การปรับปรุงของการอนุรักษ์ลักษณะเฉพาะในการเข้ารหัสวีดิทัศน์ประสิทธิภาพสูง

นายโซวัน เชน

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิศวกรรมศาสตรมหาบัณฑิต
สาขาวิชาวิศวกรรมไฟฟ้า ภาควิชาวิศวกรรมไฟฟ้า
คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย
ปีการศึกษา 2559

IMPROVEMENT OF FEATURE PRESERVATION IN HIGH EFFICIENCY VIDEO CODING

Mr. Sovann Chen

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 2016

Thesis Title	IMPROVEMENT OF FEATURE PRESERVATION IN HIGH EFFICIENCY VIDEO CODING
By	Mr. Sovann Chen
Field of Study	Electrical Engineering
Thesis Advisor	Associate 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
(Associate Professor Supot Teachavorasinskun, Ph.D.)

THESIS COMMITTEE

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

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

....External Examiner
(Associate Professor Wuttipong Kumwilaisak, Ph.D.)

โซวัน เชน : การปรับปรุงของการอนุรักษ์ลักษณะเฉพาะในการเข้ารหัสวีดิทัศน์ประสิทธิภาพสูง (IMPROVEMENT OF FEATURE PRESERVATION IN HIGH EFFICIENCY VIDEO CODING) อ.ที่ปรึกษาวิทยานิพนธ์หลัก: สุภาวดี อร่ามวิทย์, 51 หน้า.

ข้อมูลคุณลักษณะที่ได้มาจากจุดสำคัญในภาพ เป็นข้อมูลที่สำคัญสำหรับการนำไปใช้วิเคราะห์วีดิทัศน์ในรูปแบบการใช้งานต่าง ๆ เช่น การนำไปใช้ตรวจจับและติดตามวัตถุ การจำแนกวีดิทัศน์ เป็นต้น เนื่องจากลักษณะของการเข้ารหัสวิดีทัศน์ จะใช้ประโยชน์จากระบบที่เกี่ยวกับการมองเห็นของมนุษย์ ที่ส่วนประกอบความถี่สูงสามารถถูกขจัดออกไปเพื่อทำให้การบีบอัดข้อมูลได้มากขึ้น ดังนั้นจึงมีผลต่อจุดสำคัญในภาพ โดยเฉพาะอย่างยิ่งบริเวณขอบของภาพ หากจำนวนของจุดสำคัญเหลือน้อยเกินไป จะส่งผลต่อความแม่นยำของการวิเคราะห์วีดิทัศน์ เพื่อแก้ปัญหาดังกล่าววิทยานิพน์นี้นำเสนอขั้นตอนวิธีการรักษาข้อมูลคุณลักษณะเฉพาะของการสร้างวีดิทัศน์ในการเข้ารหัสวีดิทัศน์ประสิทธิภาพสูง ในการศึกษานี้เราเลือกใช้ Scale-Invariant feature transform (SIFT) คือ การเลือกจุดสำคัญจากลำดับวิดีทัศน์ดิบ จากนั้นเราจะพิจารณาจุดสำคัญที่บ่งบอกว่าหน่วยการเข้ารหัสแบบใหญ่ที่สุด นั้นมีความสำคัญ ต่อมาจะใช้ตัวปรับการเลือกหน่วยการเข้ารหัสแบบใหญ่ที่สุด โดยกำหนดหน่วยการเข้ารหัสแบบใหญ่ที่สุด ออกเป็นสองกลุ่ม คือ กลุ่มของหน่วยการเข้ารหัสแบบใหญ่ที่สุดที่มีความสำคัญ และ กลุ่มของหน่วยการเข้ารหัสแบบใหญ่ที่สุดที่ไม่มีความสำคัญ นอกเหนือจากนั้นได้นำเสนอจะแบ่งการจัดสรรบิตออกเป็นสองประเภท คือ การควบคุมอัตราในแต่ละกลุ่มภายใต้โหมดการเข้ารหัสแบบอินทราหรืออินเตอร์ ที่รักษาอัตราบิตเป้าหมาย และข้อมูลคุณลักษณะไว้ ผลการทดสอบแสดงให้เห็นว่าขั้นตอนวิธีที่เรานำเสนอสามารถรักษาจุดสำคัญในวีดิทัศน์ได้มากกว่าโปรแกรมการสร้างวีดิทัศน์ในการเข้ารหัสวีดิทัศน์ประสิทธิภาพสูงอ้างอิง ที่อัตราบิตเท่ากันภายใต้การคำนวณแบบอัตราส่วนค่าสัญญาณสูงสุดกับสัญญาณรบกวน และความเหมือนกันของ SIFT

ภาควิชา	วิศวกรรมไฟฟ้า	ลายมือชื่อนิสิต	....
สาขาวิชา	วิศวกรรมไฟฟ้า	ลายมือชื่อ อ.ที่ปรึกษาหลัก	....
ปีการศึกษา	2559		

5870290221 : MAJOR ELECTRICAL ENGINEERING
KEYWORDS: KEYPOINTS / SIFT / LCU / HEVC / PSNR

SOVANN CHEN: IMPROVEMENT OF FEATURE PRESERVATION IN HIGH EFFICIENCY VIDEO CODING. ADVISOR: ASSOC. PROF. SUPAVADEE ARAMVITH, Ph.D., 51 pp.

Feature information is known as interest points (keypoints) in an image which is very useful information for video analytic functions such as object detection and tracking, video classification, etc. Due to the characteristic of video coding, it is exploited following human vision system where high frequency component could be removed for better compression. Thus, it can affect to the keypoints, which mostly are the edge information. As a result, with less number of keypoints left, this affects the accuracy of video analytics. To solve this problem, this thesis presents an algorithm to preserve feature information of reconstructed video in the high efficiency video coding (HEVC). Scale-Invariant Feature Transform (SIFT) is chosen to extract the keypoints from raw video sequence. We then consider keypoints as an indicator of the importance of the largest coding unit (LCU). Adaptive LCU selection is defined to determine LCU into two different groups, important LCU group (IMLCU) and non-important LCU group (Non-IMLCU). Moreover, two different bit allocations are generated in rate control to each group based on coding mode, Intra or Inter mode, to achieve the target bit rate and also to keep the feature information. The experimental results show that our proposed algorithm can maintain more keypoints compared to HEVC reference software at the same bitrate based on the peak signal-to-noise ratio (PSNR) and SIFT similarity computation.

Department:	Electrical Engineering	Student's Signature	....
Field of Study:	Electrical Engineering	Advisor's Signature	....
Academic Year:	2016		

ACKNOWLEDGEMENTS

I would like to thank all the people who contributed in some ways to the work. First and foremost, I would like to express my gratitude to my thesis advisor Associate Professor Supavadee Aramvith who always guides, supports and encourages me to achieve my research. This achievement also would not have been possible without the guidance and the help of several individuals who in one way or another contributed and extended their valuable assistance.

I would also like to express my sincere thanks to committee members, Assistant Professor Suree Pumrin and Associate Professor Wuttipong Kumwilaisak, for their valuable comments and technical suggestions in the completion of this research.

Additionally, I am thankful to my seniors, Htoo Maung Maung and Watchara Ruangsang, who have spent their valuable time to give suggestion and also always help me, encourage me and share good thoughts to me. I am also thankful to all the staffs of ISE and AUN/SEED-Net, for their support.

Finally, I must express my very profound gratitude to my parents for their love, understanding, providing me with unfailing support and continuous encouragement throughout my years of study. Without their support and encouragement, this accomplishment would not have been possible. This research has been supported in part by the Collaborative Research Project entitled Video Processing and Transmission, JICA Project for AUN/SEED-Ned, Japan.

CONTENTS

Page

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1
INTRODUCTION

1.1. Motivation and Problem Statement

Technology is an application, which is used to accomplish numerous tasks in our daily lives like television, phone and computer. As well as, it is advancing from one-step to another. Nowadays, the resolution of television (TV) broadcasting is much higher than before by improving from standard definition television (SDTV) then high definition television (HDTV) after that 4K ultra high definition (UHD) and up to 8K UHD according to [1]. In addition, the video data on internet traffic targeting at mobile devices and tablet personal computers is increasing rapidly, as shown in Figure 1.1 from a report on "Trends and Analysis – Cisco" of Cisco Visual Networking Index 2016 in [2]. Hence, it strongly needs high video coding technique to reduce network traffic load with higher quality and less bit-rate.

High Efficiency Video Coding (HEVC) [3], the latest video coding standard is revealed by ITU-T and ISO/IEC to fulfil above desires. This latest one can reduce bit rates about 50 percentages at the same visual quality by comparing with the previous one, (H.264/AVC [4]); the detail is in [5]. There are some improved techniques from previous standard such as the flexible partitioning by using Quad-tree structure, flexibility in prediction modes, sample adaptive offset (SAO), and interpolation with cutting-edge techniques. In addition to achieve high compression ratio, video coding standards are generally exploited human visual system (HVS). In contrast, if compression ratio is higher and higher, it will effect to visual feature extraction and decrease the ability of video analysis. Figure 1.2 shows the decoded frame from two HEVC test video sequences by using HEVCE reference software (HM15) [6] with 265 kbps. Hierarchical-P coding structure for low delay applications is used to get those decoded frames. Mainly in this coding structure, frames are categorized into two groups of frames, core frames and common frames. Low quantization parameter (QP) values are generally assigned for core frames. In contrast, for common frames, they are encoded with high quantization parameter (QP) values which get lower quality than core frames. As a result, the important information of raw video will be lost after encoding and decoding process with low bit rates.

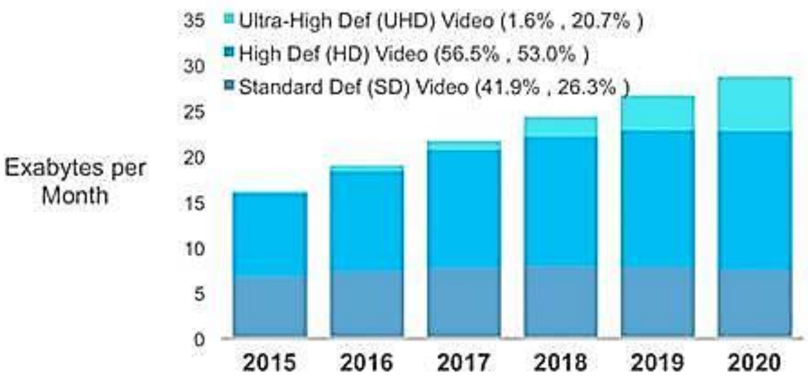
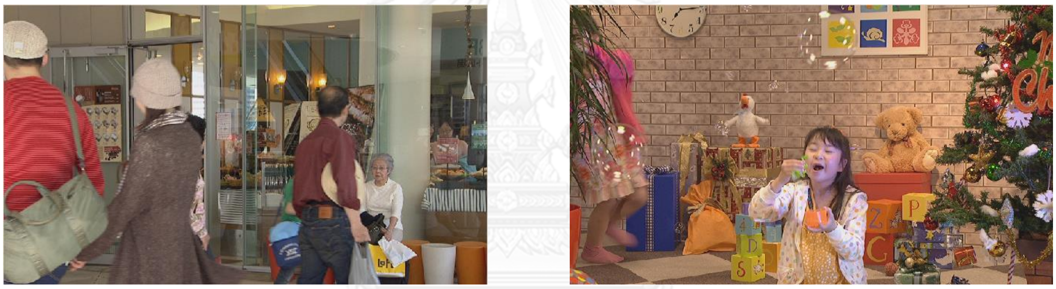


Figure 1.1. Video data traffic from Cisco VNI 2016 [2]



(a)

(b)

Figure 1.2. Losing the important information after compression with 256 kbps

(a) Raw video sequences, (b) Decoded video sequences

Moreover, the popularity of computer vision application has been increased and focused more on the object in image like applying facial detection algorithm, object recovery algorithm, object tracking algorithm, recognition algorithm, super resolution algorithm, and so on, where the feature extraction is one of the important keys to achieve the good performance in each computer vision application. Up to here, feature preservation in image or video compression becomes more interested problem where the main purpose is to maintain the feature of raw image as high as possible after compression.

1.2. Objectives

In this work, there are several objectives as follows:

- Investigate feature extraction in video coding
- Propose a modified HEVC encoder to improve features after decoding process
- Evaluate the performance of feature preservation in proposed algorithm with reference HEVC encoder software

1.3. Scope of work

- Analyse local detectors in the video by applying feature extraction technique, scale invariant feature transform (SIFT)
- Propose a feature preserved video coding algorithm to improve SIFT Similarity while maintaining target bit rate
- Examine the performance of feature preservation in proposed algorithm with reference HEVC software (HM15) based on SIFT Similarity

1.4. Expected Outputs

Improved HEVC encoder software that can preserve important feature of video which output video with good quality contains adequate keypoints that are useful for recognition in the later process.

1.5. Research Procedures

- Doing literatures review about scale invariant feature transform, image matching and video coding methods
- Collecting datasets of surveillance videos
- Doing simulation to check the performance of default HEVC reference software
- Proposing and implementing an algorithm to maintain the detectors after compression

- Simulating and analysing the performance of proposed method comparing with the default HEVC reference software
- Taking proposal examination
- Writing paper for international conference
- Writing thesis paper
- Taking thesis final defense

1.6. Outline of Thesis

There are five main chapters, which are represented in this proposal including this chapter. The rest chapters of this proposal are organized and provided brief descriptions as follows:

Chapter 2 describes some literature reviews and backgrounds such as: understanding basic of video coding, HEVC overview, briefly the conception of Rate Control, Scale Invariant Feature Transform, and review feature preservation methods for image and video compression.

Chapter 3 presents the proposed method which is described about feature analysis, largest coding unit (LCU) grouping extraction and quantization parameter (QP) computation.

Chapter 4 explains the results and discussion of experiment.

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

CHAPTER 2
LITERATURE REVIEWS AND BACKGROUND

An overview of the fundamentals video coding standard is provided in the first part of this chapter. Then High Efficiency Video Coding Standard is briefly introduced with some important new features. Next, rate control is briefly introduced. SIFT is also revealed in this chapter. The last part is focused on the review feature preservation methods on image and video compression.

2.1. Basic background of video coding standard

In general, the uncompressed video or raw video sequences in digital format are required large amount of bits to transmit or store. Practically, the raw video sequences need to be compressed before storing or transmitting because of limitation of storage device and transmission bandwidth costs. By chance, the researchers found the redundancies signal occurred in video signals which could be good criteria to get high efficiency compression. Those redundancies are group into four different types such as spatial redundancy, temporal redundancy, perceptual redundancy and statistical redundancy.

Normally a video sequence, there are a lot of similar pixel values are found in a frame are called spatial redundancy. In addition, it also has redundant pixels between subsequent frames called temporal redundancy. Based on the nature of human visual system, there is some detail information in picture that our eye could not perceive, especially high frequency components. This phenomenon causes perceptual redundancy. For that reason, the removing of high frequency component could not affect the quality of the picture for human perception. For statistical redundancy, it generally occurs in entropy coding part by checking the level of probability of content. To simplify, it comprises of two techniques, lossless compression and lossy compression, in data compression. Lossless compression is a technique to reduce bits by identifying and eliminating statistical redundancy to remain the information from original data after reconstructing from the compressed data. Lossy compression is used to reduce the large amount of bits by removing the unnecessary or less important information for storage, handling, and transmitting content.

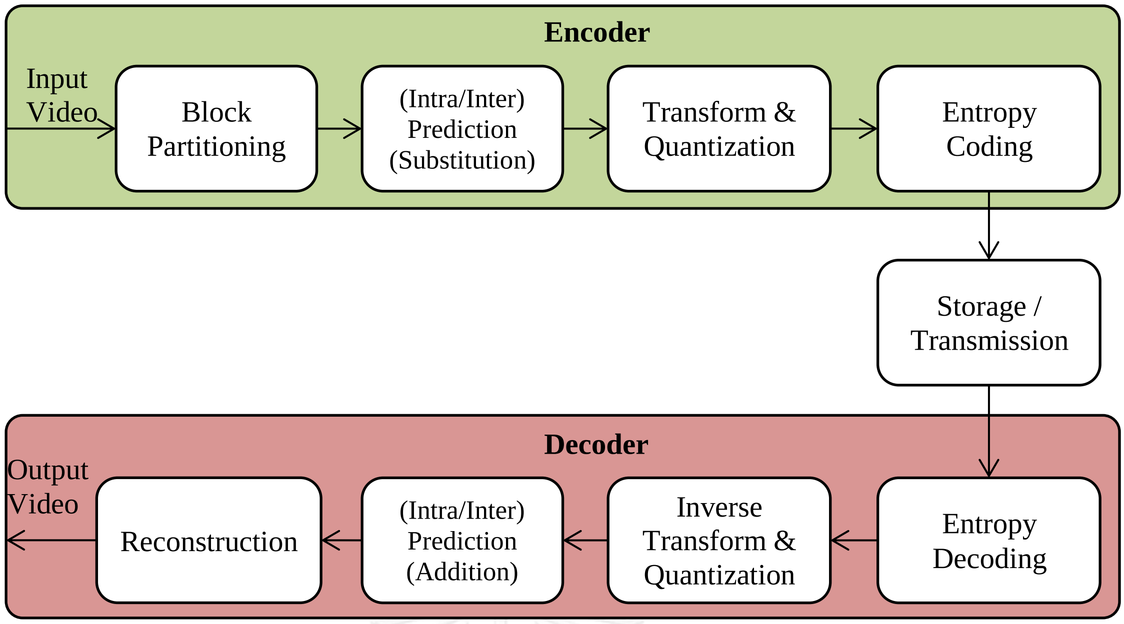


Figure 2.1. Block diagram of block-based hybrid video codec

Commonly, the video coding algorithm is designed following block-based hybrid video coding principle which has implemented to be a successful tool to compress a video signal into a smallest possible bitstream size by eliminating the redundant information from the signal. This principle video coding, four main steps are practical used including block partitioning, prediction mode, transformation and quantization, and also entropy coding as shown in Figure 2.1.

Firstly, each frame in raw video is subdivided into small non-overlapping square 16x16 luma blocks in commonly used. These blocks can be called as basic units.

Then the predicted signal is generated following Intra prediction or Inter prediction to those blocks. The prediction block to neigboring blocks in current picture is done by using Intra prediction and the prediction block in current picture from block in previous picture is done by using Inter prediction. There is much amount of blocks that can be predicted in Inter prediction based on the changing scene in video. To select prediction mode, encoder decision criterion is applied for example, rate-distortion optimization. Additionally, decoded picture buffer (DPB) is needed to use in encoder to store the decoded frame of previous encoded frame until frame is displayed. These reconstructed frames become the referent pictures which are used to estimate the current picture. After getting predicted signal, residual signal for each block is computed by applying frame different of original signal and predicted signal.

The next step is focused on conversion pixel domain to frequency domain to de-correlate signal by using discrete-cosine transform (DCT) which is applied on these residual signals of each block to get new value called coefficient. Then, these coefficient values are quantized to remove high frequency components responding to human visual perception.

Last but not least, quantized coefficient values are sent to entropy coding part to produce the output bit streams based on the statistic of contents. Statistical redundancy can be reduced in this last part by using entropy coding to generate the bit for frequently occurrence and others. Short or large code lengths are defined according to their own probability. The decoder process can be done by doing reverse process.

2.2. H.265/HEVC

High Efficiency Video Coding is also built by following the block-based hybrid video coding approach which has been successful basic structure in video coding such as MPEG-2 [7] and H.264/AVC video coding standard. Figure 2.2 shows an architecture design of HEVC encoder with build-in decoder which it has been improved and introduced several new features from the previous video coding standard to achieve half bitrate redundant with same quality. These improvements are:

1. Flexible partitioning by using Quad-tree coding
2. Flexibility in prediction modes
3. Improved interpolation and deblocking filters
4. Support parallel processing

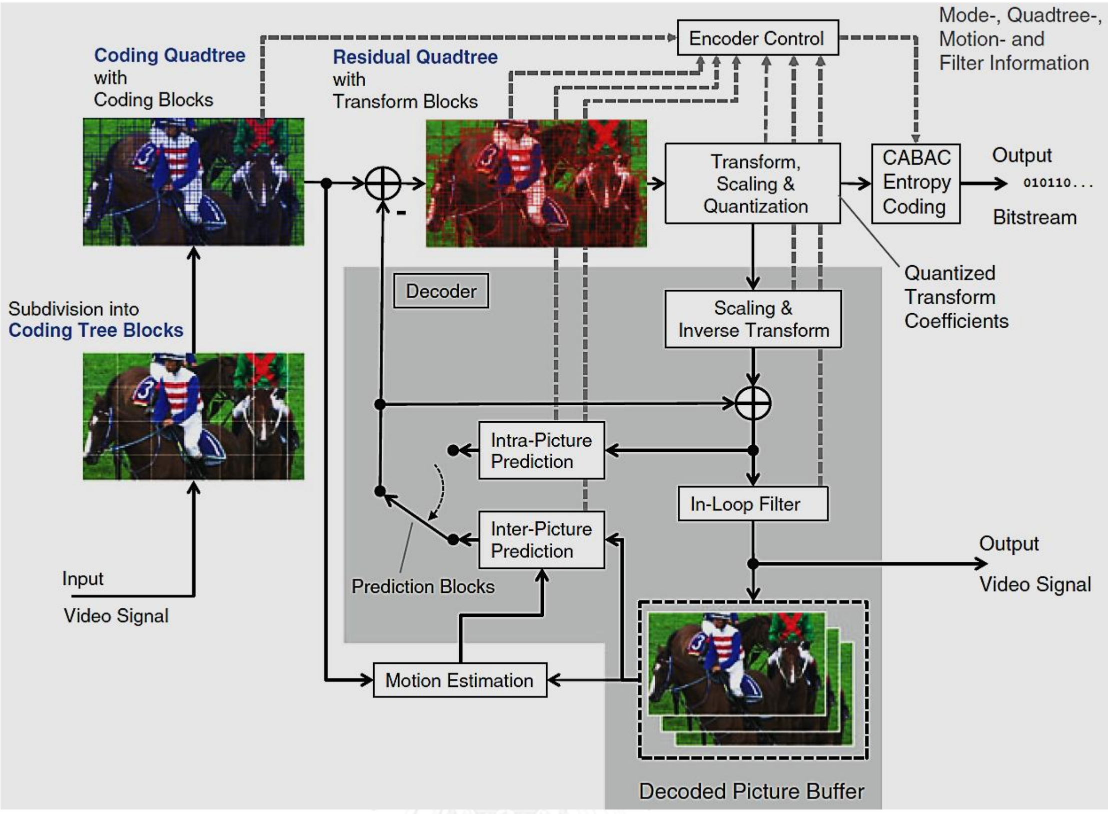


Figure 2.2: Architecture design of HEVC encoder with built-in decoder [3]

These innovative features with its own advantages will introduce in subsection below.

2.2.1. Quad-Tree Coding

In all previous video coding standards, each frame in video sequence is generally split into fixed 16x16 block of luma component and 8x8 of chroma components which are called macroblocks. However, larger block size is beneficial solution in partitioning part for high resolutions where it can support motion-compensated prediction and also transforms coding. It typically provides more suitably coding efficiency. Nevertheless, delay could be increased in this technique. Additionally, small block size is also important to support adaptive partition block based on picture properties. Hence, these ideas have been thought and applied on the latest video coding standard, HEVC, by introducing a concept of Quad-tree-coding to determine the flexible block size in a frame where these sizes is defined following Lagrangian rate-distortion cost in the encoder. In this latest video coding standard, macroblock is considered as the basic unit or coding tree unit (CTU). The flexible mechanism is applied on video frames for partitioning into CTU of variable sizes can be 16x16, 32x32 and 64x64. Then CTU is branded into coding units (CUs), where the size can be vary from 64x64 to 8x8 luma samples by using quadtree syntax as indicated in Figure 2.3. These CUs can also be more divided to square or non-square blocks beside of coding structure for main purpose such as Intra-picture prediction mode, Inter-picture prediction mode, and transform coding.

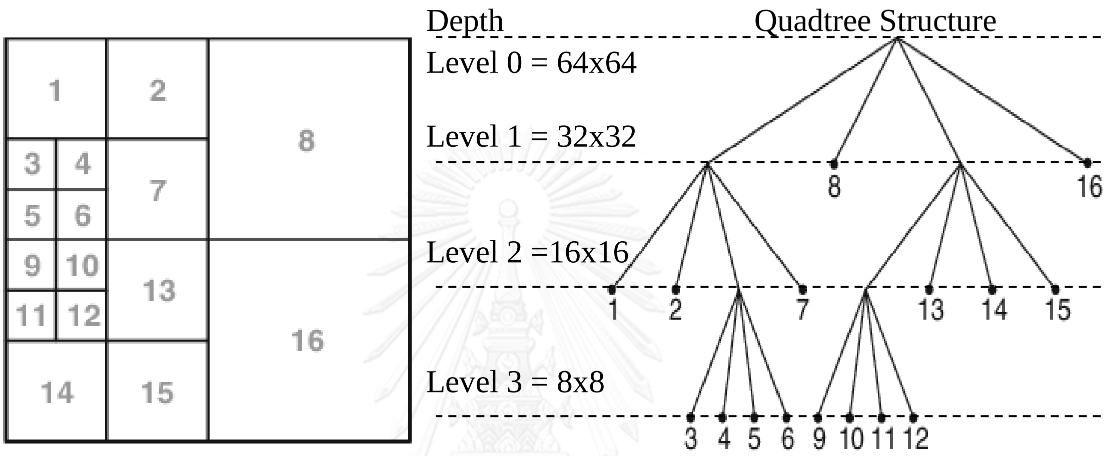


Figure 2.3. Quadtree coding structure in HEVC

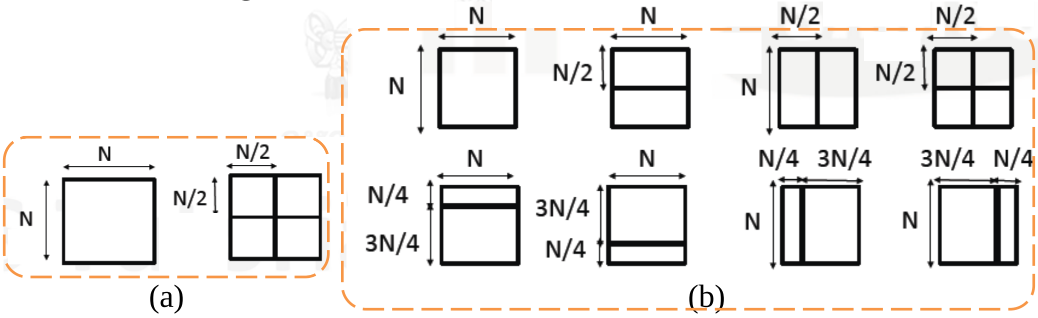


Figure 2.4. Partitioning of prediction units (PUs)

(a) intra-prediction, (b) inter-prediction

Prediction mode is signalled to indicate whether CUs are coded using Intra prediction mode or others, Inter prediction mode. In Intra prediction mode, CU can be fragmented into four square prediction units (PUs) if CU size is 8x8 block size or a single luma PU. In Inter-prediction, CU can be fragmented into square or non-square PU, but no 4x4 PU. Figure 2.4 is shown the partitioning of prediction units.

After defined prediction units, prediction residuals are computed and sent to transform coding to continue the processing. In HEVC, it supports several integer transforms block, 4x4 to 32x32 sample blocks. There are two types of transform coding methods which are integer discrete sine transform (IDST) for 4x4 block size and integer discrete cosine transform (IDCT) for other usages. This divisible is called Residual Quad-tree.

2.2.2. Intra-Picture Prediction

Intra prediction mode in HEVC can be analysed in two categories, angular prediction method which has 33 modes to provide accuracy codec with model structures of its directional edges, planar prediction and DC prediction method which can bring a good predictors estimating smooth image content. Figure 2.5 shows the modes of intra prediction.

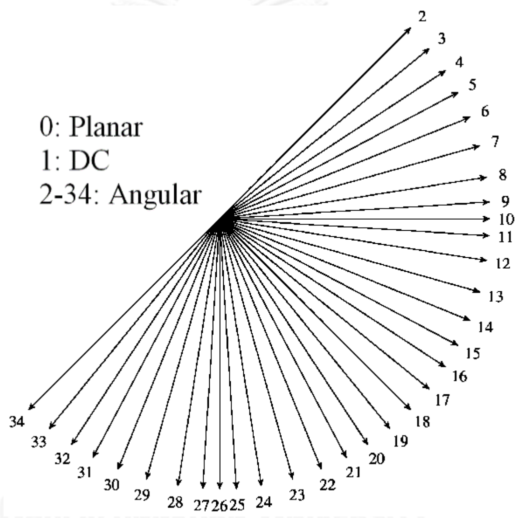


Figure 2.5. Intra prediction modes

2.2.3. Inter-Picture Prediction

While Intra prediction mode has taken the advantages of spatial correlation between neighboring samples, Inter prediction comes to focus on temporal correlation between pictures in furtherance of predicting motion data. There are two concepts in this prediction where advanced motion vector prediction is the first concept to enhance motion vector prediction and Inter-picture prediction block merging is the second concept to simplify the block-wise of motion data signalling by gathering all motions data from previous decoded blocks. Then the information after these two concepts is going to apply in next process of fractional interpolation of reference picture samples where this fractional interpolation is enlarged to 7/8 tap kernels to use in luma channel and 4-tap kernels to use in chroma channel to get high precision interpolation filtering. At the final stage, weighted sample prediction is applied.

2.2.4. Entropy Coding

Earlier video coding standard, H.264/AVC, it has introduced two main techniques of entropy coding. They are arithmetic coding and variable length coding. And this latest video coding standard, HEVC, it decides to use only arithmetic coding where this entropy coding method can challenge to parallel processing architectures and provide high coding efficiency.

2.2.5. In-Loop Filters

Reference pictures is a main important part to minimize the residual error, which can get high coding efficiency. To improve reference picture in previous coding standard, Loop filter is introduced deblocking filter and sample adaptive offset (SAO) where deblocking filter is useful to attenuate the discontinuities at the prediction and transform block boundaries and SAO is applied to improve quality of decoded pictures. It has good benefit to attenuate the noise of ringing artifacts and changes sample intensity of some arrears a picture.

2.2.6. H.264 and HEVC/H.265 standard comparison

Totally, HEVC is an advanced video coding technique which can achieve higher quality by comparing to the previous video coding standard with the same bit rates. Table 5.1 shows the differential between this new advanced video coding standard with the previous one, H.264.

Table 2.1. Comparison between H.264 and HEVC/H.265 standard

	HEVC	H.264/AVC
Partition Size	Flexible block partition from 64x64 to 8x8 by using Quad-tree coding	Macro-Blocks structure with maximum size of 16x16
Partitioning	Flexible partitioning in Prediction Unit Quad-tree down to 4x4 square, symmetric and asymmetric (only square for intra mode)	Sub-block down to 4x4
Transform	Integer DCT from 32x32 to 4x4 + DST Luma Intra 4x4	IDCT 8x8, 4x4
Intra prediction	35 directional modes	9 directional modes
Motion prediction	Advanced Motion Neighbor (3 blocks) Vector Prediction: Spatial + temporal	Spatial Median (3 blocks)
Motion-copy mode	Merge mode	Direct mode
Motion precision	¼ Pixel 7or 8 tap, ⅛ Pixel 4-tap chroma	½ Pixel 6-tap, ¼ Pixel bi-linear
Entropy coding	CABAC	CABAC and CAVLC
Filters	Deblocking Filter + Sample Adaptive Offset	Deblocking Filter

2.3. Rate Control Algorithm Overview in Video Coding

Rate control is a necessary module, which is used to control bit allocation to achieve the given bit budget after encoding process and also minimize distortion rate to get higher quality performance after decoding process. In general, there are two main objectives to discuss in rate control; they are bit allocation and quantization parameter

(QP) computation. In the bit allocation part, the bit budgets has to be generated carefully to assign into each coding level such as group of pictures (GOP) level, picture level, and basic unit level to control bits overflow. In addition to achieve the target bitrate, QP is taken into account because it has higher correlation of assigning bits. If QP is large, bit allocation will be less. In order to vary QP automatically, Rate-Distortion (R-D) performance has been considered as the prior knowledge to generate a function related to QP.

There are several rate control algorithms coming up where Q-domain rate control is the first proposed algorithm in [8] for MPEG video coding standard, then rho-domain rate control algorithm is defined in [9] for H.264/AVC, after that λ-domain rate control is come up in [10] for the current video coding standard, H.265/HEVC. This first rate control has been focused on bit rate and quantization parameter to model a rate-quantization function through R-Q curves resemble the R-D curves of Gaussian random variables. The rate-quantization modelling is determined by,

$$R = \alpha \cdot Q^{-1} + \beta \cdot Q^{-2} \tag{2.1}$$

,where R is target bitrate, Q is quantization parameter, α and β are the coefficient related with video content. This model is also called as quadratic rate-quantizer (R-D) model.

The new rate control algorithm is revealed, which focuses on bit rate and the percentage of the zeros among the quantized transform coefficient (ρ) where ρ and QP have one-to-one correspondence. This new proposed rate control is also called ρ-domain rate control algorithm which uses to achieve smaller bit rate estimation error. This model is computed by,

$$R = \theta \times (1 - \rho) \tag{2.2}$$

,where θ is a coefficient related to the video content.

As abovementioned, QP is the detracting factor which is considered in rate control model, Q-domain rate control and rho-domain rate control. QP is the only parameter which has higher effective to picture quality performance when other parameters are fixed. New rate control is publicized with latest video coding standard, HEVC, to have high flexibility in various video contents in various applications. This new rate control is called R-lambda rate control [10]. There are two flexible steps, computing a model λ of relationship between picture qualities with bitrate and analyzing QP by using λ. For first step, Hyperbolic R-D model is defined to compute λ related to bitrate R as the following Equation (2.4),

$$D(R) = CR^{-K} \tag{2.3}$$

$$\lambda = -\frac{\partial D}{\partial R} = CK \cdot R^{-K-1} = \alpha \cdot R^{\beta} \tag{2.4}$$

,where C and K are coefficient related to source characteristics. From Equation (2.4), lambda can be simplified to calculate within the bit per pixel (*bpp*) by,

$$\lambda = \alpha \cdot bpp^{\beta} \tag{2.5}$$

Next step, QP can be determined by using the Equation (2.6),

$$QP = 4.2005\, ln\,\lambda + 13.7122 \tag{2.6}$$

After encoding procedure in each frame or a CTU, all coefficients need to be updated. α and β values are updated following actual generated bits, QP value and λ value by using Equation (2.6) to (2.9).

$$\lambda_{comp} = \alpha_{old} \cdot bpp_{real}^{\beta_{old}} \tag{2.7}$$

$$\alpha_{new} = \alpha_{old} + \delta_{\alpha} \cdot (ln\,\lambda_{real} - ln\,\lambda_{comp}) \cdot \alpha_{old} \tag{2.8}$$

$$\beta_{new} = \beta_{old} + \delta_{\beta} \cdot (\ln \lambda_{real} - \ln \lambda_{comp}) \cdot \ln bpp_{real} \tag{2.9}$$

,where bpp_{real} is calculated from actual generated bits, α_{old} and β_{old} are α and β values used in coded frame, $\delta_{\alpha} = 0.1$ and $\delta_{\beta} = 0.05$.

Bit allocation proceeding including GOP level, picture level and largest coding unit level are assigned. In case of GOP level bit allocation, the target bits in a GOP can be computed by,

$$T_{AvgPic} = \frac{R_{PicAvg} \cdot (N_{coded} + SW) - R_{coded}}{SW} \tag{2.10}$$

$$T_{GOP} = T_{AvgPic} \cdot N_{GOP} \tag{2.11}$$

,where T_{AvgPic} is average target bit per picture, $R_{PicAvg} = {Target\_Bitrate}/{frame\ rate}$ is average target bit per picture, N_{coded} is number of picture already been code, R_{coded} is bit cost on the picture already been coded, N_{GOP} is the number of picture in current GOP, SW is additional number (SW = 40) and T_{GOP} is target bits for current GOP. For picture level, bit budget can be assigned in Equation (2.12).

$$T_{CurrPic} = \frac{T_{GOP} - Coded_{GOP}}{\sum_{NotCodedPictures} \omega p_i} \cdot \omega p_{CurrPic} \quad (2.12)$$

$$T_{CurrLCU} = \frac{T_{CurrPic} - Bit_{header} - Coded_{Pic}}{\sum_{NotCodedLCUs} \omega p_i} \cdot \omega p_{CurrLCU} \quad (2.13)$$

,where $T_{CurrPic}$ is target bit budget for current picture, $Coded_{GOP}$ is bits budgets for coded frames in current GOP, and $\omega p_{CurrPic}$ is the weight of each picture. The weight value depends on the position of picture in coding structure. In LCU level, suppose Bit_{header} is the estimated bits of all headers, $\omega p_{CurrLCU}$ is the weight of each LCU, and $Coded_{Pic}$ is generated bits for coded LCUs in current picture. Hence, target bit of each LCU is calculated by Equation (2.13).

2.4. Scale Invariant Feature Transform (SIFT)

SIFT is a method to figure out the interest point (Keypoints) in an image which can be strong to all sorts of image transform like: scale, rotation and illumination. It is introduced by D. Lowe in [11]. Four main procedures are deserved in this method.

- *Scale-space extrema detection* is the first start point of the whole process. First of all, Laplacian of Gaussian ($L(x,y,\sigma)$) is computed by convoluting an image ($I(x,y)$) with a variable-scale Gaussian to produce the set of scale space images shown on the left side in Figure 2.6. Then the Gaussian differential (DoG) function, $D(x,y,\sigma)$ is extracted by doing differential of two nearby scales of Laplacian of Gaussian as shown on the right side in Figure 2.6. The equations are expressed in Equations (2.14) and (2.15).

$$G(x,y,\sigma) = \frac{1}{2\pi\sigma^2} e^{-(x^2+y^2)/2\sigma^2} \quad (2.14)$$

$$D(x,y,\sigma) = L(x,y,k\sigma) - L(x,y,\sigma) \quad (2.15)$$

$$= \left(G(x,y,k\sigma) - G(x,y,\sigma)\right) * I(x,y)$$

where, *k* is a constant multiplicative factor. After doing DoG, the local maxima and minima of DoG are determined by doing comparison each sample point to its eight neighbors in the current image and nine neighbors in the scale above and below image as shown in Figure 2.7.

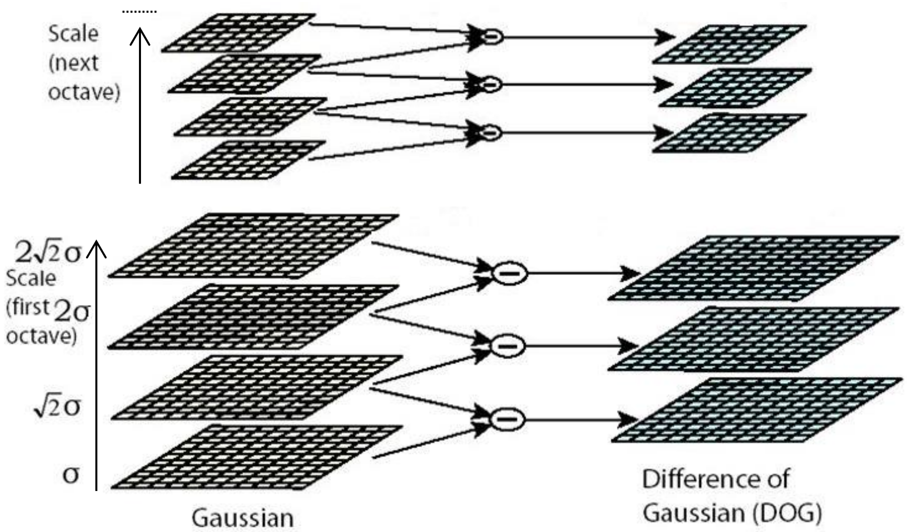


Figure 2.6. Scale-normalized Laplacian of Gaussian by DoG [12]

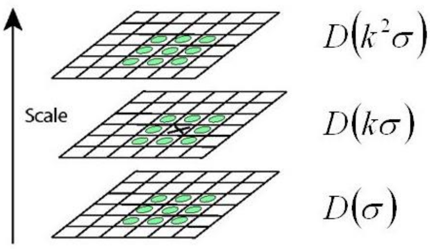


Figure 2.7. Maxima and minima of DoG [12]

- *Keypoint localization* is the second stage which determines location and scale for each interest point, and also eliminates weak keypoints including the points with low contrast and the poorly localized along an edge. To eliminate the low contrast sample point, Taylor expansion is used to simplify scale-space function, $D(x,y,\sigma)$, at the sample point,

$$D(X) = D + \frac{\partial D^T}{\partial X} X + \frac{1}{2} X^T \frac{\partial^2 D}{\partial X^2} X \tag{2.16}$$

,where $X = (x,y,\sigma)^T$ is the offset from the sample point. Then taking derivative of Equation (2.16) with respect to X and setting it to zero to define the location of the extremum, $\hat{X}$,

$$\hat{X} = -\frac{\partial^2 D^{-1}}{\partial X^2} \cdot \frac{\partial D}{\partial X} \tag{2.17}$$

If this offset $\hat{X}$ is greater than 0.5, then keypoint should be in a different sample point. Additionally, a function at the extemum for rejecting unstable extrema with low contrast can be clarified by substituting Equations (2.17) into (2.16), giving,

$$D(\hat{X}) = D + \frac{1}{2}\frac{\partial D^T}{\partial X}\hat{X} \tag{2.18}$$

Threshold is assigned on $|D(\hat{X})|$ to discard the sample point which has low contrast. To eliminate the poorly localized along an edge, Hessian matrix, H, is computed at the location and scale of the sample point (keypoint) by Equation (2.19). Then, trace of H and its determinant are determined. Let r be the ratio between the largest and the smaller magnitude eigenvalue of H, so $r = \lambda_1 / \lambda_2$. After that, another threshold r is set to remove out the poorly localized along an edge as shown in Equation (2.20).

$$H = \begin{bmatrix} D_{xx} & D_{xy} \\ D_{xy} & D_{yy} \end{bmatrix} \tag{2.19}$$

$$\frac{Trace(H)}{|H|} \langle \frac{(r+1)^2}{r} \tag{2.20}$$

- *Orientation assignment*: each keypoint can be assigned one or more orientations. This orientation is also able to involve in keypoint descriptor to achieve invariance to image rotation. The gradient magnitude, $m(x,y)$, and orientation, $\Theta(x,y)$, of keypoints from the selected Gaussian smoothed image, L, are computed using pixel differences,

$$m(x,y) = \sqrt{(L(x+1,y) - L(x-1,y))^2 + (L(x,y+1) - L(x,y-1))^2} \tag{2.21}$$

$$\theta(x,y) = tan^{-1}(L(x,y+1) - L(x,y-1) / (L(x+1,y) - L(x-1,y))) \tag{2.22}$$

- *Keypoint descriptor*: it is formed by computing the gradiant magnitude and orientation at each image sample point in a region around the keypoint location, as shown on the left side to the right side in Figure 2.8. The actual implementation uses 4x4 descriptors from 16x16 sample array, which leads in feature vectors with 128 dimensions.

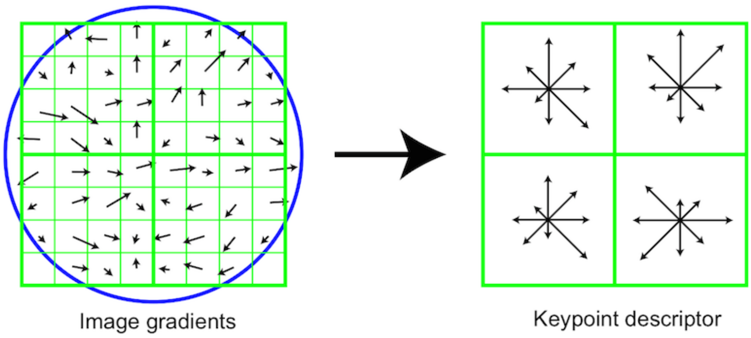


Figure 2.8. Keypoint descriptor with 128 elements vector

Totally, there are several advantages of SIFT such as:

- *Locality*: features are local, so robust to occlusion and clutter
- *Distinctiveness:* individual features can be matched to a large database of objects
- *Quantity:* many features can be generated for even small objects
- *Efficiency:* close to real-time performance

2.5. Random Sample Consensus

Random Sample Consensus (RANSAC) [12] is an algorithm which can be used to find the inliers for any type of model fitting. It is mostly applied to cancel the error of local feature detectors which often make errors. Normally, the errors are occurred in local feature detectors categorized into two types, measurement errors and classification errors. Measurement errors, they occur because of miscalculation feature itself. Generally, measurement errors form as normal distribution then they can completely remove by applying smoothing approach. Classification errors, they occurs when a feature detects a portion of an image as an existence of a feature and it has larger effect then measurement errors which could not average out. The algorithm is described following step:

a) Random selecting a sample in small range data points from the sample dataset (S) and instantiate the model from this dataset.

b) Calculate the distance in each data point of sample selection of S then determine the model based on distance threshold, and these samples selection is the consensus set of the sample and also defines the inliers of S.

c) If the size of sample selection, number of inliers, is greater than some threshold T, the model will estimate again by using the new sample data point selection.

d) After N trails, the largest consensus set of sample selection is selected, and the model is generated based on its points.

The decision to stop selecting subsets can be defined by calculating the probability of the number of trials that required selecting a subset of *N* good data points. Suppose ε is the probability of an outlier then $(1-\varepsilon)$ is the probability of an inlier. The probability that a sample of size *n* which contains all inliers is $(1-\varepsilon)^n$, so the probability of getting at least one outlier in that sample is $1-(1-\varepsilon)^n$. Finally, the probability of all samples *N* that have at least one outlier can be determined as $P_{alloutliers} = \left(1-(1-\varepsilon)^n\right)^N$. In conclusion, the probability that at least one sample has no outliers is $P_{noalloutliers} = 1-\left(1-(1-\varepsilon)^n\right)^N$ then the number of *N* can be estimated by this probability:

$$N = \frac{\log(1-P)}{\log\left(1-\left(1-(1-\varepsilon)^n\right)\right)} \quad (2.23)$$

2.6. Feature preservation methods

There are some previous works which focus on the feature preservation in image and video after compression. Those features preservation techniques can be considered into three main approaches.

a. Features are compressed directly from client side and transmitted to sever side

b. Descriptors are compressed and transmit or stored these descriptors for further processing

c. Modifying image or video compression standard to maintain the features after decompression based on the characteristic of features in uncompressed image or video.

2.6.1. First approach of feature preservation method

Generally, visual feature extraction has two main components: detector and descriptor. Coding visual features extracted from video sequences is proposed in [13] by analysing the detector and compressing it in the context of visual sensor networks which can be involved in image content including image retrieval, object tracking, etc. In addition to get high coding efficiency, they have designed a coding architecture designed for local features extracted from video sequences as show in Figure 2.9 where the visual feature is encoded other side from image or video coding standard.

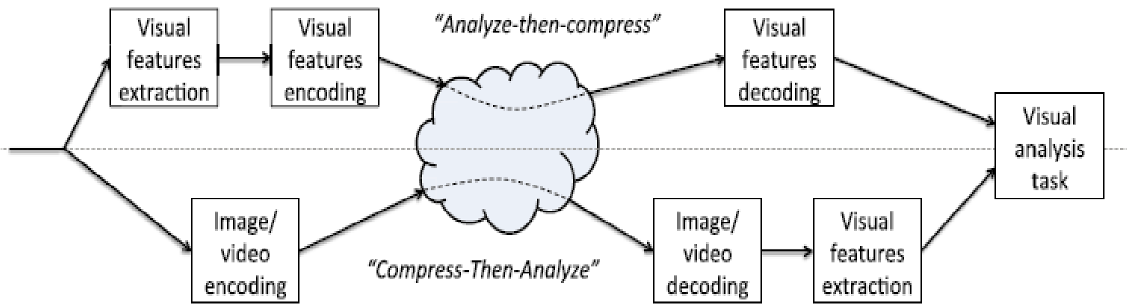


Figure 2.9. Analyze-Then-Compres & Compress-Then-Analyze paradigms piplines [13]

There is also recently work which is proposed on sending detector as the side information, "keypoint encoding and transmission for improved feature extraction from compressed video at low bitrates [14]". In this scenario, they have tried to send keypoint as side information to server side by using low bit rates. So, they analysed frame into four types of scene (same scene, scene updating, new scene and moving quickly scene) which can reduce much amount of bits to assign to keypoints. Moreover, keypoints itself can also reduce by predicting keypoint in current frame from previous frame. The system of this proposed algorithm is shown in Figure 2.10.

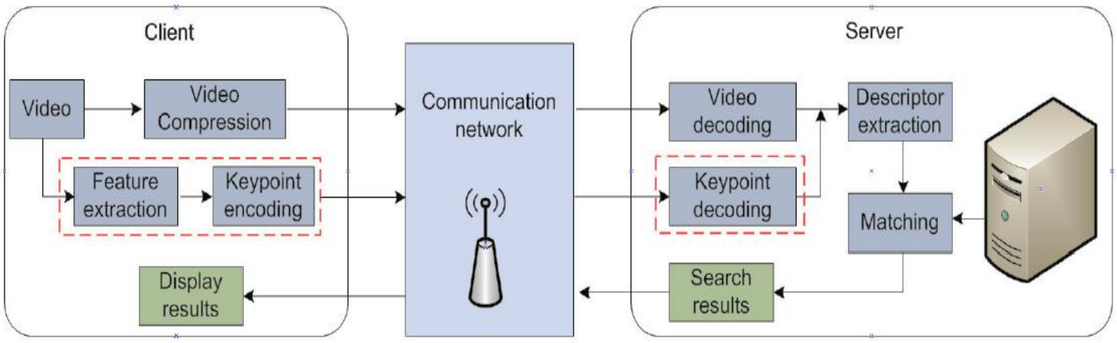


Figure 2.10. Overview system of sending keypoint as side information [14]

2.6.2. Second approach of feature preservation method

Maker et al. [15] proposed a technique to compress descriptors for resident feature extraction. They extracted the descriptors with the detected features by using SIFT detector in [11], then Adaptive Block-size Discrete Cosine Transform is used to compress those descriptors and send them as shown in Figure 2.11. After sending to the server side, descriptors are computed to check the performance of patch compression by checking the correct matching features with the database images.

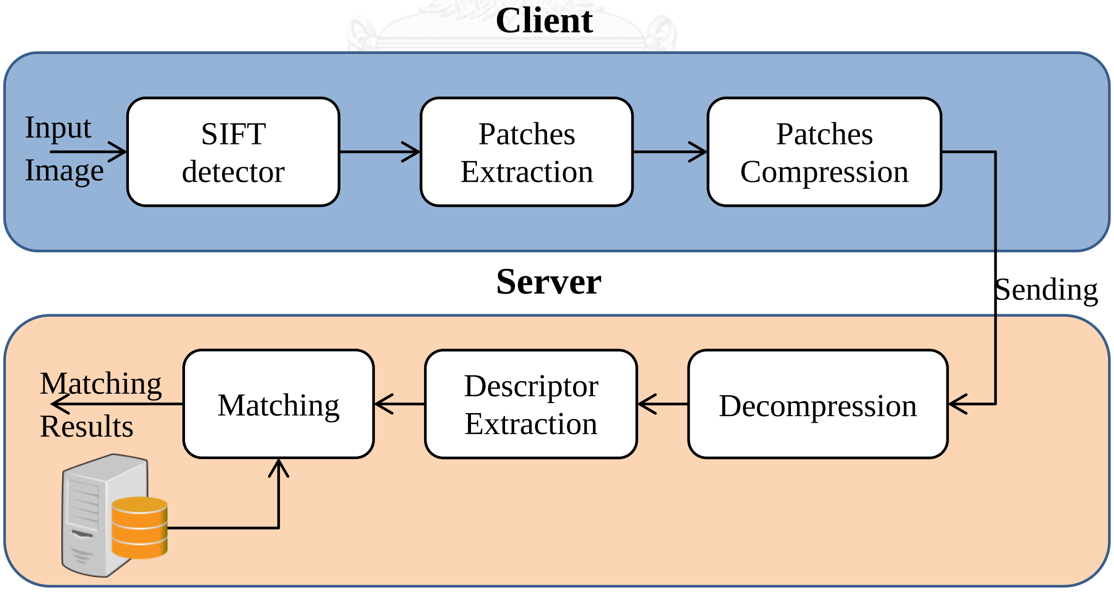


Figure 2.11. Image patch compression for resident feature extraction

Xiang et al. [16] projected a framework to compress descriptor by applying multiple hypothesis prediction to have more effectively removing redundancies from spatial and also temporal. They also apply rate-accuracy optimization technique to achieve high efficient retrieval with low bit rate. The architecture is built up as shown in Figure 2.12.

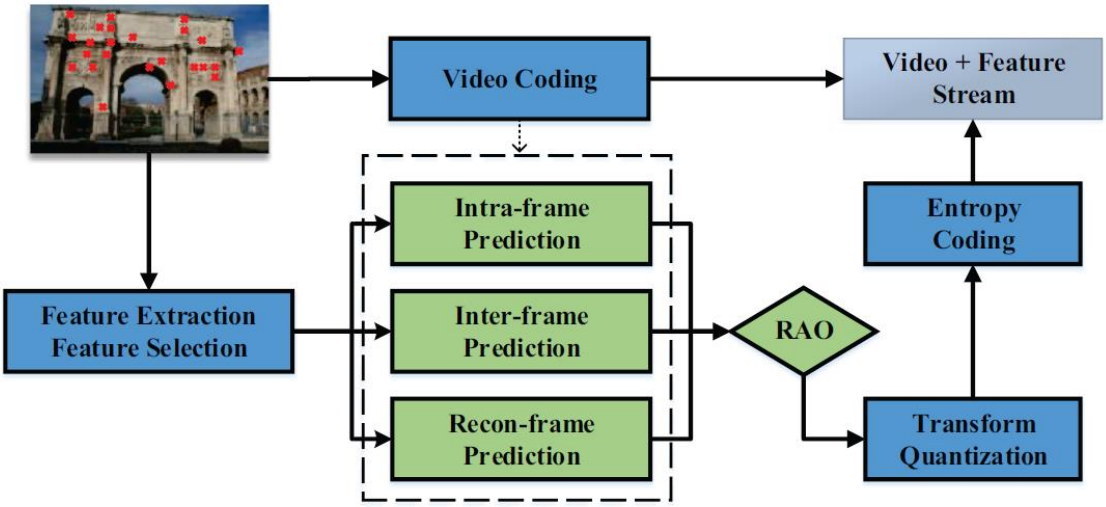


Figure 2.12. Local feature descriptor architecture representation [16]

2.6.3. Third approach of feature preservation method

Chao and Steinbach [17] proposed a new algorithm to preserve strongest SIFT features in standard image compression, JPEG-encoded image. In their work, they projected their algorithm into two approaches to allocate the bit budget in order to control the encoding process and maintain the important/relevant features at low bit rate. The first approach of their methods, they allocated more bits to SIFT detector block areas and lower bits to other blocks, which can successfully improve the feature preservation by comparing with the traditional image compression. The second approach is to apply rate distortion optimization by assuming that the matching score has a relationship with discrete cosine transform coefficient quantizer. They defined the distortion metric to be equal one minus matching score (1 – *matching score)*, and the rate distortion optimization cost function is $J = D + \lambda \times R$, where D is distortion metric, λ is Lagrangian multiplier coefficient and R is bit rate. This second approach gets the better result than the first approach.

After the improving on the image compression, Chao and Steinbach proposed SIFT feature maintain in H.264/AVC video compression in [18]. They applied similar methods from their previous work on image compression. Two bit allocation approaches and rate distortion optimization with matching score are still achieved more feature preservation in H.264/AVC video standard.

In conclusion, these three approaches have their own advantages and disadvantages. For first approach, keypoints are encoded separately from image or video coding standard to transmit or store where user could analyse the image or video by decoding these keypoints and extracting the descriptors from reconstructed image or video. However, if the reconstructed image or video has low quality or the object in the scene could not reconstructed, this approach could not work well because of the characteristic of keypoints are known, but descriptor could not extract from the scene. In addition, extra bits is needed to send these information which requires more capacity. For second approach is not different much from first approach, the researchers have figured out the important of encoding descriptor directly where the reconstructed image or video does not require anymore. That mean, if the users would like to analyse the image or video based on the patch of image (descriptor), they just decode the descriptor back. Nevertheless, bit allocation for these descriptor is higher than by encoding only keypoints where the descriptor is represented 16x16 blocks. High bandwidth is required in this case. Both first approach and second approach are not compactable standard. As the permission of modification image or video encoder, the last approach is come up by modifying the bit budget of macroblocks which contain the interest points or keypoints to guarantee the features are maintained after reconstructed image or video. In addition, extra bit budget is not required anymore in this last approach.

In this work, features preservation method is categorized in the last approach where it is compactible standard and has no extra bit budget. The latest video coding standard encoder, H.265/HEVC, is modified. The largest coding unit (LCU) is branded into two groups which are importance group and non-importance groups based on the number keypoints in each LCUs. Then quantization parameters (QPs) values are computed by using target bit from each group. To know the features are maintained or not, SIFT similarity algorithm is applied to determine the percentage between original features count and the decoded features count.

CHAPTER 3
PROPOSED METHOD TO IMPROVE FEATURE PRESERVATION IN HEVC

This chapter is separated into three main parts. The first part focuses on overall block diagram of the proposed algorithm. The second part explains on feature analysis. Quantization parameter computation is described after second part. At the last part, summary of proposed algorithm is provided.

3.1. Overall block diagram proposed algorithm

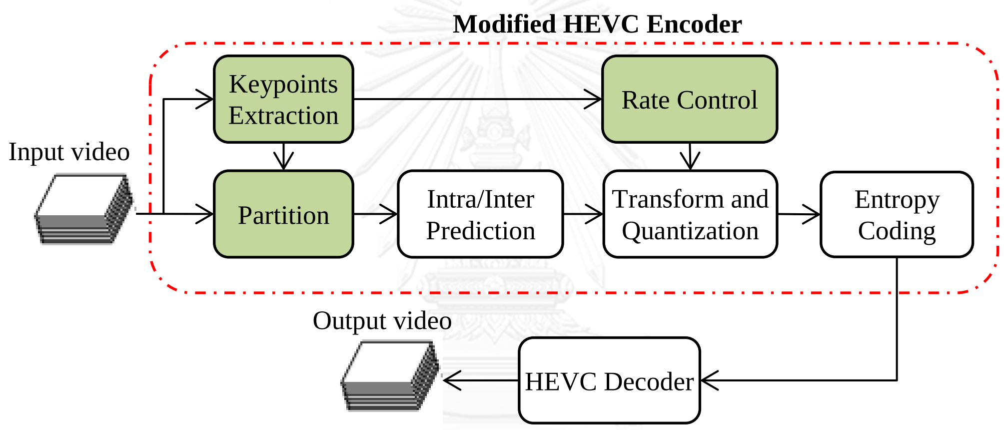


Figure 3.1 Overall proposed algorithm to improve feature preservation in HEVC

The overall block diagram to improve feature preservation in HEVC is shown in Figure 3.1. First of all, the input video goes through splits into CTUs block and keypoints extraction block to preserve the interest points. After searching the interest points, CTUs is considered to categorise into two main groups, important CTUs and non-important CTUs. Since, the proposed algorithm desires to maintain the interested keypoints as much as possible, rate control is required to modify in intra and inter mode to achieve higher keypoints preservation and also maintain target bitrate. The details are described in the following sections.

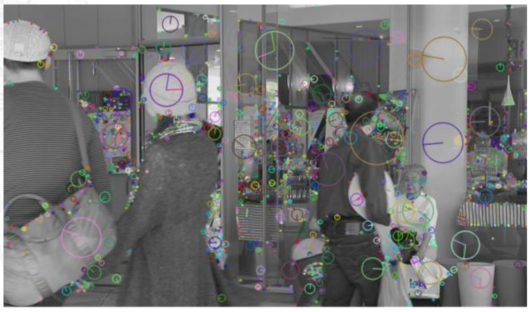
3.2. Feature analysis

At the starting point, the interest point or keypoint is extracted by using SIFT feature detection in this work. To generate keypoint by using SIFT algorithm, there are several parameters need to be set in SIFT detector including sigma value of Gaussian filter, octave levels, contrast threshold and edge threshold. According to [12], these parameters are set in Table 3.1 which can produce high accuracy point to get good keypoint in an image. Figure 3.1 shows a frame after keypoints extraction and also feature map of that frame by looking at the largest coding unit (LCU), 64x64 block sizes.

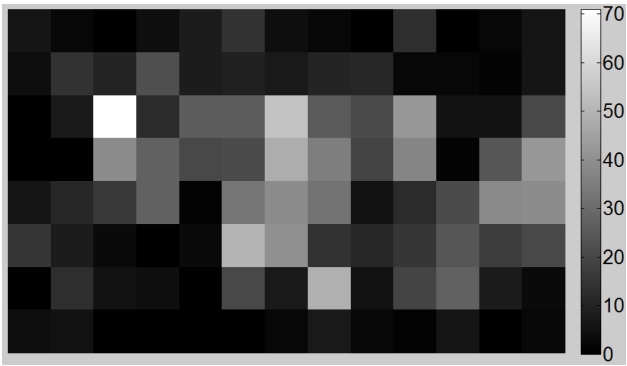
Table 3.1. Parameters setting in SIFT

Parameters Setting
contrastThreshold = 0.04
edgeThreshold = 10
nOctaveLayers = 3
Sigma = 1.6

(a)



(b)



(c)

Figure 3.2. Keypoints extraction from a frame by using SIFT

(a) Original frame, (b) Keypoints extraction frame, (c) Feature map

After features counting in each LCU, most of LCUs contains small number of features and the standard deviation is high as shown in Figure 3.2. So, in this work the highest features in LCUs are considered as the important LCUs (IMLCU) and other LCUs are considered as non-important LCUs (Non-LCU). To category LCU into each group, Mean calculation is defined by taking as the threshold.

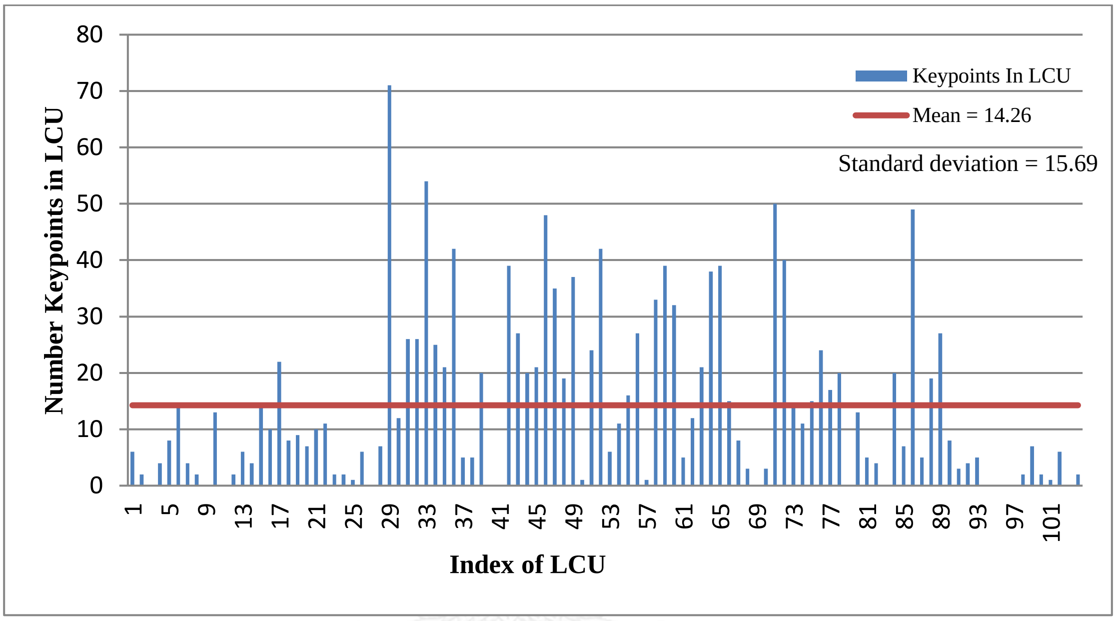


Figure 3.3. Feature count in each LCUs plot

Suppose x_i is the number of keypoints in each LCU and N is the total LCU in a frame, so the expectation $E[X]$ or Mean of keypoints in LCU can define as,

$$E[X] = \frac{1}{N}\sum_i x_i \tag{3.1}$$

Grouping LCU in the following category is finalised according to the Mean as the algorithm below,

$$Group_{LCU} = \begin{cases} 1 & \text{if } x_i > E[X] \\ 0 & otherwise \end{cases} \tag{3.2}$$

where $Group_{LCU}$ = 1 is important LCU (IMLCU), and $Group_{LCU}$ = 0 is non-important LCU (Non-IMLCU).

3.3. Quantization Parameter Computation

Before computing quantization parameter, target bit budget for each LCU group needs to be assigned. In this work, target bit budget for current picture or frame is separated into two different bit budgets as shown in Equation (3.3), T_{IM} and T_{NIM} are the target bit budgets for IMLCU group and Non-IMLCUs group in frame, respectively.

$$T_{CurrPic} = T_{IM} + T_{NIM} \tag{3.3}$$

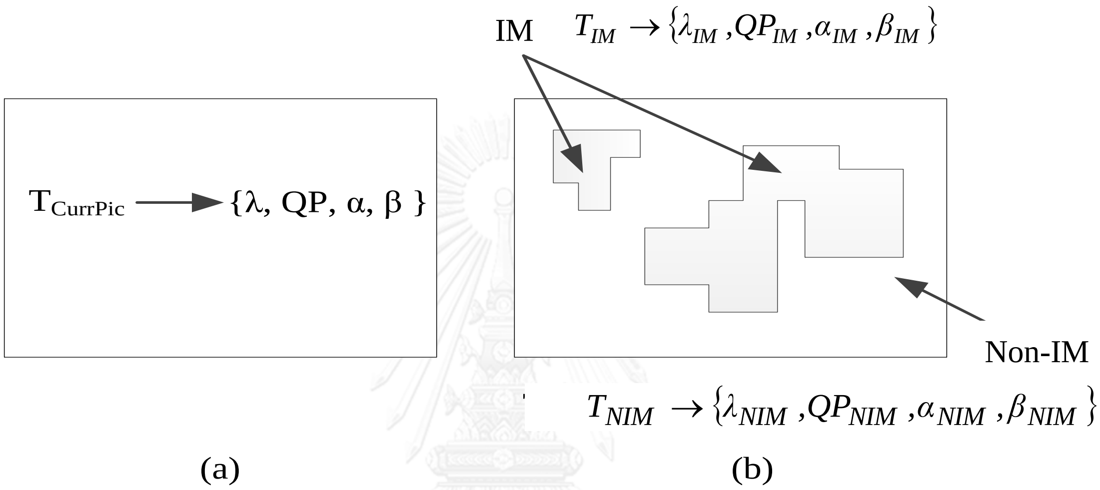


Figure 3.4. Dividing picture into two main area, IMLCUs and Non-IMLCUs
(a) Current frame and original parameters, (b) Divided frame and modified parameters

After assigning target bit to each group, there are several related parameters which are important to compute in HEVC video coding with latest rate control algorithm as shown in Figure 3.4. Those parameters are: α, β, and λ values which are required for each group in a frame to generate QP values. Parameter QP is computed following coding mode in video coding known as Intra mode and Inter mode.

3.3.1. Intra Mode

While Intra mode provides good feature preservation outcomes, it generates plenty of bits which can cause bit instability and decrease the overall performance. To avoid these problems, cost in the first picture need to be regenerated which can reduce some bits from the first picture and only target bit budget of IMLCU need to do bit refinement. In this algorithm, the LCUs in IMLCUs group are encoded by using Intra mode with period of 20 encoded frames per one time.

3.3.1.1. First Picture Coding

In reference software of HEVC (HM15) [6], the lowest Rate Distortion (RD) cost is chosen based on the value of getting Hadamard Absolute Difference (HAD) cost which can be defined as in Equation (3.4),

$$J_{HAD} = SATD + \lambda R \tag{3.4}$$

,where *SATD* is Hadamard sum absolute difference, it can be calculated by Equation (3.5). λ is Lagrangian multiplier (lambda) which can be computed by Equation (3.6) and *R* is the bit rate that is used for encoding current mode. According to [11], there is a relationship between λ and bit budget. If λ is increased, bit budget will be decreased. In this work, *SATD* is computed by looking at only the IMLCU which can reduce amount of bit from first frame and it can be advantage for the next frames.

$$SATD(A_{8X8}) = \sum_{i}^{8}\sum_{j}^{8}\left|t_{ij}\right| \tag{3.5}$$

,where *A* is current block and t_{ij} are the (i, j)th element of following *A*.

$$\lambda = \frac{\alpha}{256} \times \left(\frac{MADPerPixel}{bitsPerPixel}\right)^{\beta_1} \tag{3.6}$$

,where $\alpha = 6.7542$; $\beta_1 = 1.7860$ and,

$$MADPerPixel = \left(\frac{SATD}{TotalPixel}\right)^{1.2517} \tag{3.7}$$

After generating estimated lambda according to the IMLCU, original encoding reference software is used for this first frame.

3.3.1.2. Intra Refresh

In this section, bit budget for the whole basic unit of IMLCU group is defined as in Equation (3.8),

$$T_{IM} = \frac{1}{N_{Pic}} \times \begin{cases} T_{CurrPic} \cdot N_{IM} & \text{, if } N_{IM} < N_{NIM} \\ T_{CurrPic} \cdot N_{NIM} & \text{, otherwise} \end{cases} \tag{3.8}$$

,where N_{Pic} represents the total number of pixels in whole frame, N_{IM} is total pixels of IMLCUs group, and N_{NIM} is total pixels of Non-IMLCUs group. T_{IM} is represented as bit budget of IMLCUs group where bit refinement process is applied on this T_{IM}. So T_{IM} has to be lower than T_{NIM} before doing bit refinement process to avoid bit fluctuation. For T_{NIM} can be defined in Equation (3.9),

$$T_{NIM} = T_{CurrPic} - T_{IM} \tag{3.9}$$

3.3.2. Inter Mode

In contrast to Intra mode, the bit budget for IMLCUs should be selected lower than bit budget for Non-IMLCU; bit budget for IMLCUs has to be higher than the bit budget for Non-IMLCU in Inter mode. As a reason, there is no bit refinement process to use in Inter mode. To maintain the performance of IMLCUs, so the bit budget for each group is defined in Equation (3.10) and (3.11),

$$T_{IM} = a \times T_{CurrPic} \tag{3.10}$$

$$T_{NIM} = T_{CurrPic} - T_{IM} \tag{3.11}$$

,where a is a constant. To maintain the keypoints in IMLCUs and also the quality of the video, a is assigned 60% of total bit budget to IMLCUs group and the rest is for Non-IMLCUs group.

3.3.3. QP computation

Bit budgets assignment is done in Intra and Inter mode in section 3.3.1 and 3.3.2, respectfully. Before calculating QP values, some coefficients are computed based on its group. The coefficients in IMLCU group are denoted as α_{IM} and β_{IM} and those for Non-IMLCU group are symbolized as α_{NIM} and β_{NIM}.

In case that Intra mode is selected, α_{IM} and β_{IM} values of IMLCU group are generated the same values as it assigned for Intra frame. On behalf of Non-IMLCU group, its coefficients are produced by using the same values that computed in for current picture or frame.

$$\alpha_{IM} = \alpha_{Intra} \; ; \; \alpha_{NIM} = \alpha_{current} \tag{3.12}$$

$$\beta_{IM} = \beta_{Intra} \; ; \; \beta_{NIM} = \beta_{current} \tag{3.13}$$

With the condition that Inter mode is selected, the values of α and β are computed for current picture are used for coefficients of both groups, IMLCU and Non-IMLCU, respectfully.

$$\alpha_{IM} = \alpha_{current} \; ; \; \alpha_{NIM} = \alpha_{current} \tag{3.14}$$

$$\beta_{IM} = \beta_{current} \; ; \; \beta_{NIM} = \beta_{current} \tag{3.15}$$

QP value can be determined by using Equation (2.6) from Chapter 2 based on λ value. λ can be defined in each group according to its coefficient as shown below:

$$QP = 4.2005 \ln \lambda + 13.7122 \tag{3.16}$$

$$\lambda = \alpha \cdot bpp^{\beta} \tag{3.17}$$

After encoding process of one frame, the coefficients need to be regenerated. bpp_{real} is calculated from actual generated bits. This bpp_{real} values are separated into two values: bpp_{real_IM} and bpp_{real_NIM}, which are determined by using T_{IM} and T_{NIM}, respectively. $\lambda_{real,}$ which is used to compute QP, also has two values belonging to each LCU group: λ_{real_IM} and λ_{real_NIM} .

$$\lambda_{comp} = \alpha_{old} \times bpp_{real}^{\beta_{old}} \tag{3.18}$$

$$\alpha_{new} = \alpha_{old} + \delta_{\alpha} \times (\ln \lambda_{real} - \ln \lambda_{comp}) \times \alpha_{old} \tag{3.19}$$

$$\beta_{new} = \beta_{old} + \delta_{\beta} \times (\ln \lambda_{real} - \ln \lambda_{comp}) \times \ln bpp_{real} \tag{3.20}$$

where δ_{α} and δ_{β} equal 0.1 and 0.05, respectively.

Updating process is finalized by using (3.18)-(3.20).

3.4. Summary of proposed algorithm

The proposed algorithm can be summarized in Figure 3.5.

(1). Keypoints extraction are stored in feature map, after that feature map is used as an input to category LCUs into each group following, IMLCU and Non-IMLCU where Mean of total keypoints in LCU in current frame is a threshold, if the total keypoints in current LCU is higher than the Mean, then it will assign to be IMLCU, the others is Non-IMLCU.

(2). Coding mode is checked, if Intra mode is used, bit refinement process for IMLCU is executed.

(3). Adjusting parameters in R-lambda rate control are computed based on the bit budget in each group.

(4). Ending coding process, updated parameter process is set up for next frame.

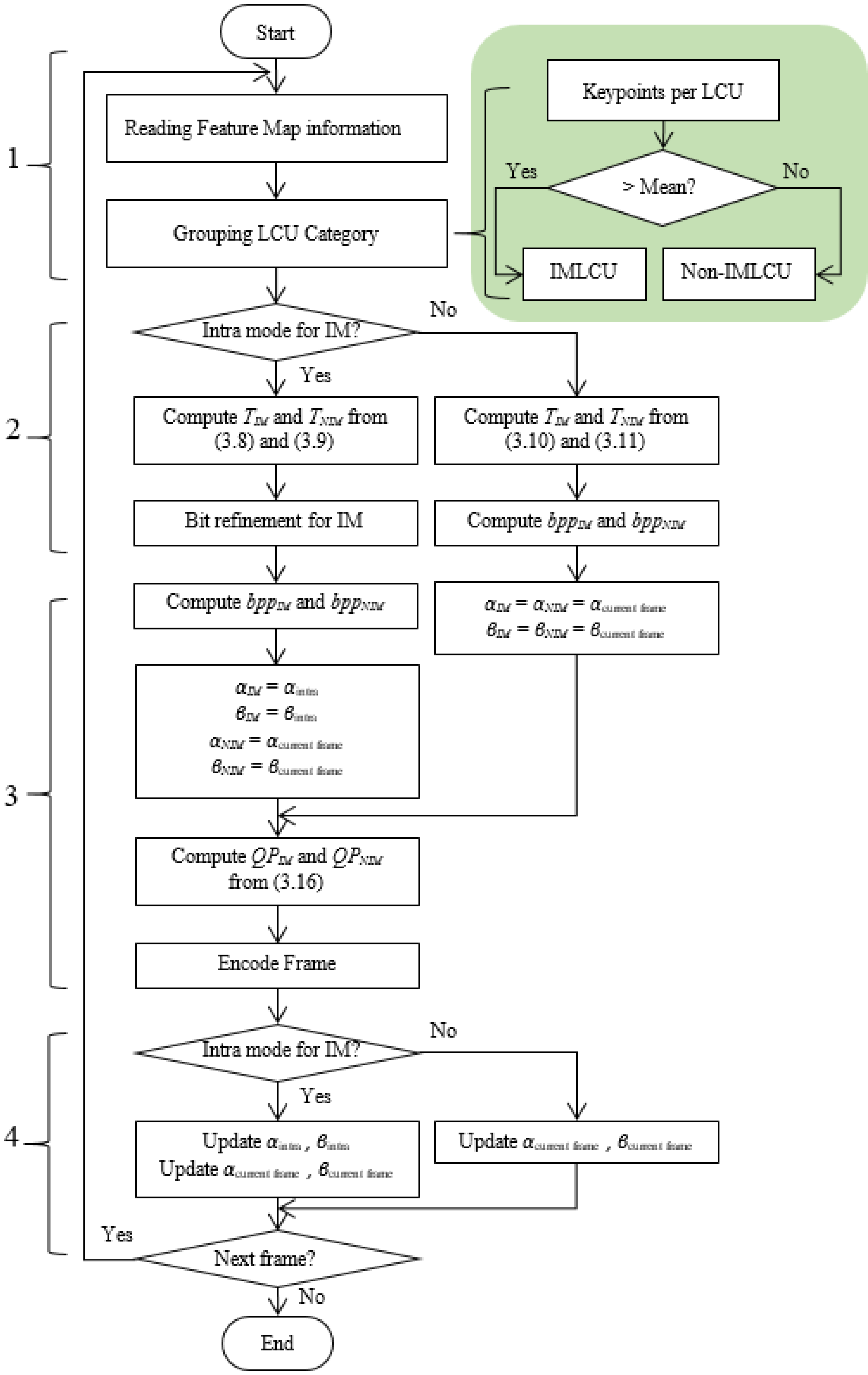


Figure 3.5. Proposed algorithm flow chart

CHAPTER 4
EXPERIMENTAL RESULTS AND DISCUSSIONS

In this chapter, the performance of proposed algorithm is evaluated by comparing with original reference software HEVC (HM15) [6]. There are two main parts presenting in this chapter. Experimental setup is described in the first part, and the second part provides the discussion of experimental results.

4.1. Experimental Setup

4.1.1. Dataset

Four HEVC test video sequences are used in the experiments, BQMall, PartyScene, KristenAndSara and FourPeople as shown Figure 3.6. Table 3.2 shows the detail of parameter in each video sequences.

(a) (b)

(c) (d)

Figure 3.6. Video sequence datasets

(a) BQMall (60fps), (b) PartyScene (50fps),

(c) KristenAndSara (60fps), (d) FourPeople (60fps)

Table 3.2. Parameter detail of each video sequence

Name	Width	Height	Frame rate	Number of frame
BQMall	832	480	60	600
PartyScene	832	480	50	500
KristenAndSara	1280	720	60	600
FourPeople	1280	720	60	600

4.1.2. Parameters Setting

"Lowdelay_P_main configureation" of HEVC is applied in experiments where 4 frames in GOP, largest block size of LCU (64x64) are chosen. The number of LCU per slice is defined in both original HEVC reference software and proposed algorithm, 12 LCUs per slice for 480p sequences and 20 LCUs per slice for 720p sequences. Three different bit rates are used to encode, 256kbps, 512kbps, and 1500kbps. Rate control is also enable and other parameters are set the same as the default in configuration file. The summary can be described in Table 3.3.

Table 3.3. Summary of encoder configurations

Encoder configuration summary	
	HM15 and Proposed algorithm
Profile	main
Maximum LCU size	64x64
Maximum LCU Partition Depth	4
Intra Frame Period	-1 # First frame only
GOP Size	4
Rate Control	Enable

4.1.3. Peak Signal to Noise Ratio Measurement

The quality of reconstructed image or video comparing with raw image or video is computed based on Peak signal to noise ratio (PSNR) measurement. Defining PSNR, it has a close relationship between mean square errors (MSE) where PSNR can be computed as the Equation (3.21),

$$PSNR = 20\log_{10}\left(\frac{2^8-1}{\sqrt{MSE}}\right) \tag{3.21}$$

where $MSE = \frac{\sum(I-R)^2}{N}$: I is raw image, R is reconstructed image, and N is the resolution of image.

4.1.4. SIFT Similarity

SIFT Similarity is used to compare of the features of two different images whether it is same or not. It is computed by Equation (3.22):

$$SIFT\ similarity = \frac{\#\,\text{correctly detected keypoints}}{\#\,\text{original detected keypoints}} \tag{3.22}$$

Correctly detected keypoints are calculated by using distance ratio from the nearest neighbor to the distance of the second nearest which is proposed in [12]. If distance ratio is higher than the threshold (0.8), the match is rejected. Moreover, mismatched feature is still occurred after apply nearest neighbor algorithm. To get high correctly keypoint, RANSAC [13] is applied after nearest neighbor algorithm where it looks on the data points around the original frame with reconstructed frame.

4.2. Results and Discussion

Test sequences are encoded within three bit rates such as 256 kbps, and 512 kbps and 1500 kbps. To evaluate the feature information after decoding process, SIFT Similarity is calculated by looking at the correct match feature in reconstructed video sequences with the original feature in raw video sequences. In addition, the comparison of SIFT Similarity focus only the IMLCU is also discussed. First of all, Figure 3.7 to 3.12 shows reconstructed frame of proposed method and reconstructed frame of original reference software HEVC (HM15). Both 480p and 720p sequences, the objects in the image mostly maintain in proposed method which can get higher keypoints.

(a) Original HM15 (b) Proposed Method

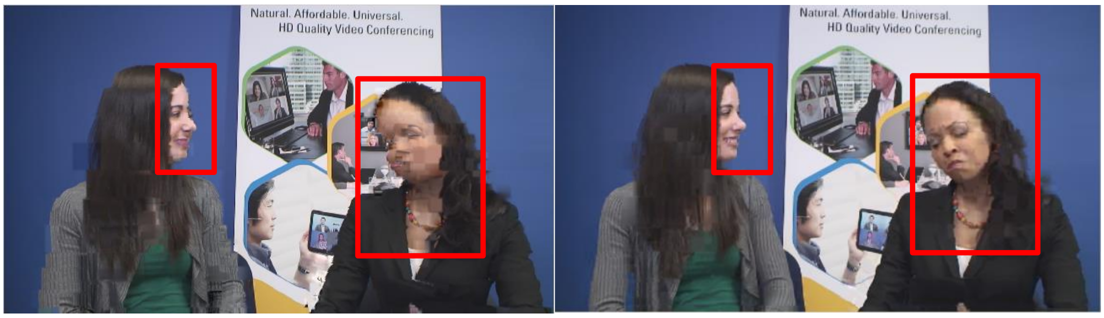
Figure 3.7. Reconstructed frame 160 of BQMall sequence by encoding 256 kbps

(a) Original HM15 (b) Proposed Method

Figure 3.8. Reconstructed frame 160 of BQMall sequence by encoding 512 kbps

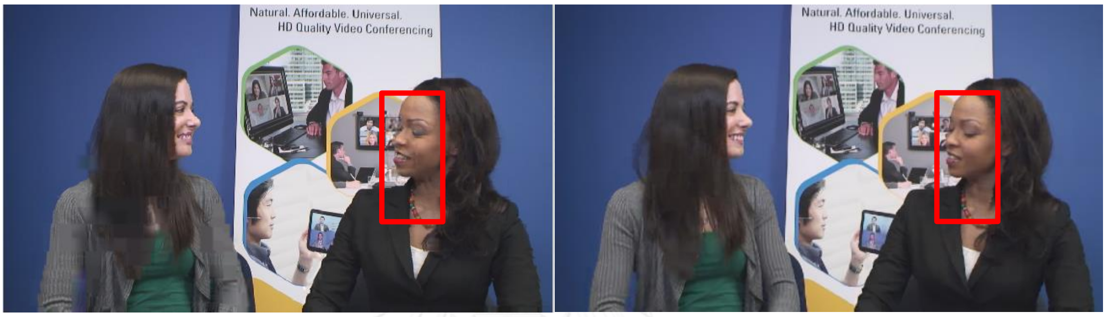
(a) Original HM15 (b) Proposed Method

Figure 3.9. Reconstructed frame 160 of BQMall sequence by encoding 1500 kbps



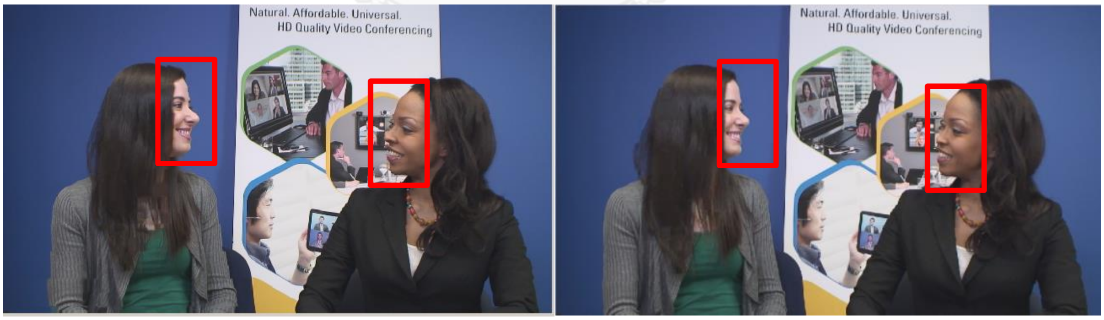
(a) Original HM15 (b) Proposed Method

Figure 3.10. Reconstructed frame 110 of KristenAndSara sequence by encoding 256 kbps



(a) Original HM15 (b) Proposed Method

Figure 3.11. Reconstructed frame 67 of KristenAndSara sequence by encoding 512 kbps



(a) Original HM15 (b) Proposed Method

Figure 3.12. Reconstructed frame 16 of KristenAndSara sequence by encoding 1500 kbps

The comparison of SIFT Similarity in full frame of video sequences are shown in Figure 3. 13. As a results, proposed algorithm can preserve more keypoints than original reference software. It can achieve more than 10% by comparing with HM15. However, the comparison of SIFT Similarity by looking at only IMLCU is also demonstrated in Figure 3.14, where the results is higher than comparing full frame because of the proposed algorithm assigns higher bit to IMLCU more than Non-IMLCU. However, 480p sequences are always getting higher performance than 720p because of the characteristic of source video. In 480p, there are more people in the scene, they also walk and play around the scene. In contrast of 720p where the scene is about broadcasting scene, the moving object in the scene is not much. As the characteristic of R-Lambda rate control, it is updated the parameter based on the previous frame. So, if there are less moving object in the scene, original reference software will get high performance of feature. However, our proposed algorithm still can preserve more feature than original reference software about 1% to 2 %.

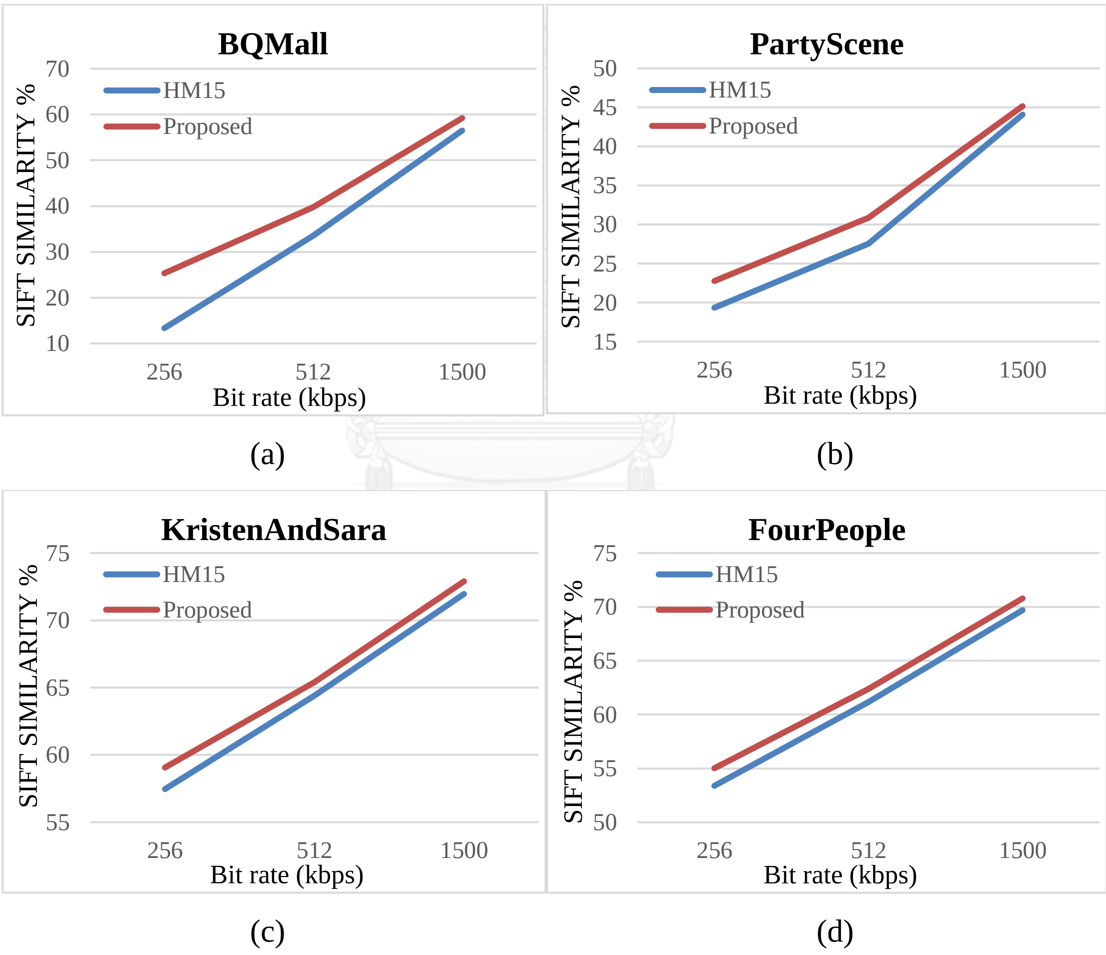


Figure 3.13. SIFT Similarity on both 480p and 720p sequences with full frame

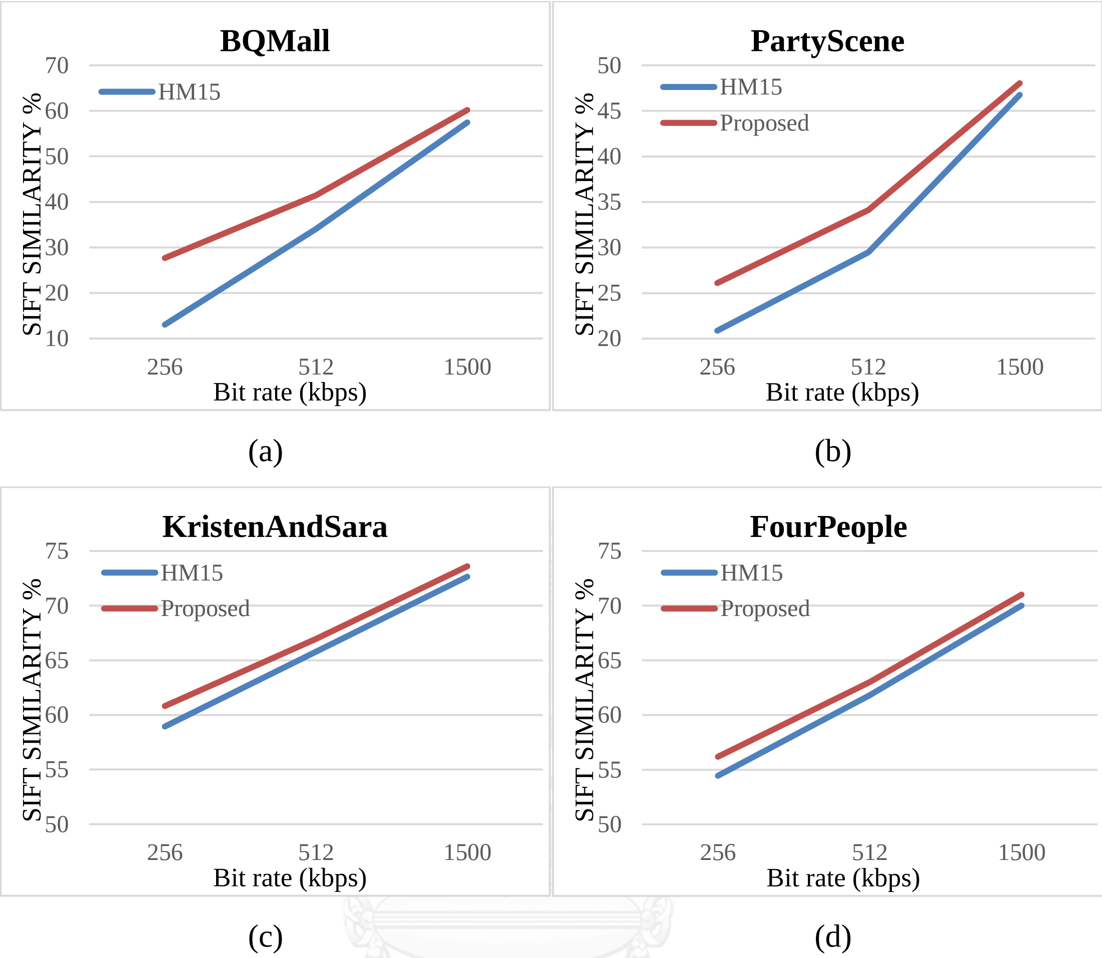


Figure 3.14. SIFT Similarity on both 480p and 720p sequences with only IMLCU

The summary results are shown in Table 3.4 to 3.7 where contains of picture quality and SIFT Similarity both on full frame and on IMLCU.

Table 3.4. Comparison results of proposed method with original HEVC reference software based on PSNR and SIFT Similarity on BQMall test sequences

	PSNR		SIFT Similarity Full Frame (%)		SIFT Similarity IMLCU (%)	
Bit rates	HM15	Proposed	HM15	Proposed	HM15	Proposed
256	24.48	**26.35**	13.33	**25.33**	13.04	**27.68**
512	29.38	**30.52**	33.52	**39.79**	34.07	**41.47**
1500	34.87	**35.56**	56.49	**59.21**	57.46	**60.19**

Table 3.5. Comparison results of proposed method with original HEVC reference software based on PSNR and SIFT Similarity on PartyScene test sequences

	PSNR		SIFT Similarity Full Frame (%)		SIFT Similarity IMLCU (%)	
Bit rates	HM15	Proposed	HM15	Proposed	HM15	Proposed
256	23.3	**23.57**	19.32	**22.77**	20.87	**26.10**
512	25.69	**26.30**	27.54	**30.88**	29.49	**34.12**
1500	30.38	**30.57**	44.08	**45.16**	46.76	**48.04**

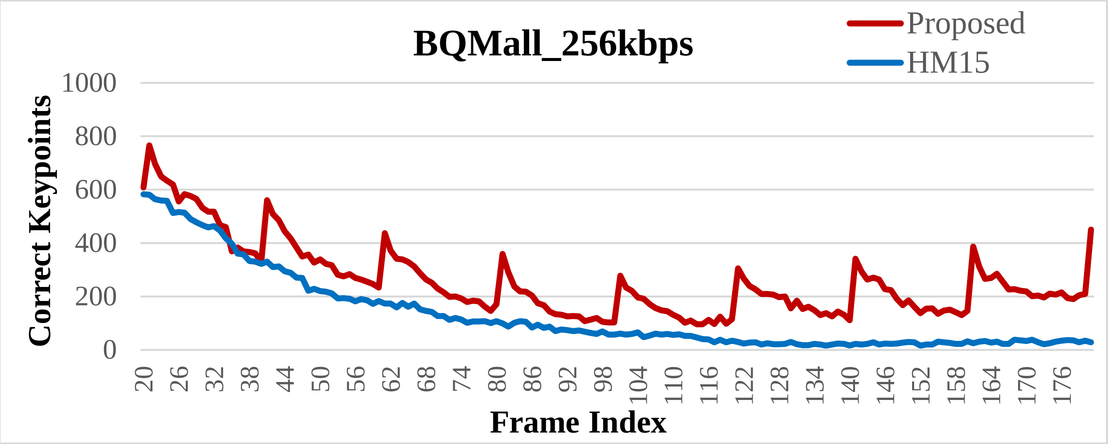
Table 3.6. Comparison results of proposed method with original HEVC reference software based on PSNR and SIFT Similarity on KristenAndSara test sequences

	PSNR		SIFT Similarity Full Frame (%)		SIFT Similarity IMLCU (%)	
Bit rates	HM15	Proposed	HM15	Proposed	HM15	Proposed
256	35.39	**35.64**	57.46	**59.06**	58.94	**60.81**
512	39.02	**39.12**	64.4	**65.4**	65.79	**66.97**
1500	42.08	**42.28**	71.96	**72.9**	72.65	**73.6**

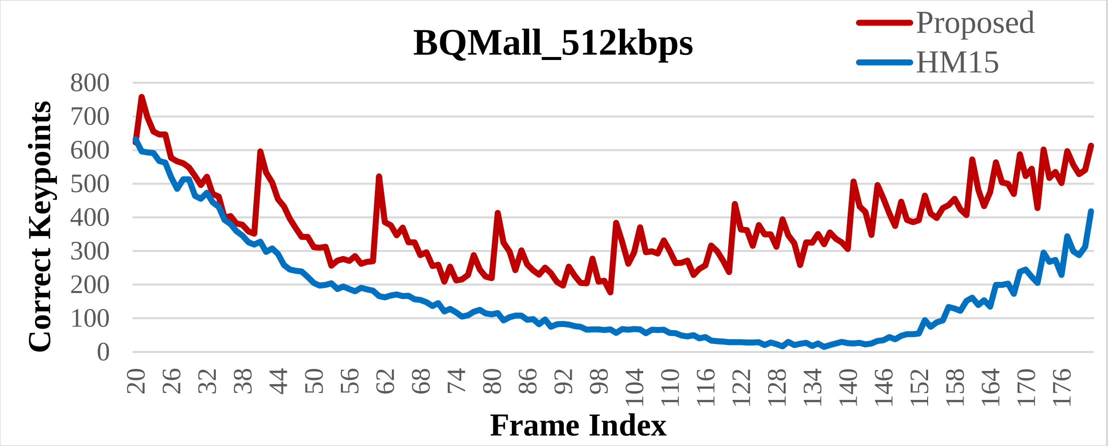
Table 3.7. Comparison results of proposed method with original HEVC reference software based on PSNR and SIFT Similarity on FourPeople test sequences

	PSNR		SIFT Similarity Full Frame (%)		SIFT Similarity IMLCU (%)	
Bit rates	HM15	Proposed	HM15	Proposed	HM15	Proposed
256	32.56	**33.78**	53.39	**55.02**	54.44	**56.18**
512	36.41	**37.72**	61.18	**62.38**	61.82	**63.02**
1500	40.38	**41.15**	69.7	**70.79**	70.01	**71.01**

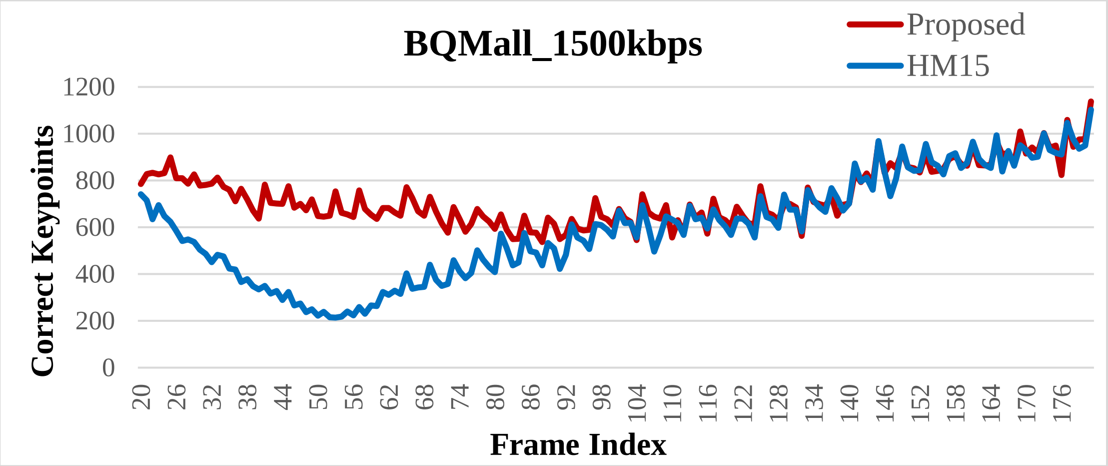
In conclusion based on the summary tables above, the proposed algorithm could achieve higher performance of feature preserving then original HEVC reference software in low bit rate. Moreover, these improvements are also depended on the characteristic of source video signal. If the video contains more moving object, R-Lambda rate control will not work well a long temporal picture. In contrast of proposed method which can achieve in above problem because of R-Lambda can know which area is important and which is not important based on bit allocation to each area. Thus, feature preservation can be maintained. The results of preserving keypoint with selected some frames are also shown in Figure 3.15 to 3.18.



(a)

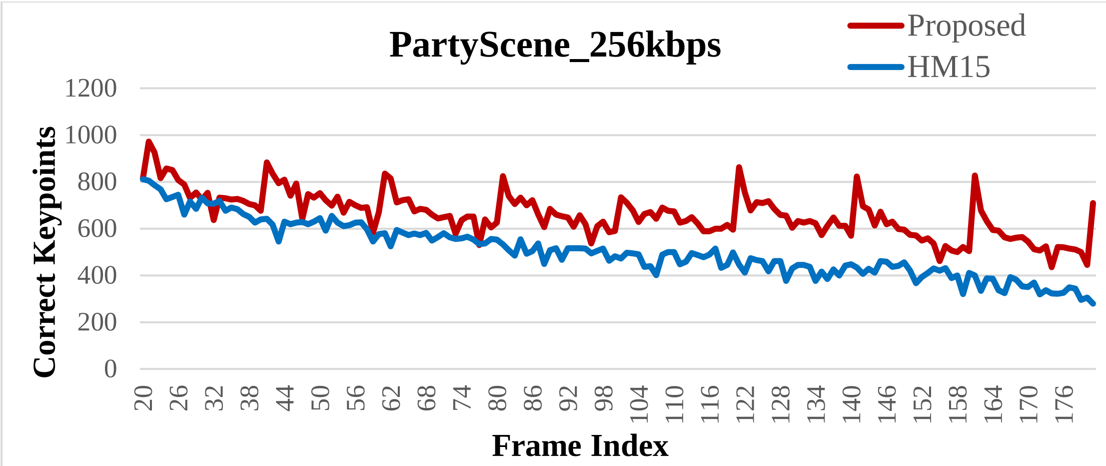


(b)

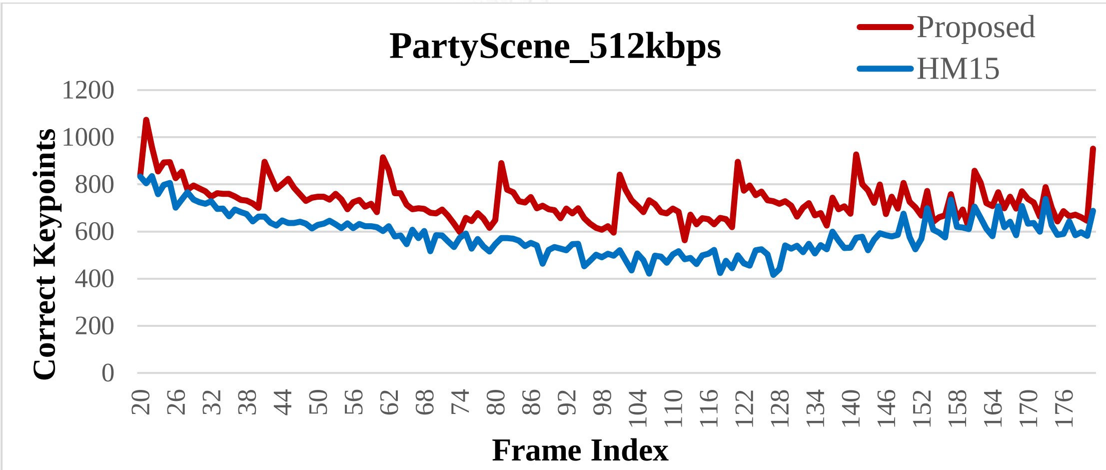


(c)

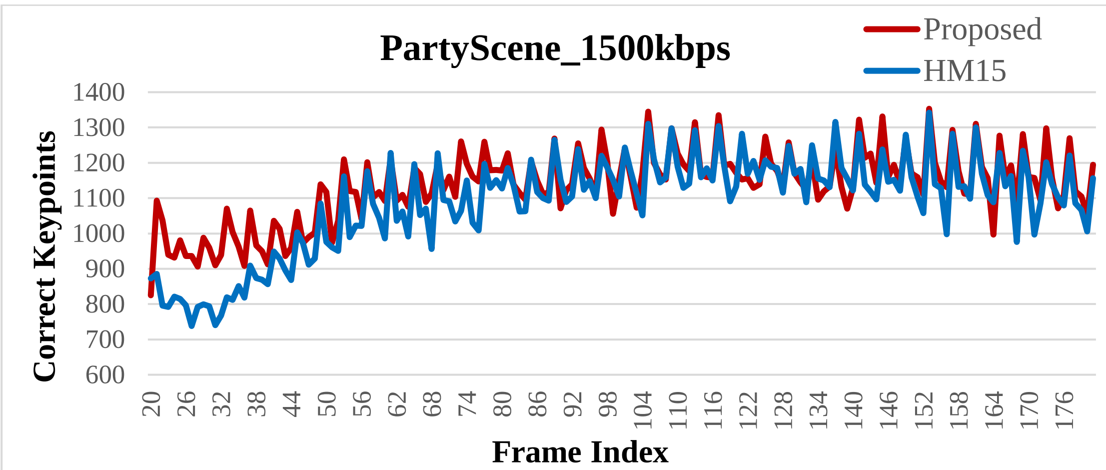
Figure 3.15. Preserving features in some selected frame of BQMall sequence



(a)

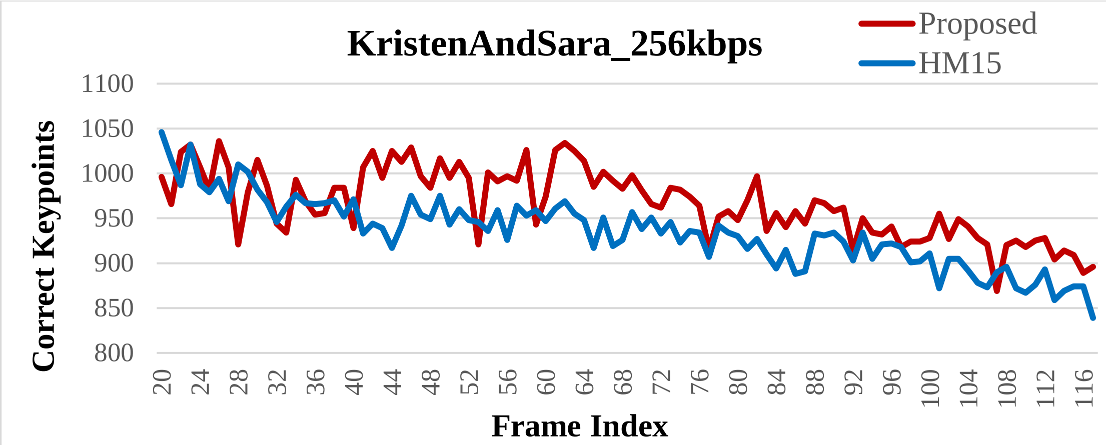


(b)

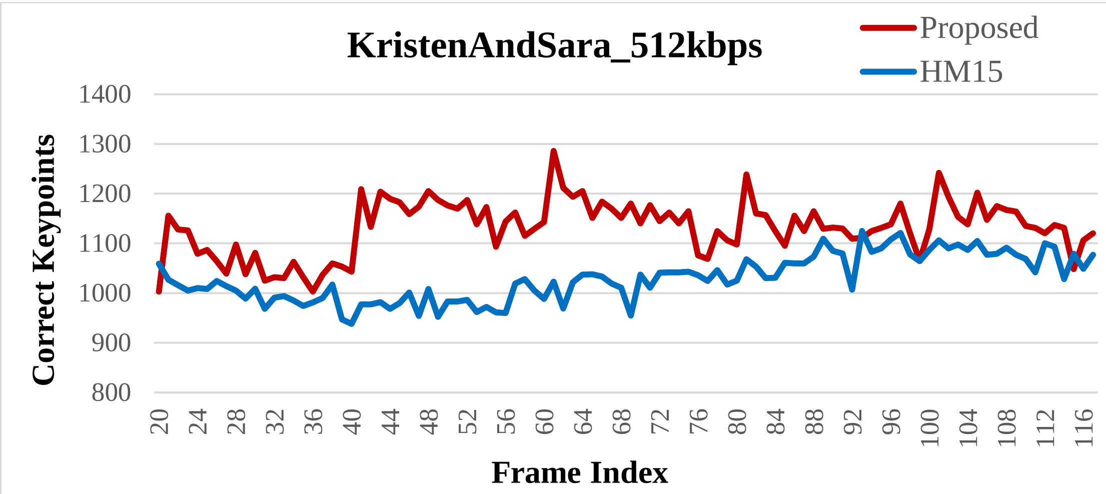


(c)

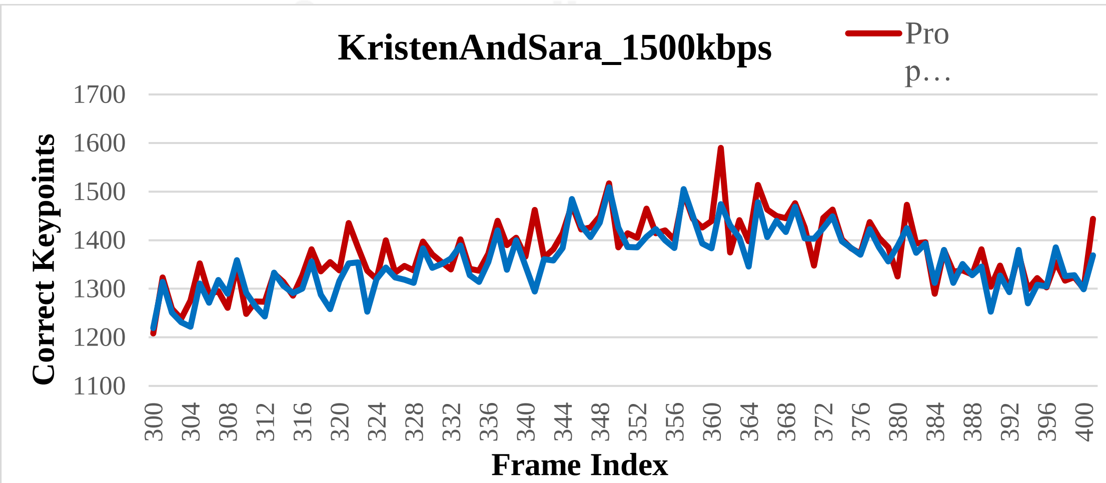
Figure 3.16. Preserving features in some selected frame of PartyScene sequence



(a)

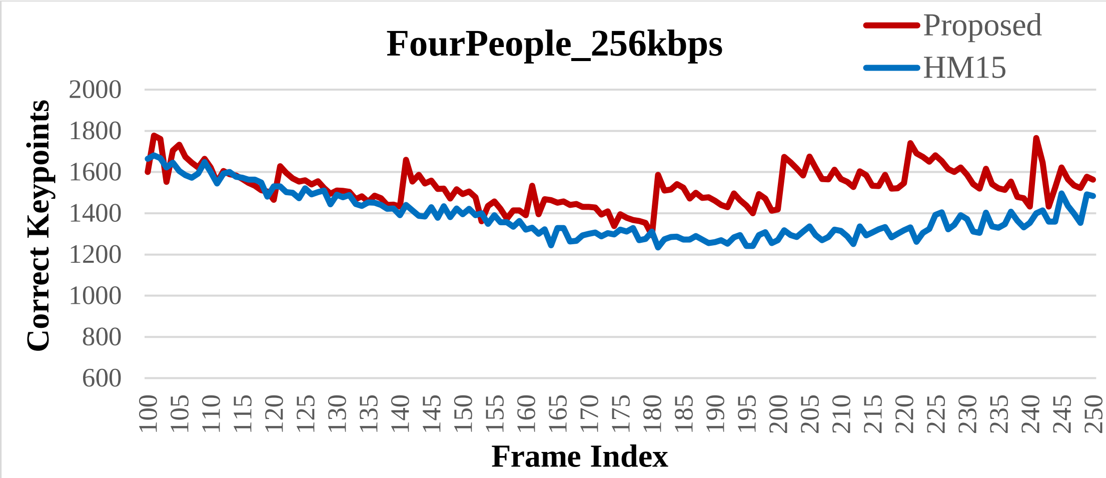


(b)

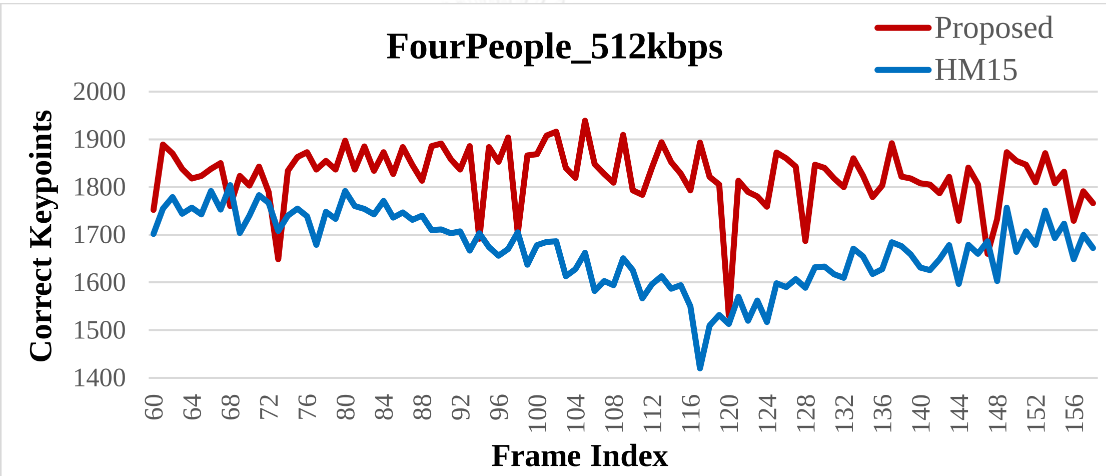


(c)

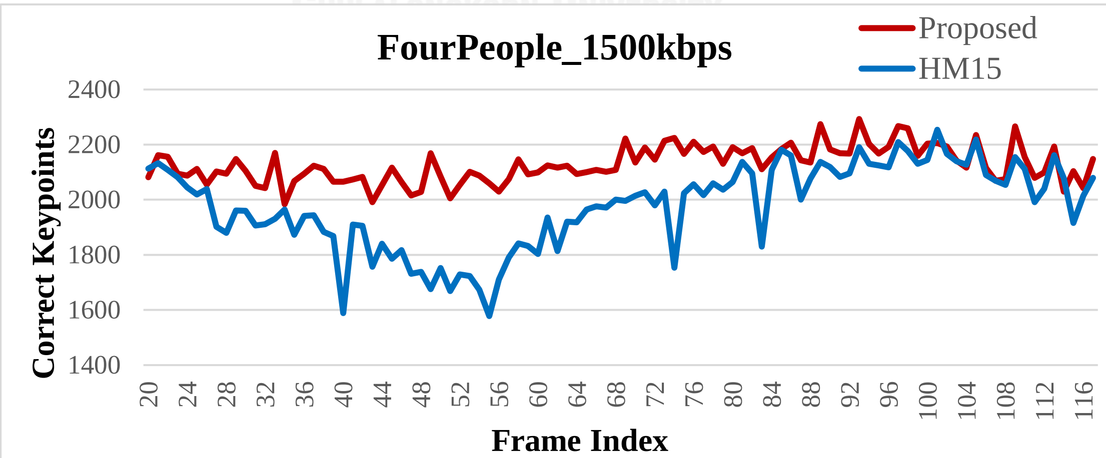
Figure 3.17: Preserving features in some selected frame of KristenAndSara sequence



(a)



(b)



(c)

Figure 3.18. Preserving features in some selected frame of FourPeople sequence

❖ Computational time

Table 3.8 shows the computational time of proposed algorithm and reference software HEVC. As a result, the proposed algorithm takes longer computation than the original reference software. The computation takes longer because of additional parameter to encoder and also modified parameter in rate control, the main parts are described following,

- Extraction keypoints in each LCU
- Mean computation
- Regenerating bit allocation belong to each group
- Doing Intra refresh where refinement bit is defined on IMLCU and cost function is computed
- Adjusting and also updating parameters in rate control are done separately on each group

Table 3.8. Computational time in average per frame of reference software HEVC and proposed algorithm

Bit rate (kbps)	Duration (sec)	
	HM15	Proposed
256	4.31	15.85
512	4.87	17.97
1500	6.25	22.79

This simulation is applied on a computer runs on Ubuntu 16.04 LTS. Specification of this computer is described following,

- Memory: 32 GB
- Processor: Intel® Core™ i7-4770 CPU @ 3.40GHz x 8
- OS type: 64-bit

CHAPTER 5
CONCLUSION AND FUTURE WORKS

Due to the performance of HEVC/H.265 video coding standard at low bit rates, some objects are gone after decoding process which can effect to further purpose in computer vision algorithms such as image retrieval, object tracking, object recognition, and some learning rate algorithms. Thus, the interest points or feature information in image or video are needed to be maintained.

To address this problem, improvement of feature preservation in HEVC/H.265 video coding standard is proposed in this thesis by dividing a frame into two groups, important largest coding units (IMLCU) and non-important largest coding units (Non-IMLCU), where the IMLCU contains higher number of interest points (keypoints) than Non-IMLCU. To consider a LCU is in IMLCU or Non-IMLCU group, the average keypoints is calculated and it is taken as a threshold to set that LCU to each group based on its keypoints. Bit allocation for a frame is also separated into two parts and assigned to each group. The bit budget of IMLCU is assigned higher than bit budget of Non-LCU to maintain the interest points as high as possible after decoded process.

The experimental results demonstrated that the performance of proposed method can preserve more keypoints than the original reference software (HM15) especially in low bit rate. In addition, proposed method can also achieve high quality in 480p sequence.

There are also possible future works in this research. First of all, keypoints locations could be exploited to be flexible following coding tree structure and redundancy keypoints can be removed by applying moving region algorithms. In addition, learning rate algorithm can also apply to estimate the relationship between interest point and lambda value. Moreover, first picture could be regenerated by using statistically modeling to estimate the coding tree structure, for example, bias estimator.

REFERENCES

[1] S. Gohshi, "Realtime Super Resolution for 4K/8K with Nonlinear Signal Processing," in SMPTE Motion Imaging Journal, vol. 124, no. 7, pp. 51-56, October 2015.

[2] Cisco VNI 2016, http://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/vni-hyperconnectivity-wp.html

[3] G. J. Sullivan, J. R. Ohm, W. J. Han and T. Wiegand, "Overview of the High Efficiency Video Coding (HEVC) Standard," in *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 22, no. 12, pp. 1649-1668, Dec. 2012.

[4] T. Wiegand, G. J. Sullivan, G. Bjontegaard and A. Luthra, "Overview of the H.264/AVC video coding standard," in *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 13, no. 7, pp. 560-576, July 2003.

[5] J. R. Ohm, G. J. Sullivan, H. Schwarz, Thiow Keng Tan and T. Wiegand, "Comparison of the Coding Efficiency of Video Coding Standards—Including High Efficiency Video Coding (HEVC)," in IEEE Transactions on Circuits and Systems for Video Technology, vol. 22, no. 12, pp. 1669-1684, Dec. 2012.

[6] https://hevc.hhi.fraunhofer.de/svn/svn_HEVCSoftware/

[7] K. Konstantinides, Cheng-Tie Chen, Ting-Chung Chen, Hown Cheng and Fure-Chinq Jeng, "Design of an MPEG-2 codec," in *IEEE Signal Processing Magazine*, vol. 19, no. 4, pp. 32-41, Jul 2002.

[8] W. Ding and B. Liu, "Rate control of MPEG video coding and recording by rate-quantization modeling," IEEE Trans. Circuits Syst. Video Technol., vol. 6, pp. 12–20, Feb. 1996.

[9] Z. He, Y. K. Kim, and S. K. Mitra, "Low-delay rate control for DCT video coding via ρ-domain source modeling," *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. 11, pp. 928-940, 2001.

[10] B. Li, H. Li, L. Li, and J. Zhang, "λ Domain Rate Control Algorithm for High Efficiency Video Coding," Image Processing, IEEE Transactions on, vol. 23, pp. 3841-3854, 2014.

[11] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vis.*, vol. 60, no. 2, pp. 91–110, Nov. 2004.

[12] M. Fischler and R. Bolles, “Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography,” *Communications of the ACM*, vol. 24, no. 6, pp. 381–395, 1981.

[13] L. Baroffio, M. Cesana, A. Redondi, M. Tagliasacchi, and S. Tubaro, “Coding Visual Features Extracted From Video Sequences,” *IEEE Transactions on Image Processing*, vol. 23, no. 5, pp. 2262–2276, May 2014.

[14] J. Chao and E. Steinbach, “Keypoint encoding for improved feature extraction from compressed video at low bitrates,” *IEEE Trans. Multi-media*, vol. 18, no. 1, pp. 25–39, Jan. 2016.

[15] M. Makar, C. L. Chang, D. Chen, S. S. Tsai and B. Girod, "Compression of image patches for local feature extraction," *2009 IEEE International Conference on Acoustics, Speech and Signal Processing*, Taipei, 2009, pp. 821-824.

[16] X. Zhang, S. Ma, S. Wang, S. Wang, X. Zhang, and W. Gao, “From visual search to video compression: A compact representation framework for video feature descriptors,” in *Proc. Data Compress. Conf. (DCC)*, 2016, pp. 407–416.

[17] J. Chao and E. Steinbach, "Preserving SIFT features in JPEG-encoded images," *2011 18th IEEE International Conference on Image Processing*, Brussels, 2011, pp. 301-304.

[18] J. Chao and E. Steinbach, "SIFT feature-preserving bit allocation for H.264/AVC video compression," *2012 19th IEEE International Conference on Image Processing*, Orlando, FL, 2012, pp. 709-712.

REFERENCES

APPENDIX

VITA

Sovann Chen received the B.Eng. degree in electrical engineering from Institute of Technology of Cambodia, Phnom Penh, Cambodia, in 2015. He has been granted a scholarship by AUN/SEED-Net, JICA to pursue his Master’s degree in electrical engineering at the Digital Signal Processing Research Laboratory, Department of Electrical Engineering, Faculty of Engineering, Chulalongkorn University, Thailand, since 2015. His research interests include multimedia compression, image/video processing, and machine learning.

List of Publications

1. S. Chen, H. M. Maung and S. Aramvith, "Improving feature preservation in high efficiency video coding standard," 2016 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA), Jeju, 2016, pp. 1-5.