國立臺灣大學電資學院電機工程學系

博士論文

Department of Electrical Engineering

College of Electrical Engineering and Computer Science

Nation Taiwan University

Doctoral Dissertation

視覺感測系統之資訊密度的解析度控制與資料獨特性的傳輸量控制

Region-of-Interest-Based Quality Control and Keyframe-Based Quantity Control in Visual Sensing System

林意淳

Yi-Chun Lin

指導教授：連豊力博士

Advisor：Feng-Li Lian, PhD

中華民國一百零三年七月

July 2014

國立臺灣大學博士學位論文
口試委員會審定書

視覺感測系統之資訊密度的解析度控制與資料獨特性的傳輸量控制

Region-of-Interest-Based Quality Control and Keyframe-Based Quantity Control in Visual Sensing Systems

本論文係林意淳君（學號 D96921002）在國立臺灣大學電機工程學系完成之博士學位論文，於民國 103 年 7 月 2 日承下列考試委員審查通過及口試及格，特此證明。

口試委員：

連豐力（指導教授）

張帆人

顏炳郎

簡忠漢

李後燦

黃正民

系主任　廖婉君

誌謝

該從何說起好，一路走來，要感謝的人太多，從中央數學研究所畢業後，投入職場兩年，發現心之所向還是學術領域這塊，很感謝當時政大統計系陳麗霞老師，願意讓我每周在研究助理工作之餘，撥出一天的時間到台大旁聽課程，當初若不是去旁聽連豊力老師的線性系統課程，就不會有機會進入這個實驗室。這七年的博士生涯，讓我認識了很多學弟妹，從一開始的歡迎他們加入，到慶祝他們畢業，也送走了好幾批，一晃眼，輪到自己了。

首要感謝我的指導教授，連豊力老師。從您身上，我學到了很多，不只是學術上的堅持、熱情和想像力，還有為人處事上的應對進退，您總是不厭其煩的對學生諄諄教誨，循循善誘，而您在教學上的熱誠，也是讓我欽佩的。真的很高興，我當初旁聽了您的線性系統課程，進而成為您的學生，這七年的時間，很感謝的您的一路提攜和指導，您總是鼓勵學生走出台灣，多多參與國際會議，進行學術的交流，也是因為這樣，我現今才能夠擁有廣闊的學術視野。

接著要感謝五位口試委員:張帆人博士、顏炳郎博士、簡忠漢博士、李後燦博士和黃正民博士，謝謝您們對論文的建議，讓我知道論文的不足之處，進而有機會完善整篇論文的內容。

再來是NCSLab的成員,謝謝志明學長不管是在學業或是生活上的支持與打氣,也謝謝各位已經畢業或是還沒畢業的學弟妹帶給我的歡樂，謝謝上瑋、士瑋和建廷，你們彈奏的吉他真的很舒壓，謝謝冠傑、一銘、執中、沛怡、子華和俊兆，讓我認識了腦力激盪的桌遊，還有其他眾多學弟妹們，我由衷的感謝你們參與了我的博士生涯，豐富了我的人生，博士生涯有你們真好。

最後，感謝我父母一路的支持，我才能夠無後顧之憂的專心於學業。

林意淳　謹誌

中華民國一百零三年八月

視覺感測系統之資訊密度的解析度控制與資料獨特性的傳輸量控制

研究生:林意淳　　指導教授:連豊力 博士

國立臺灣大學　電機工程學系

摘要

視覺感測器因其具有競爭性的價格和豐富的感測資訊量，因此，在過去幾年間被廣泛的運用在特定的區域，並且有目的性的收集感興趣目標物的影像資訊。在許多應用上都可見到其提供協助或是監視功能的蹤跡，例如: 工業機器人、軍事防禦和監控系統。隨著視覺感測系統使用的快速增加，也產生了越來越多需要被傳輸影像資料。然而，要在一個共享且頻寬有限制的網路上進行大量的影像傳輸，是非常困難且具有挑戰性的。其引起的嚴重延遲以及封包丟失會大大的降低系統表現和影像分析結果。為了兼顧系統表現和可靠穩定的傳輸，在本論文中基於資訊密度、資料獨特性和系統動態性，提出針對影像資料的解析度和傳輸量的控制方法。就理論面而言，影像資料的解析度控制被轉換成量化回饋穩定性問題並且以 Lyapunov 的方法加以證明。就

實際應用上，所設計的影像解析度控制策略被實現在 PTZ 相機上，並且在室內和室外的實驗場景都得到極佳的表現。另一方面，影像資料的傳輸量控制可視為影像摘要問題,為了確保在經過資料減量過程後，系統表現依舊維持在令人滿意的範圍內。因此，在本論文，提出基於感知運動能量來設計針對獨特性資料的取樣策略並藉此移除重複性高的資料。接著，將其實現在大量且豐富的實驗場景，不僅可以得到將近 50% 的優秀資料減量結果，更重要的是，系統表現也維持在可接受範圍內。再著，在本論文中所提出影像資料解析度跟傳輸量的控制方法也與其他方法做比較，並藉此展現其卓越的優勢。

關鍵字:

解析度控制、傳輸量控制、資料獨特性、動態取樣、量化回饋穩定性、變焦控制。

Region-of-Interest-Based Quality Control and Keyframe-Based Quantity Control in Visual Sensing System

Student: Yi-Chun Lin Advisor: Dr. Feng-Li Lian

Department of Electrical Engineering

National Taiwan University

ABSTRACT

Visual sensors are widely used in the significant area, such as industry, army and public to collect the abundant video-related data about the objects of interest in the past years due to reasonable price and unique sensing capability. They have been found in various applications such as industrial robotics, military defense and surveillance for assisting and monitoring purposes. The rapid use of visual sensing systems has led to the increased amounts of video data that impose implicit difficulties on video transmission task over a shared and bandwidth-limited communication network. Moreover, control performance or video analysis results would be greatly degraded in the presence of constraints such as severe delays and packet dropouts induced by

excessive transmitted video data. For taking desired performance and reliable transmission into account, quality and quantity control of video packet data are proposed based on information density and system dynamics in the dissertation. Quality control is modeled as the quantized feedback stabilization problem and is proved in the sense of Lyapunov. In practical applications, designed quality control policies are implemented in the camera with zoom functionality and are experimented on indoor and outdoor environment to clearly demonstrate the better human tracking and detection performance. On the other hand, quantity control is modeled as video summarization problem. In order to prevent the system performance from being influenced during data reduction process, the designed keyframe extraction rules based on perceived motion energy are proposed to remove the similar frames. The exceptional and near 50% data reduction ratio and acceptable tracking, detection and transmission results are presented with abundant typical and experimental videos. Furthermore, the proposed quality and quantity control are also compared with the other approaches to present its outstanding advantage.

Keywords:

Quality control, Quantity control, Keyframe extraction, Dynamic sampling, Quantized feedback stabilization, Zoom control

Contents

摘要 i
ABSTRACT iii
CONTENNTS vii
LIST OF FIGURES ix
LIST OF TABLES xxi
CHAPTER 1 INTRODUCTION 1
1.1 MOTIVATION 5
1.2 CONTRIBUTION 11
1.3 ORGANIZATION 18
CHAPTER 2 LITERATURE OF SURVEY 19
2.1 QUALITY CONTROL 20
2.2 QUANTITY CONTROL 23
CHATPER 3 REGION-OF-INTEREST-BASED QUALITY CONTROL 29
3.1 QUANTIZED STATE FEEDBACK STABILIZATION 32
3.2 REGION OF INTEREST BASED ZOOM CONTROL 39
3.3 EXPERIMENTAL RESULTS OF REGION-OF-INTEREST-BASED QUALITY CONTROL 43
3.3.1 DESCRIPTION 43
3.3.2 INDOOR ENVIRONMENT 46
3.3.3 OUTDOOR ENVIRONMNET 57
3.4 SUMMARY 67
CHAPTER 4 KEYFRAME-BASED QUANTITY CONTROL 69
4.1KEYFRAME EXTRACTION BASED ON PERCEIVED MOTION ENERGY (PME) 71
4.2 SAMPLING STRATEGY 82
4.3 EXPERIMENTAL RESULTS OF PME-BASED KEYFRAME EXTRACTION ANALYSIS 86
4.3.1 DESCRIPTION 86
4.3.2 EXPERIMENTAL RESULTS 89
4.3.2.1 PART I: VISUAL SENSOR IS MOBILE 89
4.3.2.2 PART II: VISUAL SENSOR IS STATIONARY 134
4.3.3 COMPARISON 171
4.4 SUMMARY 179
CHAPTER 5 CONCLUSION NAD FUTURE WORK 183
5.1 CONCLUSION 183
5.2 FUTURE WORK 190

REFERENCES 191

List of Figures

FIGURE 1.1: THE EVOLUTION OF CONTROL ARCHITECTURE AND REPRESENTATIVE PEOPLE IN THREE KINDS OF FEEDBACK CONTROL SYSTEMS 4
FIGURE 1.2: THE DIAGRAM OF VISUAL SENSING SYSTEM 6
FIGURE 1.3: THE DIAGRAM OF SENSING DATA. 7
FIGURE 1.4: THE ILLUSTRATION OF SAMPLING RATE AND QUANTIZER STEP SIZE IN THE DIFFERENT SYSTEM DYNAMIC STAGE .. 9
FIGURE 1.5: THE CONSECUTIVE CAPTURED FRAMES FROM FRAME 1 TO FRAME 15 AND FRAME 1 IS EXTRACTED AS KEYFRAME TO SUMMARIZE THE FIFTEEN FRAMES 9
FIGURE 1.6: (A) AND (B) ARE THE CAPTURED FRAMES BY USING THE QUANTIZER STEP SIZE IN (C) AND (D), RESPECTIVELY 10
FIGURE 1.7: THE FRAMEWORK OF SMART SENSING CONSISTS OF ROI-BASED CONTROLLED SENSING AND KEYFRAME-BASED COMMUNICATION MODULE 15
FIGURE 1.8: THE EXAMPLE OF KEYFRAME EXTRACTION PROCESS SHOWS THE TRACKING RESULTS OF THE ORIGINAL CAPTURED AND THE EXTRACTED FRAME SET 16
FIGURE 1.9: THE PROCEDURES OF ROI-BASED CONTROLLED SENSING AND KEYFRAME-BASED COMMUNICATION MODULE 16
FIGURE 1.10: THE DIAGRAM OF THE PROPOSED QUALITY AND QUANTITY CONTROL SYSTEM WHICH IS APPLIED TO VISUAL-SENSING SYSTEM 17

FIGURE 2.1: THE RELATIONSHIP OF SURVEYED KEYFRAME EXTRACTION APPROACHES IN THE LITERATURES 27

FIGURE 3.1: THE ILLUSTRATION OF OPERATION AND PURPOSE OF PAN TILT AND ZOOM 31
FIGURE 3.2: THE PROPOSED FRAMEWORK OF ROI-BASED CONTROLLED SENSING WITH PAN, TILT AND ZOOM CONTROLS. $Info_{dpi}$, b_{dpi}, $_y$, r, AND e REPRESENTS THE INFORMATION DENSITY OF THE TARGET, DENSITY THRESHOLD, OUTPUT, REFERENCE AND TRACKING ERROR, RESPECTIVELY 41
FIGURE 3.3: THE FLOWCHART OF PROPOSED ZOOM CONTROL 42
FIGURE 3.4: (A) AXIS P5512 PTZ CAMERA USED IN THE DISSERTATION. (B) INDOOR EXPERIMENTAL SCENE AT MD BUILDING IN NTU. (C) OUTDOOR EXPERIMENTAL SCENE IN NTU 44
FIGURE 3.5: THE CONVERGENCE ZONE DEFINITION IN IMAGE COORDINATION 45
FIGURE 3.6: THE INDOOR EXPERIMENTAL SCENE AT 5F OF Ming-Da BUILDING IN NTU 46
FIGURE 3.7: THE SNAPSHOTS OF FAILURE DETECTION RESULT WITHOUT USING ZOOM FUNCTIONALITY. 48
FIGURE 3.8: THE ZOOM VALUE IS CONSTANT IN TRADITIONAL METHOD 48
FIGURE 3.9: (A) THE SNAPSHOTS OF TRACKING RESULTS. (B) AND (C) ARE THE X-COORDINATE AND Y-COORDINATE OF CENTER POINT OF DETECTED BOX, RESPECTIVELY 49
FIGURE 3.10: THE INFORMATION DENSITY IS FIXED IN TRADITIONAL METHOD 50
FIGURE 3.11: THE SNAPSHOTS OF FAILURE DETECTION RESULT WITH USING ZOOM FUNCTION. 52
FIGURE 3.12: THE ZOOM VALUE IN INDOOR EXPERIMENTAL SCENE WITH ROI-BASED CONTROLLED SENSING METHOD 52
FIGURE 3.13: (A) THE SNAPSHOTS OF TRACKING RESULTS. (B) AND (C) ARE THE X-COORDINATE AND Y-COORDINATE OF CENTER POINT OF DETECTED BOX, RESPECTIVELY 53

FIGURE 3.14: THE INFORMATION DENSITY IN ROI-BASED CONTROLLED SENSING METHOD 54
FIGURE 3.15: PERFORMANCE COMPARISON OF (A) TRADITIONAL METHOD AND (B) ROI-BASED CONTROLLED SENSING METHOD 55
FIGURE 3.16: THE PARTIAL FIGURE OF FIGURE 3.13 FROM FRAMES 50 TO 150 56
FIGURE 3.17: THE PERCENTAGE AREA OF VIEWED PHYSICAL WORLD. (A) IN TRADITIONAL WAY. (B) IN ROI-BASED CONTROLLED SENSING 56
FIGURE 3.18: THE OUTDOOR EXPERIMENTAL SCENE AT 3F OF Ming-Da BUILDING IN NTU 57
FIGURE 3.19: THE SNAPSHOTS OF FALSE DETECTION RESULT WITHOUT USING ZOOM FUNCTION. 59
FIGURE 3.20: THE ZOOM VALUE IS CONSTANT IN TRADITIONAL METHOD 59
FIGURE 3.21: (A) THE SNAPSHOTS OF TRACKING RESULTS. (B) AND (C) ARE THE X-COORDINATE AND Y-COORDINATE OF CENTER POINT OF DETECTED BOX, RESPECTIVELY 60
FIGURE 3.22: THE INFORMATION DENSITY IS FIXED IN TRADITIONAL METHOD 61
FIGURE 3.23: THE SNAPSHOTS OF FALSE DETECTION RESULT WITH USING ZOOM FUNCTION. 62
FIGURE 3.24: THE ZOOM VALUE IN OUTDOOR EXPERIMENTAL SCENE WITH ROI-BASED CONTROLLED SENSING METHOD 62
FIGURE 3.25: (A) THE SNAPSHOTS OF TRACKING RESULTS. (B) AND (C) ARE THE X-COORDINATE AND Y-COORDINATE OF CENTER POINT OF DETECTED BOX, RESPECTIVELY 63
FIGURE 3.26: THE INFORMATION DENSITY IN ROI-BASED CONTROLLED SENSING METHOD 64
FIGURE 3.27: PERFORMANCE COMPARISON OF (A) TRADITIONAL METHOD AND (B) ROI-BASED CONTROLLED SENSING METHOD 65

FIGURE 3.28: THE PARTIAL FIGURE OF FIGURE 3.23 FROM FRAMES 100 TO 200 66
FIGURE 3.29: THE PERCENTAGE AREA OF VIEWED PHYSICAL WORLD. (A) IN TRADITIONAL WAY. (B) IN ROI-BASED CONTROLLED SENSING 66

FIGURE 4.1: THE ILLUSTRATION OF COMPUTING PERCEIVED MOTION ENERGY 74
FIGURE 4.2: THE DIAGRAM OF THE PERCEIVED MOTION VALUE OF THE VIDEO 74
FIGURE 4.3: THE DIAGRAM OF KEYFRAME EXTRACTION 78
FIGURE 4.4: THE DIAGRAM OF EXTRACTED KEYFRAMES RESULT 79
FIGURE 4.5: THE DIAGRAM OF APPROXIMATION FUNCTION SHAPED FROM THE EXTRACTED KEYFRAMES. 80
FIGURE 4.6: THE DIAGRAM OF PME-BASED KEYFRAME EXTRACTION PROCESS 81
FIGURE 4.7: EXAMPLE OF THREE DIFFERENT SAMPLING STRATEGIES, TRADITIONAL, $Fixed(\alpha,\beta)$ WITH FIXED SAMPLING PERIOD $\alpha=1$ AND $\beta=3$, AND PME-BASED WITH $\lambda_{ctrl}^{scene}=1$ 83
FIGURE 4.8: EXPERIMENTAL RESULTS OF INDOOR3PEOPLE (FROM FRAME 1 TO FRAME 15) WITH USING TRADITIONAL SAMPLING STRATEGY 84
FIGURE 4.9: EXPERIMENTAL RESULTS OF INDOOR3PEOPLE (FROM FRAME 1 TO FRAME 15) WITH USING FIXED SAMPLING STRATEGY $Fixed(\alpha,\beta)$ ($\alpha=1,\beta=3$) 84
FIGURE 4.10: EXPERIMENTAL RESULTS OF INDOOR3PEOPLE (FROM FRAME 1 TO FRAME 15) WITH USING PME-BASED SAMPLING STRATEGY ($\lambda_{ctrl}^{scene}=1$) 85
FIGURE 4.11: THE EXTRACTION FLOW OF FOREMAN WITH PERCEIVED MOTION ENERGY. $AT=\lambda_{ctrl}^{scene}\times IQR$ WHERE $\lambda_{ctrl}^{scene}=\lambda_{ctrl}^{Forman}=1$ AND $IQR=1.462$ 91

FIGURE 4.12: THE SNAPSHOTS OF FOREMAN 93
FIGURE 4.13: (A) THE MOTION ENERGY CURVE, THE APPROXIMATION MOTION ENERGY CURVE AND THE FRAME NO. OF EXTRACTED KEYFRAMES (B) THE MOTION ENERGY CURVE PRODUCED BY CAPTURED FRAMES AND EXTRACTED KEYFRAMES FROM FRAMES 150 TO 200. 93
FIGURE 4.14: THE ENERGY ERROR OF EACH FRAME. 94
FIGURE 4.15: FRAMES PER SECOND RESULTS OF TRADITIONAL AND PROPOSED PME-BASED KEYFRAME EXTRACTION ALGORITHM. 94
FIGURE 4.16: PERFORMANCE INDEX OF FACE TRACKING RESULTS OF FOREMAN (A) CORRECT RESULT (B) FALSE DETECTION RESULTS WITH SMALL OR LARGE BOX 95
FIGURE 4.17: THE ILLUSTRATION OF THE LOCATION OF LEFT/TOP OF FACE BOUNDING BOX 97
FIGURE 4.18: FACE DETECTION RESULTS OF FOREMAN VIDEO IN TWO DIFFERENT SAMPLING STRATEGIES: TRADITIONAL AND PME-BASED 97
FIGURE 4.19: FACE DETECTION RESULTS OF THE FOREMAN VIDEO IN TWO DIFFERENT SAMPLING STRATEGIES: TRADITIONAL AND PME-BASED FROM FRAMES 98 TO 189 (PARTIAL OF FIGURE 4.18) 98
FIGURE 4.20: FACE TRACKING PATH IN THE FOREMAN VIDEO WITH USING TWO DIFFERENT SAMPLING STRATEGIES 98
FIGURE 4.21: THE SNAPSHOTS OF COASTGUARD 100
FIGURE 4.22: (A) THE MOTION ENERGY CURVE, THE APPROXIMATION MOTION ENERGY CURVE AND THE FRAME NO. OF EXTRACTED KEYFRAMES (B) THE MOTION ENERGY CURVE PRODUCED BY CAPTURED FRAMES AND EXTRACTED KEYFRAMES FROM FRAMES 50 TO 100. 101
FIGURE 4.23: THE ENERGY ERROR OF EACH FRAME. 102

FIGURE 4.24: FRAMES PER SECOND RESULTS OF TRADITIONAL AND PROPOSED PME-BASED KEYFRAME EXTRACTION ALGORITHM. 103
FIGURE 4.25: MOVING OBJECTS TRACKING RESULTS OF COSTGUARD (A)(D)(E) CAPTURED FRAMES 67, 100 AND 143 (B)(E)(H) TRACKING RESULTS OF FRAMES 67, 100 AND 143 WITH TRADITIONAL WAY (C)(F)(I) TRACKING RESULTS OF FRAMES 67, 100 AND 143 WITH PME-BASED KEYFRAME EXTRACTION ALGORITHM 104
FIGURE 4.26: THE SNAPSHOTS OF INDOOR1PEOPLE..... 106
FIGURE 4.27: (A) THE MOTION ENERGY CURVE, THE APPROXIMATION MOTION ENERGY CURVE AND THE FRAME NO. OF EXTRACTED KEYFRAMES (B) THE MOTION ENERGY CURVE PRODUCED BY CAPTURED FRAMES AND EXTRACTED KEYFRAMES FROM FRAMES 50 TO 100. 107
FIGURE 4.28: THE ENERGY ERROR OF EACH FRAME..... 108
FIGURE 4.29: FRAMES PER SECOND RESULTS OF TRADITIONAL, $Fixed(1,3)$ AND PROPOSED PME-BASED KEYFRAME EXTRACTION ALGORITHM. 109
FIGURE 4.30: THREE REPRESENTED FRAMES AND THEIR ENCODED FRAMES UNDER DIFFERENT SAMPLING STRATEGY WITH ACCEPTABLE VISUAL QUALITY PSNR = 30DB IN 1.4MBPS BANDWIDTH. H: HIGH (LABEL 1 REGIONS), M: MIDDLE (LABEL 2 REGIONS); L:LOW (LABEL 3 REGIONS) 111
FIGURE 4.31: THE QUALITY OF DECODED FRAMES WITH $\lambda_{PSNR}^{scene} = 30$ IN 1.4MBPS FOR TRADITIONAL..... 113
FIGURE 4.32: THE QUALITY OF DECODED FRAMES WITH $\lambda_{PSNR}^{scene} = 30$ IN 1.4MBPS FOR $Fixed(1,3)$ 114
FIGURE 4.33: THE QUALITY OF DECODED FRAMES WITH $\lambda_{PSNR}^{scene} = 30$ IN 1.4MBPS FOR PME-BASED 114

FIGURE 4.34: THE QUALITY OF DECODED FRAMES WITH $\lambda_{PSNR}^{scene} = 30$ (2.2 MBPS) (A) TRADITIONAL (B) $Fixed(1,3)$ (C) PME-BASED..116
FIGURE 4.35: SAVING PERCENTAGE OF BANDWIDTH COMPARED WITH THAT OF THE LOSSLESS CASE.....117
FIGURE 4.36: THE SNAPSHOTS OF INDOOR3PEOPLE.....118
FIGURE 4.37: (A) THE MOTION ENERGY CURVE, THE APPROXIMATION MOTION ENERGY CURVE AND THE FRAME NO. OF EXTRACTED KEYFRAMES (B) THE MOTION ENERGY CURVE PRODUCED BY CAPTURED FRAMES AND EXTRACTED KEYFRAMES FROM FRAMES 50 TO 100. 119
FIGURE 4.38: THE ENERGY ERROR OF EACH FRAME.....120
FIGURE 4.39: FRAMES PER SECOND RESULTS OF TRADITIONAL, $Fixed(1,3)$ AND PROPOSED PME-BASED KEYFRAME EXTRACTION ALGORITHM 121
FIGURE 4.40: THREE REPRESENTED FRAMES AND THEIR ENCODED FRAMES UNDER DIFFERENT SAMPLING STRATEGY WITH ACCEPTABLE VISUAL QUALITY PSNR = 30DB IN 2.6MBPS BANDWIDTH. H: HIGH (LABEL 1 REGIONS), M: MIDDLE (LABEL 2 REGIONS); L: LOW (LABEL 3 REGIONS).....122
FIGURE 4.41: THE QUALITY OF DECODED FRAMES WITH $\lambda_{PSNR}^{scene} = 30$ IN 2.6MBPS FOR TRADITIONAL.....123
FIGURE 4.42: THE QUALITY OF DECODED FRAMES WITH $\lambda_{PSNR}^{scene} = 30$ IN 2.6MBPS FOR $Fixed(1,3)$ 124
FIGURE 4.43: THE QUALITY OF DECODED FRAMES WITH $\lambda_{PSNR}^{scene} = 30$ IN 2.6MBPS FOR PME-BASED.....124
FIGURE 4.44: THE QUALITY OF DECODED FRAMES WITH $\lambda_{PSNR}^{scene} = 30$ IN 3.2MBPS (A) TRADITIONAL (B) $Fixed(1,3)$ (C)PME-BASED. 126
FIGURE 4.45: SAVING PERCENTAGE OF BANDWIDTH COMPARED WITH THAT OF THE LOSSLESS CASE.....127
FIGURE 4.46: THE SNAPSHOTS OF VISUAL ODOMETRY 128

FIGURE 4.47: (A) THE MOTION ENERGY CURVE, THE APPROXIMATION MOTION ENERGY CURVE AND THE FRAME NO. OF EXTRACTED KEYFRAMES (B) THE MOTION ENERGY CURVE PRODUCED BY CAPTURED FRAMES AND EXTRACTED KEYFRAMES FROM FRAMES 50 TO 100. 129
FIGURE 4.48: THE ENERGY ERROR OF EACH FRAME. 130
FIGURE 4.49: FRAMES PER SECOND RESULTS OF TRADITIONAL AND PROPOSED PME-BASED KEYFRAME EXTRACTION ALGORITHM 130
FIGURE 4.50: VISUAL ODOMETRY RESULT WITH TRADITIONAL SAMPLING. 131
FIGURE 4.51: VISUAL ODOMETRY RESULT OF PME-BASED SAMPLING METHOD 132
FIGURE 4.52: THE SNAPSHOTS OF SINGLEBALL 136
FIGURE 4.53: THE MOTION ENERGY CURVE, THE APPROXIMATION MOTION ENERGY CURVE AND THE FRAME NO. OF EXTRACTED KEYFRAMES. 136
FIGURE 4.54: THE ENERGY ERROR OF EACH FRAME. 137
FIGURE 4.55: FRAMES PER SECOND RESULTS OF TRADITIONAL AND PROPOSED PME-BASED KEYFRAME EXTRACTION ALGORITHM 137
FIGURE 4.56: TRACKING ERROR OF TWO DIFFERENT TESTED SCENARIOS OF SINGLEBALL (A) INITIAL (B) SWITCH CONSTANT ACCELERATION TO CONSTANT VELOCITY 138
FIGURE 4.57: TRACKING ERROR OF TWO DIFFERENT TESTED SCENARIOS OF SINGLEBALL (A) INITIAL LOCATION IS RANDOMLY GIVEN (B) MEASUREMENT NOISE IS ADDED 139
FIGURE 4.58: THE SNAPSHOTS OF CONTAINER 141

FIGURE 4.59: (A) THE MOTION ENERGY CURVE, THE APPROXIMATION MOTION ENERGY CURVE AND THE FRAME NO. OF EXTRACTED KEYFRAMES (B) THE MOTION ENERGY CURVE PRODUCED BY CAPTURED FRAMES AND EXTRACTED KEYFRAMES FROM FRAMES 250 TO 300. 142
FIGURE 4.60: THE ENERGY ERROR OF EACH FRAME. 143
FIGURE 4.61: FRAMES PER SECOND RESULTS OF TRADITIONAL AND PROPOSED PME-BASED KEYFRAME EXTRACTION ALGORITHM 143
FIGURE 4.62: MOVING OBJECTS TRACKING RESULTS OF CONTAINER (A)(D)(E) CAPTURED FRAMES 95, 173 AND 257 (B)(E)(H) TRACKING RESULTS OF FRAMES 95, 173 AND 257 WITH TRADITIONAL WAY (C)(F)(I) TRACKING RESULTS OF FRAMES 95, 173 AND 257 WITH PME-BASED KEYFRAME EXTRACTION ALGORITHM 144
FIGURE 4.63: THE SNAPSHOTS OF FLOORS3PEOPLE 146
FIGURE 4.64: (A) THE MOTION ENERGY CURVE, THE APPROXIMATION MOTION ENERGY CURVE AND THE FRAME NO. OF EXTRACTED KEYFRAMES (B) THE MOTION ENERGY CURVE PRODUCED BY CAPTURED FRAMES AND EXTRACTED KEYFRAMES FROM FRAMES 100 TO 150. 147
FIGURE 4.65: THE ENERGY ERROR OF EACH FRAME. 148
FIGURE 4.66: FRAMES PER SECOND RESULTS OF TRADITIONAL, $Fixed(1,3)$ AND PROPOSED PME-BASED KEYFRAME EXTRACTION ALGORITHM 149
FIGURE 4.67: THE QUALITY OF DECODED FRAMES WITH $\lambda_{PSNR}^{scene} = 30$ IN 4.5MBPS (A) TRADITIONAL (B) $Fixed(1,3)$ (C) PME-BASED 150
FIGURE 4.68: SAVING PERCENTAGE OF BANDWIDTH COMPARED WITH THAT OF THE LOSSLESS CASE 151

FIGURE 4.69: THE SCENARIO EXPRESSION OF EXPERIMENTAL FLOORS3PEOPLE VIDEO 152
FIGURE 4.70: RESULTS OF LOCATION DETECTION TO THREE MOVING PEOPLE (A)(C)(E) THE ERROR OF X AND Y DIRECTION (B)(D)(E) THE DETECTED RESULTS IN THE REAL SCENE OF PARTICULAR FRAME 154
FIGURE 4.71: THE SNAPSHOTS OF HALL 155
FIGURE 4.72: (A) THE MOTION ENERGY CURVE, THE APPROXIMATION MOTION ENERGY CURVE AND THE FRAME NO. OF EXTRACTED KEYFRAMES (B) THE MOTION ENERGY CURVE PRODUCED BY CAPTURED FRAMES AND EXTRACTED KEYFRAMES FROM FRAMES 100 TO 150. 156
FIGURE 4.73: THE ENERGY ERROR OF EACH FRAME. 157
FIGURE 4.74: FRAMES PER SECOND RESULTS OF TRADITIONAL AND PROPOSED PME-BASED KEYFRAME EXTRACTION ALGORITHM 158
FIGURE 4.75: EXAMPLES OF MULTI-OBJECTS TRACKING RESULTS OF HALL (A) CORRECT DETECTION RESULT (B) FALSE DETECTION RESULT 159
FIGURE 4.76: THE ILLUSTRATIONS OF BOUNDING BOXES AND (X_1, Y_1) AND (X_2, Y_2) ARE USED TO BE THE TRACKING INDEX FOR DETECTED HUMAN 1 AND DETECTED HUMAN 2, RESPECTIVELY 159
FIGURE 4.77: MULTI-OBJECTS TRACKING RESULTS OF HALL WITH USING TRADITIONAL AND PME-BASED SAMPLING STRATEGIES. (A) X COORDINATE OF LEFT/TOP OF BOUNDING FOR HUMAN 1. (B) THE ZOOM IN VERSION FROM FRAMES 50 TO 90 161

FIGURE 4.78: MULTI-OBJECTS TRACKING RESULTS OF HALL IN TRADITIONAL AND PME-BASED SAMPLING STRATEGIES. (A) *Y* COORDINATE OF LEFT/TOP OF BOUNDING FOR HUMAN 1. (B) THE ZOOM VERSION OF (A) FROM FRAMES 50 TO 90… 162
FIGURE 4.79: MULTI-OBJECTS TRACKING RESULTS OF HALL IN TRADITIONAL AND PME-BASED SAMPLING STRATEGIES. (A) *X* COORDINATE OF LEFT/TOP OF BOUNDING FOR HUMAN 2 (B) THE ZOOM VERSION OF (A) FROM FRAMES 200 TO 250 163
FIGURE 4.80: MULTI-OBJECTS TRACKING RESULTS OF HALL IN TRADITIONAL AND PME-BASED SAMPLING STRATEGIES. (A) *Y* COORDINATE OF LEFT/TOP OF BOUNDING FOR HUMAN 2 (B) THE ZOOM VERSION OF (A) FROM FRAMES 200 TO 250 164
FIGURE 4.81: THE SNAPSHOTS OF VISIONTRAFFIC 165
FIGURE 4.82: (A) THE MOTION ENERGY CURVE, THE APPROXIMATION MOTION ENERGY CURVE AND THE FRAME NO. OF EXTRACTED KEYFRAMES (B) THE MOTION ENERGY CURVE PRODUCED BY CAPTURED FRAMES AND EXTRACTED KEYFRAMES FROM FRAMES 310 TO 360. 166
FIGURE 4.83: THE ENERGY ERROR OF EACH FRAME. 167
FIGURE 4.84: FRAMES PER SECOND RESULTS OF VISIONTRAFFIC IN TRADITIONAL AND THE PROPOSED PME-BASED KEYFRAME EXTRACTION ALGORITHM. 168
FIGURE 4.85: PERFORMANCE INDEX OF FACE TRACKING RESULTS OF VISIONTRAFFIC (A) CORRECT DETECTION RESULT (B) FALSE DETECTION RESULT 168

FIGURE 4.86: CAR COUNTING RESULTS OF VISIONTRAFFIC IN TRADITIONAL AND PME-BASED SAMPLING STRATEGIES. (A) CAR COUNTING RESULTS (B) THE ZOOM IN VERSION OF (A) FROM FRAMES 90 TO 210.....170
FIGURE 4.87: THE DIAGRAM OF TRIANGLE-BASED SAMPLING STRATEGY.....172
FIGURE 4.88: ODOMETRY RESULT OF FOUR DIFFERENT TYPES SAMPLING STRATEGY. (A) TRADITIONAL (B) PME-BASED (SAME FIGURE AS FIGURES 4.50 AND 4.51, RESPECTIVELY).....174
FIGURE 4.89: ODOMETRY RESULTS OF FOUR DIFFERENT TYPES SAMPLING STRATEGY (A) $Fixed(2,2)$ (B) $Fixed(3,3)$ (C) $Fixed(4,4)$ (D) TRIANGLE-BASED.....175
FIGURE 4.90: ODOMETRY RESULTS OF TRADITIONAL ANDPME-BASED. 176
FIGURE 4.91: ODOMETRY RESULTS OF THREE DIFFERENT FIXED SAMPLING PERIOD, NAMELY $Fixed(2,2)$, $Fixed(3,3)$ AND $Fixed(4,4)$. 177
FIGURE 4.92: ODOMETRY RESULTS OF THREE DIFFERENT SAMPLING STRATEGIES, NAMELY, PME-BASED, $Fixed(3,3)$ AND TRIANGLE-BASED.....177
FIGURE 4.93: ODOMETRY RESULTS OF SIX DIFFERENT SAMPLING STRATEGIES: TRADITIONAL, PME-ABSED, $Fixed(2,2)$, $Fixed(3,3)$, $Fixed(4,4)$,AND TRIANGLE-BASED.....178
FIGURE 4.94: THE DISTRIBUTION OF TEN TESTED VIDEOS IN SENSOR-PBJECT PLANE.....180

FIGURE 5.1: (A) TRADITIONAL CONTROL DIAGRAM. (B) PROPOSED SENSING CONTROL DIAGRAM.....187

List of Tables

TABLE 3.1: THE PARAMETERS ARE USED IN TRADITIONAL AND ROI-BASED METHOD 68
TABLE 3.2: THE TRACKING RESULTS OF TRADITIONAL AND ROI-BASED METHOD IN TWO DIFFERENT EXPERIMENTAL SCENES 68

TABLE 4.1: PART I: VISUAL SENSOR IS MOBILE 87
TABLE 4.2: PART II: VISUAL SENSOR IS STATIONARY 88
TABLE 4.3: RESULTS OF FACE TRACKING, UTILIZATION RATIO, ENERGY ERROR RATIO AND THE NUMBER OF THE EXTRACTED KEYFRAME OF THE FOREMAN VIDEO 99
TABLE 4.4: RESULTS OF MULTI-OBJECTS TRACKING, UTILIZATION RATIO, ENERGY ERROR RATIO AND NUMBER OF EXTRACTED KEYFRAME OF COASTGUARD 105
TABLE 4.5: RESULTS OF AVERAGE AVAILABLE RATE TO EACH FRAME, ENERGY ERROR RATIO AND NUMBER OF EXTRACTED KEYFRAME OF INDOOR1PEOPLE 110
TABLE 4.6: QUALITY RATIO OF THREE PHASES (1.4 MBPS) 115
TABLE 4.7: QUALITY RATIO OF THREE PHASES (2.2 MBPS) 116
TABLE 4.8: AVERAGE SAVING RATIO OF BANDWIDTH OF THREE DIFFERENT SAMPLING STRATEGIES (2.2MBPS) 117
TABLE 4.9: RESULTS OF AVERAGE AVAILABLE RATE TO EACH FRAME, ENERGY ERROR RATIO AND NUMBER OF EXTRACTED KEYFRAM OF INDOOR3PEOPLE 121
TABLE 4.10: QUALITY RATIO OF THREE PHASES (2.6 MBPS) 125
TABLE 4.11: QUALITY RATIO OF THREE PHASES (3.2 MBPS) 126
TABLE 4.12: AVERAGE SAVING RATIO OF BANDWIDTH OF THREE DIFFERENT SAMPLING STRATEGIES (3.2MBPS) 126

TABLE 4.13: RESULTS OF DISTANCE FROM ORIGINAL POINT, UTILIZATION RATIO, ENERGY ERROR RATIO AND NUMBER OF EXTRACTED KEYFRAME OF VISUAL ODOMETRY 133
TABLE 4.14: RESULTS OF TRACKING ERRROR, UTILIZATION RATIO, ENERGY ERROR RATIO AND NUMBER OF EXTRACTED KEYFRAME OF SINGLEBALL 140
TABLE 4.15: RESULTS OF MULTI-OBJECTS TRACKING, UTILIZATION RATIO, ENERGY ERROR RATIO AND NUMBER OF EXTRACTED KEYFRAME OF CONTAINER 145
TABLE 4.16: RESULTS OF AVERAGE AVAILABLE RATE TO EACH FRAME, ENERGY ERROR RATIO AND NUMBER OF EXTRACTED KEYFRAM OF FLOORS3PEOPLE 149
TABLE 4.17: QUALITY RATIO OF THREE PHASES (4.5 MBPS) 151
TABLE 4.18: AVERAGE SAVING RATIO OF BANDWIDTH OF THREE DIFFERENT SAMPLING STRATEGIES (4.5MBPS) 151
TABLE 4.19: RESULTS OF MULTI-OBJECTS TRACKING, UTILIZATION RATIO, ENERGY ERROR RATIO AND NUMBER OF EXTRACTED KEYFRAME OF HALL 165
TABLE 4.20: RESULTS OF CAR DETECTION AND COUNTING, UTILIZATION RATIO, ENERGY ERROR RATIO AND NUMBER OF EXTRACTED KEYFRAME OF VISIONTRAFFIC 169
TABLE 4.21: ERROR OF DIFFERENT SAMPLING STRATEGIES. $dist = \|(x_{end}, y_{end}) - (0,0)\|$ 178
TABLE 4.22: THE PARAMETERS ARE USED IN PART I 180
TABLE 4.23: THE PARAMETERS ARE USED IN PART II 181
TABLE 4.24: THE UTILIZATION RATIO AND EXPERIMENTAL RESULTS OF THE TESTED VIDEOS IN PART I. EXPERIMENTAL RESULTS MEAN ACCURACY RATIO, AVERAGE SAVING RATIO OF BANDWIDTH OR ERROR.THE THIRD AND FOURTH ROWS REPRESENT THE EXPERIMENTAL RESULTS OF TRADITIONAL AND PME-BASED METHOD 181

TABLE 4.25: THE UTILIZATION RATIO AND EXPERIMENTAL RESULTS OF THE TESTED VIDEOS IN PART II. EXPERIMENTAL RESULTS MEAN ACCURACY RATIO, AVERAGE SAVING RATIO OF BANDWIDTH OR ERROR. THE THIRD AND FOURTH ROWS REPRESENT THE EXPERIMENTAL RESULTS OF TRADITIONAL AND PME-BASED METHOD 181
TABLE 5.1: THE EXPERIMENTAL RESULTS OF TRADITIONAL AND ROI-BASED METHOD IN TWO DIFFERENT SCENES 188
TABLE 5.2: THE ACCURACY RATIO OF DETECTION OF TRADITIONAL AND KEYFRAME-BASED METHOD IN FIVE TYPICAL VIDEOS 188
TABLE 5.3: THE SAVING RATIO OF BANDWIDTH OF TRADITIONAL, $Fixed(1,3)$ KEYFRAME-BASED METHOD IN FIVE TYPICAL VIDEOS 189
TABLE 5.4: THE TRACKING ERROR OF TRADITIONAL, KEYFRAME-BASED, $Fixed(2,2)$, TRIANGLE-BASED METHOD IN TWO TYPICAL VIDEOS 189

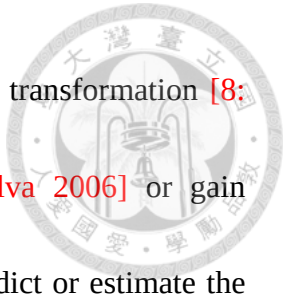
Chapter 1

Introduction

In classical control systems, controllers are built on circuit board and the electronic components are connected via hardwired connections. In technological terms, it is not convenient to redesign controllers due to the complex hardwired connections and lots of electronic components. With the development of digital computers, there is a growing interest in the use of them as a new instrumentation for design and implementation of controllers. The reasons are that digital computers could deal with complex and great arithmetical computation and more easily realize the design, simulation and implementation of control algorithms. Hence, after 1950, digital control systems have been received significant attention in various applications including manufacturing, power plants, chemical processes and airplanes [1: Baillieul & Antsaklis 2007]. However, classical or digital control systems with spatially centralized or distributed architecture are designed to transmit all information from the remote sensors to the central data servers. The architecture results in growing difficulties for maintaining and diagnosing in a widespread operation range. The situation has been changed with the

development of communication networks. The development of the networks brings more flexible architectures of spatially distributed and greatly reduces the installation, maintenance and diagnosis costs [2: Moyne & Tilbury 2007]. Thus, systems with the asynchronous communication manner among sensors, actuators, controllers and controlled plants over a shared bandwidth-limited digital network are called networked control systems (NCSs) and NCSshave been become an emerging research field for decades [3: Hespanha et al. 2007]. Figure 1.1 shows the evolution of control architecture and representative people in three kinds of feedback control systems which are classical control systems (Harry Nyquist), digital control systems (Claude Shannon) and networked control systems (Robert Bosch GmbH).

Networked control systems framework have been applied in various large-scale systems such as intelligent vehicle systems [7: Zhang et al. 2011], manufacturing systems [2: Moyne & Tilbury 2007], wireless sensor network systems [6: Schenato et al. 2007], and smart video-based surveillance systems [4: Foresti et al. 2005], [5: Hampapur et al. 2005]. Meanwhile, the constraints of communication networks such as delays, packet dropouts and limited bandwidth lead to new challenges in controller design for stabilizing controlled plants through feedback [1: Baillieul & Antsaklis 2007]. Hence, many researchers are devoted to improve and enhance the capability of controllers for dealing with the problems resulted from random transmission delays or



packet dropouts [3: Hespanha et al. 2007]. For example, scattering transformation [8: Matiakis et al. 2009], minimum-effort estimator [9: Tang & Silva 2006] or gain scheduling algorithm [10: Tipsuwan & Chow 2004] are used to predict or estimate the transmission delays and then the controllers eliminate the delay effects based on the estimation. However, the delays and packet dropouts highly depend on the varied network conditions such as congestion or channel quality. Consequently, an appropriate data transmission manner through communication networks is an important factor for avoiding congestion. Earlier related works to the network congestion problem provide some possible solutions such as task scheduling with assigned priorities [11: Lee et al. 2005] and comparing with the difference between adjacent states decides transmission or not [12: Lian et al. 2006].

Over the last few decades, the utilization of visual sensors in various applications has been increased considerably due to its reasonable price, ubiquitous and abounding measurement data of the monitored environment. Moreover, image/video processing techniques are now fairly maturation in a variety of surveillance applications. However, limited data rate implicitly imposes the challenges on the video transmission task. Hence, transmitted video data has to be appropriately reduced for reliable transmission. In the dissertation, from spatial and temporal domain point of view, video data reduction is implemented for intelligently maximizing the quality of objects of interest and

efficiently minimizing the number of frames required to be processed in image processing algorithms. Moreover, in proposed quality and quantity control framework, the tracking and detection results are maintained within the satisfactory bounds and the required bandwidth for reliable transmission is decreased.

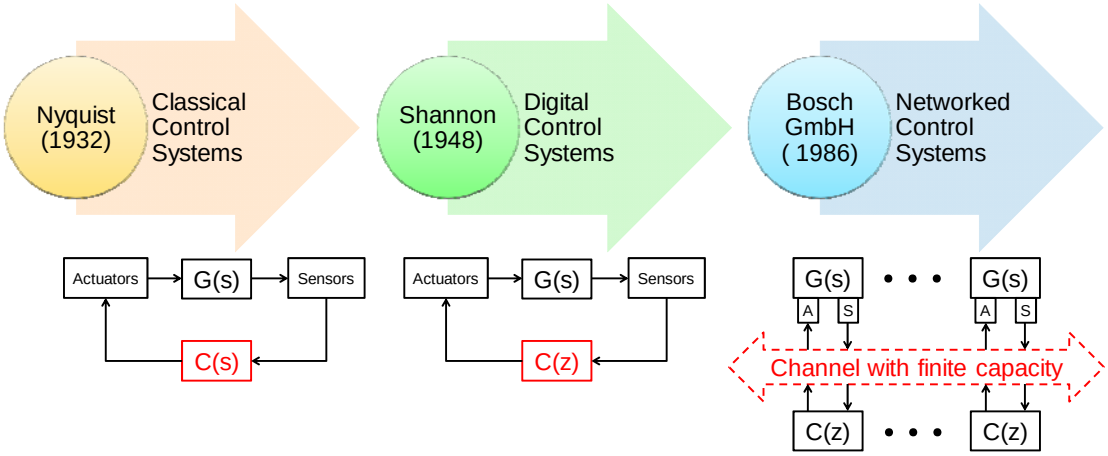


Figure 1.1. The evolution of control architecture and representative people in three kinds of feedback control systems.

1.1 Motivation

Driven by video technological developments associated with the rise of visual sensors, the use of visual sensing networks to monitor the objects or events of interest or possible threats over a wide region improves the public safety. For example, a visual sensor is located in the public area to capture the environment information and produces video-related data which consist of frames. The captured frames are then transmitted through a channel with finite capacity for different purposes such as store, monitor and further analyze as shown in Figure 1.2. The class of sensing data is shown in Figure 1.3. Visual sensing networks also have found use in various areas including habitat monitoring [13: Ghosh et al. 2007], entertainment, commercial [61: Raty 2010], military applications and industrial robotics [14: Kreucher et al. 2007] over the past decade [13: Ghosh et al. 2007], [14: Kreucher et al. 2007]. The video data are collected from distributed visual sensors and then those are typically transmitted to end-users or remote control center for being further analyzed, stored or enjoyed. However, in practical applications, finite communication bandwidth constraints associated with excessive video packet data would introduce severe delays and packet dropouts and further affect the image analysis results. In particular, for surveillance applications, the detection, recognition and tracking of potential intruders are important and immediate tasks. As consequence, a video data reduction operation is required to process the excessive

transmitted packets. One of the most popular video data size reduction procedures is compression. Although there are many approaches to compress video, the universal feature of video compression techniques is that they consume most of computation time and resources to compress the similar frames, but not all frames are needed to be compressed and/or transmitted for the video presentation, editing or storing. Hence, it is critical to design a novel procedure to find frames of interest in image collections. On the other hand, driven by a desire to make preferable commands or decisions, the quality of objects of interest is needed to increase.

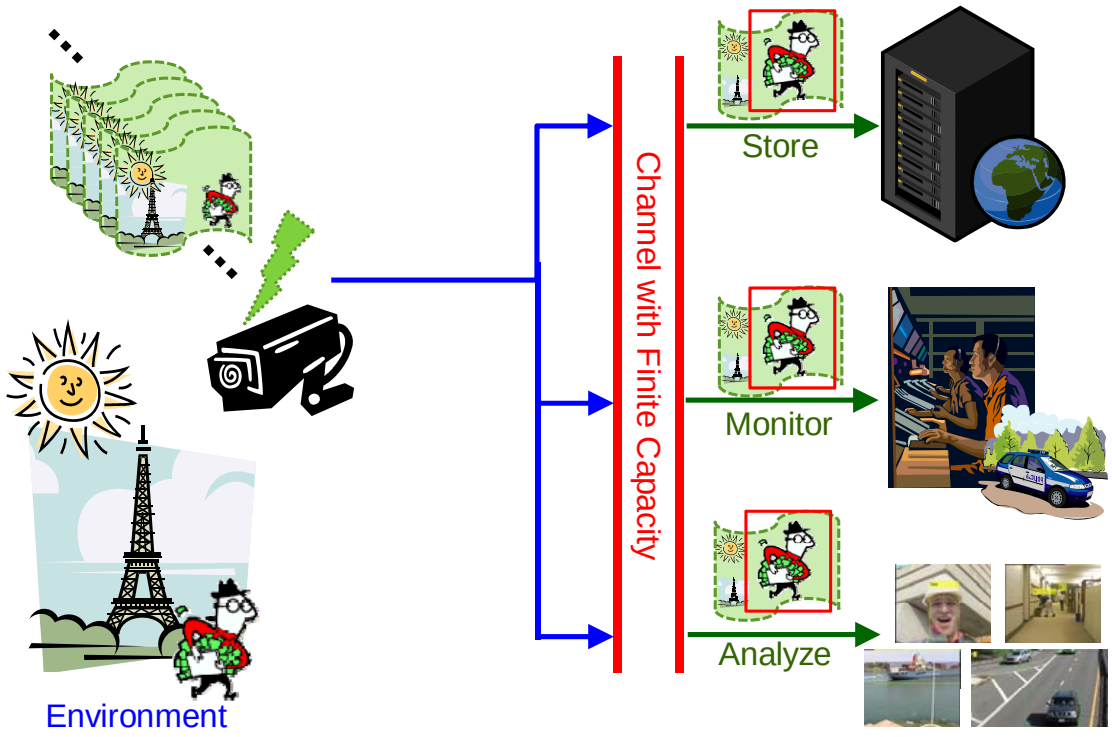


Figure 1.2. The diagram of visual sensing system.

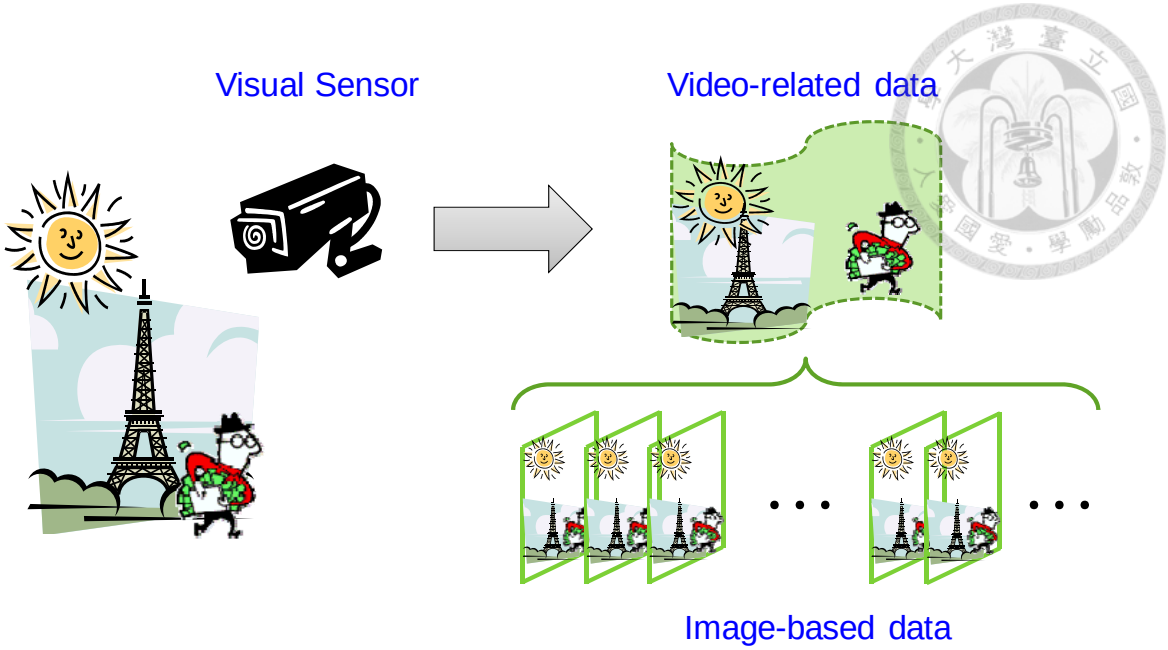


Figure 1.3. The diagram of sensing data.

Digital control appears to be one of the earliest studies along the lines of quantity and quality control. It serves two manipulations: sampling rate decides when to sample a signal for transmission and quantizer step size decides what to send through the networks based on system dynamics. For example, the curve shows the two different system dynamics: slow and fast in Figure 1.4. In the slow phase, the slow sampling rate and large quantizer step size are enough to capture the system dynamics behavior. In the fast phase, the sampling rate and the quantizer step size both are modified for obtaining more accuracy state to perform suitable controls. Moreover, with the usage of sampling and quantifying operations, one can affect the quantity and quality of transmitted data, respectively. Within the video analysis domain, keyframe extraction can be viewed as

sampling operation and zoom in/out control as a quantizer. An example of keyframe extraction is shown in Figure 1.5. Due to the video contents in the all fifteen frames are fairly similar, Frame 1 is determined as keyframe to summarize on the basis of designed extraction rules. Figure 1.6 presents an illustration of close relationship between zoom in/out and quantizer step size Δ. Zoom in/out equals to decrease/increase Δ. Based on large quantizer step size $\Delta(k)$ is shown in Figure 1.6(c), the captured image shown in Figure 1.6(a) presents ambiguous content for difficultly determining the number of people in the monitored environment. In order to obtain positive answer of number of people, it is necessary to zoom in to increase quality of objects of interest by means of small quantizer size $\Delta(k+1)$ as shown in Figures 1.6(b) and 1.6(d), respectively. Consequently, quantity and quality control can be achieved through keyframe extraction and zoom in/out control, respectively.

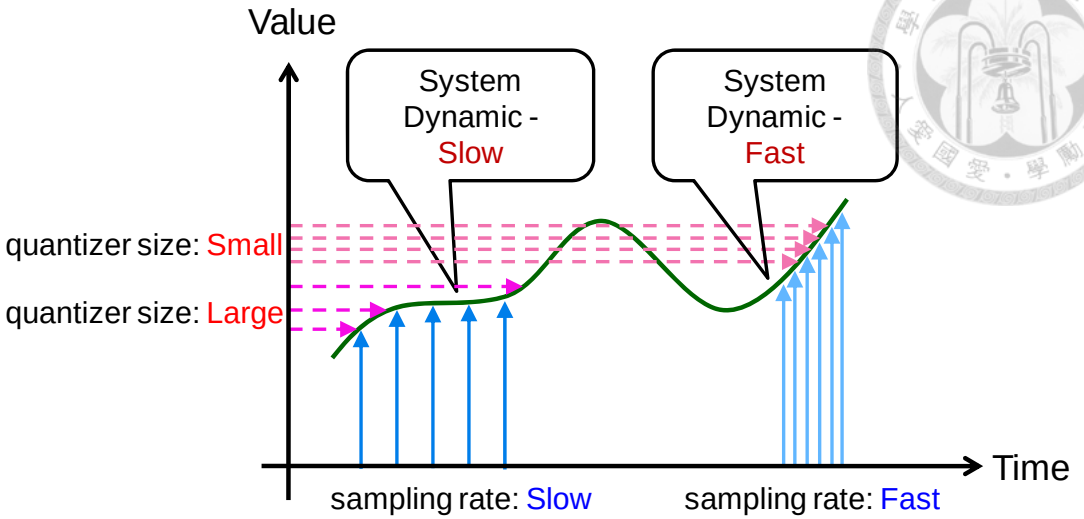


Figure 1.4. The illustration of sampling rate and quantizer step size in the different system dynamic stage.

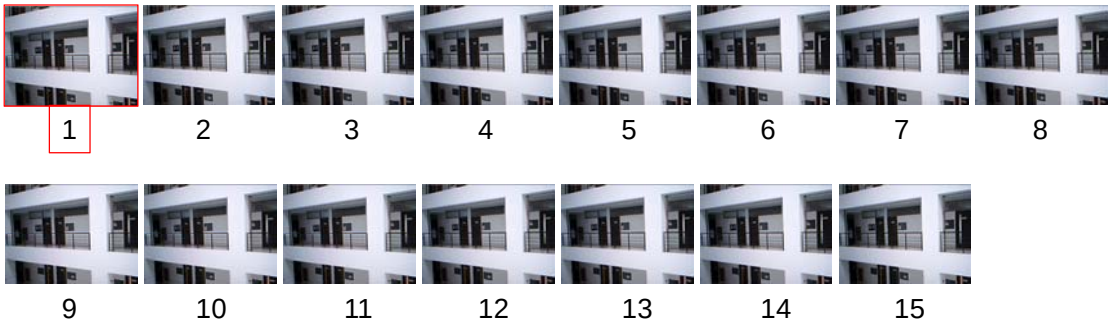


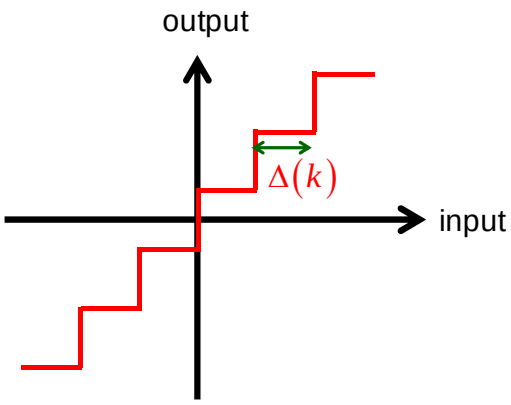
Figure 1.5. The consecutive captured frames from Frames 1 to 15 and Frame 1 is extracted as keyframe to summarize the fifteen frames.

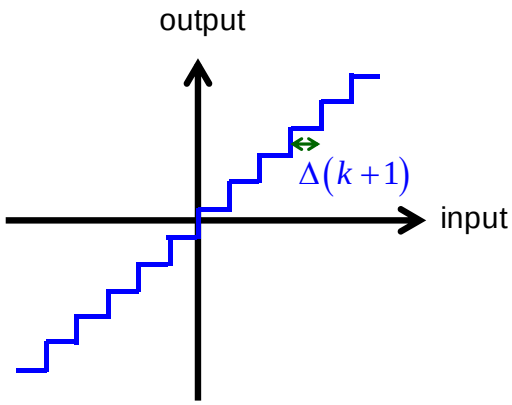
Three or Four?

Five

(a)

(b)





(c)

(d)

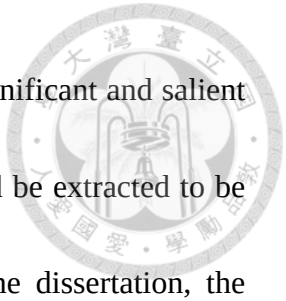
Figure 1.6. (a) and (b) are the captured frames by using the quantizer step size in (c) and (d), respectively.

1.2 Contribution

By exploiting Region-of-Interest-based (ROI-based) controlled sensing together with keyframe-based communication module it is possible to realize active smart sensing framework as shown in Figure 1.7 in which the video packet data reduction is performed by processing streams on the basis of spatial and temporal domains. Specifically, ROI-based controlled sensing consists of two parts: target acquisition and motion information analysis to achieve the desired information density of objects of interest and keyframe-based communication module also consists of two parts: data reduction and data transmission to achieve the successful transmission of important video packet data based on similar motion between frames. Furthermore, for the limitations imposed by a finite available data rate, layer-based transmission is utilized to guarantee the objects of interest with high priority to be transmitted.

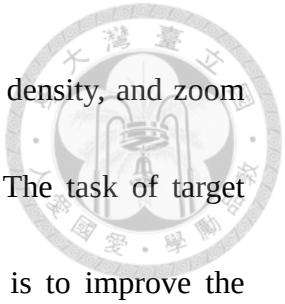
Within video technologies domain, compression based on frame connections and pixel similarities reduces the video data size. Moreover, the purpose of compression technologies is to decrease quality of objects of interest not to increase the quality. The extra-cost is needed to enhance the quality of the objects when they are represented with little pixels in the beginning. On the other hand, traditionally, Nyquist/Shannon sampling theory has been used to decide a proper sampling rate to sample a band-limited signal and then a low-pass filter is used to get the original signal [16:

Oppenheim & Willsky 1997]. However, the images are not naturally band-limited and the sampling rate for image retrieval is determined by the desired temporal or spatial resolution [17: Candès & Wakin 2008]. Hence, compressed sensing which is a novel data sensing and sampling theory recently receives increasing attention and be applied in various areas, such as signal reconstruction [18: Donoho 2006], image retrieve [19: Lustig & Donoho 2008], medical imaging devices, radio receivers and so on [17: Candès & Wakin 2008]. It overcomes the limitations of Nyquist/Shannon sampling theory that requires the system bandwidth to be limited and the sampling rate has to be at least twice the maximum frequency. However, compressed sensing focuses on the theoretical development so far and the digital image/video camera for implementing compressed sensing is still a very open research issue. Moreover, the sampling process is achieved offline and is applied to medical images retrieval, such as CT or MRI [19: Lustig & Donoho 2008]. From a time-critical point of view, however, compressed sensing is not a proper method for immediate image analysis tasks and real-time control applications. In particular, compression algorithms and compressed sensing employ fewer pixels or measurements to retrieve the image with lower quality. In comparison to them, the proposed ROI-based controlled sensing is able to increase or decrease the quality of the objects of interest based on the demand of human perception or video analysis algorithm.



On the other hand, video summarization is comprised of the significant and salient frames called keyframes. To decide whether or not the frame should be extracted to be the keyframe, the extraction criterion is critically important. In the dissertation, the proposed keyframe extraction rules for quantity control are based on the evolution of system dynamics. Moreover, the extraction threshold is determined by the robustness of controller and the statistical index, namely inter quartile range (IQR) for adapting various applications. To demonstrate the applicability of the proposed keyframe extraction method, ten typical and experimental videos are used to show the satisfied results. An example of keyframe extraction with typical tested video, Foreman, can be seen in Figure 1.8. In this figure, F is original captured frame set and $\tilde{F}$ represents the extracted frame set. Traditionally, face tracking algorithm is implemented in F and generates the corresponding face detection results. Because the channel capacity is finite, using $\tilde{F}$ saves the computation resource for processing the frames and reduces the required bandwidth for transmitting those. More importantly, the proposed keyframe-based quantity reduction method has similar or better performance compared with those using all frames. For example, similar face detection results can be seen in Frames 97 and 98 and Frame 109 in $\tilde{F}$ shows better result compared with that in F.

The whole procedures of ROI-based controlled sensing and keyframe-based communication module are depicted in Figure 1.9. In this figure, ROI-based controlled



sensing utilizes three parts: target acquisition, calculate information density, and zoom in/out control to achieve the quality control of object of interest. The task of target acquisition, calculate information density and zoom in/out control is to improve the target detection and tracking results by executing appropriate zoom in/out control based on the information density estimation of the target. The contribution of ROI-based controlled sensing module for quality control is presented in Chapter 3. In keyframe-based communication module, first of all is to compute the perceived motion energy (PME) value and then to sample the keyframes based on the proposed extraction rules with prior given threshold λ_{ctrl}^{scene} that is determined by motion characteristic in the monitored scene. In order to reliably transmit the important video packet data to server, end-users and controllers, layered-based transmission mechanism with assigned image quality value λ_{PSNR}^{scene} performs progressive transmission. The contribution of keyframe-based communication module for quantity control is described in Chapter 4. Furthermore, the proposed ROI-based controlled sensing and keyframe-based communication modules are experimented on different indoor and outdoor environments. Figure 1.10 shows an example of experimental scenario in which the visual sensor captures many images $F=\{F(k),k=1,\ldots,N\}$ of the environment E based on the proposed ROI-based quality control and then the salient images $\tilde{F}=\{F(j),j\in\Theta\}$ are sampled for further store, monitor or analyze based on proposed

keyframe-based quantity control.

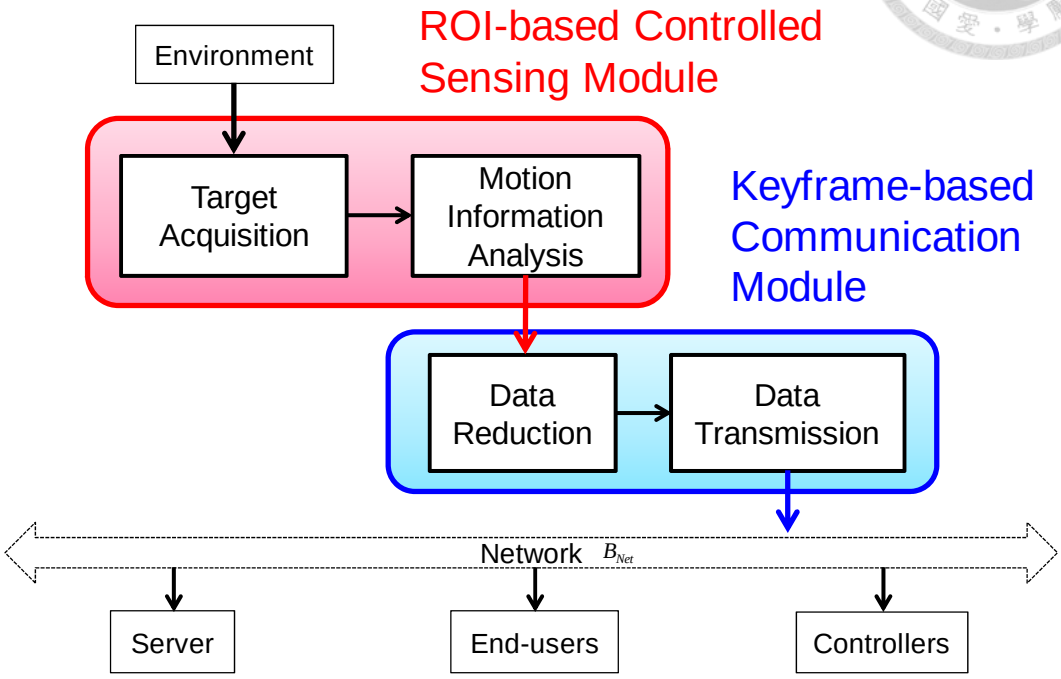


Figure 1.7. The framework of smart sensing consists of ROI-based controlled sensing and keyframe-based communication module.

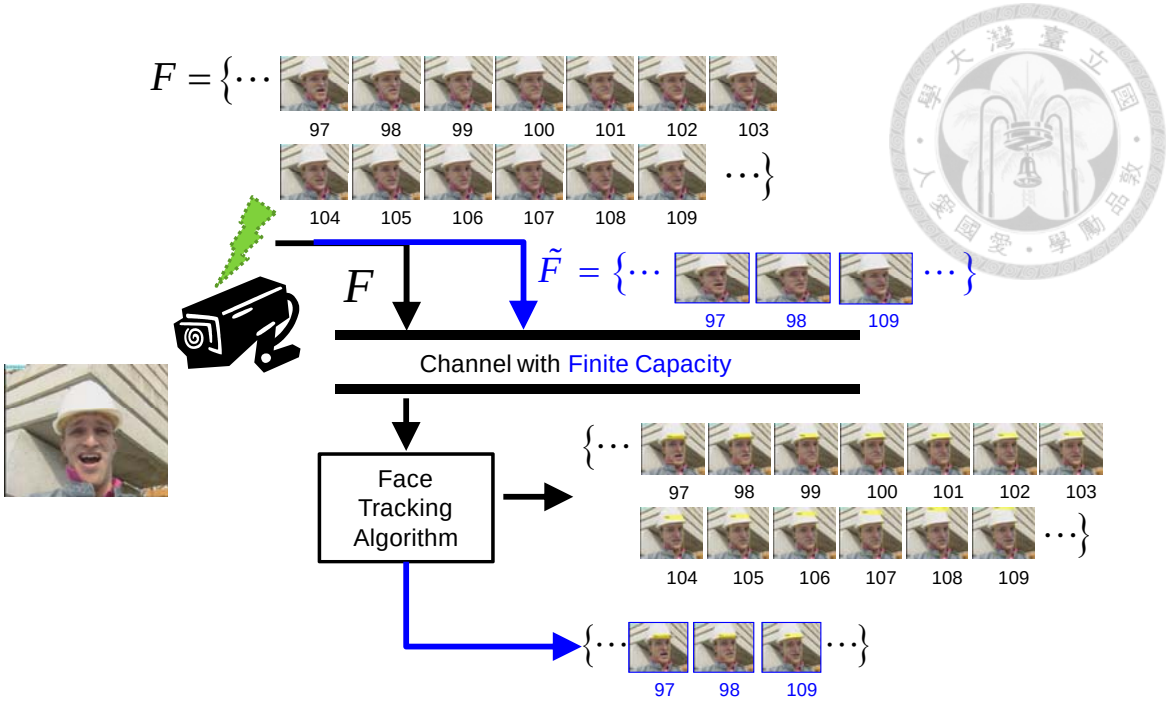


Figure 1.8. The example of keyframe extraction process shows the tracking results of the original captured and the extracted frame set.

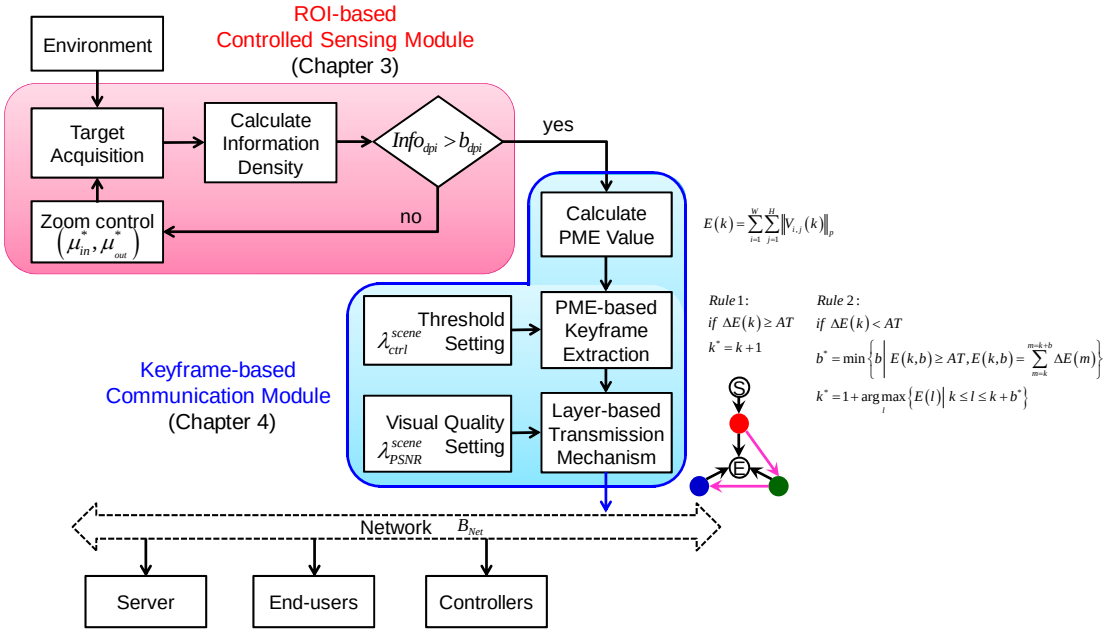


Figure 1.9. The procedures of ROI-based controlled sensing and keyframe-based communication module.

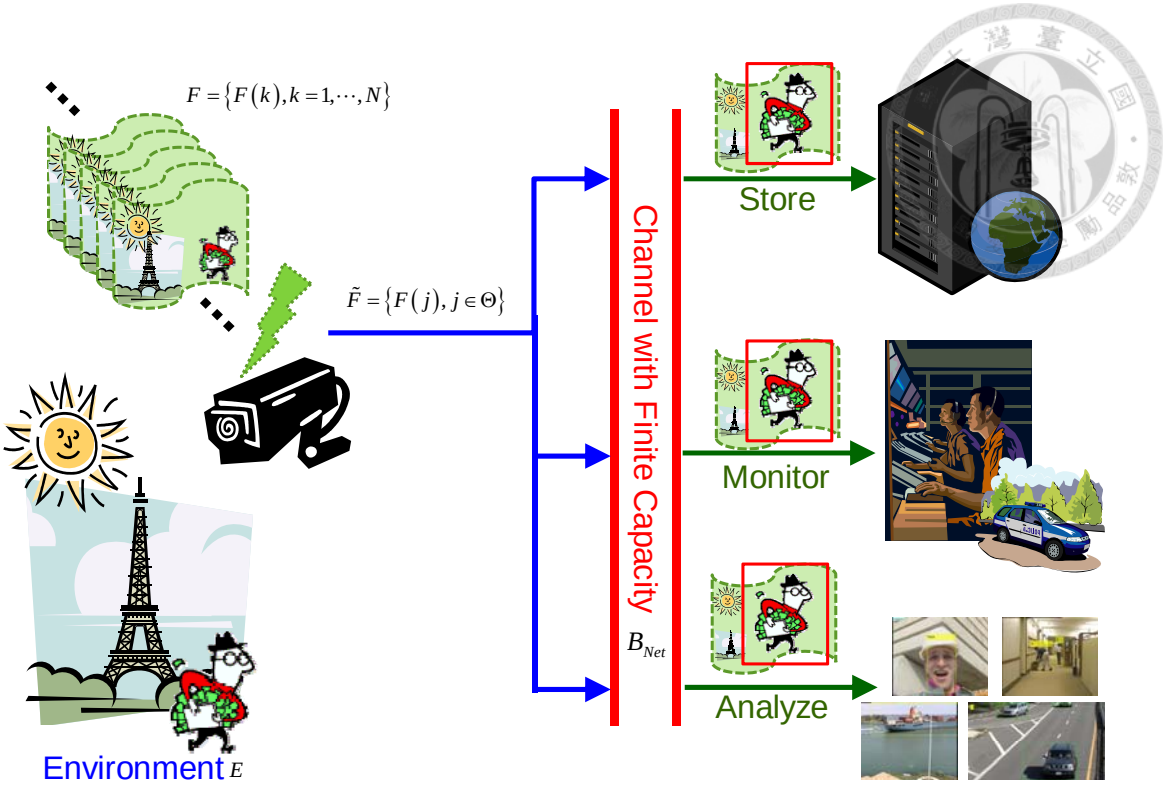


Figure 1.10. The diagram of the proposed quality and quantity control system which is applied to visual-sensing system.

1.3 Organization

The remainder of the dissertation is organized as follows. Chapter 2 reviews the related works about quality and quantity control. ROI-based controlled sensing for quality control is presented in Chapter 3. The quantizer step size is considered as zoom parameter and ***Theorem 1*** is proposed to prove the stability of linear system to be guaranteed with the proposed control policy, namely, zoom in or zoom out. Specifically, in each time step, the optimal zoom operation is performed based on the required information density of controller for stabilizing the system. Moreover, camera with zoom functionality is used to demonstrate the quality control operation in tracking object and compares the results to those without quality control.

Keyframe-based communication module for quantity control is presented in Chapter 4. On the basis of perceived motion energy resulted from motion vectors, the keyframe extraction rules are designed based on the evolution of system dynamics. Experimental videos consist of abundant scenes with different application goals are presented, such as tracking, detection or transmission in indoor or outdoor environments. For each typical or experimental video, the proposed keyframe extraction rules could greatly reduce the video data size and it is important that system performances or image analysis results are maintained in a satisfied level. Chapter 5 concludes the dissertation and suggests the future works.

Chapter 2

Literature Survey

Video-related applications have been grown significantly within the last decades due to the ubiquitous visual equipments and mature image processing techniques. They have found in various fields for assisting and/or monitoring, such as industrial robot, security and ecology surveillance [5: Hampapur et al. 2005], entertainment business, school [20: Karpenko & Aarabi 2011]. However, for practical purposes such as object detection, tracking and recognition, the insufficient quality of the objects of interest would affect the results [21: Tordoff & Murray 2003]. Hence, there is a tradeoff between the required quality and sensor coverage in building a visual sensor network [22: Chen et al. 2008], [23: Micheloni et al. 2010]. Moreover, video sharing and transmission become more convenient due to the development of Ethernet and telecommunication technologies [24: Han et al. 2011]. However, in the face of constraints on communication bandwidth, it is very difficult to transmit excessive video packet data over communication networks [25: Huang et al. 2009]. The methodologies and strategies in quality and quantity control are surveyed and discussed in the following for enhancing video quality and reducing video quantity.

2.1 Quality Control

Pan-tilt-zoom (PTZ) cameras have been utilized for increasing the resolution of moving targets and adapting sensor coverage in surveillance applications [4: Foresti et al. 2005], [5: Hampapur et al. 2005], [23: Micheloni et al. 2010]. The challenges in the scope and implementation of PTZ cameras involve controlling PTZ parameters and calibrating the moving cameras. The controllable zoom lens is the significant parameters for providing the desired quality of targets to improve the results of detection [26: Dinh et al. 2011], tracking [21: Tordoff & Murray 2003], [27: Micheloni & Foresti 2005] or recognition [26: Dinh et al. 2011], so called purposeful zooming [28: Tordoff & Murray 2004]. However, the field of view of the PTZ camera is changed after performing zoom action and it imposes the difficulties for relocating the moving objects. Hence, zoom parameter has been integrated into the motion model of the objects in earlier works. For example, the measurement model is scaled by the focal length f when Kalman filiter is used to do tracking tasks [21: Tordoff & Murray 2003]. For preserving the consistent target size in the image domain, a proper local length f is computed with perspective projection method [28: Tordoff & Murray 2004], [29: Dinh et al. 2009]. The major disadvantage of PTZ camera compared with the other static cameras is the narrow angle view when it stills. Hence, a number of researchers combine static and PTZ cameras to establish smart active camera networks [4: Foresti et al. 2005], [5:

Hampapur et al. 2005]. The major objective of static cameras is to find the potential targets and then send their coordinates to the corresponding PTZ cameras to obtain clearer target images by performing functionality of zooming in and out [22: Chen et al. 2008], [30: Ding et al. 2012], [23: Micheloni et al. 2010].

On the other hand, on the basis of binary representation of videos or frames, the problem of increasing resolution of the targets could be regarded as quantization of signals. How to adjust quantizer step size to modify the resolution of measurements by employing adaptive quanitzer is studied in [31: Goodman & Gersho 1974]. For avoiding the system stability is considerably affected by quantized output/state feedback, several studies propose proper control policies [32: Brockett and Liberzon 2000], [33: Delchamps 1990], [34: Liberzon 2003]. The control policies consist of two stages in which first stage is increasing the quantizer step size and second one is decreasing to guarantee the asymptotically stable with restricted quantization regions [32: Brockett and Liberzon 2000]. The quantization regions restriction and stability result in [32: Brockett and Liberzon 2000] has been extended to arbitrary shapes and nonlinear systems by integrating the quantizer step size into system states as hybrid systems [34: Liberzon 2003]. Moreover, a logarithmic quantizer comes from the Lyapunov function that is used to verify the stability of SISO linear discrete-time systems [35: Elia & Mitter 2001].

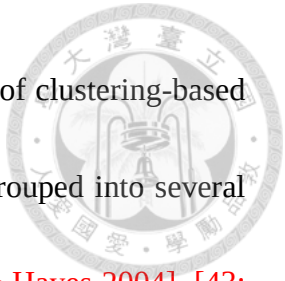
With the development of networks, the lower bound of the channel rate for achieving asymptotically stable in linear discrete-time systems associated with a primitive quantizer is presented in [36: Tatikonda & Mitter 2004]. Moreover, a unified framework of networked and quantized control systems (NQCS) for discussing the stability properties in the presence of quantization and/or time-scheduling protocols is provided in [37: Nesci & Liberzon 2009].

The quantizer step size can be regarded as zoom parameter of PTZ camera. Moreover, on the basis of appropriate data resolution, system stability would be guaranteed [32: Brockett and Liberzon 2000]. In the dissertation, on the basis of desired information density of the objects of interest, the lower bound of quantizer step size for stabilizing is provided and zoom controls are designed and implemented for adapting to the motion behaviors of them. More details would be presented in Chapter 3.

2.2 Quantity Control

Recently, video abstraction is one of methods for solving the problem caused by excessive data in a variety of video applications, such as retrieval, browsing, editing [38: Chen et al. 2010], annotation and indexing [39: Chatzigiorgaki & Skodras 2009]. It can be further divided into two categories: video skimming and video summarization. Video skimming generally extracts portions of video and audio data for producing a short synopsis, such as DVD program list. Video summarization generates a summarization which consists of salient and significant frames so called keyframes and rapidly receives attention in multimedia applications [22: Chen et al. 2010], [40: Li et al. 2006]. Because the characters of video skimming are that the whole video is required and the skimming process is offline, it is not appropriate manner for time-critical video applications. Hence, the concept of the video summarization is adopted in the dissertation to reduce video data in temporal domain.

Video summarization highly depends on the selection of keyframes. Keyframe extraction method has two different concepts, namely high-level and low-level [41: Jiang & Qin 2010]. High-level concept means to extract the keyframes as well as human cognition do. Hence, for modeling the human perceptive, visual attention clues are necessary to be considered and [41: Jiang & Qin 2010] designs a visual attention index (VAI) for extracting the keyframes. However, most of keyframe extraction approaches



are based on low-level concept which can be divided into two types of clustering-based and individual-based. In clustering-based type, the video is firstly grouped into several shots and then extracts the keyframes from the each shot [42: Lee & Hayes 2004], [43: Zhuang et al. 1998]. Hence, how to group the video, how to extract and how many extracted keyframes are three important issues in clustering-based type.

The intuitive grouping method is to divide the video with equi-distance into several shots and then extracting keyframes from them. The drawback to intuitive manner is that the distance needs to be assigned prior which results in less adaptability for various applications. Recently, K-Means is usually used to group the video in which the number of clusters is firstly assigned and then frames are assembled on the basis of the specific distance measurement manner between low features. After grouping stage, keyframe is the frame which is nearest the cluster center [42: Lee & Hayes 2004], [43: Zhuang et al. 1998]. The low-level features include color, texture, and/or sophisticated motion analysis, video content and even shot activities [41: Jiang & Qin. 2010]. Examples of distance measurement manner include Gaussian mixture model (GMM) with color histograms [44: Hammound & Mohr 2000], shot reconstruction degree interpolation (SRDI) algorithm with motion vectors [45: Liu et al. 2004], genetic algorithm based on information theory [46: Zeinalpour 2009] and iso-content principles consist of iso-content distance, iso-content error, ans iso-content distortion [47: Panagiotakis et al.

2009]. Other works map the keyframe extraction issue into tasks of normalized cut [48: Ngo et al. 2005], MinMax distortion optimization [49: Li et al. 2005], and manifold-learning [50: Zhang et al. 2010]. Moreover, the number of keyframes can be automatically determined in supervised or unsupervised manner by using Bayes information criterion [51: Schwarz 1978] or combination of GMM and expectation–maximization algorithm (EM) [52: Celeux & Govaert 1995].

In individual-based type, on the basis of the similarity between two continuous frames in features, the keyframes are extracted without clustering process [53: Guan et al. 2013]. Examples of method with local features include scale-invariant feature transform (SIFT) [53: Guan et al. 2013] and feature-based sparse representation [55: Kang et al. 2011] and with global features, such as color [38: Chen et al. 2010], motion [56: Liu et al. 2003], texture, and edge [54: Anagnastopoulos et al. 2009] are usually used for evaluating similarity in images. Moreover, the combination of image quality measurement and human excitement is presented in [58: Ratsamee et al. 2013].

For shortening the semantic gap between low-level and high-level, researchers combine with low-level concept and objects segmentation which aims to segment the foreground that invokes human attention such as human or car. The keyframes are extracted on the basis of the objects segmentation results [59: Liu & Fan 2005], [60: Song & Fan 2006].

For time-critical video reduction and video analysis, clustering-based is not a proper method due to its processing stages, conversely, individual-based is more appropriate based on the keyframe extraction [60: Song & Fan 2006]. However, the amount of the keyframes mainly depends on the similarity threshold. Specifically, large threshold would result in a coarse representation and small one would decrease the video reduction performance [53: Guan et al. 2013]. In the dissertation, the proposed keyframe extraction based on motion feature which belongs to individual-based type is discussed further in Chapter 4. The proposed keyframe extraction rules are on the basis of the evolution of motion energy and more flexible and adaptable for various video applications and video analysis algorithms in comparison to other approaches. Figure 2.1 shows the relationship of surveyed methodologies for keyframe extraction.

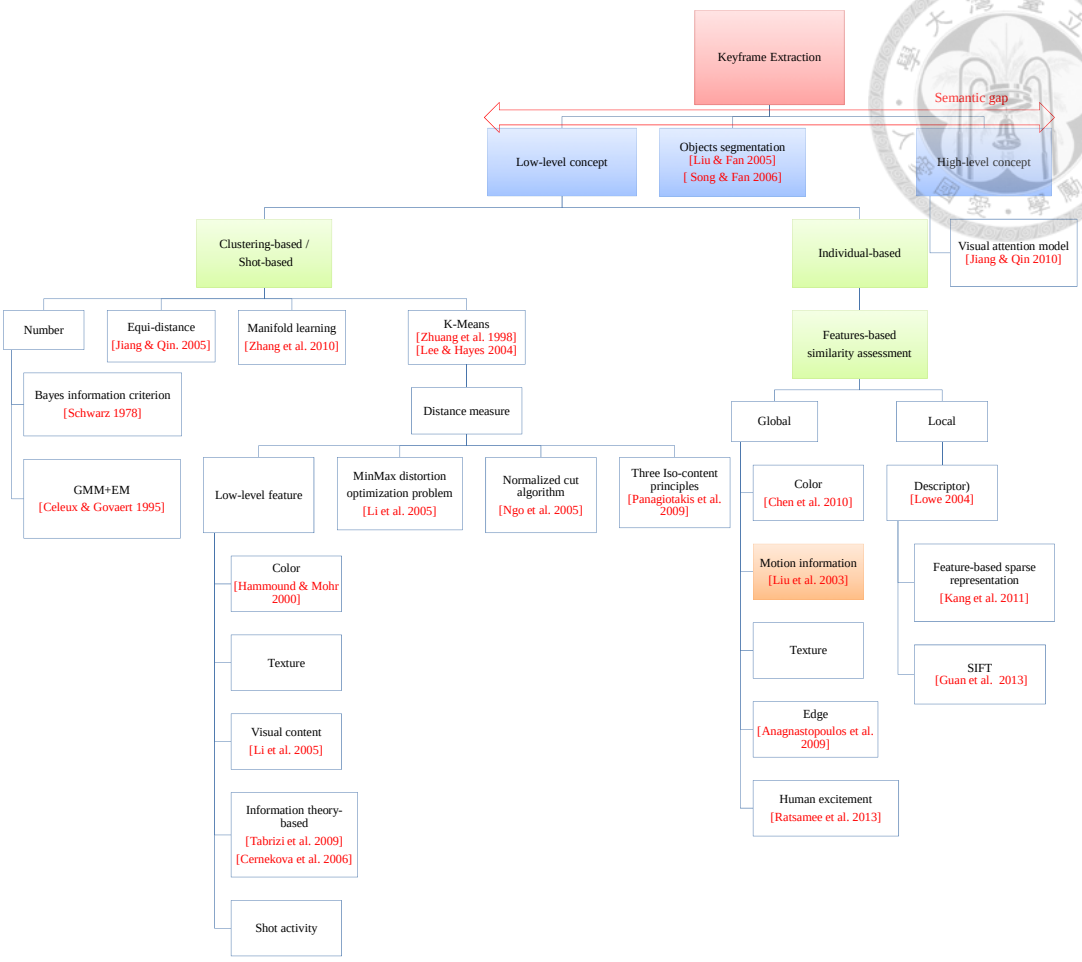
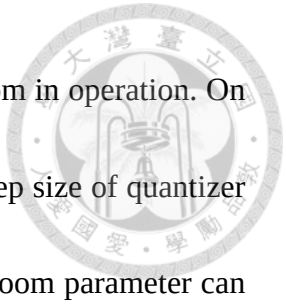


Figure 2.1. The relationship of surveyed keyframe extraction approaches in the literatures.

Chapter 3

Region-of-Interest-Based Quality Control

For the growing demands in homeland security and public safety, visual sensors have received significant attention and been applied widely in variety of applications such surveillance, military defense and law enforcement [61: Raty 2010]. The essential limitation in the visual sensors is the quality of providing surveillance footage that is insufficiently to provide demanding clearness for identity recognition in surveillance applications. The reason results from the balancing among sensor coverage, visual sensor costs, and surveillance footage quality. For obtaining wide sensor coverage and low cost, the quality has to be sacrificed. However, the insufficient quality of potential targets has led to incorrect recognition results. With the development of PTZ cameras, combination of PTZ and traditional surveillance cameras has received particular attention and the combination establishes a leader-follower system for considering the monitoring coverage and the quality demanding of the objects [61: Raty 2010]. PTZ cameras enable to change their field of view (FOV) by controlling the pan and tilt



parameters and to obtain better quality of objects of interest with zoom in operation. On the other hand, the zoom parameter of PTZ can be regarded as the step size of quantizer [32: Brockett and Liberzon 2000]. Hence, the problem of selecting zoom parameter can be modeled as a quantizer redesigning problem to select a suitable step size that stabilizes the system. The proposed ROI-based controlled sensing is inspired by the similar functionality between adjustable zoom and variable step size. In the dissertation, a zoom control policy is proposed which relies on the demanding quality of object of interest for video analysis algorithm and wide monitoring coverage while system evolves. The usage of pan and tilt mainly aim to keep the object of interest in the desired convergence zone and the usage of zoom focus to provide the demanding quality of the object for tracking, detection and recognition tasks. For example, the control objective of pan, tilt and zoon can be seen in Figure 3.1. In this figure, the goal for pan and tilt is to make the detected box of the object move into the desired convergence zone and the quality of the object of interest is increased by using zoom in

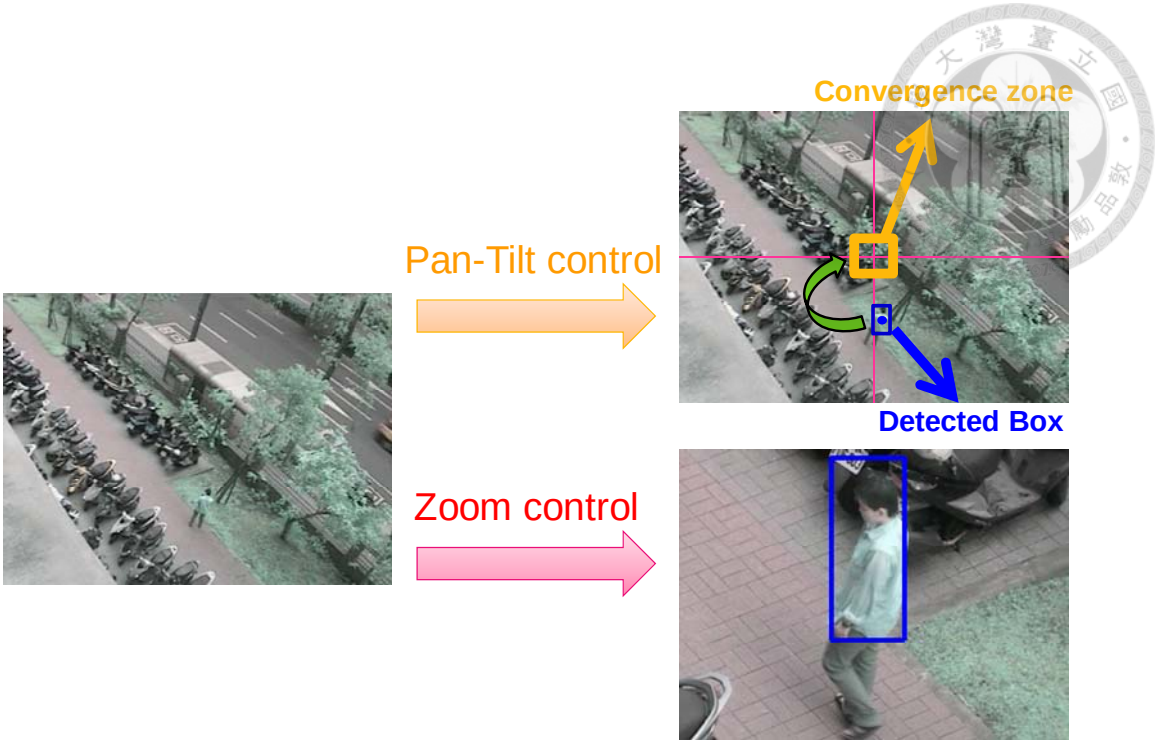


Figure 3.1. The illustration of operation and purpose of pan tilt and zoom.

3.1 Quantized State Feedback Stabilization

In order to implement the ROI-based controlled sensing, the zoom functionality is necessary for obtaining precise information of the object of interest. The zoom functionality not only acquires the demanding quality, but also improves the system performance. On the basis of the connection between zoom and step size of quantizer, the quantized state feedback stability of discrete time linear system is claimed with variable step size in the dissertation. The discussed discrete time linear system is as follows:

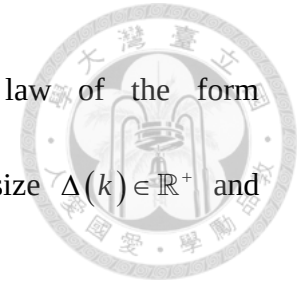
$$x(k+1) = Ax(k) + Bu(k) \tag{3.1}$$

where $x \in \mathbb{R}^n$, $u \in \mathbb{R}^m$ and A and B are matrices of suitable dimensions. After coordinate transformation, matrix A can be transformed into the form

$$A = \begin{bmatrix} A_s & \\ & A_u \end{bmatrix} \tag{3.2}$$

where A_s and A_u blocks correspond to the stable and unstable subspace. Without loss of generality, the marginally stable and unstable eigenvalues are mainly concerned in the dissertation. Next, used notations are introduced and defined for proving quantized state feedback stability with adjustable step size of quantizer.

If the controllable system is in the unstable mode, then there exists a matrix K such that all eigenvalues of $A - BK$ have negative real parts. In the situation, it is



reasonable to perform a quantized state feedback control law of the form $u(k) = -K\Delta(k)q(x(k))$ where $q:\mathbb{R}\to\mathbb{Z}$ is quantizer with step size $\Delta(k)\in\mathbb{R}^+$ and saturation $M\in\mathbb{Z}^+$ as follows:

$$q(x)=\begin{cases} M & if\ x>(M+\frac{1}{2})\Delta \\ -M & if\ x\le -(M+\frac{1}{2})\Delta \\ \left\lfloor \frac{x}{\Delta}+\frac{1}{2}\right\rfloor & if\ -(M+\frac{1}{2})\Delta < x \le (M+\frac{1}{2})\Delta \end{cases} \tag{3.3}$$

where $\lfloor x\rfloor \triangleq \max\{s\in\mathbb{Z}: s<x\}$. Next, let I_χ denote the indicator of a set $\chi\in\mathbb{R}$

$$I_\chi(\tau)=\begin{cases}1 & if\ \tau\in\chi \\ 0 & if\ \tau\notin\chi\end{cases} \tag{3.4}$$

The performance of a quantizer is measured by a distortion measure as follows:

$$D(q)=d(x,\hat{x})=d(x,q(x))=\|x-q(x)\|_2 \tag{3.5}$$

If quantizer q has $Q\in\mathbb{N}$ levels, then it needs $R(q)=\log Q$ bits to code for each input sample. Based on the distortion tolerant of the controller, says D, the operational rate-distortion function $\rho(D)$ that is the required infimum rate when quantizer distortion less than D could be defined. The formula of operational rate-distortion function is as follows:

$$\rho(D)=\inf_{D(q)\le D} R(q)=\inf_{D(q)\le D}\log Q \tag{3.6}$$

On the other hand, the controller performance depends on the data precision that relies on the step size of quantizer. From a dot density point of view, data precision can be written as information density. Within the image domain, information density Ξ is

defined to represent the quality of each pixel/unit as follows:

$$\Xi = \frac{\rho(D)}{(2M+1)\Delta} \tag{3.7}$$

that is, how many bits are used to describe it. Hence, improving video quality means to increase Ξ by adjusting Δ through controlling zoom variable $\mu \in \mathbb{N}$. The relationship between Δ and μ is assumed as $\Delta = \Psi(\mu)$ and Ψ is one to one and onto function. Zoom in and out mean to decrease and increase the step size, respectively.

Moreover, from a stabilizing point of view, an information rate is required for controller to achieve the goal. The lower bound of the required information rate only depends on the system eigenvalues as follows [36: Tatikonda & Mitter 2004]:

$$R > \sum_{\lambda(A)} \max\left\{0, \log\left|\lambda(A)\right|\right\} \tag{3.8}$$

It implies

$$R_{\min} = \sum_{|\lambda(A)|>1} \log\left|\lambda(A)\right| \tag{3.9}$$

If total transmitted pixels/units are Ω at a time, then total transmitted information rate to the controller is $\Omega\Xi$. For stabilizing the plant, the following inequality should be satisfied based on Equation (3.9).

$$\Omega\Xi > R_{\min} \tag{3.10}$$

Substitution Equation (3.7) into Equation (3.10), the upper bound of the step size is

obtained as follows:

$$\Delta < \frac{\Omega\rho(D)}{R_{\min}(2M+1)} \tag{3.11}$$

In order to decrease consumption of channel rate and control energy, an optimal step size Δ^{opt} is chosen as follows and then the optimal zoom variable $\mu^{opt} = \Psi^{-1}(\Delta^{opt})$ is obtained.

$$\Delta^{opt} = \min\left\{\Delta_s \in \mathbb{R}_{>0} : \Delta_s < \frac{\Omega\rho(D)}{R_{\min}(2M+1)}\right\} \tag{3.12}$$

Using μ^{opt} in each zoom control step, it could greatly increase the date resolution and further improve system.

The proposed control policy with information density could stabilize the quantized state feedback system and proposed Theorem 1 is described as follows for verifying the proposed control policy [32: Brockett and Liberzon 2000].

Theorem 1:

Suppose that all eigenvalues of $A-BK$ lie in the unit circle. Then there exists a control policy

$$\begin{aligned}\Delta(k+1) &= G\left(\Delta(k), q\left(x(k)\right), \Xi(k), k\right)\\ \mu(k+1) &= \Psi^{-1}\left(\Delta(k+1)\right)\\ u(k) &= -KI_{[k_0,\infty]}(k)\Delta(k)q\left(x(k)\right)\end{aligned}$$

where q is quantizer with step size Δ and $k_0 \in \mathbb{N}$, such that the solution of the closed-loop system is bounded.

$$\begin{aligned}x(k+1) &= Ax(k) - BKI_{[k_0,\infty]}(k)\Delta(k)q\left(x(k)\right),\ x(0)\ \textit{arbitrary}\\ \Delta(k+1) &= G\left(\Delta(k), q\left(x(k)\right), \Xi(k), k\right),\ \Delta(0) = \Delta_0\end{aligned}$$

Proof:

Consider the system $x(k+1) = Ax(k) + Bu(k)$ and $u(k) = -K\Delta(k)q\left(x(k)\right)$, hence the system can be rewritten as

$$\begin{aligned}x(k+1) &= Ax(k) - BK\Delta(k)q\left(x(k)\right)\\ &= (A-BK)x(k) + BK\eta(k)\end{aligned} \qquad (3.13)$$

where $\eta(k) = x(k) - \Delta(k)q\left(x(k)\right)$

On the basis of the Lyapunov stability theorem for discrete time linear system [83: Chen 1999], $\exists$positive symmetry definite matrices F and H such that

$$\left(A-BK\right)^T F\left(A-BK\right) - F = -H \qquad (3.14)$$

Based on the assumption of the quantizer does not achieve saturation and $k = k_0$ is

chosen until the system state is adequately measured as follows:

$$\|x\|_\infty \le (M-\frac{1}{2})\Delta \tag{3.15}$$

$$k_0 = \min\left\{k \ge 1 \middle| \left\|q\left(x\left(k\right)\right)\right\| < M-1\right\} \tag{3.16}$$

Based on Equation (3.16), the magnitude of the representing quantization error η is obtained as follows:

$$\|\eta\| = \|x-\Delta q(x)\| \le \sqrt{n}\left\|x-\Delta\left\lfloor\frac{x}{\Delta}+\frac{1}{2}\right\rfloor\right\|_\infty \le \sqrt{n}\frac{\Delta}{2} \tag{3.17}$$

Next, in order to claim the stability in the sense of Lyapunov, standard Lyapunov function $V(x)=x^TFx$ is chosen. Whenever Equation (3.17) holds $x^T\left(k+1\right)Fx\left(k+1\right)-x^T\left(k\right)Fx\left(k\right)$ is given by

$$\begin{aligned}
&x^T\left(k+1\right)Fx\left(k+1\right)-x^T\left(k\right)Fx\left(k\right)\\
&=x^T\left(k\right)\left(A-BK\right)^T F\left(A-BK\right)x\left(k\right)+x^T\left(k\right)\left(A-BK\right)^T FBK\eta\left(k\right)\\
&\quad+\eta^T\left(k\right)K^TB^TF\left(A-BK\right)x\left(k\right)+\eta^T\left(k\right)K^TB^TFBK\eta\left(k\right)-x^T\left(k\right)Fx\left(k\right)\\
&=-x^T\left(k\right)Hx\left(k\right)+2x^T\left(k\right)\left(A-BK\right)^T FBK\eta\left(k\right)+\eta^T\left(k\right)K^TB^TFBK\eta\left(k\right)\\
&\le-\lambda_{\min}(H)\left\|x\left(k\right)\right\|^2+\left\|x\left(k\right)\right\|\left\|\left(A-BK\right)^T FBK\right\|\Delta\sqrt{n}\\
&\quad+\left\|K^TB^TFBK\right\|\frac{\Delta^2n}{4}
\end{aligned} \tag{3.18}$$

Let $\beta_r\left(x\right)$ denote the ball with radius r as follows:

$$\beta_r\left(x\right)=\left\{x\in\mathbb{R}^n \middle| \left\|x\right\| \le r,\ r=\Sigma\Delta\sqrt{n}\right\} \tag{3.19}$$

where $\Sigma=\dfrac{1}{2\lambda_{\min}(H)}\left(\left\|\left(A-BK\right)^T\right\|FBK+\sqrt{\left\|\left(A-BK\right)^TFBK\right\|^2+\lambda_{\min}(H)\left\|K^TB^TFBK\right\|}\right)$

Then, the last expression in Equation (3.18) is negative outside the ball $\beta_r\left(x\right)$. For each time, there exists a control policy to make the solution bounded. Hence, it could

conclude that the bounded property of the system solution with the proposed control policy. □

3.2 Region of Interest based Zoom Control

For implementing the adjustable quality control, the ROI-based controlled sensing framework is proposed in the dissertation as shown in Figure 3.2. ROI-based controlled sensing mainly consists of target acquisition, information density computation and pan-tilt-zoom camera controls. The proposed control policy in ***Theorem 1*** is to adequately detect target state at first and then gradually adjust the step size of quantizer. The corresponding relationships between ***Theorem 1*** and practical application are as follows:

$$G(k) = zoom(k) \tag{3.20}$$

$$u(k) = pan(k) + tilt(k) \tag{3.21}$$

In terms of PTZ camera controls, pan and tilt functions are used to keep the target in the center of sensed frame and then zoom variable is adjusted gradually base on the information density $Info_{dpi}$ and tracking error e of the target. The purpose of zoom control is to increase the quality of the target to improve the performance. On the other hand, in the literature, the PTZ cameras are used to integrate with Omnidirectional camera which provides the wide monitoring area to only provide the partial body image of the targets through zoom in [22: Chen et al. 2008]. In the dissertation, both quality and surrounding objects of the target are provided through the proposed zoom control which is applied in PTZ camera for surveillance applications.

The flowchart of proposed zoom control is shown in Figure 3.3. Zoom in is executed to increase quality of the target for performing more accurate pan and tilt controls. Zoom out is performed for including the surrounding objects of the target and/or preventing from false tracking result which might be caused by zoom action in the complex environment. Convergence zone b_{conv} is designed as the threshold for determining pan and tilt control convergence. Zoom operation would be executed until tracking error e lies at prescribed critical zone b_{conv}. In order to avoid the tracking result to go wrong due to the variation of background, tracking error is checked after each zoom operation. Zoom operations are divided into two parts: zoom in and zoom out with different increment δ_1 and δ_2, respectively and $\delta_1 < \delta_2$. The extreme value μ_{in}^* in zoom in mode is determined by prescribed desired information density b_{dpi} which depends on the ability of controller, image processing algorithm or the demand of human operators. More, the extreme value in zoom out mode is prior decided by μ_{out}^* which depends on the height of PTZ camera position. The main designing purposes of δ_1, δ_2, b_{dpi} and μ_{out}^* are to ensure the convergence in a progressive way and to take into account the information about surrounding objects. Two thresholds of the number of frames b_{Cf} and b_{Lf} are designed for switching between zoom in and zoom out. b_{Cf} and b_{Lf} represent the number of consecutive frames where the tracking error belongs convergence zone and the target does not be tracked, respectively. The optimal

zoom value is decided by

Zoom Control

$$
\begin{aligned}
&(1)\ Zoom\ In \\
&\mu(k+1) = \mu(k) + \delta_1 \\
&\mu_{in}^{*} = \min\left\{\mu \mid Info_{dpi}(\mu) > b_{dpi}\right\} \\
&(2)\ Zoom\ Out \\
&\mu(k+1) = \mu(k) - \delta_2 \\
&\mu(k) \to \mu_{out}^{*} \ \ as\ k \to \infty
\end{aligned}
\tag{3.22}
$$

in two types of zoom operations: zoom in and zoom out.

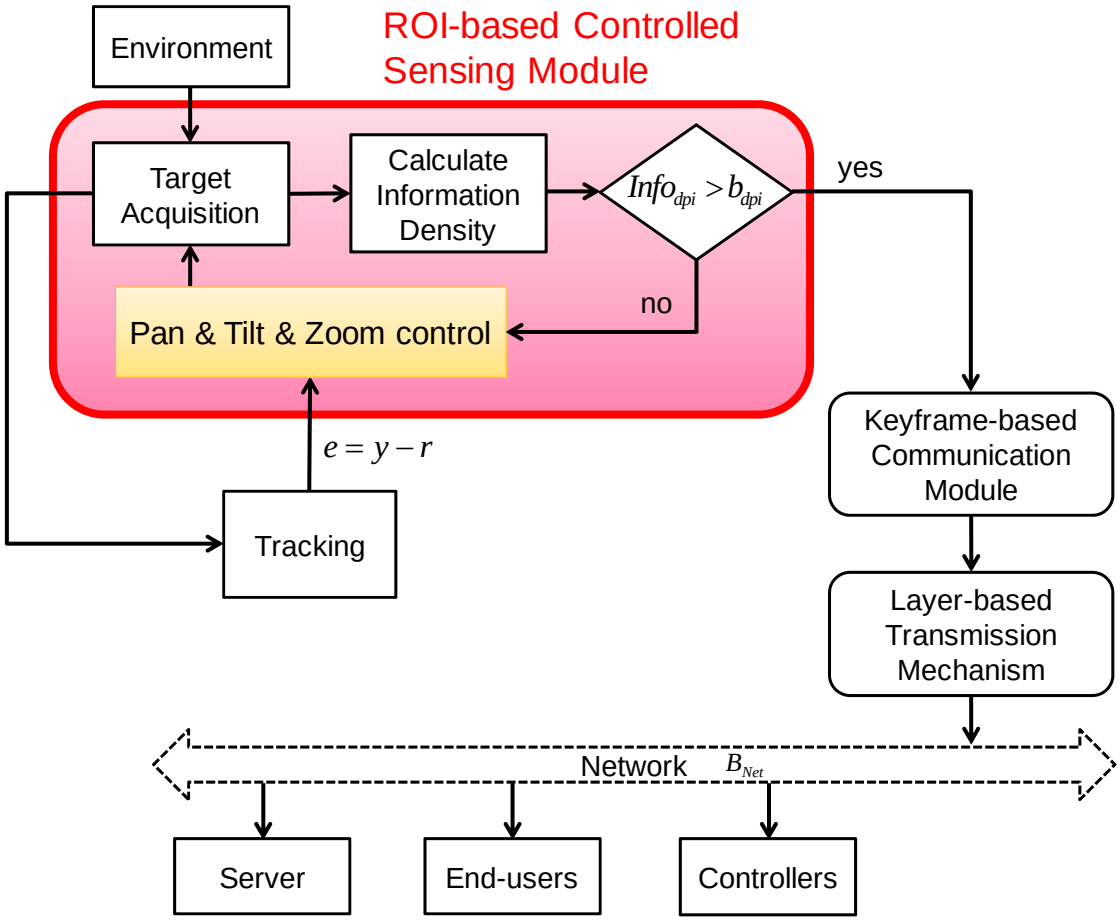


Figure 3.2. The proposed framework of ROI-based controlled sensing with pan, tilt and zoom controls. $Info_{dpi}$, b_{dpi}, y, r, and e represents the information density of the target, density threshold, output, reference and tracking error, respectively.

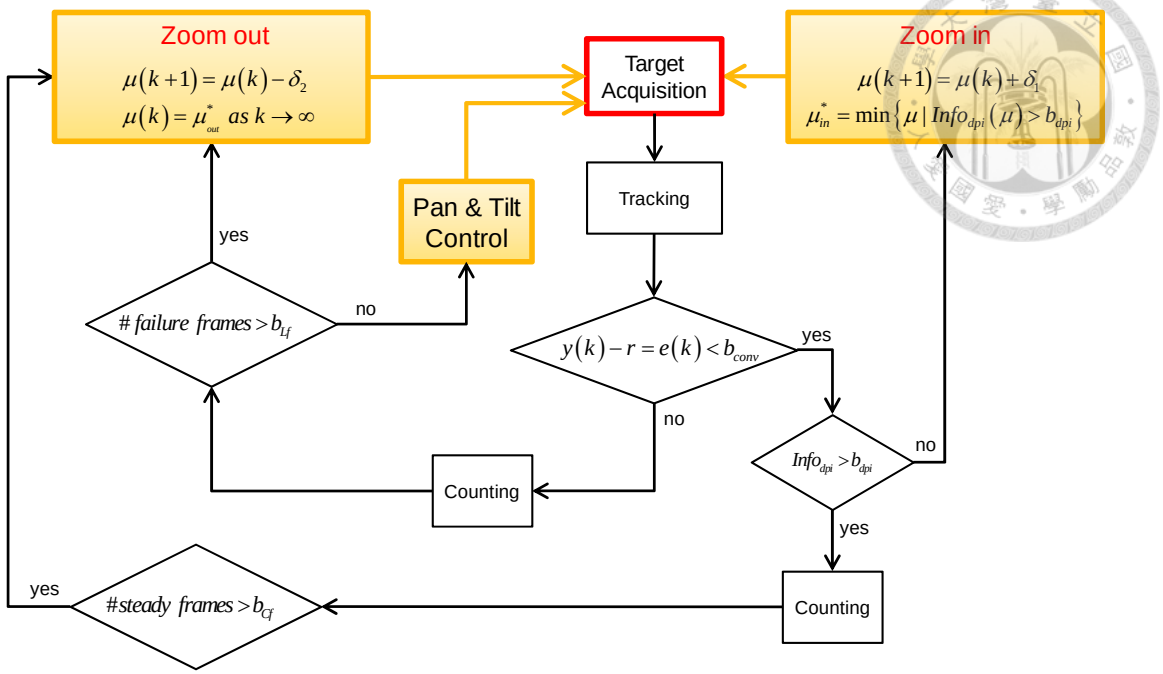


Figure 3.3. The flowchart of proposed zoom control.

3.3 Experimental Results of Region-of-Interest-Based Quality Control

3.3.1 Description

For demonstrating the efficiency of the proposed ROI-based controlled sensing framework, the designed zoom control policies are applied in AXIS P5512 PTZ camera as shown in Figure 3.4(a). Two experimental scenes: indoor and outdoor are used to present the applicability of proposed zoom control policies. Figures 3.4(b) and 3.4(c) show the indoor and outdoor experimental scene, respectively, at Ming-Da building in NTU and the view of PTZ. Moreover, the tracking algorithm used in the dissertation is based on the color information to track the particular target. However, when color information is insufficient or similar color occlusion happens, the tracking algorithm without zoom control easily has incorrect result. With the proposed zoom control policies, the tracking performance and the quality of the particular target could be greatly improved. It is especially useful for further video analysis such as recognition in criminal investigation.

(a) (b) (c)

Figure 3.4. (a) AXIS P5512 PTZ camera used in the dissertation. (b) Indoor experimental scene at MD building in NTU. (c) Outdoor experimental scene in NTU.

For measuring the information density, the percentage area of view of physical world FOV is used and defined as follows:

$$FOV(k) = \frac{1}{(\gamma \cdot \mu(k))^2} \times 100\%, \quad \gamma = 1 \tag{3.23}$$

where γ is the scale factor of view and for simplicity, let $\gamma = 1$ in the dissertation. Then information density $Info_{dpi}$ is defined as the reciprocal of FOV to represent how

many pixels are used to describe one unit

$$Info_{dpi}(k) = \frac{1}{FOV(k)} \quad \left(pixels/unit\right) \tag{3.24}$$

Furthermore, the convergence zone $b_{conv} = [160-\zeta_X, 160+\zeta_X] \times [120-\zeta_Y, 120+\zeta_Y]$ is shown in Figure 3.5. In this figure, the image center is $[160,120]$t and the convergence boundary of the center coordination (O_X, O_Y) of detected box is $[160-\zeta_X, 160+\zeta_X]$ for O_X and $[120-\zeta_Y, 120+\zeta_Y]$ for O_Y.

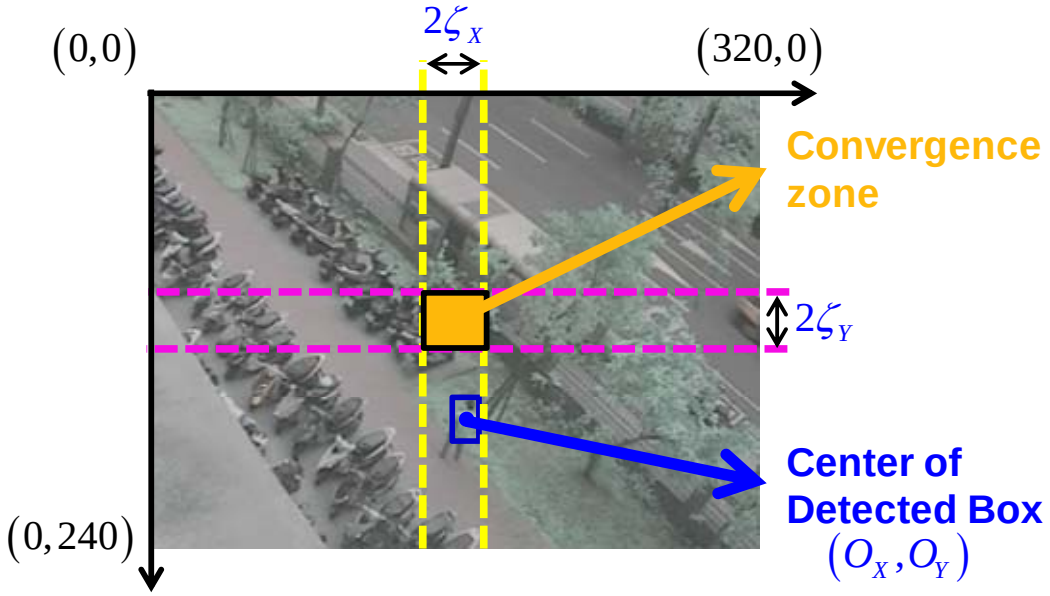


Figure 3.5. The convergence zone definition in image coordination.

3.3.2 Indoor Environment

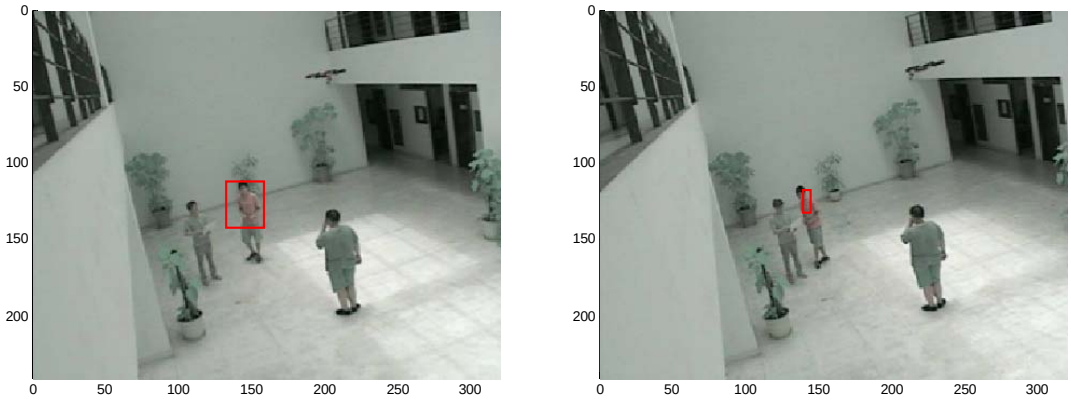
The indoor experimental scene at 5F of Ming-Da building in NTU is shown in Figure 3.6. Most of the background color is white with some green and brown and the color of target is red. Next, the experimental results are presented in two types: traditional (without zoom) and ROI-based controlled sensing (with zoom control).

Figure 3.6. The indoor experimental scene at 5F of Ming-Da building in NTU.

A. Traditional

Due to without zoom in/out control, hence the zoom value stays in the initial condition, i.e. $\mu=1$ as shown in Figure 3.8 and the red star represents the false detection results. The false condition is decided by the size of detected box. When the size of detected box is too small or too large, the detection results are classified to failure. An example of the false detection results can be seen in Figure 3.7. Figures 3.7(a) and 3.7(b) show too large and too small detected box, respectively. The reason for small detected box is insufficient sensed color information and the reason for large detected box is that the color of background objects are similar to the target color to be included and considered as target. Figure 3.9 (a) shows the snapshots during the tracking process. In this figure, the face and the gesture of the target are ambiguity and blur. Figures 3.9 (b) and 3.9(c) show the detection results, the performance of pan and tilt controls and the error bar in the figures is the width and height of detected box, respectively. In the figures, the black lines represent the designed convergence range and the trend of center coordination of detected box. In Figure 3.9(b), the black lines are $160-\zeta_X$ and $160+\zeta_X$ and $\zeta_X=13.33$. In Figure 3.9(c), the black lines are $120-\zeta_Y$ and $120+\zeta_Y$ and $\zeta_Y=12$. Although the trend of the center coordination stays in the designed convergence range, the most of detection results are failure. The designed convergence zone is $b_{conv}=[160-\zeta_X,160+\zeta_X]\times[120-\zeta_Y,120+\zeta_Y]$ where

$\zeta_X = 13.33$ and $\zeta_Y = 12$. Moreover, Figure 3.10 shows the information density that is low and fixed due to without zoom control. Although the center coordination of detected box stays in the designed boundary, the resolution of the target is too low to recognize the identity that is a serious problem for surveillance applications.



(a) (b)

Figure 3.7. The snapshots of failure detection result without using zoom functionality.

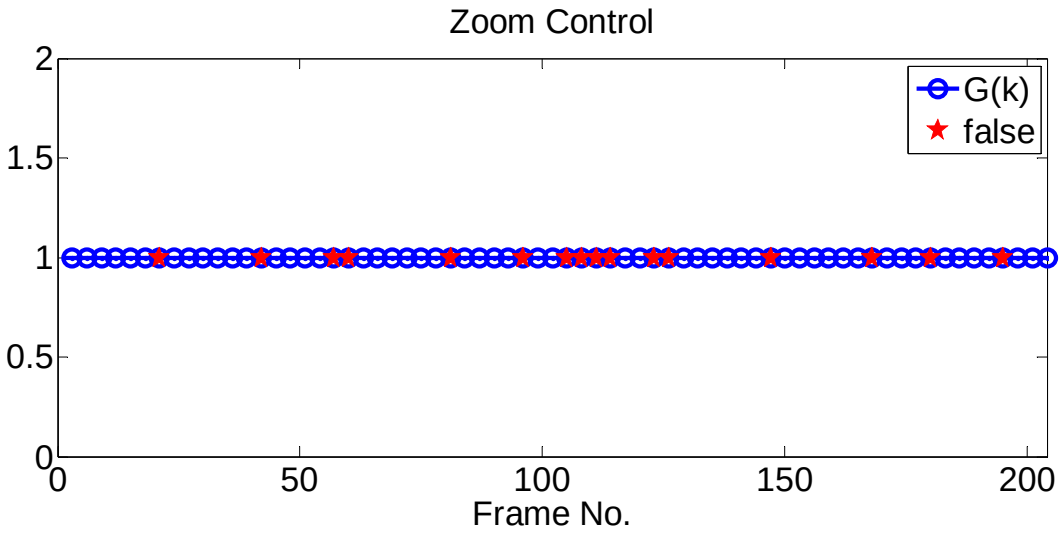
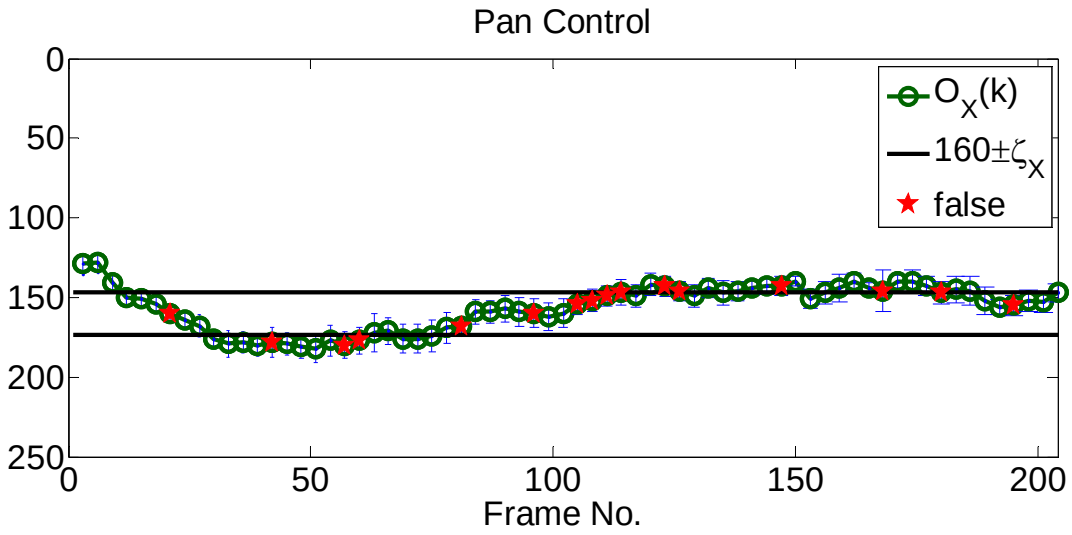
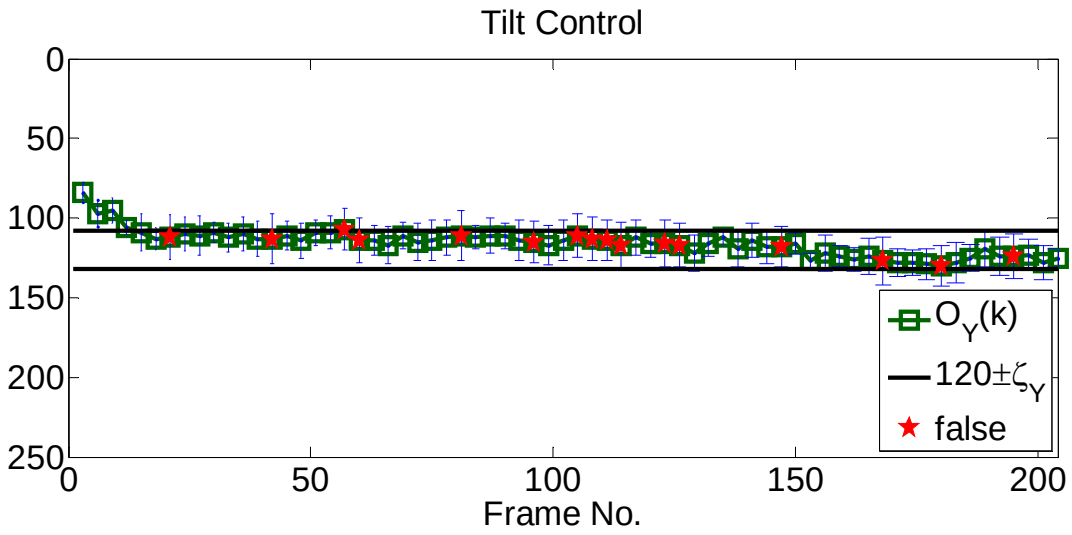


Figure 3.8. The zoom value is constant in traditional method.

(a)



(b)



(c)

Figure 3.9. (a) The snapshots of tracking results. (b) and (c) are the X-coordinate and Y-coordinate of center point of detected box, respectively.

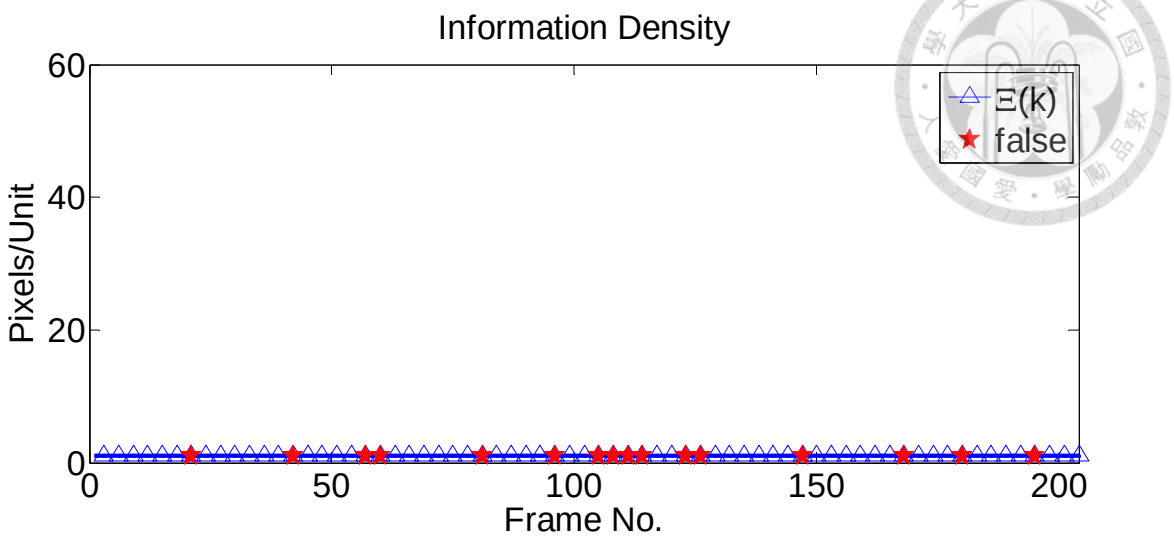


Figure 3.10. The information density is fixed in traditional method.

B. ROI-based controlled sensing

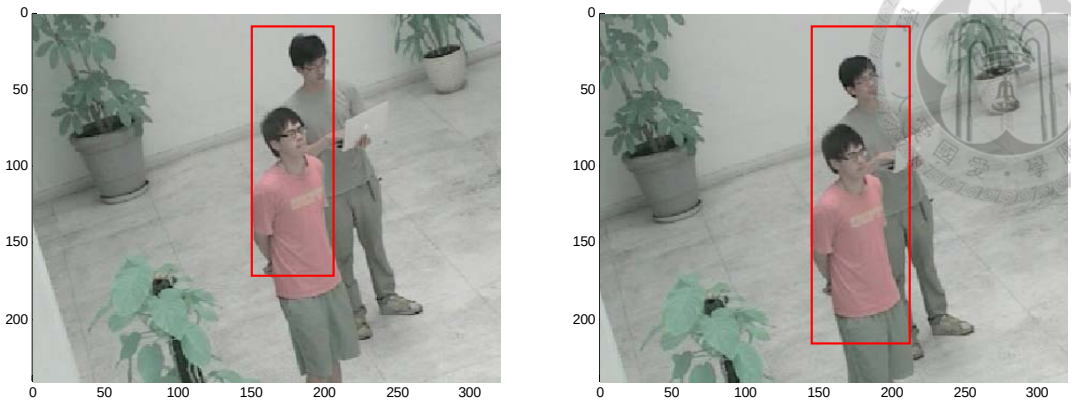
In the same experimental scene as shown in Figure 3.6, the proposed zoom control policies are applied to the tracking task. The parameters in the zoom control are $\delta_1 = 0.1$, $\delta_2 = 0.2$, $b_{dpi} = 20$ $b_{Cf} = 15$, $b_{Lf} = 3$ and $\mu^*_{out} = 1.2$. In particular, for adapting the effect caused by the zoom operation, the dynamic convergence zone $b_{conv} = [160 - \zeta_X, 160 + \zeta_X] \times [120 - \zeta_Y, 120 + \zeta_Y]$ is designed for effective eliminating the vibration from pan and tilt control in the process of zoom control. Moreover, ζ_X and ζ_Y are function of μ as follows:

$$\zeta_X(k)=\frac{320}{h_{pan}(k)} \qquad \zeta_Y(k)=\frac{240}{h_{tilt}(k)}$$
$$h_{pan}(k)=h_{pan}(k-1)+\Delta h \qquad h_{tilt}(k)=h_{tilt}(k-1)+\Delta h$$
$$h_{pan}(0)=24 \qquad h_{tilt}(0)=20 \qquad (3.25)$$

$$\Delta h=\begin{cases}-\delta_h & \mu(k)-\mu(k-1)>0\\ \delta_h & \mu(k)-\mu(k-1)<0\end{cases}$$

where $\delta_h=0.2$ in the dissertation.

Hence, the convergence zone becomes larger when zoom in is operated. Figure 3.12 shows the evolution of zoom behavior. At first zoom in is taken to increase the quality of target to the desired one and then zoom out is performed to take into account the behavior of surrounding objects. An example of the false detection results can be seen in Figure 3.11. Figure 3.13(a) shows the snapshots in the process of tracking. At first the target is detected and then the face and the gesture of the target become clear through zoom in. The last snapshot shows the detected result after zoom out. Figures 3.13(b) and 3.13(c) presents the tracking results and the performance of pan and tilt control with dynamic convergence boundary. The tracking results are marked improved, convergence is still persevered and the quality of targets is raised to significant enough for identity recognition. Figure 3.14 shows the trend of information density with zoom control. With the raise of information density, enough color information of target is provided to obtain better tracking results which are shown in Figures 3.13(b) and 3.13(c).



(a) (b)

Figure 3.11. The snapshots of failure detection result with using zoom function.

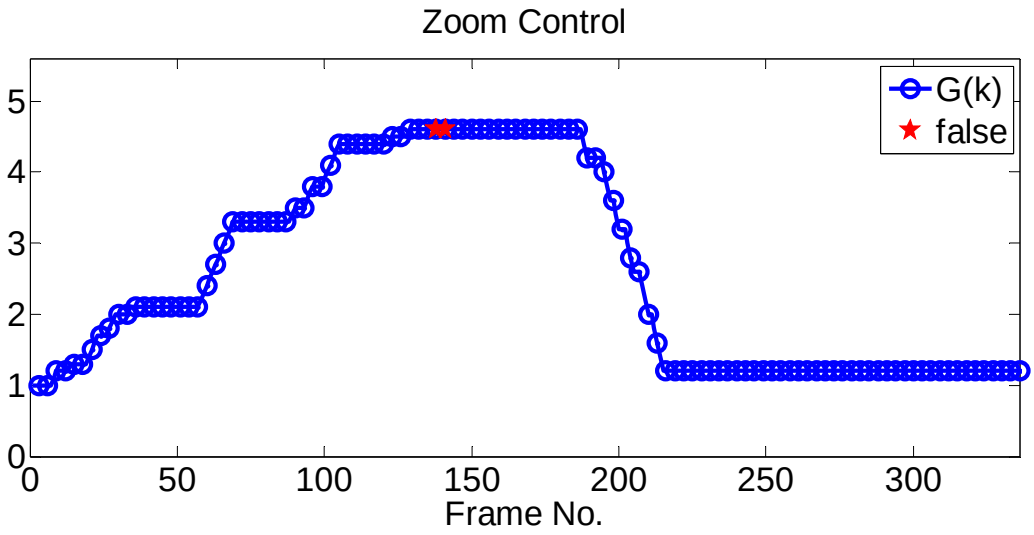
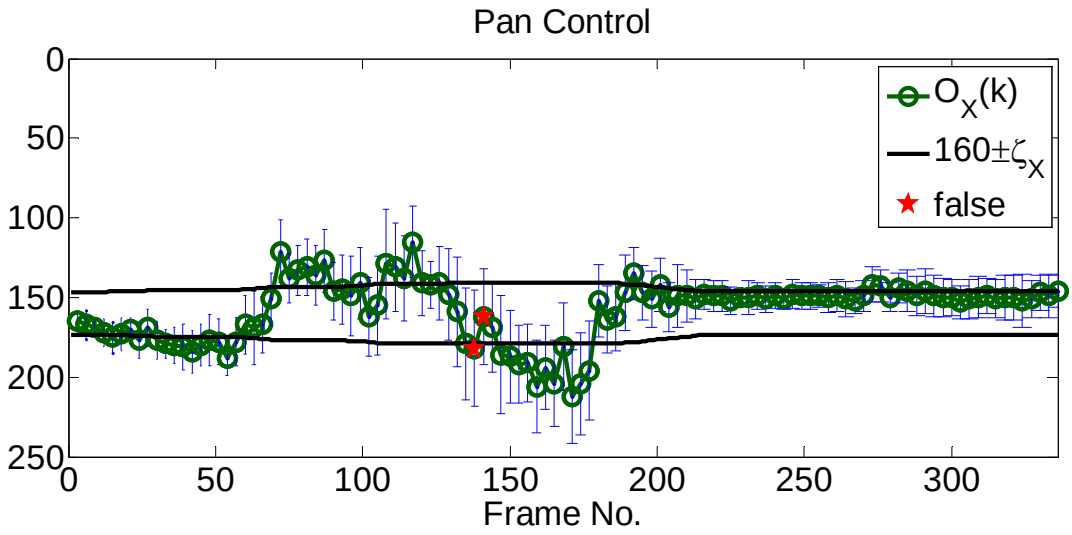
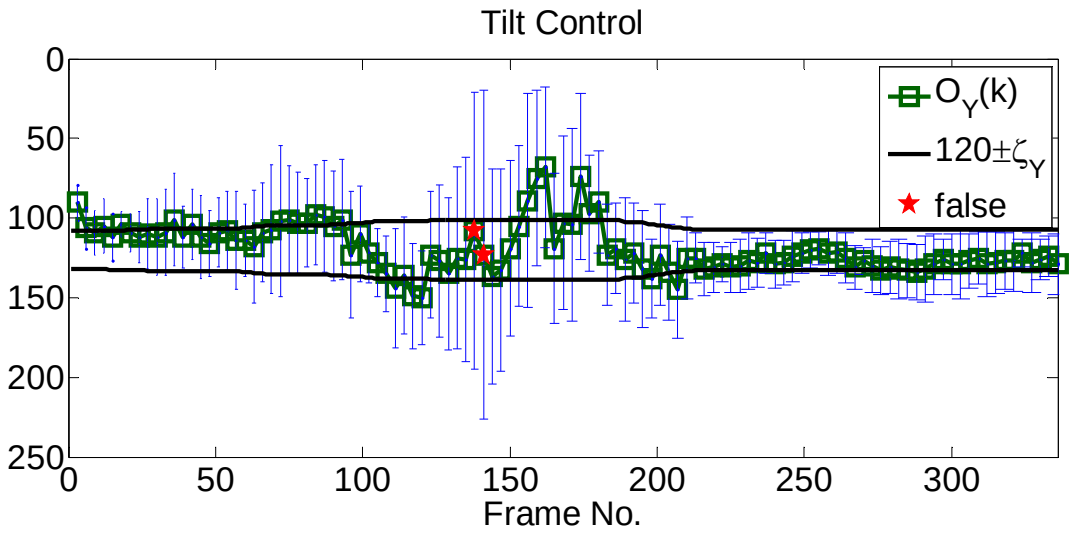


Figure 3.12. The zoom value in indoor experimental scene with ROI-based controlled sensing method.

(a)



(b)



(c)

Figure 3.13. (a) The snapshots of tracking results. (b) and (c) are the X-coordinate and Y-coordinate of center point of detected box, respectively.

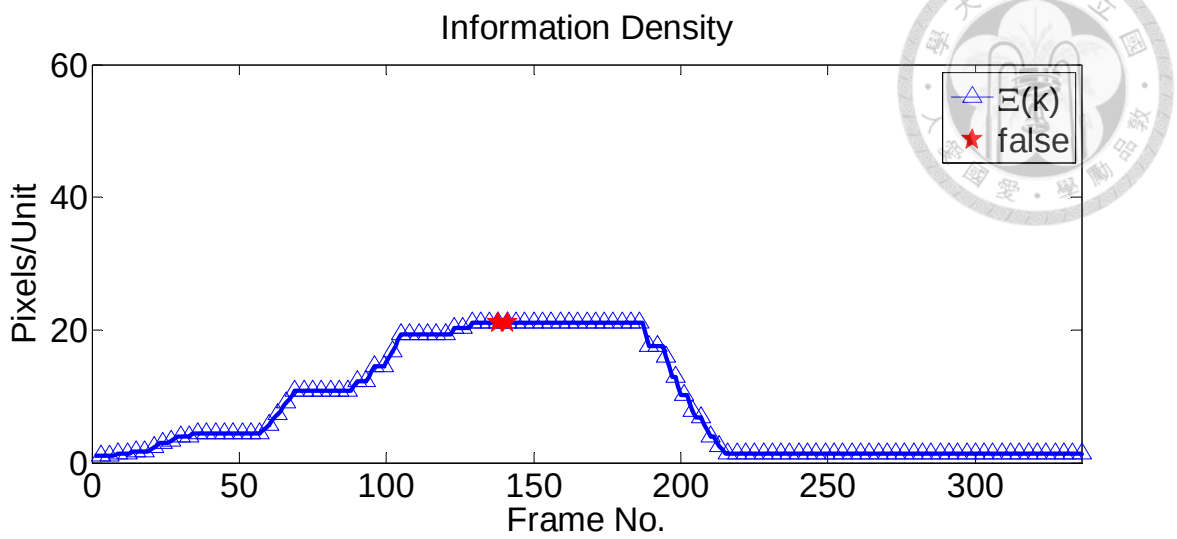


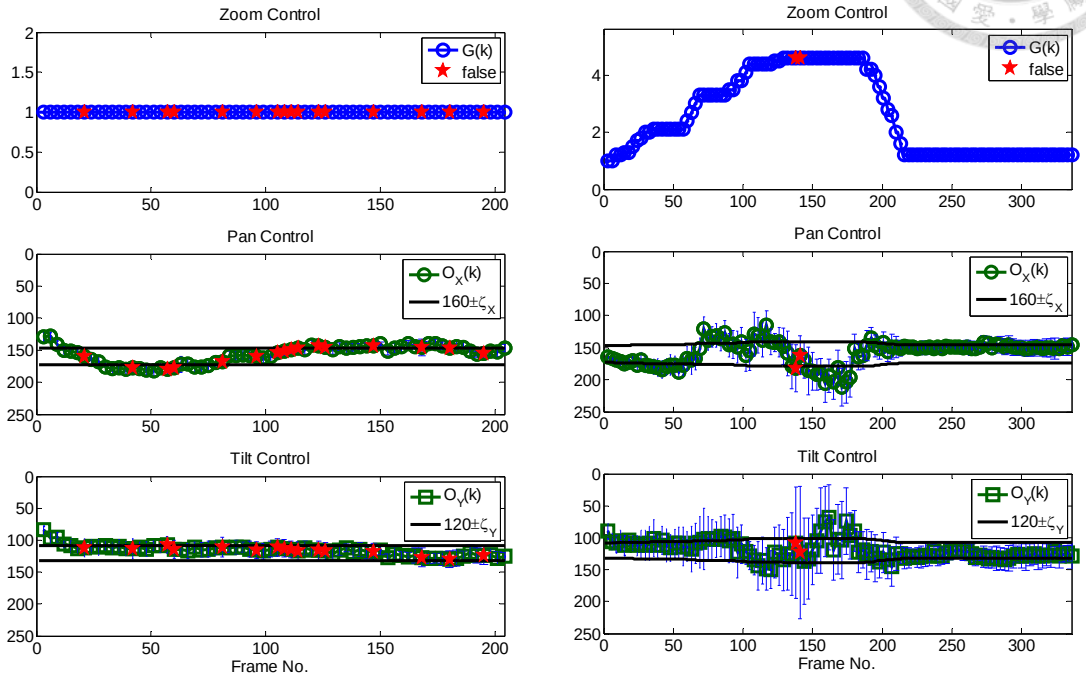
Figure 3.14. The information density in ROI-based controlled sensing method.

C. Comparison

In comparison of traditional approach, the proposed ROI-based controlled sensing method has better and satisfactory tracking results, adjustable quality of target for demands of human operator or video analysis algorithm, flexibility monitoring range for taking into account the behavior of target and other objects as shown in Figures 3.15 and 3.17. Figure 3.16 is the results from Frames 50 to 150 to demonstrate tracking results in detail.

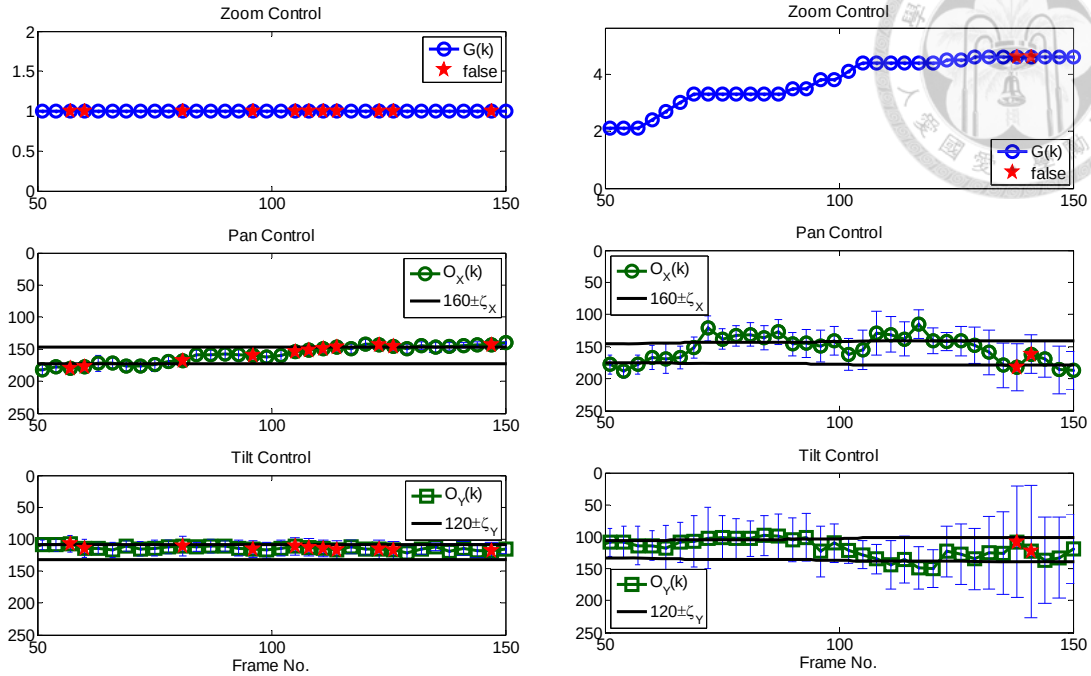
Because the indoor experimental scene is simple and color of target is salient than other objects, the difference of tracking result between traditional and ROI-based controlled sensing is not salient. For demonstrating the remarkable predominance of the proposed ROI-based controlled sensing, the complex outdoor environment scene is

adopted as experimental scene.



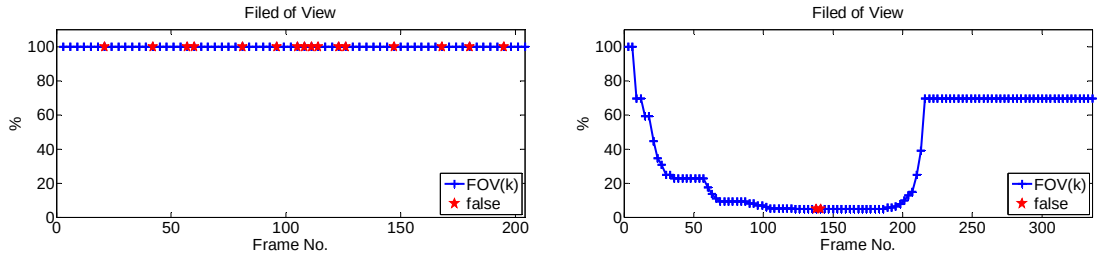
(a) (b)

Figure 3.15. Performance comparison of (a) traditional method and (b) ROI-based controlled sensing method.



(a) (b)

Figure 3.16. The partial figure of Figure 3.15 from Frames 50 to 150.



(a) (b)

Figure 3.17. The percentage area of viewed physical world. (a) Traditional way (b) ROI-based controlled sensing.

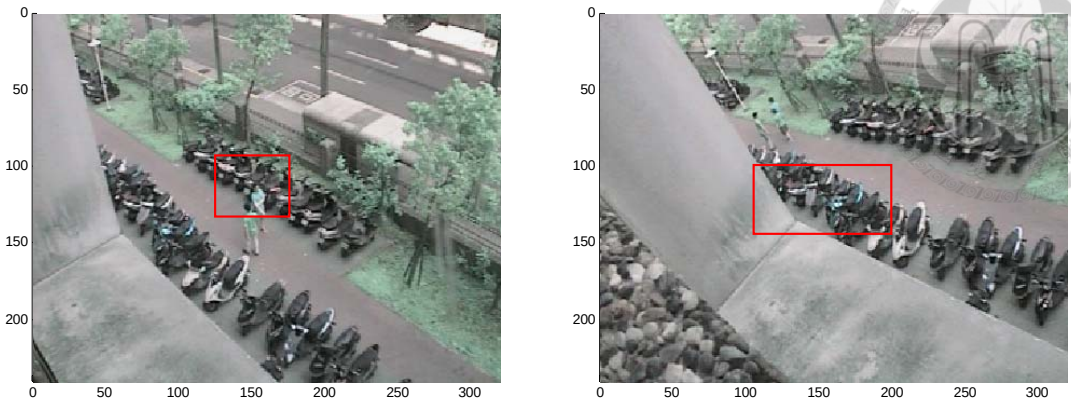
3.3.3 Outdoor Environment

The outdoor experimental scene of Ming-Da building in NTU is a motorcycle parking area as shown in Figure 3.18. The experimental scene has more complex color information and objects than indoor environment. The experimental results are also presented in terms of traditional (without zoom) and ROI-based controlled sensing (with zoom control).

Figure 3.18. The outdoor experimental scene at 3F of Ming-Da building in NTU.

A. Traditional

In the traditional method, zoom value stays in constant one that can be seen in Figure 3.20 and information density is too low to obtain satisfactory performance. The information density is low and fixed due to without zoom control as shown in Figure 3.22. Examples of false detection result are depicted in Figures 3.19. The low quality can be seen in the snapshots as shown in Figure 3.21(a). Because the color of target is green that is similar to the surrounding objects such as tree and parking floor and the color information of target is insufficient, the tracking results quickly are false and the width of detected box is divergent in the end as shown in Figures 3.21(b) and 3.21(c). In the figures, the lines are designed convergence range and the error bar is width and height of detected box, respectively. The convergence zone is $b_{conv} = [160 - \zeta_X, 160 + \zeta_X] \times [120 - \zeta_Y, 120 + \zeta_Y]$ where $\zeta_X = 13.33$ and $\zeta_Y = 12$.



(a) (b)

Figure 3.19. The snapshots of false detection result without using zoom function.

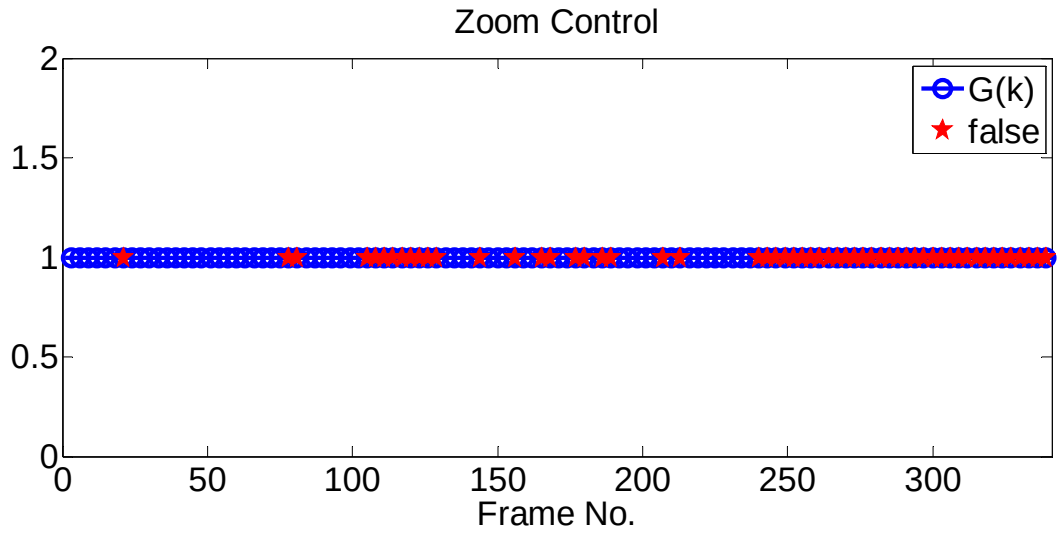
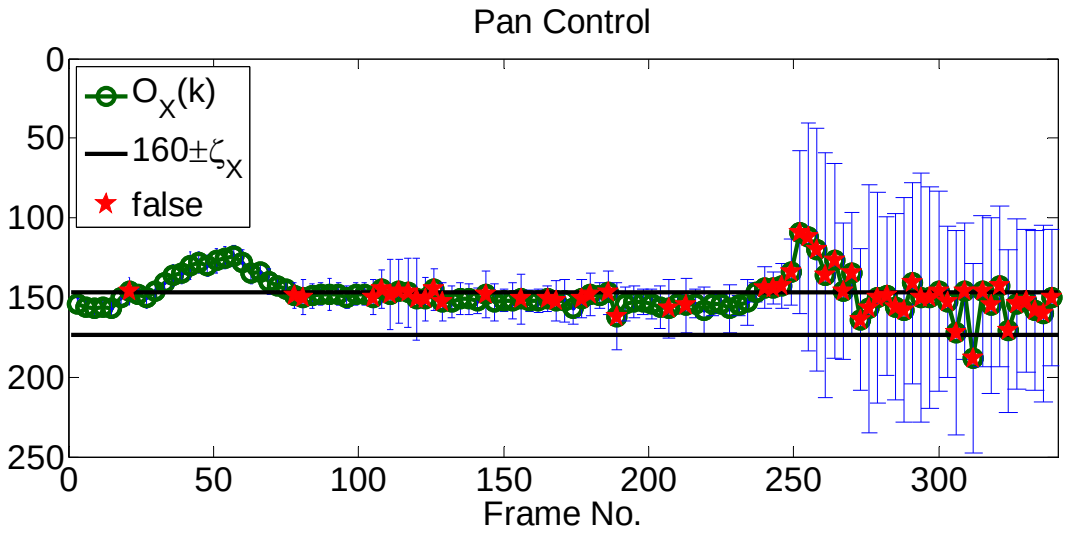
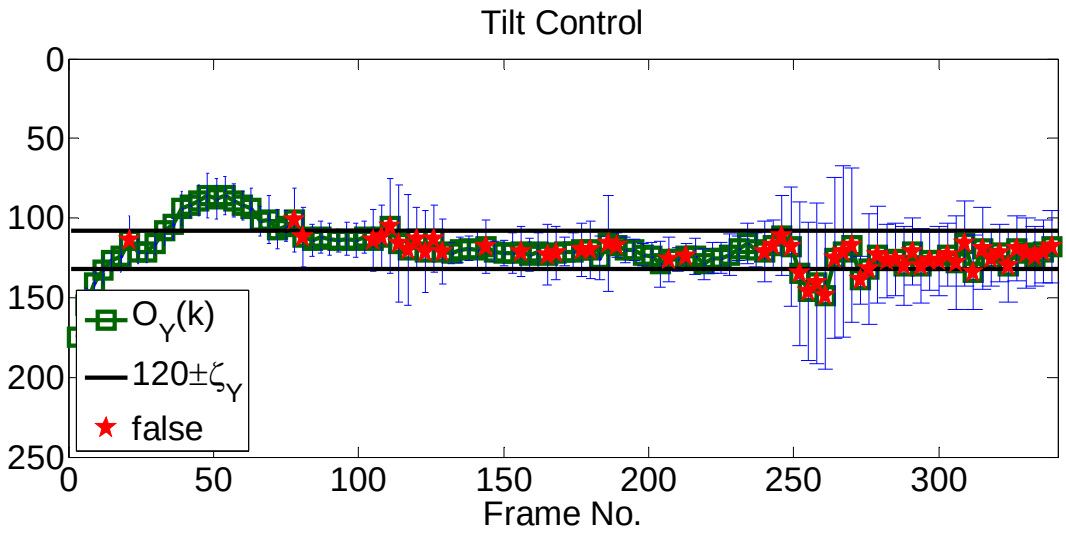


Figure 3.20. The zoom value is constant in traditional method.

(a)



(b)



(c)

Figure 3.21. (a) The snapshots of tracking results (b) and (c) are the X-coordinate and Y-coordinate of center point of detected box, respectively.

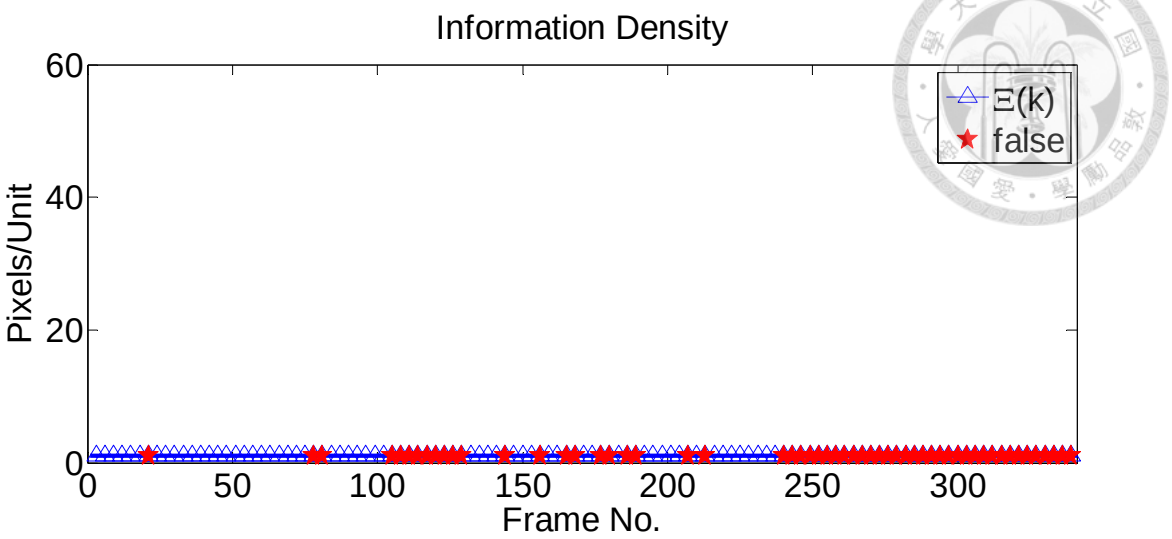
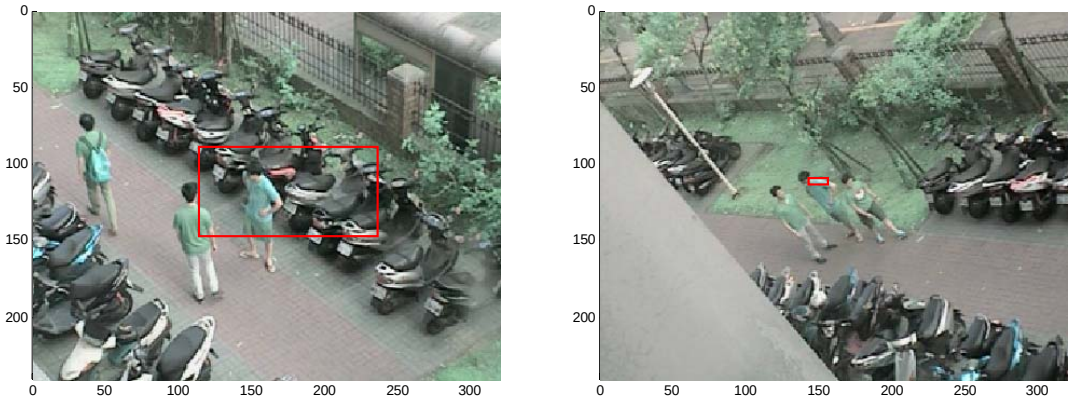


Figure 3.22. The information density is fixed in traditional method.

B. ROI-based controlled sensing

In the same outdoor experimental scene as shown in Figure 3.18, the proposed zoom control policies are implemented to improve the tracking result. The parameters of the zoom control in outdoor experimental scene are $\delta_1 = 0.1$, $\delta_2 = 0.2$, $b_{dpi} = 49$ $b_{Cf} = 15$, $b_{Lf} = 3$ and $\mu^*_{out} = 2.4$. The dynamic convergence zone is the same as Equation (3.25). Figure 3.24 shows the evolution of zoom behavior. For achieving the desired information density, zoom in is operated at first and then zoom out until the condition of $b_{Cf} = 15$ is met. Examples of false detection results are presented in Figure 3.23. Figure 3.25(a) shows the snapshots during the tracking process and the evolution of quality of the target. Because the color information of the target is sufficiently provided as shown in Figure 3.26, the tracking results are significantly

improved and convergence behavior of width and height of detected box is maintained as presented in Figures 3.25(b) and 3.25(c).



(a) (b)

Figure 3.23. The snapshots of false detection result with using zoom function.

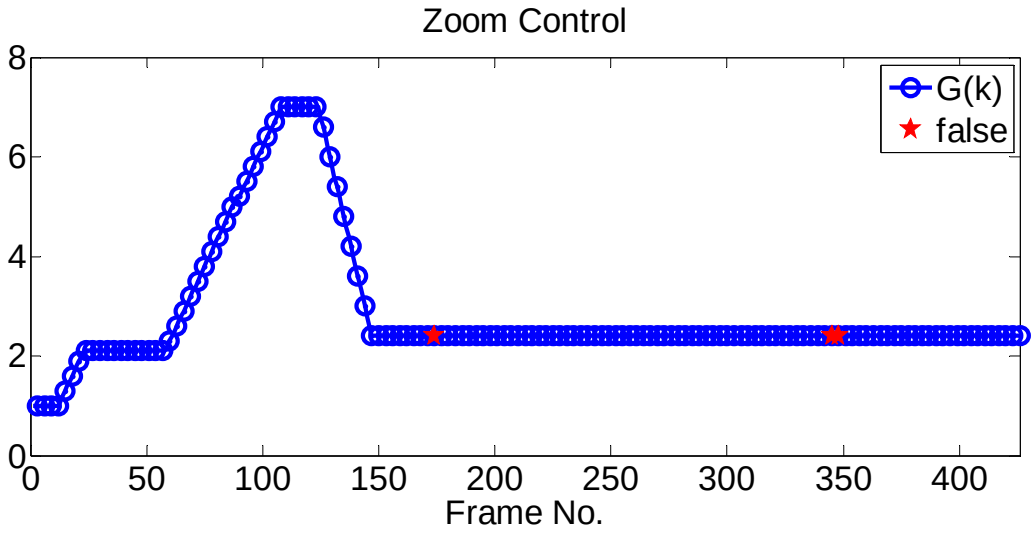
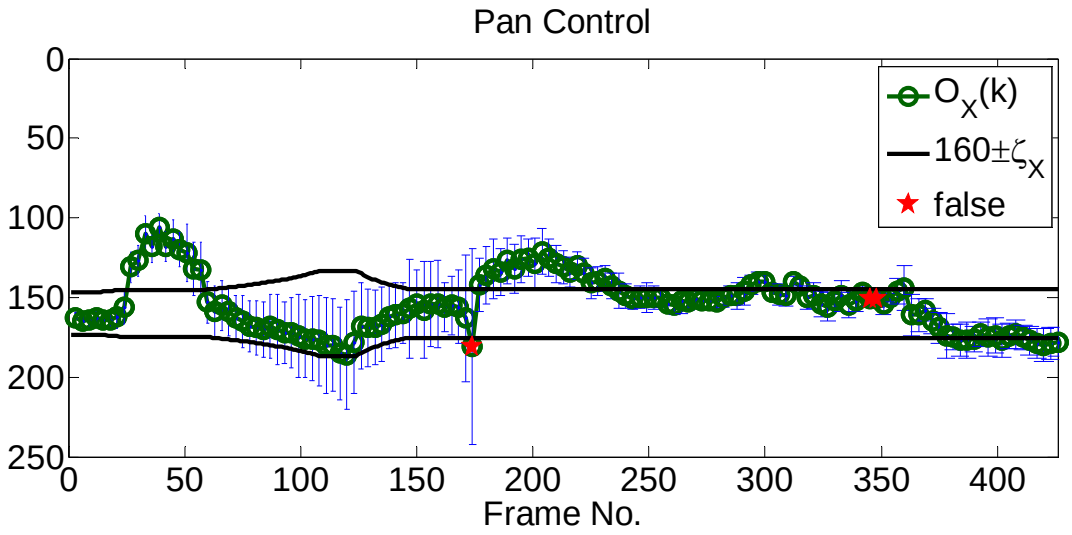
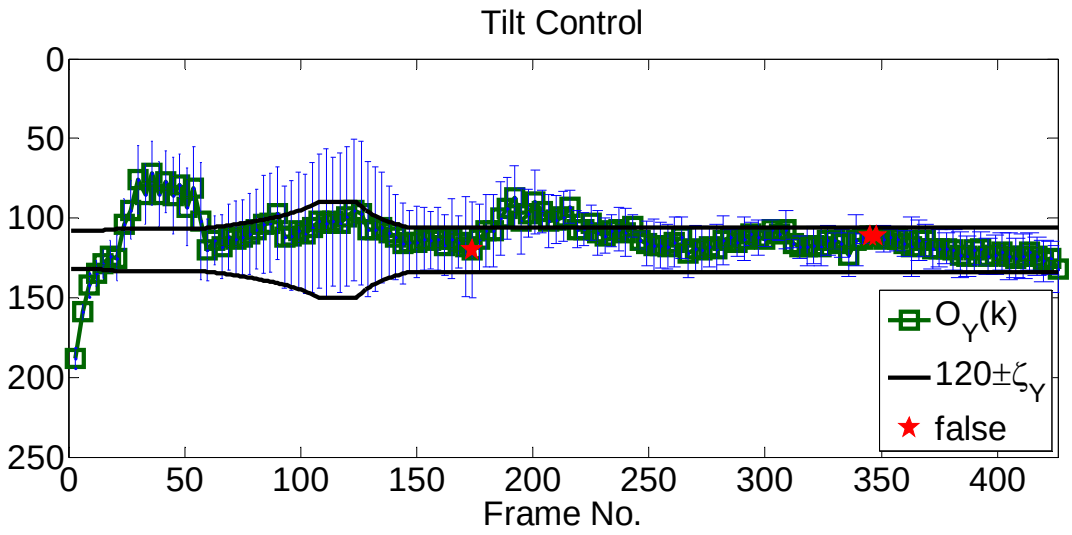


Figure 3.24. The zoom value in outdoor experimental scene with ROI-based controlled sensing method.

(a)



(b)



(c)

Figure 3.25. (a) The snapshots of tracking results (b) and (c) are the X-coordinate and Y-coordinate of center point of detected box, respectively.

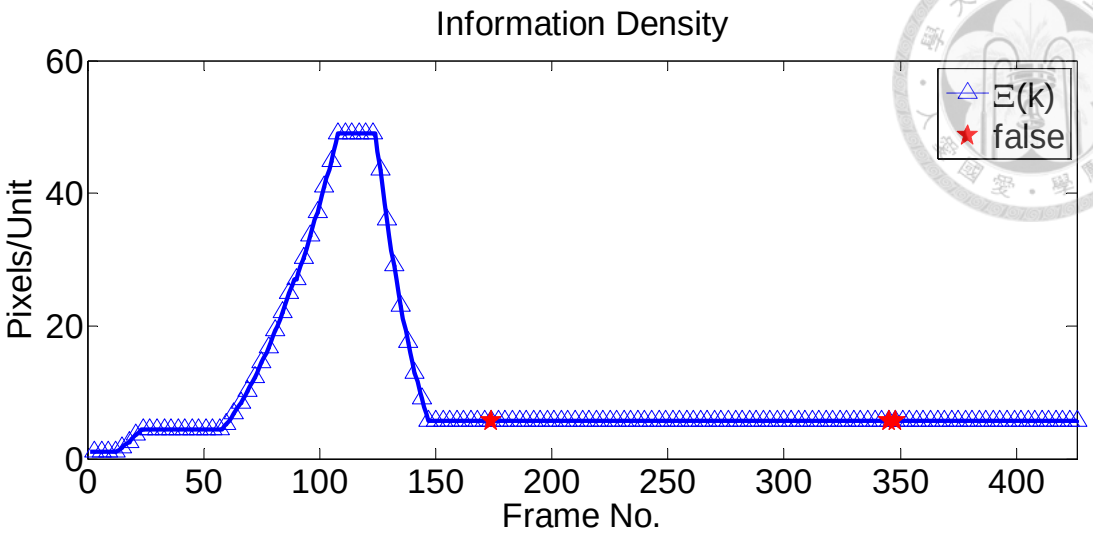


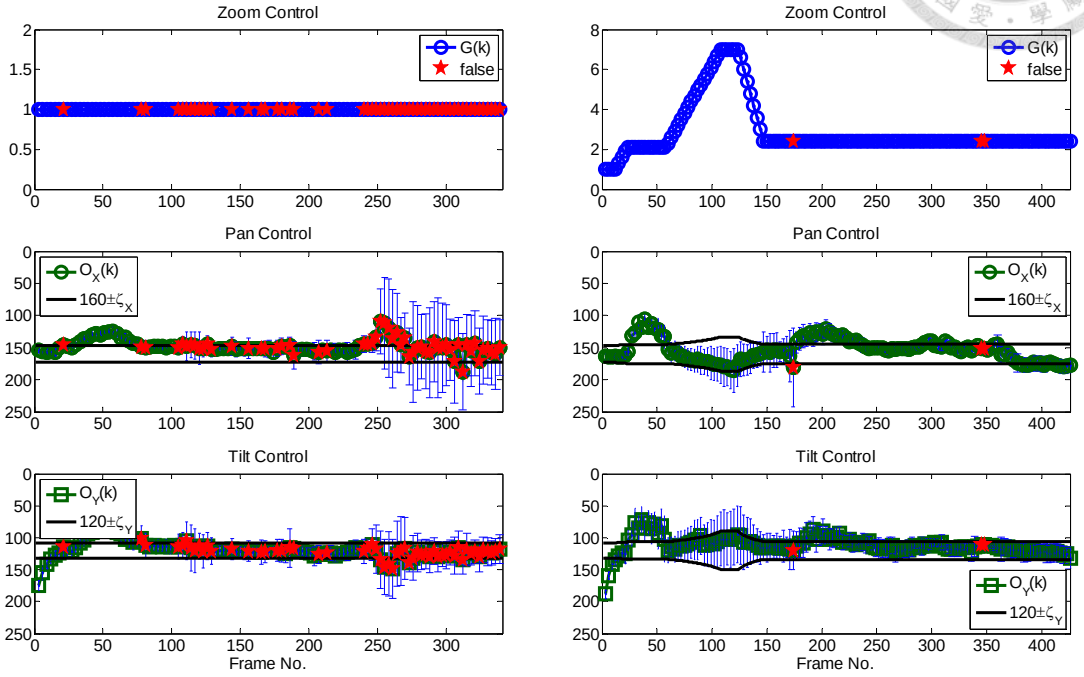
Figure 3.26. The information density in ROI-based controlled sensing method.

C. Comparison

In comparison of traditional approach, the proposed ROI-based controlled sensing method has excellent tracking results based on the adjustable quality of target. On the other hand, flexibility monitoring range enables to taking into account the behavior of target and other objects as shown in Figures 3.27 and 3.29. Figure 3.28 is the part of Figures 3.27 from Frames 100 to 200 to present the tracking results in detail.

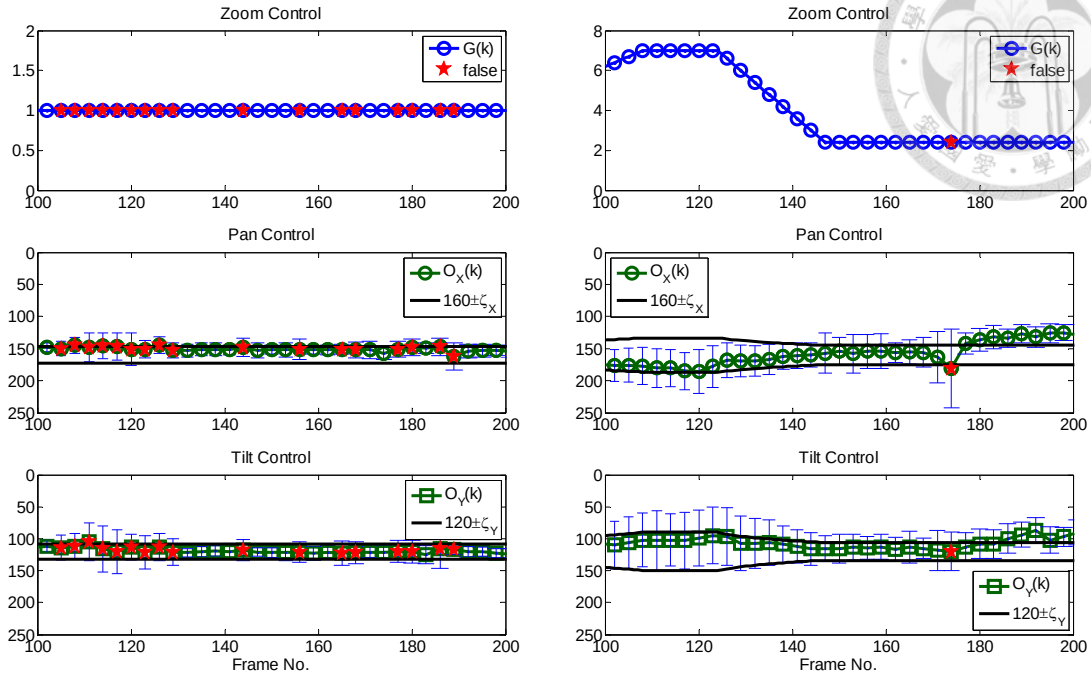
From the experimental results of indoor and outdoor environment, it can be seen that tracking results and quality of the target can be greatly improved by using proposed ROI-based controlled sensing method. Moreover, the information about other objects is not to lose based on the flexibility monitoring capability. Hence, the proposed ROI-based controlled sensing with designed zoom control policies indeed has

outstanding predominance.



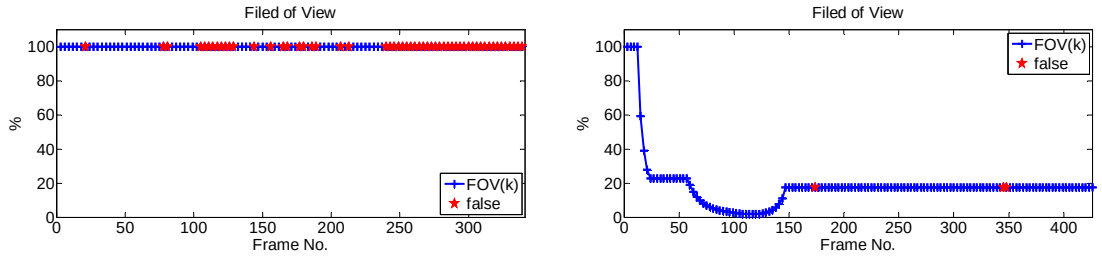
(a) (b)

Figure 3.27. Performance comparison of (a) traditional method and (b) ROI-based controlled sensing method.



(a) (b)

Figure 3.28. The partial figure of Figure 3.27 from Frames 100 to 200.



(a) (b)

Figure 3.29. The percentage area of viewed physical world (a) traditional way (b) ROI-based controlled sensing.

3.4 Summary

Quality control can be modeled as the quantized feedback problem and further be proved the stability with adjustable step size of quantizer. The proposed ROI-based controlled sensing is designed based on the information density of the object of interest by changing zoom functions. The advantage of proposed zoom control policies is that it not only decreases the tracking error but also take into account surrounding objects behavior. The parameters used in the indoor and outdoor experimental scene of traditional and ROI-based are summarized in Table 3.1. In particular, the ζ_X and ζ_Y are related to the zoom variable in the ROI-based controlled sensing method and are referred to Equation (3.25). The tracking results of traditional and ROI-based in two experimental scenes are summarized in the Table 3.2. False detection is greatly decreased by using proposed Region-of-Interest-based controlled sensing and the convergence behavior of detected box size is also maintained. The proposed ROI-based method only has 1.49% and 2.82 false detection ratio those are lower than 22.06% and 48.68% of traditional in indoor and outdoor experimental scene, respectively.

Table 3.1. The parameters are used in traditional and ROI-based method.

	Indoor Experimental Scene		Outdoor Experimental Scene	
	Traditional	ROI-based	Traditional	ROI-based
δ_1		0.1		0.1
δ_2		0.2		0.2
ζ_X	13.33	Equation (3.25)	13.33	Equation (3.25)
ζ_Y	12	Equation (3.25)	12	Equation (3.25)
δ_h		0.2		0.2
b_{Cf}		15		15
b_{Lf}		3		3
b_{dpi}		20		49
μ^*_{out}		1.2		2.4

Table 3.2. The tracking results of traditional and ROI-based method in two different experimental scenes.

	Indoor Experimental Scene		Outdoor Experimental Scene	
	Traditional	ROI-based	Traditional	ROI-based
False Ratio	22.06%	1.49%	48.68%	2.82%
Convergence	Yes	Yes	No	Yes

Chapter 4
Keyframe-Based Quantity Control

For improving the public safety, upgrading the visual experience in entertainment and sports, and assisting teaching in education, the demands for video-related information are rapidly increasing [38: Chen et al. 2010], [4: Foresti et al. 2005]. Visual sensors are deployed in the particular area including manufactory, mall, station, highway, stadium and museum to sense and monitor the possible objects, targets or events. Then their video data are transmitted through the bandwidth-limited network to control station and/or the end-users for further recording, analyzing and/or watching [62: Huu et al. 2010], [63: Misra et al. 2008], [64: Si et al. 2012]. However, if the requested video quantity exceeds the available channel capacity, the packet congestion caused by excessive network traffic might impose severe transmission delay and high ratio of packet loss to degrade the received video quality [65: Huang et al. 2009], [66: Wu et al. 2000]. The solution to the problems usually encounters a tradeoff between quality and transmission quantity of the received video [67: Lian et al. 2006]. That is, removing too much video packet data would affect the quality of the received video or lose some important frames to degrade system performance [68: Andersson et al. 2008], [67: Lian et al. 2006]. Therefore, in the dissertation, a solution is proposed to reduce the quantity

and/or complexity of the sensed video packet data while maintaining the characters of the original data to keep system performance in an acceptable level.

In order to reduce the quantity of video packet data, video abstraction is a common method to be considered and adopted [69: Li et al. 2006]. One of the video abstraction types is video summarization which extracts a number of significant and salient frames or shots [38: Chen et al. 2010], [40: Li et al. 2006]. The extracted frames are so called keyframes that are used to represent the whole video content in a compact manner [38: Chen et al. 2010], [40: Li et al. 2006]. An efficient, intelligent and dynamic keyframes extraction process is proposed to reduce the number of sensed frames based on the perceived motion energy of each frame.

4.1 Keyframe Extraction Based On Perceived Motion Energy (PME)

Because visual experience depends on their neural activities in the brain, different people would have different visual perception for the same video [70: Buche et al. 1998]. In other words, if the keyframes are decided by human, then the number and content of them might be different for the same video. Hence, a general and feasible method is required to develop for determining the keyframes based on the significance and characters of the content. That is, keyframes are decided by the dynamics of the monitored system or targets or the information quantity of the video. In the dissertation, motion feature is used to represent the information quantity contained in one single frame. The two reasons for using motion feature are as follows: (1) in our brain system, it deals with huge information from outside every day. How can it do that? Brain would give different visual attention weight to different objects based on their motion dynamic. Specifically, brain would preferentially process input information including motion character [71: Kanwisher & Wojciulik 2000]. Motion perception is an important faculty of the brain for paying attention to and dealing with sudden events [70: Buche et al. 1998]. For example, driver pays more attention to the moving cars which are approaching than immobile cars. Hence, motion is an important visual input for our brain system [72: Duncan et al. 1997]. (2) In the image processing technique, moving

objects are always the goal to find, and a lot of mature algorithm could be used to compute motion estimation of the objects between two continuous frames in different real scene [74: Irani & Anandan 1999], [73: Liu et al. 2008], [75: Torr & Zisserman 1999]. Therefore, motion is a reasonable index for representing the information quantity in the frame and it also could stand for the dynamics of the monitored system or the objects. Perceived motion energy (PME) is used to model the dynamic of motion information in the video [56: Liu et al. 2003].

Figure 4.1 shows the illustration of computing perceived motion energy. Firstly, the received video consists of a series of frames and can be written as follows

$$\begin{aligned} F &= \{F(k), k = 1, \cdots, N\} \\ &= \{F(1), F(2), F(3), \cdots, F(k), F(k+1), \cdots\} \end{aligned} \tag{4.1}$$

where F and k denote frame set and frame no., respectively. Motion vectors could be obtained by using motion estimation algorithm, such as block matching algorithm [76: Lian et al. 2013] or optical flow [75: Irani & Anandan 1999].. Optical flow consists of velocity filed which is formed by relative movement between the objects and visual sensors or the changing in the light [77: Aires et al. 2008]. Frame filled with the motion vectors $V_{i,j}(k)$ is called motion vector frame as shown in Figure 4.1 and motion vectors are obtained by

$$V_{i,j}(k) = f_{motion}\left(F(k), F(k+1)\right) \tag{4.2}$$

where i and j denote the (i, j) pixel of the frame and $f_{motion} : \mathbb{R}^{W \times H} \times \mathbb{R}^{W \times H} \to \mathbb{R}^{W \times H}$ is the motion estimation function and $f_{motion} = f_{optical\ flow}$ in the dissertation. Next, each motion vector frame $V_{i,j}(k)$ is used to calculate corresponding perceived motion energy value $E(k)$

$$E(k) = \sum_{i=1}^{W} \sum_{j=1}^{H} \left\| V_{i,j}(k) \right\|_2 \tag{4.3}$$

where W and H denote width and height of the frame, respectively.

In [56: Liu et al. 2003], MPEG video format is used for easily obtaining motion estimation from B-frames and perceived motion energy is built on the motion vectors with spatial consistency of angles. However, it might loss the other inconsistency but important motion pattern. In the dissertation, simplifying computation of perceived motion energy could have possibility for online application and flexible for various video format. Moreover, the whole motion energy of the frame is computed for corresponding to entirety sensing function of the brain system.

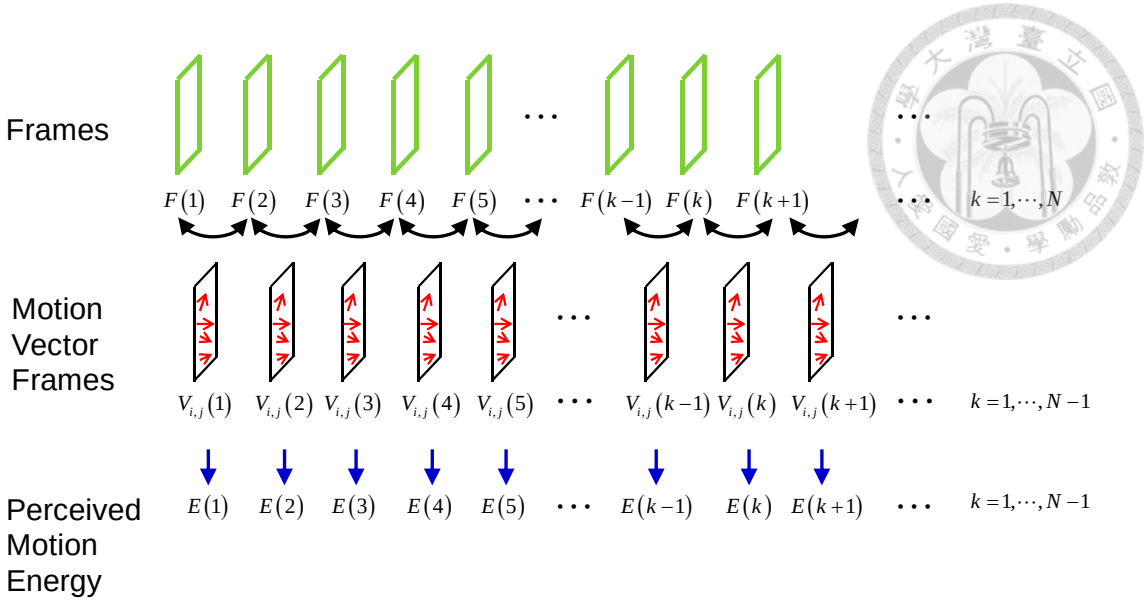


Figure 4.1. The illustration of computing perceived motion energy.

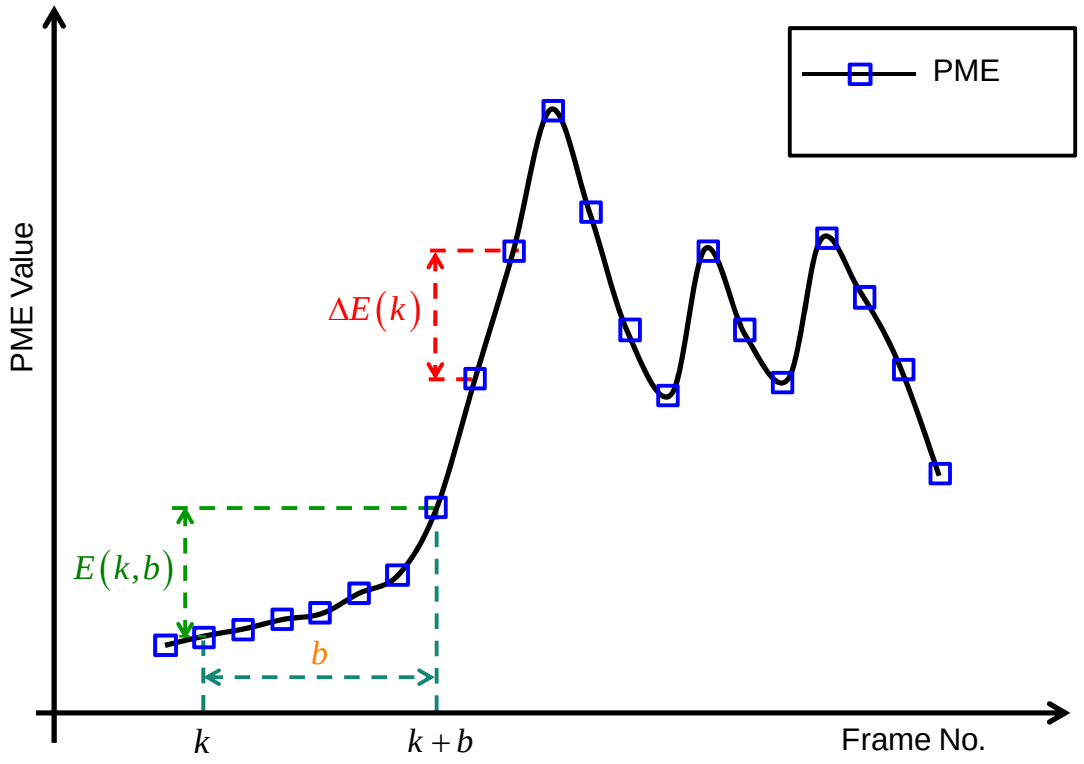


Figure 4.2. The diagram of the perceived motion value of the video.

Figure 4.2 shows the illustration of perceived motion energy of the sensed frame

by Equation (4.3). Based on the observation of motion pattern, there are two phases in it as rapidly increase (decrease) and steadily stay in the certain level. Hence, difference between adjacent perceived motion energy is used to measure the similarity of consecutive frames for rapidly increasing (decreasing) phase.

$$\Delta E(k) = \left| E(k+1) - E(k) \right| \tag{4.4}$$

For steady staying in the certain level phase, the accumulation difference of motion energy from k_{th} to $(k+b)_{th}$ Frame is computed by

$$E(k,b) = \sum_{m=k}^{m=k+b} \Delta E(m) \tag{4.5}$$

When the monitored system dynamic becomes fast, more frames are necessary to be extracted for following the system transferring. On the other hand, when it stays in the certain level, a few frames are required to be extracted for updating the motion information based on accumulation difference of motion energy. Hence, based on Equations (4.4) and (4.5), two extraction rules (Rule 1) and (Rule 2) are designed to determine the frame no. of keyframes for two different phases as follows:

(Rule 1)

$$if\ \Delta E(k) \geq AT\ then\ frame\ no.\ of\ keyframe\ is\ k^* = k+1 \tag{4.6}$$

(Rule 2)

if $\Delta E(k) < AT$

then

(1) *accumulation interval size*

$$b^* = \min\left\{ b \,\middle|\, E(k,b) \geq AT, E(k,b) = \sum_{m=k}^{m=k+b} \Delta E(m) \right\} \tag{4.7}$$

(2) *frame no. of keyframe*

$$k^* = 1 + \arg\max_{l}\left\{ E(l) \,\middle|\, k \leq l \leq k+b^* \right\}$$

In Equation (4.6), if similarity between two neighbor frames over the threshold AT, then the frame no. of keyframe is determined as $k^* = k+1$. For avoiding losing too much motion energy to affect system performance, accumulative perceived motion energy is considered and computed as Equation (4.5) where b is the size of accumulation interval. For adapting to state evolution, b is not constant. Hence, the appropriate b^* is firstly decided with AT in Equation (4.7). Next, the frame no. with maximum perceived motion energy is extracted as keyframe from k_{th} to $\left(k+b^*\right)_{th}$ Frame. An example of keyframe extraction by using Equations (4.6) and (4.7) on two different motion patterns can be seen in Figure 4.3.

In the extraction process, extraction threshold AT plays an important role over the whole keyframe extraction process. For not being affected by the extreme value of perceived motion energy, AT is decided with two parameters λ_{ctrl}^{scene} and inter quartile range(IQR) as follows:

$$AT = \lambda_{ctrl}^{scene} \times IQR$$
$$IQR = Q_3 - Q_1 \quad (4.8)$$

where Q_1 is first quartile (lower quartile), Q_3 is third quartile (upper quartile) and λ_{ctrl}^{scene} is designed to cooperate with tolerance of controller to noise or disturbance of sensing data from different monitored scene. For preventing the AT from greatly influencing by outliers as extreme perceived motion energy, IQR is used as the baseline of threshold AT. Inter quartile range (IQR) is a measure of statistical dispersion [78: Graham & Ian 1996] and equals the difference between the upper and lower quartiles [79: Zwillinger & Kokoska 2000]. Moreover, IQR is the common robust measure of scale which is a robust statistic that quantifies the statistical dispersion in a set of numerical data [80: Mizera & Müller 2004], [81: Rousseeuw & Croux 1993].

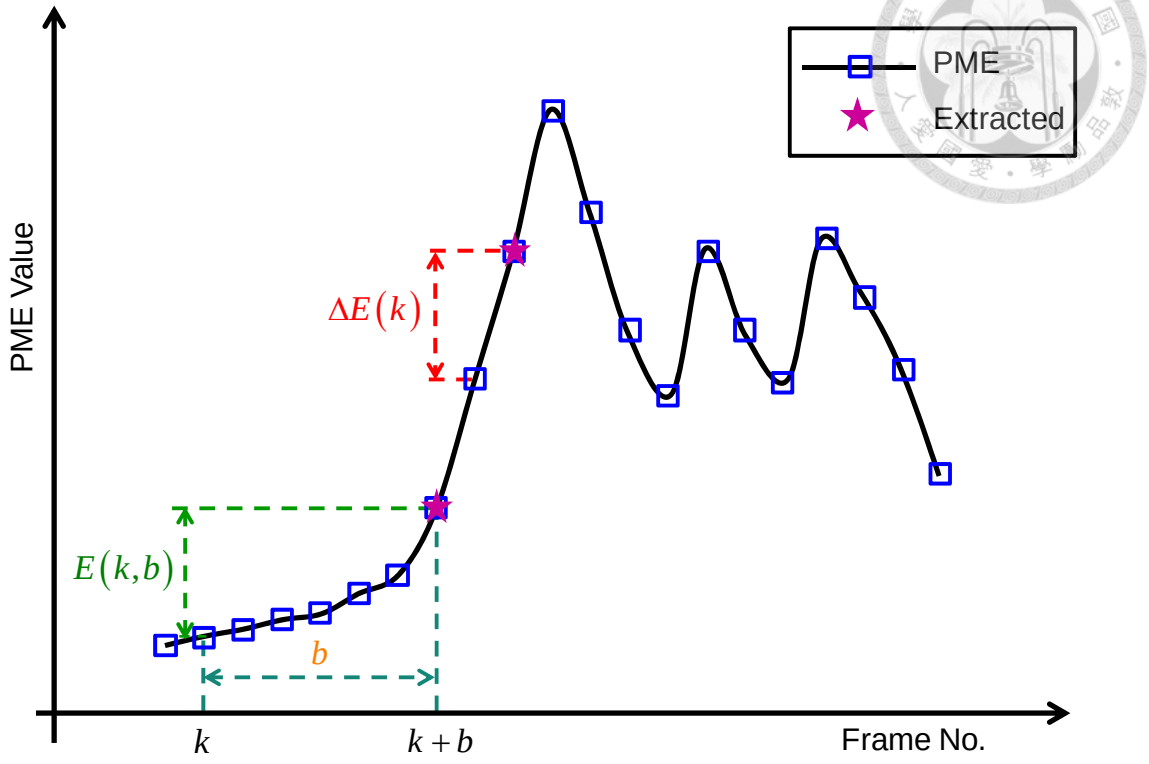


Figure 4.3. The diagram of keyframe extraction.

An example of keyframe extraction result by implementing two extraction rules as Equations (4.6) and (4.7) on the Figure 4.2 is depicted in Figure 4.4. Moreover, the utilization ratio and energy error ratio are computed to stand for the efficiency of the proposed keyframe extraction rules based on perceived motion energy value as follows:

$$Utilization\ Ratio = \frac{\sigma(extracted\ \ frames)}{\sigma(total\ \ frames)} \times 100\% \tag{4.9}$$

where σ is a counting function, namely $\sigma: S \to \mathbb{Z}^+$, S is a real set. For instant, if $A=\{a_1, a_2, \cdots, a_n\}$, $a_i \neq a_j$ then $\sigma(A)=n$.

$$Energy\ Error\ Ratio = \frac{\tilde{E}_{total}}{E_{total}} \times 100\%$$

$$\tilde{E}_{total} = E_{total} - \hat{E}_{total}$$

$$E_{total} = \sum_{j_1 \in \Omega} E(j_1), \Omega = \{1, 2, \cdots, k, k+1, \cdots\}$$

$$\hat{E}_{total} = \sum_{j_2 \in \Theta} E(j_2), \Theta = \left\{ \underbrace{1 \cdots 1}_{k_1^* - 1}, \underbrace{k_1^* \cdots k_1^*}_{k_2^* - k_1^*}, \underbrace{k_2^* \cdots k_2^*}_{k_3^* - k_2^*}, \cdots, \underbrace{k_m^* \cdots k_m^*}_{k_{m+1}^* - k_m^*}, \underbrace{k_{m+1}^* \cdots k_{m+1}^*}_{k_{m+2}^* - k_{m+1}^*}, \cdots \right\}$$

(4.10)

where Ω is the captured frame set and Θ is the extracted keyframe set.

Hence, for example, the utilization ratio is $^{13}\!/_{21} \times 100\% = 62\%$ in Figure 4.4. Figure 4.5 shows the approximation motion energy curve shaped from the perceived motion energy value of extracted keyframes and the original motion energy curve produced by the perceived motion energy value of each captured frame.

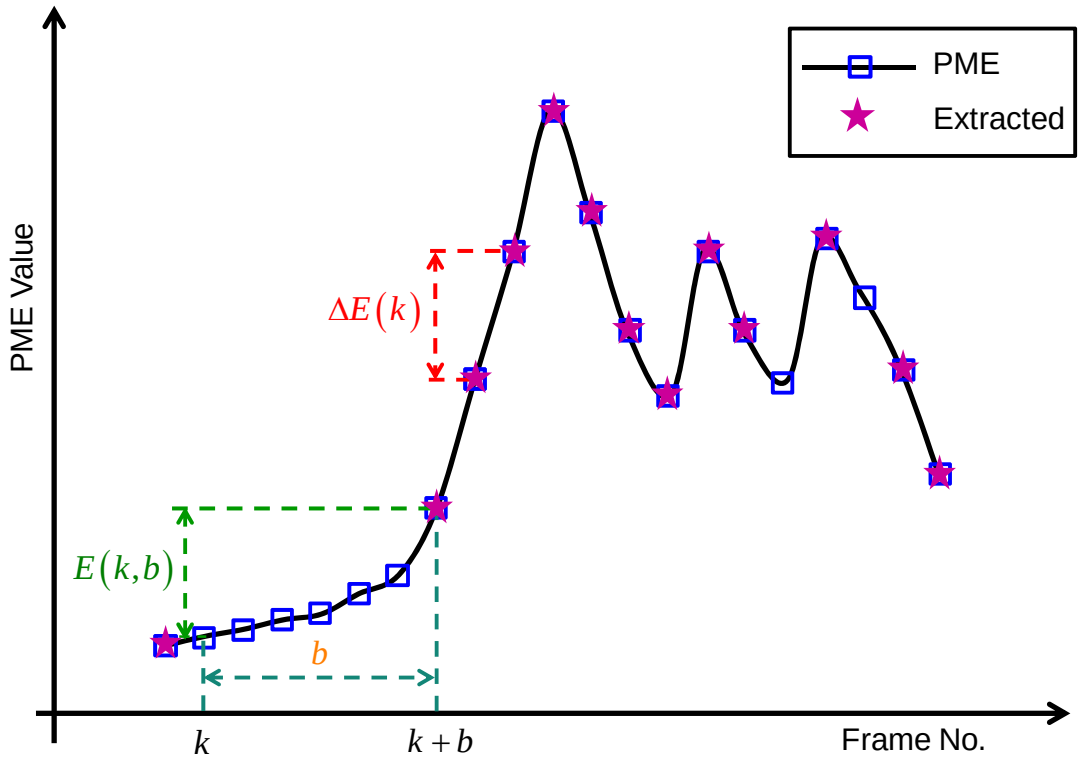


Figure 4.4. The diagram of extracted keyframes result.

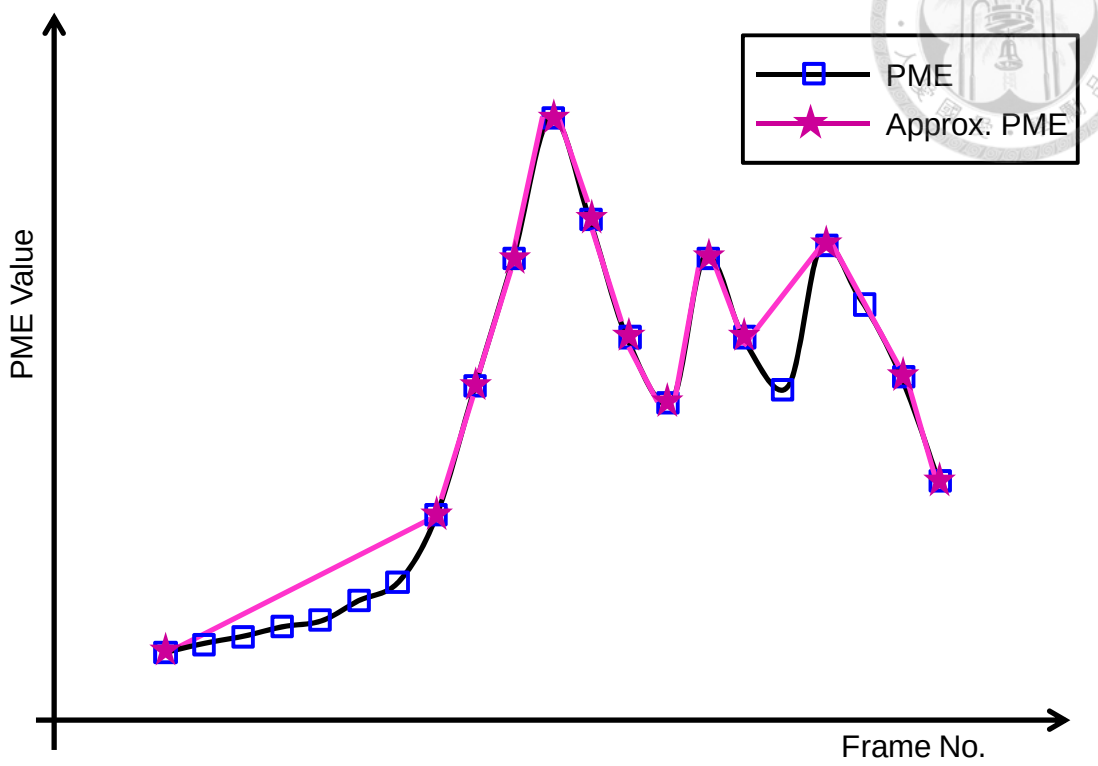


Figure 4.5. The diagram of approximation function shaped from the extracted keyframes.

After controlled sensing module that is presented in Chapter 3, the perceived motion energy (PME) is calculated to represent the motion information quantity contained in the captured frame. Then, two designed extraction rules (Rule 1) and (Rule 2) with threshold $AT = \lambda_{ctrl}^{scene} \times IQR$ are performed to extract the keyframes to represent the whole video. In general, $\lambda_{ctrl}^{scene} = 1$ is sufficient to extract enough number of keyframes to keep system performance unless the robustness of controller is low. Figure 4.6 shows the framework of PME-based keyframe extraction process that consists of PME value computation and proposed keyframe extraction rules based on

motion energy for rapidly increasing (decreasing) phase and steady staying in the certain level phase. This extraction process is automatic and be able to extract the frames with significant motion energy as keyframes. In order to preserve the most information, such as moving objects to be successfully transmitted to controllers or end-users under condition of rate limitation, the progressive decoding process is implemented [82: Lian et al. 2012]. Layer-based transmission module would divide image content into three regions with different priority. Region with high priority gets to be decoded firstly and then is region with middle priority, last is region with low priority [82: Lian et al. 2012]

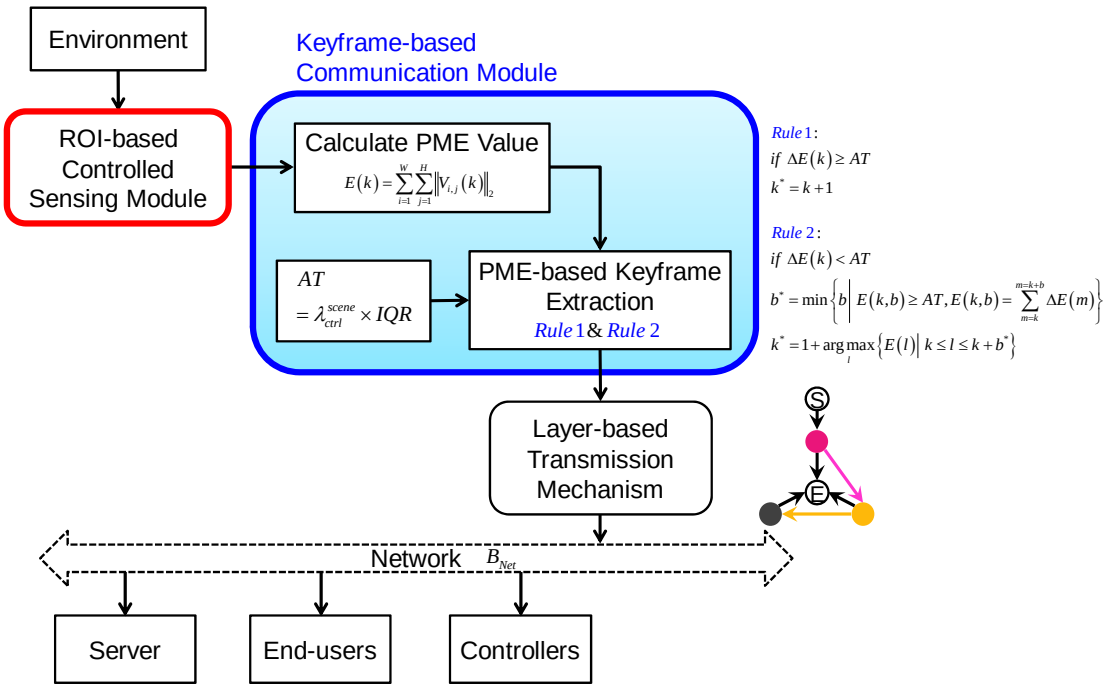


Figure 4.6. The diagram of PME-based keyframe extraction process.

4.2 Sampling strategy

Before presenting experimental results, three different sampling strategies which produce keyframes are firstly described in the section. First, the proposed perceived motion energy based (PME-based) keyframe extraction algorithm with threshold $AT = \lambda_{ctrl}^{scene} \times IQR$ can sample dynamically according to the motion energy induced by the motion of video content and/or visual sensor. Second, the fixed sampling strategy $Fixed(\alpha,\beta)$ is based on prior given sampling periods where α and β are sampling period for moving and stationary objects, respectively. Specifically, when moving people in the scene are detected, the sampling period would be α, otherwise the sample period would be assigned as β. Moreover, the extracted keyframe would be the one having most edge number to represent possessing most information within α or β continuous frames. Third, all data are sampled that is traditional manner. For example, keyframe extraction results by using the three different sampling strategies are depicted in Figure 4.7. In the figure, captured frames are all keyframes in traditional manner, due to captured frame without detecting any moving objects, keyframes are sampled within 3 continuous frames in fixed sampling strategy $Fixed(\alpha,\beta)$ with α=1 and β=3, and only one frame is sampled based on motion energy of the captured frames in the proposed PME-based keyframe extraction algorithm with designed extraction rules as Equations (4.6)-(4.7).

In order to demonstrate the performance of the proposed PME-based, $Fixed(\alpha,\beta)$ and traditional manner are executed to compare. Figures 4.8-4.10 show the experimental results of the three different sampling strategies, respectively. In traditional as shown in Figure 4.8, each frame is extracted as keyframe and total extracted number is 15. In $Fixed(\alpha,\beta)$ as shown in Figure 4.9 with α=1 and $\beta=3$. Then the frame no. of extracted keyframes are 1, 4, 7, 10, and 13, and total extracted number is five. In the proposed PME-based as shown in Figure 4.10 with $\lambda_{ctrl}^{scene}=1$, only one frame is extracted as keyframe, i.e. Frame 1. In all fifteen frames, no moving people appear but camera is moving to cause the movement of background. Furthermore, video content in the frames appears alike for human eyes. Hence, transmitting all frames would consume bandwidth of network or computation.

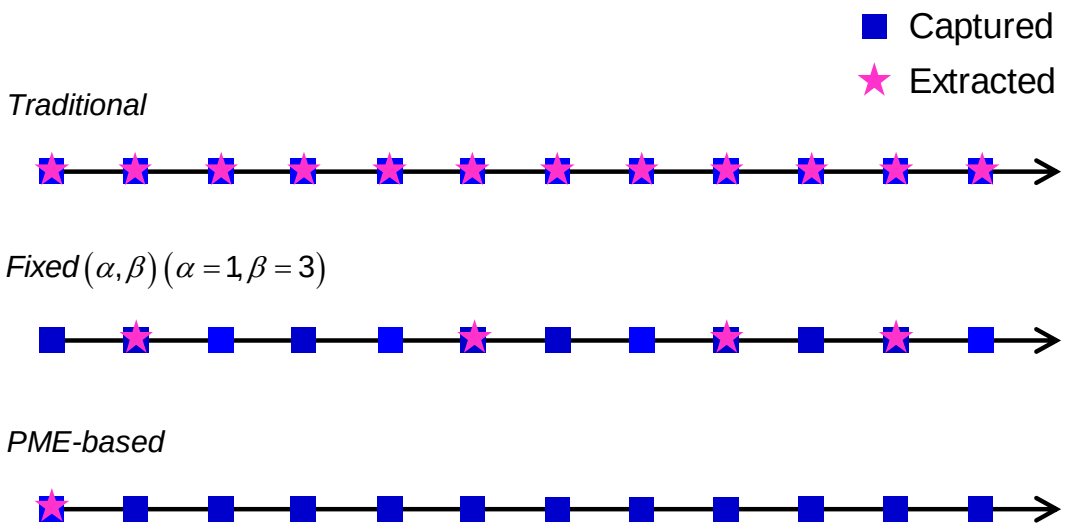


Figure 4.7. Example of three different sampling strategies, traditional, $Fixed(\alpha,\beta)$ with fixed sampling period α=1 and $\beta=3$, and PME-based with $\lambda_{ctrl}^{scene}=1$.

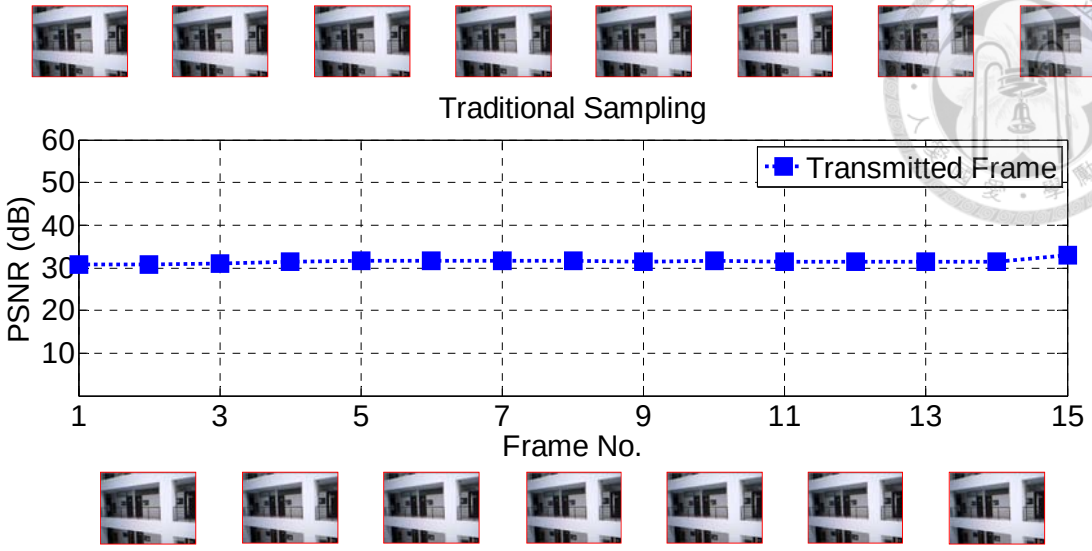


Figure 4.8. Experimental results of Indoor3people (from Frame 1 to Frame 15) with using traditional sampling strategy.

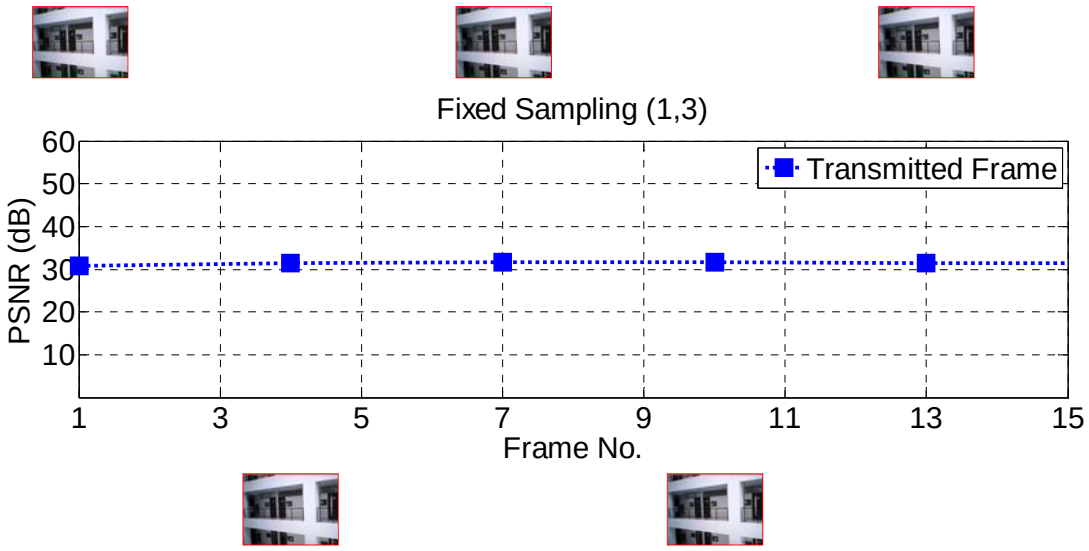


Figure 4.9. Experimental results of Indoor3people (from Frame 1 to Frame 15) with using fixed sampling strategy $Fixed(\alpha,\beta)$ (α=1,β=3).

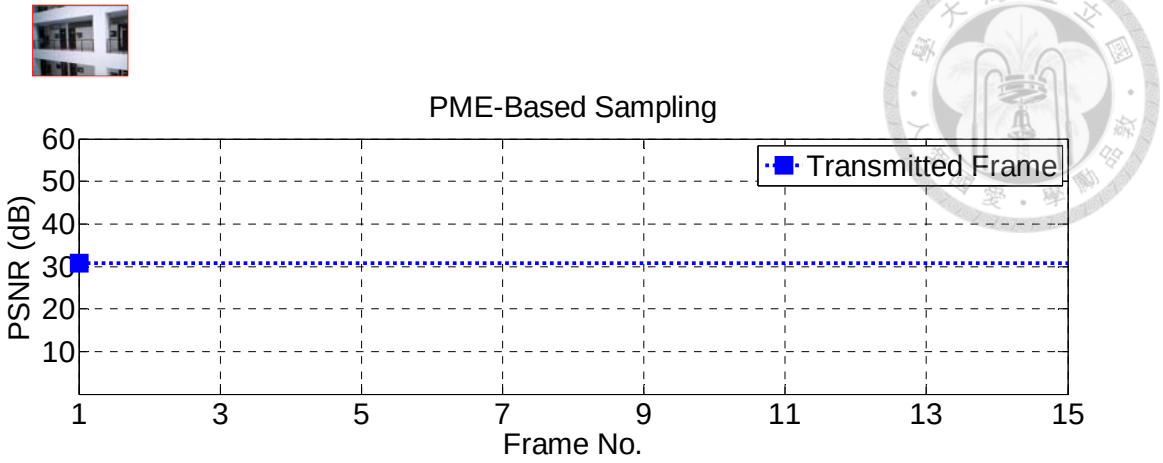


Figure 4.10. Experimental results of indoor3people (from Frame 1 to Frame 15) with using PME-based sampling strategy ($\lambda_{ctrl}^{scene} = 1$).

4.3 Experimental Results of PME-based Keyframe Extraction and Analysis

4.3.1 Description

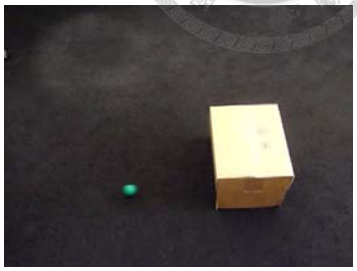
In order to present the flexibility and efficiency of proposed PME-based keyframe extraction algorithm with threshold $AT = \lambda_{ctrl}^{scene} \times IQR$, ten videos are used in different application tasks. The experimental results shows object tracking and detection results are still maintained in an acceptable level or even as well as the origin one after video data reduction based on proposed motion energy extraction rules. Ten videos are categorized into two parts based on motion character: Part I: visual sensor is mobile and Part II: visual sensor is stationary. In Part I, perceived motion energy is produced by the relative motion of visual sensor and objects and in Part II, it is caused only by absolute motion of objects. In order to cover the possibility of motion energy induced way in the real world, the object moving direction is divided into single and multiple in Part I and Part II. The video name, snapshot and frames per second (fps) of the ten videos in Part I and Part II are shown in Tables 4.1-4.2, respectively. Six of them are typically tested videos, namely, Forman, Hall, VisionTraffic, Container, Coastguard and SingleBall. Other two indoor experimental videos are captured by the mobile camera that is moving as horizontal or ego-rotation way on the 5F and 6F of the Ming-Da building at NTU with walking, stationary and running peoples. Floors3people is captured by the

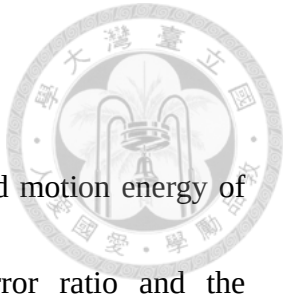
stationary pantoscopic camera on the 6F of the MD building at NTU to emulate the real surveillance scene. Odometry video is captured by the webcam mounted in wheeled robot that performs a circle trajectory.

Table 4.1. Part I: Visual sensor is mobile.

<table>
<tr><td>Object motion direction</td><td colspan="3">Snapshot</td></tr>
<tr><td rowspan="2">Single</td><td colspan="2">  </td><td>  </td></tr>
<tr><td colspan="2">Foreman (fps: 25)</td><td>Coastguard (fps: 25)</td></tr>
<tr><td rowspan="2">Multiple</td><td>  </td><td>  </td><td>  </td></tr>
<tr><td>Indoor3people (fps: 15)</td><td>Indoor1people (fps: 15)</td><td>Visual Odometry (fps: 15)</td></tr>
</table>

Table 4.2. Part II: Visual sensor is stationary.

Object motion direction	Snapshot		
Single			
	Container (fps: 25)	SingleBall (fps: 30)	
Multiple			
	Hall (fps: 25)	Floors3people (fps: 23)	VisionTraffic (fps: 30)



4.3.2 Experimental Results

In this section, the motion energy curve produced by perceived motion energy of each frame, the number of extracted keyframe, the energy error ratio and the experimental results for different application of the ten tested videos are shown. Moreover, the efficiency of the proposed PME-based keyframe extraction algorithm with threshold $AT = \lambda_{ctrl}^{scene} \times IQR$ is demonstrated by comparing to the experimental results with using $F(\alpha, \beta)$ and traditional way.

4.3.2.1 Part I: Visual sensor is mobile

A. Object motion direction: Single

In this category, two tested videos, Foreman and Coastguard, are used to demonstrate the experimental results of the proposed keyframe extraction algorithm in Part I (a). For each video, the mentioned Equations (4.3), (4.6)-(4.7) and (4.9)-(4.10) in Chapter 4.1 are performed to obtain the motion energy value by Equation (4.3), the frame no. of extracted keyframe by Equations (4.6)-(4.7), the utilization ratio by Equation (4.9) and the energy error ratio by Equation (4.10).

$$E(k) = \sum_{i=1}^{W} \sum_{j=1}^{H} \left\| V_{i,j}(k) \right\|_2 \qquad (4.3)$$

(Rule 1)

$$if\ \Delta E(k) \geq AT\ then\ frame\ no.\ of\ keyframe\ is\ k^* = k+1 \tag{4.6}$$

(Rule 2)

$$if\ \Delta E(k) < AT$$

$$then$$

$$(1)\ accumulation\ interval\ size$$

$$b^* = \min\left\{ b \,\middle|\, E(k,b) \geq AT, E(k,b) = \sum_{m=k}^{m=k+b} \Delta E(m) \right\} \tag{4.7}$$

$$(2)\ frame\ no.\ of\ keyframe$$

$$k^* = 1 + \arg\max_{l} \left\{ E(l) \,\middle|\, k \leq l \leq k+b^* \right\}$$

$$Utilization\ Ratio = \frac{\sigma(extracted\ \ frames)}{\sigma(total\ \ frames)} \times 100\% \tag{4.9}$$

where $\sigma : S \to \mathbb{Z}^+ ; \sigma(A) = n, where\ A = \{a_1, a_{2,} \cdots, a_n\}$

$$Energy\ Error\ Ratio = \frac{\tilde{E}_{total}}{E_{total}} \times 100\%$$

$$\tilde{E}_{total} = E_{total} - \hat{E}_{total}$$

$$E_{total} = \sum_{j_1 \in \Omega} E(j_1), \Omega = \{1, 2, \cdots, k, k+1, \cdots\} \tag{4.10}$$

$$\hat{E}_{total} = \sum_{j_2 \in \Theta} E(j_2), \Theta = \left\{ \underbrace{1 \cdots 1}_{k_1^* - 1}, \underbrace{k_1^* \cdots k_1^*}_{k_2^* - k_1^*}, \underbrace{k_2^* \cdots k_2^*}_{k_3^* - k_2^*}, \cdots, \underbrace{k_m^* \cdots k_m^*}_{k_{m+1}^* - k_m^*}, \underbrace{k_{m+1}^* \cdots k_{m+1}^*}_{k_{m+2}^* - k_{m+1}^*}, \cdots \right\}$$

where Ω is the captured frame set and Θ is the extracted keyframe set.

The first is Foreman video that has one man who is talking to camera, sometimes swinging his head, raising his hand and in the end of video camera is moved to right side to shoot the construction site. Figure 4.11 shows the keyframe extraction flow of

Foreman. Optical flow is utilized to find the motion vectors frame $V_{i,j}(k+1)$ with $F(k+1)$ and $F(k+2)$ and then perceived motion energy is computed by using Equation (4.3). Then Equations (4.6) and (4.7) adding $E(k)$ are executed to obtain the frame no. of extracted keyframe. The important extraction threshold $AT = \lambda_{ctrl}^{scene} \times IQR$ of Foreman is equal to 1.462 with $\lambda_{ctrl}^{scene} = \lambda_{ctrl}^{Foreman} = 1$ and $IQR = 1.462$. Following the process in Figure 4.11 extracts the keyframes with significant motion energy.

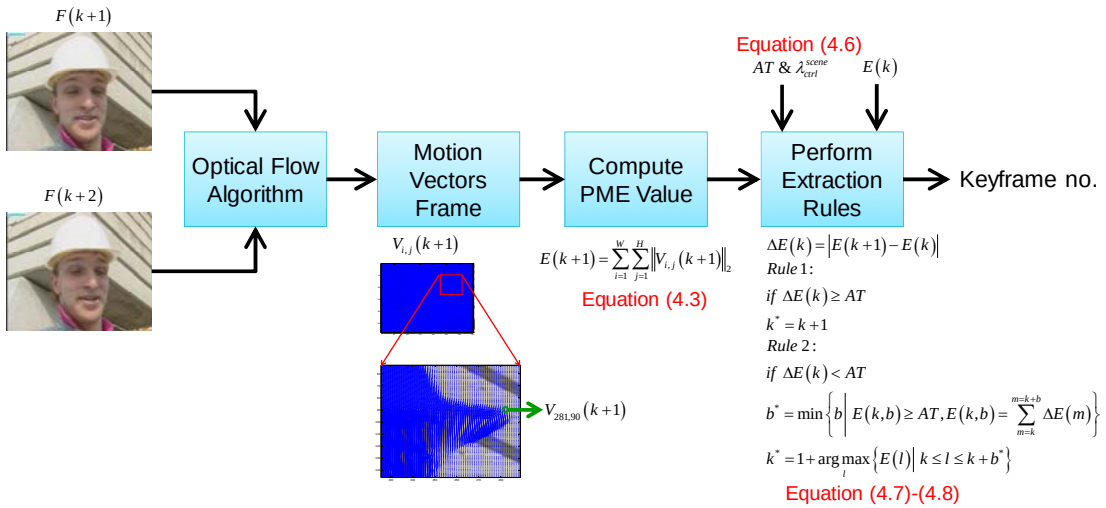


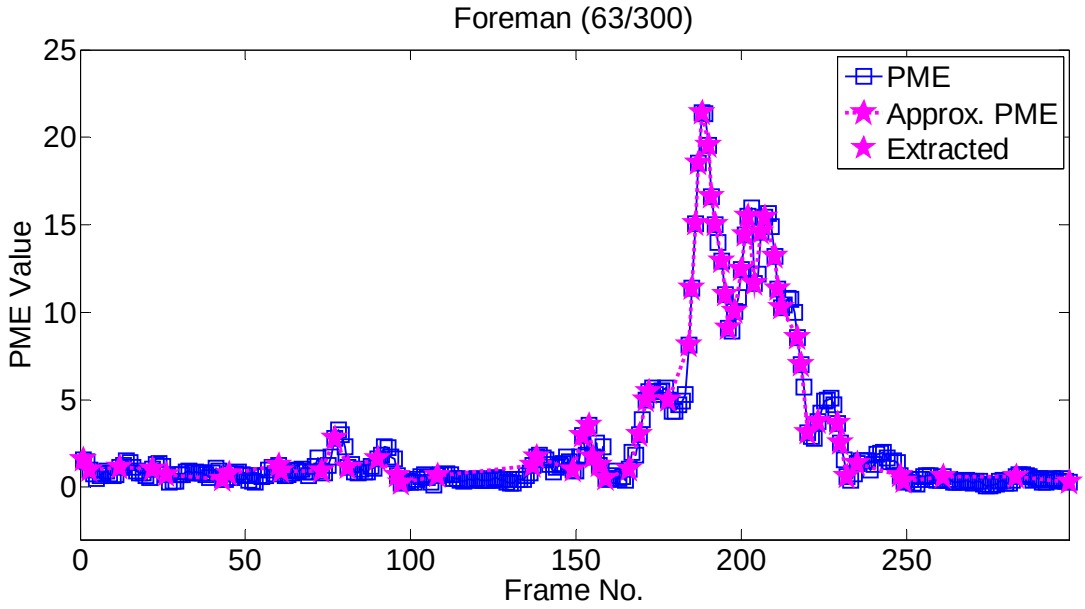
Figure 4.11. The extraction flow of Foreman with perceived motion energy. $AT = \lambda_{ctrl}^{scene} \times IQR$ where $\lambda_{ctrl}^{scene} = \lambda_{ctrl}^{Forman} = 1$ and $IQR = 1.462$.

Figure 4.12 shows five snapshots of Foreman and Figure 4.13(a) presents the motion energy curve produced by perceived motion energy value of captured frames, the approximation motion energy curve shaped from perceived motion energy value of extracted keyframes and the frame no. of extracted keyframe. Before Frame 150, because the man keeps talk and move head and hand, motion similarity is high in

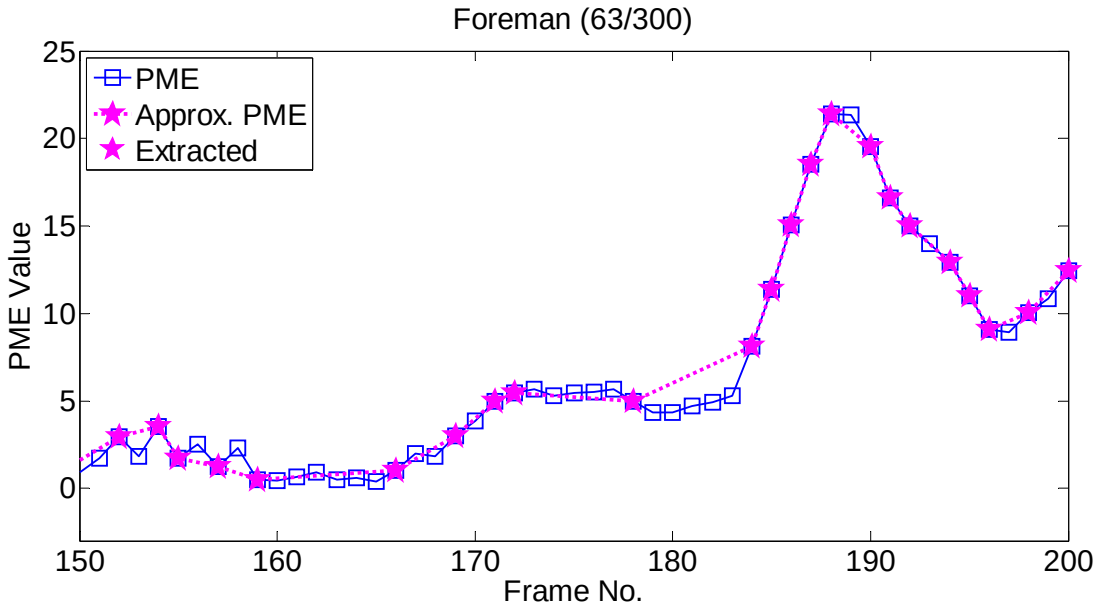
consecutive frames. Hence, perceived motion energy value stays between 0 and 2. From Frame 160 to Frame 220, perceived motion energy value suddenly has a great variation due to visual sensor is fast moved to right to capture the construction site to cause large relative motion between visual sensor and workman. After Frame 220, perceived motion energy value is around zero because camera motion tends toward still and no moving objects in the scene. Figure 4.13(b) shows the motion curve from Frames 150 to 200 to present that the approximation motion curve is close to original motion curve by only using 63 of 300 frames. After finishing the keyframes extraction process, the energy error of each frame is computed by Equation (4.10) as shown in Figure 4.14.

Figure 4.15 shows the frame rate results of traditional and proposed keyframe extraction algorithm. From sixth second to tenth second, which is from Frames 150 to 250, motion energy in video increases rapidly as shown in Figure 4.13(a), frame rate should become fast in order to adapt to the changing in video content. Unlike the traditional method, fixed frame rate (25) is used to generate video data without considering system dynamic. Using the proposed PME-based sampling strategy, it could dynamically adjust the frame rate to follow the system dynamics to guarantee the performance and meanwhile to remove data with high similarity to decrease the consumption of computation resource.

Figure 4.12. The snapshots of Foreman.



(a)



(b)

Figure 4.13. (a) The motion energy curve, the approximation motion energy curve and the frame no. of extracted keyframes (b) The motion energy curve produced by captured frames and extracted keyframes from Frames 150 to 200.

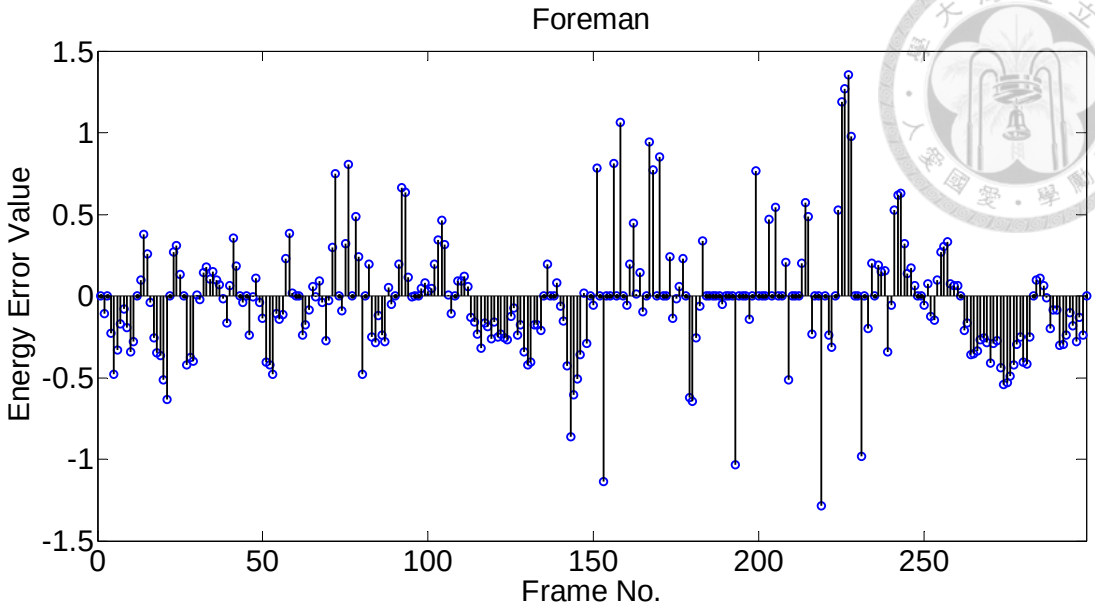


Figure 4.14. The energy error of each frame.

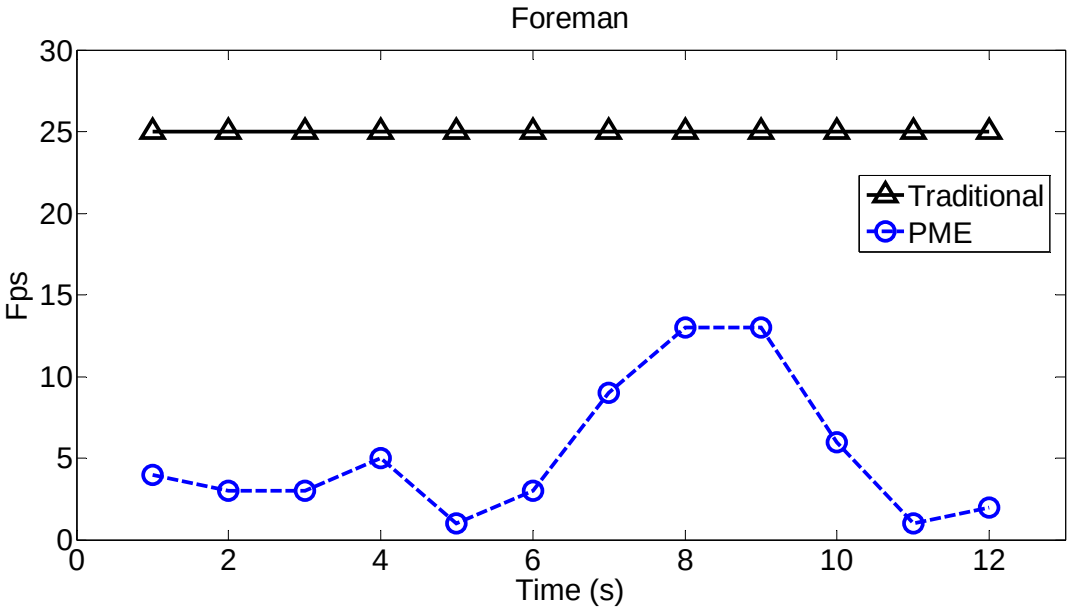


Figure 4.15. Frames per second results of traditional and proposed PME-based keyframe extraction algorithm.

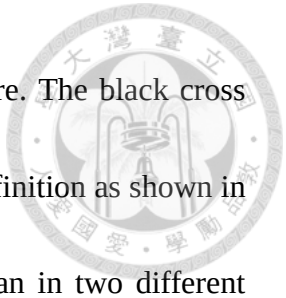
However, removing too many frames could affect the application or system performance because too much motion information of the whole video is lost. In order

to guarantee the performance stays in an acceptable range by using the proposed PME-based keyframe extraction algorithm as dynamical sampling strategy. The face detection results of Foreman are demonstrated with two different sampling strategies: traditional way and PME-based. Figure 4.16 shows the face tracking results of Foreman. Figure 4.16(a) presents the good tracking result and Figure 4.16(b) shows the bad results. The bounding boxes in Figure 4.16(b) are too small to cover the entire face range or too large including other objects.

(a) (b)

Figure 4.16. Performance index of face tracking results of Foreman. (a) Correct result (b) False detection results with small or large box.

For presenting the face tracking results, using (X,Y) denotes the coordinate of left/top of bounding box to represent the location of the detected box as shown in Figure 4.17. Based on the definition of face tracking performance as shown in Figure 4.16, the face tracking results of PME-based and traditional are shown in Figure 4.18. Traditional way uses all captured frames (300 frames) to perform the face tracking, and PME-based



only uses 63 keyframes to perform the same face tracking procedure. The black cross and diamond are false markers in Figure 4.18 based on the results definition as shown in Figure 4.16. Figure 4.18 shows the face tracking results of Foreman in two different sampling strategies. Because the face tracking algorithm requires a few frames for initializing, the face tracking results of traditional and PME-based manners are false in the beginning. However, the trend of X and Y of PME-based is more consistency than traditional. It is not reasonable for the Y coordinate lower than 50 due to the face locates in the center of frame before Frame 150. After Frame 150, the face moves to left/top corner due to camera motion. The false results of PME-based occur due to the size of the detected box. Figure 4.19 shows the face tracking results from Frame 98 to Frame 189 including the camera motion. The results show that the proposed PME-based algorithm has better performance than traditional. Figure 4.20 shows the face tracking path in the Foreman video by using traditional and PME-based sampling strategies. It also shows the path result of using PME-based is close to real face motion trajectory in comparison with using traditional manner.

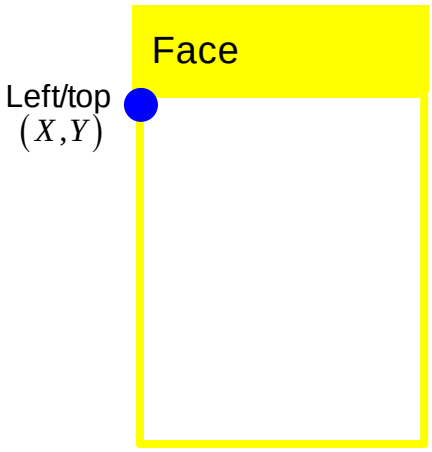


Figure 4.17. The illustration of the location of left/top of face bounding box.

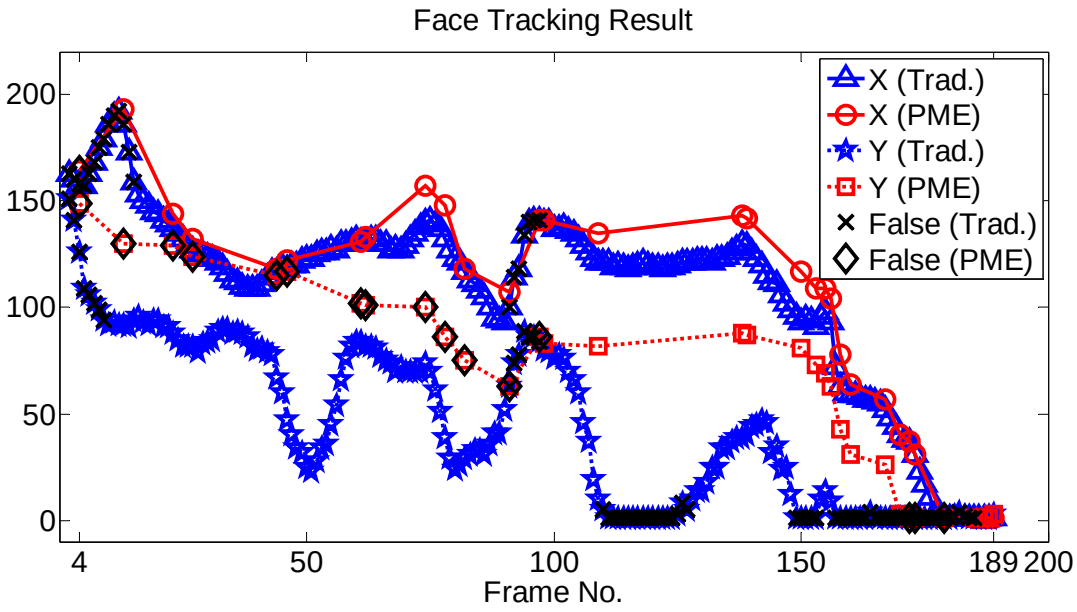


Figure 4.18. Face detection results of Foreman video in two different sampling strategies: traditional and PME-based.

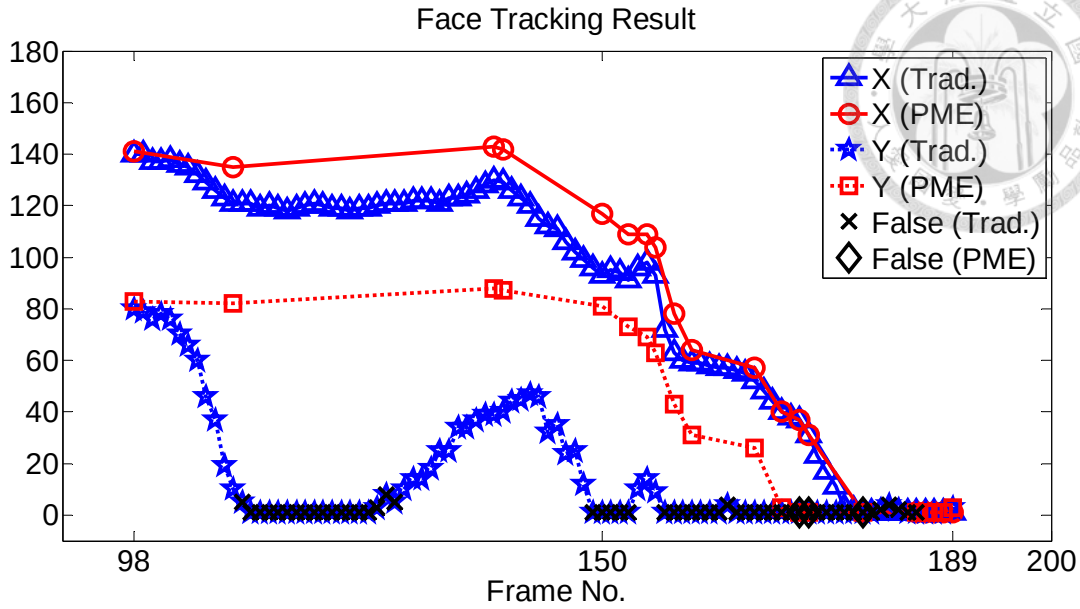


Figure 4.19. Face detection results of the Foreman video in two different sampling strategies: traditional and PME-based from Frames 98 to 189 (partial of Figure 4.18).

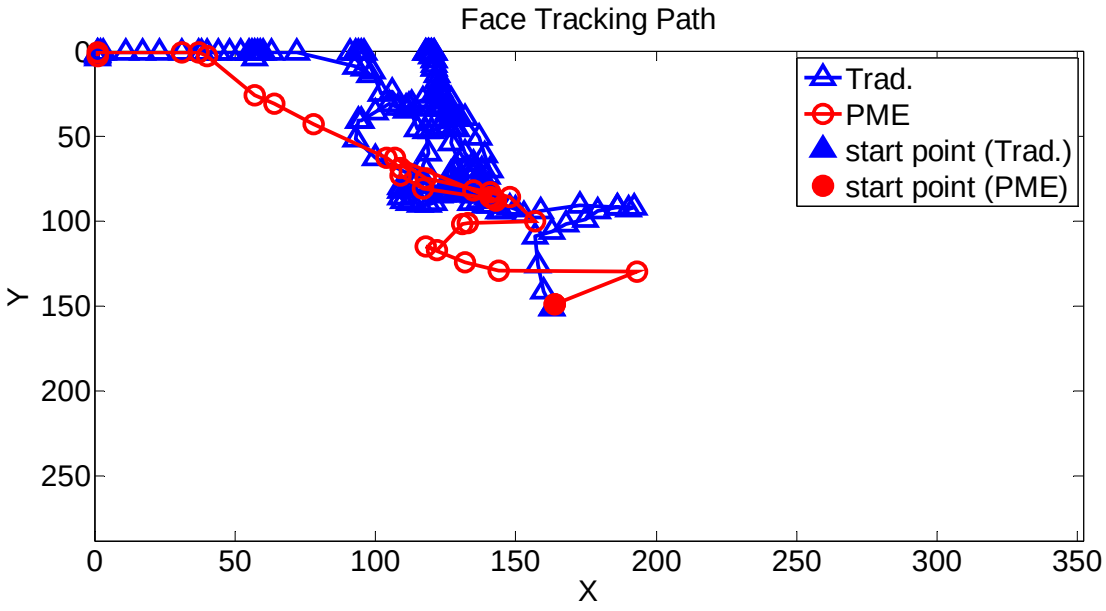


Figure 4.20. Face tracking path in the Foreman video with using two different sampling strategies.

Table 4.3. Results of face tracking, utilization ratio, energy error ratio and the number of the extracted keyframe of the Foreman video.

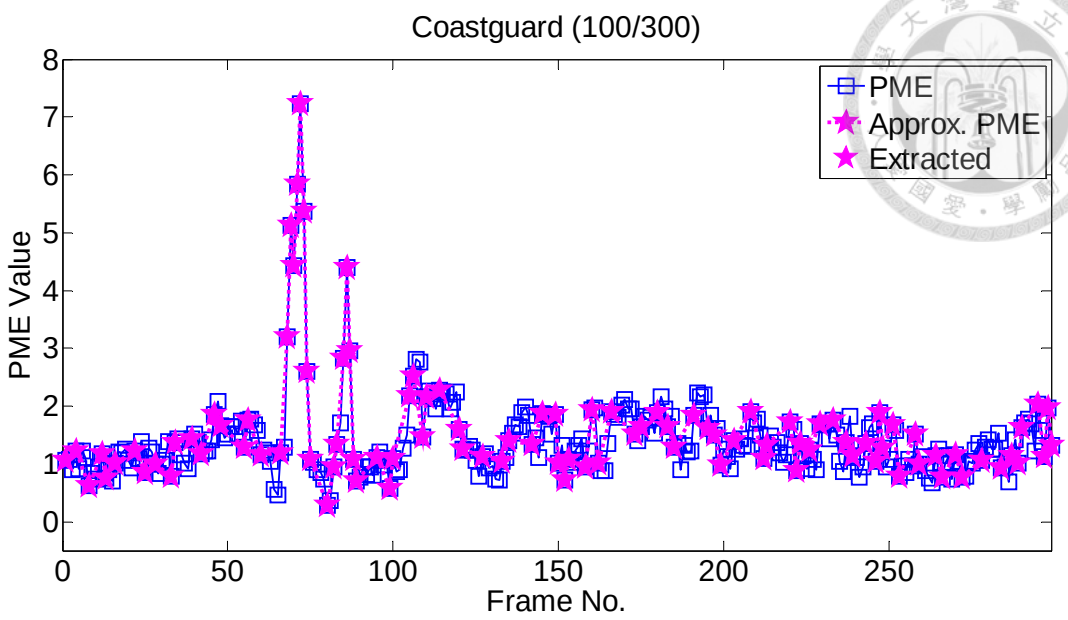
Accuracy Percentage of Face Tracking				
Traditional	PME-based	Utilization Ratio	Energy Error Ratio	#Extracted Frames/ #Total Frames
61.17%	79.6%	21%	-0.83%	63/300

Based on the performance definition as shown in Figure 4.16 and the tracking results as shown in Figure 4.18, Table 4.3 shows that the face tracking accuracy ratio for the Foreman video of traditional and PME-based $\left(\lambda_{ctrl}^{scene}=\lambda_{ctrl}^{Foreman}=1\right)$ is 61.17% and 79.6%, respectively. Moreover, PME-based sampling strategy only utilizes 21% video data to achieve 79.6% accuracy ratio and absolute value of energy error ratio is less 1%.

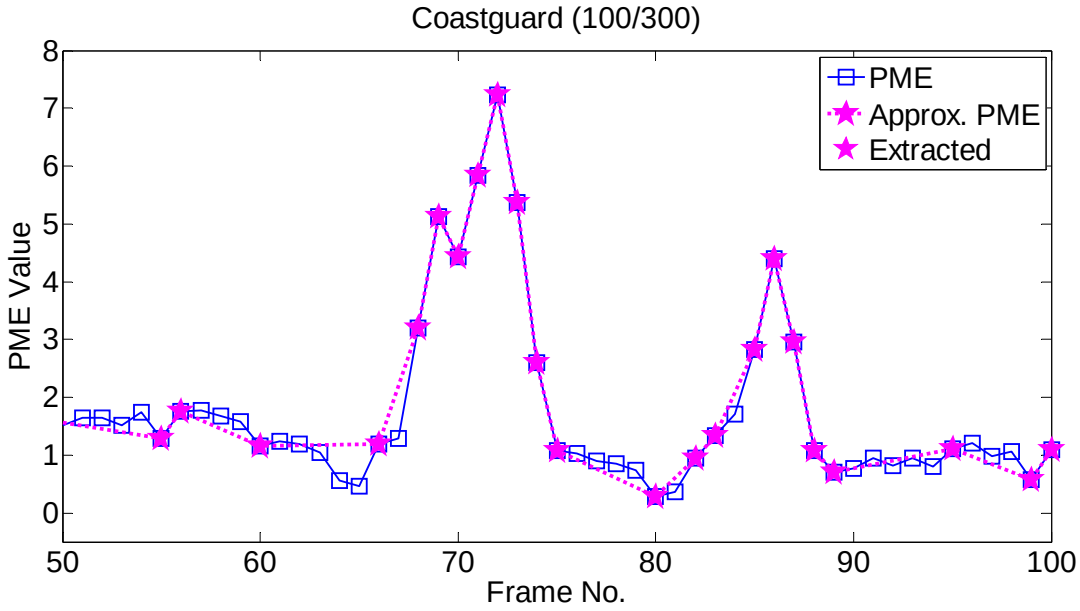
The second tested video is Coastguard that has one boat and one mosquito boat moving toward different direction and the camera is moved to follow the boat and keep the boat in the center of image. The keyframe extraction process of Coastguard is same as Foreman as shown in Figure 4.11. The parameters $\lambda_{ctrl}^{scene}=\lambda_{ctrl}^{Coastguard}=1$ and $IQR=0.657$ are used in the Coastguard video. Figure 4.21 shows five snapshots of the Coastguard and Figure 4.22(a) presents the motion curve produced by perceived motion energy value of each captured frame and the extracted keyframes and the frame no. of extracted keyframe. From Frames 50 to 100, the boat and the mosquito boat are passing through each other and the camera is suddenly moved toward up in order to follow the

boat. Thus the motion curve has the peak. In the rest the frames, because the camera is moved to maintain the boat in the center of frame, the motion between two continuous frames is only from the background, such as tree, the perceived motion energy value is around one. Moreover, the current perceived motion energy value could be considered as the camera motion energy. Next, the energy error of each frame is presented in Figure 4.23.

Figure 4.21. The snapshots of Coastguard.



(a)



(b)

Figure 4.22. (a) The motion energy curve, the approximation motion energy curve and the frame no. of extracted keyframes (b) The motion energy curve produced by captured frames and extracted keyframes from Frames 50 to 100.

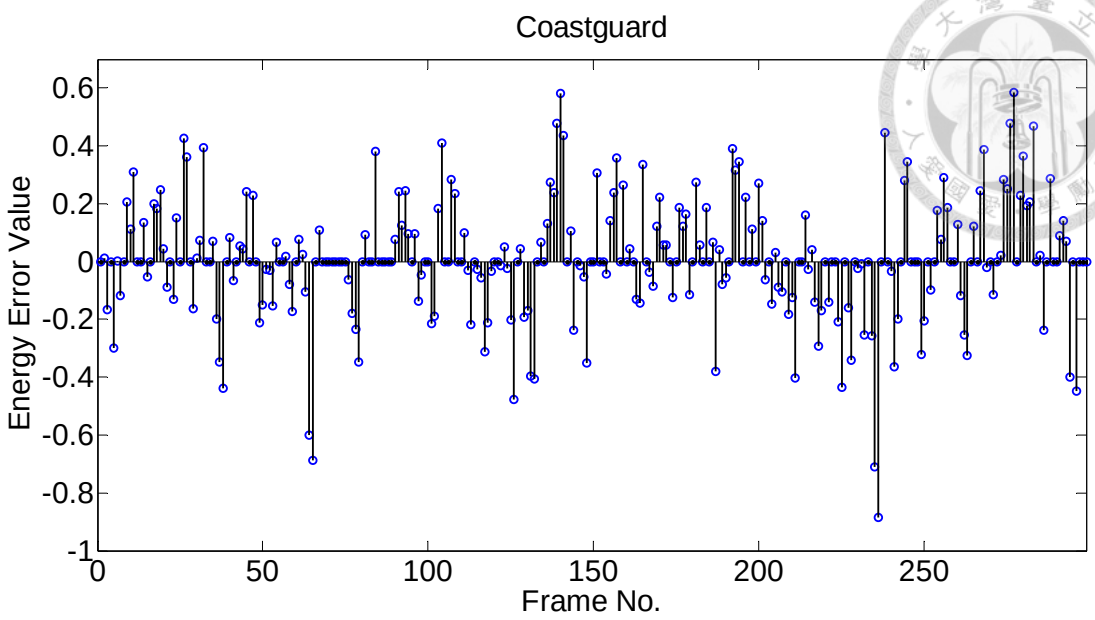


Figure 4.23. The energy error of each frame.

Figure 4.24 shows the frames per second (fps) results of traditional and proposed keyframe extraction algorithm of the Coastguard. From next second to fourth second (from Frames 50 to 100) corresponds to peak of motion curve as shown in Figure 4.22(a), thus the frame rate increases to 10. In the rest of time, frame rate stays between 5 and 10 due to the motion energy comes from background. After PME-based sampling process, a large number of frames are removed to save computation resource further.

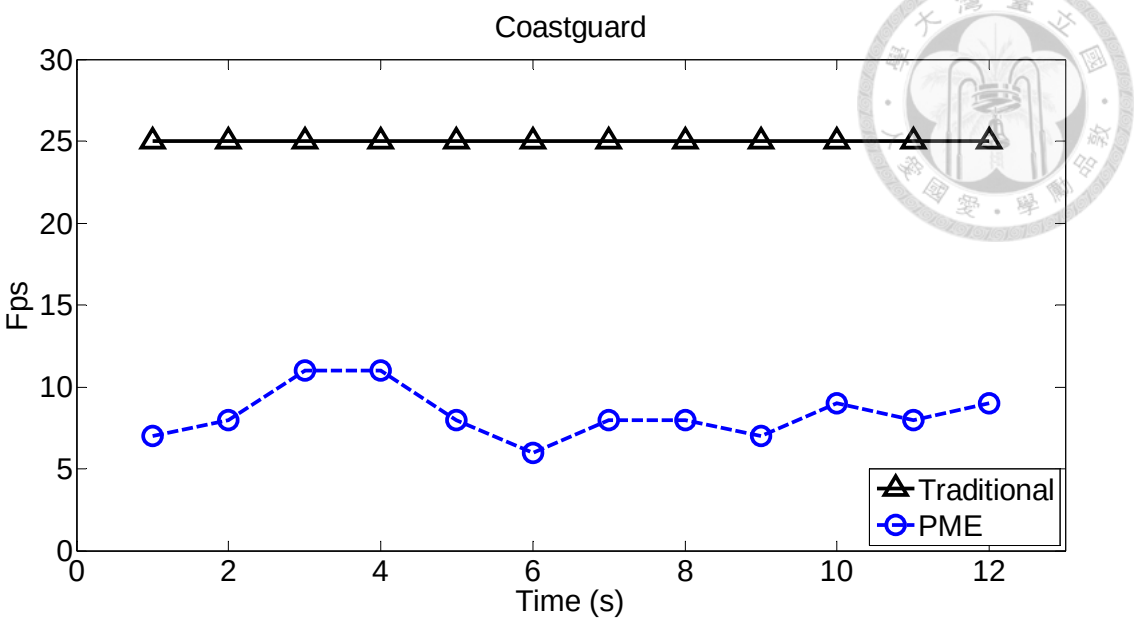
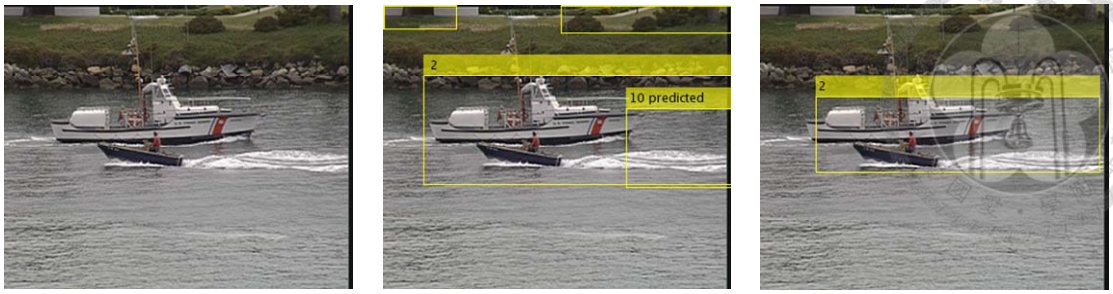


Figure 4.24. Frames per second results of traditional and proposed PME-based keyframe extraction algorithm.

Figure 4.25 presents the multiple objects tracking results of Frames 67, 100 and 143 by using two different sampling strategies, namely, traditional and PME-based. Figures 4.25(a)(d)(g) show the original snapshots of Frames 67, 100 and 143.Figures 4.25(b)(e)(h) and Figures 4.25(c)(f)(i) show the tracking results of traditional and PME-based, respectively. In Frame 143, because the camera follows the boat, small motion energy is contributed from the boat. Moreover, wave is kind of the moving object in this scenario. The moving objects tracking results show that PME-based sampling strategy has better performance than traditional as shown in Figure4.25.



(a) Captured (Frame 67) (b) Traditional (Frame 67) (c) PME-based (Frame 67)

(d) Captured (Frame 100) (e) Traditional (Frame 100) (f) PME-based (Frame 100)

(g) Captured Frame 143 (h) Traditional (Frame 143) (i) PME-based (Frame 143)

Figure 4.25. Moving objects tracking results of Costguard (a)(d)(e) Captured Frames 67, 100 and 143 (b)(e)(h) Tracking results of Frames 67, 100 and 143 with Traditional way (c)(f)(i) Tracking results of Frames 67, 100 and 143 with PME-based keyframe extraction algorithm..

Accuracy ratio of multi-object tracking results of traditional and PME-based sampling manner are both 100%. Table 4.4 shows that PME-based strategy only samples 100 keyframes from 300 frames to achieve the high accuracy ratio. Moreover, the utilization ratio is 33.33% and absolute value of energy error ratio is only 0.44%.

Table 4.4. Results of multi-objects tracking, utilization ratio, energy error ratio and number of extracted keyframe of Coastguard video.

Accuracy Percentage of Multi-Objects Tracking				
Traditional	PME-based	Utilization Ratio	Energy Error Ratio	#Extracted Frames/ #Total Frames
100%	100%	33.33%	0.44%	100/300

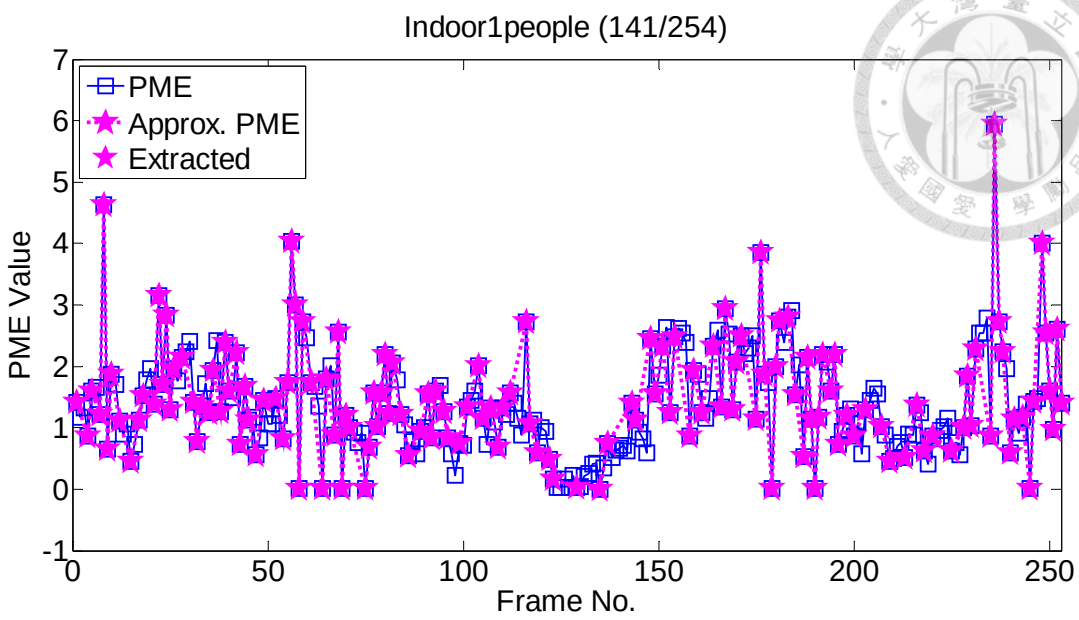
B. Object motion direction: Multiple

Three tested videos, namely, Iindoor1people, Indoor3people, and Visual Odometry, are used to perform the keyframe extraction process as shown in Figure 4.11. The application task for Indoor1people and Indoor3people is video transmission and three different sampling strategies are compared, that is, traditional, $Fixed(\alpha,\beta)$ where α=1, $\beta=3$ and PME-based with threshold $AT=\lambda_{ctrl}^{scene}\times IQR$ where $\lambda_{ctrl}^{scene}=\lambda_{ctrl}^{Indoor1}=0.5$, $IQR=1.046$ in Indoor1perople and $\lambda_{ctrl}^{scene}=\lambda_{ctrl}^{Indoor3}=1$, $IQR=0.816$ in Indoor3perople, respectively. The task for odometry is to do localization by analyzing the associated camera images. Moreover, $\lambda_{ctrl}^{scene}=\lambda_{ctrl}^{VO}=0.01$ and $IQR=2.263$. For each video, Equations (4.3), (4.6)-(4.7) and (4.9)-(4.10) are performed to obtain motion energy curve, frame no. of the extracted keyframe, utilization ratio and energy error ratio.

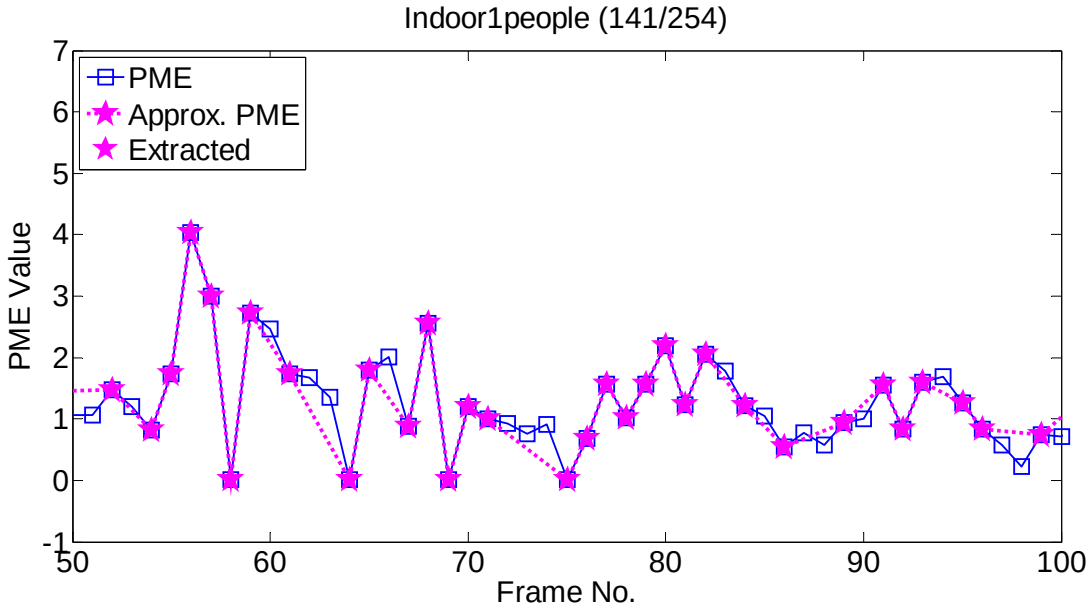
First, Indoor1people video is captured by the mobile camera doing horizontal motion and one walking people is in the scenario. Figure 4.26 shows fives snapshots of Indoor1people and Figure 4.27(a) presents the motion energy curve produced by

perceived motion energy value of captured frames and the extracted keyframes and the frame no. of extracted keyframes. When motion direction of camera and walking people is the same, the perceived motion energy value would be decreased. If motion direction is opposite, length of motion vectors increases to make large perceived motion energy value. Figure 4.27(b) shows the motion curve from Frames 50 to 100 to present that the approximation motion curve is close to original motion curve by only using 141 of 524 frames. Next, the energy error of each frame is computed by Equation (4.10) as shown in Figure 4.28.

Figure 4.26. The snapshots of Indoor1people.



(a)



(b)

Figure 4.27. (a) The motion energy curve, the approximation motion energy curve and the frame no. of extracted keyframes (b) The motion energy curve produced by captured frames and extracted keyframes from Frames 50 to 100.

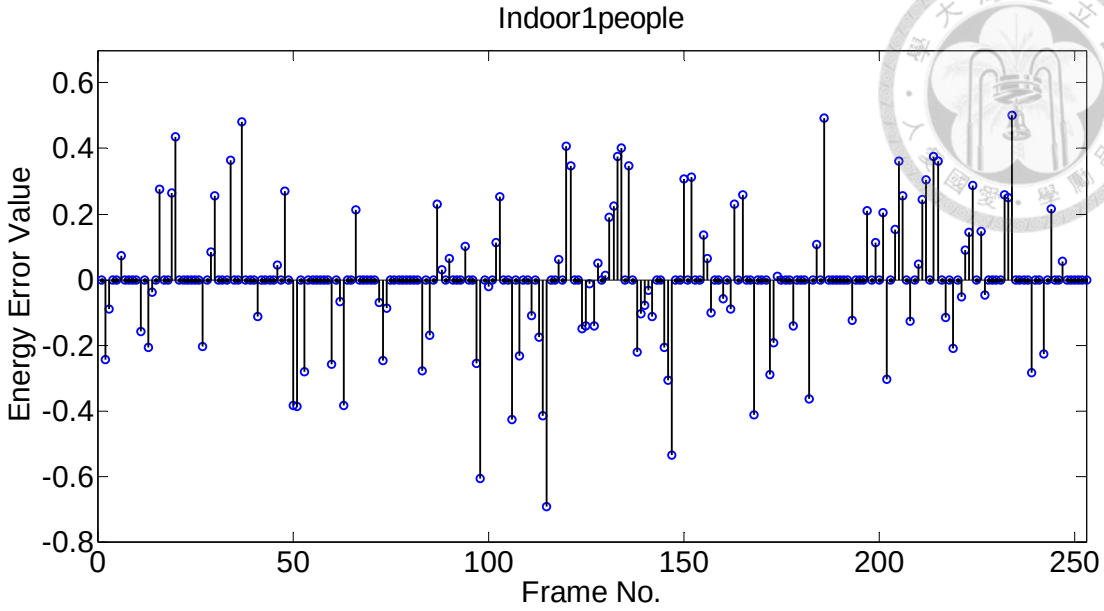


Figure 4.28. The energy error of each frame.

Figure 4.29 shows the frames per second (fps) results of traditional, $Fixed(1,3)$, the proposed PME-based keyframe extraction algorithm of indoor1people. From first second to fifth second and from tenth second to thirteenth second, one people walks in the monitored scene and moves in an opposite direction of mobile camera. Hence, perceived motion energy increases. Moreover, from sixteenth second to seventeenth second, the other appears from afar. In these three periods, sampling rate should increase to catch the changing of motion energy. PME-based sampling strategy performs well in these three periods. Moreover, PME-based sampling method dynamically adjusts frame rate based on the complexity of background in the rest of time. If background consists of various decorations, the motion energy would be

generated. If background is just a simple wall without anything on it, less motion energy is produced even when camera is mobile. Thus, from fifth second to ninth second and from fourteenth second to fifteenth second, Based on richness of background, PME-based method dynamically extract appropriate keyframes. In $Fixed(1,3)$, when people walks in the scene, each captured frame is sampled. When no moving people in the scene, fixed sampling strategy selects one frame having most edges from three consecutive frames. However, no moving persons in the scene do not represent the background is stationary. Hence, fixed sampling would lose background information when camera moves fast.

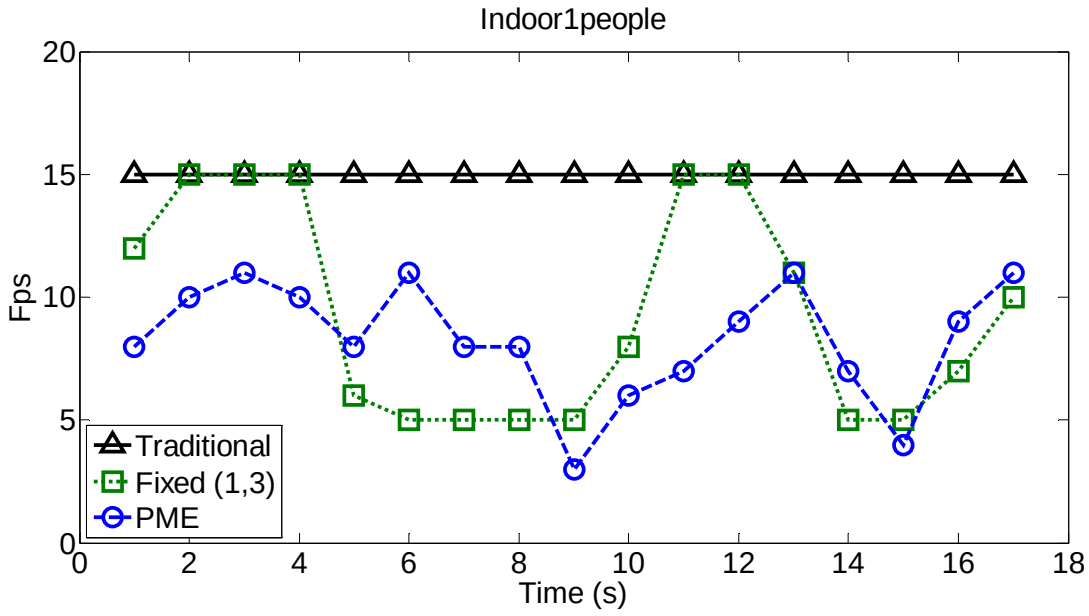


Figure 4.29. Frames per second results of traditional, $Fixed(1,3)$ and proposed PME-based keyframe extraction algorithm.

Table 4.5 shows the results of average available bandwidth to one frame based on

frames per second result as shown in Figure 4.29. For traditional and PME-based sample method, average frame rate is fifteen and eight during 18 seconds, respectively. PME-based method only samples 141 keyframes from 254 frames and has 55.51% utilization ratio. The absolute value of energy error ratio is 0.2%.

Table 4.5. Results of average available rate to each frame, energy error ratio and number of extracted keyframe of Indoor1people video.

Average available Rate for each frame				
Traditional	PME-based	Utilization Ratio	Energy Error Ratio	#Extracted Frames/ #Total Frames
$B_{Net}/15$	$B_{Net}/8$	55.51%	0.19%	141/254

For the purpose of video transmission, the extracted keyframes are processed with layer-based transmission mechanism before being placed into network. Layer-based transmission mechanism assigns different priorities to moving people (Label 1), background objects with edges (Label 2) and background without having edges (Label 3), for example, walls. Moreover, visual quality parameter λ_{PSNR}^{scene} in the Layer-based transmission mechanism is designated by users. Based on the sensing limitation of human eyes, λ_{PSNR}^{scene} is given as 30 in the dissertation. Figure 4.30 shows the results of three represented frames for three different sampling strategies, and their decoded visual quality. The visual quality of high priority regions is encoded first. If any bandwidth is

left, it is used to encode the middle-priority and/or low-priority regions. Therefore, the visual quality of high-priority region including moving objects is guaranteed and the required bandwidth for transmission is efficiently decreased for the frames without having moving objects. Under condition of same bandwidth (1.4 Mbps), PME-based sampling method achieves the best visual quality of three labels.

Server		End-users ($\lambda_{PSNR}^{scene} = 30$)								
		Bandwidth (1.4Mbps)								
		Received Frame								
No.	Origin captured	Traditional			Fixed Sampling			PME-Based		
26										
	Priority	H	M	L	H	M	L	H	M	L
	PSNR	34.9	31.0	16	34.0	31.0	16.4	34.9	31.0	32.0
110										
	Priority	H	M	L	H	M	L	H	M	L
	PSNR		31.2	17.0		31.2	32.0		31.2	32.0
172										
	Priority	H	M	L	H	M	L	H	M	L
	PSNR	31.8	31.7	2.0	31.8	31.7	2.0	31.8	31.7	32.2

Figure 4.30. Three represented frames and their encoded frames under different sampling strategy with acceptable visual quality PSNR = 30dB in 1.4Mbps bandwidth. H: high (Label 1 regions), M: middle (Label 2 regions); L: low (Label 3 regions).

Figures 4.31-4.33 show the visual quality of decoded images for three different

sampling strategies under condition of same bandwidth. In traditional manner, only the regions of label 1 and 2 achieve the given visual quality value, 30. In $Fixed(1,3)$ sample manner, the regions of label 1 and 2 also achieve desired visual quality value. Moreover, some regions of label 3 achieve the visual quality. In proposed PME-based sample manner, all labels achieve the given visual quality value. Based on the video content, Indoor1people is divided into three phases. Phase I has one moving people from Frames 5 to61, Phase II has no moving people from Frames 62 to 145 and Phase III has one moving people who is close to camera from Frames 146 to 187. For presenting the quality of decoded images, quality ratio is computed in three phases for traditional, $Fixed(1,3)$ and PME-based. Quality ratio is defined as follows:

$$QualityRatio_{\phi(j)} = \frac{\sum_{\mu \in \Psi(j)} \left(I_{L_1}^{F(\mu)} + I_{L_2}^{F(\mu)} + I_{L_3}^{F(\mu)} \right)}{\sigma\left(L_{\phi(j)}\right) \times \sigma\left(\Psi(j)\right)} \times 100\%, \quad j = \{1,2,3\}$$
$$\Psi(j) = \{\eta \in \Theta \mid F(\eta) \in \phi(j)\} \tag{4.11}$$
$$L_{\phi(1)} = \{Labels \text{ in Phase I}\} \qquad \phi(1) = \{F(k) \text{ in Phase I}\}$$
$$L_{\phi(2)} = \{Labels \text{ in Phase II}\} \qquad \phi(2) = \{F(k) \text{ in Phase II}\}$$
$$L_{\phi(3)} = \{Labels \text{ in Phase III}\} \qquad \phi(3) = \{F(k) \text{ in Phase III}\}$$

$$I_{L_\rho}^{F(k)} = \begin{cases} 1 & if \ quality \ of \ L_\rho \ achieves \ \lambda_{PSNR}^{scene} \ in \ F(k) \\ 0 & otherwise \end{cases} \tag{4.12}$$

Table 4.6 shows the quality ratio results of traditional, $Fixed(1,3)$ and PME-based

sample method in three phases are computed by Equations (4.11)-(4.12). The proposed PME-based sample method gets 100% in all phases.

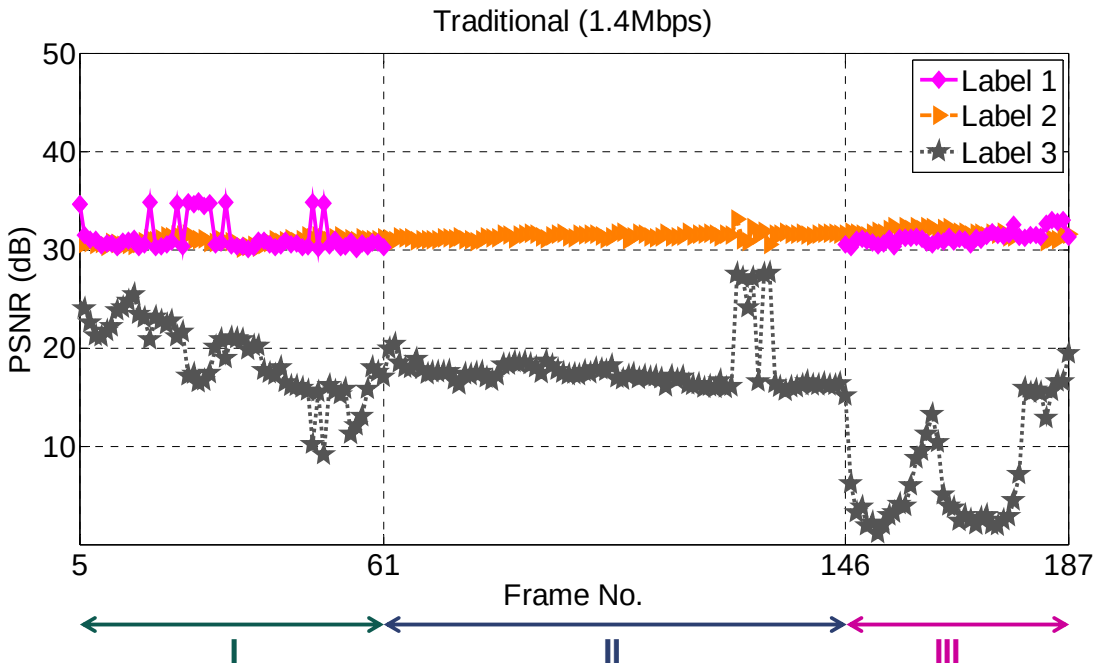


Figure 4.31. The quality of decoded frames with $\lambda_{PSNR}^{scene} = 30$ in 1.4Mbps for traditional.

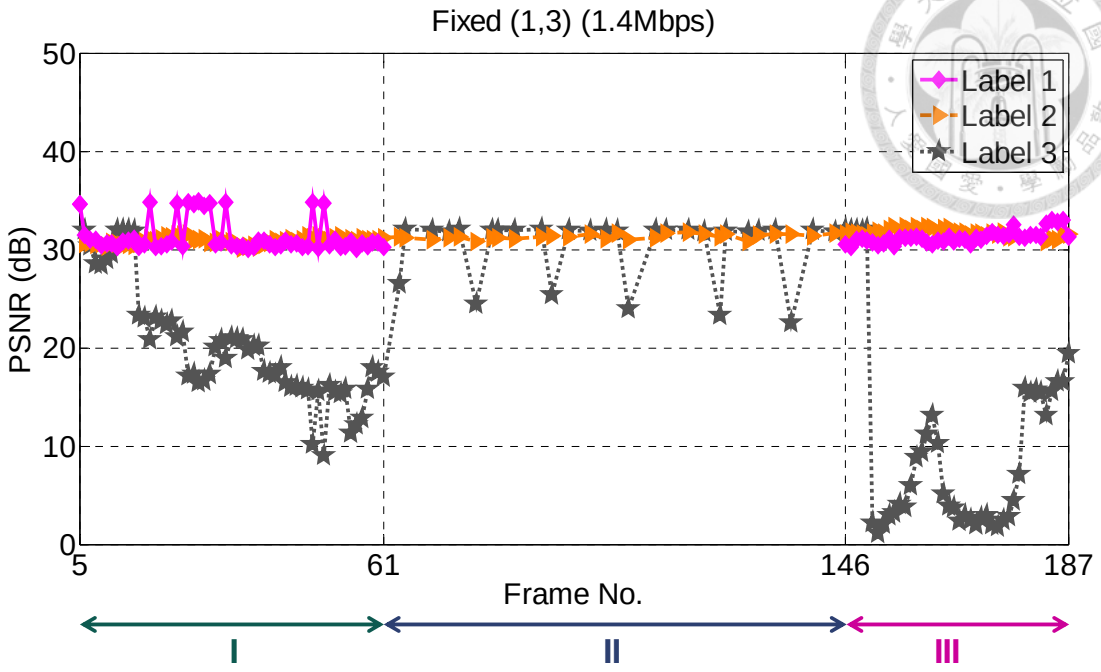


Figure 4.32. The quality of decoded frames with $\lambda_{PSNR}^{scene} = 30$ in 1.4Mbps for $Fixed(1,3)$.

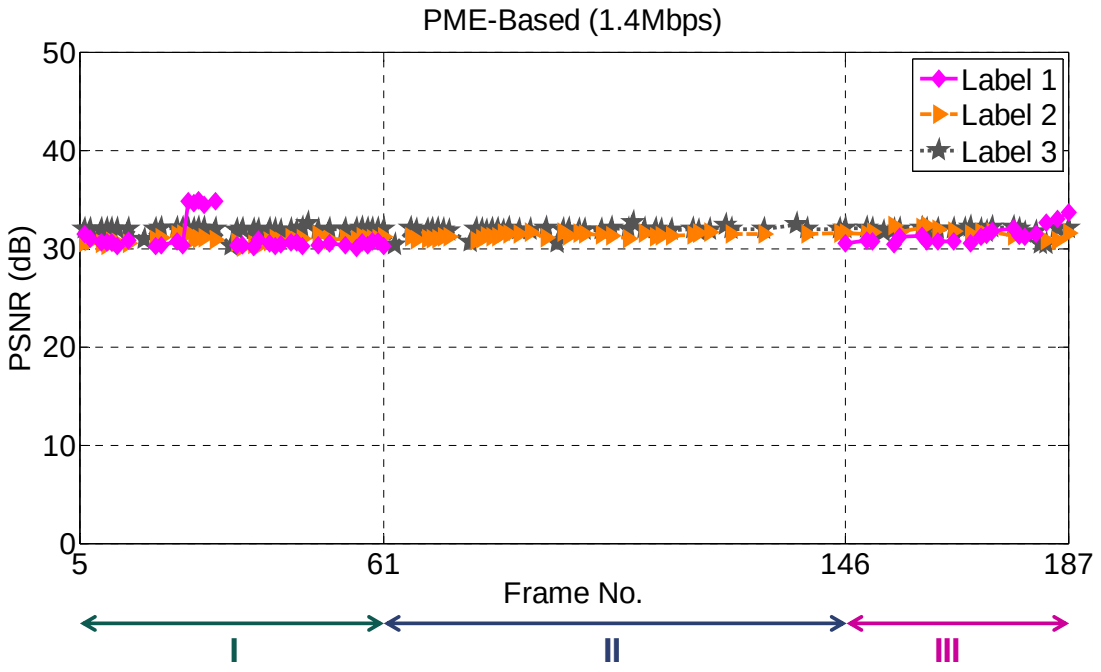


Figure 4.33. The quality of decoded frames with $\lambda_{PSNR}^{scene} = 30$ in 1.4Mbps for PME-based .

Table 4.6. Quality ratio of three phases (1.4 Mbps) in Indoor1people video.

Phase	I	II	III
Traditional	67%	50%	63%
Fixed Sampling	71%	89%	67%
PME-based	100%	100%	100%

Next, another important issue in video transmission is bandwidth limitation. Figure 4.34 shows that all labels achieve PNSR=30 in traditional, $Fixed(1,3)$ and PME-based manner under condition of bandwidth 2.2Mbps. Table 4.7 shows the quality ratio results of three phases computed by Equations (4.11)-(4.12) and they all achieve 100%. Figure 4.35 shows the saving percentage in traditional, $Fixed(1,3)$ and proposed PME-based methods under the condition of 2.2 Mbps. $Fixed(1,3)$ gets better saving percentage than proposed PME-based when no moving people in the scene. However, PME-based has better saving result overall as shown in Table 4.8. Moreover, PME-based has over-half saving ratio in bandwidth.

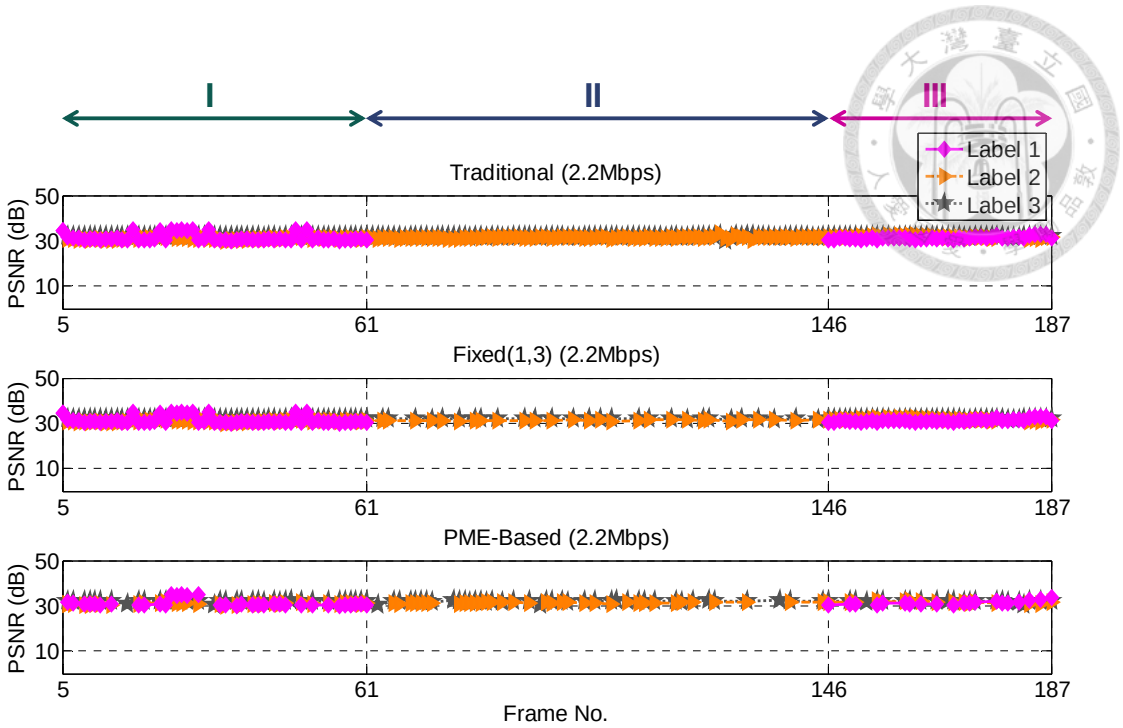


Figure 4.34. The quality of decoded frames with $\lambda_{PSNR}^{scene} = 30$ (2.2 Mbps). (a) Traditional (b) $Fixed(1,3)$ (c) PME-based.

Table 4.7. Quality ratio of three phases (2.2 Mbps) in Indoor1people video.

Phase	I	II	III
Traditional	100%	100%	100%
Fixed Sampling	100%	100%	100%
PME-based	100%	100%	100%

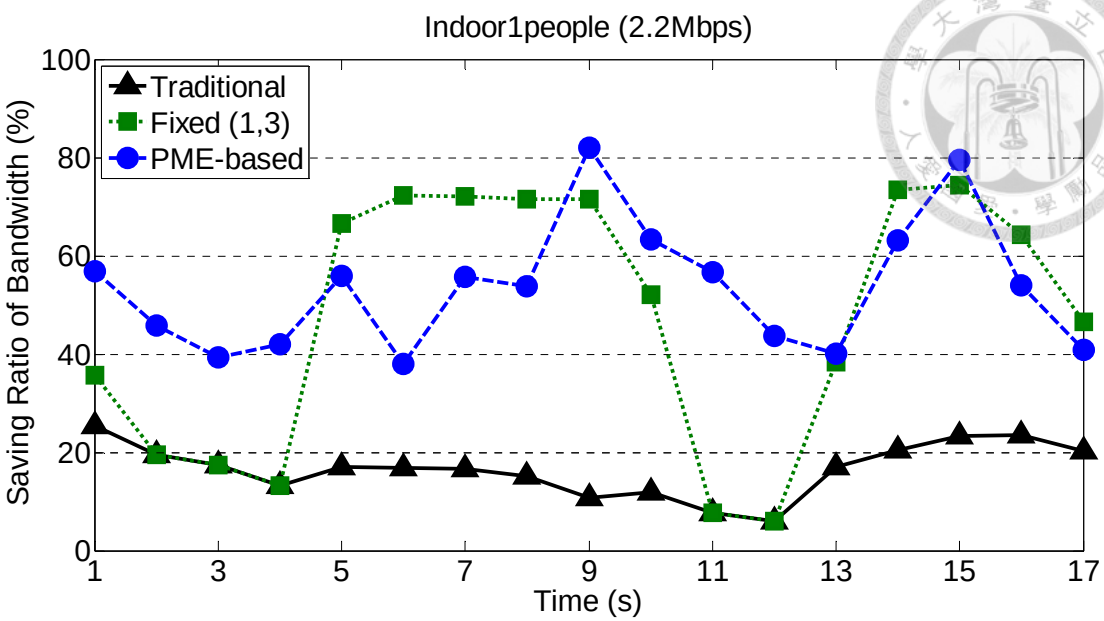


Figure 4.35. Saving percentage of bandwidth compared with that of the lossless case.

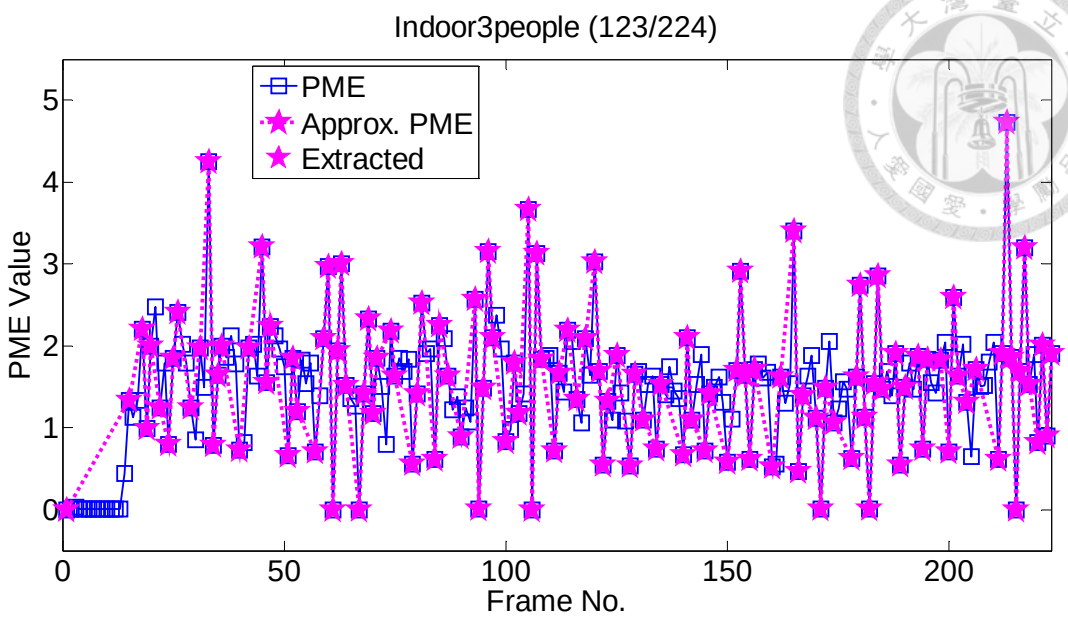
Table 4.8. Average saving ratio of bandwidth of three different Sampling Strategies in Indoor1people video (2.2Mbps).

	Traditional	Fixed Sampling	PME-based
Average Saving Ratio	17%	47%	54%

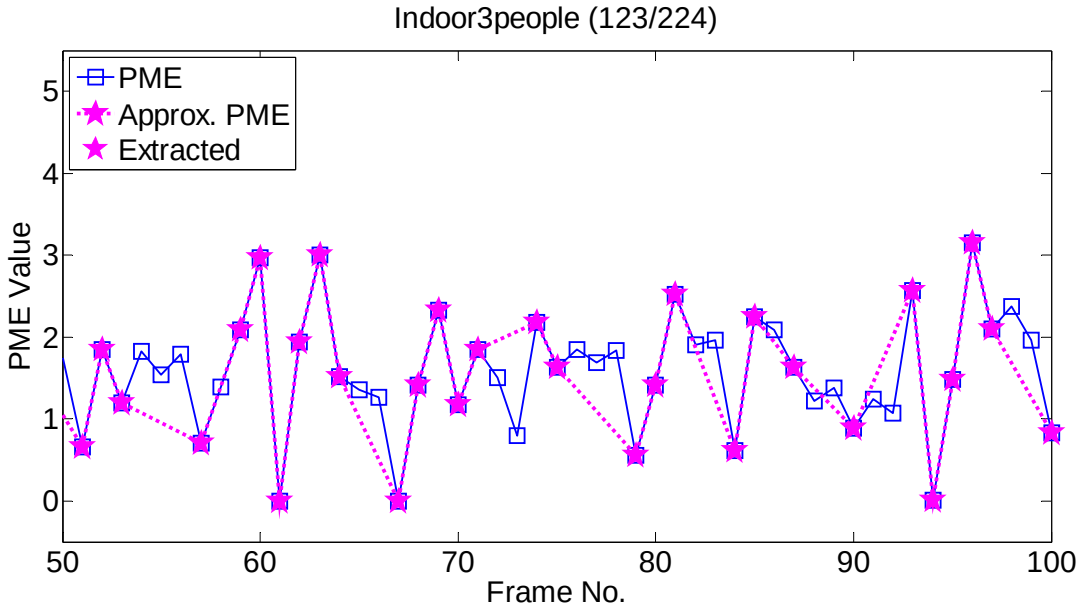
The second experimented video is Indoor3people which has two walking and one stationary people captured by the mobile camera doing ego-rotation movement from left to right. Figure 4.36 shows five snapshots of Indoor3people and Figure 4.37(a) presents the motion curve produced from perceived motion energy of each captured frame and the extracted frames and the frame no. of extracted keyframe. Before Frame 15, the camera is in the initialization stage without moving and no moving people in the scene. Hence, perceived motion energy value is zero in that period. Figure 4.37(b) shows the

motion curve from Frames 50 to 100 to present that the approximation motion curve is close to original motion curve by only using 123 of 224 frames. Next, the energy error of each frame is computed by Equation (4.10) as shown in Figure 4.38.

Figure 4.36. The snapshots of Indoor3people.



(a)



(b)

Figure 4.37. (a) The motion energy curve, the approximation motion energy curve and the frame no. of extracted keyframes (b) The motion energy curve produced by captured frames and extracted keyframes from Frames 50 to 100.

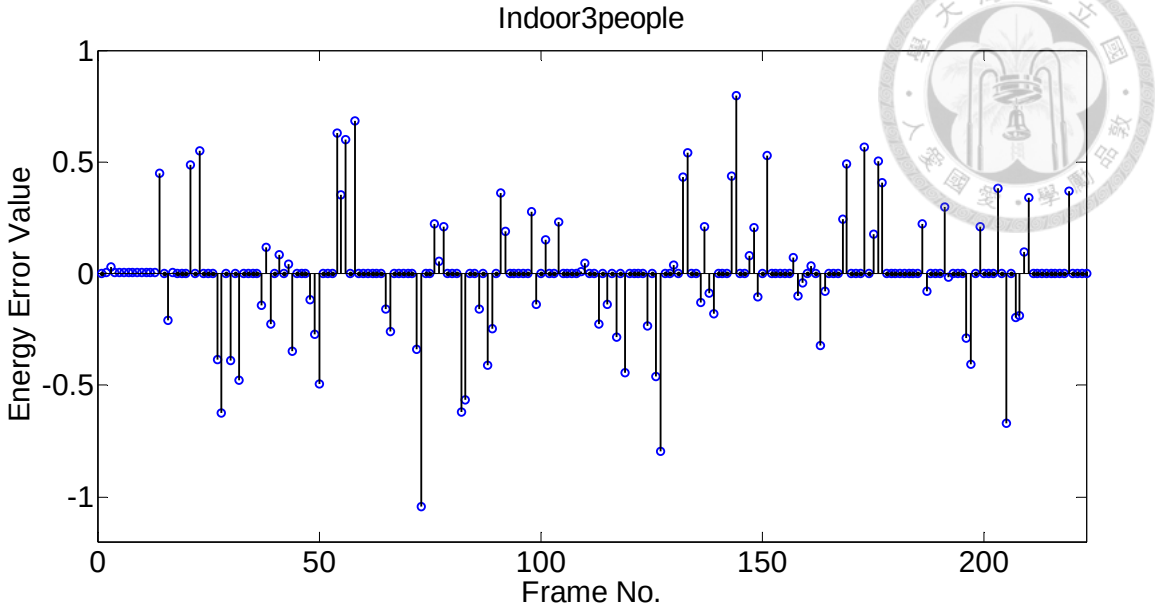


Figure 4.38. The energy error of each frame.

Figure 4.39 shows the frames per second (fps) results of traditional, $Fixed(1,3)$ proposed PME-based keyframe extraction algorithm of Indoor3people. Moving people show from seventh second to eleventh second. Frame rate increases due to the motion energy caused by complexity background and moving camera in PME-based method. Table 4.9 shows the results of average available bandwidth to one frame based on fps as shown in Figure 4.39. For traditional and PME-based sample method, average frame rate is fifteen and eight during 16 seconds, respectively. PME-based manner only samples 123 keyframes from 224 frames, and the utilization ratio is 54.91% and absolute value of energy error ratio is 0.13%. Figure 4.40 shows the results of three represented image frames for three different sampling strategies, and their encoded

visual quality. Under condition of same bandwidth (2.6 Mbps), PME-based sampling method gets the best visual quality of three labels.

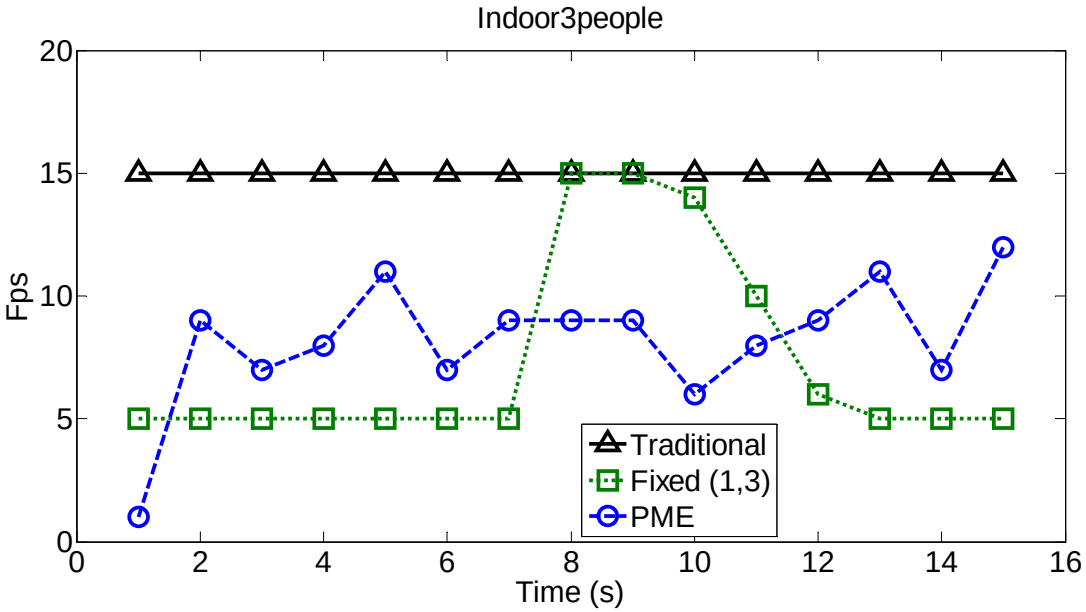


Figure 4.39. Frames per second results of traditional, *Fixed*(1,3) and proposed PME-based keyframe extraction algorithm.

Table 4.9. Results of average available rate to each frame, energy error ratio and number of extracted keyfram of Indoor3people video.

Average available Rate for each frame				
Without extraction	PME-based extraction	Utilization Ratio	Energy Error Ratio	#Extracted Frames/ #Total Frames
$B_{Net}/15$	$B_{Net}/8$	54.91%	0.13%	123/224

Server		End-users ($\lambda_{PSNR}^{scene} = 30$)								
			Bandwidth (2.6Mbps)							
			Received Frame							
No.	Origin captured		Traditional			Fixed Sampling			PME-Based	
53										
		Priority	H	M	L	H	M	L	H	M	L
		PSNR		32.1	21.4		32.1	33.0		32.1	33.0
123										
		Priority	H	M	L	H	M	L	H	M	L
		PSNR	33.8	32.3	5.6	33.8	32.3	5.6	33.8	32.3	33.5
189										
		Priority	H	M	L	H	M	L	H	M	L
		PSNR		31.5	4.4		31.5	30.1		31.5	30.1

Figure 4.40. Three represented frames and their encoded frames under different sampling strategy with acceptable visual quality PSNR = 30dB in 2.6Mbps bandwidth. H: high (Label 1 regions), M: middle (Label 2 regions); L: low (Label 3 regions).

Figures 4.41-4.43 show the visual quality of decoded images for three different sampling strategies under condition of same bandwidth. Based on the video content, Indoor3people is divided into three phases. Phase I has no moving people from Frames 16 to 104, Phase II has two moving people and one stationary man laying on balustrade from Frames 105 to 167 and Phase III has the stationary man from Frames 168 to 225. In traditional manner, only the regions of label 1 and 2 achieve the given visual quality

value, 30. In $Fixed(1,3)$ sample manner, the regions of label 1 and 2 also achieve desired visual quality value. Moreover, regions of label 3 achieve the visual quality in Phase I and III. In proposed PME-based sample manner, all labels achieve the given visual quality value in all three phases. For presenting the quality of decoded images, quality ratio is computed by Equations (4.11)-(4.12) in three phases for traditional, $Fixed(1,3)$ and PME-based method Table 4.10 shows the quality ratio results of three phases and PME-based method gets 100% in all phases under condition of 2.6Mbps bandwidth.

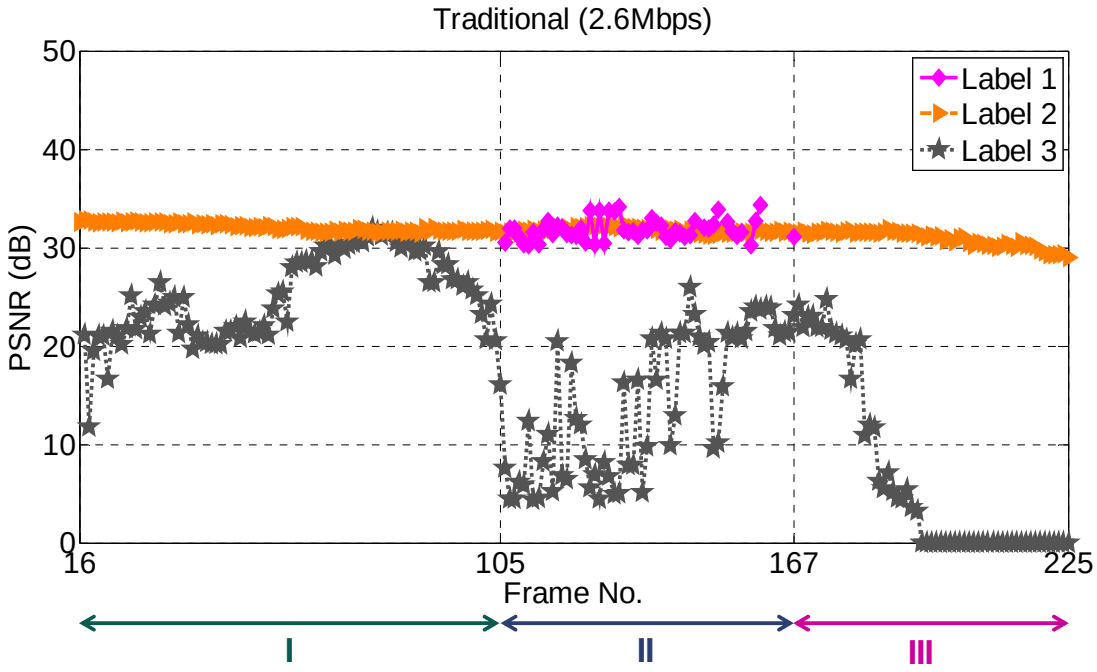


Figure 4.41. The quality of decoded frames with $\lambda_{PSNR}^{scene} = 30$ in 2.6Mbps for traditional.

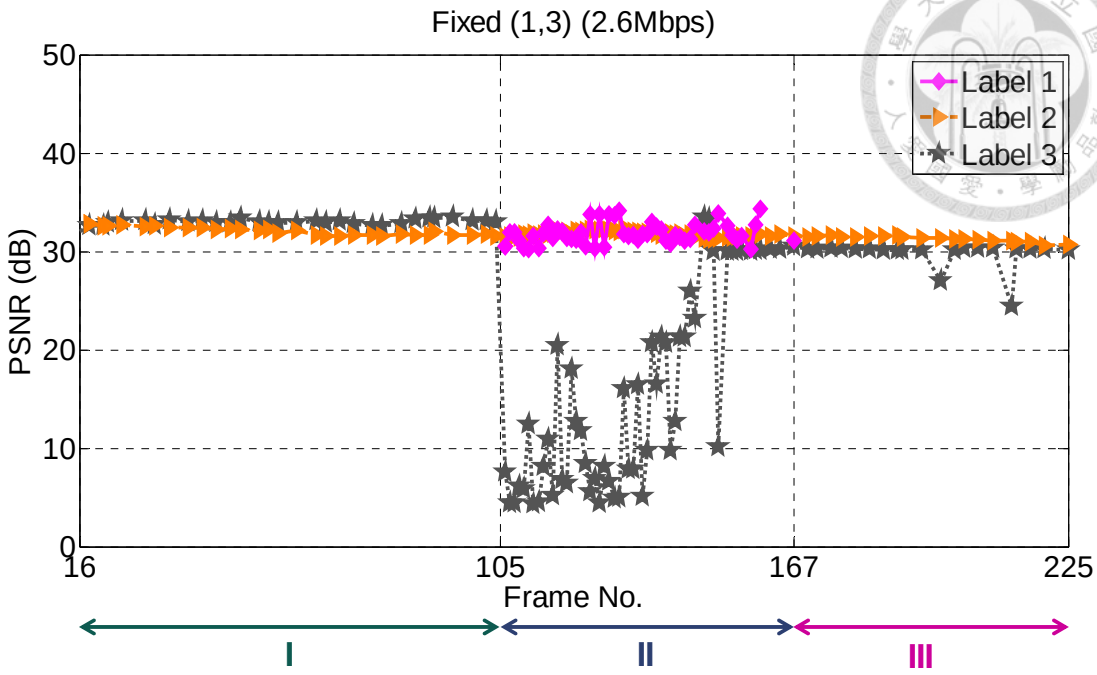


Figure 4.42. The quality of decoded frames with $\lambda_{PSNR}^{scene} = 30$ in 2.6Mbps for $Fixed(1,3)$.

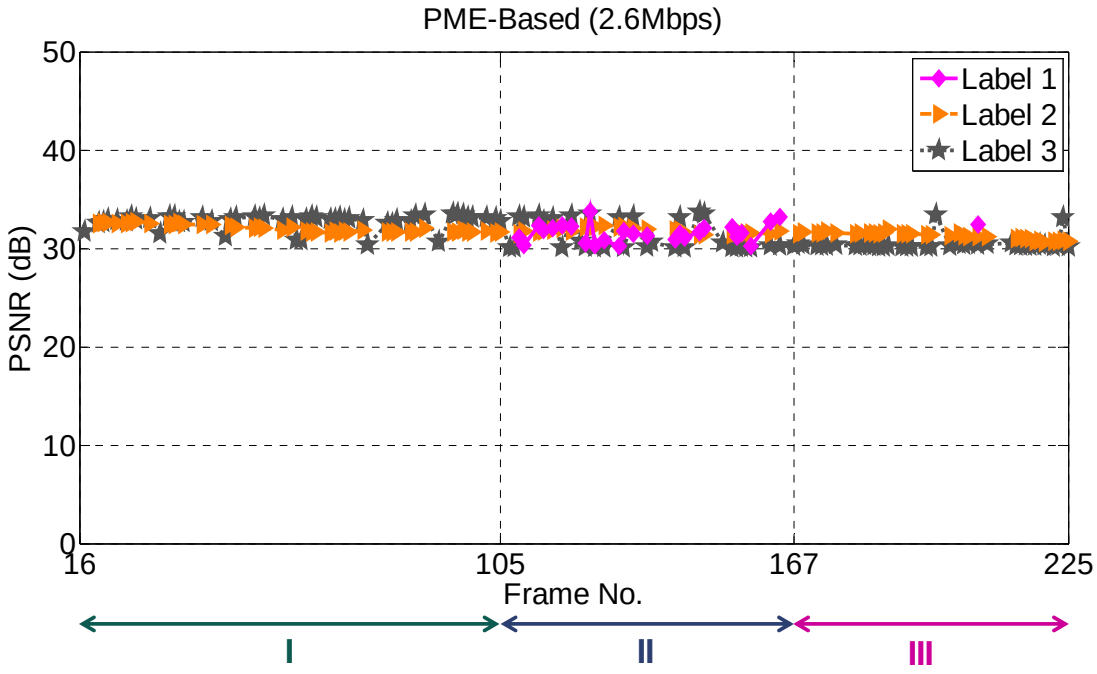


Figure 4.343. The quality of decoded frames with $\lambda_{PSNR}^{scene} = 30$ in 2.6Mbps for PME-based .

Table 4.10. Quality ratio of three phases (2.6 Mbps) in Indoor3people video.

Phase	I	II	III
Traditional	63%	67%	46%
Fixed Sampling	100%	75%	95%
PME-based	100%	100%	100%

Figure 4.44 shows that all labels achieve PNSR =30 of traditional, $Fixed(1,3)$ and PME-based in all phases under condition of bandwidth 2.2Mbps. Table 4.11 shows the quality ratio results of three phases computed by Equations (4.11)-(4.12) and they all achieve 100% in all phases. Figure 4.45 shows the saving percentage in traditional, $Fixed(1,3)$ and PME-based under bandwidth condition of 3.2Mbps. $Fixed(1,3)$ gets better saving result 59% overall as shown in Table 4.12. However the saving ratio in $Fixed(1,3)$ has large gap in two phases: with moving people and without moving people. On the other hand, PME-based method has more smooth saving curve and also has over-half saving ratio in bandwidth. For the purpose of increasing saving percentage of bandwidth, it is obviously to place less frames into network. However, if too little frames are transmitted to controller, system performance might be affected and get bad result. Another captured experimental video is used to verify the suggestion in Part II.

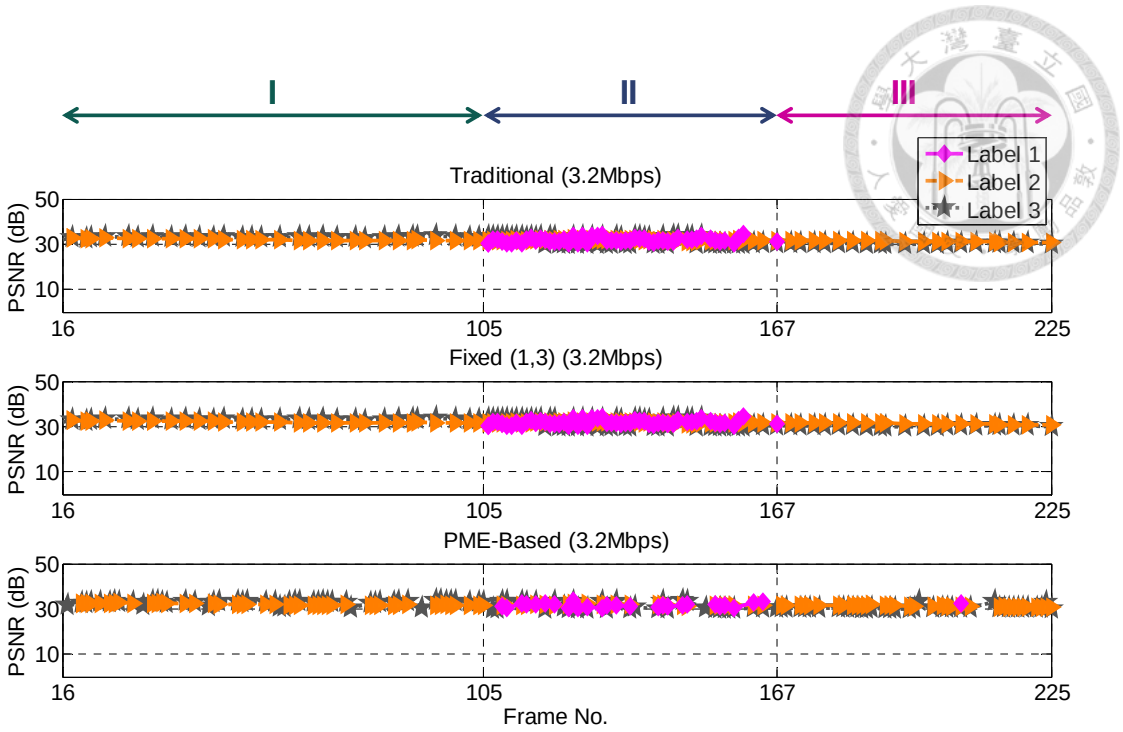


Figure 4.44. The quality of decoded frames with $\lambda_{PSNR}^{scene} = 30$ in 3.2Mbps. (a) Traditional (b) $Fixed(1,3)$ (c) PME-based .

Table 4.11. Quality ratio of three phases (3.2 Mbps).

Phase	I	II	III
Traditional	100%	100%	100%
Fixed Sampling	100%	100%	100%
PME-based	100%	100%	100%

Table 4.12. Average Saving Ratio of Bandwidth of three different sampling strategies (3.2Mbps).

	Traditional	Fixed Sampling	PME-based
Average Saving Ratio	10%	59%	51%

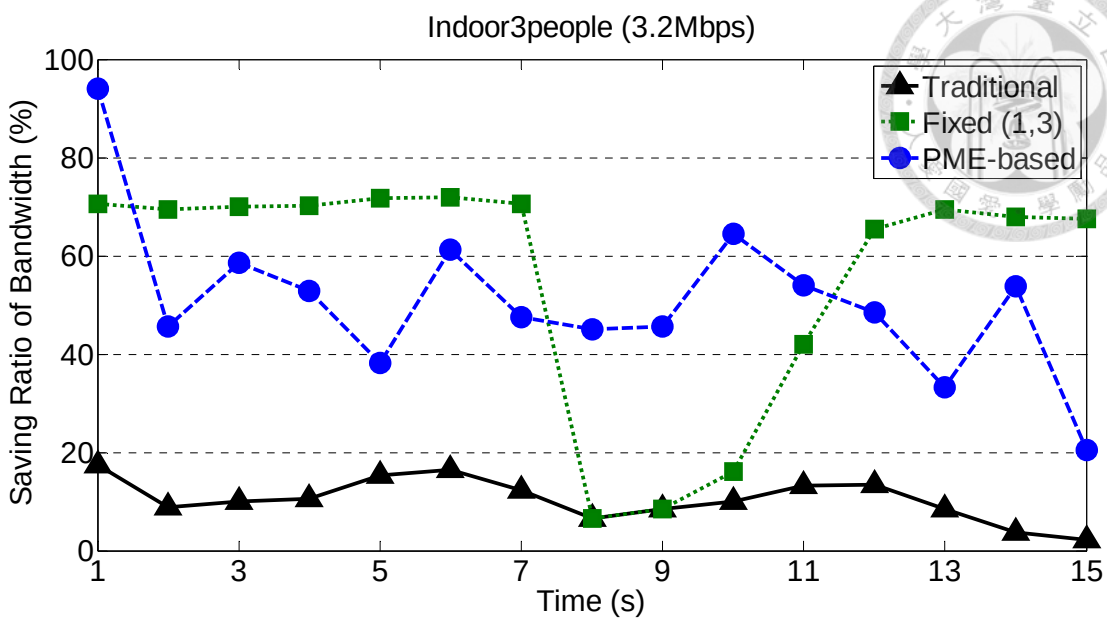


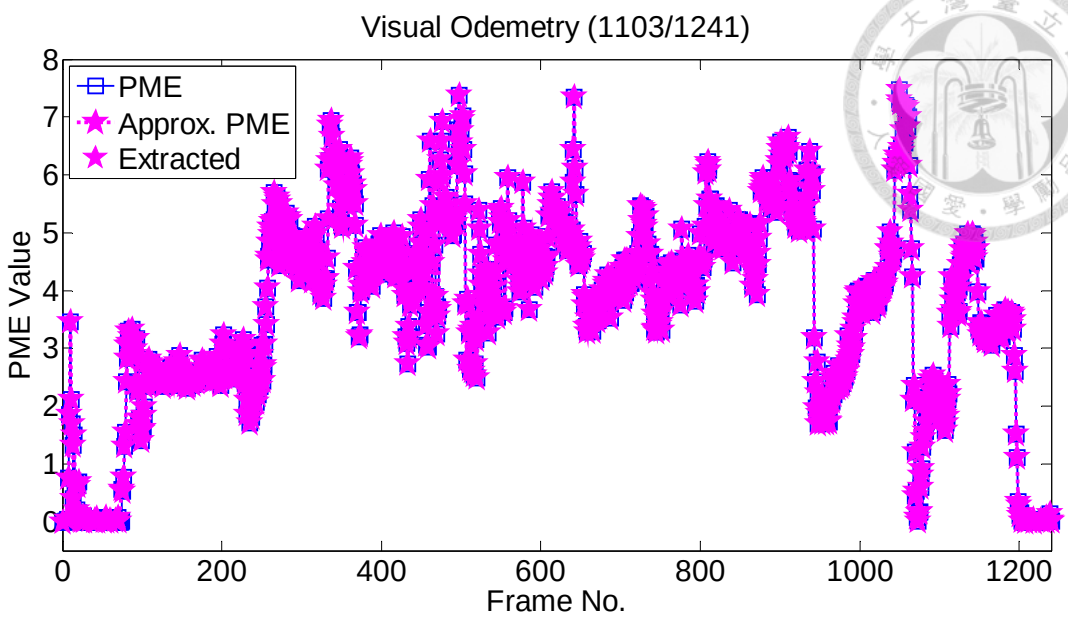
Figure 4.45. Saving percentage of bandwidth compared with that of the lossless case.

The third experimented video is Visual Odoemtry which is captured by the camera mounted in a wheeled robot to circuit a pillar as circle trajectory. Figure 4.46 shows five snapshots of Visual Odometry and Figure 4.47(a) presents the motion curve produced from perceived motion energy of each captured frame and the extracted frames and the frame no. of extracted keyframe. Because wheeled robot runs a patrol trajectory with non-constant velocity and the floor has complex pattern, perceived motion energy value has smoothly and rapidly increasing and decreasing stages as shown in Figure 4.47(a). Figure 4.43(b) shows the motion curve from Frames 50 to 100 to present that the approximation motion curve is close to original motion curve by only using 1103 of 1241 frames. Next, the energy error of each frame is computed by Equation (4.10) as

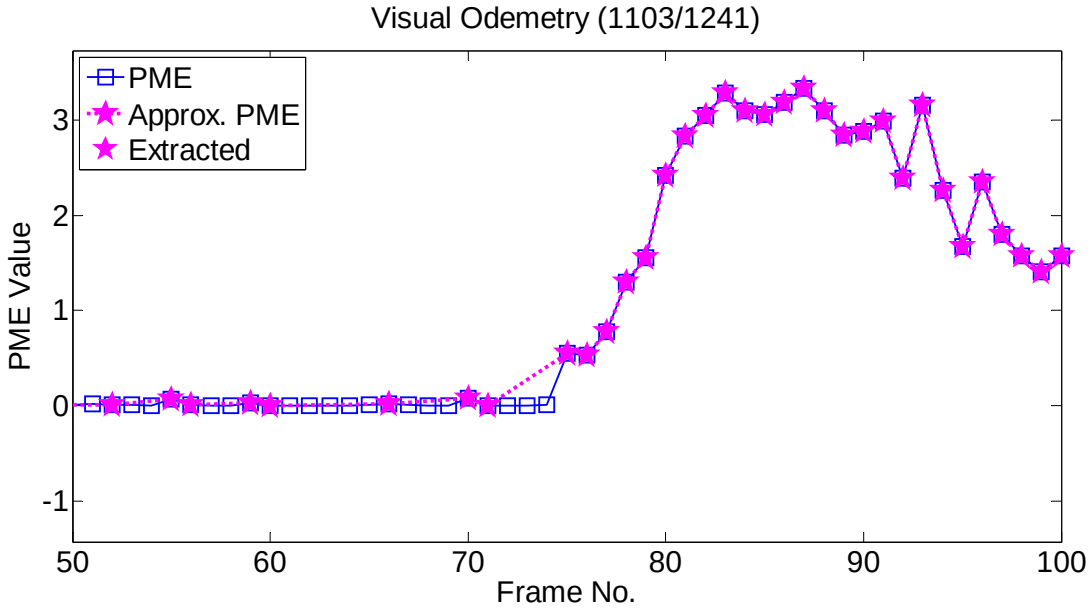
shown in Figure 4.48.

Figure 4.49 shows the frames per second (fps) results of traditional and the proposed PME-based keyframe extraction algorithm of Visual Odometry. Wheeled robot is in the initialization stage before fifth second (Frame 75). Wheeled robot starts to perform the circle trajectory from sixth second. Because the dynamics of perceived motion energy is high, frame rate stays over 10 to follow the changing in perceived motion energy value. Figure 4.50 and Figure 4.51 show the odometry results of two different sampling strategies: traditional and PME-based. The goal of odometry is to go back to the start point $(0,0)$. The Coordinates of end is $(-0.05,-0.03)$ in traditional sampling manner as shown in Figure 4.50 and is $(-0.09,0.06)$ in PME-based strategy as shown in Figure 4.51. It shows that the result of PME-based manner is close to $(0,0)$ and more important is that PME-based manner only utilizes 88.88% of origin video data to achieve it.

Figure 4.46. The snapshots of Visual Odometry.



(a)



(b)

Figure 4.47. (a) The motion energy curve, the approximation motion energy curve and the frame no. of extracted keyframes (b) The motion energy curve produced by captured frames and extracted keyframes from Frames 50 to 100.

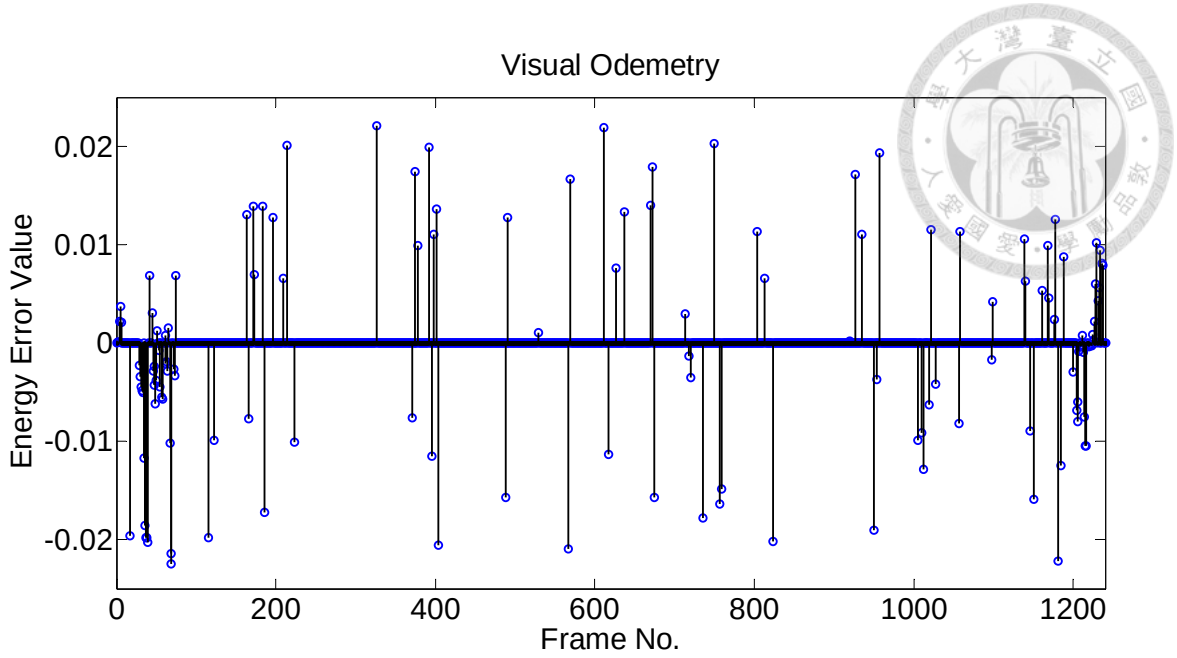


Figure 4.48. The energy error of each frame.

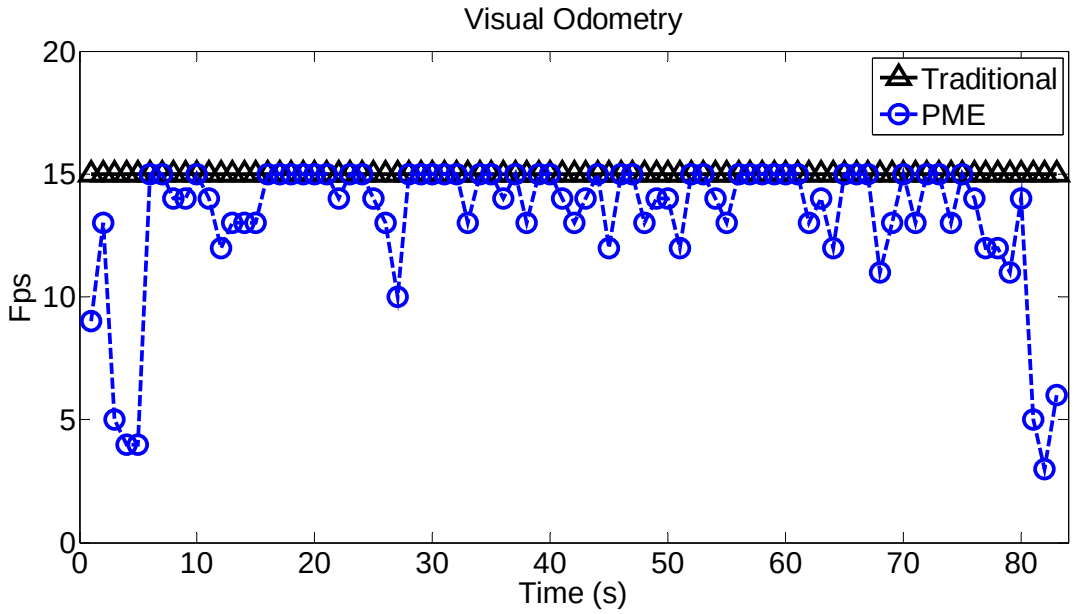


Figure 4.49. Frames per second results of traditional and proposed PME-based keyframe extraction algorithm.

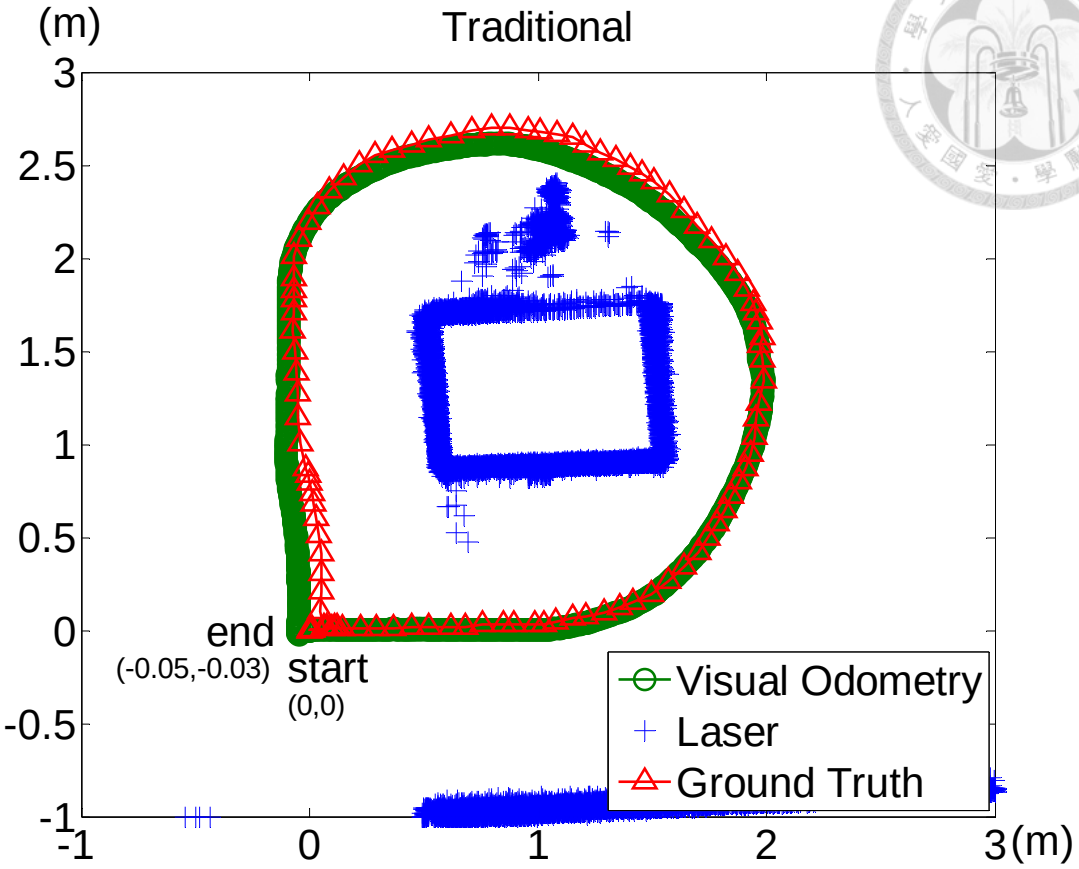


Figure 4.50. Odometry result of traditional sampling.

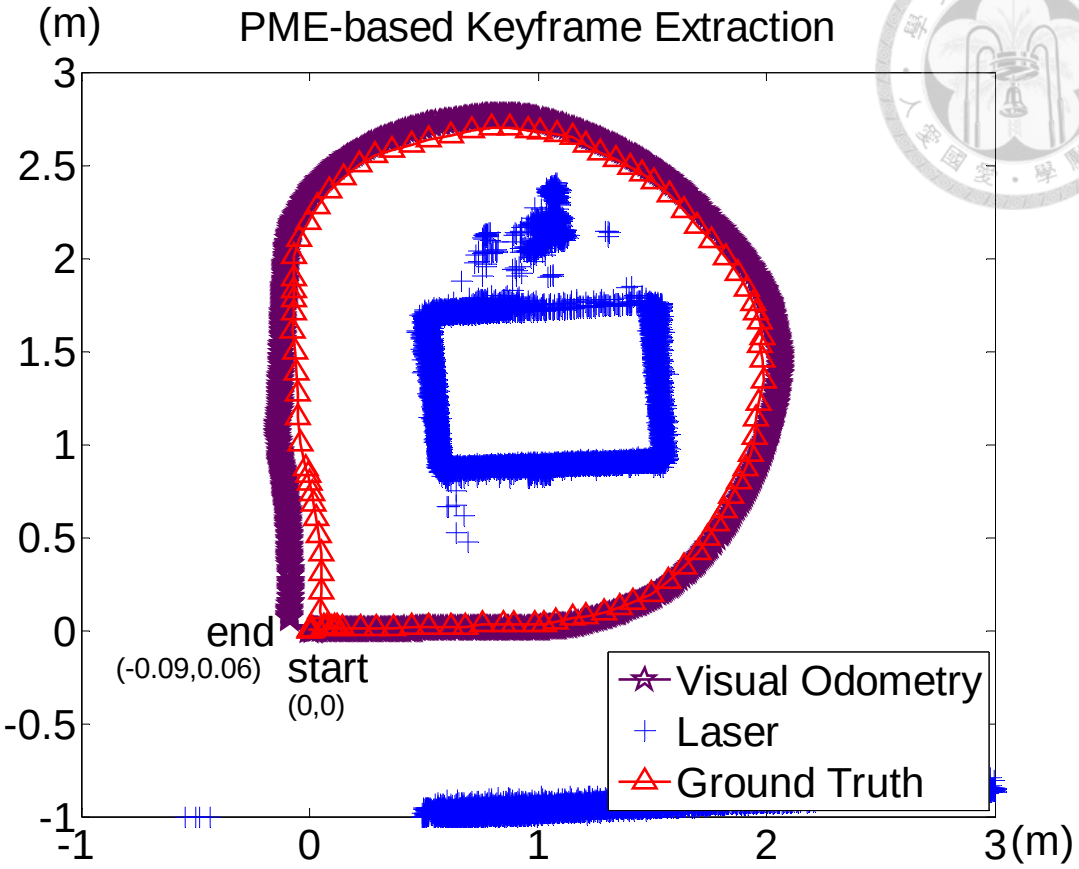


Figure 4.51. Odometry result of PME-based sampling method.

In order to compare the performance in traditional and PME-based manner, distance between start and end point is used as performance index. Distance is 0.05 and 0.11 in traditional and PME-based sampling strategy, respectively. The distance result exhibits the adaptability of PME-based keyframe extraction algorithm. Table 4.13 shows that PME-based strategy only samples 1103 keyframes from 1241 frames to achieve expected performance and has 88.88% utilization ratio. The absolute value of energy error ratio is almost zero.

Table 4.13. Results of distance from original point, utilization ratio, energy error ratio and number of extracted keyframe of Visual Odometry.

Distance results of Visual Odometry $dist = \left\|(x_{end}, y_{end}) - (0,0)\right\|$				
Traditional	PME-based	Utilization Ratio	Energy Error Ratio	#Extracted Frames/ #Total Frames
$dist = 0.05$	$dist = 0.11$	88.88%	0.0026%	1103/1241
Traditional: $[x_{end} \quad y_{end}] = [-0.05 \quad -0.03]$; PME-based: $[x_{end} \quad y_{end}] = [-0.09 \quad 0.06]$				

4.3.2.2 Part II: Visual sensor is stationary

A. Object motion direction: Single

Two tested videos, namely, SingleBall and Container are used to perform the keyframe extraction process as shown in Figure 4.11. The application task for SingleBall is to track the ball and for Container is moving objects tracking. Two different sampling strategies are implemented in two tested videos, that is, traditional and PME-based with threshold $AT = \lambda_{ctrl}^{scene} \times IQR$ where $\lambda_{ctrl}^{scene} = \lambda_{ctrl}^{Ball}$ and $\lambda_{ctrl}^{scene} = \lambda_{ctrl}^{Container}$ in SingleBall and Container, respectively. For SingleBall, $\lambda_{ctrl}^{Ball} = 1$ and $IQR = 0.029$ and for Container $\lambda_{ctrl}^{Container} = 1$ and $IQR = 0.005$. For each videos, Equations (4.3), (4.6)-(4.7) and (4.9)-(4.10) are performed to obtain the motion energy curve, the frame no. of extracted keyframe, the utilization ratio and the energy error ratio.

The first tested video is SingleBall which is captured by tossing a ball to pass through a box. Tracking algorithm with kalman filter is utilized to track the ball in four different scenarios; (a) initial (b) switch motion model from constant acceleration to constant velocity (c) initial location is random given (d) measurement noise is added. Figure 4.52 shows five snapshots of SingleBall and Figure 4.53 presents the motion energy curve produced from perceived motion energy of each captured frame and the extracted frames and the frame no. of extracted keyframe. The reason for motion energy

peak is that the ball appears in the frame. Next, the energy error of each frame is computed by Equation (4.10) as shown in Figure 4.54. Figure 4.55 shows the frame rate results of traditional and PME-based sampling strategy of SingleBall. When ball is out of the monitored range, frame rate decreases after 1 second. Figures 4.56 and 4.57 show the ball tracking error results of two different sampling strategies, traditional and PME-based in four different scenarios. Tracking error is defined as follows:

$$Tracking\ Error = \left\| Location_{\text{detected}} - Location_{\text{tracked}} \right\|_2 \qquad (4.13)$$

In initial scenario, kalman parameters are obtained when ball tracking works well. Figure 4.56(a) shows the tracking error in extracted keyframes. In fifth tracked location, ball starts to pass through the box, thus the error becomes large. When ball starts to appear in the scene, it immediately tracked and then error goes down. Figure 4.56(b) shows tracking error in the switching motion model scenario and big error still occurs when ball goes to pass through the box. Figure 4.57(a) shows the tracking error in the scenario of initial location that is not given as actual detected one. Figure 4.57(b) shows tracking error when measurement noise is added and PME-based sampling strategy performs well. The performance of four scenarios by using PME-based sampling strategy is satisfactory.

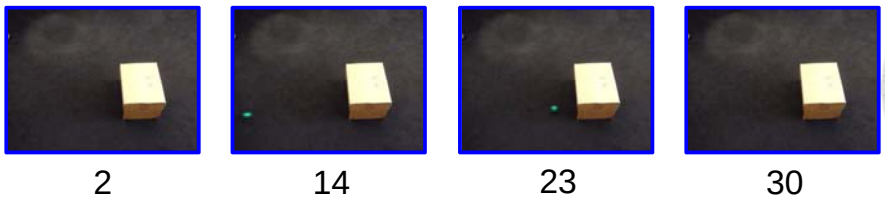


Figure 4.52. The snapshots of SingleBall.

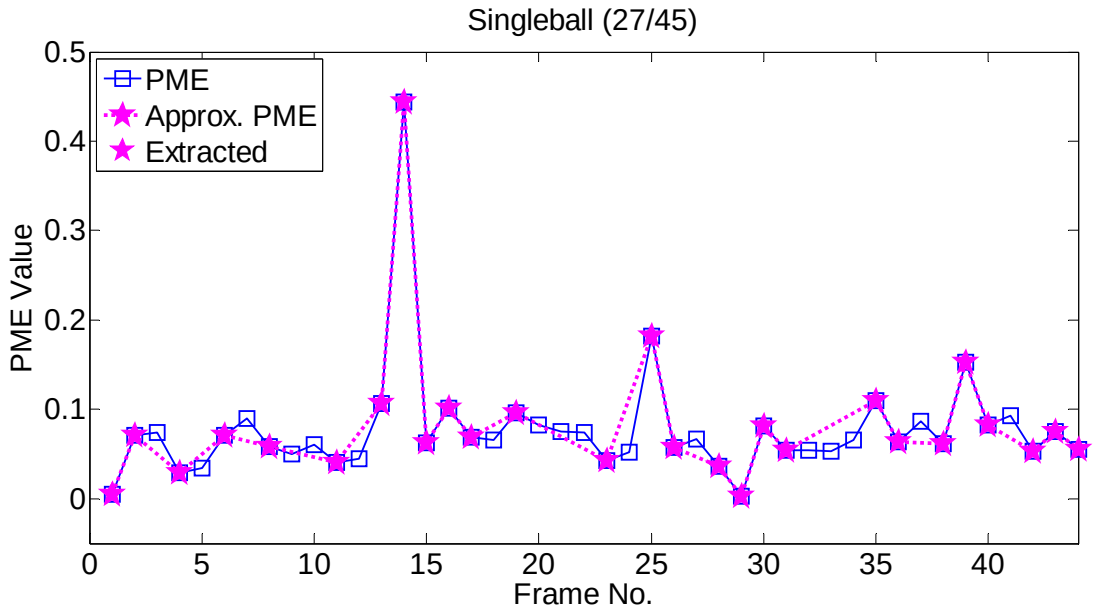


Figure 4.53. The motion energy curve, the approximation motion energy curve and the frame no. of extracted keyframes.

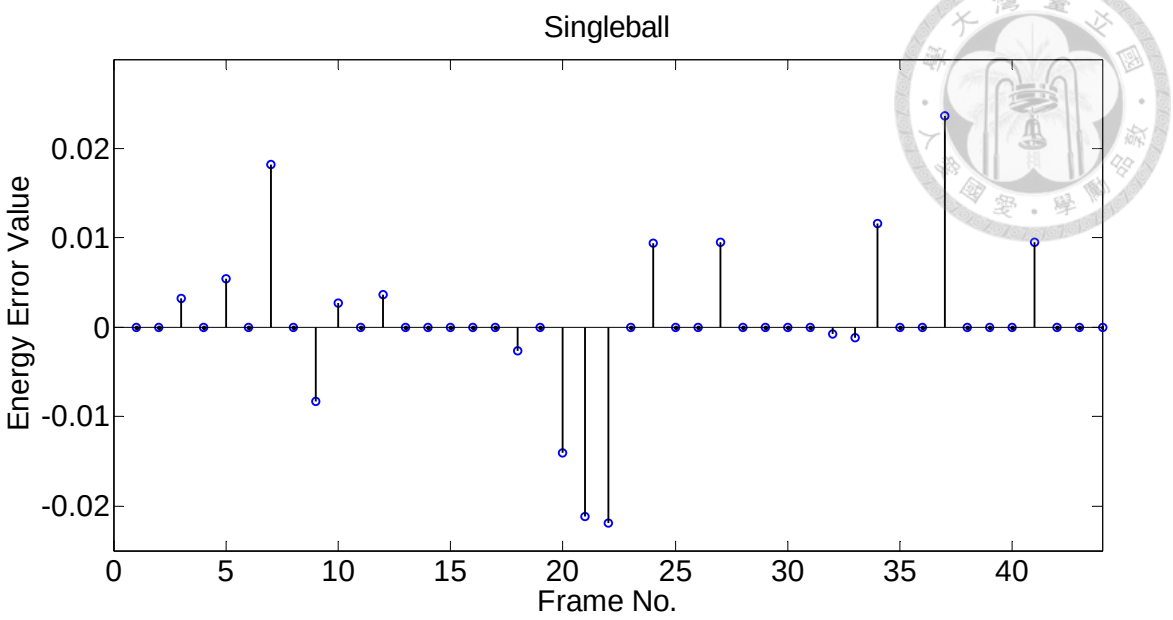


Figure 4. 54. The energy error of each frame.

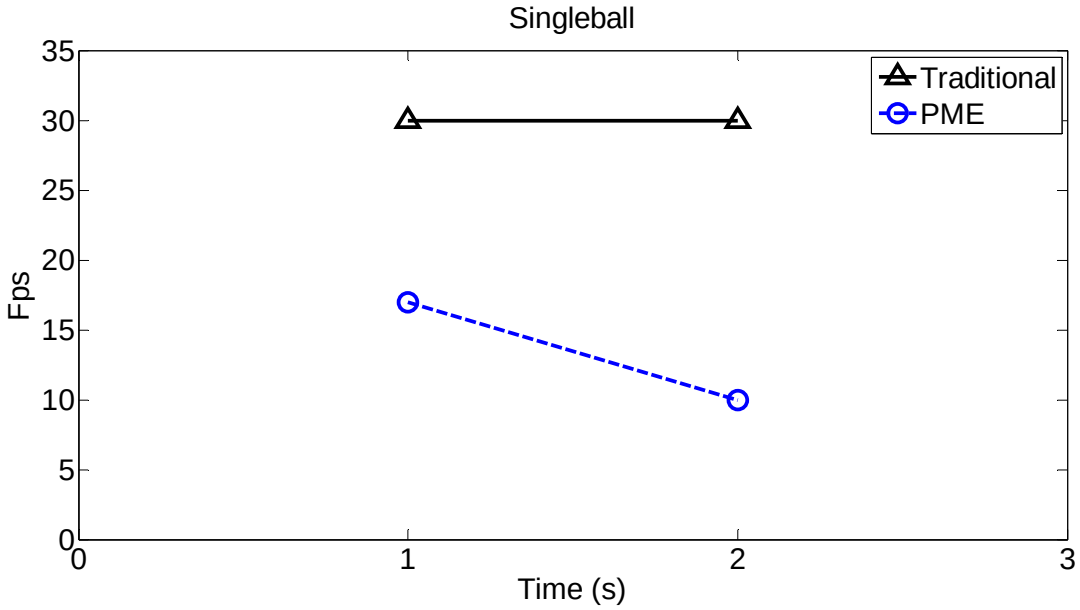
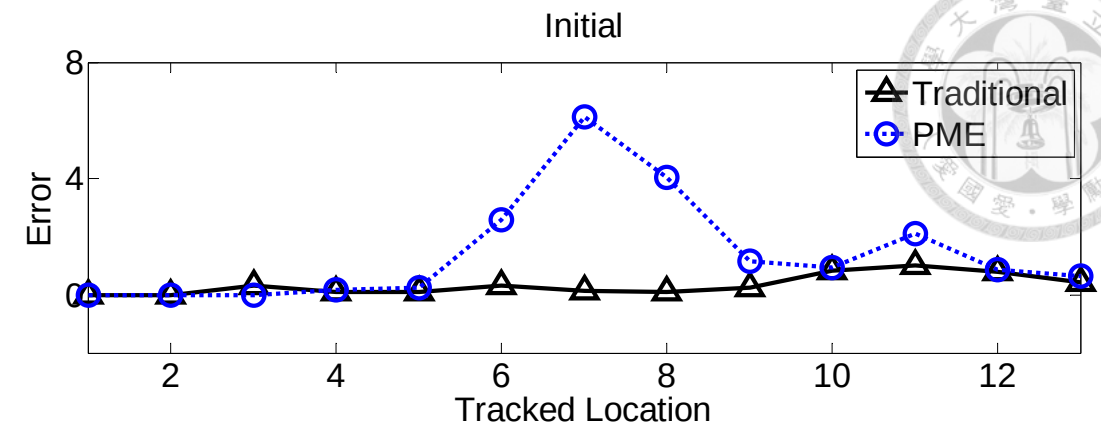
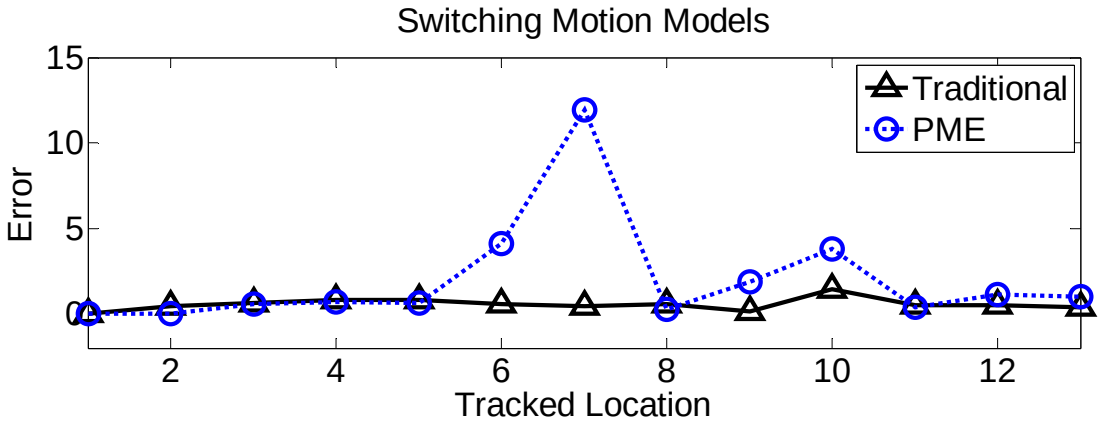


Figure 4.55. Frames per second results of traditional and proposed PME-based keyframe extraction algorithm.

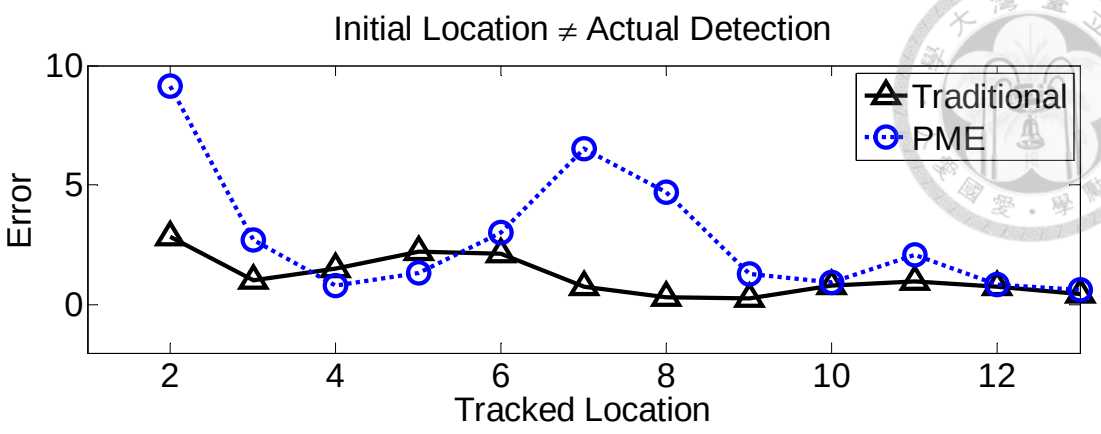


(a)

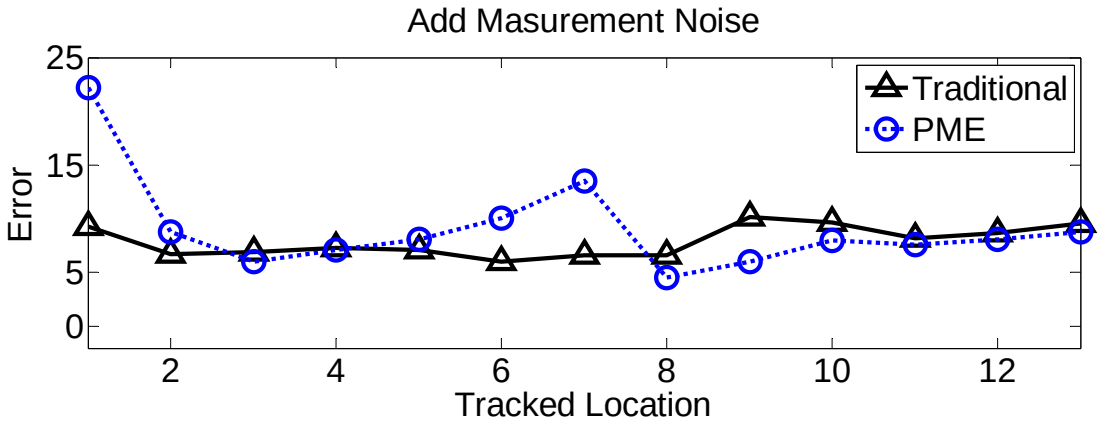


(b)

Figure 4.56. Tracking error of two different tested scenarios of SingleBall. (a) Initial (b) Switch constant acceleration to constant velocity.



(a)



(b)

Figure 4.57. Tracking error of two different tested scenarios of SingleBall. (a) Initial location is randomly given (b) Measurement noise is added.

The performance index is defined as follows:

$$\|e\|_2 = \sum_i \|e_i\|_2 \text{ , where } e_i = \left(e_{i,x}, e_{i,y}\right) \tag{4.14}$$

Tracking error is computed in each scenario as shown in Table 4.14. The proposed PME-based strategy only samples 27 keyframes from 45 frames to achieve expected and satisfactory performance and has the 60% utilization ratio. The absolute value of energy error ratio is almost 1%.

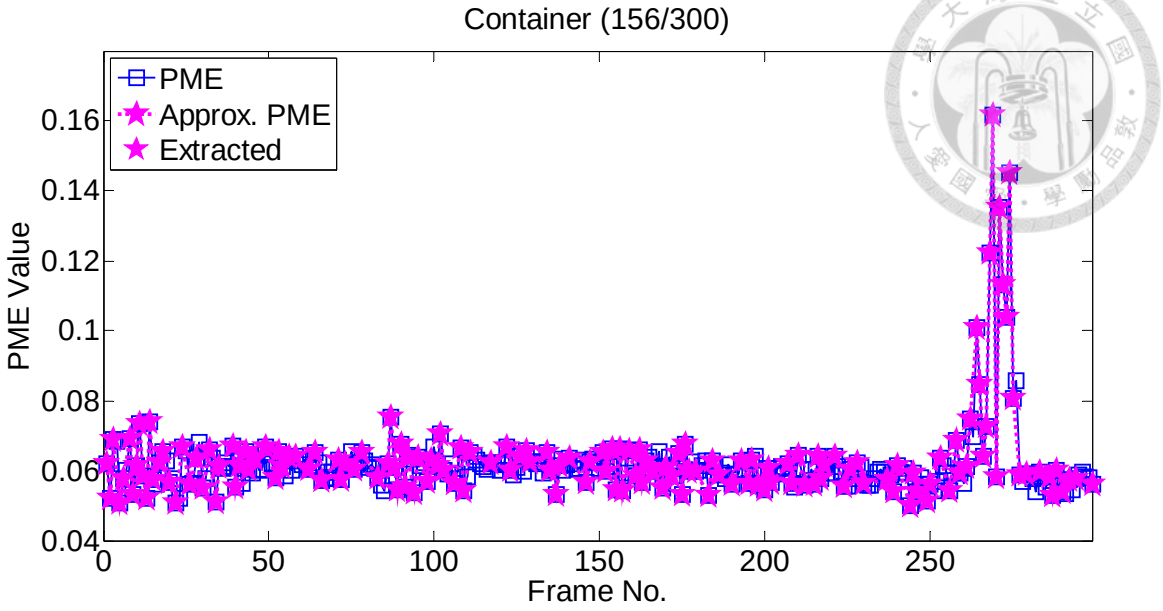
Table 4.14. Results of tracking errror, utilization ratio, energy error ratio and number of extracted keyframe of SingleBall.

<table>
<tr><td colspan="4">Tracking error of SingleBall Tracking
$\|e\|_2 = \sum_i \|e_i\|_2 ; e_i = (e_{i,x}, e_{i,y})$</td><td></td><td></td><td></td></tr>
<tr><td colspan="2">Traditional</td><td colspan="2">PME-based</td><td>Utilization Ratio</td><td>Energy Error Ratio</td><td>#Extracted Frames/ #Total Frames</td></tr>
<tr><td>S1</td><td>3.9</td><td>S1</td><td>18.91</td><td rowspan="4">60%</td><td rowspan="4">0.8%</td><td rowspan="4">27/45</td></tr>
<tr><td>S2</td><td>8.4</td><td>S2</td><td>26.41</td></tr>
<tr><td>S3</td><td>22.1</td><td>S3</td><td>34.11</td></tr>
<tr><td>S4</td><td>109.9</td><td>S4</td><td>118.97</td></tr>
<tr><td colspan="7">S1: Initial; S2: Switching Motion Models;
S3: Initial Location ≠ Actual Detection; S4: Add Measurement noise.</td></tr>
</table>

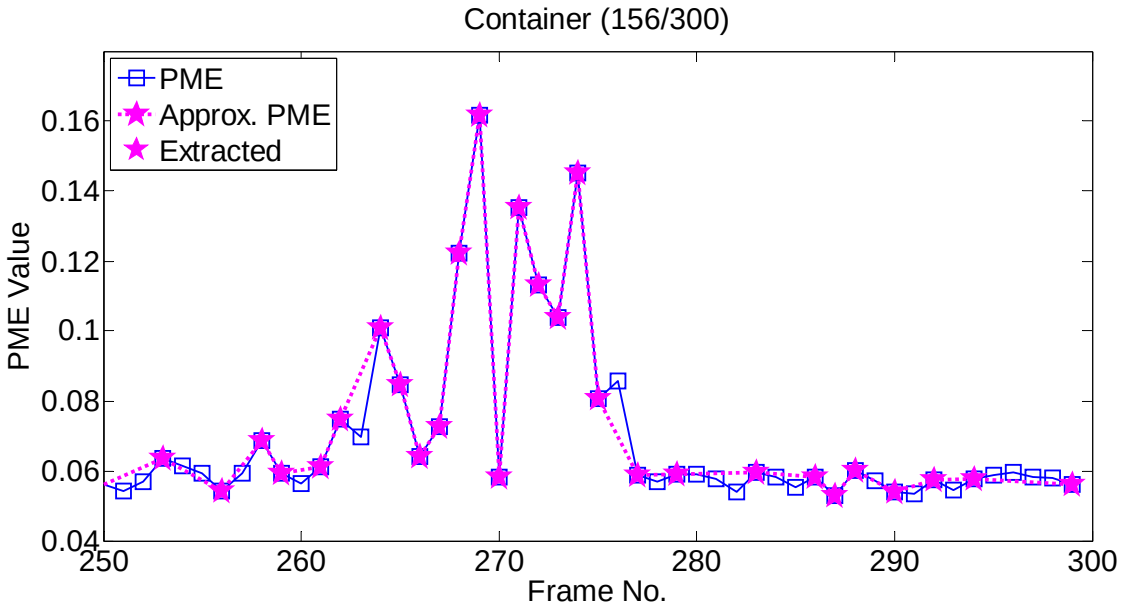
The second tested video is Container that has one cargo ship slowly departing from port and another boat is chasing it and camera is stationary. However, wave motion from the ocean produces unexpected environment disturbance and that might affect the results of multi-objects detection. Figure 4.58 shows five snapshots of Container and Figure 4.59 presents the motion energy curve produced from perceived motion energy of each captured frame and the extracted frames and the frame no. of extracted keyframe. Before Frame 250, the cargo ship leaves slowly and boat motion seems slow due to the visual sensor is far, hence average PME value stays around 0.06. From Frame 250, two birds suddenly fly into the scene to cause great motion energy around the birds

as shown in Figure 4.58. Hence, Figure 4.59(a) has a high motion energy peak in the last. Figures 4.59(b) shows the motion energy curve from Frames 250 to 300 to present that the approximation motion energy curve is close to original motion energy curve by only using 156 of 300 frames. The energy error of each frame is computed by Equation (4.10) as shown in Figure 4.60. Figure 4.61 shows the frames per second (fps) results of traditional and proposed keyframe PME-based extraction algorithm of Container. The perceived motion energy value has small changing from Frames 1 to 250 (from first second to tenth second) as shown in Figure 4.59(a). Hence the frame rate stays below 15. At eleventh second, birds fly into scene, thus the frame rate increases over 15 to catch the motion changing in the environment. The flexibility of PME-based sampling strategy in small perceived motion energy value situation can be seen in Figure 4.61.

Figure 4.58. The snapshots of Container.



(a)



(b)

Figure 4.59. (a) The motion energy curve, the approximation motion energy curve and the frame no. of extracted keyframes (b) The motion energy curve produced by captured frames and extracted keyframes from Frames 250 to 300.

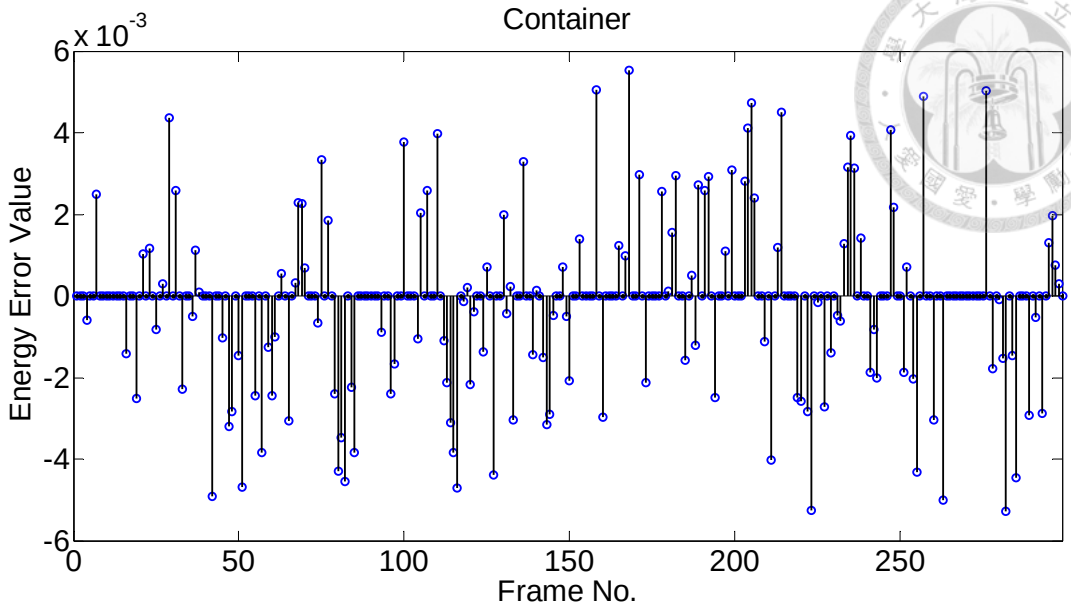


Figure 4.60. The energy error of each frame.

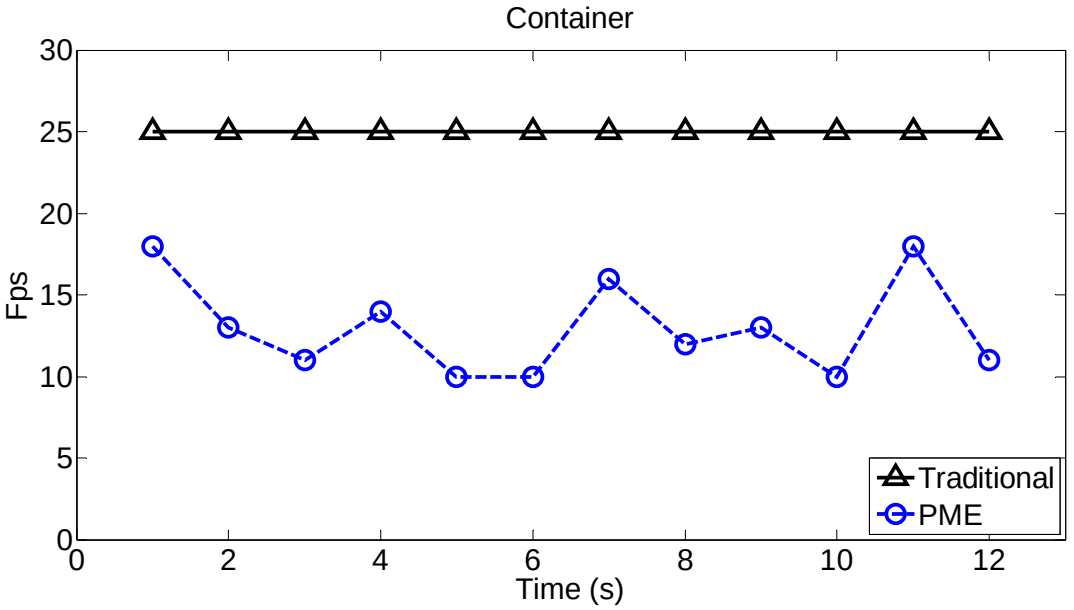
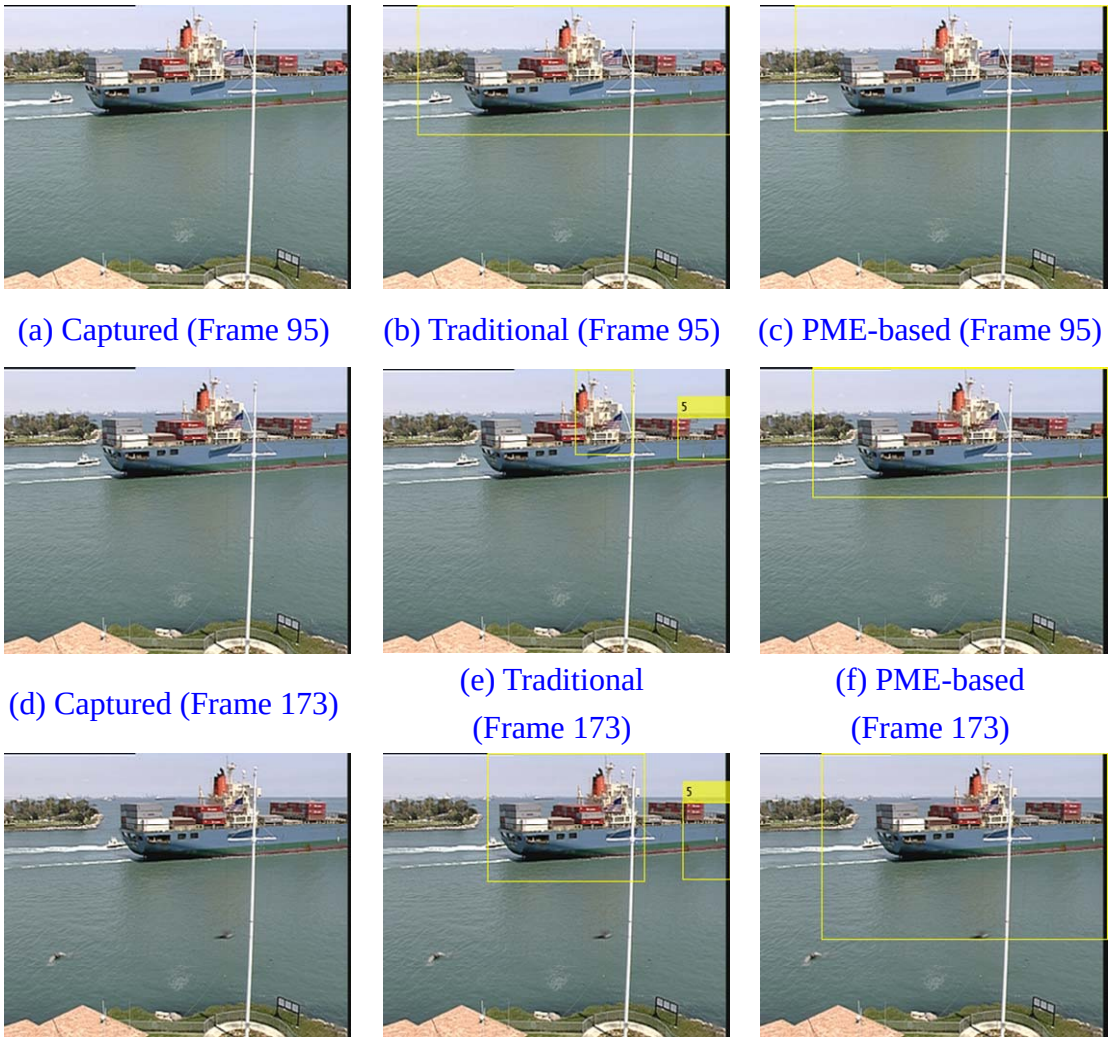


Figure 4.61. Frames per second results of traditional and proposed PME-based keyframe extraction algorithm in Container.

.

Figure 4.62 presents the multiple objects tracking results of Frames 95, 173 and 257 by using two different sampling strategies, namely, traditional and PME-based.

Figures 4.62(a)(d)(g) show the snapshots of captured Frames 95, 173 and 257 and Figures 4.62(b)(e)(h) and Figures 4.62(c)(f)(i) present the tracking results of using traditional and PME-based manner. Moreover, the results show that PME-based sample methodis better than traditional way.



(a) Captured (Frame 95) (b) Traditional (Frame 95) (c) PME-based (Frame 95)

(d) Captured (Frame 173) (e) Traditional (Frame 173) (f) PME-based (Frame 173)

(g) Captured Frame 257 (h) Traditional (Frame 257) (i) PME-based (Frame 257)

Figure 4.62. Moving objects tracking results of Container (a)(d)(e) Captured Frames 95, 173 and 257 (b)(e)(h) Tracking results of Frames 95, 173 and 257 with Traditional way (c)(f)(i) Tracking results of Frames 95, 173 and 257 with PME-based keyframe extraction algorithm .

Accuracy ratio of multi-object tracking results of traditional and PME-based extraction are both 100%. Moreover, Table 4.15 shows that PME-based strategy only samples 156 keyframes from 300 frames to achieve the high accuracy ratio. The utilization ratio is 52% and absolute value of energy error ratio is only 0.3%.

Table 4.15. Results of multi-objects tracking, utilization ratio, energy error ratio and number of extracted keyframe of Container.

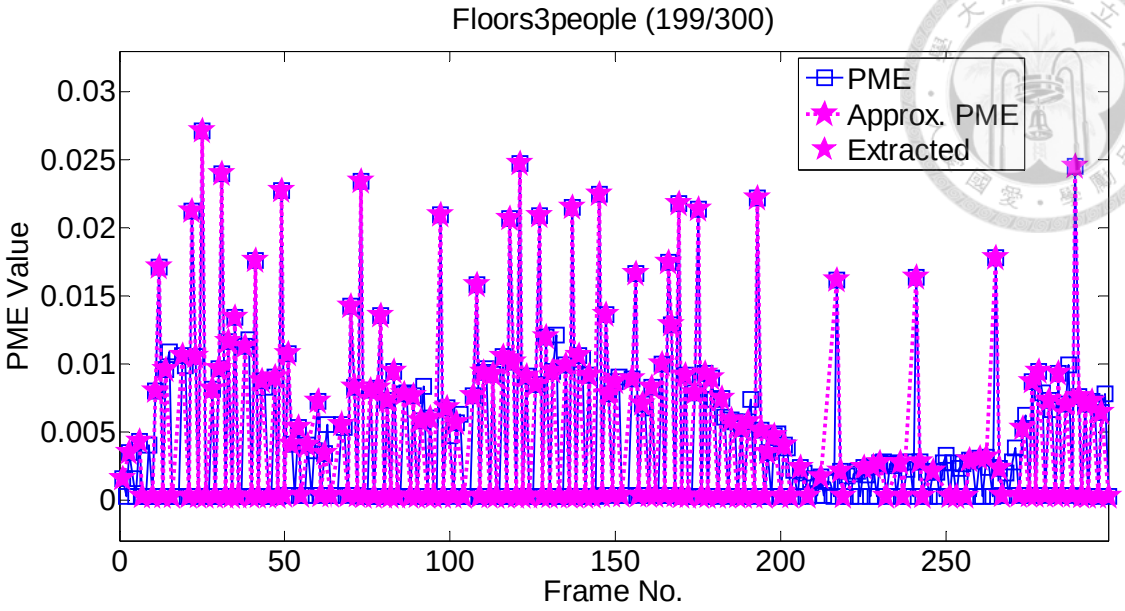
Accuracy Percentage of Multi-Objects Tracking				
Traditional	PME-based	Utilization Ratio	Energy Error Ratio	#Extracted Frames/ #Total Frames
100%	100%	52%	-0.26%	156/300

B. Object motion direction: Multiple

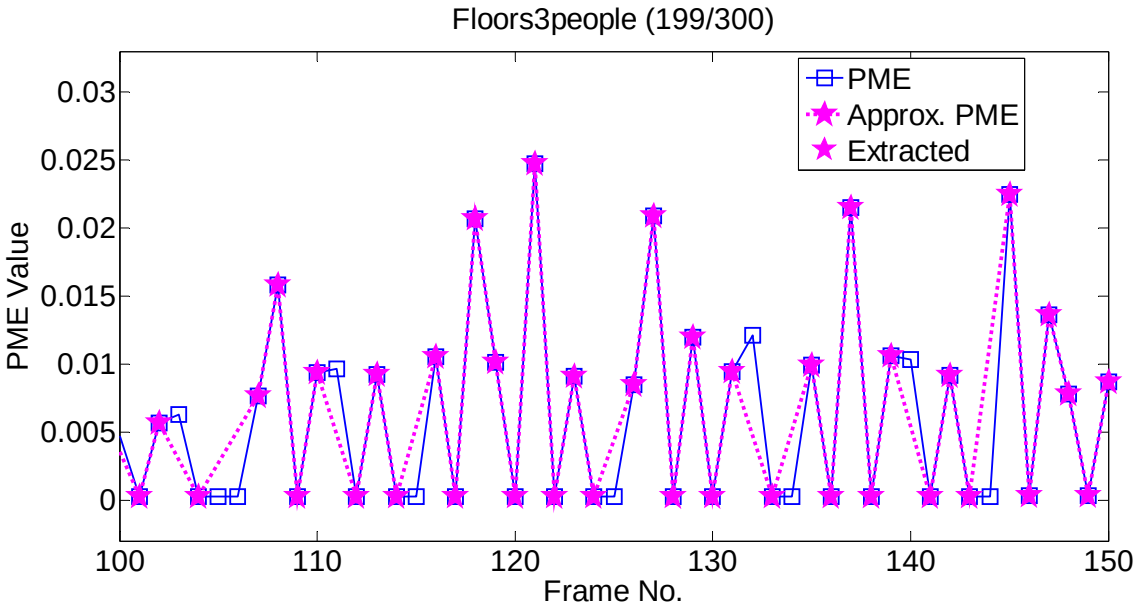
Three tested videos, namely, Floors3people, Hall and Traffic, are used to perform the keyframe extraction process as shown in Figure 4.11. The application task is moving people tracking for Floors3people and Hall and is car detection and counting for Traffic. Three different sampling strategies: traditional, PME-based with threshold $AT = \lambda_{ctrl}^{scene} \times IQR$ where $\lambda_{ctrl}^{scene} = \lambda_{ctrl}^{Floors} = 0.5$, $IQR = 0.007$ and $\lambda_{ctrl}^{scene} = \lambda_{ctrl}^{Hall} = 0.5$, $IQR = 0.11$ and $Fixed(\alpha, \beta)$ where α=1 and $\beta = 3$. For each video, Equations (4.3), (4.6)-(4.7) and (4.9)-(4.10) are performed to obtain the motion energy curve, the extracted keyframe no., the utilization ratio and the energy error ratio.

The first experimented video is Floors3people which has three walking people in three different floors captured by the stationary visual sensor. Figure 4.63 shows five snapshots of Floors3people and Figure 4.64(a) presents the motion energy curve produced from perceived motion energy of each captured frame and the extracted frames and the frame no. of extracted keyframe. Because the distance between moving people and camera is long, the motion energy caused by moving people is little. Moreover, background has no contribution to motion energy because the visual sensor is stationary. Hence, perceived motion energy value is smaller than other tested videos. Figure 4.64(b) shows the motion energy curve from Frames 100 to 150 to present that the approximation motion energy curve is close to original motion energy curve by only using 199 of 300 frames. The energy error of each frame is computed by Equation (4.10) as shown in Figure 4.65.

Figure 4.63. The snapshots of Floors3people.



(a)



(b)

Figure 4.64. (a) The motion energy curve, the approximation motion energy curve and the frame no. of extracted keyframes (b) The motion energy curve produced by captured frames and extracted keyframes from Frames 100 to 150.

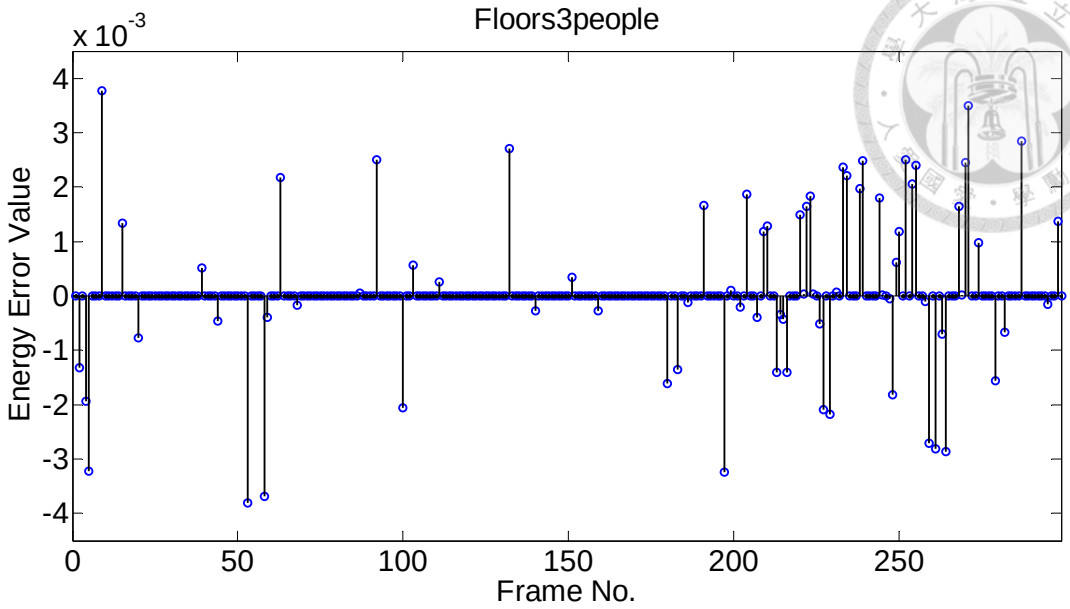


Figure 4.65. The energy error of each frame.

Figure 4.66 shows the frames per second (fps) results of traditional, $Fixed(1,3)$ proposed PME-based keyframe extraction algorithm of Floors3people. $Fixed(1,3)$ manner select one frame from three continuous frames because the moving people detection process does not determine any moving ones in a low motion energy value circumstance. Hence, frames per second value in the $Fixed(1,3)$ is eight in all duration. The applicability of the proposed PME-based sampling method can be seen in Figure 4.66.

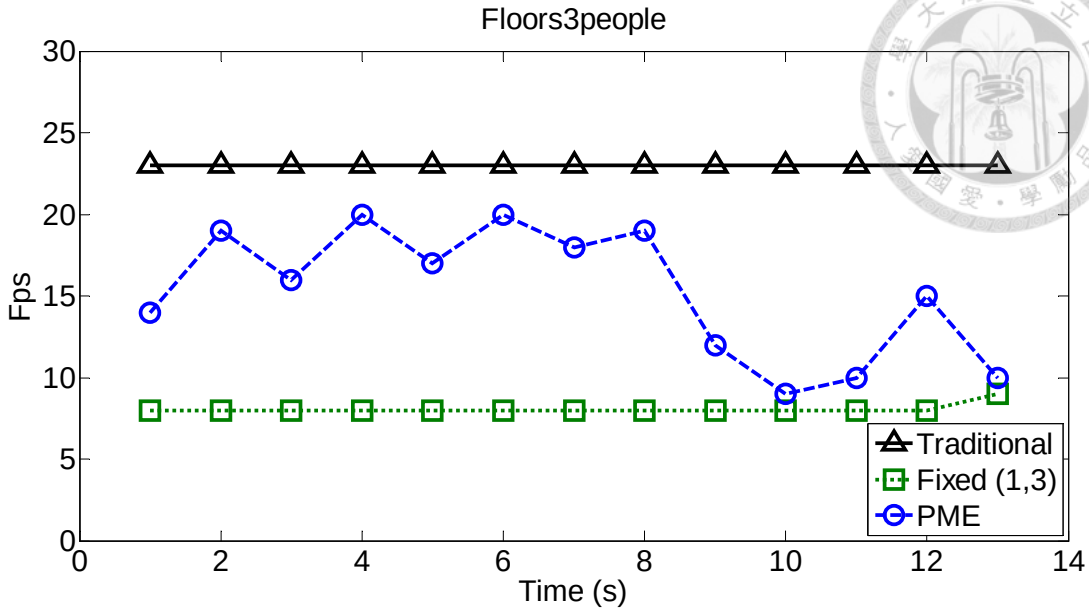


Figure 4.66. Frames per second results of traditional, *Fixed*(1,3) and proposed PME-based keyframe extraction algorithm.

Table 4.16 shows the results of average available bandwidth to one frame based on frames per second as shown in Figure 4.66. For traditional and PME-based sample method, average frame rate is twenty-three and fifteen, respectively. The proposed PME-based method only samples 199 keyframes from 300 frames and has the 66.33% utilization ratio. The absolute value of energy error ratio is 0.72%.

Table 4.16. Results of average available rate to each frame, energy error ratio and number of extracted keyfram of Floors3people.

Average available Rate for each frame				
Traditional	PME-based	Utilization Ratio	Energy Error Ratio	#Extracted Frames/ #Total Frames
$B_{Net}/23$	$B_{Net}/15$	66.33%	0.72%	199/300

Figure 4.67 shows that three labels achieve desired video quality PNSR =30 by using traditional, $Fixed(1,3)$ and PME-based manners in all phases under condition of bandwidth 4.5Mbps. Table 4.17 shows the quality ratio results of three phases are computed by Equations (4.11)-(4.12) and quality ratio achieve 100% in all phases. Next, Figure 4.68 shows the saving ratio of bandwidth with using traditional, $Fixed(1,3)$ and PME-based manners under the condition of 3.2 Mbps. $Fixed(1,3)$ gets average saving results 62% higher than PME-based 47% as presented in Table 4.18.

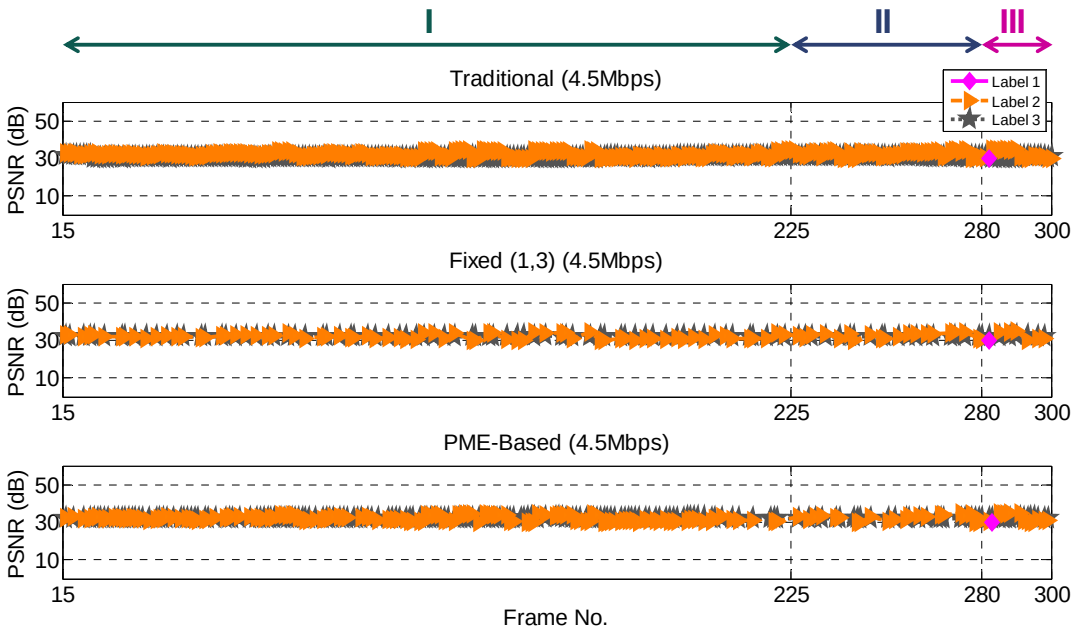


Figure 4.67. The quality of decoded frames with $\lambda_{PSNR}^{scene} = 30$ in 4.5Mbps. (a) Traditional (b) $Fixed(1,3)$ (c) PME-based.

Table 4.17. Quality ratio of three phases (4.5 Mbps) in three different sampling strategies: Traditional, $Fixed(1,3)$ and PME-based.

Phase	I	II	III
Traditional	100%	100%	100%
Fixed Sampling	100%	100%	100%
PME-based	100%	100%	100%

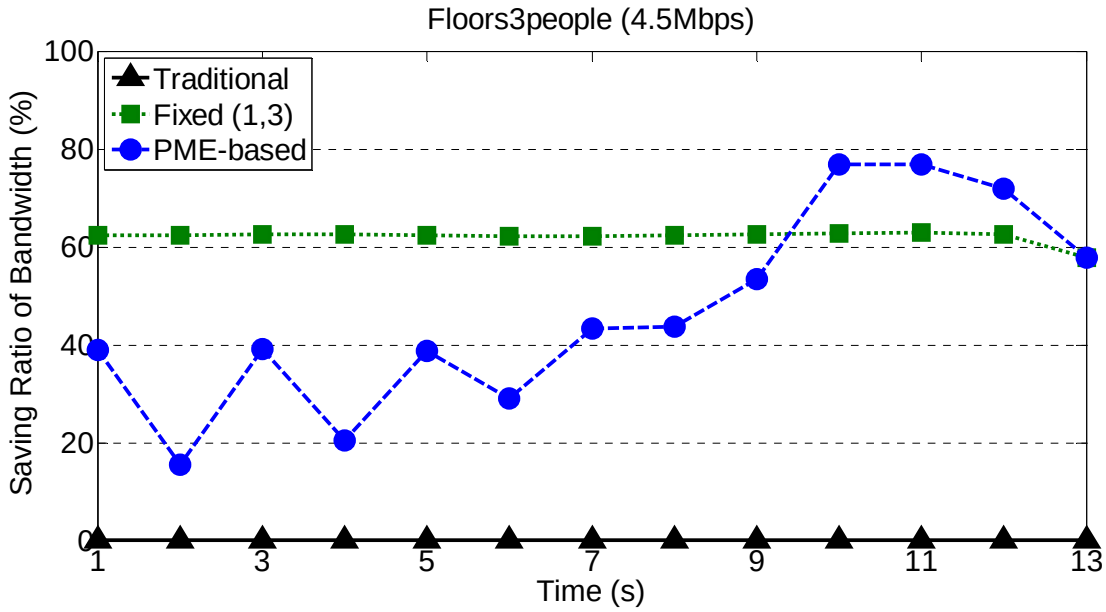


Figure 4.68. Saving percentage of bandwidth compared with that of the lossless case.

Table 4.18. Average Saving Ratio of Bandwidth of three different sampling strategies (4.5Mbps).

	Traditional	$Fixed(1,3)$	PME-based
Average Saving Ratio	0.04%	62%	47%

Although $Fixed(1,3)$ sampling strategy gets 62% saving ratio of bandwidth, what would happen in system performance. On the other hand, the proposed PME-based

sampling strategy not only has about 50% saving ratio but also considers the motion characters in environment to maintain tracking results. The simple expression of Floors3people scenario is depicted in Figure 4.69. Next, moving object detection and location estimation algorithm is performed to get the coordinates of moving objects using two different extracted keyframes sets from $Fixed(1,3)$ and the proposed PME-based. For measuring the system performance in two sampling strategies: $Fixed(1,3)$ and PME-based, the performance index is defined as follows:

$$Error_{(x,y),SamplingType} = \left\| Traditional_{(x,y)} - SamplingType_{(x,y)} \right\| \quad (4.15)$$

where $SamplingType = \{Fixed\ (1,3), \text{PME-based}\}$

The performance index is designed to present which sampling manner loses too much motion energy and further degrades system performance.

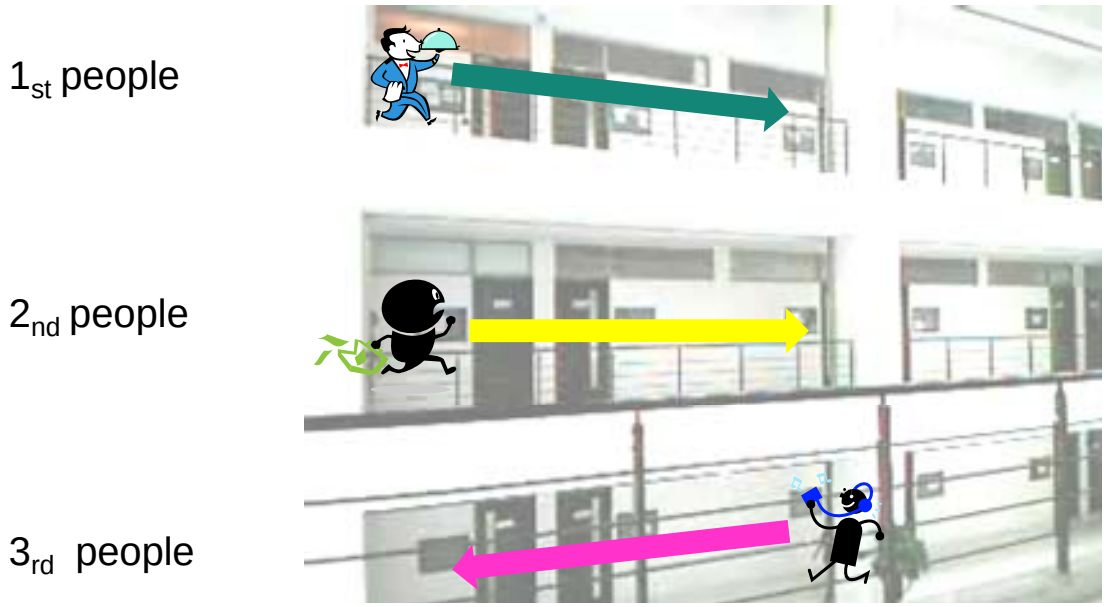
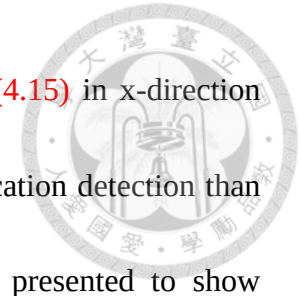
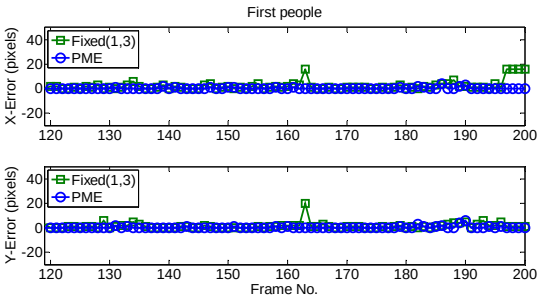
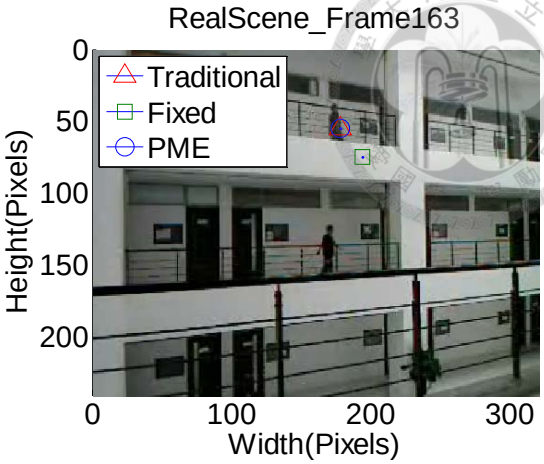


Figure 4.69. The scenario expression of Floors3peoplevideo.

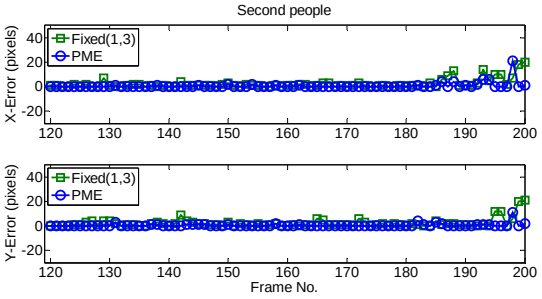


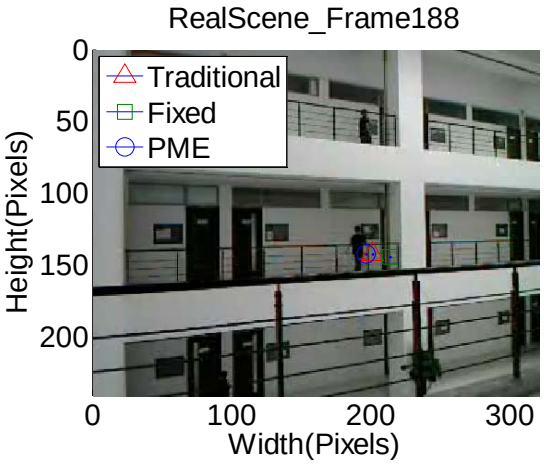
Figures 4.70(a)(c)(e) show the error computed by Equation (4.15) in x-direction and y-direction. The proposed PME-based gets better results in location detection than $Fixed(1,3)$. Moreover, the particular frames with large error are presented to show location detection results in the real scene in different sampling strategies: traditional, $Fixed(1,3)$ and PME-based as shown in Figures 4.70(b)(d)(f). Although $Fixed(1,3)$ sampling strategy gets 62% saving ratio of bandwidth, it has bad location detection results. Hence, the proposed PME-based not only has high efficiency in saving bandwidth but also maintain system performance.



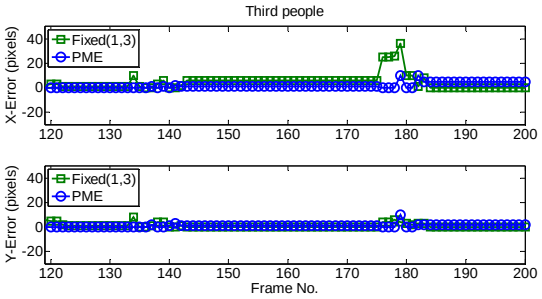


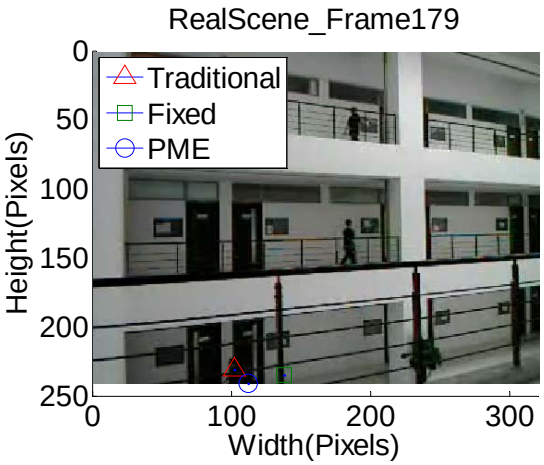
(a) (b)





(c) (d)





(e) (f)

Figure 4.70. Results of location detection to three moving people. (a)(c)(e) The error of X and Y direction (b)(d)(e) The detected results in the real scene of particular frame.

The second tested video is Hall that has two men walks toward different directions in an office environment and the visual sensor is stationary. Figure 4.71 shows five snapshots of Hall. Because the captured frame rate is higher than motion in the scene, the length of motion vectors is short. Hence, the perceived motion energy value of Hall is small as shown in Figure 4.72(a). Figure 4.72(b) shows the motion energy curve from Frames 150 to 200 to present that the approximation motion energy curve is close to original motion energy curve by only using 182 of 300 frames. The energy error of each frame is computed by Equation (4.10) as shown in Figure 4.73.

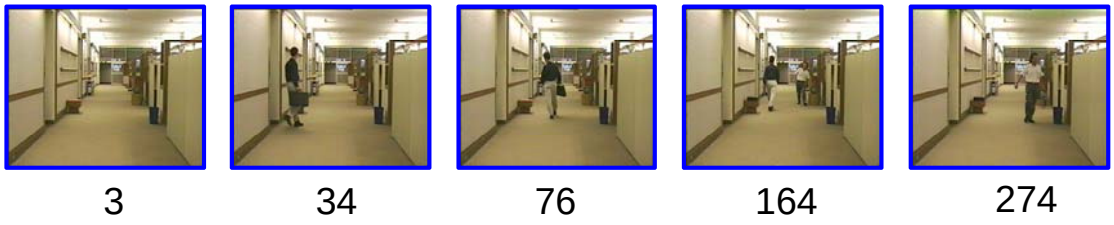
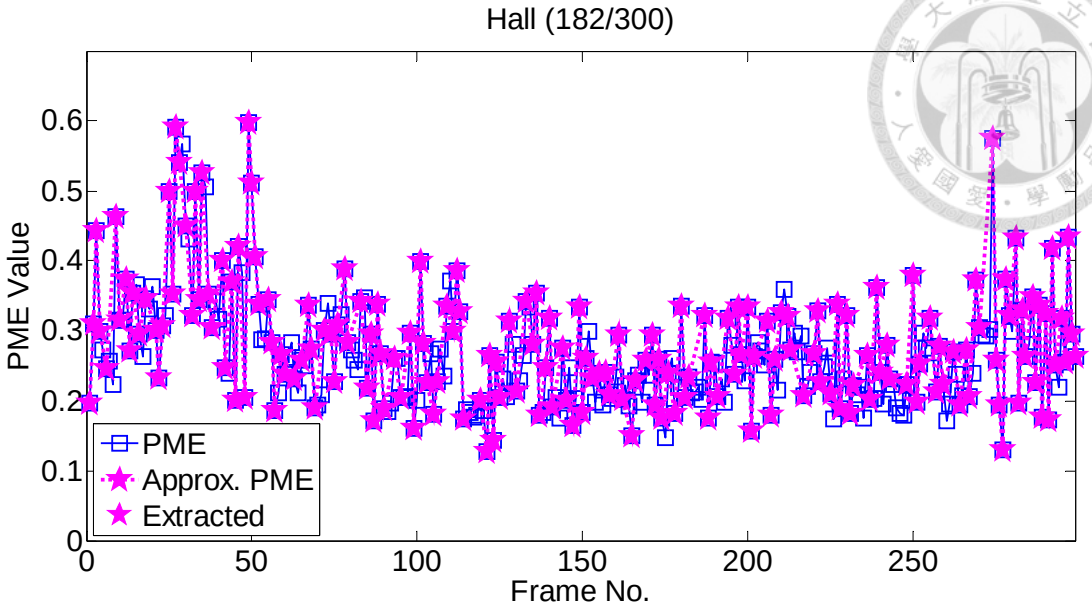
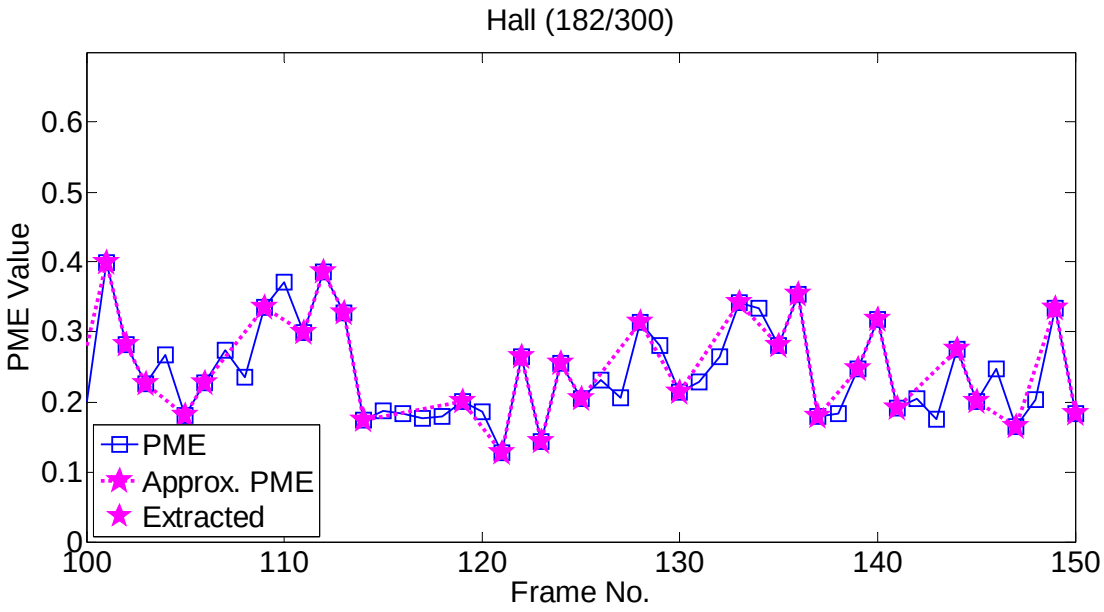


Figure 4.71. The snapshots of Hall.



(a)



(b)

Figure 4.72. (a) The motion energy curve, the approximation motion energy curve and the frame no. of extracted keyframes (b) The motion energy curve produced by captured frames and extracted keyframes from Frames 100 to 150.

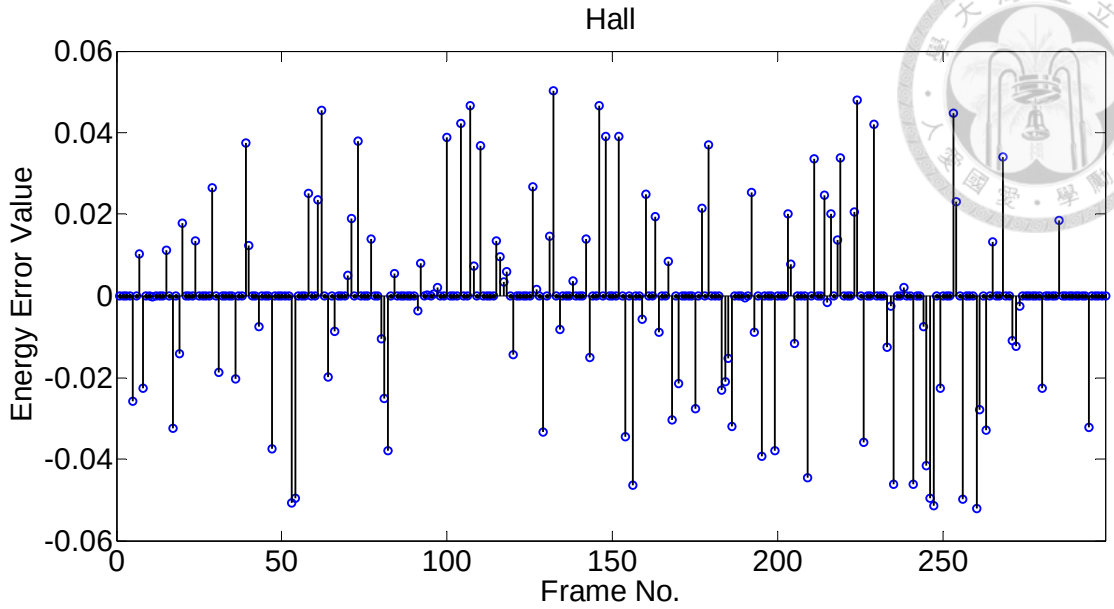


Figure 4.73. The energy error of each frame.

Figure 4.74 shows that the frames per second (fps) results of traditional and proposed PME-based keyframe extraction algorithm. From first second to eleventh second (from Frames 1 to 275 as shown in Figure 4.72(a)), motion dynamic in video is variation but still has average value around about 0.3. It means the environment stays in the steady state which no large motion occurs suddenly. Hence, frame rate value should be maintained in a certain value. In the last second, another man suddenly turns his direction into the corner. In order to follow this changing, frame rate value increases. Original frame rate of Hall is 25, using PME-based sampling strategy, it could greatly decrease frame rate to less than 15 to save the computation resource in the steady state.

Figure 4.75 shows the snapshots of multi-objects tracking results of Hall. Figure

4.75(a) presents the good tracking result and Figure 4.75(b) shows the bad tracking result. For presenting the multi-objects tracking results, using (X,Y) denotes the coordinate of left/top of bounding box to represent the location of the detected box as shown in Figure 4.76.

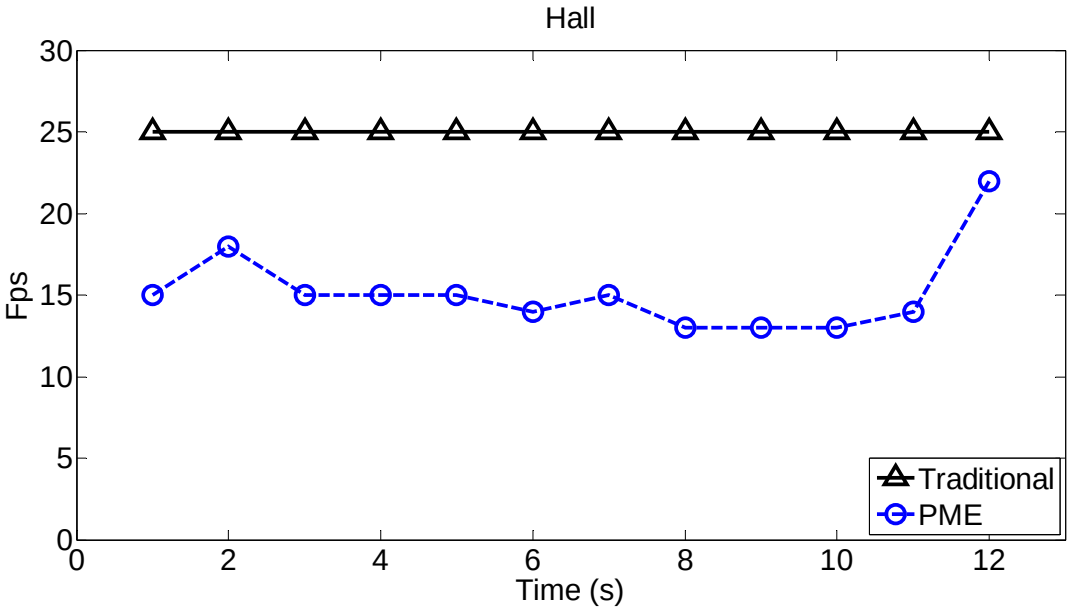


Figure 4.74. Frames per second results of traditional and proposed PME-based keyframe extraction algorithm in Hall video.

Correct

(a)

False

(b)

Figure 4.75. Examples of multi-objects tracking result of Hall. (a) Correct detection result (b) False detection result.

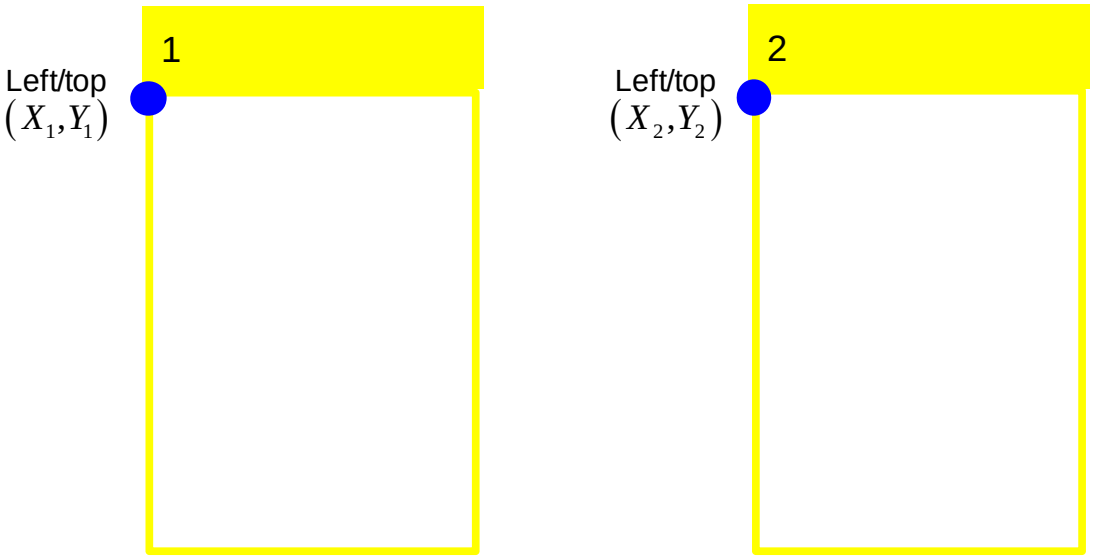
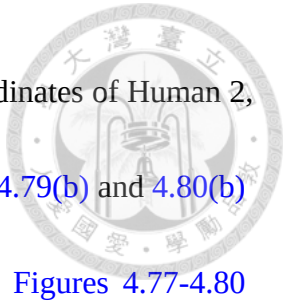
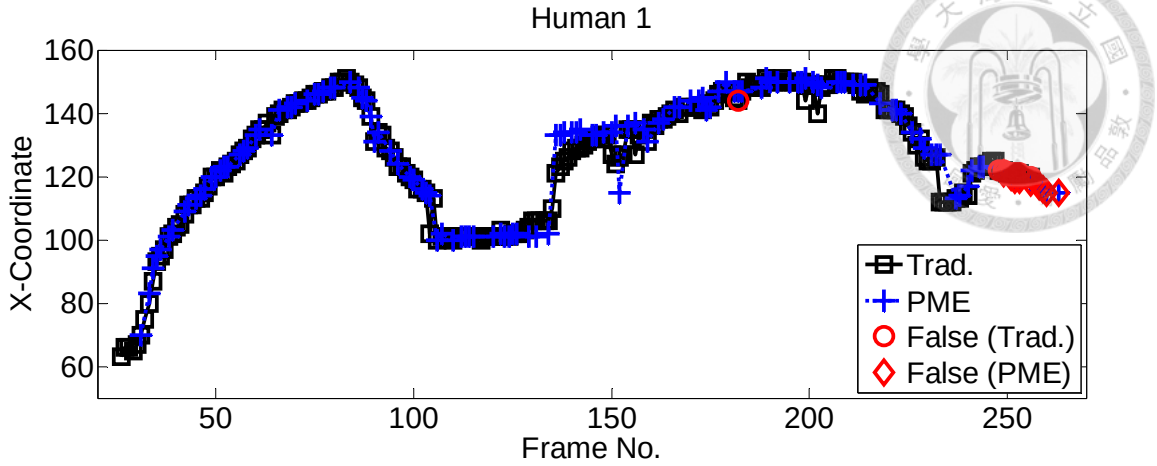


Figure 4.76. The illustrations of bounding boxes and (X_1, Y_1) and (X_2, Y_2) are used to be the tracking index for detected Human 1 and detected Human 2, respectively.

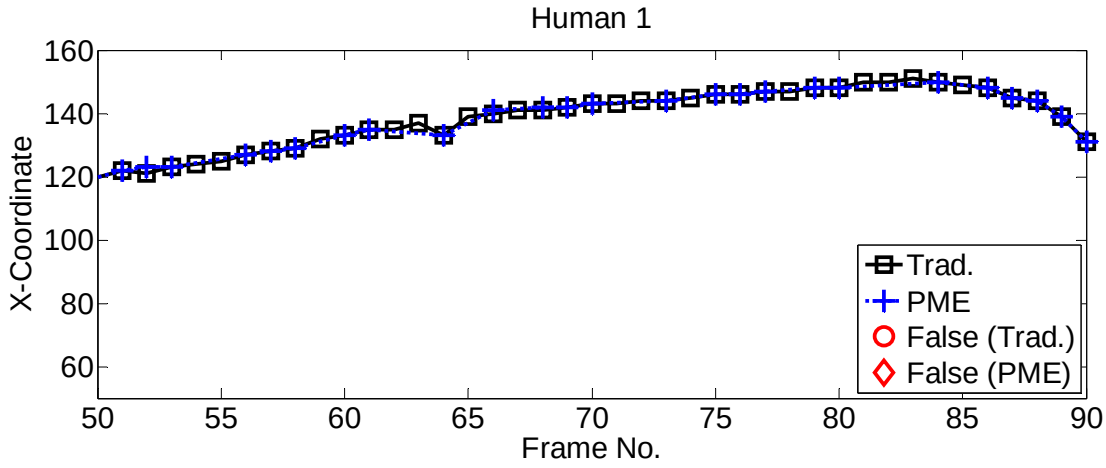
The multi-objects tracking results of PME-based and traditional are shown in Figures 4.77-4.80. Figures 4.77(a) and 4.78(a) show the X and Y coordinates of Human 1, and false markers are based on the definition as shown in Figure 4.76. Figures 4.77(b) and 4.78(b) are the zoom in version of Figures 4.77(a) and 4.78(a),



respectively. Figures 4.79(a) and 4.80(a) show the X and Y coordinates of Human 2, and false markers are based on the definition in Figure 4.76. Figures 4.79(b) and 4.80(b) are the zoom version of Figures 4.79(a) and 4.80(a), respectively. Figures 4.77-4.80 show the powerful of PME-based sampling strategy. The tracking results of Humans by using PME-based manner are as good as the results of traditional. It is clearly to see the tracking results with using PME-based in Figures 4.77(b), 4.78(b), 4.779b) and 4.80(b). The proposed PME-based only uses fewer keyframes to obtain the same tracking trend of traditional.

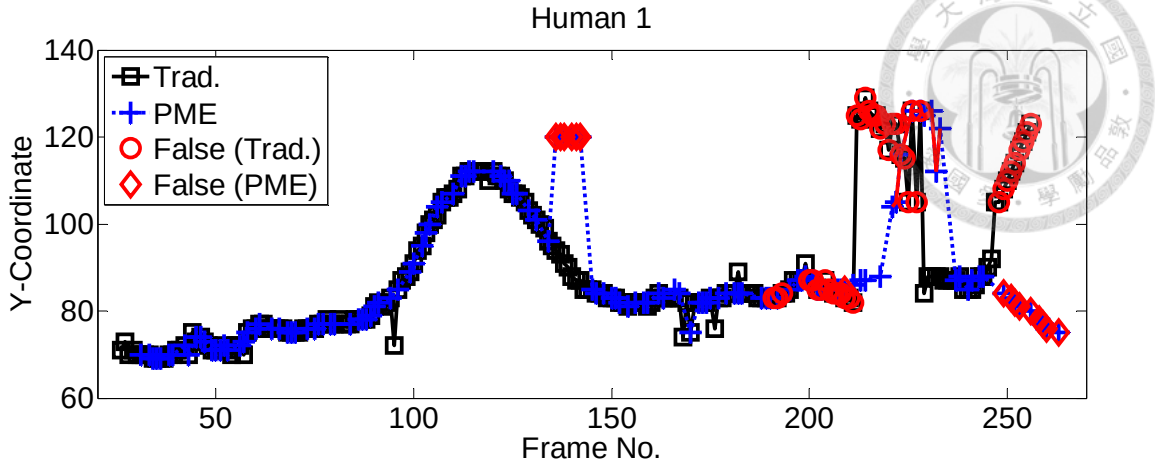


(a)

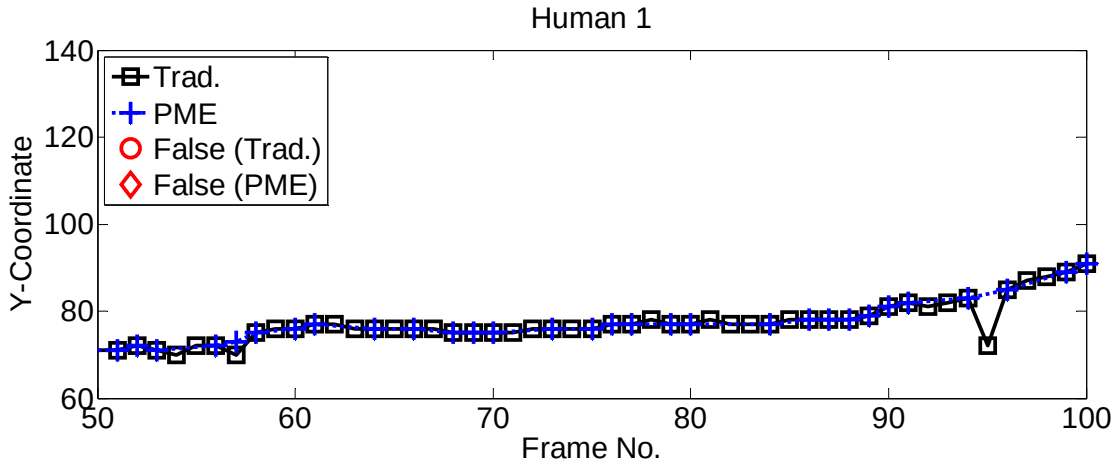


(b)

Figure 4.77. Multi-objects tracking results of Hall with using traditional and PME-based sampling strategies. (a) X coordinate of left/top of bounding for Human 1 (b) The zoom in version from Frames 50 to 90.

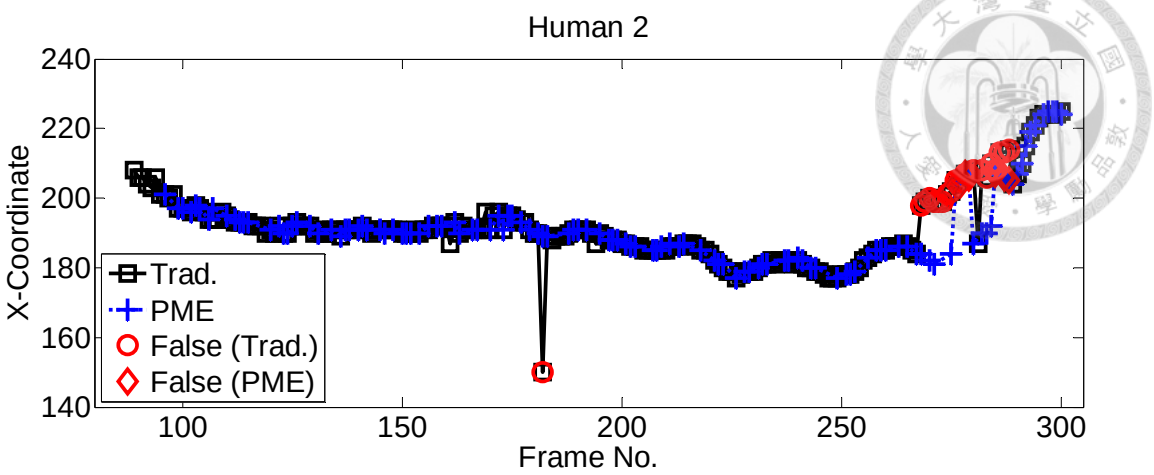


(a)

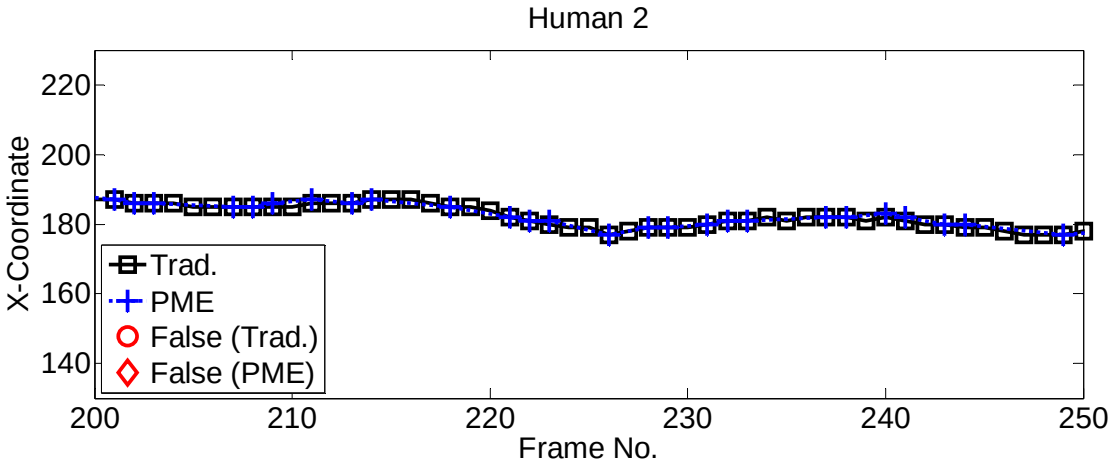


(b)

Figure 4.78. Multi-objects tracking results of Hall in traditional and PME-based sampling strategies. (a) Y coordinate of left/top of bounding for Human 1 (b) The zoom in version of (a) from Frames 50 to 90.

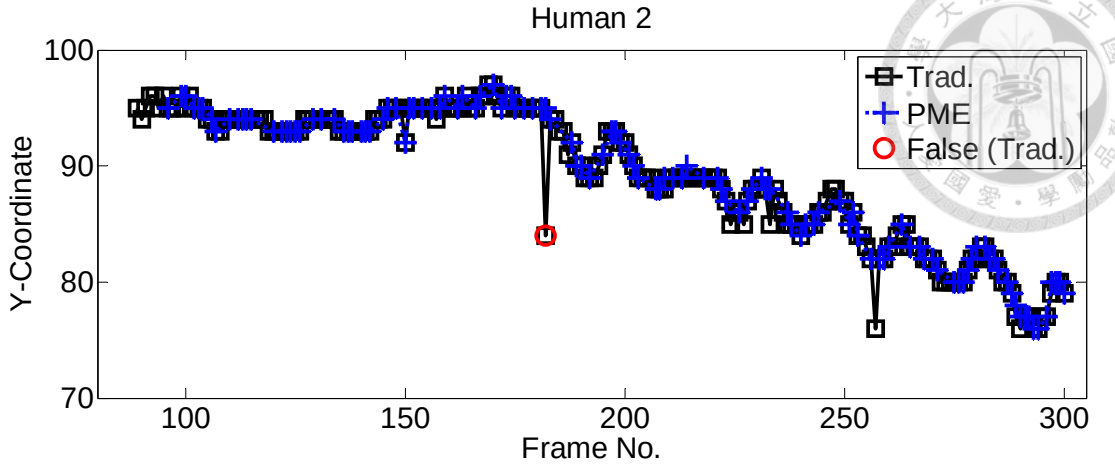


(a)

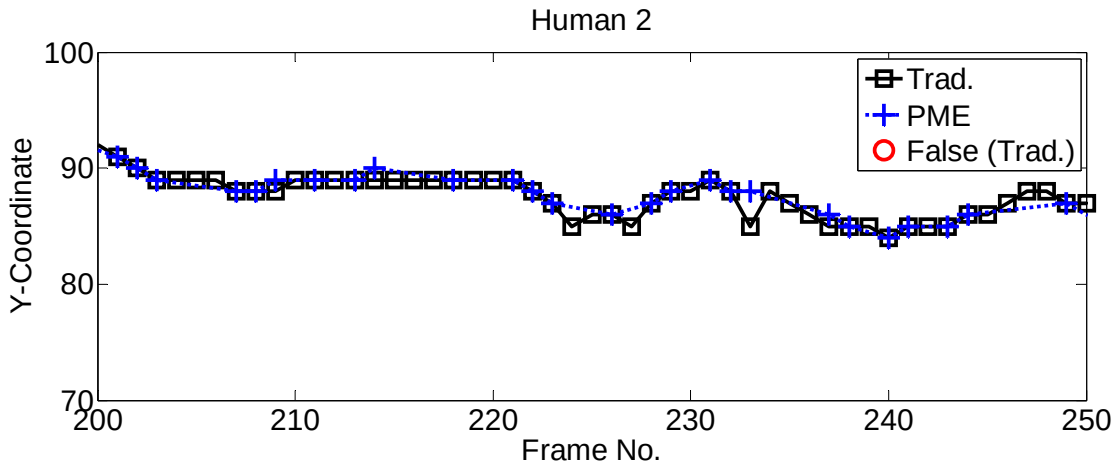


(b)

Figure 4.79. Multi-objects tracking results of Hall in traditional and PME-based sampling strategies. (a) X coordinate of left/top of bounding for Human 2 (b) The zoom in version of (a) from Frames 200 to 250.



(a)



(b)

Figure 4.80. Multi-objects tracking results of Hall in traditional and PME-based sampling strategies. (a) Y coordinate of left/top of bounding for Human 2 (b) The zoom version of (a) from Frames 200 to 250.

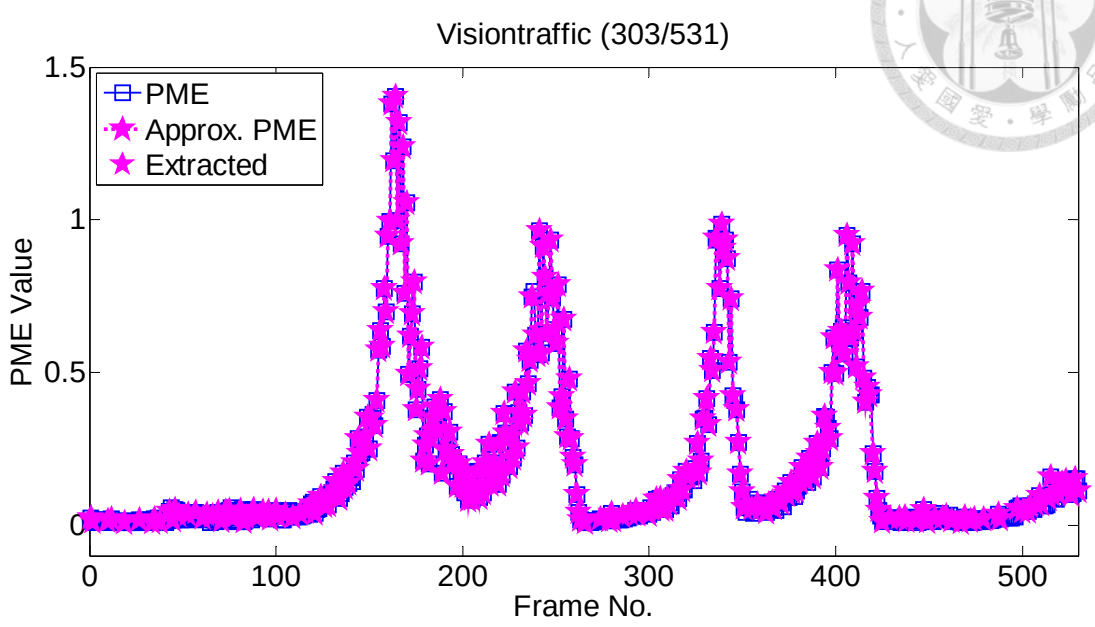
Based on the performance index as shown in Figure 4.76 accuracy ratio of multi-objects tracking results of traditional and PME-based is 79.33% and 80.04%, respectively. Moreover, Table 4.19 shows that PME-based strategy only samples 182 keyframes from 300 frames to achieve 80.04% accuracy ratio better than the traditional and has 60.67% utilization ratio. The absolute value of energy error ratio is 0.2%.

Table 4.19. Results of multi-objects tracking, utilization ratio, energy error ratio and number of extracted keyframe of Hall.

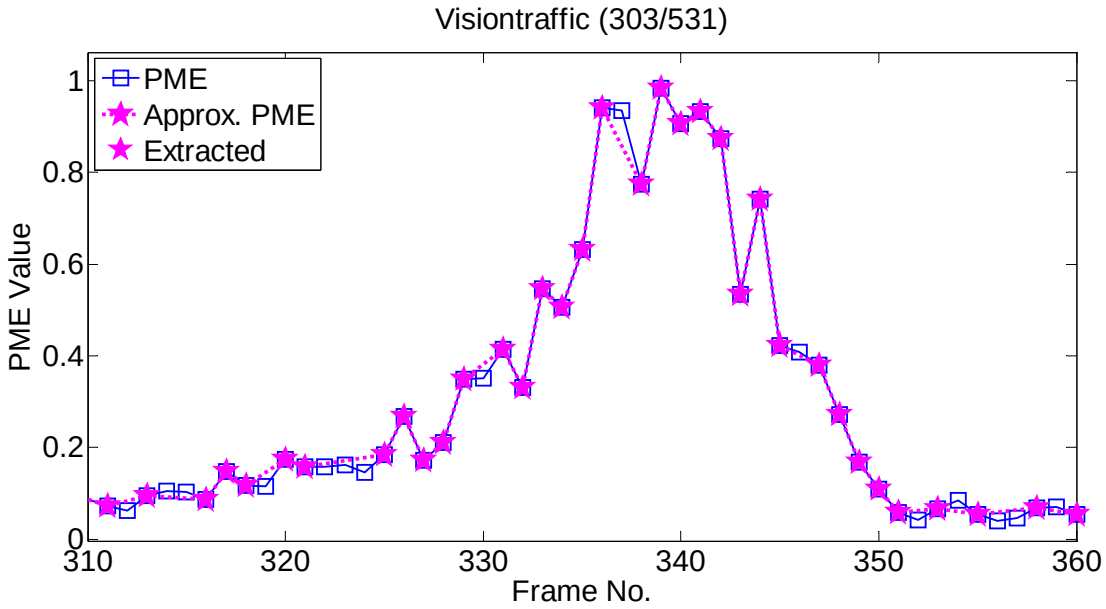
Accuracy Percentage of Multi-Objects Tracking				
Traditional	PME-based	Utilization Ratio	Energy Error Ratio	#Extracted Frames/ #Total Frames
79.33%	80.04%	60.67%	-0.17%	182/300

The third tested video is VisionTraffic that displaces a stationary visual sensor to monitor the traffic. The goal is to detect car and count car number. Figure 4.81 shows five snapshots of VisionTraffic. In VisionTraffic, there are no cars passing before Frame 100 and after Frame 420, thus perceived motion energy value stays near zero in the duration as shown in Figure 4.82(a). Moreover, the peaks represent the status of cars passing through. For example, at first peak, three cars are passing through. Figures 4.82(b) shows the motion energy curve from Frams 310 to 360 to present that the approximation motion energy curve is close to original motion energy curve by only using 303 of 531 frames. Next, the energy error of each frame is computed by Equation (4.10) as shown in Figure 4.83.

Figure 4.81. The snapshots of VisionTraffic.



(a)



(b)

Figure 4.82. (a) The motion energy curve, the approximation motion energy curve and the frame no. of extracted keyframes (b) The motion energy curve produced by captured frames and extracted keyframes from Frames 310 to 360.

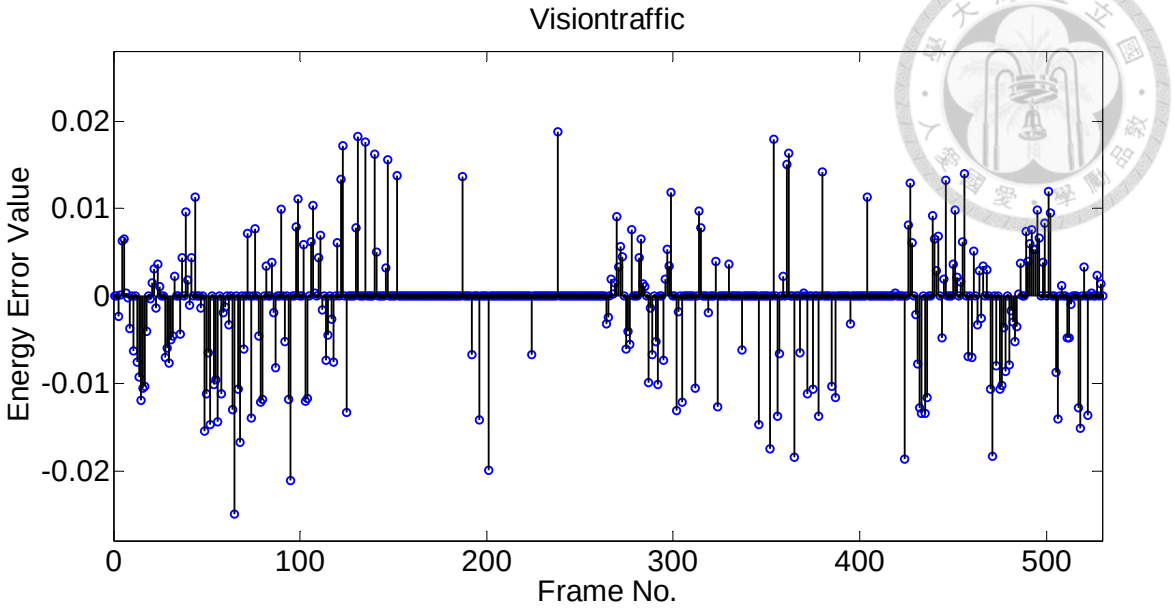


Figure 4.83. The energy error of each frame.

Figure 4.84 shows the frames per second (fps) results of traditional and proposed PME-based keyframe extraction algorithm of VisionTraffic. From fourth second to seventh second (from Frames 120 to 210) corresponds to first peak occurring as shown in Figure 4.82(a). Moreover, frame rate results from seventh second to ninth second, from eleventh second to thirteenth second and from thirteenth second to fifteenth second, they all correspond to the time of happening at second, third and fourth peak. It shows powerful ability of adapting to the motion evolution in the monitored environment of PME-based sampling strategy.

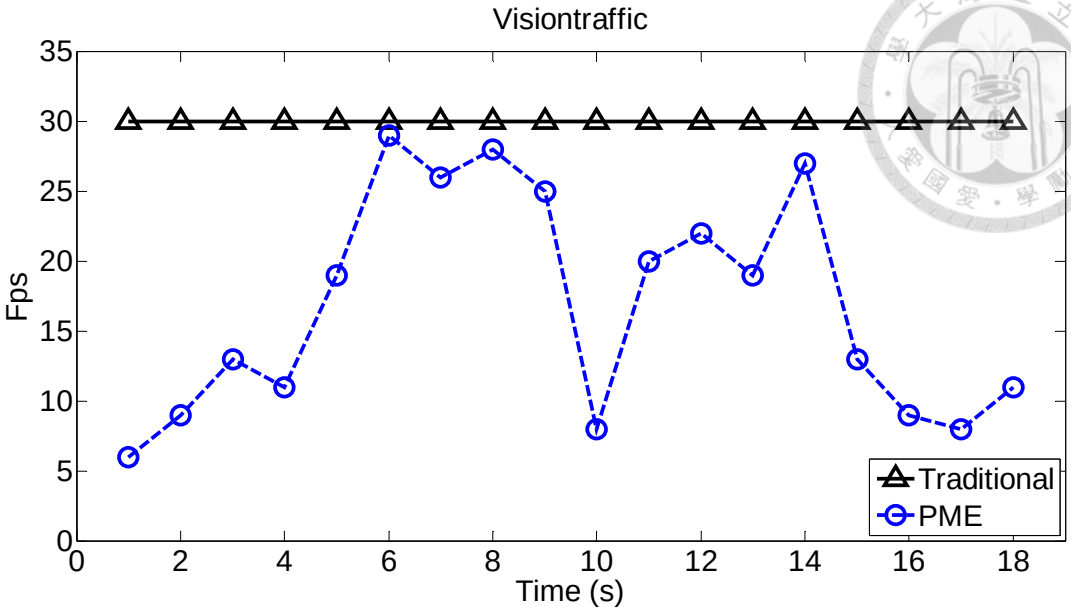
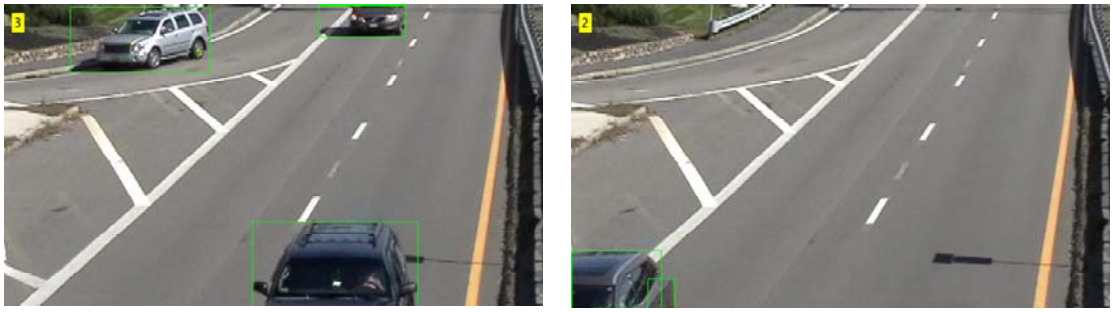


Figure 4.84. Frames per second results of Visiontraffic in traditional and the proposed PME-based keyframe extraction algorithm.

Figure 4.85 shows the car detection and counting results of VisionTraffic. Figure 4.85(a) presents the correct detection and counting results and Figure 4.85(b) shows the false results.



Correct

False

(a)

(b)

Figure 4.85. Performance index of face tracking results of VisionTraffic. (a) Correct detection result (b) False detection result.

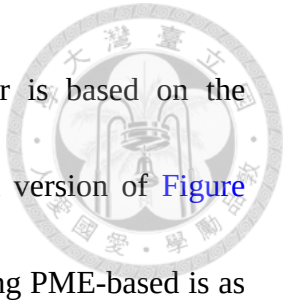
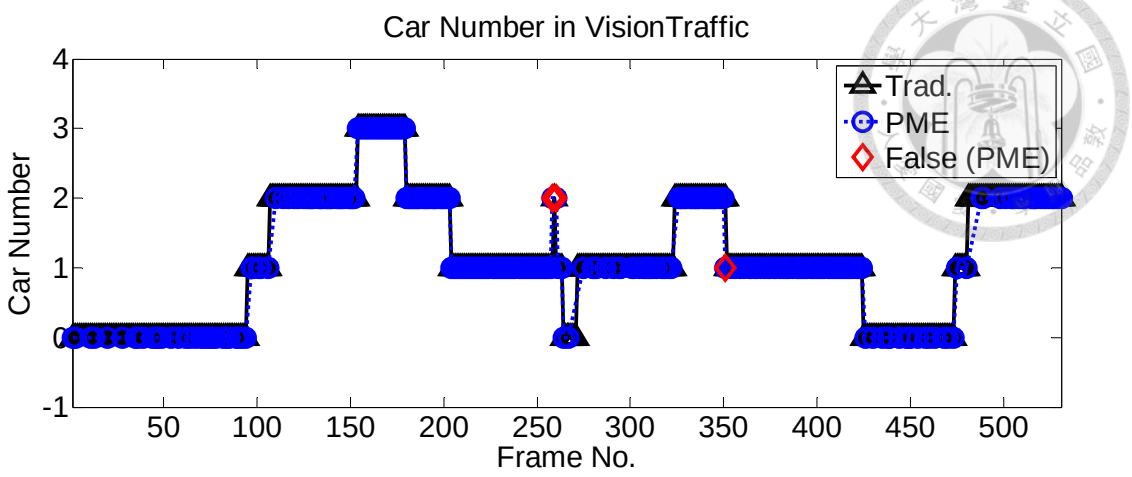


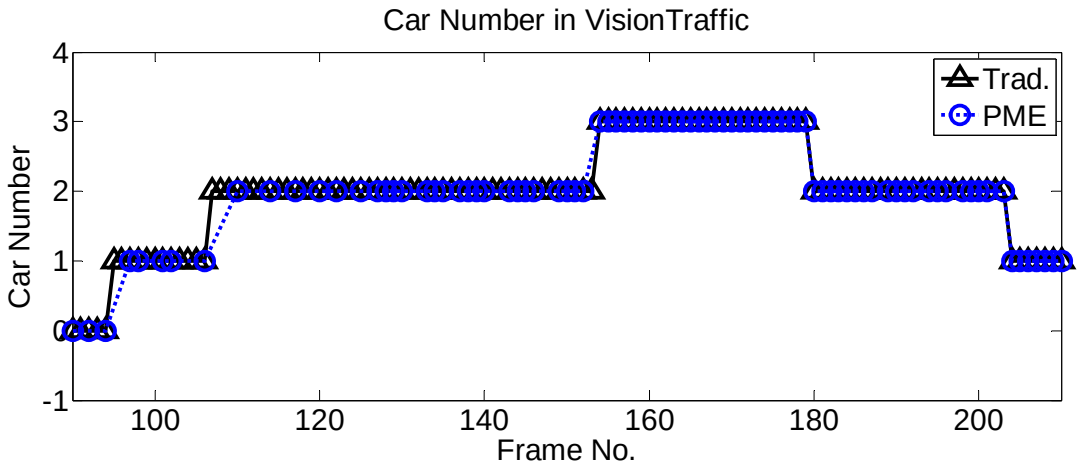
Figure 4.86(a) shows car counting results, and false marker is based on the definition as shown in Figure 4.85. Figure 4.86(b) is the zoom in version of Figure 4.86(a). In Figure 4.86, the car detection and counting results by using PME-based is as good as traditional. The proposed PME-based manner has 99.01% detection ratio. It is clearly to see the sampling behavior with using PME-based in the zoom in version in Figures 4.86(b). Moreover, that the proposed PME-based strategy only samples 303 keyframes from 531 frames to achieve the high accuracy ratio 99.01% with 57.06% utilization ratio as depicted in Table 4.20. The absolute value of energy error ratio is almost 0.3%.

Table 4.20. Results of car detection and counting, utilization ratio, energy error ratio and number of extracted keyframe of VisionTraffic.

Accuracy Percentage of car detection and counting				
Traditional	PME-based	Utilization Ratio	Energy Error Ratio	#Extracted Frames/ #Total Frames
100%	99.01%	57.06%	-0.31%	303/531



(a)



(b)

Figure 4.86. Car counting results of VisionTraffic in traditional and PME-based sampling strategies. (a) Car counting results (b) the zoom in version of (a) from Frames 90 to 210.

4.3.3 Comparison

For demonstrating the applicability and efficiency of the proposed PME-based keyframe extraction algorithm, it compares to three different sampling strategies: traditional, $Fixed(\alpha,\beta)$ and triangle-based. Triangle-based sampling strategy [56: Liu et al. 2003] firstly divides the motion energy curve into motion triangles, and then the frame no. of keyframs are the apexes of motion triangles, for example, Figure 4.87 shows triangle-based sampling strategy. On the other hand, the intuition way is to do sampling with constant period for reducing the data. Three different sampling periods are implemented in $Fixed(\alpha,\beta)$ those are $\alpha=\beta=2$, $\alpha=\beta=3$, and $\alpha=\beta=4$. The odometry results of six different sampling strategies which are traditional, $Fixed(2,2)$, $Fixed(3,3)$, $Fixed(4,4)$, PME-based and triangle-based are presented and compared to demonstrate the outstanding performance of the proposed PME-based method.

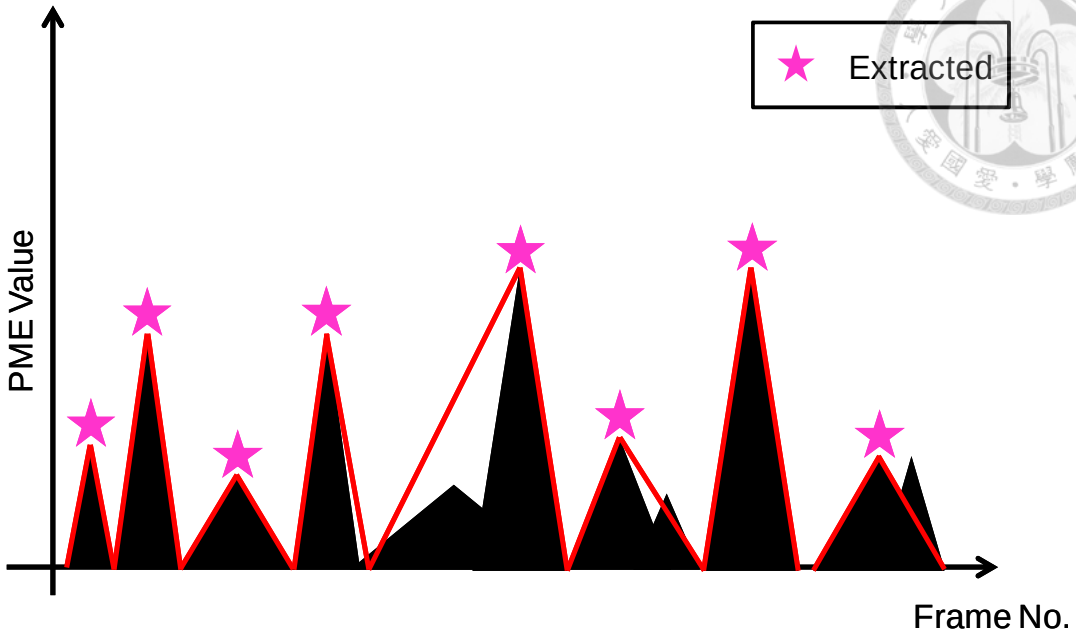
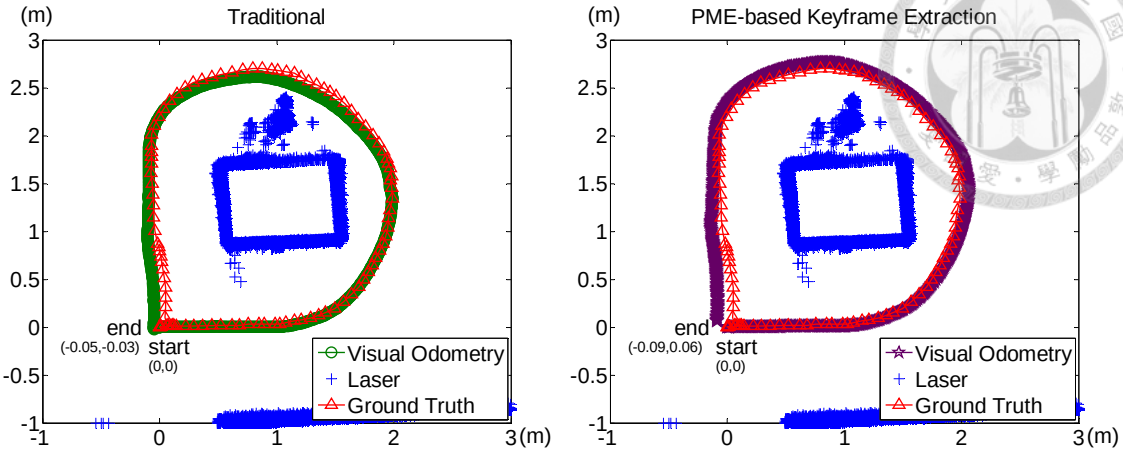


Figure 4.87. The diagram of triangle-based sampling strategy.

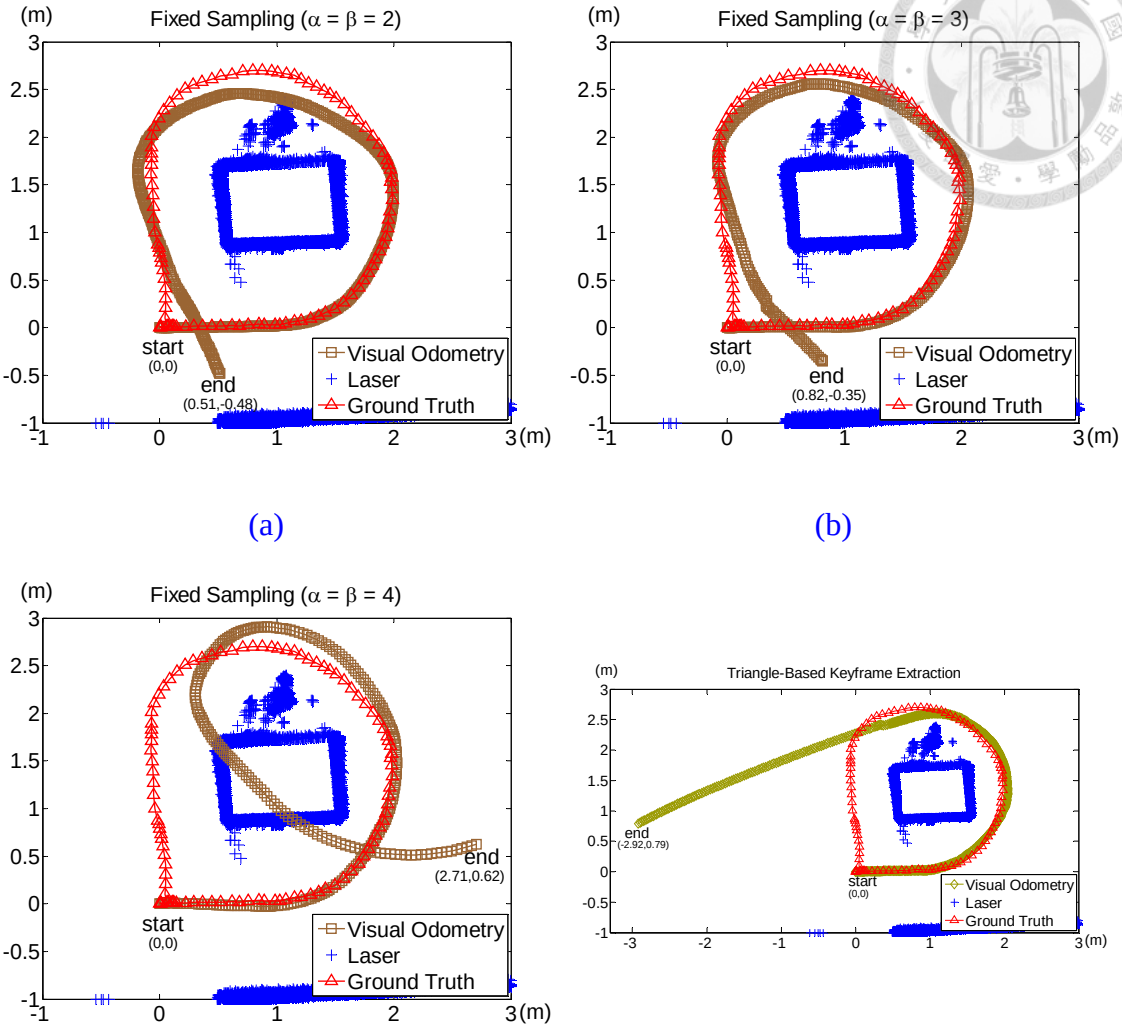
For comparison, the odometry results by using traditional, $Fixed(2,2)$, $Fixed(3,3)$, $Fixed(4,4)$, PME-based and triangle-based are shown in Figures 4.88 and 4.89. Figures 4.88(a) and 4.88(b) show the odometry results of using traditional and the proposed PME-based sampling strategy (same figures as shown in Figures 4.50 and 4.51, respectively). Figures 4.89(a)(b)(c) show the odometry results of using fixed sampling strategy with sampling period two, three and four, respectively. The performance gets worse when sampling period increases as shown in Figure 4.91. Figure 4.89(d) shows the results of using triangle-based sampling strategy. The end point (-2.92, 0.79) has long distance to the start point (0,0) compared with PME-based manner. Figure 4.90 shows the odometry results of traditional and PME-based. The two endpoints are very

close and PME-based manner only utilizes 88.88% of video data. Figure 4.92 presents the odometry results of PME-based, $Fixed(3,3)$, and triangle-based and it is clear to see that the PME-based manner has better performance than $Fixed(3,3)$ and triangle-based manners. Figure 4.93 and Table 4.21 show the odometry results and error distance to start point $(0,0)$ of six different sampling strategies, respectively. The error distance of traditional, PME-based, $Fixed(2,2)$, $Fixed(3,3)$, $Fixed(4,4)$, and triangle-based is 0.05, 0.11, 0.7, 0.89, 2.89 and 3.03, respectively. Hence, the endpoints of traditional and PME-based are close to start point $(0,0)$ than other sampling strategies. Triangle-based sampling strategy gains insufficient motion information for performing matching. Therefore, the proposed PME-based extraction algorithm indeed has the best result compared with $Fixed(2,2)$, $Fixed(3,3)$, $Fixed(4,4)$, and triangle-based. Moreover, the proposed PME-based keyframe extraction has flexible and adaptable for various scenarios and it could reduce greatly the required computation resource and decrease bandwidth utilization.



(a) (b)

Figure 4.88. Odometry result of four different types sampling strategy. (a) Traditional (b) PME-based (same Figure as Figures 4.50 and 4.51, respectively)



(a) (b)

(c) (d)

Figure 4.89. Odometry results of four different types sampling strategy. (a) $Fixed(2,2)$ (b) $Fixed(3,3)$ (c) $Fixed(4,4)$ (d) Triangle-based.

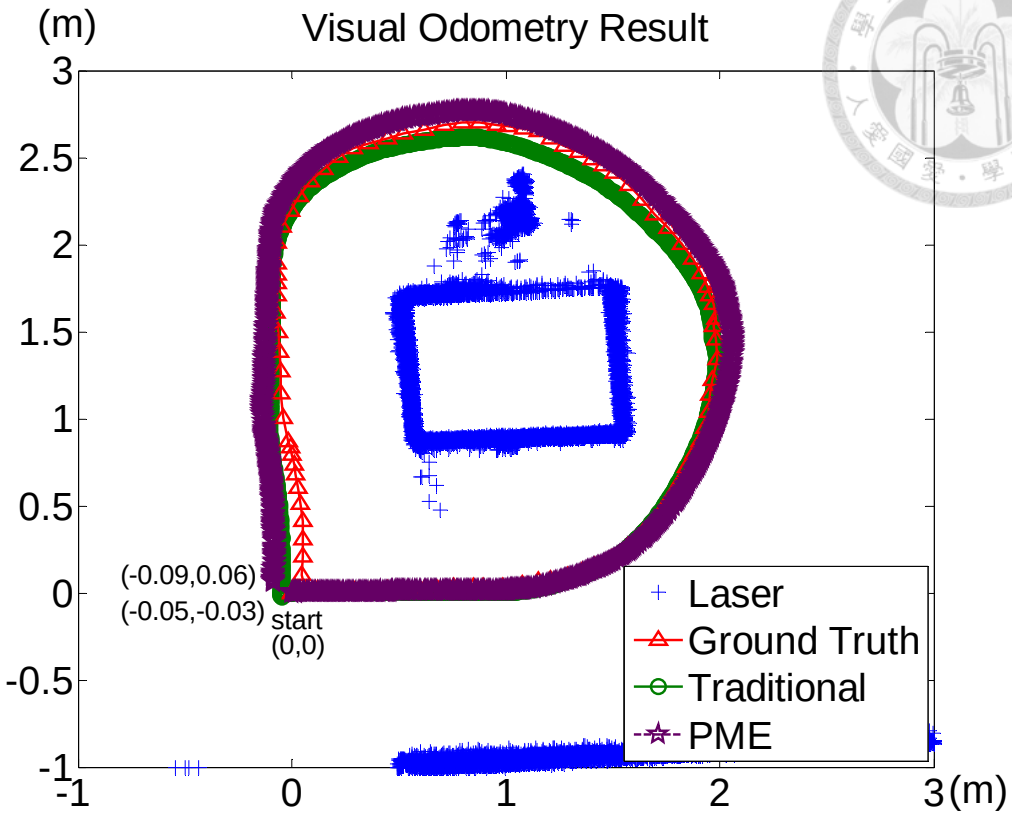


Figure 4.90. Odometry results of traditional and PME-based.

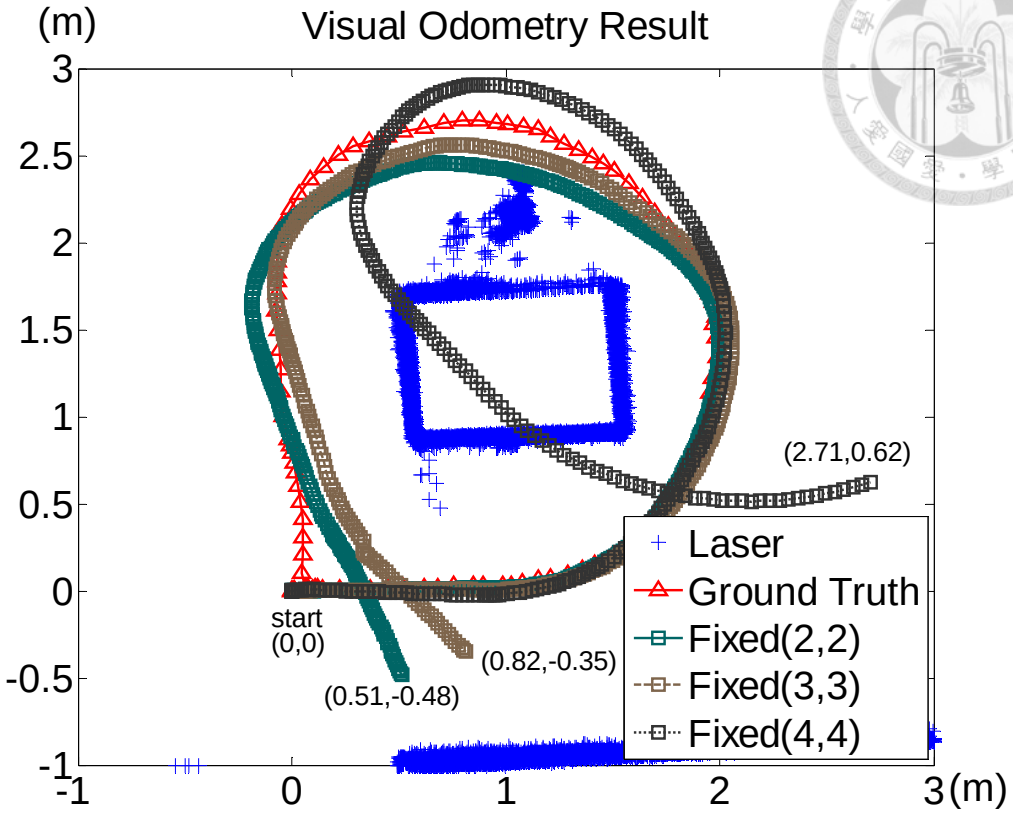


Figure 4.91. Odometry results of three different fixed sampling period, namely, $Fixed(2,2)$, $Fixed(3,3)$ and $Fixed(4,4)$.

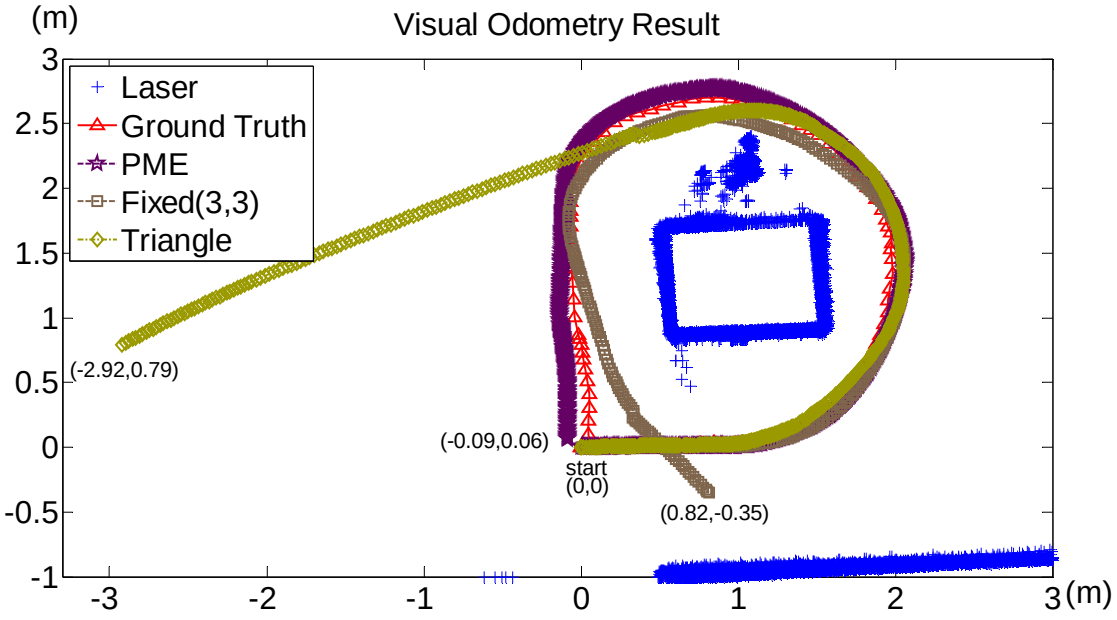


Figure 4.92. Odometry results of three different sampling strategies, namely, PME-based, $Fixed(3,3)$ and triangle-based.

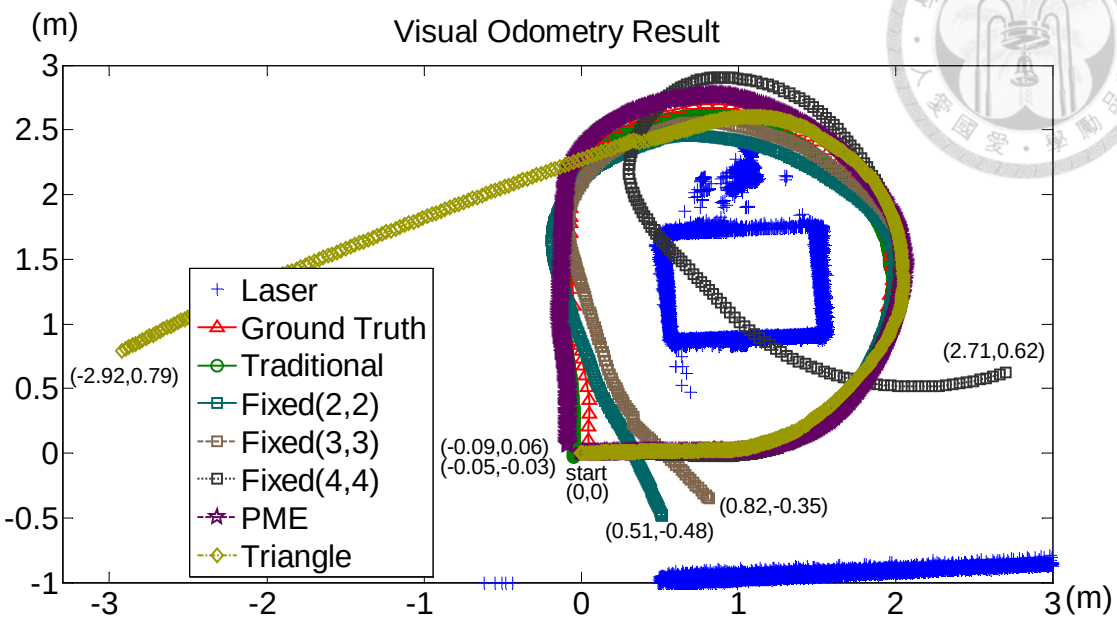


Figure 4.93. Odometry results of six different sampling strategies: traditional, PME-based, $Fixed(2,2)$, $Fixed(3,3)$, $Fixed(4,4)$, and triangle-based.

Table 4.21. Error of different sampling strategies. $dist = \|(x_{end}, y_{end}) - (0,0)\|$.

Traditional	PME-based	$Fixed(2,2)$	$Fixed(3,3)$	$Fixed(4,4)$	Triangle-based
0.05	0.11	0.7	0.89	2.78	3.03

4.4 Summary

The parameters used in Part I and II are summarized in Tables 4.22-4.23 including λ_{ctrl}^{scene}, IQR and AT. Tables 4.22-4.23 are considered as the reference table for performing keyframe extraction algorithm in various environments. The utilization ratio and the experimental results of the ten tested videos in Part I and Part II are summarized in Tables 4.24-4.25. The proposed PME-based sampling strategy only uses about 50% utilization ratio to achieve at least 80% accuracy ratio of human detection, face tracking, and car number counting, low tracking error and 50% saving ratio of bandwidth. Figure 4.94 shows the distribution of ten tested videos in visual sensor × object motion direction plane. Visual sensor is mobile in the right half plane and stationary in the left half plane. Hence motion energy is generated by relative motion between visual sensor and monitored environment or objects in the right half plane. In the left plane, motion energy is derived from the motion of monitored objects due to visual sensor is stationary and the motion could consider as absolute motion. In absolute motion area, the five tested videos are in the same column due to the motion energy all comes from objects motion. They are placed in different row due to the magnitude of the motion energy. The tested video with large motion energy value is placed in top row and vice versa. In relative motion area, the five tested videos are placed as two columns, first column means the motion energy is derived from the interaction of motion in the objects and

visual sensors. Second column means the motion energy is generated from objects motion or visual motion. They are placed in different row due to the magnitude of the motion energy as absolute area.

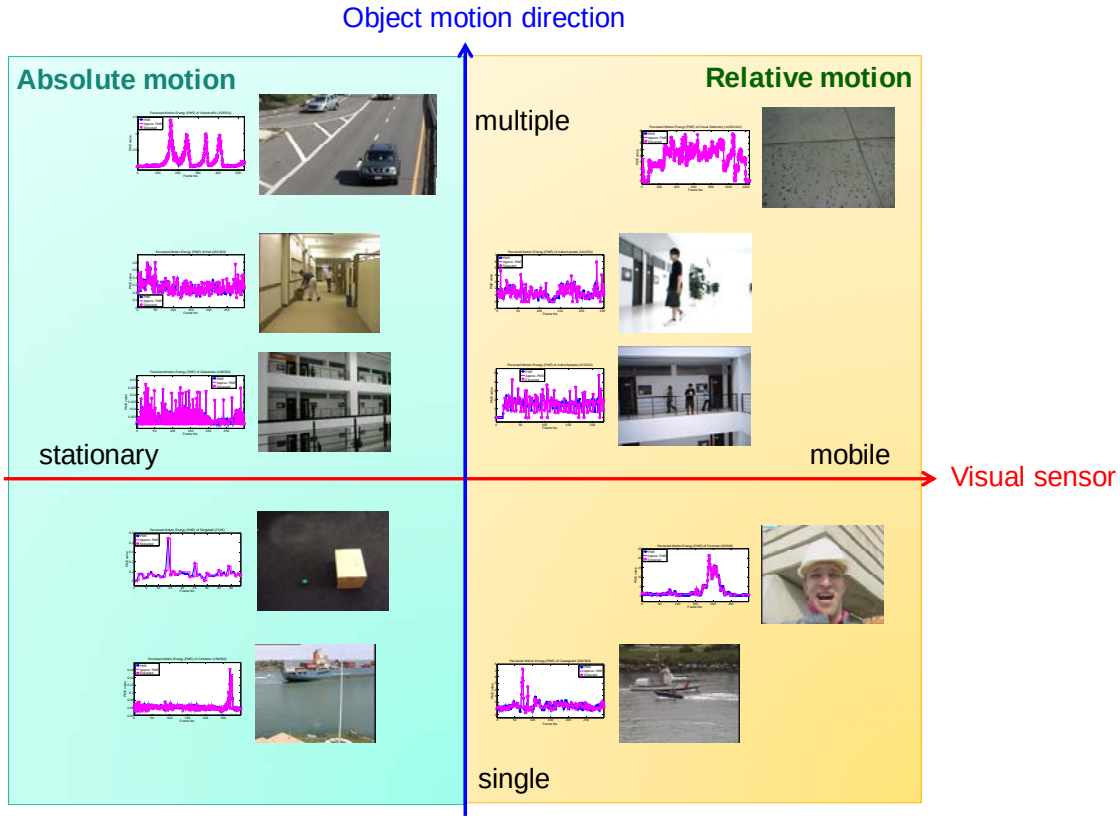


Figure 4.94. The distribution of ten tested videos in sensor-object plane.

Table 4.22. The used parameters in Part I.

Video	Foreman	Coastguard	Indoor1people	Indoor3people	Visual Odometry
λ_{ctrl}^{scene}	1	1	0.5	1	0.01
IQR	1.462	0.657	1.046	0.816	2.263
AT	1.462	0.657	0.523	0.816	0.023

Table 4.23. The used parameters in Part II.

Video	SingleBall	Container	Floors3people	Hall	Vision Traffic
λ_{ctrl}^{scene}	1	1	0.5	0.5	0.1
IQR	0.029	0.005	0.008	0.111	0.212
AT	0.029	0.005	0.004	0.055	0.021

Table 4.24. The utilization ratio and experimental results of the tested videos in Part I. Experimental results mean accuracy ratio, average saving ratio of bandwidth or error. The third and fourth rows represent the experimental results of traditional and PME-based method.

Video	Foreman	Coastguard	Indoor1people	Indoor3people	Visual Odometry
Utilization Ratio	21%	33.33%	55.51%	54.91%	88.88%
Traditional	61.17%	100%	17%	10%	0.05
PME-based	79.6%	100%	54%	51%	0.11

Table 4.25. The utilization ratio and experimental results of the tested videos in Part II. Experimental results mean accuracy ratio, average saving ratio of bandwidth or error. The third and fourth rows represent the experimental results of traditional and PME-based method.

Video	SingleBall		Container	Floors3people	Hall	Vision Traffic
Utilization Ratio	60%		52%	66.33%	60.67%	57.06%
Traditional	3.9	8.4	100%	0.04%	79.33%	100%
	22.1	109.9				
PME-based	18.91	26.41	100%	47%	80.04%	99.01%
	34.11	118.97				

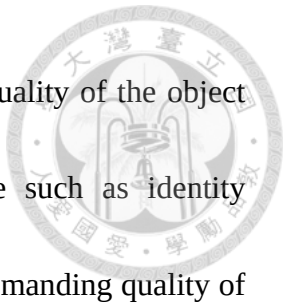
Chapter 5

Conclusion and Future Work

5.1 Conclusion

Visual sensing systems have rapidly received particular attention in the past few years and have been found in various video-related applications such as habitat monitoring, remote-surveillance, military defense, commercial entertainment, sporting competition broadcasting. Visual sensors are displayed in the particular area to collect image-based data of objects of interest and further to transmit them to human operators and/or controller through a shared communication network. Unfortunately, the image analysis results and control performance would be considerably affected in the face of severe constraints on communication bandwidth, delays and packet dropouts. The main reason is the number of transmitted video packets data greatly exceeds the available channel capacity. In the dissertation, video data control is proposed and implemented from two different perspectives: quality and quantity to solve the problem.

Within the quality domain, compression techniques are commonly used to decrease the resolution of frame by discarding video packet data which contains sophisticate details in the recent literature. However, the quality of the object cannot be enhanced by using compression which is post-processing techniques. The quality of object of interest

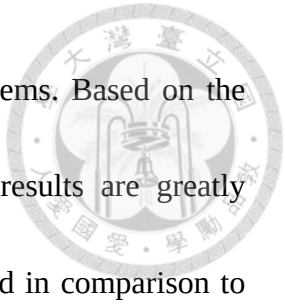


is determined at capturing instant. In practical applications, lower quality of the object would affect the video analysis results and system performance such as identity recognition and visual odometry. Therefore, for actively acquiring demanding quality of the object, the ROI-based controlled sensing framework is proposed, proved and implemented in the PTZ camera to dynamically adjust the zoom value to increase the information density of the object in the dissertation. In comparison of traditional method, the tracking results are greatly improved by using the proposed ROI-based controlled sensing with designed zoom control policies in both simple indoor and complex outdoor experimental scenes. Moreover, it has flexible monitoring capability for avoiding losing the information about surrounding objects.

Within the quantity domain, keyframe extraction is generally used to remove the similar frames by selecting thoses which contain salient content in terms of video summarization. However, the extraction rules are offline and needed to prior decide the number of keyframe in the related works. They are not suitable for time-critical surveillance applications and lack adaptability when system dynamics changes suddenly. Hence, the PME-based keyframe extraction rules are designed to dynamically and adaptively extract the salient frames which contain the significant dynamic information. Enormous typical and experimental videos are use to demonstrate the near 50% reduction ratio and at the same time the performance of tracking, detection and visual

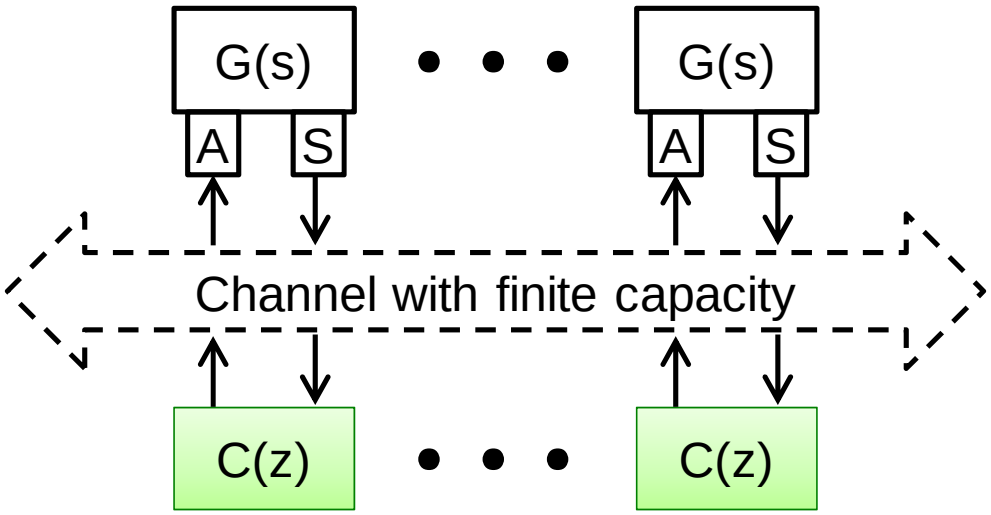
odometry is maintained in a satisfactory range after data reduction process. Moreover, the saving ratio of bandwidth is close to 50%.

Figure 5.1 shows the difference between the traditional control diagram and the proposed sensing control diagram. The traditional control focuses on the design of controller in the face of constraints on noises, disturbances, delays and data dropouts for solving tracking and regulation problems as shown in Figure 5.1(a). Most research work has been devoted to the development of controllers with complex control theories, however sometimes the realization of the controllers is not a simple task. Furthermore, the constraints are derived from the inaccuracy and huge amounts of sensing data. Hence, in the dissertation an intelligent, efficient and flexibility control method in sensing part is proposed for increasing the accuracy and reducing the amount of transmitted sensing data. The proposed quality and quantity control consists of ROI-based controlled sensing, keyframe-based communication module and layer-based transmission mechanism as shown in Figure 5.1(b). ROI-based controlled sensing is responsible for improving the accuracy of data by increasing the information density of the object of interest. The keyframe-based communication module and layer-based transmission mechanism are in charge of reducing the amount of measurements by only transmitting significant packet data. For demonstrating the flexibility and adaptability of the proposed sensing control architecture, it has been experimented on many different

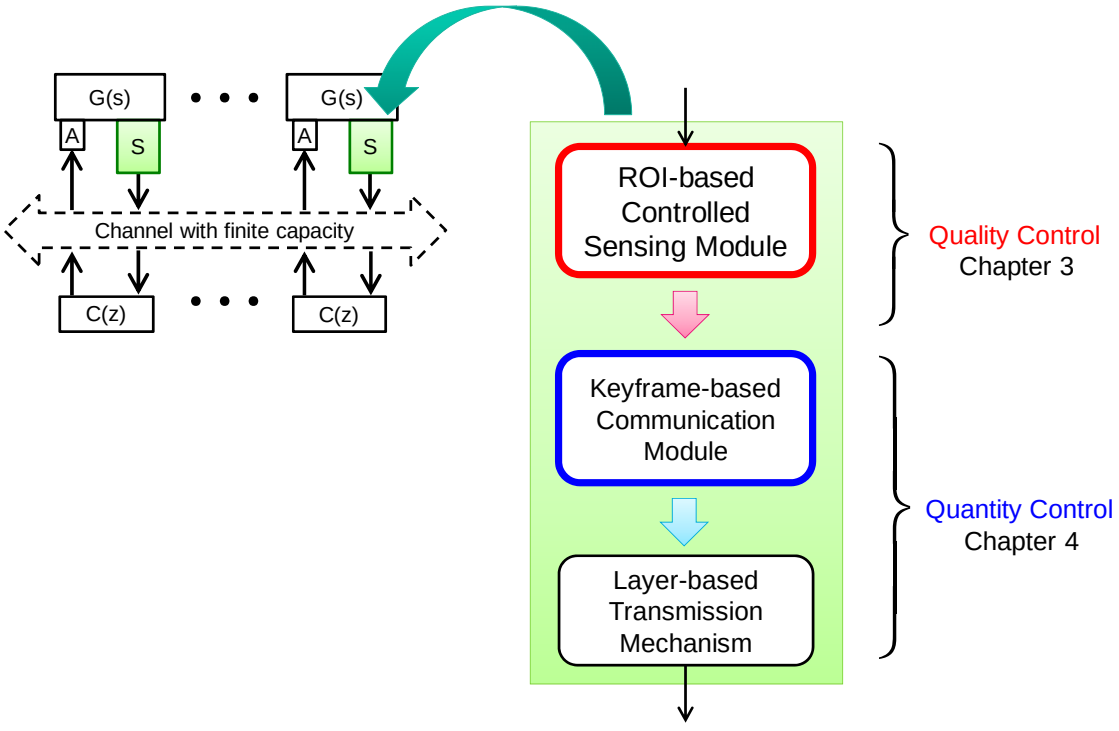


scenes including indoor and outdoor for the different control problems. Based on the hypothesis that control algorithm is the same, the experimental results are greatly improved by using the proposed quality and quantity control method in comparison to other sampling strategies: traditional, fixed sampling and triangle-based as shown in Tables 5.1-5.4. Table 5.1 shows the comparison results of traditional and proposed ROI-based quality control method in indoor and outdoor environment. The proposed ROI-based quality control method only has 1.49% and 2.82% false ratio that are lower than 22.02% and 48.68% of traditional in indoor and outdoor environment, respectively. Tables 5.2-5.4 show the comparison results of detection, bandwidth saving and tracking in ten videos by using different quantity control strategies: traditional, proposed keyframe-based, fixed sampling and triangle-based. The proposed keyframe-based quantity control method only utilizes half of original data to achieve 80% above accuracy ratio of detection in the videos of Foreman, Hall, Container, Coastguard and Visual Traffic as shown in Table 5.2. In Table 5.3, the proposed keyframe-based quantity control method also has average 50% saving ratio of bandwidth in the videos of Indoor1people, Indoor3people and Floors3people that is higher than traditional 10%. Although fixed sampling method has average 56% saving ratio, the detection results of that are worse than the proposed keyframe-based quantity control method as shown in Chapter 4. Table 5.4 shows the proposed keyframe-based quantity control method only

utilizes 60% and 88% to achieve small tracking error that is as good as traditional and better than fixed sampling and triangle-based in the videos of SingleBall and Visual Odometry.



(a)



(b)

Figure 5.1. (a) Traditional control diagram. (b) Proposed sensing control diagram.

Table 5.1. The experimental results of traditional and ROI-based method in two different scenes.

	Indoor Experimental Scene		Outdoor Experimental Scene	
	Traditional	ROI-based	Traditional	ROI-based
pixels/unit	1	[1,20]	1	[1,49]
False ratio	22.06%	1.49%	48.68%	2.82%
Convergence	Yes	Yes	No	Yes

Table 5.2. The accuracy ratio of detection of traditional and keyframe-based method in five typical videos.

Video	Utilization Ratio	Results	
		Traditional	Keyframe
Foreman	21%	61.17%	79.6%
Hall	60.67%	79.33%	80.04%
Container	52%	100%	100%
Coastguard	33.33%	100%	100%
Visual Traffic	57.06%	100%	99.01%

Table 5.3. The saving ratio of bandwidth of traditional, $Fixed(1,3)$ keyframe-based method in five typical videos.

Video	Utilization Ratio	Results		
		Traditional	$Fixed(1,3)$	Keyframe
Indoor1people	55.51%	17%	47%	54%
Indoor3people	54.91%	10%	59%	51%
Globalview	66.33%	0.04%	62%	47%

Table 5.4. The tracking error of traditional, keyframe-based, $Fixed(2,2)$, triangle-based method in two typical videos.

Video	Utilization Ratio	Results			
		Traditional		Keyframe	
SingleBall	60%	3.9	8.4	18.91	26.41
		22.1	109.9	34.11	118.97
Visual Odometry	88.88%	Traditional	Keyframe	$Fixed(2,2)$	Triangle
		0.05	0.11	0.7	3.03

5.2 Future Work

Based on the extensive development in the hardware and software of visual sensors, the video-related applications would rapidly increase in the next decade and becomes future trends in various fields such industrial robotics, manufacturing, intelligent vehicle systems and surveillance systems. The bandwidth limitation is a major constraint for video transmission task. Video data controls from quality and quantity point of view have been proposed in the dissertation. Future studies should integrate the ROI-based controlled sensing and PME-based keyframe extraction to become a complete active and smart sensing systems. On the other hand, a solid theoretical foundation of video data sampling should be developed through a collection of information theory, computer science and control theory for providing reduction criterions in different visual systems.

References

[1: Baillieul & Antsaklis 2007]

J. Baillieul and P. J. Antsaklis, “Control and Communication Challenges in Networked Real-Time Systems,” Proceedings of the IEEE, Vol. 95, No. 1, pp. 9-28, January 2007.

[2: Moyne & Tilbury 2007]

J. R. Moyne, and D. M. Tilbury, “The Emergence of Industrial Control Networks for Manufacturing Control, Diagnostics, and Safety Data,” Proceedings of the IEEE, Vol. 95, No. 1, pp. 29-47, January 2007.

[3: Hespanha et al. 2007]

J. P. Hespanha, P. Naghshtabrizi, and Y. Xu, “A Survey of Recent Results in Networked Control Systems,” Proceedings of the IEEE, Vol. 95, No. 1, pp. 138-162, January 2007.

[4: Foresti et al. 2005]

G. L. Foresti, C. Micheloni, L. Snidaro, P. Remagnino, and T. Ellis, “Active Video-Based Surveillance System: the Low-Level Image and Video Processing Techniques Needed for Implementation,” IEEE Signal Processing Magazine, Vol. 22, No. 2, pp. 25-37, March 2005.

[5: Hampapur et al. 2005]

A. Hampapur, L. Brown, J. Connell, A. Ekin, N. Haas, M. Lu, H. Merkl, S. Pankanti, A. Senior, C.-F. Shu, and Y. L. Tian, “Smart Video Surveillance: Exploring the Concept of Multi-scale Spatiotemporal Tracking,” IEEE Signal Processing Magazine, Vol. 22 , No. 2, pp. 38-51, March 2005.

[6: Schenato et al. 2007]

L. Schenato, B. Sinopoli, M. Franceschetti, K. Poolla, and S. Shankar Sastry, “Foundations of Control and Estimation over Lossy Networks,” Proceedings of the IEEE, Vol. 95, No. 1, pp. 163-187, January 2007.

[7: Zhang et al. 2011]

J. Zhang, F.-Y. Wang, K. Wang, W.-H. Lin, X. Xu, and C. Chen, "Data-Driven Intelligent Transportation Systems: A Survey," IEEE Transactions on Intelligent Transportation Systems, Vol. 12, No. 4, pp. 1624-1639, December 2011.

[8: Matiakis et al. 2009]

T. Matiakis, S. Hirche, and M. Buss, "Control of Networked Systems using the Scattering Transformation," IEEE Transactions on Control Systems Technology, Vol. 17, No. 1, pp. 60-67, January 2009.

[9: Tang & Silva 2006]

P. L. Tang, and C. W. de Silva, "Compensation for Transmission Delays in an Ethernet-Based Control Network using Variable-Horizon Predictive Control," IEEE Transactions on Control Systems Technology, Vol. 14, No. 4, pp. 707-718, July 2006.

[10: Tipsuwan & Chow 2004]

Y. Tipsuwan and M.-Y. Chow, "Gain Scheduler Middleware: A Methodology to Enable Existing Controllers for Networked Control and Teleoperation-Part I: Networked Control," IEEE Transactions on Industrial Electronics, Vol. 51, No. 6, pp. 1218-1227, December 2004.

[11: Lee et al. 2005]

K. C. Lee, S. Lee, and M. H. Lee, "QoS-Based Remote Control of Networked Control Systems Via Profibus Token Passing Protocol," IEEE Transactions on Industrial Informatics, Vol. 1, No. 3, pp. 183-191, August 2005.

[12: Lian et al. 2006]

F.-L. Lian, J. K. Yook, D. M. Tilbury, and J. Moyne, "Network Architecture and Communication Modules for Guaranteeing Acceptable Control and Communication Performance for Networked Multiagent Systems," IEEE Transactions on Industrial Informatics, Vol. 2, No. 1, pp. 12-24, February 2006.

[13: Ghosh et al. 2007]

B. K. Ghosh, A. D. Polpitiya, and W. Wang, "Bio-Inspired Networks of Visual Sensors, Neurons, and Oscillators," Proceedings of the IEEE, Vol. 95, No. 1, pp. 188-214, January 2007.

[14: Kreucher et al. 2007]

C. M. Kreucher, A. O. Hero, K. D. Kastella, and M. R. Morelande, "An Information-Based Approach to Sensor Management in Large Dynamic Networks," Proceedings of the IEEE, Vol. 95, No. 5, pp. 978-999, May 2007.

[15: Lobaton et al. 2010]

E. Lobaton, R. Vasudevan, R. Bajcsy, and S. Sastry, "A Distributed Topological Camera Network Representation for Tracking Applications," IEEE Transactions on Image Processing, Vol. 19, No. 10, pp. 2516-2529, October 2010.

[16: Oppenheim & Willsky 1997]

A. V. Oppenheim, A. S. Willsky, and S. Hamid, "Signals and Systems," 2nd Ed, Prentice Hall, 1997.

[17: Candès & Wakin 2008]

E. J. Candès and M. B. Wakin, "An Introduction to Compressive Sampling," IEEE Signal Processing Magazine, Vol. 25 , No. 2, pp. 21-30, March 2008.

[18: Donoho 2006]

D. L. Donoho, "Compressed Sensing," IEEE Transactions on Information Theory, Vol. 52. No. 4, pp. 1289-1306, April 2006.

[19: Lustig & Donoho 2008]

M. Lustig, D. L. Donoho, J. M. Santos, and J. M. Pauly, "Compressed Sensing MRI," IEEE Signal Processing Magazine, Vol. 25, No. 2, pp. 72-82, March 2008.

[20: Karpenko & Aarabi 2011]

A. Karpenko, and P. Aarabi, "Tiny Videos: A Large Data Set for Nonparametric Video Retrieval and Frame Classification," IEEE Transition on Pattern Analysis and Machine Intelligence, Vol. 33, No. 3, pp. 618-630, March 2011.

[21: Tordoff & Murray 2003]

B. J. Tordoff and D. W. Murray, "Resolution vs. Tracking Error: Zoom as a Gain Controller," IEEE Conference on Computer Vision and Pattern Recognition, Vol. 1, pp. I-273-I-280, Madison, Wisconsin, 18-20 June 2003

[22: Chen et al. 2008]

C.-H. Chen, Y. Yao, D. Page, B. Abidi, A. Koschan, and M. Abidi, "Heterogeneous Fusion of Omnidirectional and PTZ Cameras for Multiple Object Tracking," IEEE Transactions on Circuits and Systems for Video Technology, Vol. 18, No. 8, pp. 1052-1063, August 2008

[23: Micheloni et al. 2010]

C. Micheloni, B. Rinner, and G. L. Foresti, “Video Analysis in Pan-Tilt-Zoom Camera Networks,” IEEE Signal Processing Magazine, Vol. 27, No. 5, pp. 78-90, September 2010.

[24: Han et al. 2011]

J. Han, D. Farin, and P. H. N. de With, “A Mixed-Reality System for Broadcasting Sports Video to Mobile Devices,” IEEE Multimedia, Vol. 18, No. 2, pp. 72-84, 2010.

[25: Huang et al. 2009]

Y. Huang, S. Mao, and S. F. Midkiff, “A Control-Theoretic Approach to Rate Control for Streaming Videos,” IEEE Transactions on Multimedia, Vol. 11, No. 6, pp. 1072-1081, October 2009.

[26: Dinh et al. 2011]

Thang Ba Dinh, Nam Vo and G´erard Medioni, “High Resolution Face Sequences from A PTZ Network Camera,” IEEE International Conference on Automatic Face & Gesture Recognition and Workshops, pp. 531-538, Santa Barbara, CA, 21-25 March 2011.

[27: Micheloni & Foresti 2005]

C. Micheloni and G.L. Foresti, “Zoom on Target while Tracking,” IEEE International Conference on Image Processing (ICIP), vol. 3, pp. 117-120, Genoa, Italy, 11-14 September 2005.

[28: Tordoff & Murray 2004]

B. Tordoff and D. Murray, “Reactive Control of Zoom while Fixating using Perspective and Affine Cameras,” IEEE Transactions on Pattern Analysis And Machine Intelligence, Vol. 26, No. 1, pp. 98-112, January 2004.

[29: Dinh et al. 2009]

T. Dinh, Q. Yu, and G. Medioni, “Real Time Tracking using an Active Pan-Tilt-Zoom Network Camera,” IEEE/RSJ International Conference on Intelligent Robots and Systems, St. Louis, USA, pp. 3786-3793, 11-15 October 2009.

[30: Ding et al. 2012]

C. Ding, B. Song, A. Morye, J. A. Farrell, and A. K. Roy-Chowdhuryrray, "Collaborative Sensing in a Distributed PTZ Camera Network," IEEE Transactions on Image Processing, Vol. 21, No. 7, pp. 98-112, July 2012.

[31: Goodman & Gersho 1974]

D. J. Goodman and A. Gersho, "Theory of Adaptive Quantizer," IEEE Transactions on Communications, Vol. COM-22, No. 8, pp. 1037-1045, August 1974.

[32: Brockett and Liberzon 2000]

R. W. Brockett and D. Liberzon, "Quantized Feedback Stabilization of Linear Systems," IEEE Transactions on Automatic Control, Vol. 45, No. 7, pp. 1279-1289, July, 2000.

[33: Delchamps 1990]

D. Delchamps, "Stabilizing a Linear System with Quantized State Feedback," IEEE Transactions on Automatic Control, Vol. 35, No. 8, pp. 916-924, August, 1990.

[34: Liberzon 2003]

D. Liberzon, "On Stabilization of Linear Systems with Limited information," IEEE Transactions on Automatic Control, Vol. 48, No. 2, pp. 304-307, February, 2003.

[35: Elia & Mitter 2001]

N. Elia and S. K. Mitter, "Stabilization of Linear Systems with Limited Information," IEEE Transactions on Automatic Control, Vol. 46, No. 9, pp. 1384-1400, September, 2001.

[36: Tatikonda & Mitter 2004]

S. Tatikonda and S. Mitter, "Control under Communication Constraints," IEEE Transactions on Automatic Control, Vol. 49, No.7, pp. 1056-1068, July, 2004.

[37: Nesci & Liberzon 2009]

D. Nesci and D. Liberzon, "A Unified Framework for Design and Analysis of Networked and Quantized Control Systems," IEEE Transactions on Automatic Control, Vol. 54, No. 4, pp. 732-747, April, 2009.

[38: Chen et al. 2010]

Y. Chen, Y. Deng, Y. Guo, W. Wang, Y. Zou, and K. Wang, “A Temporal Video Segmentation and Summary Generation Method Based on Shots' Abrupt and Gradual Transition Boundary Detecting,” Second International Conference on Communication Software and Networks, Singapore, pp. 271-275, 26-28 February 2010.

[39: Chatzigiorgaki & Skodras 2009]

M. Chatzigiorgaki and A. N. Skodras, “Real-Time Keyframe Extraction Towards Video Content Identification,” 16th International Conference on Digital Signal Processing, Santorini-Hellas, pp. 1-6, 5-7 July 2009.

[40: Li et al. 2006]

Y. Li, S.-H. Lee, C.-H. Yeh, and C.-C. J.Kuo, “Techniques for movie content analysis and skimming: Tutorial and overview on video abstraction techniques,” IEEE Signal Processing Magazine, Vol. 23, No. 2, pp. 79-89, March 2006.

[41: Jiang & Qin 2010]

P. Jiang and X.-L. Qin, “Keyframe-Based Video Summary using Visual Attention Clues,” IEEE Multimedia, Vol. 17, No. 2, pp. 64-73, 2010.

[42: Lee and Hayes 2004]

S. Lee and M. H. Hayes, ‘‘Properties of the Singular Value Decomposition for Efficient Data Clustering,’’ IEEE Signal Processing Letters, Vol. 11, No. 11, pp. 862-866, 2004.

[43: Zhuang et al. 1998]

Y. Zhuang, Y. Rui, T. S. Huang and S. Mehrotra, “Adaptive Keyframe Extraction using Unsupervised Clustering,” International Conference on Image Processing (ICIP) ,Chicago, IL, Vol.1, pp. 866-870, 4-7 October 1998.

[44: Hammound & Mohr 2000]

R. Hammoud and R. Mohr, “A Probabilistic Framework of Selecting Effective Key Frames for Video Browsing and Indexing,” International Workshop on Real-Time Image Sequence Analysis, pp. 79-88, 2000.

[45: Liu et al. 2004]

T.-Y. Liu et al., ‘‘Shot Reconstruction Degree: A Novel Criterion for Key Frame Selection,’’ Pattern Recognition Letters, Vol. 25, No. 12, pp. 1451-1457, 2004.

[46: Zeinalpour 2009]

Z. Zeinalpour, B. M. Bidgoli, and M. Fathi, “Video Summarization using Genetic Algorithm and Information Theory,” 14th International CSI Computer Conference, pp. 158-163, Tehran, 20-21 October 2009.

[47: Panagiotakis et al. 2009]

C. Panagiotakis, A. Doulamis, and G. Tziritas, “Equivalent Key Frames Selection based on Iso-Content Principles,” IEEE Transactions on Circuits and Systems for Video Technology, Vol. 19, No. 3, pp. 447-451, March 2009.

[48: Ngo et al. 2005]

C.-W. Ngo, Y.-F. Ma, and H.-J. Zhang, “Video Summarization and Scene Detection by Graph Modeling,” IEEE Transactions on Circuits and Systems for Video Technology, Vol. 15, No. 2, pp. 296-305, February 2005.

[49: Li et al. 2005]

Z. Li, G. M. Schuster, and A. K. Katsaggelos, “MINMAX Optimal Video Summarization,” IEEE Transactions on Circuits and Systems for Video Technology, Vol. 15, No. 10, pp. 1245-1256, October 2005.

[50: Zhang et al. 2010]

J. Zhang, H. Huang, and J. Wang, “Manifold Learning for Visualizing and Analyzing High-Dimensional Data,” IEEE Intelligent Systems, Vol. 25, No. 4, pp. 54-61, 2010.

[51: Schwarz 1978]

G. Schwarz, “Estimating the Dimension of a Model,” The Annals of Statistics, Vol. 6, No. 2, pp. 461-464, 1978.

[52: Celeux & Govaert 1995]

G. Celeux and G. Govaert, “Gaussian Parsimonious Clustering Models,” Pattern Recognition, Vol. 28, No. 5, pp. 781-783, 1995.

[53: Guan et al. 2013]

G. Guan, Z. Wang, S. Lu, J. D. Deng, and D. Dagan, “Keypoint-Based Keyframe Selection,” IEEE Transactions on Circuits and Systems for Video Technology, Vol. 23, No. 4, pp.729-734, April 2013.

[54: Anagnastopoulos et al. 2009]

V. Anagnastopoulos, N. Doulamis and A. Doulamis, “Edge-motion Video Summarization: Economical Video Summarization for Low Powered Devices,” 10th Workshop on Image Analysis for Multimedia Interactive Services (WIAMIS), pp. 284-287, Brunei Gallery, London, 6-8 May 2009.

[55: Kang et al. 2011]

L.-W. Kang, C.-Y. Hsu, H.-W. Chen, C.-S. Lu, C.-Y. Lin, and S.-C. Pei, “Feature-Based Sparse Representation for Image Similarity Assessment,” IEEE Transactions on Multimedia, Vol. 13, No. 5, pp. 1019-1030, October 2011.

[56: Liu et al. 2003]

T. Liu, H.-J. Zhang, and F. Qi, “A Novel Video Key-Frame-Extraction Algorithm Based on Perceived Motion Energy Model,” IEEE Transactions on Circuits and Systems for Video Technology, Vol. 13, No. 10, pp. 1006-1013, October, 2003.

[57: Lowe 2004]

D. G. Lowe, “Distinctive image features from scale-invariant keypoints,” International Journal of Computer Vision, Vol. 60, No. 2, pp. 91-110, November 2004.

[58: Ratsamee et al. 2013]

P. Ratsamee, Y. Mae, A. Jinda-apiraksa, J. Machajdik, K. Ohara, M. Kojima, R. Sablatnig, and T. Arai, “Lifelogging Keyframe Selection Using Image Quality Measurements and Physiological Excitement Features,” IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pp. 5215-5220, Tokyo, Japan, 3-7 November, 2013.

[59: Liu & Fan 2005]

L. Liu and G. Fan, “Combined Key-Frame Extraction and Object-Based Video Segmentation,” IEEE Transactions on Circuits and Systems for Video Technology, Vol. 15, No. 7, pp. 869-884, June 2005.

[60: Song & Fan 2006]

X. Song and G. Fan, “Joint Key-Frame Extraction and Object Segmentation for Content-Based Video Analysis,” IEEE Transactions on Circuits and Systems for Video Technology, Vol. 16, No. 7, pp. 904-914, July 2006.

[61: Raty 2010]

T. D. Raty, “Survey on Contemporary Remote Surveillance Systems for Public Safety,” IEEE Transactions on Systems, Man, and Cybernetics-part C: Applications and Reviews, Vol. 40, No. 5, pp. 493-515, September 2010.

[62: Huu et al. 2010]

P. N. Huu, V. Tran-Quang, and T. Miyoshi, “Image Compression Algorithm Considering Energy Balance on Wireless Sensor Networks,” IEEE International Conference on Industrial Informatics (INDIN), pp. 1005-1010, Osaka, Japan, 13-16 July 2010.

[63: Misra et al. 2008]

S. Misra, M. Reisslein, and G. Xue, “A Survey of Multimedia Streaming in Wireless Sensor Networks,” IEEE Communications Surveys & Tutorials, Vol. 10, No. 4, pp. 18-39, January 2008.

[64: Si et al. 2012]

Y. Si, J. Mei, and H. Gao, “Novel Approaches to Improve Robustness, Accuracy and Rapidity of Iris Recognition Systems,” IEEE Transactions on Industrial Informatics, Vol. 8, No. 1, pp. 110-117, February 2012.

[65: Huang et al. 2009]

Y. Huang, S. Mao, and S. F. Midkiff, “A Control-Theoretic Approach to Rate Control for Streaming Videos,” IEEE Transactions on Multimedia, Vol. 11, No. 6, pp. 1072-1081, October 2009.

[66: Wu et al. 2000]

D. Wu, Y. T. Hou, and Y. Q. Zhang, “Transporting Real-Time Video Over the Internet: Challenges and Approaches,” Proceedings of IEEE, Vol. 88, No. 12, pp. 1855-1877, December 2000.

[67: Lian et al. 2006]

F.-L. Lian, J. K. Yook, D. M. Tilbury, and J. Moyne, “Network Architecture and Communication Modules for Guaranteeing Acceptable Control And Communication Performance for Networked Multiagent Systems,” IEEE Transactions on Industrial Informatics, Vol. 2, No. 1, pp. 12-24, February 2006.

[68: Andersson et al. 2008]

B. Andersson, N. Pereira, W. Elmenreich, E. Tovar, F. Pacheco, and N. Cruz, "A Scalable and Efficient Approach for Obtaining Measurements in CAN-Based Control Systems," IEEE Transactions on Industrial Informatics, Vol. 4, No. 2, pp. 80-91, May 2008.

[69: Li et al. 2006]

Y. Li, S.-H. Lee, C.-H. Yeh, and C.-C. J. Kuo, "Techniques for Movie Content Analysis and Skimming: Tutorial and Overview on Video Abstraction Techniques," IEEE Signal Processing Magazine, Vol. 23, No. 2, pp. 79-89, March 2006.

[70: Buche et al. 1998]

C. Buchel, O. Josephs, G. Rees, R. Turner, C. D. Frith and K. J. Friston, "The Functional Anatomy of Attention to Visual Motion: A Functional MRI Study," Brain, Vol. 121, No. 7, pp. 1281-1294, 1998.

[71: Kanwisher & Wojciulik 2000]

N. Kanwisher and E. Wojciulik, "Visual Attention: Insights from Brain Imaging," Nature Reviews Neuroscience, Vol. 1, No. 2, pp. 91-100, November 2000.

[72: Duncan et al. 1997]

J. Duncan, G. Humphreys and R. Ward, "Competitive Brain Activity in Visual Attention," Current Opinion in Neurobiology, Vol. 7, No. 2, pp. 255-261, April 1997.

[73: Liu et al. 2008]

C. Liu, W. T. Freeman, E. H. Adelson, and Y. Weiss, "Human-assisted motion annotation," IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp.1-8, Anchorage, AK, 23-28 June 2008.

[74: Irani & Anandan 1999]

M. Irani and P. Anandan, "About Direct Methods," International Conference on Computer Vision (ICCV), Kerkyra, Corfu, Greece, pp. 267-277, September 20-25, 1999.

[75: Torr & Zisserman 1999]

P. H. S. Torr and A. Zisserman, "Feature Based Methods for Structure and Motion Estimation," ICCV Workshop on Vision Algorithms, Corfu, Greece, pp. 278-294, 1999.

[76: Lian et al. 2013]

F.-L. Lian, Y.-C. Lin, C.-T. Kuo, and J.-H. Jean, “Voting-Based Motion Estimation for Real-Time Video Transmission in Networked Mobile Camera Systems,” IEEE Transactions on Industrial Informatics, Vol. 9, No. 1, pp. 172-180, February 2013.

[77: Aires et al. 2008]

K. R. T. Aires, A. M. Santana, and A. A. D. Medeiros, “Optical Flow Using Color Information,” 23rd Annual ACM Symposium on Applied Computing, Fortaleza, Ceará, Brazil, pp. 1607-1611, March 16 -20, 2008.

[78: Graham & Ian 1996]

U. Graham and C. Ian, “Understanding Statistics,” Oxford University Press, p. 55, 1996.

[79: Zwillinger & Kokoska 2000]

D. Zwillinger and S. Kokoska, “CRC Standard Probability and Statistics Tables and Formulae,” Journal of the Royal Statistical Society: Series D (The Statistician), Vol. 50, No. 2, pp. 239-241, 2001.

[80: Mizera & . Müller 2004]

I. Mizera and C. H. Müller, “Location-Scale Depth,” Journal of the American Statistical Association, Vol. 99, No. 468, pp. 949-966, December 2004.

[81: Rousseeuw & Croux 1993]

P. J. Rousseeuw and C. Croux, “Alternatives to the Median Absolute Deviation,” Journal of the American Statistical Association, Vol. 88, No. 424, pp. 1273-1283, December 1993.

[82: Lian et al. 2012]

F.-L. Lian, Y.-C. Lin, C.-T. Kuo, and J.-H. Jean, “Rate and Quality Control with Embedded Coding for Mobile Robot with Visual Patrol,” IEEE Systems Journal, Vol. 6, No. 3, pp. 368-377, September 2012.

[83: Chen 1999]

C.-T. Chen, “Linear System Theory and Design,” 3rd Ed., Oxford, 1999.

Yi-Chun Lin（林意淳）

Email: d96921002@ntu.edu.tw

CURRICULUM VITAE

EDUCATION

National Taiwan University
Ph.D. in Electrical Engineering (2014)
Dissertation: “Region-of-Interest-Based Quality Control and Keyframe-Based Quantity Control in Visual Sensing Systems”
Advisor: Dr. Feng-Li Lian

National Central University
M.S. in Mathematics (2005)

National Central University
B.S. in Mathematics (2003)

PUBLICATIONS

Journal articles:

1. Feng-Li Lian, Yi-Chun Lin, Chien-Ting Kuo, and Jong-Hann Jean, “**Voting-Based Motion Estimation for Real-Time Video Transmission in Networked Mobile Camera Systems**,” *IEEE Transactions on Industrial Informatics*, Vol. 9, No. 1, pp. 172-180, February 2013.

2. Feng-Li Lian, Yi-Chun Lin, Chien-Ting Kuo, and Jong-Hann Jean, “**Rate and Quality Control with Embedded Coding for Mobile Robot with Visual Patrol**,” *IEEE Systems Journal*, Vol. 6, No. 3, pp. 368-377, September 2012.

Conference papers:

1. Yi-Chun Lin and Feng-Li Lian, “**Data Reduction Based on Keyframe with Motion Energy Extraction Rules**,” (Best Paper in Information Award) *2014 IEEE International Conference on Information and Automation (ICIA)*, Hulun Buir, China, July 2014.

2. Yi-Chun Lin, Chung-Yi Hung, Yung-Cheng Huang, Bo-I Wu, and Feng-Li Lian, "**Extracting Key-Frame for Maximizing Transmission Quality in Visual Sensing System and Multi-Dimensional Reconstruction for Endoscopic Video**," (Best Student Paper Award) *Proceedings of 2013 CACS International Automatic Control Conference*, Sun Moon Lake, Nantou, Taiwan, December 2013.
3. Yi-Chun Lin and Feng-Li Lian, "**Reducing Redundant Video Data by Extracting Key-Frame with Significant Perceived Motion Energy for Mobile Visual Sensing System**," *Proceedings of the 2013 IEEE International Conference on Information and Automation (ICIA)*, pp. 19-24, Yinchuan, China, August 2013.
4. Yi-Chun Lin and Feng-Li Lian, "**Data Reduction and Bandwidth Allocation for Video-Based Network System**," *Proceedings of the 2012 IEEE International Conference on Information and Automation (ICIA)*, pp. 116-121, Shenyang, China, June 2012.
5. Yi-Chun Lin, Feng-Li Lian, and Chih-Ming Hsu, "**Design Framework of Controlled Sensing for Data Reduction Under Data Rate Constraints**," *Proceedings of The International Conference on Automatic Control and Artificial Intelligence (ACAI)*, Xiamen, China, March 2012.
6. Yi-Chun Lin and Feng-Li Lian, "**Information Reduction Based on Temporal Similarity and Spatial Importance for Video Transmission in Mobile Surveillance System**," (Best Paper Finalist) *Proceedings of the 2011 IEEE International Conference on Information and Automation (ICIA)*, pp. 91-96, Shenzhen, China, June 2011.
7. Feng-Li Lian, Yi-Chun Lin, and Jong-Hann Jean, "**Information Rate Control for Characterizing Moving Objects in Networked Visual Applications**," *Proceedings of the 2010 IEEE International Conference on Industrial Informatics*, pp. 384-389, Osaka, Japan, July 2010.
8. Chien-Ting Kuo, I-Iming Chen, Feng-Li Lian, Yi-Chun Lin, and Jong-Hann Jean, "**Temporal Sampling and Spatial Coding for Rate Control of Video Transmission On Mobile Cameras**," *Proceedings of the 2010 IEEE International Conference on Control Applications*, pp.71-76, Yokohama, Japan, September 2010.
9. Yi-Chun Lin and Feng-Li Lian, "**On Characterizing the Interaction between Dynamical Behavior and Information Flow in Networked Control Systems**," (Best Presentation Paper Award) *Proceedings of the 2008 IEEE International Conference on Industrial Informatics*, pp. 1574-1579, Daejeon, Korea, July 2008.