based rate control **[2-6,21-24]** uses perfect segmentation files for testing purposes. These works fail to consider the fact that noises and artifacts present in a segmented frame/*VOP* can severely affect the quality of a segmented video both spatially and temporally. Hence their algorithms are unable to perform optimally when confronted with this real-world phenomenon. To verify the performance of the *CQRC* algorithm in a practical real-time situation, imperfect segmentation masks are used in some of the experiments. A real-time object segmentation algorithm is used to generate the segmentation masks of each *VO* **[25]**. These are then fed into the MPEG-4 encoder along with its video frames. Because of strict computational and time constraints, output segmentation masks from the real-time object segmentation algorithm will be imperfect and artifacts (both spatial and temporal) will be present.

3.7   Extension of *CQRC* Algorithm to Multiple Video Objects

The *CQRC* mechanism for multiple video objects (*MVO*) is a non-trivial extension of the single video object (*SVO*) algorithm. The main addition to the *SVO* scheme is the target bitrate distribution function, which enables individualized *R-D* control over separate *VO*s. Generally speaking, an object-based coder attempts to code each object with a different quantization parameter to exploit the fact that each object need not be coded with the same parameters to achieve comparable quality. For example, a stationary background object coded with a Q of 25 will very likely have a higher quality decoded output than a more complex moving object that was coded with a Q of 12. To find appropriate Q values for every object in the scene, it is necessary to extend the *SVO* algorithm to analyze object-based data and distribute the total allocated bit budget among multiple objects. This chapter addresses the problem of how to

distribute the available resources (bitrate) among different *VO*s with different characteristics.

The bitrate distribution for the *MVO* problem has been addressed in many researches **[5,21-24]**. The simplest solution will be to assign a pre-defined bitrate to each *VO* and perform bitrate control for each *VO* independently. The output bitrate is the sum of the individual bitrates for the various *VO*s. However this approach is obviously sub-optimal, as characteristics of the objects are not taken into account. Since *VO*s may change significantly over time (e.g. scene changes, disappearing *VO*s), it is desirable that an algorithm considers the evolution of the relevant characteristics of the various objects and dynamically allocate available resources among them. This distribution algorithm will assign an instantaneous bitrate to each *VOP*, depending on its priority and a set of *VO* characteristics. This allocated bitrate will then be further shaped by the *CQRC* algorithm to yield a final target bitrate for each *VO*.

In **[22,23]**, measures such as variance, contrast, size and motion were incorporated to locate areas of interest, and hence prioritize the perceptive importance of each different *VO*. The available bitrate is then distributed accordingly. In the statistical multiplexing (StatMux) problem **[24]**, it has been shown that adjustments can be made on the Q of several encoders to ensure that the channel capacity is being efficiently utilized. Their proposed scheme attempts to achieve uniform quality among different video encoders by comparing the statistical variation between the different programs.

In this research the target bitrate distribution algorithm is adopted from [5], which is a combination of philosophies from the perceptual efficient approach and the StatMux approach. Size, motion and variance (defined in this case as MAD^2 instead of MAD because of its better performance [5]) are used as the factors for calculating the distribution ratio between different VOs. The target bitrate (T_i) for object i is given by:

$$T_i = T_{total} \cdot (\omega_S \cdot SIZE_i + \omega_M \cdot MOTION_i + \omega_V \cdot VARIANCE_i) \qquad (3.18)$$

where $SIZE_i$, $MOTION_i$ and $VARIANCE_i$ are the size, motion and variance of video object i, normalized by the total $SIZE$, $MOTION$ AND $VARIANCE$ of all objects in the sequence.

$SIZE$ is the number of MBs in the VO.

$MOTION$ is the sum of the absolute values of motion vectors within the VO.

VARIANCE is MAD^2 of the VO.

T_{total} is the total bitrate allocated to all the video objects for that time instant.

The weights $\{\omega_S, \omega_M, \omega_V\} \in [0,1]$ and satisfy the constraint $\omega_S + \omega_M + \omega_V = 1$

From the experimental results presented in [21], it can be concluded that size is not an important criteria as large objects with low activity will take most of the available bits and achieve a much higher $PSNR$ than other subjectively more important objects that may be small. Since target application of this research is surveillance, which usually has a large static background, it is desirable to spend more bits on the foreground, or objects with relatively larger motion. The complexity of the VO is also a very important consideration for bit distribution, as more complex images require more bits to achieve the desired quality.

[5] puts more weight on variance ($\omega_S = 0.25$, $\omega_M = 0.25$ and $\omega_V = 0.5$.) while **[21]** emphasizes more on motion ($\omega_S = 0.2$, $\omega_M = 0.5$ and $\omega_V = 0.3$). Based on the above arguments and experimental trial and error, the final weights used for this research are a combination of both works: $\omega_S = 0.2$, $\omega_M = 0.35$ and $\omega_V = 0.45$. While testing the proposed algorithm, it is found out that the encoder performance is insensitive to the specific weighting factors so long as the heuristics discussed above is followed.

MPEG-4 allows each *VO* to be coded at different frame rates. However when objects are coded at different frame rates, it is highly possible that undefined pixels will be present in the composite image. To avoid these composition problems, we impose the restriction to code each object at the same frame rate. Although a large amount of savings (in terms of bitrate) can be achieved by coding *VO*s at different frame rates, there is currently no benchmark method (to the knowledge of the author) to effectively overcome the composition problem. Another modification on the *SVO* algorithm concerns the frame skipping mechanism. For situations where buffer overflow may occur and the rate control algorithm decides to skip the *VOP* coding, frames of all other *VO*s will be skipped for that time instant.

3.8   Performance Measure

Measuring video quality is a difficult task as there are many factors influencing the results. Since visual quality is inherently subjective, some quality measurements are based on subjective tests where assessors are invited to grade the sequences (e.g. the

Double Stimulus Continuous Quality Scale [9]). However, this process is both time-consuming and expensive (in resources).

Objective video quality measurements are preferred over subjective video quality by many primarily because of their ability to produce accurate, repeatable results and because of the high cost for subjective video quality assessments. *PSNR* is a popular quality measure due mainly to its simplicity. The objective of many standard reference video rate control schemes [3,5,6,13,24] is to maximize average *PSNR* given real-world constraints.

$$PSNR_{DB} = 10\log_{10}\frac{(2^n - 1)^2}{MSE} \qquad (3.19)$$

where *n* is number of object bits in a frame.

MSE is the mean squared error between the original frame and the reconstructed frame.

However a high *PSNR* value does not equate to high video quality as it does not necessarily correlate with 'true' subjective quality. Tests have shown that observers are much more sensitive to temporal fluctuations in quality than actual frame quality [9]. Flickering artifacts and fluctuating video quality have a significant impact on perceived video quality whilst their influence on average *PSNR* is not significant. Despite the fact that the change in *PSNR* does not correspond fully to flickering, it is noted that by keeping the image quality of each frame almost constant, the flickering effect can be drastically reduced [4]. As yet, there is no objective measurement system that can completely reproduce the subjective experience of a human observer.

Since the main objective of this research is to maximize subjective video quality by minimizing quality fluctuations, the variance of the *PSNR* (σ^2_{PSNR}), which has high correlation with subjective video quality, is chosen to be the evaluation metric as it has been proven to be an extremely useful measure of spread because of its mathematically tractable nature. The formula for an unbiased estimate of the variance is:

$$\sigma^2_{PSNR} = \frac{\sum_{i=0}^{N}(PSNR_i - PSNR_{mean})^2}{N-1} \qquad (3.20)$$

where σ^2_{PSNR} is the variance of the *PSNR* of the video sequence

$PSNR_i$ is the *PSNR* of the ith frame

$PSNR_{mean}$ is the mean *PSNR* of all frames in the sequence

N is the total number of frames in the sequence.

To improve the accuracy of the performance measurements and to include an element of qualitative assessments into this research, subjective visual tests are conducted by the author to grade the visual quality of all re-constructed output video sequences.

3.9 Conclusion

In this chapter, the inner mechanisms of the *SRC* scheme and the ρ-domain rate control model are studied in detail. These serves as the fundamental framework for the proposed object-based rate control algorithm, which aims to minimize quality fluctuations. A suitable distortion index, average *DCT* reconstruction error, is chosen for the constant quality management mechanisms of the proposed algorithm. Issues such as video segmentation problems and the extension of the proposed algorithm to include multiple video objects are also presented in this chapter. Finally a suitable performance measure is proposed to evaluate the effectiveness of the proposed algorithm in minimizing visual quality variations.

CHAPTER 4: THE CONSTANT QUALITY RATE

CONTROL SCHEME

4.1   Introduction

The flow of the *CQRC* algorithm is presented below (see *Fig. 3*). The modules for the algorithm are very similar to that of the *SRC* scheme. The main difference in the program flow is that for *SRC*, the model is reset after every *GOP*. Since the objective of the *CQRC* is to achieve consistent quality within the whole sequence, and not constant quality within each *GOP*, this reset model function is discarded. The algorithm has been described in stages so that individual additions and modifications to the existing SRC algorithm are justified. This will be discussed in the following subsections.
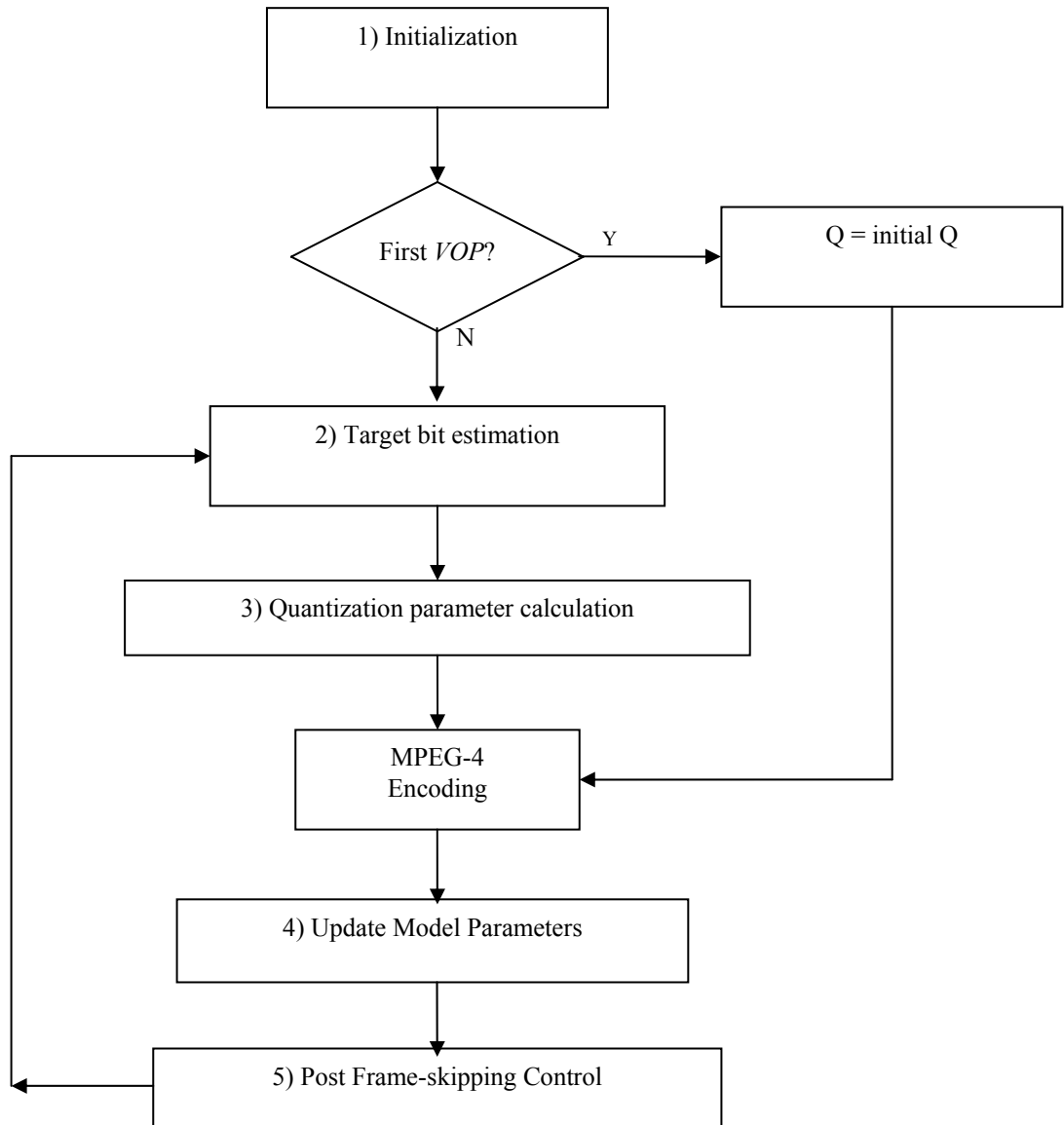
Fig. 3: Flowchart of the CQRC scheme

4.2   Initialization

Similar to *SRC* scheme, this is the stage where model parameter values, initial quantization values, buffer occupancy and other encoding parameters are initialized. The proposed *CQRC* scheme will also need to initialize the ρ-domain rate control model. After the 1st *PVOP* is encoded, the number of non-zero *DCT* coefficients before and after quantization is used to calculate current frame distortion index *D(Q)* using

equation (16). This distortion index will be set as the target distortion index (D_{target}) for the next (2nd) *PVOP*. A separate thread will run using (Q+1) as the quantization parameter to find $N_{NZ}(Q+1)$ and subsequently $D(Q+1)$. The 3 model parameters (θ, β and σ^2) can then be calculated by using these 2 pairs of $\{N_{NZ}, D\}$: the $\{N_{NZ}(Q), D(Q)\}$ pair, and the $\{N_{NZ}(Q+1), D(Q+1)\}$ pair. *Equation (3.16)* is reproduced below:

$$D(Q) = \frac{1}{N}\sum_{u=0}^{N-1}[DCT_{ORIG}(u) - DCT_{RECON}(u)]^2 \qquad (3.16)$$

Note that rate control for *SRC* will commence after the 1st *PVOP* of *every GOP* is encoded (the 1st *PVOP* of each *GOP* is encoded using a user-determined initial Q) whereas the *CQRC* scheme will commence rate control straight after the 1st frame of the sequence is encoded.

4.3 Computation of the Target Bitrate before Encoding

The main difference between the *CQRC* scheme and the *SRC* scheme is in bit allocation. The *SRC* scheme allocates a constant amount of bits for each *GOP*, and it allocates bits to each *VOP* based on the remaining available bitrate and the bitrate used by the previous *VOP*. Bit allocation based on availability of bits will however not produce a video sequence with constant quality as mentioned in the previous section. The *CQRC* scheme however will allocate bits based on the ρ-domain rate control model to achieve uniform video quality. The target bitrate estimation for the *CQRC* scheme is determined in 2 stages:

4.3.1 Initial Bit Estimation

Based on D_{target} and the model parameters α and σ^2 found in the initial stage, the desired number of non-zero *DCT* coefficients ($N_{NZ,Desired}$) can be estimated using *equation (4.1)*. The initial desired bit budget $R_{initial}$ for the current frame can then be approximated using *equation (4.2)*. Note that the target bitrate estimated here is bit allocation for *texture* only. Bits for motion and shape for the current *VOP* is assumed to be the same as it's previous *VOP* and is added to the estimated texture bit budget to yield the initial target bit estimation for the current frame. Tests have shown that the actual bit-count for non-texture information (such as headers, motion vectors, and shape information) is invariant with the value of Q used.

Derived from *equation (3.15)*:

$$N_{NZ,Desired} = -\beta \cdot \ln(\frac{D_{Target}}{\sigma^2}) \qquad\qquad (4.1)$$

Derived from *equation (3.14)*:

$$R_{initial} = \theta \cdot (N_{NZ,Desired}) \qquad\qquad (4.2)$$

The Joint Buffer Control mechanism used by the *SRC* scheme is omitted for the *CQRC* scheme as it is not our objective to maintain buffer occupancy at 50%.

4.3.2 Buffer Management Strategy

The bit allocation problem is jointly treated with the buffer control problem, thus realizing constraints set forth by the network. To avoid buffer overflow and underflow, the target bitrate is further shaped by current buffer occupancy. The final bit budget allocation scheme is given as follows:

$$R_{final} \quad = TH_{over} \text{ x } B_{max} - B_c + \frac{C}{F} \quad if \quad (B_c + R_{initial}) \quad > TH_{over} \text{ x } B_{max}$$

$$= TH_{under} \text{ x } B_{max} - B_c + \frac{C}{F} \quad if (B_c + R_{initial} - \frac{C}{F}) < TH_{under} \text{ x } B_{max}$$

$$= R_{initial} \qquad\qquad otherwise \qquad\qquad (4.3)$$

where C is the target (channel) bitrate.

F is the target frame rate.

TH_{over} and TH_{under} are set to 90% and 10% respectively, same as the *SRC* scheme.

A similar buffer management strategy is used for both schemes. However it is noted that the buffer used in the *SRC* scheme is not the actual transmission buffer (as the buffer is re-initialize to 50% occupancy after the end of each *GOP*) but a virtual one. For the *CQRC* scheme, actual channel buffer used for transmission is used as the condition both for frame skipping and buffer over/underflow checks.

SRC's virtual buffer assumes that the number of bits removed from the buffer is the average allocated bits of each *PVOP*. To cater to real-world channel constraints, R_p for the implemented *SRC* scheme is set to be $\dfrac{C}{F}$. *Equation (3.4)* is reproduced below:

$$R_p = (R_s - R_{IVOP})/N \qquad\qquad (3.4)$$

4.4   Quantization Parameter Calculation

Quantization parameter calculation of the *CQRC* scheme is identical to the one used by the *SRC* scheme. The quadratic Rate-Quantization model is used to calculate the ideal *Q*. The same constraints on *Q* used for the *SRC* scheme still apply here. *Equation (3.5)* for the Quadratic *R-Q* Model is reproduced below:

$$T_{texture} = \frac{X_1 \times MAD_c}{Q_c} + \frac{X_2 \times MAD_c}{Q_c^{\,2}} \qquad\qquad (3.5)$$

4.5   Updating the Model Parameters

The model parameters for both the quadratic Rate-Quantization model and the ρ-domain rate control model are updated continually in this stage, by using both the encoded results of the current frame and a specified number of past frames.

4.5.1 Quadratic Rate-Quantizer Model

The *CQRC* scheme follows strictly the update procedures used by the *SRC* scheme. By using a sliding window data-point selection module, applying statistical removal of data outliers and using the least-mean square approximation, X_1 and X_2 can be determined. Since the *CQRC* scheme does not reset the sliding window data points after each *GOP* (unlike the *SRC* scheme), current sliding window data points may cover those before scene changes, resulting in w being excessively large. Hence to improve the model accuracy, an additional constraint $w_{i+1} \le w_i + 1$ is applied.

4.5.2 ρ-domain Rate Control Model

By monitoring the number of *DCT* zeros before and after quantization, the distortion index of the current *VOP* can be calculated. Target distortion index ($D_{TARGET,}$) is then updated. The $\{N_{NZ}(Q), D(Q)\}$ of the previous and the current *VOP* will then be used to update the model parameters (θ, σ^2 and β). *Equation (3.17)* for calculating $D_{TARGET,k}$ is reproduced below:

$$D_{TARGET,k} = \frac{1}{k-1}\sum_{i=1}^{k-1} D_i \qquad\qquad (3.17)$$

4.6   Post Frame-Skipping Control

The encoder will skip subsequent frames until B_c is less than 80% of total buffer size. This post frame-skipping mechanism will prevent overflowing of the channel buffer. The frame-skipping control of the *CQRC* scheme is identical to that of the *SRC* scheme discussed in the above section with the exception that the virtual buffer (used by *SRC*) is substituted by an actual channel buffer. Since the *SRC* scheme resets the buffer occupancy back to 50% after encoding each *GOP* (which most probably will *not* coincide with the actual channel buffer), a separate channel buffer is simulated for the *SRC* scheme and post frame-skipping issues will be formulated based on this actual channel buffer. Adjustment of the allocated bit budget by *SRC*'s buffer management mechanisms will however still use its original virtual buffer. For the case of *MVOs*, whenever the rate control scheme decides to skip a *VOP*, *VOPs* of **all** other *VOs* will also be skipped for that time instant.

4.7 Conclusion

In this chapter, the proposed *CQRC* algorithm is explained in stages. It involves the combination of the *SRC* algorithm and the ρ-domain rate control model. By allocating bit budget of *VO*s based on quality instead of buffer occupancy, the *CQRC* algorithm aims to achieve minimization of visual quality fluctuations for the reconstructed (encoded and decoded) video sequence. Modifications and additions to the existing *SRC* algorithm are presented to facilitate implementation of this proposed algorithm for real-world problems.

CHAPTER 5:   EXPERIMENT AND RESULTS

5.1   Introduction

In this section, the performance of the proposed *CQRC* algorithm is illustrated using 3 different video sequences with different characteristics and different number of *VO*s. For comparison reasons, the MPEG-4 Annex L **[3]** frame-level rate control scheme (*SRC* scheme) is used as a reference. Experimental results are provided to evaluate the performance of the algorithms within each individual *VO*s under hard real-time conditions and low bandwidth constraints. The first set of experiment aims to demonstrate the effectiveness of the proposed *CQRC* scheme to limit video quality fluctuations. The second set of experiment aims to justify the performance of the proposed scheme under realistic real-life constraints by using imperfect segmentation masks for encoding. The final set of experiment will show the ability of the proposed algorithm in handling multiple video objects.

5.2   Experimental Setup

From the classical works on bit allocation, it is clear that rate control is dependent on the application. For this research, the target application chosen is real-time video streaming for video surveillance information. All the test sequences used are 300 frames long, in *CIF* format (352 x 288) and the shape information is binary-coded. Since the target application for this research is on video surveillance, all test sequences

will have a static background with no camera movement. 3 different video sequences with different characteristics are used for testing. The number of objects in each sequence is as follows: "*Akiyo*" (2 objects), "*Silent*" (2 objects) and "*News*" (3 objects). For this research, *VO*0 is defined as the background object while *VO*1 and *VO*2 are defined as the foreground objects.

The target bitrate (*C*) and frame rate (*F*) is set to 128 Kbps at 10 fps and the size of the buffer is chosen to be a second's length (i.e. $B_{max} = C = 128$ Kbits). A constant frame rate is used for all encoded *VO*s to avoid problems associated with composition as mentioned in *section 3.7*. The *GOP* structure used for this scheme consists of 5 *PVOP*s for each IVOP. Similar to the *SRC* scheme, bitrate control is only applied to *PVOP*s. In this work, B-type *VOP*s are not mentioned but since they operate similarly to *PVOP*s, they can be tackled in a similar fashion. Mean *PSNR* and the *PSNR* variance of the reconstructed output *VO*s will be used as the evaluation criteria for the performance of the algorithm as discussed in the previous section. All tests are done on a Pentium 4 PC with a CPU speed of 2.6 GHz and a RAM of 512 MB.

The proposed bitrate control scheme is applied only to texture coding and is not extended to motion and shape data. This is because motion and shape are of higher priority, and also because the Rate-Quantization model and Rate-Distortion model used are purely texture-based. Hence the necessary bits for motion and shape coding will always be allocated (lossless coding), using the remaining bits for texture coding. Concerning texture information, the dominant rate control parameter is the quantization parameter for the *DCT* coefficients and this will be adjusted to achieve the objective of minimizing variation of frame distortion.

5.3 Experimental Results for "*Akiyo*"

5.3.1 Introduction

The sequence "*Akiyo*" is encoded under the MPEG-4 framework using the *CQRC* scheme and the *SRC* scheme separately. Perfect segmentation files are used for encoding (see *Fig. 4 – 6* below for a sample of the original frame, the segmentation mask used and the reconstructed VO). The *PSNR* characteristics of their reconstructed *VO*s (*VO*0 and *VO*1) are illustrated in *Fig. 7* and *Fig. 8* (respectively) while the results are summarized in *Table 1*.

Fig. 4   Original 'Akiyo' frame 1

Fig. 5 Perfect segmentation mask for 'Akiyo' VO1 frame 1
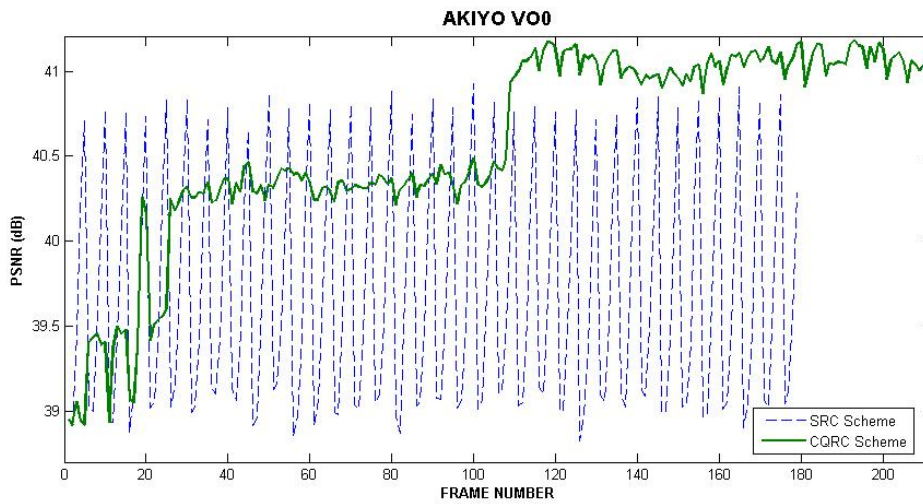

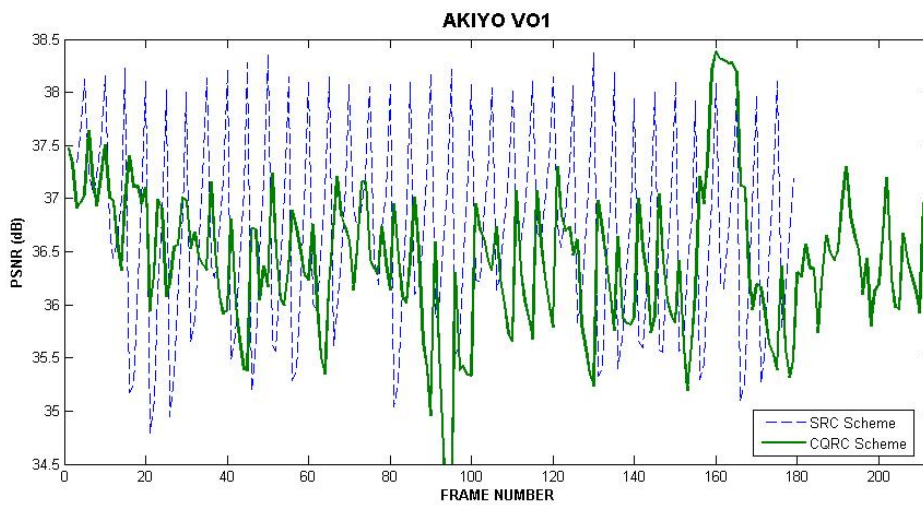*Fig. 6 Reconstructed "Akiyo' VO1 frame 1*

Fig.7 Performance of "Akiyo" VO0

Fig.8 Performance of "Akiyo" VO1

	SRC scheme			CQRC scheme		
	Average PSNR (dB)	$\sigma^2_{PSNR,SRC}$	Frames skipped	Average PSNR (dB)	$\sigma^2_{PSNR,CQRC}$	Frames skipped
VO0	39.702679	0.486559	85	40.571693	0.325386	45
VO1	36.704742	0.865658		36.402752	0.493912	

Table 1: Performance comparison between CQRC scheme and SRC scheme for "Akiyo"

5.3.2 Drawback of the *SRC* Scheme

The behaviour of the *PSNR* characteristics for both *VO*0 and *VO*1 shows clearly that the *SRC'*s bit allocation approach (which is based on bits availability and buffer occupancy) will inevitably lead to significant video quality fluctuations. It also demonstrates the inadequacy of using a *GOP*-based bit allocation principle. As more bits are available for allocation after each *PVOP* in the *GOP* is encoded, more bits will be allocated to the next *PVOP* resulting in an increase in *PSNR*. The very last *PVOP* in the *GOP* will be allocated *all* the remaining bits allocated to the *GOP* leading to very high frame quality. At the start of the next *GOP*, the 1st *PVOP* will be quantized by the user-determine initial *Q*, resulting in a steep drop in *PSNR*, and the cycle continues. This will lead to a 'zigzag' curve, which visually corresponds to flickering effects and a temporally distorted video sequence output. This phenomenon, when using the *SRC* scheme, is not only true for "*Akiyo*", but also to the other video sequences used in this research.

5.3.3 Characteristics of *VO0*

From *Fig. 7*, it can be seen that the *CQRC* scheme is highly effective in controlling the variation of the video quality for *VO*0. There are a few relatively large fluctuations in *PSNR* at the initial 20 or so frames due to the fluctuations in target distortion (average distortion of previous *K* encoded frames) in the beginning of the sequence. As *K* gets larger, the target distortion will converge to a stable value and the fluctuations of the *PSNR* will cease, leading to a fairly-constant *PSNR* value. At the 109th *PVOP*, the algorithm decides to decrease *Q* for *VO*0 from 7 to 6 both to maintain buffer stability and to achieve the target distortion. This leads to a 0.5 dB increase in *PSNR*. Preceding *PVOP*s shows minimal quality fluctuations.

It is noted that even though $\sigma^2_{PSNR,CQRC}$ of *VO*0 is 0.325386 (66.9% of $\sigma^2_{PSNR,SRC}$), the actual improvement in video quality is much higher. This can be seen from the characteristics of *Fig. 4*. In the 1st half of the sequence (ignoring the initial stages), video quality is very consistent with a $\sigma^2_{PSNR,CQRC}$ of 0.244800 from the 25th *PVOP* to the 108th *PVOP*. After the *PSNR* surge, $\sigma^2_{PSNR,CQRC}$ of the 110th *PVOP* to the end of the sequence is 0.164417. Both σ^2_{PSNR} (0.244800 and 0.164417) are much less than σ^2_{PSNR} of the whole sequence (0.325386). This phenomenon shows that σ^2_{PSNR} is unable to reproduce the actual visual quality of the video sequence. One must study its characteristics over time to effectively measure the visual fluctuation of a video sequence. Nevertheless, σ^2_{PSNR} is still an effective tool to measure visual quality fluctuations.

5.3.4 Characteristics of *VO*1

For the foreground object (*Fig. 8*), the variation of the *PSNR* is comparatively larger than that of its background. This is expected, as more motion vectors are present in the foreground object. Since previous encoded data are used to estimate the rate control models, the non-stationary nature of *VO*1 will result in model inaccuracy leading to non-uniform video output quality. However when compared to the *SRC* scheme, it can be clearly seen that the *CQRC* algorithm is much more successful in containing and reducing the fluctuations of video quality. Apart from a few *PVOP*s with large changes in *PSNR* (e.g. *PSNR* of the 94th *PVOP* drop by 0.86 dB) due to model inaccuracy, *PSNR* for most of the *VOP*s are fairly stable.

5.3.5 Comparison of Performance between *CQRC* and *SRC*

From *Table 1* it is observed that for the *CQRC* scheme, *PSNR* fluctuation has been greatly reduced. $\sigma^2_{PSNR,CQRC}$ of *VO*0 and *VO*1 dropped by 33.1% and 42.9% respectively when compared to the *SRC* Scheme. Average *PSNR*s for both schemes are comparable for both objects. When using the *CQRC* scheme, *PSNR* of *VO*0 increases by 0.87 dB whereas *PSNR* of *VO*1 decreases by 0.30 dB when compared to the *SRC* scheme. This shows that the *CQRC* scheme can perform as well as the *SRC* scheme in terms of delivering the optimal average *PSNR* given buffer constraints. The *CQRC* scheme also shows a tremendous improvement in reducing the number of frames skipped. Using the *SRC* scheme, 28.3% of frames are skipped whereas only 15% of frames are skipped when using the *CQRC* scheme. This means that for the same

bandwidth and buffer constraints, the *CQRC* scheme is able to code more frames, significantly improving video quality in terms of motion continuity.

Subjective tests are conducted to compare actual subjective visual quality of the composed scene (i.e. the 2 *VO*s are combined) using both rate control algorithms. The improvement in subjective video quality is obvious. Using the *SRC* scheme, there are flickering artifacts in both the background and the foreground. This is very distracting especially in the background because of the lack of motion. Comparatively, the *CQRC* scheme produces a background with very stable video quality. Subjective tests have proven that the abrupt change in *PSNR* at the 109th *PVOP* for *VO*0 is not noticeable and overall visual quality of '*Akiyo*' when using the *CQRC* scheme is much more stable than using the *SRC* scheme. Although video quality fluctuations are still present in the moving foreground object when using the *CQRC* scheme, the extent of these variations in quality is much lower than that of the *SRC* scheme.

5.4 Experimental Results for "*Silent*"

5.4.1 Introduction

To verify the robustness of the *CQRC* algorithm under real-life constraints, the sequence "S*ilent*" is encoded using imperfect segmentation mask (unlike in the earlier experiment where "*Akiyo*" is encoded using perfect segmentation files). A real-time object segmentation algorithm is used to segment the different *VO*s in the video sequence, and the results are fed as input into the MPEG-4 encoder. Segmented *VO*s are noisy (especially around the boundaries) and artifacts (spatial and temporal) are present in both the foreground and background. *Fig. 9 – 11* presents a sample of the "Silent" sequence and an imperfect segmentation mask used for encoding. It can be seen that the movement of the woman and her shadow leads to erroneous segmentation results. Performance of both the *CQRC* scheme and the *SRC* scheme, when faced with real-life segmentation problems like noise and artifacts, are shown and analyzed. *Fig. 12* and 13 below illustrates the *PSNR* characteristics of '*Silent*' while *Table 2* summarizes the results.

Fig. 9 Original 'Silent' frame 30

Fig. 10 Reconstructed "Silent' VO1 frame 30

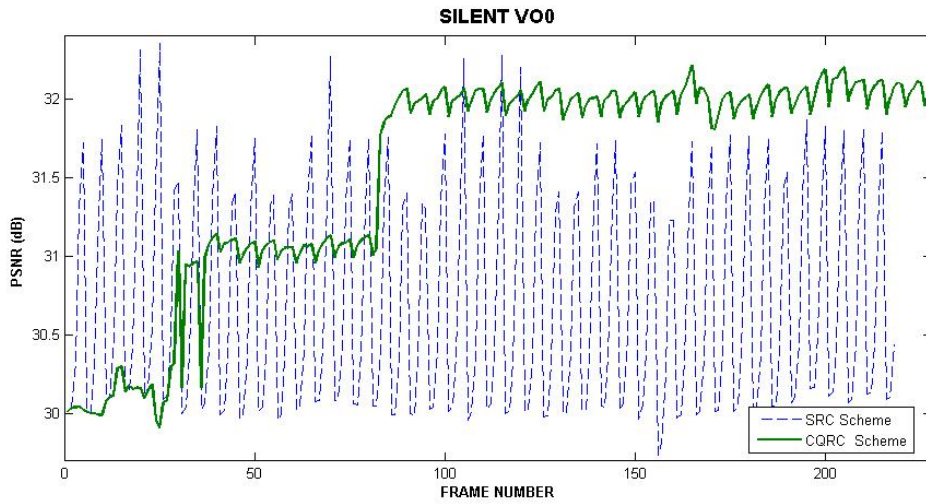*Fig. 11 Imperfect segmentation mask for 'Silent' VO1 frame 30*
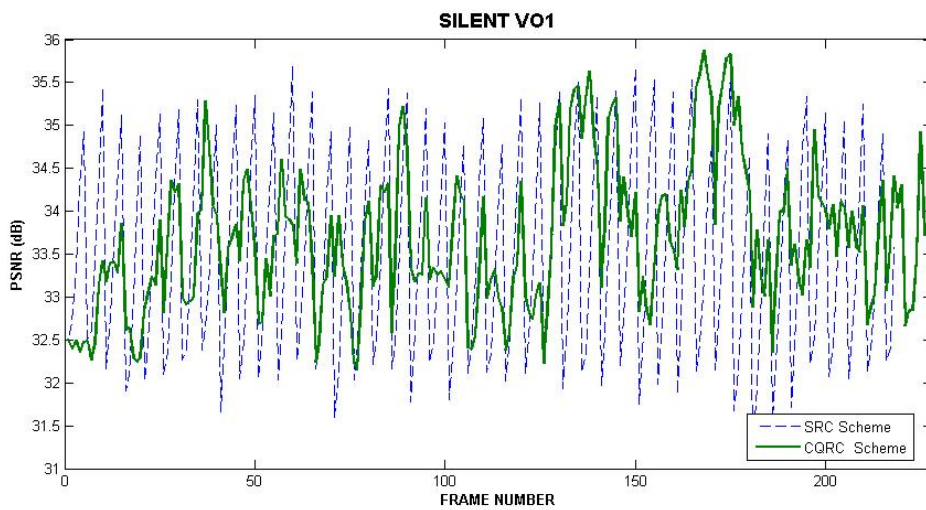

Fig.12 Performance of "Silent" VO0

Fig.13 Performance of "Silent" VO1

Silent					
	SRC scheme			CQRC scheme		
	Average PSNR (dB)	$\sigma^2_{PSNR,SRC}$	Frames skipped	Average PSNR (dB)	$\sigma^2_{PSNR,CQRC}$	Frames skipped
VO0	30.705177	0.511949	38	31.53688	0.483082	26
VO1	33.485985	1.430744		33.682316	0.736513	

Table2: Performance comparison between CQRC scheme and SRC scheme for "Silent"

5.4.2 Effects of using Imperfect Segmentation Files

As expected, the video qualities (both in terms of average *PSNR* and σ^2_{PSNR}) of both *VO*s of '*Silent*' using both schemes are comparatively much poorer than the earlier case which uses perfect segmentation files. This is due to the additional overheads, complexity and noise introduced by the imperfect segmentation. An interesting phenomenon is that the background quality for "*Silent*" is actually lower than that of its foreground quality. Although the distribution algorithm for most *MVO* schemes usually allocates lesser bits to the background object, background quality is usually much higher than the foreground objects because it is static (lack of motion and overheads) and thus bits allocated will be used to code it with higher quality. The fact that the background object has a higher visual quality will have a negative impact on viewers, as they will find that quality of the foreground objects even worse. Hence many distribution algorithms **[21]** allocate even lesser amount of bits for background objects. However for this case, due to the imperfect segmentation, additional bits are

needed to encode the noise and artifacts present in the background, resulting in it having a lower video quality than its foreground. This effect due to segmentation must be taken into account when formulating any practical bit distribution algorithm for multiple arbitrary shaped *VO*s.

5.4.3 Comparison of Performance between *CQRC* and *SRC*

From *Table 2* above, it is observed that by using the *CQRC* scheme, number of skipped frames are reduced (from 12.7% to 8.7%) and video sequence of much more consistent quality ($\sigma^2_{PSNR,CQRC}$ of *VO*0 and *VO*1 drop by 5.6% and 48.5% respectively when compared to the *SRC* scheme) is produced. Similar to the case of "*Akiyo*", the slight improvement in the σ^2_{PSNR} of the background (VO0) does not coincide with actual visual quality improvement (see characteristics of the curve in *Fig. 6*). It is clear, to viewers, that the flickering effects are greatly reduced when using the *CQRC* algorithm. The average *PSNR* for "*Silent*" using the *CQRC* scheme shows an slight improvement of 0.83 dB and 0.20 dB for both *VO*s respectively. These experimental results are consistent with the previous experiment.

5.5 Experimental Results for "*News*"

5.5.1 Introduction

To verify the performance of the *CQRC* scheme for *MVO* (the previous 2 sequences consist of only 1 foreground and 1 background *VO*), the video sequence "*News*" is used. "*News*" is chosen because of the significant coding qualities difference between *VO*1 and *VO*2 due to their different characteristics in time: *VO*1 is of a large size but with low activity whereas *VO*2 is of a smaller size but with high activity and several scene cuts (see *Fig. 14 -16* below). Similar to the "*Silent*" sequence, imperfect segmentation masks are generated for the three *VO*s and fed into the encoder. *Fig. 17-19* illustrates the *PSNR* characteristics of the *VO*s of '*News*' for both rate control schemes while *Table 3* summarizes the results.

Fig. 14:" News" VO0: Background
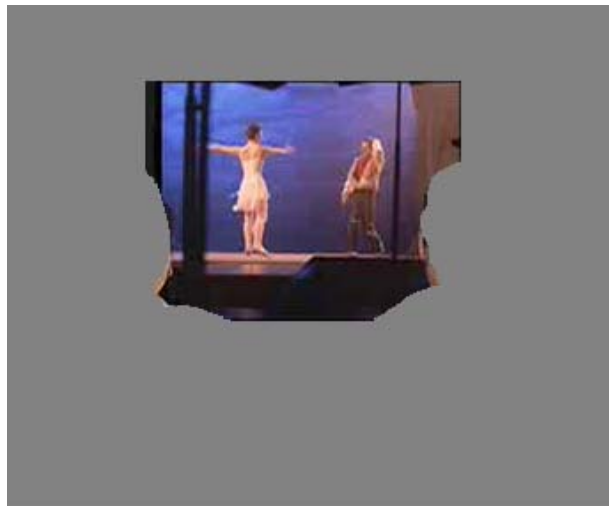

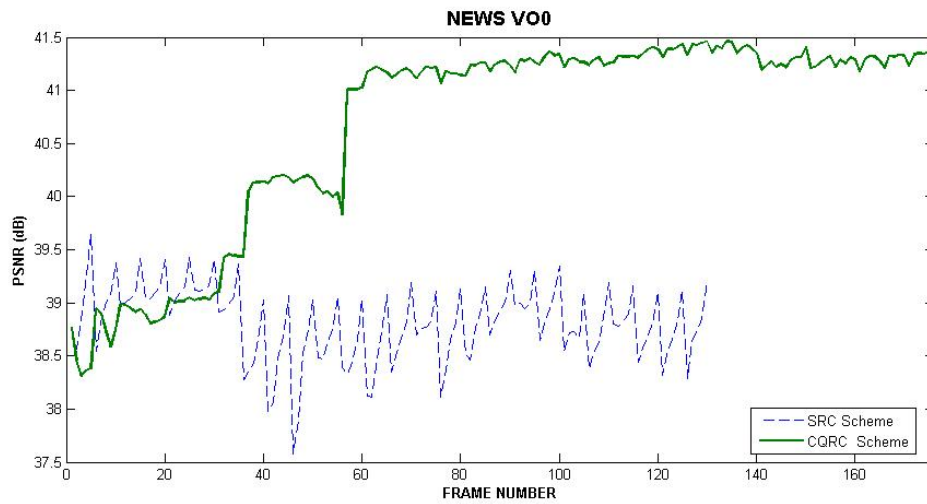*Fig. 15: "News" VO1: Speakers*

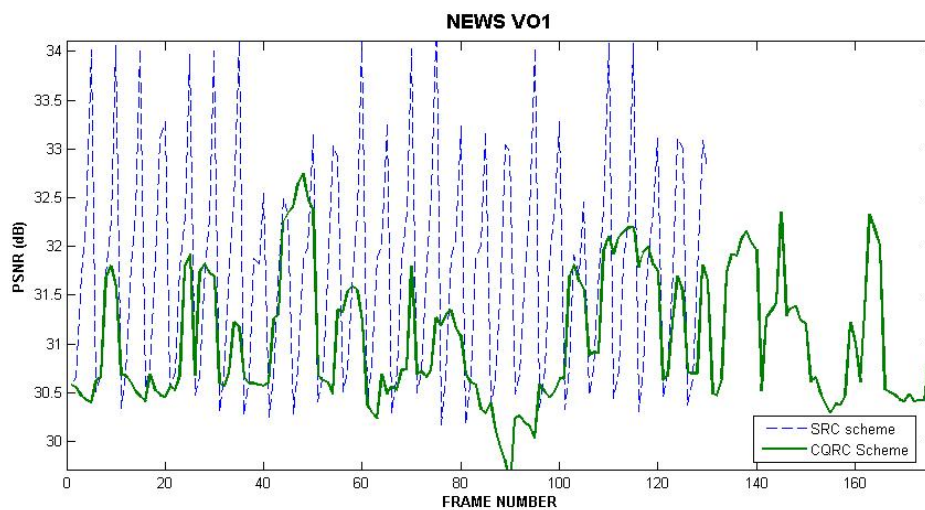*Fig. 16: "News" VO2: Monitor*

Fig.17 Performance of "News" VO0

Fig.18 Performance of "News" VO1
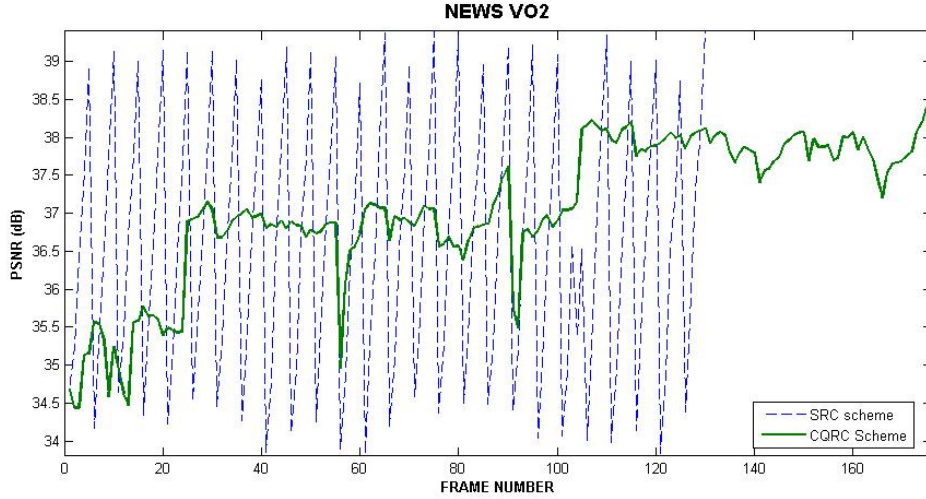
Fig.19 Performance of "News" VO2

	News					
	SRC scheme			*CQRC* scheme		
	Average PSNR (dB)	$\sigma^2_{PSNR,SRC}$	**Frames skipped**	**Average PSNR (dB)**	$\sigma^2_{PSNR,CQRC}$	**Frames skipped**
*VO*0	38.812168	0.123761		40.673882	0.952602	
*VO*1	31.755045	1.385077	145	31.03488	0.458992	88
*VO*2	36.554653	3.024824		37.043655	0.886841	

Table 3: Performance comparison between CQRC scheme and SRC scheme for "News"

5.5.2 Comparison of Performance between *CQRC* and *SRC*

For the foreground objects *VO*1 and *VO*2 (see *Fig. 12* and *13*), $\sigma^2_{PSNR,SRC}$ is very high due to its high activity and the limited bandwidth allocated to it. By using the *CQRC* scheme, fluctuations in video quality is effectively controlled by allocating bits based on maintaining constant video quality. $\sigma^2_{PSNR,CQRC}$ of *VO*1 and *VO*2 dropped by 66.8% and 70.7% respectively when compared to the *SRC* scheme. For *VO*0, $\sigma^2_{PSNR,CQRC}$ is

higher than that of $\sigma^2_{PSNR,SRC}$, which seems to contradict the previous conclusion that the *CQRC* scheme can greater reduce σ^2_{PSNR} and consequently quality variation. For this case, the value of σ^2_{PSNR} does not fully corresponds to quality variations. By observing the characteristics of the σ^2_{PSNR} curve in *Fig. 17*, it can be seen that quality variation is in actual fact much reduced compared to the *SRC* scheme. The two abrupt changes in *Q* (the 36th *PVOP* and the 58th *PVOP*) will lead to a high value of σ^2_{PSNR}. However apart from this abrupt change in *PSNR* at those 2 points, the rest of the *PVOP*s exhibit very consistent *PSNR* behaviour. Subjective quality tests also states that the visual quality of the background object is much more stable when using the *CQRC* scheme. All these are consistent with the previous discussion in the last 2 experiments.

From *Table 3*, it is can be seen that there is a tremendous improvement in video quality when using the *CQRC* scheme to encode multiple video objects, both in σ^2_{PSNR} and number of frames skipped. Comparing both rate control schemes, it is observed that the *SRC* scheme performs better in producing a higher average *PSNR* for *VO*1, the visually most important *VO* in the sequence, compared to the *CQRC* scheme. The average *PSNR* of *VO*1 when using the *SRC* scheme is 0.72 dB higher than that of *VO*1 when using the *CQRC* scheme. However this slight improvement in average *PSNR* is overwhelmed by the large fluctuations in video quality ($\sigma^2_{PSNR,SRC}$ = 1.385077 vs. $\sigma^2_{PSNR,CQRC}$ = 0.458992, a drop of 70.7%) and the additional 57 frames dropped (an improvement of 19%). It is noted that there is a very significant amount of frames skipped for both algorithms. This is due mainly to poor object segmentation, where bits needed to encode each *VO* are greatly increased leading to more frames being

discarded to maintain buffer stability. The bitrate distribution mechanism also contributes to this increase in number of frames skipped. The post frame-skipping control mechanism of both rate control algorithms will skip frames from all VOs if one VOP fails to meet the frame-skipping condition. When *VO*1 (which incidentally is the visually most important *VO* and the one with the most noise due to segmentation) is not allocated enough bits, frame skipping will occur, resulting in the algorithm skipping frames from the other 2 VOs. From *Fig. 17-19*, it is apparent that *VO*0 and *VO*2 has a much higher average *PSNR* compared to *VO*1 due to the target bitrate distribution algorithm allocating 'more than enough' bits to them. If more bitrate is diverted from *VO*0 and *VO*2 (which does not require as much bits as *VO*1) to *VO*1, lesser frames will be dropped.

By observing the characteristics of the *PSNR* performance of all the 3 *VO*s, it can be concluded that video quality has become much more stable especially for *VO*0 and *VO*2. Even though there are still significant quality variations in *VO*1, it is contained and minimized. These conclusions are verified by subjective assessment tests. On a frame-by-frame basis, there are no obvious differences in frame quality between *VOP*s of either output sequence from both schemes. However, when comparing visual performance of both schemes in a video-by-video basis, it is observed that visual performance of the *CQRC* scheme performs much better than that of the *SRC* scheme due to the great reduction in flickering artifacts and motion jerkiness. This observation is consistent with the conclusions derived from the qualitative experimental results. Hence, we can conclude that the *CQRC* scheme outperforms the *SRC* scheme when encoding multiple *VO*s.

5.6   Conclusion

Our experimental results have shown that with the *CQRC* algorithm, we can accurately estimate the *R-D* curve and robustly control the output bitrate or picture quality of the video coder. The proposed *CQRC* algorithm can achieve a constant quality output video sequence with much less flicker and motion jerkiness compared to the *SRC* scheme. It has better temporal video quality due to the lesser number of frames dropped and performs better in the scenario of imperfect foreground masks and multiple video objects. Although the proposed algorithm is not by any means optimal, experimental results shown here are promising.

CHAPTER 6: CONCLUSIONS AND FUTURE WORKS

6.1 Conclusion

In this project, the concept of maintaining constant visual quality of reconstructed (decoded) video sequences along time for scenes with multiple arbitrary shaped video objects is explored. The ρ-domain rate control model is combined with the quantization calculation scheme of MPEG-4 Annex L *SRC* scheme to improve modeling accuracy and robustness. By using a linear *R-ρ* model, an exponential *D-ρ* model and a quadratic *R-Q* model, the *D-Q* relationship of the video sequence can be approximated and texture rate control can be done to achieve a target distortion value.

The proposed *CQRC* scheme is applied to real-time surveillance monitoring applications and experiments are run under various real-life constraints like limited bandwidth, limited buffer size, use of imperfect segmentation masks for encoding and existence of multiple video objects. Experimental results for the different test sequences show that in terms of quality fluctuations (measured by the variance of *PSNR* and the characteristic of the *PSNR-frame* curve) and temporal quality (number of frames skipped), the proposed *CQRC* algorithm clearly outperforms the *SRC* algorithm. When using the *CQRC* scheme over the *SRC* scheme, σ^2_{PSNR} of the foreground objects decreases between 42.9% (for "*Akiyo*" *VO*1) and 70.7% (for "*News*" *VO*1) whereas number of frames skipped decreases between 4% (12 more frames encoded for "*Silent*") and 19% (57 more frames encoded for "*News*"). This is supported by subjective assessments, which states that flickering is much reduced

when using the *CQRC* scheme for all test sequences. Although the objective is not to maximize average video quality (measured in terms of average *PSNR* value), the *CQRC* algorithm is able to perform as well as the *SRC* scheme in providing the optimal average video quality given buffer and bandwidth constraints.

Overall, the algorithm does not experience any buffer overflow/underflow and video sequences are coded with reasonable quality and with much improved smoothness in video quality. The resulting visual quality of the decoded frames is significantly better than that of the *SRC* algorithm.

Main contributions of this project are thus as follows:

1. Proposed and formulated a novel object-based rate control scheme with the objective of maintaining constant visual quality along time for each individual *VO*.

2. Incorporate the ρ-domain rate control model with MPEG-4 Annex L *SRC* scheme in order to allocate bit budget among different *VO*s optimally in terms of maintaining constant visual quality.

3. Extension of the proposed *CQRC* scheme to include multiple arbitrary shaped *VO*s by means of a target bitrate distribution algorithm based on each individual *VO*'s variance, motion and size.

4. Comparison of performance between the proposed *CQRC* scheme and *MPEG*-4 Annex L *SRC* scheme under numerous real-life constraints including the use of imperfect segmentation masks for encoding.

6.2　Drawbacks of the *CQRC* Scheme

By observing the characteristics of the various *PSNR* curves for the *CQRC* scheme, it can be seen that there are still quality fluctuations present, especially for the foreground objects. This is due mainly to the fact that the rate control models used in this research are purely texture-based models. In general, the rate control problem is much simpler when the texture bits comprise of a very large percentage of the total bitrate. Bits generated by shape and motion can hence be considered as overheads and the major impact on frame quality will be in the distribution of texture bits among each object. However, the situation changes drastically when this assumption does not hold in the case of object-based *CODEC*s. A significant amount of complexity is added since the header bits no longer comprise of a small percentage of the total bitrate (because of the additional shape bits). This is especially true when using imperfect segmentation masks for encoding as noises and artifacts will significantly increase the bits for shape and motion. Hence when constant quality rate control methods are applied to non-object-based video coding standards like H263 and MPEG-2, it can be very successful in controlling quality fluctuation (as can be seen from **[11]**). However, MPEG-4 is an object-based encoder where a significant amount of bits is used to code shape and motion information. Both the ρ-domain rate control model and the quadratic rate-quantization model used are purely texture-based models. Hence, they are unable to perform optimally due to their inability to take into account the significant non-texture information, resulting in model inaccuracy. Nevertheless when compared with the *SRC* scheme, the *CQRC* scheme has higher modeling accuracy and can be generally applied in frame level bit allocation to alleviate the quality fluctuation problem.

The *CQRC* scheme also has a greater computational complexity compared to the *SRC* scheme as it needs to calculate the number of non-zeros *DCT* coefficients for every *VOP*, and it needs to formulate the ρ-domain rate control model. This increase in complexity is however not a big disadvantage as it is insignificant compared to the entire computation involved in the video encoding process of the MPEG-4 *CODEC*. There is no memory/computational overhead and it does not require a multi-pass framework. The time difference used to encode video sequences between the *SRC* scheme and the *CQRC* scheme is negligible.

6.3   Recommendation for Future Works

For improved performance, an algorithm should be devised to verify that current model parameters produce near-optimal results. To arrive at these decisions, it is necessary to have good models for describing the shape and texture of a video object. The Rate–Distortion and Rate–Quantizer model used in the proposed work is quite adequate for texture coding; however, effective models for shape have yet to be developed. It is expected that such a model will depend on the geometric attributes of the object. Rate control for shape should be studied in more detail to achieve a good tradeoff between shape and texture coding accuracy. Our experimental results show that shape rate control does not have a significant impact on the coding efficiency in terms of average *PSNR*. It however plays an important role in improving temporal resolution and reducing quality variations. This is especially important when imperfect segmentation masks are used for encoding as shape bits encompass a significant percentage of bits used by current *VO* (especially for foreground objects). Even though

the background is of lower subjective importance, its quality fluctuations have been proven to be very distracting, especially in boundaries of poorly segmented areas. A good understanding of the perceptual weighting between shape and texture distortion is required to exercise joint shape and texture rate control. Finally, some means of overcoming the composition problem should be developed so that different objects can be encoded at different frame rates. Although this would require a more complex buffering scheme, the potential savings in terms of time and computational complexity are enormous.

Experimental results show that there is a significant difference in video quality between the background, and the foreground objects. This is because the background is mostly static and hence the allocated bitrate is used to encode it to a higher quality (as opposed to the other objects, which uses a significant part of the allocated bitrate for motion and size differences). This is undesirable as the background object usually has little significance to observers (especially for surveillance applications), and bits that can be utilized to improve foreground quality is wasted. A possible solution is the choice (dependent on application) between either a user-input *VO* priority factor, or an improved *VO* distribution algorithm that takes into account the lower priority of background objects. By putting less weight on the distribution ratio allocated to the background, more bits will be freed and these will be used to improve the quality of other, more important foreground objects.

REFERENCES

[1] C. E. Shannon: A mathematical theory of communication. Bell System Technical Journal, vol. 27, pp. 379-423 and 623-656, July and October, 1948.

[2] Hung-Ju Lee, Tihao Chiang and Ya-Qin Zhang, "*Scalable Rate Control for Very Low Bit Rate (VLBR) Video*", Proceedings of the 1997 International Conference on Image Processing (ICIP '97)

[3] MPEG-4 video verification model v8.0 ISO/IEC JTC1/SC29/WG11, "*Coding of Moving Pictures and Associated Audio*", MPEG97/N1796, Stockholm, Sweden, July 1997.

[4] Hwangjun Song and C.-C. Jay Kuo, "*Rate Control for Low-Bit-Rate Video via Variable-Encoding Frame Rates*", IEEE Transactions on Circuits and System for Video Technology, Vol. 11, No. 4, Apr 2001

[5] Anthony Vetro, Huifang Sun, and Yao Wang, "*MPEG-4 Rate Control for Multiple Video Objects*", IEEE Transactions on Circuits and System for Video Technology, Vol. 9, No. 1, Feb. 1999.

[6] Feng Pan, Z. Li, K. Lim, and G. Feng, "*A Study of MPEG-4 Rate Control Scheme and Its Improvements*", IEEE Transactions on Circuits and System for Video Technology, Vol. 13, No. 5, May 2003

[7] C. Y. Hsu, A. Ortega, and A. R. Reibman, "*Joint selection of source and channel rate for VBR transmission under ATM policing constraints*," IEEE J. Select. Areas Commun., vol. 15, pp. 1016–1028, Aug. 1997.

[8] W. Ding, "*Joint encoder and channel rate control of VBR video over ATM networks*," IEEE Trans. Circuits Syst. Video Technol., vol. 7, pp. 266–278, Apr. 1997.

[9] Iain E. G. Richardson, "*H.264 and MPEG-4 Video Compression*", 2003 Wiley and Sons, Ltd. ISBN: 0-470-84837-5.

[10] LiFeng Zhao, JongWon Kim and C.-C. Jay Kuo, "*Constant quality rate control for streaming MPEG-4 FGS video*", Proceedings of the 15th international conference on Computer communication, 2002.

[11] Junqiang Lan, Xinhua Zhuang, Wenjun Zeng, "*Single-pass frame-level constant distortion bit allocation for smooth video quality*", ICME 2004.

[12] "*Test Model 5*", Draft, ISO-IEC/JTC1/SC29/WG11/N0400, April 1993.

[13] ITU-T SG16/Q15, Video Coding Expert Group, "*Video CODEC Test Model TMN8,*" Portland, June 1997.

[14] B. Xie and W. Zeng, "*A sequence-based rate control framework for constant quality video,*" submitted to IEEE T Transactions on Circuits and System for Video Technology, 2003.

[15] J. Ribas-Corbera and S. Lei, "*Rate control in DCT video coding for low-delay video communications,*" IEEE Transactions on Circuits and System for Video Technology, vol. 9, Feb. 1999, pp. 172–185.

[16] Smoot S. R. and Rowe L. "*A Study of DCT coefficient distributions,*" Proceedings of the SPIE Symposium on Electronic Imaging, volume 2657, San Jose, CA, Jan. 1996.

[17] Jinho Choi, "*Distortion Policy of Buffer-Constrained Rate Control for Real-Time VBR*", IEEE Transactions on image processing, Vol. 8, No. 4, April 1999.

[18] Zhihai He and Sanjit K. Mitra, "*A Unified Rate-Distortion Analysis Framework for Transform Coding*", IEEE Transactions on Circuits and System for Video Technology, Vol. 11, No. 12, Dec 2001

[19] 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 System for Video Technology. Volume: 12, pp.840-849, Oct. 2002

[20] T. Chiang and Y.-Q. Zhang, "*A new rate control scheme using quadratic rate-distortion modeling*," IEEE Trans. Circuits Syst. Video Technol., Feb. 1997.

[21] Paulo Nunes, Fernando Pereira, "*Rate Control for Scenes with Multiple Arbitrarily Shaped Video Objects*", Picture Coding Symposium, Berlin – Germany, September 1997

[22] P. Fleury, J. Reichel, and T. Ebrahimi, "*Image quality prediction for bit rate allocation*," in Proc. IEEE Int. Conf. Image Processing, Lausanne, Switzerland, Oct. 1996, pp. 339–342.

[23] X. Marichal, T. Delmot, C. DeVleeschouwer, V. Warscotte, and B. Macq, "*Automatic detection of interest areas of an image or of a sequence of images*," in Proc. IEEE Int. Conf. Image Processing, Lausanne, Switzerland, Oct. 1996, pp. 371–374.

[24] L. Wang and A. Vincent, "*Joint rate control for multi-program video coding*", IEEE Trans. Consumer Electron., vol. 42, pp. 300–305, Aug. 1996.

[25] L. Li, W. Huang, I.Y.H. Gu, Q. Tian, "Statistical Modeling of Complex Backgrounds for Foreground Object Detection," IEEE Trans. Image Processing, vol. 13, no. 11, pp. 1459-1472, 2004.