การปรับปรุงการเข้ารหัสวีดิทัศน์ตรวจตราโดยตั้งอยู่บนมาตรฐาน IEEE 1857

นางสาวทิพเกสร บุญนาคม

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิศวกรรมศาสตรมหาบัณฑิต
สาขาวิชาวิศวกรรมไฟฟ้า ภาควิชาวิศวกรรมไฟฟ้า
คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย
ปีการศึกษา 2558

IMPROVEMENT OF SURVEILLANCE VIDEO CODING
BASED ON IEEE 1857 STANDARD

Miss Thipkesone Bounnakhom

A Thesis Submitted in Partial Fulfillment of the Requirements
for the Degree of Master of Engineering Program in Electrical Engineering
Department of Electrical Engineering
Faculty of Engineering
Chulalongkorn University
Academic Year 2015

Thesis Title	IMPROVEMENT OF SURVEILLANCE VIDEO CODING BASED ON IEEE 1857 STANDARD
By	Miss Thipkesone Bounnakhom
Field of Study	Electrical Engineering
Thesis Advisor	Assistant Professor Supavadee Aramvith, Ph.D.

Accepted by the Faculty of Engineering, Chulalongkorn University in Partial Fulfillment of the Requirements for the Master's Degree

....Dean of the Faculty of Engineering
(Professor Bundhit Eua-arporn, Ph.D.)

THESIS COMMITTEE

....Chairman
(Assistant Professor Widhyakorn Asdornwised, Ph.D.)

....Thesis Advisor
(Assistant Professor Supavadee Aramvith, Ph.D.)

....Examiner
(Assistant Professor Suree Pumrin, Ph.D.)

....External Examiner
(Narut Soontranon, Ph.D.)

ทิพเกสร บุญนาคม : การปรับปรุงการเข้ารหัสวีดิทัศน์ตรวจตราโดยตั้งอยู่บนมาตรฐาน IEEE 1857 (IMPROVEMENT OF SURVEILLANCE VIDEO CODING BASED ON IEEE 1857 STANDARD) อ.ที่ปรึกษาวิทยานิพนธ์หลัก: ผศ. ดร. สุภาวดี อร่ามวิทย์, 50 หน้า.

การเข้ารหัสวิดีโอ ตามมาตรฐาน IEEE 1857 ที่ถูกนำมาใช้เพื่อจุดมุ่งหมายที่จะทำให้อัตราการบีบอัดเพิ่มขึ้นสองเท่าให้มีประสิทธิภาพสูงและทำให้ความซับซ้อนน้อยลงสำหรับวีดิทัศน์เฝ้าระวัง อย่างไรก็ตามในการบีบอัดสัญญาณวีดิทัศน์ที่มีอัตราส่วนการอัดที่สูงจะมีผลต่อการนำไปใช้รู้จำสัญญาณวิดีทัศน์

ในงานวิจัยนี้เรานำเสนอวิธีการใหม่ในการปรับปรุงความคล้ายคลึงกันแบบ Scale Invariant Feature Transform (SIFT) โดยใช้การหาจุดที่มีลักษณะเด่น ซึ่งใช้การตรวจหาจุดที่มีลักษณะเด่นจากการตรวจหาแบบ SIFT ซึ่งขึ้นอยู่จุดที่มีลักษณะเด่นในมาโครบล็อกของเฟรมนั้นๆ โดยสามารถแบ่งอออกเป็นสองกลุ่ม คือ กลุ่มมาโครบล็อกที่มีความสำคัญ และ กลุ่มมาโครบล็อกที่มีความสำคัญน้อย ถ้ามาโครบล็อกปัจจุบันอยู่ในกลุ่ม มาโครบล็อกที่มีความสำคัญจะมีการเข้ารหัสโดยใช้กับพารามิเตอร์การควอนไทซ์ให้มีค่าน้อย เพื่อที่จะสามารถเก็บรักษาคุณลักษณะของข้อมูลในมาโครบล็อกหลังจากการบีบอัด สำหรับมาโครบล็อกอื่น ๆ จะถูกบีบอัดโดยการใช้ค่าพารามิเตอร์การควอนไทซ์ให้มีค่ามาก เพื่อที่จะสามารถรักษาประสิทธิภาพของข้อมูลโดยรวมให้คงอยู่

ในการทดสอบประสิทธิภาพของวิธีการที่นำเสนอเราได้ประเมินค่าสำหรับประสิทธิภาพของการทำงานของอัลกอริทึมที่นำเสนอนี้โดยการเปรียบเทียบการหาค่าความคล้ายคลึงกันแบบ SIFT กับซอฟต์แวร์อ้างอิงที่ใช้ในปัจจุบันภายใต้มาตรฐาน IEEE 1857 โดยปิดการใช้งานแบบจำลองพื้นหลังและนำเสนอขั้นตอนวิธีการที่จะปรับปรุงความถูกต้องให้มากขึ้น สำหรับผลการทดลองเห็นว่าวิธีการที่เราได้นำเสนอมานั้น สามารถปรับปรุง ค่าของความคล้ายคลึงกันแบบ SIFT ได้ถึง 49% เมื่อเข้ารหัสด้วยค่าอัตราบิตสูงและ พารามิเตอร์ quantization ต่ำ ใน MBs สำคัญที่เฉพาะเจาะจงเพื่อเก็บรักษาคุณลักษณะของข้อมูล ซึ่งแสดงให้เห็นว่าการทำงานของอัลกอริทึมที่นำเสนอนี้มีประสิทธิภาพที่ดีกว่าซอฟต์แวร์อ้างอิงในปัจจุบันในส่วนของการประเมินค่าของความคล้ายคลึงกันแบบ SIFT ในขณะที่อัตราส่วนค่าสัญญาณต่อค่ารบกวน และค่าอัตราบิตที่คงที่.

ภาควิชา	วิศวกรรมไฟฟ้า	ลายมือชื่อนิสิต	....
สาขาวิชา	วิศวกรรมไฟฟ้า	ลายมือชื่อ อ.ที่ปรึกษาหลัก	....
ปีการศึกษา	2558		

5670545121 : MAJOR ELECTRICAL ENGINEERING
KEYWORDS: IEEE 1857; VIDEO COMPRESSION; FEATURE PRESERVATION; SIFT; RECOGNITION;

THIPKESONE BOUNNAKHOM: IMPROVEMENT OF SURVEILLANCE VIDEO CODING BASED ON IEEE 1857 STANDARD. ADVISOR: ASST. PROF. SUPAVADEE ARAMVITH, Ph.D., 50 pp.

IEEE 1857 Video Coding standard is introduced with the aims to double compression rate for high efficiently and low complexity for video surveillances. However, to compress with high compression ratio is affected to recognition.

In this research, we propose a new algorithm to improve SIFT similarity by using feature map, which is detected keypoint features from SIFT detection. Based on the number of keypoints, macroblock (MB) of frames are divided into two groups: the MB which has more features uses as an important MB group and others for non-important MB group. If the current MB is in important MB group, that MB is encoded with lower quantization parameter (QP) so the features in that MB can preserve after compression. For other MBs, the high QP is used so the overall performance is maintained.

On the experiment, the effectiveness of the proposed algorithm is evaluated by comparing the SIFT similarity values between current reference software of IEEE 1857 standard with disable background modeling and proposed algorithm. According to the experimental results, our proposed algorithm can improve the SIFT Similarity values up to 49% when encoded with high target bits and Low QP in the specific important MBs based on to improve the MBs feature preservation, which is shown that the proposed algorithm outperforms to the current reference software in terms of improve SIFT similarity while maintain the same PSNR and bitrate values.

Department:	Electrical Engineering	Student's Signature	....
Field of Study:	Electrical Engineering	Advisor's Signature	....
Academic Year:	2015		

ACKNOWLEDGEMENTS

First of all, I would like to express my sincerest gratitude and thanks to my supervisor Asst. Prof. Dr. Supavadee Aramvith for her tireless efforts, encouragement, support, constant guidance and reliable advice during two years guidance throughout the entire course of my study.

My thanks also goes to members of the committee, Asst. Prof. Dr. Widhyakorn Asdornwised, Asst. Prof. Dr. Suree Pumrin , and Dr. Narut Soontranon for their time, critical review ad advice for this work.

Special thanks go to all the members of Video Processing Research Group for their assistance and companionship especially Mr. Htoo Maung Maung for his effort to explain and guiding me while developing some of the contributions described in this thesis without his friendly support I could not have completed this work. Thanks to Ms. Kanokphan for give many suggestions and others help to me, thanks to the others members especially Mr. Watchara for encourage of my thesis work and help me to process all the works which is related to my research.

I would like to express my heartfelt thanks to all my colleagues at the Department of Information Technology and Computer Engineering, Faculty of Engineering, University of Laos, and also Lao student Association in Chulalongkorn University Group for all supports.

Thanks to AUN/SEED-NET scholarship for giving me the opportunity of pursuing Master degree in Digital Signal Processing Research Laboratory (DSPRL), Department of Electrical Engineering, Chulalongkorn University and also to the staff in ISE and AUN//SEED-Net: Ms. Benjaporn Raiseangphor and Mrs. Kunaporn Phuntunilfor for their supports and advices.

Finally, I must thank to my family in Laos for their all supports, encourage and motivate to me when I am felling tried.

This research has been supported in part by the Collaborative Research Project of JICA Project for AUN/SEED-NET, Japan.

CONTENTS

Page

THAI ABSTRACT iv

ENGLISH ABSTRACT v

ACKNOWLEDGEMENTS vi

CONTENTS vii

LIST OF FIGURES 1

LIST OF TABLES 3

CHAPTER 1 INTRODUCTION 4

1.1. Motivations and Significance of the Research Problems. 4

1.2. Research contributions. 6

1.3. Objectives 7

1.4. Scope of the thesis 7

1.5. Expected Outputs 7

1.6. Research Procedures 7

1.7. Thesis Organization 8

CHAPTER 2 BACKGROUND AND LITERATURE REVIEWS 9

2.1. Basic of Video Coding Standard. 9

2.2. Brief Review of Block-Based Hybrid Video Coding. 10

2.3. Overview of IEEE 1857 standard 11

2.4. Comparison between H.264/AVC and IEEE 1857 standards. 14

2.5. Background Modeling 16

2.6. The Importance of Rate Control 17

2.7. Rate Control 18

2.7.1. GOP Level Rate Control 19

2.7.2. Frame Level Rate Control 20

2.7.3 Basic Unit Level Rate Control 21

2.8. Scale Invariant Feature Transform (SIFT) 21

2.9. Review of Related Works 23

2.9.1. Background Model 23

Page

2.9.2. Feature Preservation 24

2.9.3. Region of Interests (ROI) 24

CHAPTER 3 PROPOSED METHOD 26

3.1. Overall Block Diagram of Proposed Algorithms 26

3.2. Feature Analysis and MB Categorizing 27

3.3. MB Level for QP Adjustment 31

CHAPTER 4 EXPERIMENTS AND RESULTS 32

4.1. Initial the Experiments 32

4.1.1. SIFT Similarity measurement 33

4.1.2. PSNR measurement 34

4.2. Experimental Results 34

4.3. Result of Analysis 38

4.3.1. Comparisons in SIFT Similarity with QP 38

4.3.2. Comparisons in SIFT Similarity with Generated Bitrate 40

CHAPTER 5 CONCLUSION AND FUTURE WORKS 42

5.1. Summary and Conclusion 42

5.2. Future work 42

REFERENCES 43

VITA 50

LIST OF FIGURES

Figure 1.1The Problems of SIFT Similarity on Compression (a) Original video (b) compressed Video 5

Figure 1.2 Effect of QP increasing to the SIFT similarity for difference resolutions ...5

Figure 2.1 Block diagram of a typical block based hybrid coder [5] 10

Figure 2.2 The architecture of IEEE 1857 Standard with background model based coding framework [2] 12

Figure 2.3 Framework of Background modeling...16

Figure 2.4 Illustration of Rate control for Video Encoder[12] 17

Figure 2.5 Rate and Distortion relationship of hybrid video encoders[12] 18

Figure 2.6 Show the Scale-space extrema detection [13]...22

Figure 2.7 The Keypoint descritor from SIFT[13] 23

Figure 3.1 The process of conducting proposed method. 26

Figure 3.2 The Flowchart of proposed algorithm 27

Figure 3.3 Detect Keypoint features by SIFT detection 28

Figure 3.4 Analysis the Feature extraction in compressed video 28

Figure 3.5 SIFT features of *Crossroad* and *Classover* sequences. 29

Figure 3.6 The average number of features MB for two sequences, (a) for crossroad sequence and (b) for Intersection sequence 30

Figure 4.1 Video Test Squences from PKU-SVD-A dataset......................................33

Figure 4.2 The improvement of SIFT Similarity for Crossroad 256kbps with frame number 150 to 250 35

Figure 4.3 The improvement of SIFT Similarity for Classover 512kbps with frame number 150 to 250...35

Figure 4.4 The improvement of SIFT Similarity for Intersection 512kbps with frame number 150 to 250...36

Figure 4.5 The improvement of SIFT Similarity for Mainroad 512kbps with frame number 300 to 400 36

Figure 4.6 Achieve of SIFT Similarity values based using QP to compress Crossroad 256kbps video sequence between RM09 and Proposed method....39

Figure 4.7 Achieve of PSNR values based on improve SIFT Similarity to compress Crossroad 256kbps video sequence between RM09 and Proposed method....39

Figure 4.8 The comparison using QP to compress Crossroad 256kbps between RM09 and Proposed method....40

Figure 4.9 The distortion of compress video frame for Crossroad 256kbps from frame number 1485 between RM09 and Proposed method....40

Figure 5.1 The comparison of improvement SIFT Similarity for Crossroad sequence with 256 Kbps....46

Figure 5.2 The improvement of SIFT Similarity with QP for Crossroad sequence with 256 Kbps....47

Figure 5.3 The improvement of PSNR with QP for Crossroad sequence with 256 Kbps....48

Figure 5.4 The improvement of SIFT Similarity with Generated bit for Crossroad sequence with 256 Kbps....49

LIST OF TABLES

Table 2.1 Differentials of H.264 and IEEE 1857 Standards......................................15

Table 4.1 Test Sequences parameter..32

Table 4.2 Result of comparison PSNR, Bitrate and SIFT Similarity between RM09 and Proposed algorithm for Crossroad sequence.......................................37

Table 4.3 Result of comparison PSNR, Bitrate and SIFT Similarity between RM09 and Proposed algorithm for *Classover* sequence.......................................37

Table 4.4 Result of comparison PSNR, Bitrate and SIFT Similarity between RM09 and Proposed algorithm for Overbridge sequence....................................37

Table 4.5 Result of comparison PSNR, Bitrate and SIFT Similarity between RM09 and Proposed algorithm for Intersection sequence. 37

Table 4.6 Result of comparison PSNR, Bitrate and SIFT Similarity between RM09 and Proposed algorithm for Mainroad sequence. 38

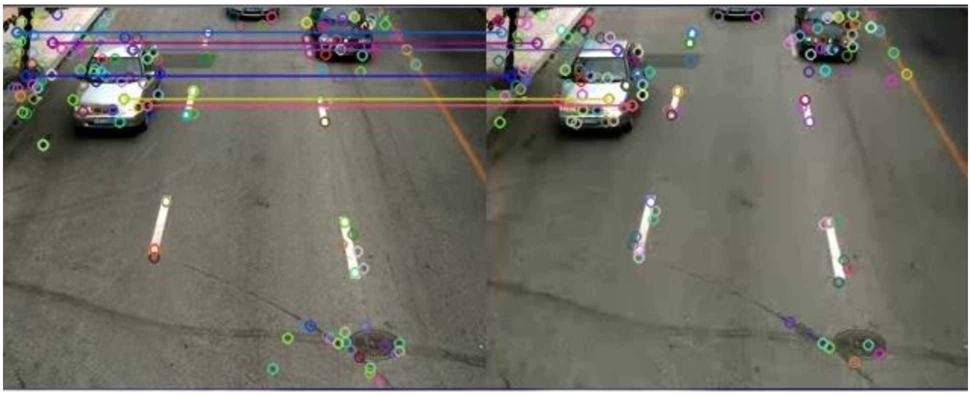
CHAPTER 1
INTRODUCTION

1.1. Motivations and Significance of the Research Problems.

Nowadays, video surveillance is commonly used in several surveillance systems for public security such as: traffic controlling, home monitoring, office monitoring, and so on. Every huge amount of video data is being stored and transmitted in surveillance systems. To reduce the load of data, video coding standard plays on the important role to compress of these huge surveillance data. The traditional video coding standards such as H.264/AVC[1] are originally designed for generic videos, which is widely used in surveillance video compression. In general, the main characteristic of video surveillance is using stationary cameras which locate at a certain position. For good quality of capture videos or image data is used high camera resolutions. Consequently, it produces a large amount of data that influence on transmission and storage capacity. However, to use high compression rate to compress the video, it will effect in difficulty of visual feature extraction and decrease ability of video analysis especially in video recognition.

To evaluate the effect of recognition in video compression by using high compression rate, we compress video with different Quantization Parameters (QPs) with several resolutions. In this case, SIFT Similarity is used to compare the keypoints matching between the original frame with the compression frame as you can see in the Figure. 1.1 Some keypoints in the compress frame are missing cause of distortion. Generally, The QP determines spatial detail of compression when we compress with low QP, almost detail is remained; otherwise using high QP some detail is loss of quality because of distortion as shown in the Figure 1.2.

To address this issue, to make the better of characteristics in surveillance videos codec has been considered. So, IEEE 1857 standard[2] are developed for making double compression rate comparing to H.264/AVC with same complexity and improved the compression efficiency by utilize the characteristics on surveillance video coding that included of two features such as: Recognition-friendliness, High-efficiency and low-complexity video coding technologies.



(a) (b)

Figure 1.1The Problems of SIFT Similarity on Compression

(a) Original video (b) compressed Video

This standard is introduced background modeling technique which can reduce bitrate significantly for the sequence with constant background. So that IEEE 1857 standard is capable used for high compression and be able to add some techniques to improve the coding standard base on achieve better accuracy of recognition especially SIFT Similarity.

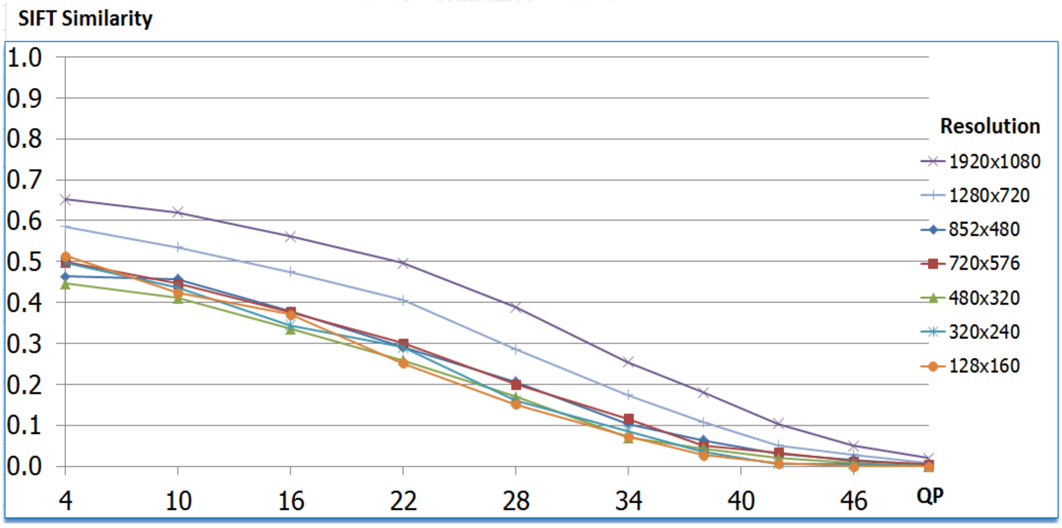


Figure 1.2 Effect of QP increasing to the SIFT similarity for difference resolutions

As in many other computer vision applications, feature extraction plays an important role in analyzing the surveillance videos effectively. So, preserving the feature information becomes equally or even more important than optimizing the visual quality especially for surveillance application. However, due to the high compression of video coding standards, feature extraction performance of compressed video is affected and consequently, the accuracy of recognition is reduced.

There are some previous works on feature preserving image or video compression techniques. M. Maker et al. [3] proposed the compression of relevant local patches after feature extraction. In this approach, the canonical images patches are compressed and transmit these patches as side information to decoder for further processing. But this method is not a standard compatible one so the decoder needs to be modified. Jianshu Chao et al. [4] proposed SIFT feature-preserving image compression for JPEG standard. Two bit-allocation methods are proposed to control encoding process for the most important and relevant features are preserved.

In IEEE 1857 standard is mainly focused on reduce the bitrate by using background model and improved the recognition performance by encoding region-of-interest with intra mode. However, using excessive intra coded blocks can reduce quality especially encoding under rate constraint. This can also affect the recognition performance of compress videos. To address this problem, we propose a simple yet effective algorithm to preserve features in compress video and also to improve the recognition performance.

1.2. Research contributions.

The main contribution in this research, we mainly focus on improving recognition performance of surveillance videos by propose a new algorithm to improve SIFT similarity by using feature map, which is detected keypoint features from SIFT detection. Based on the number of keypoints, macroblock (MB) of frames are divided into two groups: the MB which has more features uses as an important MB group and others for non-important MB group. If the current MB is in important MB group, that MB is encoded with lower quantization parameter (QP) so the features in that MB can preserve after compression. For other MBs, the high QP is used so the overall performance is maintained.

The effectiveness of the proposed algorithm is evaluated by comparing the SIFT similarity values between current reference software of IEEE 1857 and our proposed algorithm. According to the experimental results, the performance of proposed algorithm compare with RM09 for the current reference software in IEEE 1857 Video Coding Standard, in terms of SIFT Similarity value for our proposed algorithm outperforms the current reference software to improve the better recognition on SIFT Similarity values while maintain the same PSNR and bitrate vales.

1.3. Objectives

- To propose algorithm for adjust Quantization Parameter (QP) on the encoder process in IEEE 1857 standard based on improve video quality for feature preservation.
- Evaluate the performance of the proposed algorithm to improve the accuracy recognition for SIFT Similarity.

1.4. Scope of the thesis

- Propose an algorithm to improve both video quality and SIFT similarity while maintaining the PSNR and target bit rate in terms of disable background modeling.
- Evaluate the performance of the proposed algorithm by comparing with reference software for original encoder in order to improve the accuracy recognition for SIFT Similarity.

1.5. Expected Outputs

- Understanding the new features such as background model based coding and rate control of IEEE 1857 Standard.
- Achieving better video quality in terms of SIFT Similarity and recognition friendliness using SIFT detection.

1.6. Research Procedures

- Reviewing literatures about IEEE1857 standard and related video coding methods
- Testing the performance of the IEEE 1857 reference software.
- Finding some techniques to propose an algorithm by modified coding.
- Downloading datasets from the provider to prepare for conduct experiment.
- Summarizing the results, and comparing with previous researches
- Proposal examination
- Submitting papers and conference
- Writing thesis
- Thesis defense

1.7. Thesis Organization

This thesis is organized into five main chapters including this chapter. The following paragraphs provide brief description of the remaining chapter of this thesis.

Chapter 2 we introduce some basic knowledge of video coding standard concepts, and brief review for block based hybrid video coding, overview of IEEE 1857standard which is use in this research based on main feature of IEEE 1857 and also summaries literature review of related works.

Chapter 3 presents the overall diagram of proposed algorithm and describe how to get feature map and use as an input to improve the preservation feature in IEEE 1857 standard and some idea to adjust QP.

Chapter 4 presents the experimental and results of proposed method compared to current reference software.

Chapter 5 concludes the conclusions and future works of the research.

CHAPTER 2
BACKGROUND AND LITERATURE REVIEWS

In this chapter, we introduce some basic background of video coding standard and the benefits of using compression, brief overview of Block based hybrid video coding, and overview of IEEE 1857 video coding standard are introducing in this chapter to provide some basic understanding of the concepts and summarize for related work in this research.

2.1. Basic of Video Coding Standard.

Normally, for the original video sequences or uncompressed video names" Raw video" are required the large number of bits. To compact the video data is used video coding to compress the video into small data size and number of bits (bit stream) which are necessary for transmission and practical storage of digital video.

There are included of two techniques for compressions are Lossless and Lossy compression techniques. Lossless compression is used to reduce image or video data while remaining the quality of the original images and the decoded image quality is required to be identical to the image quality prior to encoding. In contrast, for Lossy compression is use to compress data to meet a given target bit-rate for storage and transmission. In these technique uses high video compression by degrading the video quality is reduced compared to the quality of the original images prior to encoding.

Video compression is very useful to reduce the data resource usage by using elimination the redundancy in the video data then achieve small data storage space and improve the transmission capacity. To eliminate redundancy data, there are included of two types in video compression such as spatial and temporal redundancies. For spatial redundancy to remove a spatial redundancy is looked within a frame and refer the correlation that present in different parts of frame as Intra Coding. For temporal redundancy to remove the redundancies, it will look between frames at a sufficiently high frame rate in similar successive frames in the video sequence.

2.2. Brief Review of Block-Based Hybrid Video Coding.

The main goal of video coding technology is that of increasing as much as possible the compression ratio between the size of the original data and the size of the bitstream. A video coding standard typically consists of a set of specifications which regulate the steps necessary to decode a bitstream into a reconstructed sequence ready for consumption. Most standards are developed by experts in the field working together with the goal of providing the best technology to achieve compression under a wide range of conditions and to address the relevant industry needs. In most cases only the decoder (and the corresponding bitstream) is defined in the specifications. Each manufacturer or service provider can independently develop a suitable encoder as long as it produces compliant bitstreams. The standard process is essential to ensure that the best compression technologies can be quickly and widely adopted by the industry. The encoder and decoder framework is of fundamental importance for video coding especially in case the compressed signals need to be shared, for instance in content distribution or storage in physical media. Clearly the users that receive the compressed signal should be able to decode it. For this reason, and in order to avoid the coexistence of an endless variety of compression schemes each requiring a different decoder, many efforts in developing video coding technology converge in the definition and ratification of video coding standards.

The traditional hybrid video coding frame work are including of two main parts such as: Encoder and Decoder as shown in the Figure 2.1.

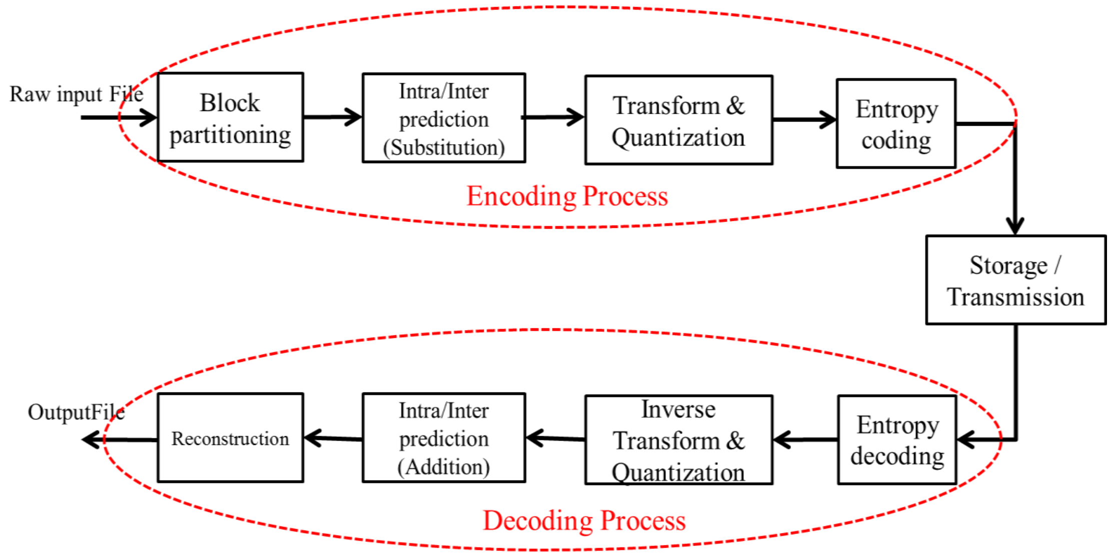


Figure 2.1 Block diagram of a typical block based hybrid coder [5]

A block-based hybrid video encoder is included of four steps:

- Block partitioning
- Intra/Inter prediction(Substitution)
- Transform and Quantization coding
- Entropy coding.

For the Encoding process, first, every raw pictures are divided into several small square blocks by the common size of 16x16.These are the basic unit of blocks for the latter encoding process. Then carried out for each block to get the prediction signal on the intra or inter prediction, the residual signal for each blocks are applied DCT (discrete-cosine transform) transform to get frequency coefficients will obtain by subtracting the prediction signal from the original signal. After that the coefficients are quantized and entropy coded to get the output bit stream.

A block-based hybrid video decoder is reversed process which including of:

- Entropy decoding.
- Inverse Transform and De-quantization coding
- Intra/Inter prediction(Addition)
- Reconstruction

For the decoder simply as applies the reverse process to get the reconstructed output. Inter prediction or motion compensated prediction is one of the key components in hybrid codec. On the encoder process to do the inter prediction, the reconstructed version of previously encoded pictures will store in the buffer is called decoded picture buffer (DPB). These pictures are used as reference pictures for motion estimation of the current block to be encoded. Motion estimation produce motion vectors of the current block, there are the important information with the coefficient values which is encode and send to the decoder together. The decoder also has a DPB in to store the reference pictures which are the same as encoder to get the correct output videos.

2.3. Overview of IEEE 1857 standard

IEEE 1857 Video coding Standard is originated from Advanced Audio and Video Coding Standard (AVS) [6], which was established in year 2006 and approved on March 2013. This standard is designed in an application-oriented way especially for video surveillance by introducing many efficient tools. Among many profiles such as Main Group, Portable Group, Enhance Group, Broadcasting Group, Surveillance

Baseline and Surveillance Groups; Surveillance baseline profile and surveillance profile are specifically designed for video surveillance applications. Since the main interest of this research is to enhance the recognition friendliness of surveillance videos, the above mentioned two profiles are the most interesting profiles. For this work, surveillance baseline profile is selected to use because this profile includes same coding tools as surveillance profile except the background based coding method. The coding tools include background picture prediction, background modeling, core frame coding, and region-of- interest based coding, and so on.

Compared with several video coding standards as H.264/AVC [7] and H.265/HEVC[8] standards, for the most important contribution in IEEE 1857 standard is integrated the model-based coding with traditional hybrid entropy coding framework as you can see in the Figure 2.2 .

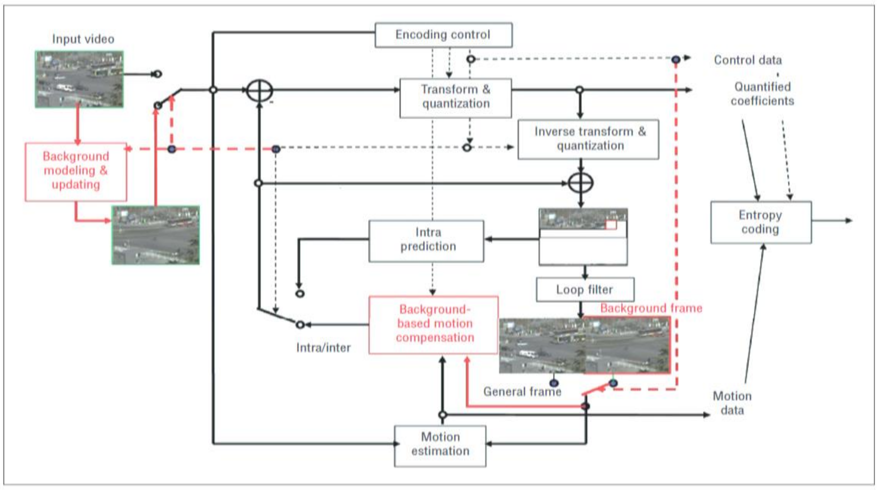


Figure 2.2 The architecture of IEEE 1857 Standard with background model based coding framework [2]

In the theory, video sequences are consisting of several pictures; each picture is split into small MB called basic unit which is contains the samples of luminance component and their associated chrominance as a size of 16x16. To predict the video data on predictive coding process, it will perform for each MB which can be categorized into either intra-frame prediction or inter-frame prediction. After that, predict on the blocks of residues signal between difference between original pixel in the current image and predicted pixel image. Then, transform all coefficients and quantize before process to entropy coding, and finally the entropy coded all information in MBs into the bitstream as an output format.

IEEE 1857 video coding standard is the first standard is designed on the aims to achieve double compression rate for high efficiency and low complexity to reduce the scene redundancy (Background Scene) in the video surveillance systems.

These are the key techniques feature coding efficiency for IEEE 1857 standard.

- Background Modeling and Updating

In IEEE 1857 there are three different types of background modeling algorithm are supported such as: Background modeling using running average algorithm [9], Background modeling using median algorithm [10], weight based running average algorithm[11].

- Background Based Motion compensation

In IEEE 1857 there are five different types of motion compensation modes are supported: Forward, backward, symmetry based bi-predictive, skip and direct. On the symmetric prediction are coded only forward motion vector in each block and then backward motion vectors can be derived from this motion as a forward vector. The direct mode used forward motion vector of current block and motion vector of the collocated block to solve the problem by spatial direct mode technique.

- Entropy coding

In IEEE 1857 Video Coding Standard, there are included of two methods for entropy coding such as Two-dimensional variable length coding (2DVLC) for coding the transformed residuals, and the other one is improved Arithmetic Entropy coding (AEC) for two of these methods provides both higher coding efficiency and low computational complexity.

- Transform and Quantization

The residual samples are computed as the difference between original MB and prediction before being transformed. The transform is not computed for the entire 16x16 samples at once, but instead the MB is split in 4x4 blocks that are independently transformed. To limit the complexity of this additional step, Hadamard transform is used instead of DCT. After transform, the coefficients are quantized. The quantize step is determined by the value of a quantization parameter (QP), which may be set at a sequence, frame or MB level depending on the coding configuration. High values of the QP correspond to low expected quality of the decoded signal. Finally, the quantized coefficients are input to the entropy coder. On this standard are utilized the integer transforms and Adaptive block- size transforms by using 4x4, 8x8, 16x16 blocks transforms. On the 4x4 blocks transform will transform only 4x4 blocks, while 8x8 blocks can be used to guide the adaptively by apply both 4x4 and 8x8 block transform in the current blocks.

- Intra prediction

In IEEE 1857, Intra-frame prediction (intra-prediction) is used to decode information in the current frame and predict only MB within the same stage as the reference prediction, exploit the statistically for spatial dependence between the pixel in a frame. To reduce bits average is based on neighboring blocks for 8x8 intra prediction modes and to achieve better coding efficiency is predicted by using 4x4 Intra-prediction which is focus to predict only smaller block size.

- Motion estimation

For the motion estimation unit, original frame from input image is accepted on the buffers and the reconstruct frames are able to code from forward- backward reference which store in the encoder. For this estimation is used to produces motion vector by using motion compensation of forward prediction or interpolate in the current frame which is used for transform by first predictive encoder then uses in the entropy coding.

- Loop Filter

To produces blocking artefacts especially at low bit rates similar as H.264/AVC standard is encoded by using Block-based video coding. In IEEE 1857 standard is implemented as loop adaptively in the de-blocking filter to improve the decoded visual quality. On the boundary between two 8×8 Luma or Chroma blocks are accepted which is located in portion of boundaries or in the image boundaries. For the edge of image for boundary strength (Bs) is derived based on the MB type, motion vector, and reference index of different filters will be considered.

2.4. Comparison between H.264/AVC and IEEE 1857 standards.

The difference of both H.264/AVC and IEEE 1857 video coding standard is background modeling. For the background frame is will be updated based motion compensation module. For the advantage of these two modules can increases the compression rate up to two times comparing to H.264/AVC with additional encoding complexity similar as H.265/HEVC as show in the Table 2.1.

For the main contribution of IEEE 1857 compared with H.264/AVC is background model based coding framework and these are the characteristic of IIEEE 1857 standard.

1. Background model based coding framework is including of background and updating model, background based motion compensation and background buffer.
2. Low- complexity background modeling is used a segment and weight based running average (SWRA) to approximate the background by assign the larger weight values in the averaging process to get the model background.
3. G-Pictures based background prediction: G-picture can be quantified with smaller QP and encoded as non-display frame to realize the better efficiency of background pixels in the current frame.

Table 2.1 Differentials of H.264 and IEEE 1857 Standards

Coding Tools	**IEEE 1857**	**H.264**
Entropy Coding	AEC and UVLC	CAVLC and CABAC
Transform & QP	8x8 integer transform	4x4, 8x8 integer transform
Intra Prediction	Adaptive intra-prediction by 4x4 intra-predictions and can be applied along 8x8 intra-predictions.	Two different types of intra prediction are possible 4×4 and 16×16 intra-predictions
Motion Estimation	Variable block size 4x4,8x8, 16x16	
Loop Filter	The filter is 8×8 block based	The filter is 4×4 block based
Background modeling	Yes	No

4. Optional difference coding for mixed macroblocks (MBs): The difference between the current MB and its corresponding background can be predicted by using data in the current reference frame and G pictures.
5. Improved motion vector prediction: To derive the predicted motion vector when using G-picture as direct or in direct prediction reference.
6. Improve BBV buffer management: To deal with G-picture so as support to no delay, no frame drop and real time encoding.

7. Error resilience coding tools: IEEE 1857 standard to get some errors resilience is including of: flexible slice set, core picture coding, the constrained DC intra prediction, the non-reference P-picture, the optimized motion vector which is scaling for interlaced video coding and adoptive weighting quantization.

2.5. Background Modeling

Background modeling methods is used to classify the background and foreground frames. In this method is category into two groups: a parametric method such as Gaussian mixture model (GMM) is used to represent a weighted sum of Gaussian component densities and commonly used as continuous measurements on features in a biometric systems and nonparametric methods such as Bayesian modeling, kernel density estimation, temporal median filter, mean-shift are all very useful in a wide variety of settings, including of density estimation, clustering, and classification. This is the works flow of Background Modeling based coding framework as shown in Figure 2.3.

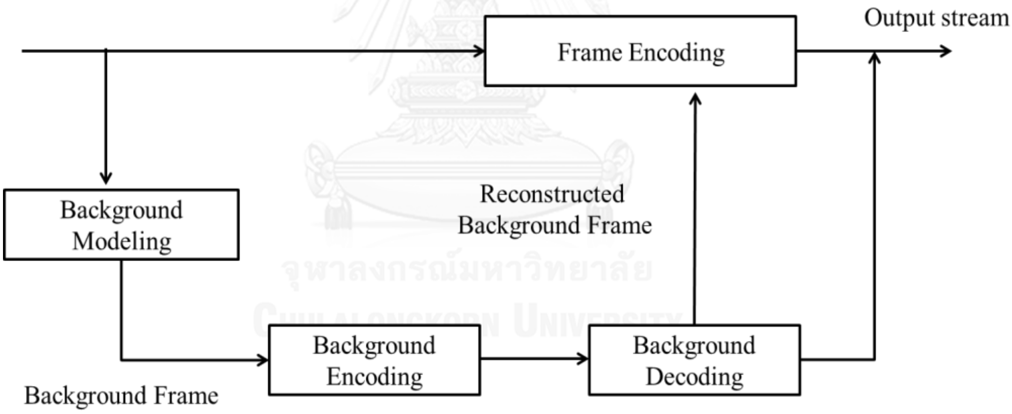


Figure 2.3 Framework of Background modeling

Background modeling method is generally designed to improve on analysis of video content as object detection. On the aim to using this method is to make the best use of capture the static properties of a scene as foreground scene which is easier to detect. However, for the different of situation in surveillance video, background model need to be high-quality of video coding standard to encode with large amount of bits only one time on the process of update in a long period. In this method is very useful for surveillance video coding that include of important feature which is requirements for video codec such as periodically updating by using small prediction residual error variance with low computational and low complexity in small memory and especially for the process of encoding is should make for no delay modeling.

The background picture prediction is defined to additional achievement to remove background noise, temporal redundancy, and facilitate to use object segmentation and motion detection to generate video events in video surveillances. In IEEE 1857 standard is presented the special coded as I-picture and reconstructed background picture will be stored in the separable memory. For a special P-picture is coded being predictable from the reconstructed background picture only. In the background picture prediction process, start with generate the background frame by using background modeling procedure to encode with small Quantization parameter based on retrain the quality of background frame. In the other hand, to generate a foreground frame is used the difference of reconstructed background frame and encoded in the encoder. Finally, on the encoding process, to encode background and foreground frames are written into the bit-stream. In decoding process, to reconstruct the video sequence is decoded background and difference frames to get the output video by using the difference compensation module.

2.6. The Importance of Rate Control

A rate control algorithm is used to dynamically adjusts parameters to achieve a target bitrate [12] in the encoder process. To allocate a bit budget of each group of pictures, individual picture and/or sub-picture in a video sequence is controlled by approximate the complexity of frames and adjust QP parameter to achieve the desired rate. The QP must be dynamically adjusted such that the bitrate and demanded rate and the same as shown in Figure 2.4

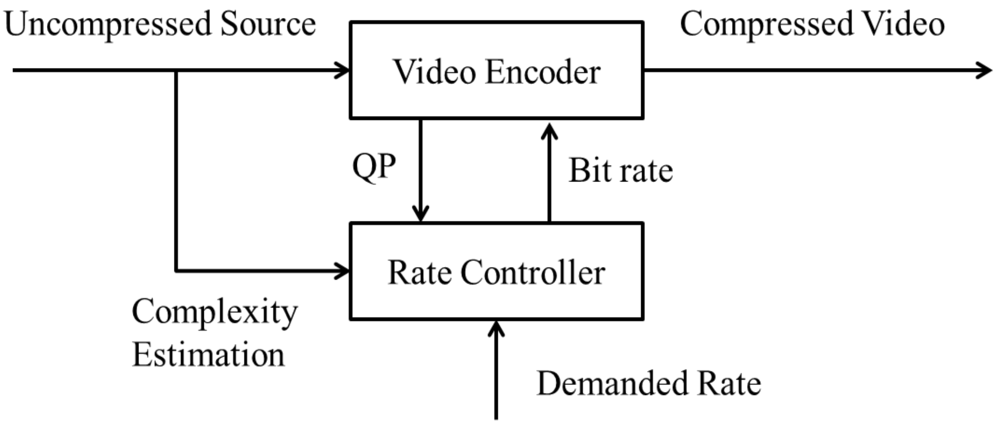


Figure 2.4 Illustration of Rate control for Video Encoder[12]

To prevent the over flow or underflow buffer, and achieved good video quality, a rate control scheme is applied to dynamically adjust the encoder parameter in order to achieve the target bit rate. One method to control the bitrate is to vary the degree of quantization for each frame, which is a very challenge in rate control. To determine a combination of suitable encoding parameters, such as quantization step size, macroblock mode, motion search, etc… which can help to achieved the best video quality and at the same time satisfy the network and application constants.

In particular, the quantization parameter (QP) regulates the spatial details when QP is very small almost the details are retained and missed when the QP is very high that caused to bit rate is dropped and loss the quality. The rate and distortion relationship of a particular input picture is illustration in Figure 2.5. To achieve lower bit rate is used to increasing the number of QP at the cost of increasing distortion and then rate control algorithm also take into consideration the buffer fullness, the available channel and the video frame complexity in the determining QP.

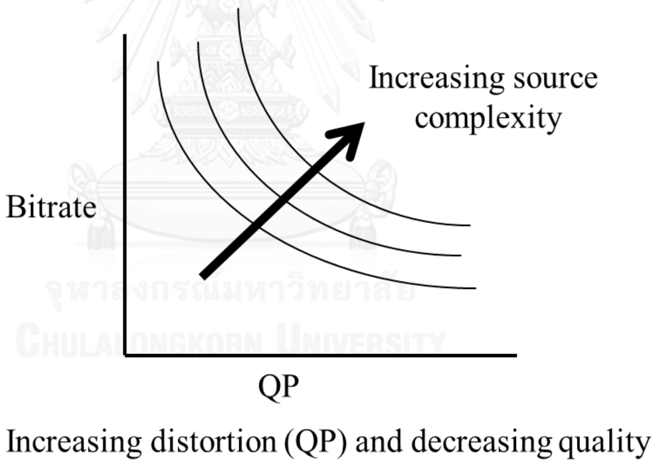


Figure 2.5 Rate and Distortion relationship of hybrid video encoders[12]

2.7. Rate Control

The IEEE 1857 rate control algorithm implements a linear prediction model to predict the Mean Absolute Differences (MADs) of the remaining basic units in the current frame by co-located basic units in the previous, then compute target bits by fluid flow traffic model and derived two values for considered by hypothetical reference decoder (HRD). For the corresponding quantization parameter is computed by quadratic ate-distortion model then applied rate distortion optimization (RDO) to perform all MBs in current basic unit by quantization parameter.

There are basically three levels of rate control in IEEE 1857 standard such as: GOP level rate control, Frame level rate control and Basic unit level rate control.
To compute the total numbers of basic units in one frame is defined by equation:

$$N_{\text{Total numbers of basic units}} = \frac{N_{\text{Total numbers of frame}}}{N_{\text{Basic unit}}} \tag{1}$$

- A linear MAD prediction model is proposed to predict the MAD of current macroblock in the current frame that defined by equation:

$$MAD_C = a_1 * MAD_P + a_2 \tag{2}$$

where, MAD_C is predicted MAD of current MB in a current frame, MAD_P is the actual MAD of MB in the co-located position of previous frame, a_1 and a_2 are the coefficients of prediction model will be updated after coding each basic unit.

According to aforementioned analyses, the prediction model based on frame MAD is more precise than that based on macroblock MAD. Therefore, the following problem is how to establish a new quadratic R-D model based on frame MAD instead of on macroblock MAD correspondingly. For that, we must analyze firstly the quadratic R-D model in encoder described as follows:

- A quadratic rate-distortion (R-D) model is used to compute the quantization parameter of current macroblock, the quantization step corresponding to the target bits which computed by the equation:

$$T_i = c_1 * \frac{MAD_C}{Q_{step,i}} + c_2 * \frac{MAD_c}{Q^2_{step,i}} - m_{h,i} \tag{3}$$

where, $m_{h,i}$ is the total number of header bits and motion vector bits, c_1 and c_2 are two coefficients of the quadratic model, and $Q_{step,i}$ is the quantization step for the i^{th} basic unit.

2.7.1. GOP Level Rate Control

Group of Pictures (GOP) is allocated bit budgets based on target bit rate and the total number of frames in that group. For three difference frame type (I, P or B) of images are used to selective to allocates bits before encoding each frame in that GOP and also estimate the complexity of measurement, and current bits remaining number of GOP. To compute the total number of bit allocation for the current GOP by the equation:

$$B_{gop}(1) = \frac{R(1)}{f} * N_{gop} + \mathrm{B}_{gop-1}(\mathrm{N}_{gop-1}) \tag{4}$$

where, N_{gop} is the number of GOP and R is the instant available bit rate.

For the initial quantization parameter is will decided in the GOP level rate control. To predefine the number of quantization parameter in the first GOP is specified for motion estimation/mode decision and use as the initial quantization parameter

$$QP_i = \frac{\sum QP(i-1)}{1 + N_{i-1}^{p}} \tag{5}$$

where, $\sum QP(i-1)$ is the sum average QP for all I/P pictures in the (i-1)th GOP, and N_{i-1}^{p} is the total number of P pictures in the (i-1)th GOP.

2.7.2. Frame Level Rate Control

Frame level rate control or Picture level rate control allocates target bits for each picture based on the remaining bits, picture's complexity and virtual buffer's occupancy. Getting the target bits and MAD of current picture, the quantization parameter can be obtained based on the R-D model.

- Target bit rate computation: To determine the target bits to allocated for the *j*th frame in the *i*th GOP are based on the level of target buffer, the number of frame rate, the channel bandwidth and the actual buffer occupancy.

$$\tilde{f}(n_{i,j}) = \frac{u(n_{i,j})}{F_r} + \gamma(\mathrm{Tbl}(\mathrm{n}_{i,j}) - \mathrm{B}_c(\mathrm{n}_{i,j})) \tag{6}$$

where, γ is a constant in typical value 0.75 when there is no B frame and 0.25 otherwise.

- To compute the quantization parameter and perform RDO: The MAD of current P frame is predicted by using the actual MAD of previous P frame. The quantization parameter $\tilde{Q}_{pc}$ corresponding to the target bit is then computed by using the quadratic model and the final quantization parameter $\tilde{Q}_{pc}$ is further bounded by:

$$\tilde{Q}_{pc} = \min\{51, \max\{\tilde{Q}_{pc}, 1\}\} \tag{7}$$

2.7.3 Basic Unit Level Rate Control

For basic unit is used to define as continuous MBs of group. To achieve a trade-off between the whole coding efficiency and bits fluctuation, it is similar to the picture level rate control, which include of MAD prediction, bit allocation, and the desired quantization parameter. For the initial of first basic unit in the current picture, the QP can be derived from average QP of all basic units in the previously coded picture.

$$QP_j = \alpha \times \frac{\sum QP(\mathrm{j}-1)}{N_{unit}} \tag{8}$$

where, α is a constant and N_{unit} is the number of basic unit in the picture.

2.8. Scale Invariant Feature Transform (SIFT)

Scale invariant feature transform (SIFT) algorithm [13] is introduced by D. Lowe, which is widely used in robust local feature detection for scale and rotation descriptor. This algorithm detects distinctive invariant features from scale invariant keypoints, which is using to extract and compute its descriptors. To extract the features by using SIFT algorithm are applied four stage filtering approach such as: Scale-space extrema detection, keypoints localization, orientation assignment, and keypoint descriptor.

- Scale-space extrema detection: to identify possible invariant to scale and orientation of interest points by using scale-space extreme in the image resulting from the Gaussian function convolved with the input image.

$$\mathrm{L}(\mathrm{x}, \mathrm{y}, \partial) = \mathrm{G}(\mathrm{x}, \mathrm{y}, \partial) * \mathrm{I}(\mathrm{x}, \mathrm{y}) \tag{9}$$

$$G(\mathrm{x}, \mathrm{y}, \partial) = \frac{1}{2\pi\partial^2} e^{-((x^2+y^2)/2\partial^2)} \tag{10}$$

where, L(x, y, ∂) is an image, G(x, y, ∂) is the result from the Gaussian function and I(x, y). is an input image.

By using scale-space extrema in the difference-of-Gaussian (DoG) function convolved with the image for the efficiently detection stable on the keypoint locations in scale space. D(x, y, ∂) which can be computed from the difference of two nearby scales separated by a constant multiplicative factor k:

$$\begin{aligned} D(x, y, \partial) &= \big(G(x, y, k\partial) - G(x, y, k\partial)\big) * I(x, y) \\ &= L(x, y, k\partial) - L(x, y, \partial) \end{aligned} \tag{11}$$

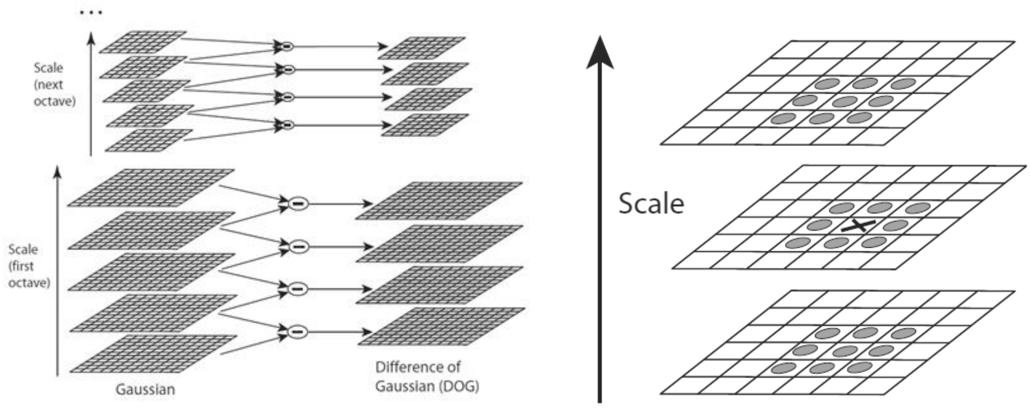


Figure 2.6 Show the Scale-space extrema detection [13]

- Key point localization: to determine location and scale, a keypoint candidate has been found by comparing a pixel to its neighbors, then to perform a detailed fit to the nearby data for location, scale, and ratio of principal curvatures. This information allows points to remove keypoint with low contrast and edges

Orientation assignment: For every keypoint location will be assigned orientations based on local image gradient directions. The future operations perform on the data image by transform and provide the invariance relative to an orientation, scale, and location for each keypoint features. To do the scale of keypoint, the Gaussian smoothed image L with closet scale is used to compute and perform the scale-invariant features manner. For compute the orientation is calculated by pixel differences [13] .

$$m(x,y) = \sqrt{(L(x+1,y)-(L(x-1,\mathrm{y}))^2+(L(x,y+1)-(L(x,y-1))^2} \qquad (12)$$

$$\theta(x,y) = \tan^{-1}((L(x,y+1)-L(x,y-1))/(L(x+1,y)-L(x-1,y)))$$

where, L(x, y), is image sample , m(x, y) is gradient magnitude, and $\theta(x,y)$ is an orientation

- Keypoint descriptor: A keypoint descriptor is created by first computing the gradient magnitude and orientation at each image sample point in a region around the keypoint location, as shown on the Figure 2.6 illustrates the computation of the keypoint descriptor. In each key point will be selected as scale in the region to measure the local image gradients and represent the significant levels of local shape distortion and change in illumination as you can see in the Figure 2.7.

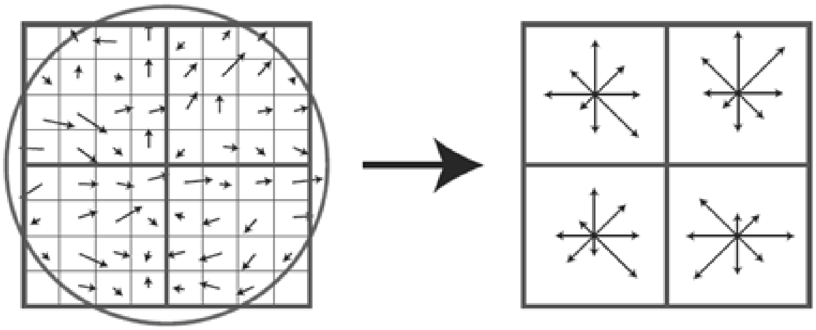


Figure 2.7 The Keypoint descritor from SIFT[13]

2.9. Review of Related Works

2.9.1. Background Model

Recently, Background modeling is very useful for video coding standard with high efficiency and low complexity in surveillance video, meanwhile, there are many researches are presented some specific requirement for background model such as low memory cost, low computation and improve the background and foreground prediction efficiency.

Xianguo Zhang *et al.[11]* is proposed the segment-and weight based running average (SWRA) method in surveillance video coding. For the step of using SWRA method, first divides pixels into several temporal segments at each position in the training frames, then the corresponding mean values and weights are calculated. For this method is used to reduce the effect of foreground pixels and obtain the modeling results which is achieved to use less modeling time and memory cost.

Xianguo Zhang *et al.* [14] conducted analyses on improved the efficiencies of using background and foreground prediction in surveillance video coding by proposed a BMAP method. For the process of BMAP all encoded blocks are utilized to classify into FB, GB or HB then the BRP and BDP modes were utilized to determine all candidate partitions of each category of blocks, and predicts each block selectively. For BRP mode is used background model from the original input as long reference and then BDP will predict the current data in the background difference domain. The BRP is improved the prediction efficiency by using high quality of background as the reference where the foreground –background hybrid blocks the BDP are provided the better reference after subtract background pixels.

Xianguo Zhang *et al.* [15] proposes a macro-block-level selective background difference coding method (MSBDC) to encoded of two ways such as encoded with each macro-block (MB) as the original MB and direct coded to the difference data between the MB and the corresponding background, and also employed the features classification of MBs to facilitate the selection which turns more accurate both on foreground and background in the prediction and motion compensation. The result of MSBDC decreased the total bitrate and gain on performance of obtained foreground.

2.9.2. Feature Preservation

Jianshu Chao *et al.* [16] proposed to improved rate control framework for H.264/advanced video coding based to improves the preservation of gradient-based features with analysis of matching score by grouped into difference feature type before encoding then utilized quantization with difference quantization parameter (QP), so the result of this method can reaches the desired target bit rate and preserved more feature information.

Jianshu Chao and Eckehard Steinbach [17] proposed to allocate the bit budget for preserved the important features such as strongest SIFT features by applied matching score for bit allocate of categorize the macro blocks in GOP then investigate the detector features of the correct matching pairs and proposed a R-D optimization method to do repeatability metric of H.264 encoder based on reduces the computational complexity and achieve better feature preservation of optimize the maximum picture quality compared to standard video encoding.

2.9.3. Region of Interests (ROI)

ROIs is the main feature which is very useful for feature analysis and recognition in surveillance video. Traditionally, aims of video coding are used to reduce the rate compression of frames while not individual the background or ROIs in that frame. To represent the regions of interest for real time identifying to extract object feature, clear facial and track any objects can be process in the coding process.

Yang Liu *et al.* [18] is proposed region-of-interest (ROI) based bit allocate bit and reduce the computational power schemes. On the ROI process, first detect the direct frame difference and skin-tone information including quantization parameter, candidates which is used in the mode decision, then the number of frame reference, motion vectors accuracy and range of motion estimation search are used to adjust MB.

At the encoder process, bits will be allocated and compute the power of that region. On the decoding process is utilized an ROI based rate-distortion-complexity (R-D-C) cost function then it will achieve the better of visual quality.

A. Pietrowcew *et al.*[19] is presented the bit-rate control algorithm for video sequence encoding with Region Of Interest (ROI).For this algorithm distributed bit budget in the image layers and take it into consideration both the distance and the local complexity to improves the image quality in ROI by lowering the image quality outside ROI with the preservation of the global constraint of the encoded stream bit-rate. The result of this algorithm is guarantees of degradation outside ROI for image quality.

Xiaoyu Wang *et al.*[20]is proposed cascaded boosting classifier which integrated the various types of features from competing regionlets. For the regoinlets is established in small groups on one-dimensional and positions stable to delineate fine-grained spatial layouts inside an object based on feature extraction. On that region is defined as the portion of window as size and aspect ratio. To estimate the object bounding box segmentation in the chosen search range by limit the locations of thousand that performs on multi-class detection then achieves the high detection mean average precision.

CHAPTER 3
PROPOSED METHOD

In this chapter, we present the overall diagram for propose algorithm which mainly focus on improving recognition performance of surveillance videos. And then, interview the way to categorize the MBs in each frame are categorized into importance and non-importance groups. Then simply way to assigning QP values to each MB based on their important levels.

The rest of this chapter is organized as follows. In section 3.1, we present the overall block diagram of proposed algorithms, describe of Feature Analysis and MB Categorizing is explained in section 3.2, and finally, to adjust Quantization parameter in MB level is explained in section3.3.

3.1. Overall Block Diagram of Proposed Algorithms

In this section, discuss about the proposed method that included of two parts, first is pre-processing process for feature analysis and MB categorize and second, QP adjustment for each MB Levels as Figure. 3.1 for the process of conducting proposed method.

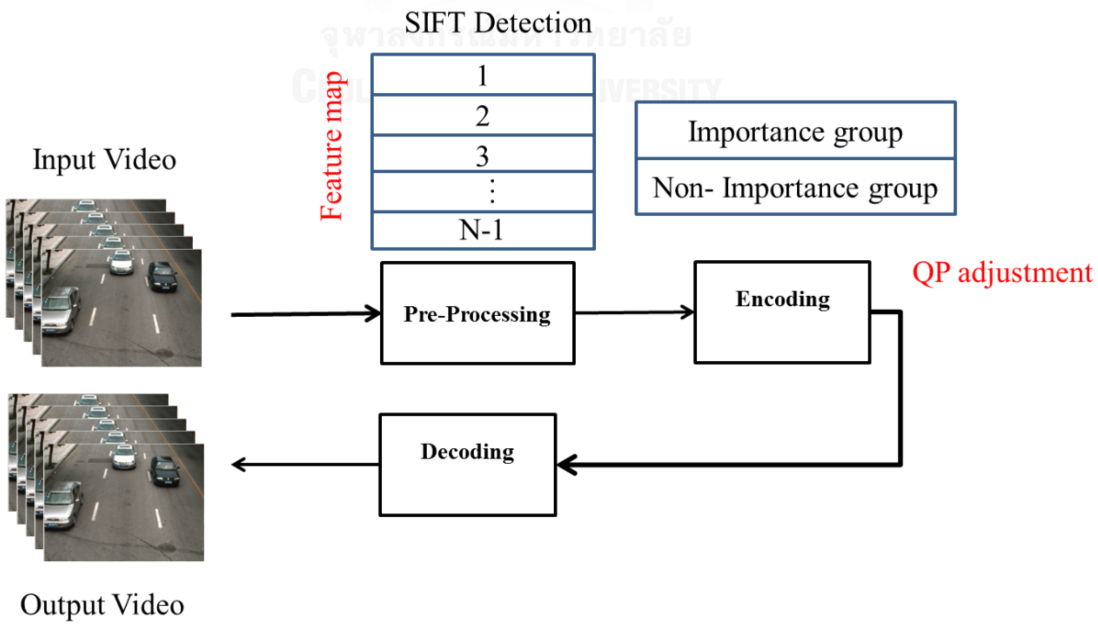


Figure 3.1 The process of conducting proposed method.

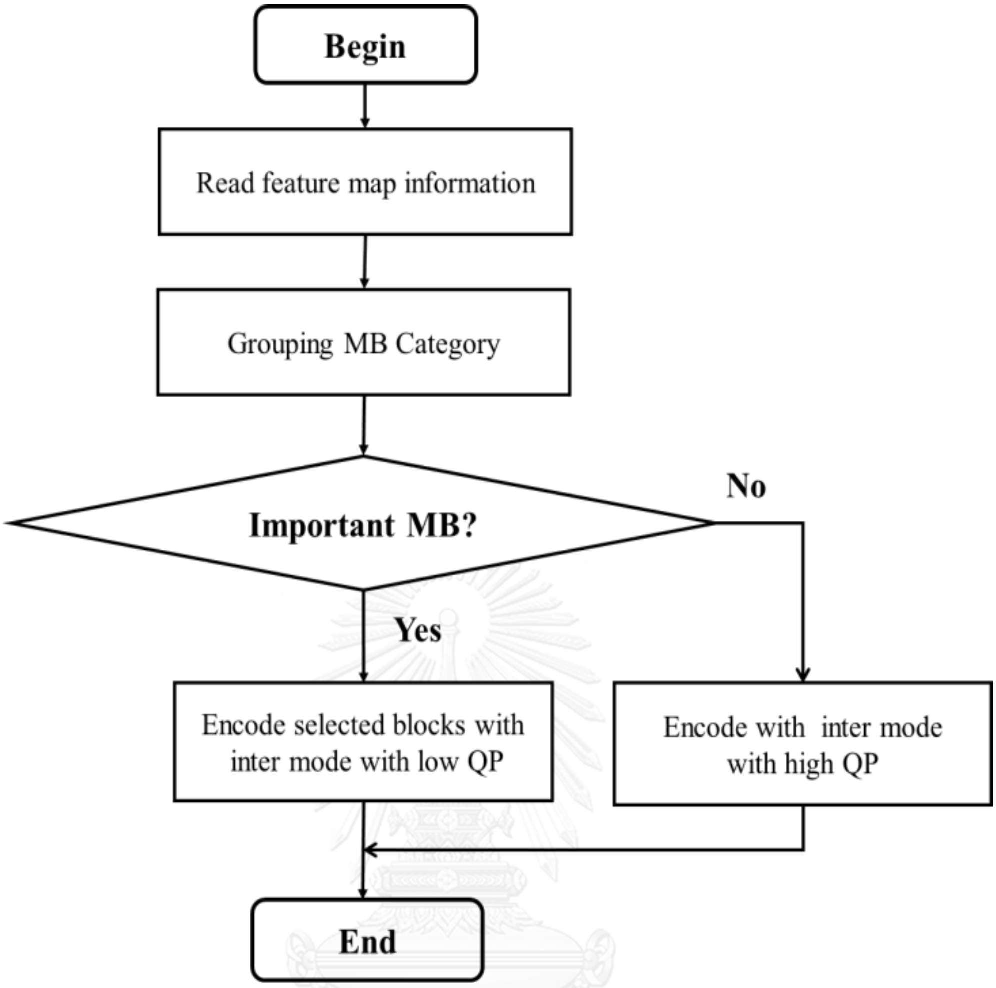


Figure 3.2 The Flowchart of proposed algorithm

The proposed algorithm, as you can see the Figure 3.2. The entire of proposed algorithm, for keypoint map information is proposed. Keypoint map for each frame is input before encoding a frame. Based on the keypoints information, macroblocks (MB) of a frame are divided into two groups: important MB group and non-important MB group. If the current MB is in important MB group, that MB is encoded with lower quantization parameter (QP) so that the features in that MB can preserve after compression. For other MBs, the high QP is used so that the overall rate performance is maintained.

3.2. Feature Analysis and MB Categorizing

In this research, SIFT feature detection algorithm as you can see in the Figure 3.3 is used to detect features of each frame of input sequences. To identify possible invariant of scale and orientation for interest points by using scale-space extreme in the image, location and scale for localize the keypoints.

A candidate keypoint has been found by compare a pixel to its neighbors, and perform a detailed fit to nearby location, scale and ratio data. After that, assign one or more orientations for every keypoint to reach of invariance on the image rotation. A neighborhood is taken around the keypoint based on scale, gradient magnitude and direction which is used to calculate the region. After that the keypoint is created with equal scale but different directions and same location to give the stability of matching and represent as a vector to form keypoint descriptor.

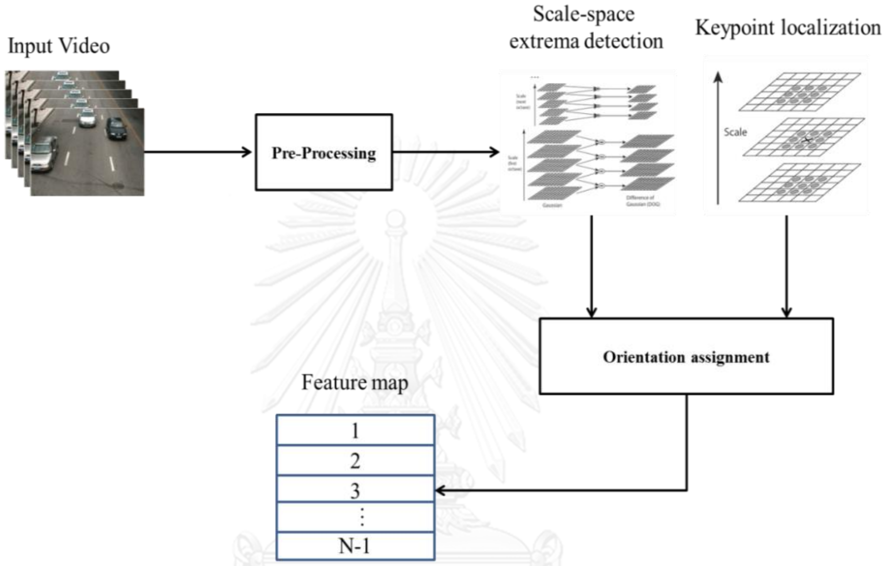


Figure 3.3 Detect Keypoint features by SIFT detection

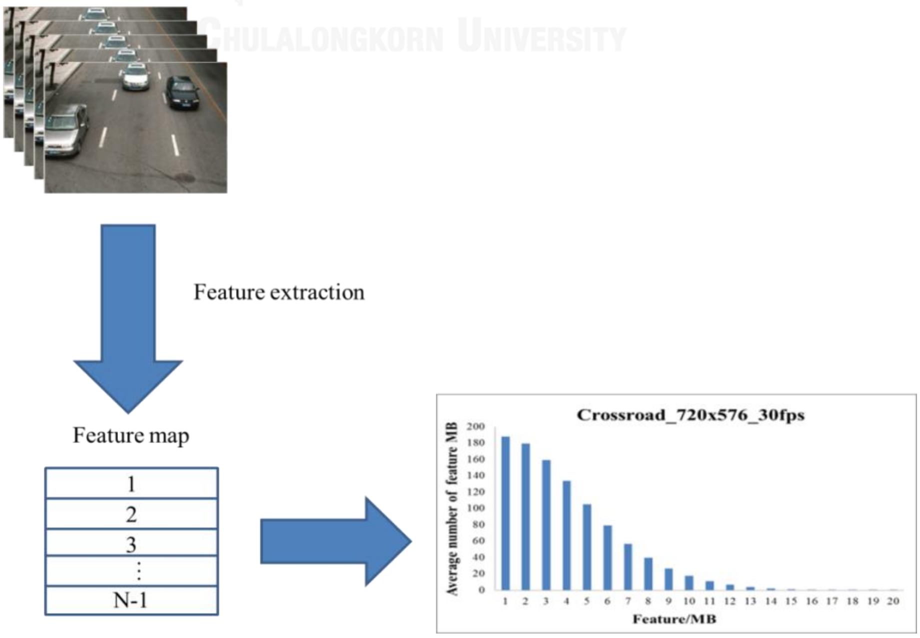


Figure 3.4 Analysis the Feature extraction in compressed video

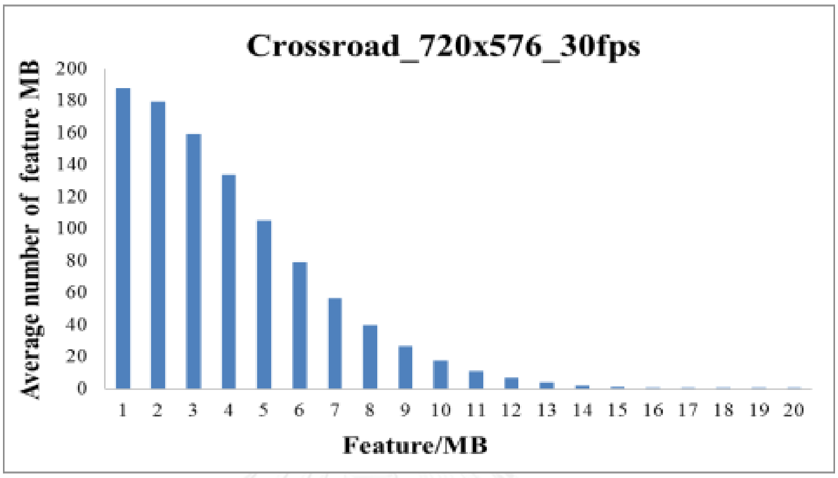
Ideally, to conduct feature extraction to recognize the location of the keypoint features in the images before compression is conducted. For the keypoint information will be stored as MB information and category into two features are considered as you can see in the Figure 3.4.

To differentiate the importance level of MBs, observations concerning SIFT features are carried out. Figure 3.5 shows the SIFT features of frame 9 of *Crossroad* sequence and frame 31 of *Classover* sequence respectively. It is obvious that some MBs contain more than one feature per block while others have no feature.

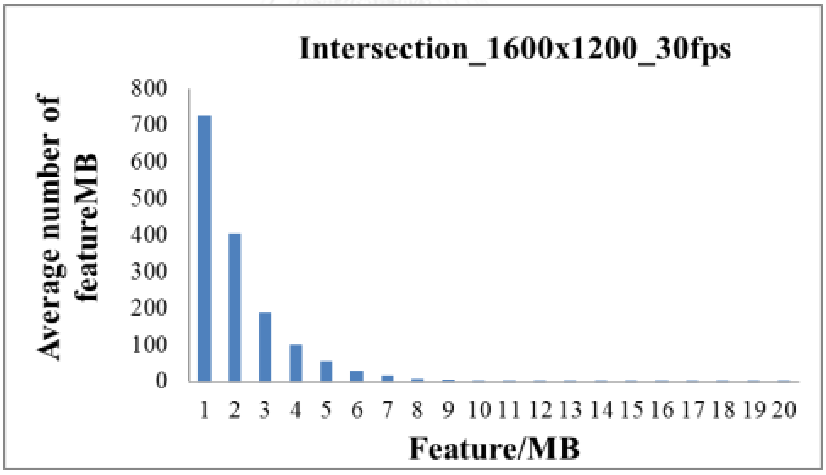
Figure 3.5 SIFT features of *Crossroad* and *Classover* sequences.

The MB distribution of each frame based on number of feature per MB is also shown in Figure. 3.6. It can be noticed that there are very few MBs that contain very high number of features and shows average number of feature MB for two sequences from PKU-SVD-A Dataset[21].

The number of MB is decreased as the number of feature per MB is increased. According to our experiments with the sequences from mention dataset, the average number of feature MB per frame is ranging from 20% - 60% of total number of MB per frame. But most of these feature MBs contain small number of feature per MB. Therefore, it is decided to select 25% of total MBs as important MB in this work. The MB with highest number of feature will be selected as important MB until the total number of selected MB reaches 25% of total MBs.



(a)



(b)

Figure 3.6 The average number of features MB for two sequences,

(a) for crossroad sequence and (b) for Intersection sequence

In summary, for the first step of our proposed algorithm is used to extract keypoint features in each MBs of each frame and categorize the MBs into different groups as important and Non-important groups. This can be done on pre-processing which is used for reference on the encoding process based on improved the recognition performance in SIFT Similarity values.

3.3. MB Level for QP Adjustment

In the proposed algorithm, on the encoding process MB categorizing is firstly carried out based on the concept that discussed in section III.

The quality of MB in important MB group should be high enough such that ideally no feature point is loss due to compression. Therefore, low QP is used for encoding these MBs. And QP of each MB is decided by using equation 13.

$$QP_{MB}(\mathrm{p}_k) = \begin{cases} QP_{frame}(k) - \delta_{qp}, & \text{if } \mathrm{MB}_p \in I_k \\ QP_{frame}(k) + \delta_{qp}, & \text{otherwise} \end{cases} \tag{13}$$

where, $QP_{frame}(k)$ is the frame level QP of frame number k

$QP_{MB}(\mathrm{p}_k)$ is the block level QP of block p of frame k

I_k is the set of important MBs for frame k

and δ_{qp} is the QP adjustment values. In this work, δ_{qp} is heuristically chosen as 6 if current MB is in important MB group. Otherwise, it is chosen as 2.

In this process, the QP is required for encoded every MBs in a frame at certain feature preservation. Based on the experimental, to achieve certain feature preservation, the MBs in the importance group will be assigned an initial QP value subtract with adjustment value for heuristically chosen as 6. For subsequent QPs in further groups, the initial QP will be added by the adjustment values for heuristically chosen as 5. According to the theory, to use the small compression ratio will achieve the better recognition as feature preservation. For the default RM.09 reference software, on the encoding process, QPs will be assigned in varying from 1 to 63 ranges to all MBs in every frame. In contrast, to allocate bit budget in this proposed algorithm, the QP for each group is assigned as initial QP subtract with 6 as an importance group and for Non-important group is used initial QP adding with 5 for respectively. At this time, we have designed the MB categorization and assign QP for each group; in these experimental results are performed to evaluate the efficiency of our methodology.

CHAPTER 4
EXPERIMENTS AND RESULTS

In this chapter, we show the experimental and results for the proposed algorithms base on improved the feature preservation for SIFT detection by using feature map and also adjust some quantization parameter is an importance technique to improve the recognition performance of video surveillance. Discussions for result of all test sequences are introduced in this chapter.

4.1. Initial the Experiments

In this section, the recognition performance of proposed algorithm is evaluated by using SIFT similarity values. For the experiments, three standard definition resolution (720x576 pixels) test sequences: *Classover*, *Overbridge* and *Crossroad*, High definition (1280x720 pixels) test sequence for *Intersection*, and two ultra-extended graphic array (UEGA) resolution (1600x1200 pixels) test sequences: *Mainroad* and *Intersection*, from PKU-SVD-A dataset [21] are used as you can see in the Table 4.1 the parameter of all test sequences from PKU-SVD-A dataset.

Table 4.1 Test Sequences parameter

Video Sequences	**Width**	**Height**	**Frame-rate**	**Total frames**
Crossroad	720	576	30	3000
Classover	720	576	30	3000
Overbridge	720	576	30	3000
Intersection	1280	720	30	1000
Mainroad	1600	1200	30	1000
Intersection	1600	1200	30	1000

The current reference software of IEEE 1857 (RM09)[22] and the encoder with proposed algorithm are used in these experiments. All test sequences are firstly encoded with RM09 by using three target bitrates, 256kbps, 512kbps, and 1500kbps. IPPP… coding structure is used for all cases. Background modeling and rate distortion optimization modes are disable. MB level rate control is enable.

In this research, the recognition performance of proposed algorithm is evaluated by using three main parameters such as: Bitrate, PSNR and SIFT Similarity values. These are snapshot of each test sequence is shown in Figure. 4.1. All these sequences have a frame rate of 30fps. Each standard definition sequence has 3000 frames whereas each UEGA sequence has 1000 frames and HD sequence has 1000 frames.

Classover 720x576

Crossroad 720x576

Overbridge 720x576

Mainroad 1600x1200

Intersection 1600x1200

Intersection 1280x720

Figure 4.1 Video Test Squences from PKU-SVD-A dataset

For the proposed algorithm, a feature information file “Feature map” is required as input. The feature information for each sequences is obtaining by SIFT detection method. Each test sequence is encoded by the encoder with proposed algorithm using same configurations as mention earlier for RM09 case.

4.1.1. SIFT Similarity measurement

To evaluate the SIFT Similarity, there are including of three Steps: Detect the key points using SIFT detection, calculate descriptors (feature vectors), and matching descriptor vectors with a brute force matcher with two video as input such as original and compressed videos.

$$SIFT\ similarity = \frac{\#\ \text{correctly detected keypoint}}{\#\ \text{original keypoint}} \tag{14}$$

4.1.2. PSNR measurement

Peak signal-to-noise ratio (PSNR) is a measure of the "noise" in a signal relative to the peak intensity. When comparing the two images, PSNR is calculated by taking the mean squared error (MSE) between the pixel intensities and taking the ratio of the maximum possible intensity in relation to typical 8-bit value this is $2^8 - 1 = 255$.

- To compute the mean-squared error using the equation:

$$MSE = \frac{\sum (f - F)^2}{N} \tag{15}$$

where, f is the original image, F is the reconstructed images, and N is the size of image (Weight x Height).

- To compute PSNR using the equation:

$$PSNR = 20\log_{10}\left(\frac{255}{\sqrt{MSE}}\right) \tag{16}$$

4.2. Experimental Results

For the experiment and results show the improvement of feature preservation for SIFT Similarity by using feature extraction or feature map as a reference with proposed algorithms compare with the original encoder RM.09 to improve the recognition performance of video surveillances especially on SIFT Similarity values.

Crossroad video sequence with 720 x 576 resolutions is encoded with 256 kbps target bitrate to compare the improvement of SIFT Similarity of frame number 100 to 200. For this result, SIFT Similarity value of the proposed algorithm is up to about 0.56 is equal 56% compared with original encoder is 0.46 equal 46%. It means on these frames in this video sequence for proposed algorithm, SIFT Similarity score is increasing about 10% to preserved the number of keypoint features because of these MBs which is stored in important MBs feature which is reached 25% and encoded with low QP to preserve the information. For frame number 160 it contains more distortion because of encoded with high QP So SIFT Similarity is decreased than the original as you can see in the Figure 4.2.

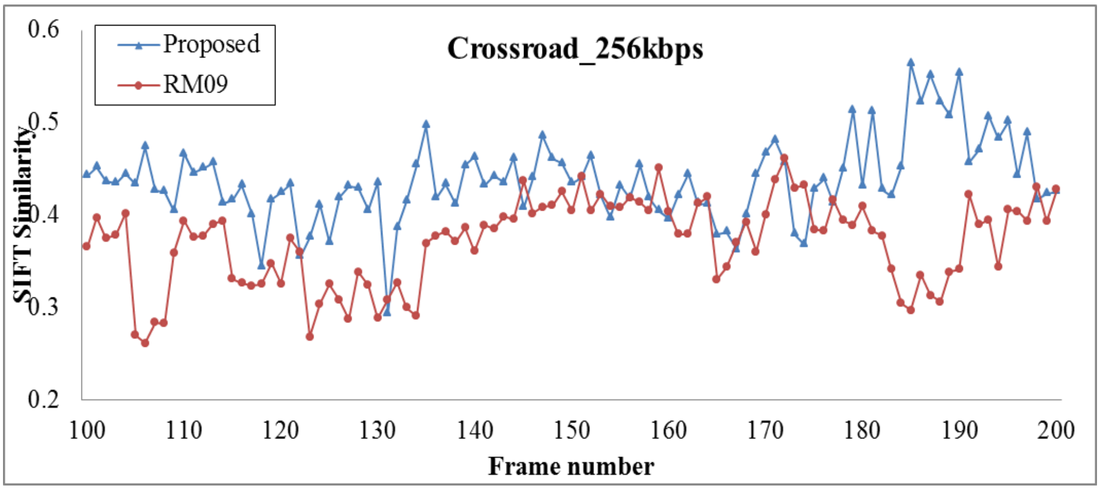


Figure 4.2 The improvement of SIFT Similarity for *Crossroad* 256kbps with frame number 150 to 250

Classover video sequence with 720 x 576 resolutions is encoded with 512 kbps target bitrate of frame number 100 to 200. For this result, SIFT Similarity value of the proposed algorithm is up to about 0.6 is equal 60% compared with original encoder is 0.45 equal 45%. It means on these frames in this video sequence for proposed algorithm, SIFT Similarity score is increasing about 15% to preserved the number of keypoint features because of these MBs which is stored in important MBs feature which is reached 25% and encoded with low QP to preserve the information as you can see in the Figure 4.3.

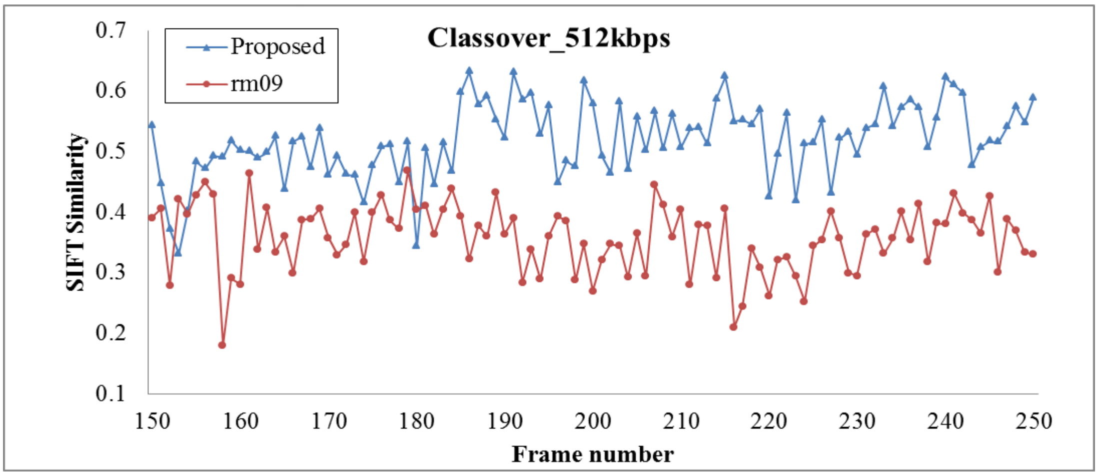


Figure 4.3 The improvement of SIFT Similarity for *Classover* 512kbps with frame number 150 to 250

Intersection video sequence with 1280 x 720 resolutions is encoded with 512 kbps target bitrate of frame number 150 to 250. For this result, SIFT Similarity value of the proposed algorithm is up to about 0.8 is equal 80% compared with original encoder is 0.3 equal 30%. So SIFT Similarity score from this result is increased 50% to preserved the number of keypoint features because of these MBs which is stored in important MBs feature which is reached 25% and encoded with low QP to preserve the information as you can see in the Figure 4.4.

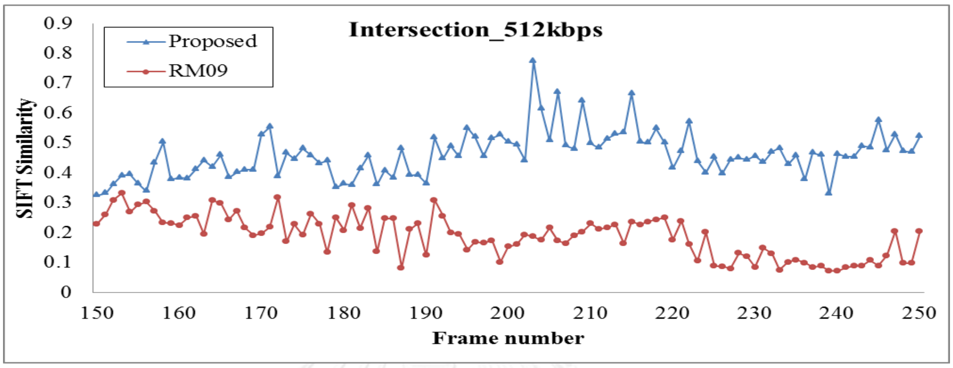


Figure 4.4 The improvement of SIFT Similarity for *Intersection* 512kbps with frame number 150 to 250

Mainroad video sequence with 1600 x 1200 resolutions is encoded with 512 kbps target bitrate of frame number 150 to 250. For this result, SIFT Similarity on this video sequence of the proposed algorithm is about 0.08 is equal 8% compared with original encoder is 0.05 equal 5% So SIFT Similarity score from this result is increased 3% to preserved the number of keypoint features because of these MBs which is stored in important MBs feature which is reached 25% and encoded with low QP to preserve the information.as you can see in the Figure 4.5.

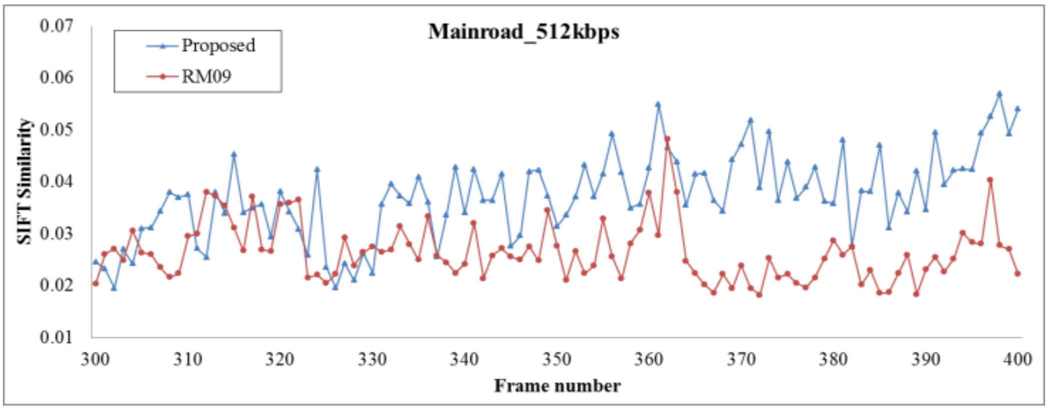


Figure 4.5 The improvement of SIFT Similarity for *Mainroad* 512kbps with frame number 300 to 400

In the other hand, to show the result of overall video sequences are encoded with three difference target bitrate 256,512, and 1500 kbps to compared of PSNR, Generated Bit and the SIFT Similarity score with proposed algorithm and the original RM09 reference software as you can see the detail in these tables.

Table 4.2 Result of comparison PSNR, Bitrate and SIFT Similarity between RM09 and Proposed algorithm for *Crossroad* sequence.

Crossroad (720x576, 30fps)									
Method	PSNR			Bitrate			SIFT		
Bitrates (kbps)	256	512	1500	256	512	1500	256	512	1500
RM09	24.54	27.02	30.30	256.20	512.12	1500.27	0.43	0.52	0.40
Proposed	**24.56**	**27.02**	**30.30**	**256.19**	**512.13**	**1500.30**	**0.60**	**0.84**	**0.89**

Table 4.3 Result of comparison PSNR, Bitrate and SIFT Similarity between RM09 and Proposed algorithm for *Classover* sequence.

Classover (720x576, 30fps)									
Method	PSNR			Bitrate			SIFT		
Bitrates (kbps)	256	512	1500	256	512	1500	256	512	1500
RM09	35.42	36.78	38.99	256.77	512.11	1500.35	0.33	0.34	0.39
Proposed	**35.43**	**36.78**	**38.99**	**256.77**	**512.11**	**1500.35**	**0.71**	**0.84**	**0.93**

Table 4.4 Result of comparison PSNR, Bitrate and SIFT Similarity between RM09 and Proposed algorithm for *Overbridge* sequence.

Overbridge (720x576, 30fps)									
Method	PSNR			Bitrate			SIFT		
Bitrates (kbps)	256	512	1500	256	512	1500	256	512	1500
RM09	29.41	31.45	34.02	256.14	512.11	1500.23	0.48	0.50	0.53
Proposed	**29.41**	**31.46**	**34.02**	**256.14**	**512.10**	**1500.22**	**0.58**	**0.61**	**0.65**

Table 4.5 Result of comparison PSNR, Bitrate and SIFT Similarity between RM09 and Proposed algorithm for *Intersection* sequence.

Intersection (1280x720, 30fps)									
Method	PSNR			Bitrate			SIFT		
Bitrates (kbps)	256	512	1500	256	512	1500	256	512	1500
RM09	30.57	34.45	39.48	256.03	512.03	1500.24	0.61	0.86	0.92
Proposed	**30.63**	**34.5**	**39.48**	**256.03**	**512.04**	**1500.24**	**0.90**	**0.95**	**0.97**

Table 4.6 Result of comparison PSNR, Bitrate and SIFT Similarity between RM09 and Proposed algorithm for *Mainroad* sequence.

Mainroad (1600x1200, 30fps)									
Method	PSNR			Bitrate			SIFT		
Bitrates (kbps)	256	512	1500	256	512	1500	256	512	1500
RM09	29.38	31.78	34.12	273.90	512.55	1500.69	0.08	0.10	0.20
Proposed	**29.39**	**31.80**	**34.13**	**270.70**	**512.53**	**1500.72**	**0.49**	**0.52**	**0.57**

For the experimental and results of all video sequences are coded with both original encoders of RM09 reference software and proposed algorithm to compare the improvement of SIFT Similarity score according to improve feature preservation based on scheme as we have described in Chapter 3. The initial QPs are varying on the range of 1 to 63. Base on the analysis for feature map information as we analyses on the pre-processing, the QP adjustment values are used to subtract with 6 for the the important group and adding with 5 for the Non-important group. These are the heuristically values which is used to preserved the keypoint features for MBs which is contained of 25% of MBs feature on these sixth video sequences. It can be seen that the proposed approaches lead to improve SIFT Similarity significantly for feature preservation and both retain PSNR and bitrate values same as current reference software.

4.3. Result of Analysis

4.3.1. Comparisons in SIFT Similarity with QP

On this figure describes an obtaining of SIFT Similarity values based on compress *Crossroad* video sequence with 256Kbps target bitrate with several QP on the proposed algorithm compare with original encoder. Based on the results shows that using lower QP are achieved the better SIFT similarity value up to 1.0 equal to 100% of Similarity in some frame but in original encoder even using low QP but SIFT Similarity still low because of in that frame the MBs is not store in the important MBs so the overall of improve SIFT Similarity values in this sequence is about 49% and the video sequence as you can see in the Figure 4.6.

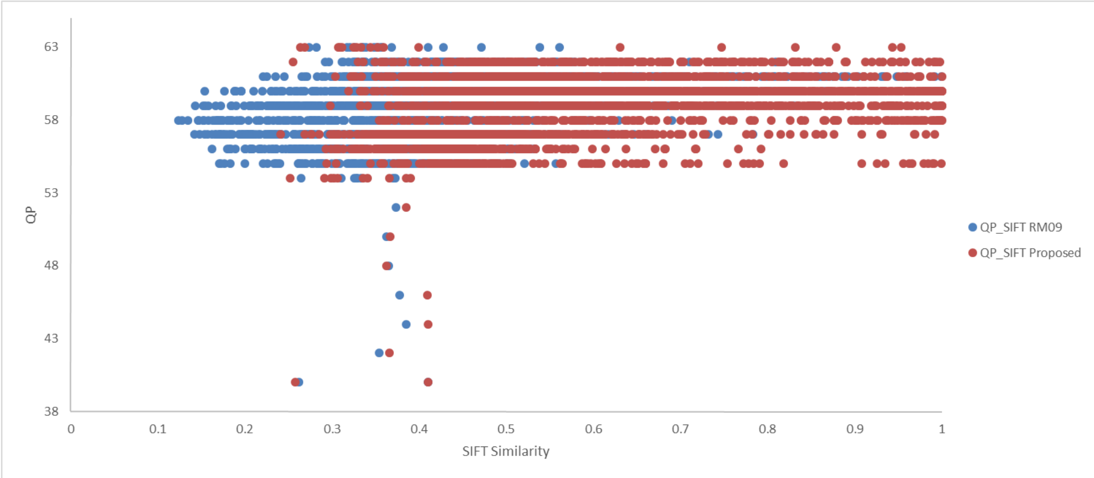


Figure 4.6 Achieve of SIFT Similarity values based using QP to compress *Crossroad* 256kbps video sequence between RM09 and Proposed method

On this figure describes an obtaining of SIFT Similarity values based on using PSNR on proposed algorithm compare with original encoder. Based on the results shows that using higher PSNR up to 24dB are achieved the better SIFT similarity value up to 0.8 equal to 80% of Similarity values as you can see in the Figure 4.7.

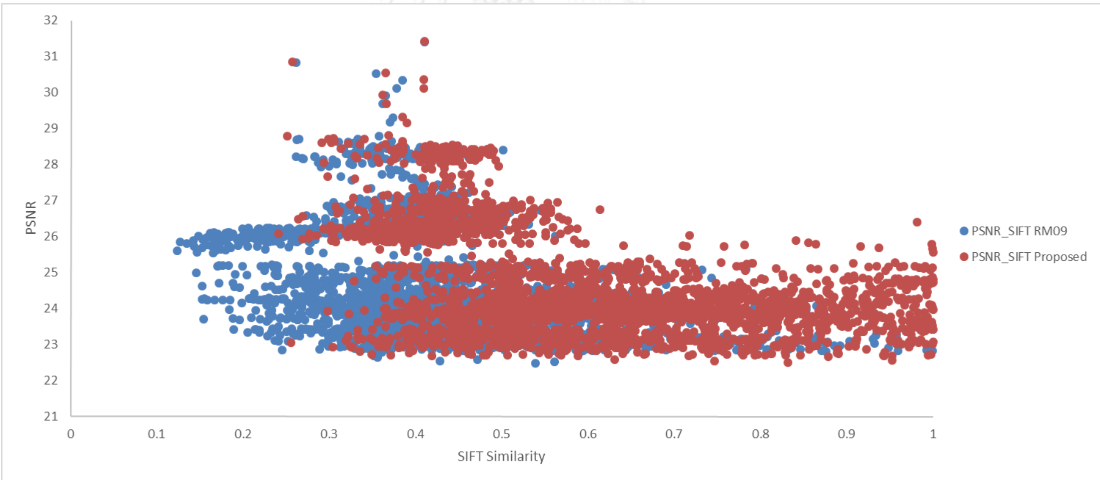


Figure 4.7 Achieve of PSNR values based on improve SIFT Similarity to compress *Crossroad* 256kbps video sequence between RM09 and Proposed method

On this figure describes an obtaining of SIFT Similarity values from frame number 1200 to 1800 on proposed algorithm compare with original encoder. Based on the results shows that on the proposed algorithm is achieved the SIFT Similarity value up to 0.9 equal to 90% compare with original encoder from 0.4 equal to 40% mean the proposed algorithm can achieve the better recognition for 50% of SIFT Similarity as you can see in the Figure 4.8

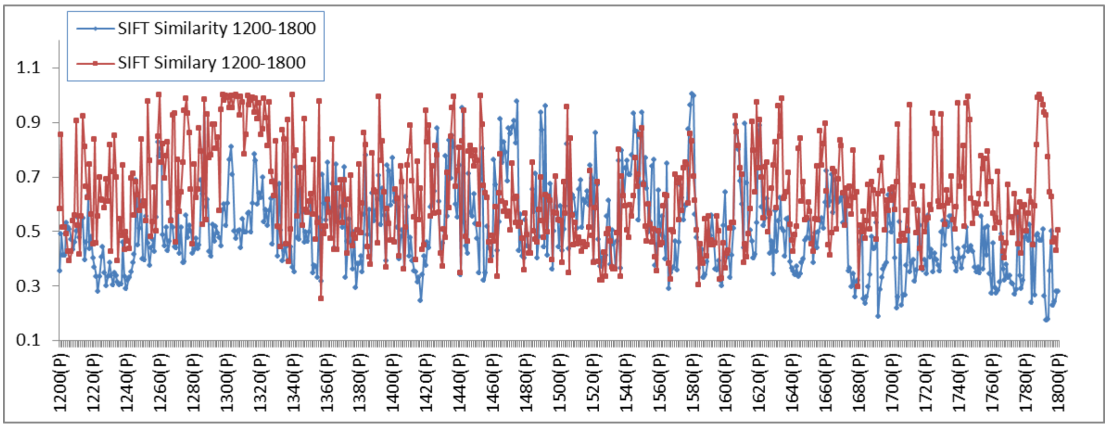


Figure 4.8 The comparison using QP to compress *Crossroad* 256kbps between RM09 and Proposed method

On this figure shows the effect of distortion from encoded with higher QP from frame number 1485, the red color part as you can see in the Figure 4.9 some distortion occur which is effected to SIFT Similarity of proposed algorithm is lower than RM09 about 5% because of this frame is included small number of MBs feather encode with high QP which effect to missing some information while frame number 1489 the value of SIFT Similarity is increased about 20%.

Figure 4.9 The distortion of compress video frame for *Crossroad* 256kbps from frame number 1485 between RM09 and Proposed method

4.3.2. Comparisons in SIFT Similarity with Generated Bitrate

On this figure describes the number of generated bitrate with using QP between RM09 and Proposed algorithm to compress *Crossroad* sequence with 256Kbps target bitrate. For frame number 1400 to 1500, the generated bitrate is increase compare to RM 09 original encoder and the SIFT Similarity is also increase about 30% in frame number 1404 as you can see in figure 4.10.

According to the analysis result we can say that the proposed algorithms is achieved to improve the recognition of SIFT Similarity because of we focus to improve in MBs level that can detect more edge and only focus to encoded the MB with have more feature. Due to the similarity score is calculated by the ratio between the number of keypoints from compressed image divided by the number of keypoints in original frame. According to the result for over all the sequence, the Standard Resolution for C*rossroad*, *Classover* and *Overbridge* video sequences are improved the SIFT Similarity score up to 49% when encoded with high target bits and for the UEGA resolution for *Mainroad* and *Intersection* video sequences are up to 41%. So SIFT Similarity of the proposed algorithm is achieved the better quality while some MBs or frame has more Distortion but without important part so the average of PSNR and Bitrate are retaining the same values of original encoder.

CHAPTER 5
CONCLUSION AND FUTURE WORKS

5.1. Summary and Conclusion

In this research, we have proposed to improve the preservation of SIFT detection and maintain the quality of surveillance videos and bitrates in chapter 3.

In section 3.1 we introduce the overall diagram of proposed algorithm in IEEE 1857 standard. However, many researches have been done about improved of preservation.

In section 3.2 we introduce how to do the feature Analysis by introduce the process of SIFT feature detection to detect the keypoints and MBs in each frame are categorized into importance and non-importance groups.

In section 3.3 we introduce the simply assigning QP values to each MB based on their important levels.

For the experimental results, the performance of proposed algorithm is compared with RM09 are introduces in chapter 4.

In section 4.1 Initialize the test sequences from PKU-SVD-A dataset and explain some parameter which is use to evaluated the performance of original encoder and proposed algorithm.

In section 4.2 Shows the experimental and results with discussion the result form proposed method compared with current reference software of IEEE 1857, in term of SIFT Similarity. Based on the experimental results, our proposed algorithm generally outperformed to improve feature recognition performance for surveillance video compared with current reference software of IEEE 1857 standard while maintaining the recognition performance at same quality for PSNR and bitrate.

5.2. Future work

According to the main contribution of IEEE 1857 standard is background modeling. For the future work, we will extend our algorithm by taking into account the background modeling feature of IEEE 1857 and optimize rate control base on using background modeling.

REFERENCES

[1] I. E. G. Richardson, "The H.264/AVC Video Coding Standard for the Next Generation Multimedia Communication " *Chichester, U.K. Wiley,* pp. 136–138, 2003.

[2] W. Gao, Y. Tian, T. Huang, S. Ma, and X. Zhang, "The IEEE 1857 Standard: Empowering Smart Video Surveillance Systems," *Smart Surveillance,* pp. 30-32, 2014.

[3] C.-L. C. Mina Makar, David Chen, Sam S. Tsai and Bernd Girod "Compression Of Image Patches For Local Feature Extraction," *International Conference on Acoustics, Speech and Signal Processing(ICASSP),* pp. 821-824, April 2009.

[4] J. C. a. E. Steinbach, "Preserving Sift Features In JPEG-Encoded Images," *IEEE International Conference on Image Processing (ICIP),* pp. 301-304, 2011.

[5] I. Richardson, "An Overview of H.264 Advanced Video Coding," *Vcodex White Paper,* 2007-2013.

[6] S. Ma, S. Wang, and W. Gao, "Overview of IEEE 1857 video coding standard," *The International Conference on Image Processing (ICIP),* pp. 1500-1504, 2013.

[7] T. Wiegand, G. J. Sullivan, G. Bjøntegaard, and A. Luthra, "Overview of the H.264/AVC Video Coding Standard," *IEEE Transactions On Circuits And Systems For Video Technology,* vol. 13, pp. 560-575, July 2003.

[8] J.-R. O. Gary J. Sullivan, Woo-Jin Han, and Thomas Wiegand, "Overview of the High Efficiency Video Coding (HEVC) Standard," *IEEE Transactions On Circuits And Systems For Video Technology,* vol. 22, pp. 1649-1668, December 2012.

[9] Z. M. Zhen Tang, and Yanli Wan, "Background Subtraction Using Running Gaussian Average and Frame Difference " *IFIP International Federation for Information Processing 2007* pp. 411–414, 2007.

[10] B. P. L. L. a. S. A. Velastin, "Automatic Congestion Detection System for Underground Platforms " *International Symposium on Intelligent Multimedia Multimedia, Video and Speech Processing* pp. 158-161, May 2-4 2001

[11] X. Zhang, Y. Tian, T. Huang, and W. Gao, "Low-complexity and high-efficiency background modeling for surveillance video coding," *Proc. IEEE Int’l Conf. Visual Communication and Image Processing,* pp. 1-6, 2012.

[12] R. D.Cajole, "Joint Source Channel Error Resilient Video Coding For Wireless Video Transmission," *PhD. Thesis Chulalongkorn University,* 2009.

[13] D. G. Lowe, "Distinctive Image Features from Scale-Invariant Keypoints," January 5, 2004.

[14] T. H. Xianguo Zhang, Yonghong Tian and Wen Gao, "Background-Modeling-Based Adaptive Prediction for surveillance Video Coding," *IEEE TRANSACTIONS ON IMAGE PROCESSING,* vol. 23, February 2014.

[15] Y. T. Xianguo Zhang, Luhong Liang, Tiejun Huang, Wen Gao, "Macro-Block-Level Selective Background Difference Coding For Surveillance Video," *IEEE International Conference on Multimedia and Expo,* pp. 1067-1072, 2012.

[16] R. H. Jianshu Chao, Eckehard Steinbach,Damien Schroeder, "A Novel Rate Control Framework for SIFT/SURF Feature Preservation in H.264/AVC Video Compression," *Ieee Transactions On Circuits And Systems For Video Technology,* vol. 25, pp. 958-972, JUNE 2015.

[17] J. C. a. E. Steinbach, "SIFT Feature-Preserving Bit Allocation For H.264/Avc Video Compression," *IEEE International Conference on Image Processing (ICIP)*, pp. 709 - 712, 2012.

[18] Z. G. L. Yang Liu, Yeng Chai Soh, "Region-of-Interest Based Resource Allocation for Conversational Video Communication of H.264/AVC," *IEEE Transactions On Circuits And Systems For Video Technology*, vol. 18, pp. 134-139, January 2008.

[19] A. B. A. Pietrowcew, W. Skarbek, "Bit-rate Control Algorithm Based on Local Image Complexity for Video Coding with ROI," *Advanced Video and Signal Based Surveillance (AVSS),* pp. 582-587, 2005.

[20] M. Y. Xiaoyu Wang, Shenghou Zhou, Yuanqing Lin, "Regionlets for Generic Object Detection," *IEEE International Conference on Computer Vision(ICCV),* pp. 17-24, 2013.

[21] PKU-SVD-A-Dataset [ftp://222.29.140.51/] [Online].

[22] I. C. Society, "http://www.ieee1857.org/software/RM12.4.zip," ed, 1857-2013.

APPENDIX

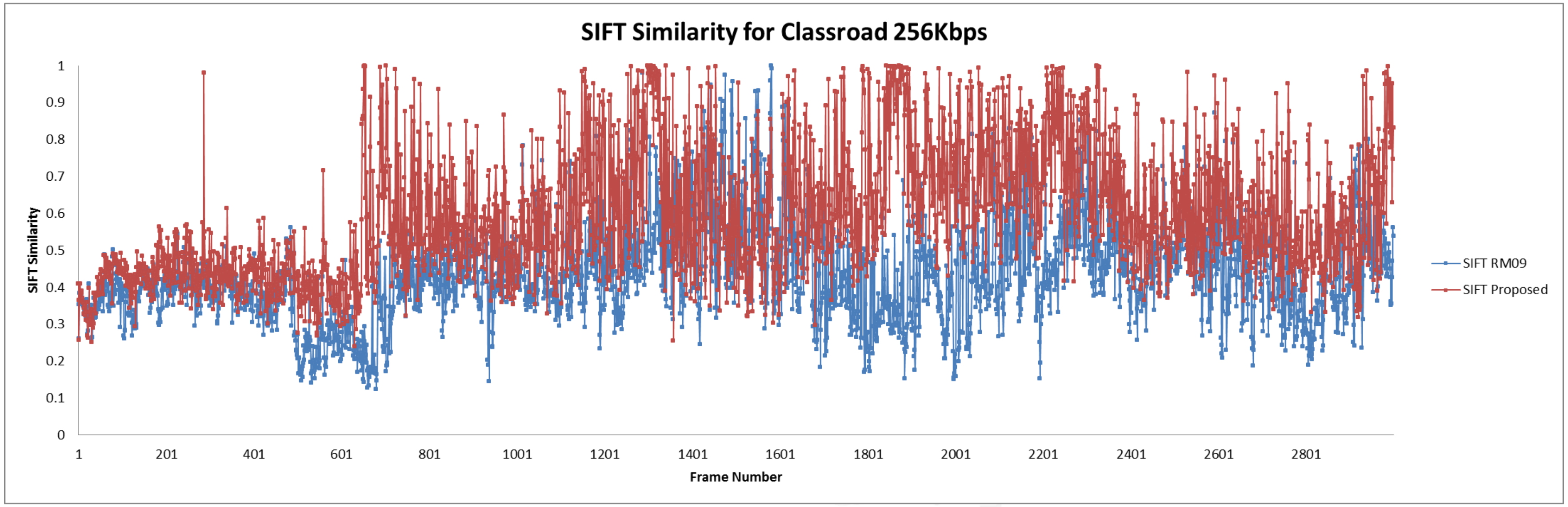


Figure 5.1 The comparison of improvement SIFT Similarity for *Crossroad* sequence with 256 Kbps

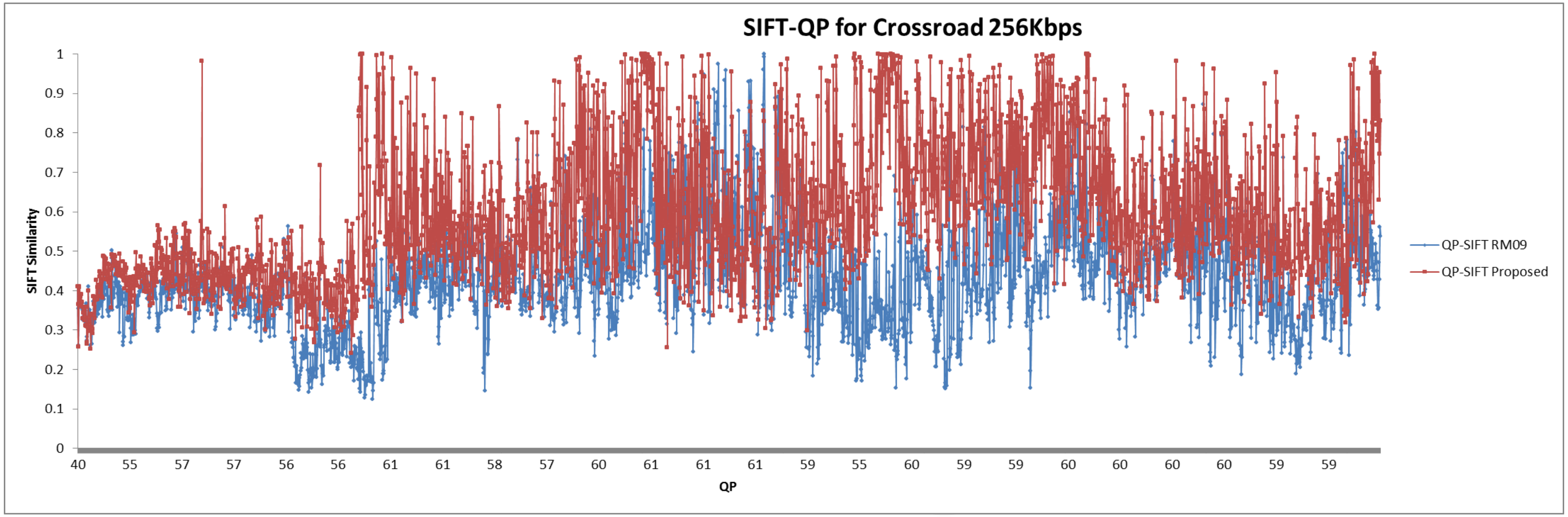


Figure 5.2 The improvement of SIFT Similarity with QP for *Crossroad* sequence with 256 Kbps

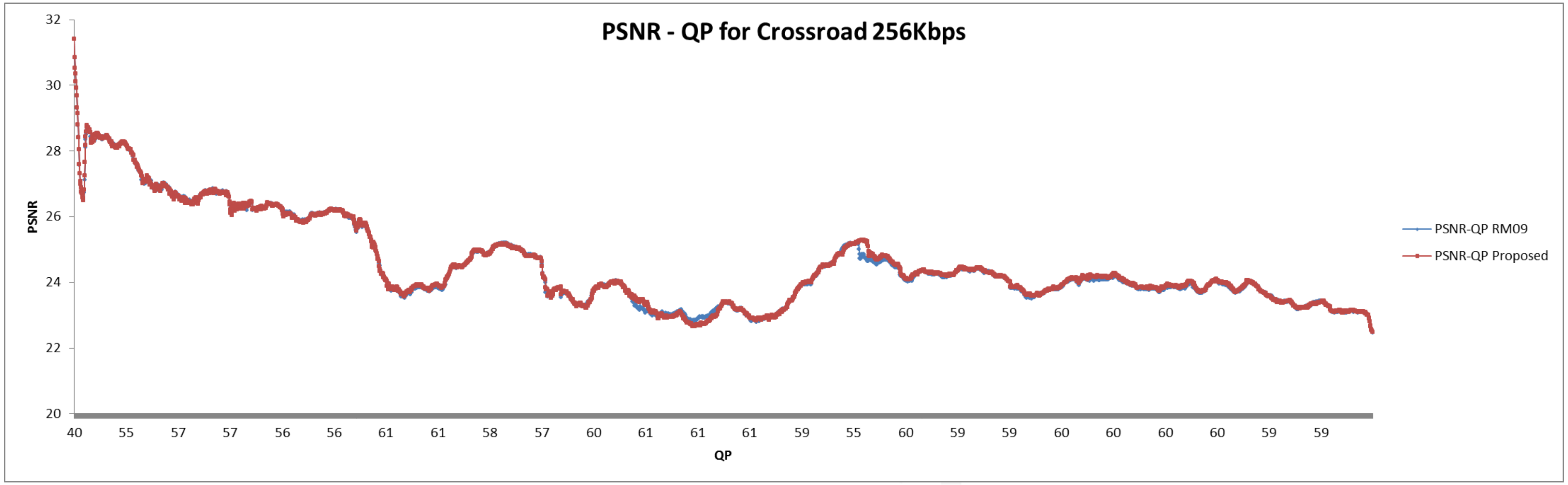


Figure 5.3 The improvement of PSNR with QP for *Crossroad* sequence with 256 Kbps

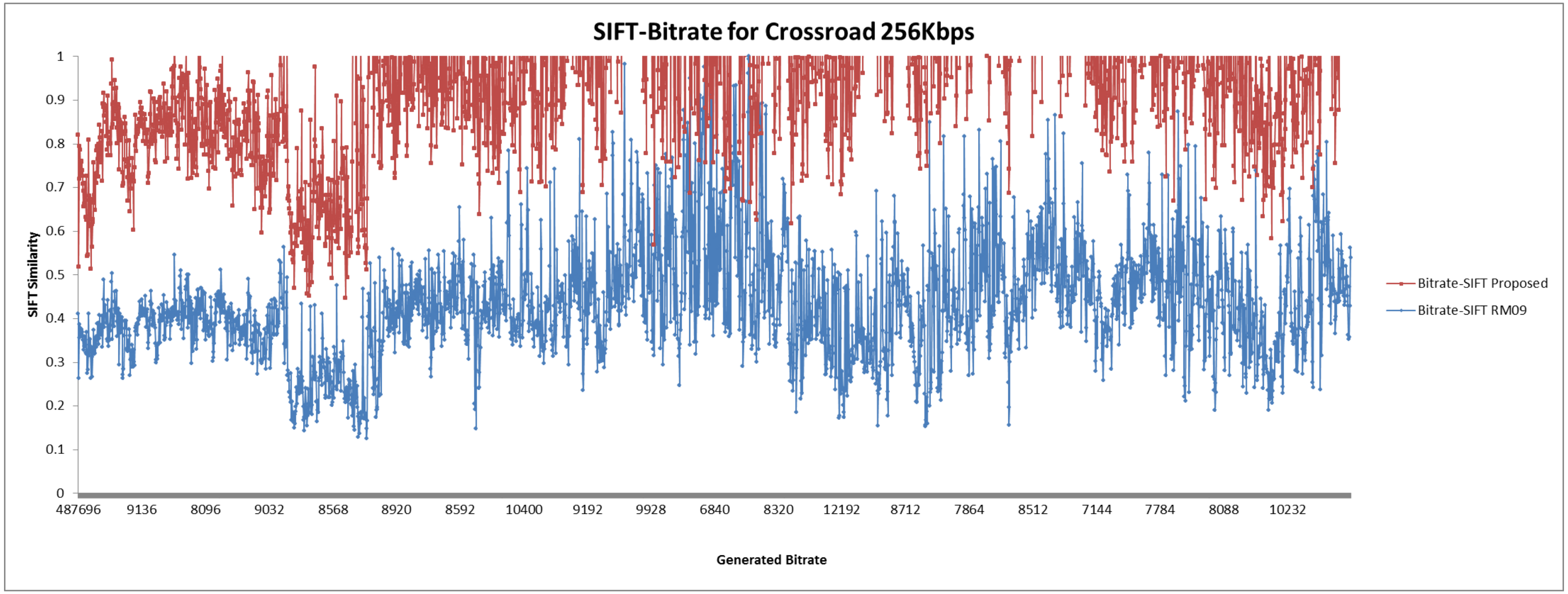


Figure 5.4 The improvement of SIFT Similarity with Generated bit for *Crossroad* sequence with 256 Kbps

VITA

Thipkesone BOUNNAKHOM was born in Vientiane, Lao PDR, in 1989. She hold Bachelor degree of Computer Science in 2011 in Faculty of Natural Science, National University of Laos and IT Specialist course in 2013 (Software Engineering) in Faculty of Engineering ,University of Laos.

In 2013, she is received scholarship from the AUN/Seed-Net to continue Master degree on the field of Multimedia and Signal Processing in the division of Digital Signal Processing Research Laboratory (DSPRL), Department of Electrical Engineering, Faculty of Engineering, Chulalongkorn University.

For the research area is video coding based on improve feature preservation on IEEE 1857 Standard.