



Title	ビデオ符号化を用いた複数チャネル環境における映像品質の向上とリアルタイム通信への適用に関する研究
Author(s)	小桐, 康博
Citation	大阪大学, 2014, 博士論文
Version Type	VoR
URL	<a href="https://doi.org/10.18910/50570">https://doi.org/10.18910/50570</a>
rights	
Note	

*The University of Osaka Institutional Knowledge Archive : OUKA*

<https://ir.library.osaka-u.ac.jp/>

The University of Osaka

ビデオ符号化を用いた複数チャネル環境における映像品質の  
向上とリアルタイム通信への適用に関する研究

小桐 康博

2014年7月

## 関連発表論文

学術論文誌（査読付き）

1. 小桐康博,山口博幸,栄藤稔, “量子化操作による MPEG-4 映像符号化の多重記述”, 画像電子学会誌, Vol.37, No.5, pp.663-670, 2008

学術研究集会報告

1. 小桐康博,山口博幸,栄藤稔, “MPEG-4 を用いた多重記述映像符号化の実現とその評価”, 電子情報通信学会総合大会シンポジウム, SD-3-16, pp.403-404, 2002
2. 小桐康博,山口博幸,正満峰夫,名古博幸, “実網を用いたエラー耐性を有する画像通信方式の性能評価”, 電子情報通信学会総合大会, D-11-75, pp.75, 1998
3. 小桐康博,山口博幸,正満峰夫,名古博幸, “GOB 毎にエラー耐性機能を有した画像通信方式の性能評価”, 電子情報通信学会ソサイエティ大会, D-11-31, pp.146, 1998
4. 小桐康博,筒井朗博,高橋誠, “ユーザビリティを考慮したインドア環境における GPS 測位について”, GPS/GNSS シンポジウム, p139-141, 2009
5. Han Li, Yasuhiro Ogiri, Hiroyuki Yamaguchi, “A Low Computational Complexity Video Transcoding Algorithm”, The7th international Work on Mobile multimedia Communications (MoMuC2000), P14, pp.1-6, 2000
6. 李瀚,小桐康博,山口博幸, “DCT 係数を活用した直接トランスコーディング”, 電子情報通信学会総合大会, D-11-64, pp.64, 2000
7. 山口博幸,小桐康博,大矢智之, “IMT-2000 モバイルマルチメディア信号処理, ビジュアルフォン, 映像クリップ配信の高機能化”, 電子情報通信学会技術研究報告, DSP, デジタル信号処理 102(41), pp.69-72, 2002

社内研究集会報告

1. 小桐康博,山口博幸, “Quantizer Method を用いた MD の一手法”, 2001
2. 小桐康博,山口博幸, “Packet Loss が発生する NW 環境における Temporal MD の品質”, 2001
3. 小桐康博,名古博幸,山口博幸,正満峰夫, “リアルタイム画像通信(NEWPRED)”, 1998
4. 小桐康博,名古博幸,山口博幸,正満峰夫, “リアルタイム画像通信における最適動ベクトル探索方法”, 1998
5. 小桐康博,山口博幸,正満峰夫,名古博幸,市川忠嗣,一之瀬進, “高エラー耐性動画像符号化ソフトモジュールの開発”, 1997
6. 小桐康博,李瀚,山口博幸, “MPEG-2⇒MPEG-4 トランスコーディングの検討”, 2000
7. 小桐康博,李瀚,山口博幸, “双方向トランスコーディング(H.261,H.263⇔MPEG-4)の検討”, 2001
8. 小桐康博,山口博幸, “IMT2000/PHS ビジュアルフォンとの接続方式および 3G-324F の実装に伴う技術情報”, 2001

## 特許

1. 特許公開 2012-044263 情報提供装置及び情報提供方法
2. 特許公開 2002-077867 コンテンツ視聴システムおよびコンテンツ視聴方法
3. 特許公開 2001-359104 動画像符号化データのトランスコーディング方法およびトランスコーディング装置
4. 特許公開 2001-359103 動画像データ符号化装置,動画像データ伝送方法および動画像データ復号装置
5. 特許公開 2001-169295 動画像符号化データのトランスコーディング方法およびトランスコーディング装置
6. 特許公開平 11-164296 映像表示方法及び装置
7. WO03/026315 符号化方法,復号方法,符号化装置,復号装置,画像処理システム,符号化プログラム及び復号プログラム

# ビデオ符号化を用いた複数チャネル環境における映像品質の向上とリアルタイム通信への適用に関する研究

## 内容梗概

近年，サーバからデータを受信しながら映像を再生するビデオストリームには，ビデオ符号化が用いられており，ビデオストリームの低ビットレート化および高品質化の重要な技術要素となっている．一方，ネットワーク環境は，セルラーや無線 LAN などの異種無線アクセス網を組み入れた複合的無線サービスが普及してきている．そのような複合的無線サービスでは，複数の通信経路を持つことが可能となるので，ネットワークの負荷分散，通信障害耐性向上などに効果が期待されている．このような複数チャネルを有するネットワーク環境によって，ビデオストリーム品質が向上してきている．一般的には，同じ品質の映像を異なるネットワークで 2 重化して配信することで，映像配信の信頼性や品質を確保する完全冗長 2 重化方式が使われていることが多い．しかし，ネットワークが異なると帯域などのネットワーク特性が異なるため，それらのネットワークを用いて同じ品質の映像を配信するには，遅延が発生して，リアルタイム性を損なう問題がある．ネットワーク特性に応じたビデオストリームを生成する方法の 1 つとして，ビデオ符号化のスケーラビリティ機能がある．例えば，2 つのビデオストリームをそれぞれに基本レイヤと拡張レイヤと定義し，基本レイヤは，低品質の画像が復号できるもの，拡張レイヤでは，基本レイヤの品質を補強する高品質なデータを送信する．基本レイヤのビデオストリームが受信機で正常に受信できないと拡張レイヤの映像データも復号できない．送信側に通知して基本レイヤと拡張レイヤを入れ替える必要があり，既存の伝送構成に大幅な変更が必要であった．

そこで，複合的無線サービスに適した多重記述のビデオ符号化を用いた映像品質の向上を目的とした．ビデオストリームに対して多重記述ビデオ符号化を行うことで，1 つの映像ソースを 2 つのエンコーダによってそれぞれ送信し，相互補完することで途切れの無いデータ通信が行われ，さらに 2 つのチャネルが受信できれば，高品質なデータ通信が行われる．次に，高品質なビデオストリームの受信ができるようになると，受信したビデオデータを解析することによって様々なアプリケーションが実現すると考えられる．本研究では，その中でも，ビデオ中の歩行者数のリアルタイム推定に着目する．これまで，センサや専用ハードウェアを用いてリアルタイムに歩行者数を推定していたが，設置場所に限りがあった．また，ビデオストリームからデコード後の映像を処理して，歩行者数を推定する技術はあるが，いずれも一旦別処理で歩行者数の推定処理を行うので，リアルタイム処理には難しい技術的な課題があった．

そこで，ビデオストリームのビデオ符号化から歩行者数をリアルタイムに推定する．ビデオストリームを復号する際に抽出される予測残差信号からフレーム毎に回帰分析を用いて歩行者数を推定し，フレーム内の可変ブロックを参照して歩行者数を補正するところに特徴がある．従来手法と同等の性能が得られ，かつリアルタイムに推定で

きることを確認した。これら 2 つの研究の成果は、例えば、商業ビルなどに設置されたカメラで撮影されている映像をネットワークを介して収集し、それをを用いて監視やマーケティングに用いることが考えられる。地理的に離れた様々な箇所に設置されたカメラから映像を受信するためには、無線通信を含んだ様々なネットワークが用いられると考えられる。本論文の成果を用いることで、ネットワーク環境によらず高品質な映像を得ることができるようになる。さらに、得られた映像中の歩行者数に関する情報をリアルタイムに得ることができるため、イベント等における人流把握や警備計画の立案、また災害時の避難経路の確保などの道路環境の改善に活用できると考えられる。

## 謝辞

本研究は、大阪大学大学院情報科学研究科情報ネットワーク学専攻 松岡茂登教授の御指導のもとに進められました。

松岡教授には、素晴らしい研究の機会と充実した環境を与えて頂き、日頃から終始熱心に御指導を賜り、また温かく御激励頂きました。ここに改めて厚く御礼申し上げます。また、モバイルコンピューティング講座の東野輝夫教授、先進ネットワークアーキテクチャ講座の村田正幸教授、情報流通プラットフォーム講座の長谷川亨教授、インテリジェントネットワークング講座の渡辺尚教授をはじめ平成 25 年 3 月に退職されました同研究室中野博隆教授や長谷川准教授を含む諸先生方には、筆者の博士後期課程在学中に適切な御指導を賜り、また本論文をまとめるに際して懇切な御教示を賜りました。最後に、これまで研究指導や的確なアドバイスのみならず、研究者としての心構えについて御指導頂いた、株式会社 NTT ドコモの栄藤稔執行役員、山口博幸担当部長、倉掛正治担当部長を含む諸先輩方に深く感謝の意を表します。

# 目次

関連発表論文	i
内容梗概	iii
謝辞	V
第1章 序論	1
1.1 ビデオストリームに使われるビデオ符号化の概要	1
1.1.1 ビデオ符号化の概要	1
1.1.2 MPEG-4 の主な機能	3
1.1.3 H.264/AVC の主な機能	3
1.2 多重記述に使われているビデオ符号化の概要	5
1.2.1 異種無線アクセス網を組み入れた複合的な無線サービスの現状	5
1.2.2 複数通信経路に適した多重記述ビデオ符号化の概要	6
1.2.3 2章で説明する提案内容	7
1.3 歩行者数の推定に使われるビデオ符号化の概要	8
1.3.1 歩行者や自動車の追跡技術の現状	8
1.3.2 歩行者数の推定に使われるビデオ符号化の概要	8
1.3.3 3章で説明する提案内容	9
1.4 歩行者数の推定に使われるネットワークカメラ	9
第2章 多重記述のビデオ符号化を使った映像品質の向上	10
2.1 序言	10
2.2 時間スケラビリティ機能を用いた方式(Temporal MD)の評価	11
2.2.1 序言	11
2.2.2 提案方法	12
(1) 参照画像変更の概念	12
(2) 参照画像の変更処理	14
(3) ポストフィルタ	15
2.2.3 シミュレーションおよび結果	16
(1) 評価諸元	16
(2) 評価構成	17
(3) 実験結果	17
(4) 考察	25
2.2.4 まとめ	25
2.3 SNR スケラビリティ機能を用いた方式(Quantizer MD)の評価	26
2.3.1 序言	26
2.3.2 提案方法	27
(1) 量子化パラメータの操作方法	27
(2) 量子化パラメータの算出方法	28
(3) 参照画像の変更に起因する画質劣化の低減(ドリフト補償)	29
2.3.3 シミュレーションおよび結果	31
(1) 評価諸元	31

(2) 評価構成	32
(3) 実験結果	34
(4) 考察	40
2.3.4 まとめ	44
<b>第3章 ビデオ符号化に基づく歩行者数の推定</b>	<b>45</b>
3.1 序言	45
3.2 ビデオ符号化に基づく歩行者数のリアルタイム推定方式の評価	45
3.2.1 H.264/AVC の動き補償と歩行者数の相関	45
(1) H.264/AVC の可変ブロック構成	45
(2) 量子化パラメータによる予測残差信号の変化	46
(3) 予測残差信号と歩行者数の相関	46
3.2.2 提案方法	48
(1) 全体構成	48
(2) 前処理	49
(3) 歩行者数の推定処理	50
(4) 符号化に依存した抽出画総数の補正	51
3.2.3 シミュレーションおよび結果	52
(1) 評価諸元	52
(2) 評価構成	53
(3) 実験結果	54
(4)-1 可変ブロックによる補正に関する考察	55
(4)-2 他手法との比較に関する考察	62
3.3 まとめ	69
<b>第4章 結論</b>	<b>70</b>
<b>参考文献</b>	<b>73</b>

## 図リスト

図 1-1	ビデオ符号化技術の歴史と相関	2
図 1-2	異種無線アクセス網	5
図 1-3	Multiple Description Video Coding の構成	8
図 2-1	Multiple Description Video Coding の構成	11
図 2-2	MPEG-4 Scalable Coding の構成	11
図 2-3	参照画像変更の概念	13
図 2-4	参照画像変更処理	14
図 2-5	ポストフィルタの概念	15
図 2-6	映像ソースのスナップショット	16
図 2-7	評価に使った実験ネットワークの構成	17
図 2-8	Overtime の R-D 曲線(Packet Loss 0%)	18
図 2-9	Overtime の R-D 曲線(Packet Loss 1%)	18
図 2-10	Overtime の R-D 曲線(Packet Loss 3%)	19
図 2-11	Overtime の R-D 曲線(Packet Loss 5%)	19
図 2-12	News の R-D 曲線(Packet Loss 0%)	20
図 2-13	News の R-D 曲線(Packet Loss 1%)	20
図 2-14	News の R-D 曲線(Packet Loss 3%)	21
図 2-15	News の R-D 曲線(Packet Loss 5%)	21
図 2-16	Foreman の R-D 曲線(Packet Loss 0%)	22
図 2-17	Foreman の R-D 曲線(Packet Loss 1%)	22
図 2-18	Foreman の R-D 曲線(Packet Loss 3%)	23
図 2-19	Foreman の R-D 曲線(Packet Loss 5%)	23
図 2-20	Multiple Description Video Coding の構成	26
図 2-21	Quantizer MD の基本概念	27
図 2-22	真値の近似方法( <i>range1</i> と <i>range2</i> が重なり合った場合)	28
図 2-23	真値の近似方法( <i>range1</i> が <i>range2</i> に包含された場合)	29
図 2-24	映像ソースのスナップショット	31
図 2-25	復号化のブロック図	32
図 2-26	復号化のフローチャート	33
図 2-27	$D_1$ と $D_2$ および $D_0$ との PSNR 比較 (評価ビデオストリームは Foreman)	36
図 2-28	$D_1$ と $D_2$ および $D_0$ との PSNR 比較 (評価ビデオストリームは Mobile&Calendar)	38
図 2-29	$D_0$ における <i>range1</i> と <i>range2</i> の包含関係(Foreman)	41
図 2-30	$D_0$ における <i>range1</i> と <i>range2</i> の包含関係(Mobile&Calendar)	41
図 2-31	量子化パラメータを設定した場合の包含関係(Foreman)	43
図 2-32	量子化パラメータを設定した場合の包含関係(Mobile&Calendar)	43
図 3-1	SAD 総和値と歩行者数の相関	47
図 3-2	ビデオストリームのスナップショット	47
図 3-3	歩行者 1 人当たりの SAD 総和値	47
図 3-4	全体構成のフローチャート	48
図 3-5	前処理のフローチャート	49
図 3-6	歩行者数の推定	50

図 3-7 I フレームと P フレームの符号量の推移	51
図 3-8 評価ビデオストリームのスナップショット	52
図 3-9 各ビデオストリームの VBS on, VBS off と真値の比較	59
図 3-10 Albiol, Conte と VBSon との MAE 比較	64
図 3-11 評価したビデオストリームのスナップショット	65
図 3-12 各フレームの処理時間	66
図 3-13 ビデオストリーム(a)(b)の各手法における歩行者数の推定と真値の比較	67
図 3-14 実際の歩行者数に対する推定した歩行者数の比較	68
図 4-1 研究成果の利用イメージ	71

## 表リスト

表 2-1 評価の諸元	11
表 2-2 評価ビデオストリームも諸元	31
表 3-1 PETS2009 データセット	52
表 3-2 ビデオ符号化方式の条件	53
表 3-3 MAE/MRE 比較結果	54
表 3-4 提案手法の処理時間	54

## 第1章 序論

本論文は、サーバからデータを受信しながら映像を再生するビデオストリームには、ビデオ符号化が用いられており、ビデオストリームの低ビットレート化および高品質化の重要な技術要素となっている。一方、ネットワーク環境は、セルラーや無線 LAN などの異種無線アクセス網を組み入れた複合的無線サービスが普及してきている。そのような複合的無線サービスでは、複数の通信経路を持つことが可能となるので、ネットワークの負荷分散、通信障害耐性向上などに効果が期待されている。このような複数チャンネルを有するネットワーク環境によって、ビデオストリーム品質が向上してきている。一般的には、同じ品質の映像を異なるネットワークで 2 重化して配信することで、映像配信の信頼性や品質を確保する完全冗長 2 重化方式[1]が使われていることが多い。しかし、ネットワークが異なると帯域などのネットワーク特性が異なるため、それらのネットワークを用いて同じ品質の映像を配信するには、遅延が発生して、リアルタイム性を損なう問題がある。

そこで、複合的無線サービスに適した多重記述のビデオ符号化を用いた映像品質の向上を目的とした。ビデオストリームに対して多重記述ビデオ符号化を行うことで、1 つの映像ソースを 2 つのエンコーダによってそれぞれ送信し、相互補完することで途切れの無いデータ通信が行われ、さらに 2 つのチャンネルが受信できれば、高品質なデータ通信が行われる。次に、高品質なビデオストリームの受信ができるようになると、受信したビデオデータを解析することによって様々なアプリケーションが実現すると考えられる。本研究では、その中でも、ビデオ中の歩行者数のリアルタイム推定に着目する。これまで、センサや専用ハードウェアを用いてリアルタイムに歩行者数を推定していたが、設置場所に限りがあった。また、ビデオストリームからデコード後の映像を処理して、歩行者数を推定する技術はあるが、いずれも一旦別処理で歩行者数の推定処理を行うので、リアルタイム処理には難しい技術的な課題があった。

そこで、ビデオストリームのビデオ符号化から歩行者数をリアルタイムに推定する。ビデオストリームを復号する際に抽出される予測残差信号からフレーム毎に回帰分析を用いて歩行者数を推定し、フレーム内の可変ブロックを参照して歩行者数を補正するところに特徴がある。従来手法と同等の性能が得られ、かつリアルタイムに推定できることを確認した。

本章では、ビデオストリームに含まれるビデオ符号化の概要を述べる。

### 1.1 ビデオストリームに使われるビデオ符号化の概要

#### 1.1.1 ビデオ符号化の概要

近年、インターネットの発展によってサーバからデータを受信しながら映像を再生するビデオストリームを使った様々な映像配信サービスが発展してきた。インターネットの急速な普及により、文字や写真だけでなく映像の利用も活発になり、インターネット経由で世界各地からのニュース映像やコンサート中継が視聴できるようになった。従来は、デジタル映像のファイルをパソコンにダウンロードした後、再生する必要があったため、文字や静止画に比べてデータ量の大きなデジタル映像のダウンロードに時間が掛かっていた。そのため、ビデオストリーミング技術が開発され、待ち時間なしに映像の視聴ができるようになった。ビデオストリームは、テレビ映像を単純にデジタル化すると数 100Mbps もの帯域が必要となる。しかし、インターネットで利用できる帯域は数百 k~数 Mbps であった。そこで、ビデオ符号化技術により 1/100 程

度に圧縮を実現することで狭い帯域でも滑らかな動きやノイズの少ないビデオ符号化技術の開発が行われてきた。

図 1-1 に、これまでのビデオ符号化方式の歴史と相関を示す。1990 年に ITU-T の国際標準機関において、世界最初のデジタル動画像圧縮方式として H.261 が標準化された。この時、採用された圧縮技術として動き補償，離散コサイン変換，量子化の組み合わせがあり，以降の標準化作業における基礎となっている。ISO/IEC 国際標準である Moving Picture Experts Group(以下，MPEG)においては，コンパクトディスク等の蓄積媒体に 1.5Mbps の符号化レートで蓄積することを目的として標準化された MPEG-1[4]，放送，通信などの領域に適用されることを目的として標準化された MPEG-2[5]がある。さらに 1996 年に標準化された H.263[6]は，従来の MPEG-1，MPEG-2 の領域とは異なり，インターネットを使った映像配信サービスを想定した低ビットレートに向けて標準化された。1998 年に標準化された MPEG-4[2]は，低ビットレートを維持しつつ，H.263 に 1/4 画素動き補償や B フレーム採用により高画質化を目指すと共に誤り耐性技術が導入されているので移動通信のような無線区間でエラーが発生するネットワークに適したビデオ符号化として標準化された。2003 年には，ITU-T の Video Coding Experts Group(以下，VCEG)と MPEG の 2 つの団体が共同で H.264/AVC[3]の標準化を行った。MPEG-4 に複数参照フレーム予測や可変ブロック機能，イントラフレーム内予測の採用により MPEG-4 と比較して，2 倍以上の圧縮率で同程度の画質を実現できるビデオ符号化を目的に標準化された。

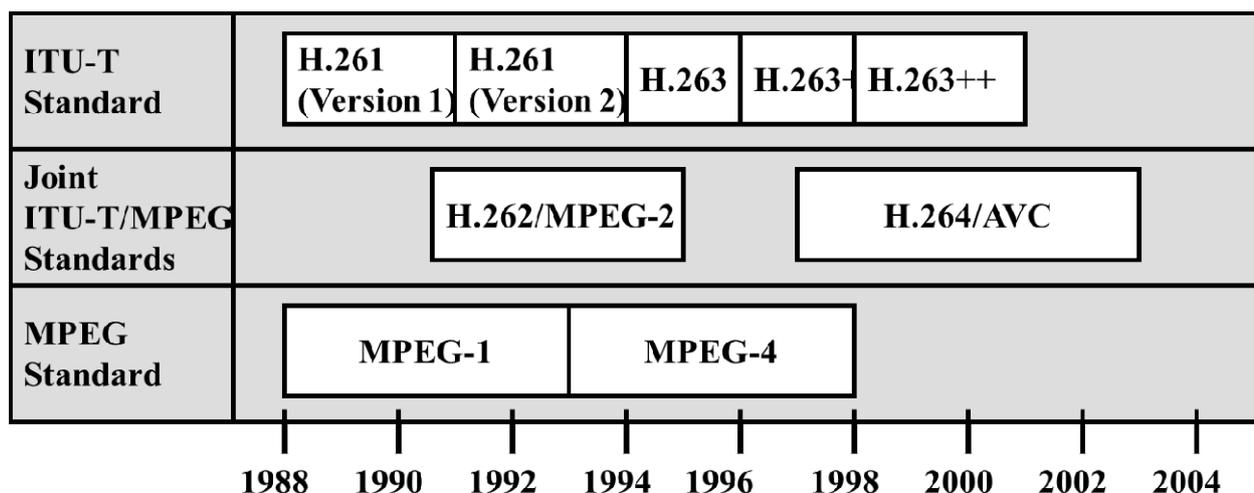


図 1-1 ビデオ符号化技術の歴史と相関

次に，本章での研究対象となる MPEG-4 および H.264/AVC の機能について述べる。

### 1.1.2 MPEG-4 の主な機能

MPEG-1 ではビデオ CD, MPEG-2 では放送や HDTV での使用を想定しているのに対して, MPEG-4 では低ビットレートでの使用にまで用途を拡大することを目標として規格化が開始された. 圧縮アルゴリズムの基本原理は, MPEG-1, MPEG-2, H.263 などと基本的には同様であり, 空間変換やフレーム間予測, 量子化, エントロピー符号化を採用している. MPEG-4 の大きな特徴として高能率符号化と誤り耐性能力の向上である. 高能率符号化については, 当時急速に普及していたインターネットを使った映像配信サービスへの応用を想定しており, 当初の転送レートは, 数百 k~数 Mbps 程度なので, これまでのビデオ符号化技術では送信できなかった. そのため, 高能率な圧縮技術が求められ, 先に規格化が進んでいた H.263 を基に幾つかのツールを追加した構成になっている. 特に符号化効率の改善に効果があったツールとしては, 8×8 ブロック動き補償, 直接予測, イントラ予測などがある. 次に誤り耐性能力の向上については, MPEG-4 の主要アプリケーションは移動体通信を想定しており, データは無線区間を通信することになり, 外部ノイズによるデータ欠損対策が重要となる. MPEG-4 で採用された手法は, 巡回冗長検査やリードソロモン符号のようなエラー訂正符号のアプローチではなく, エラー隠蔽の技術であった. つまり, ビデオストリームにエラーが混入しても, 見た目にその影響をわからなくする技術である. このために生成ビット長を基準としたパケット分割, データ・パーティショニング, スタッフィング・バイトの改善などが行われた. 最後に H.263 との相違点は, フレーム間予測における B フレームの採用, 離散コサイン変換(Discrete Cosine Transform 以下, DCT)係数のイントラ予測の導入などが挙げられる. また, MPEG-4 の利用シーンとして, ネットワークを介して受信機に伝送され, 受信機で復号される. ネットワークには, 高レートの伝送路である光ファイバや低レートの伝送路である無線通信など, 多様な伝送レートの通信路が混在している. このように異なるレートが混在するネットワークで映像データを伝送するためにスケーラブル映像符号化 (Scalable Video Coding 以下, SVC) 方式が準備されている. この方式は, SVC 方式に対応した符号化器によって, 基本レイヤと拡張レイヤに符号化し, 基本レイヤのみを復号すると低品質の映像が復号され, 基本レイヤと拡張レイヤを復号すると高品質の復号映像が得られる符号化方式である. SVC 方式の特徴として, 受信機側の符号化器が 1 つで済むこと, 映像配信用にサーバの容量が少なく済むというメリットがある. デメリットとしては, SVC 方式に対応した特別な復号器が必要な場合があることである. SVC 方式には, 「時間スケーラビリティ機能」, 「空間スケーラビリティ機能」, 「SNR スケーラビリティ機能」の 3 つが規定されている.

### 1.1.3 H.264/AVC の主な機能

圧縮アルゴリズムの原理は, 従来方式の MPEG-4 などと基本的に同様で, 空間変換やフレーム間予測, 量子化, エントロピー符号化を採用している. H.264/AVC ではこれらのツールに対して非常に多数の改良が施されており, 算術符号化やフィルタなどのツールも追加されている. さらに, 画像特徴に応じて多彩なモードを適応的に使い分けることで, 従来方式を遥かに凌ぐ圧縮効率を達成している. 従来規格である MPEG-1, MPEG-2 や H.261 では 16×16 画素, H.263, MPEG-4 では 8×8 画素のマクロブロック(Macro Block 以下, MB)を単位として, 原画像ないしフレーム間予測の予測誤差画像の DCT 係数を求め, その係数を量子化している. このとき, コサイン関数を用いるため, 実数精度の演算が必要となる. これに対し H.264/AVC では, 16 ビット整数精度で演算が可能な整数変換を採用している. この整数変換は, 加減算とビットシフトのみによって演算可能となるように設計されているため, ソフトウェア, ハードウェアいずれの場合でも実装が非常に容易となる. これまでのビデオ符号化ではフレーム

間予測において参照フレームとして指定できるフレームは、P フレームについては直前の I, P フレーム、B フレームについては直前および直後の I, P フレームに固定されている。H.264/AVC では、複数の参照フレームを持つことによって、例えばシーンチェンジや移動物体を考慮してより前のフレームを参照フレームとして指定することが可能となっている。また、B フレームについては未来方向のフレームを使わずに過去の 2 フレームを参照フレームとして指定したり、別の B フレームを参照フレームとして指定することが可能となっている。動き補償の単位は  $16 \times 16$  画素のマクロブロックが基本であり、H.263 および MPEG-4 においては  $8 \times 8$  画素ブロック単位の動き補償も利用できた。H.264/AVC ではさらに単位ブロックサイズを追加し、 $16 \times 16$ ,  $16 \times 8$ ,  $8 \times 16$ ,  $8 \times 8$  の 4 種類から選択可能となっている。さらに、 $8 \times 8$  画素ブロックについては、 $8 \times 8$ ,  $8 \times 4$ ,  $4 \times 8$ ,  $4 \times 4$  の 4 種類のサブマクロブロック(Sub Macro Block 以下, SubMB)分割も指定できる。このように多数のブロックサイズを利用することで、形状や動きに適したブロックから予測が可能である。これは、原理的には符号化効率が上がることとなる。ただし、サブブロックを指定することは余分なヘッダが付加されることになり、これがオーバーヘッドとなって符号化効率に影響を与える可能性もある。シーンに適した動き補償ブロックサイズを選択することが、エンコーダには求められる。

## 1.2 多重記述に使われているビデオ符号化の概要

### 1.2.1 異種無線アクセス網を組み入れた複合的な無線サービスの現状

近年、セルラーや無線 LAN、WiMAX など、多種多様な無線アクセス網が確立されている。種々の無線アクセス網を個々に利用するだけでなく、これらのシステムを統合し相互連携した異種無線アクセス網の構築が考案されている(図 1-2)。そのような複合的な無線サービスでは、複数の通信経路を持つことができるので、ネットワークの負荷分散、通信障害への耐性向上、映像品質の向上などに効果が期待されている。一般的には、複数のビデオストリームに分けてビデオデータを送信するので、その方式と問題点をまとめた。

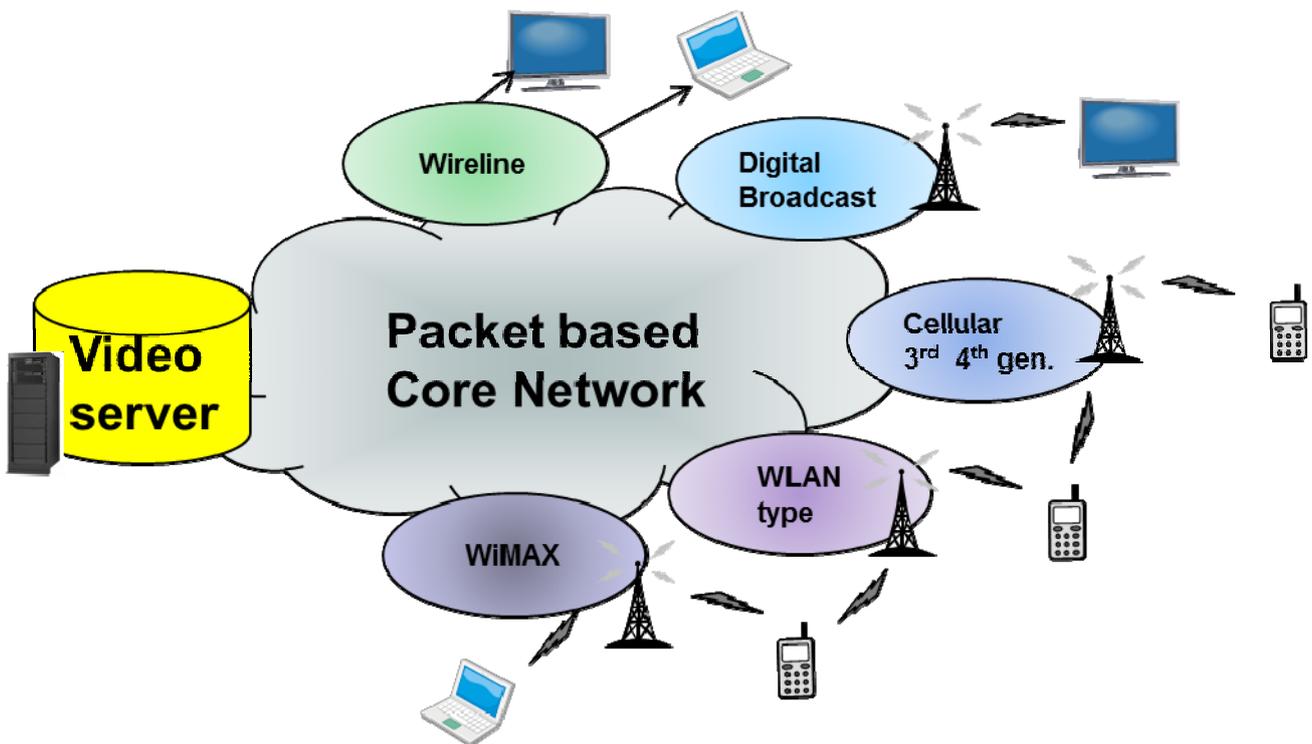


図 1-2 異種無線アクセス網

#### ①完全冗長 2 重化方式

完全冗長 2 重化方式は、現用チャンネルと予備チャンネルに同じビデオデータを送信し、どちらのチャンネルで受信してもビデオが再生できる方式である。映像配信サービスの高信頼性確保のため、同じビットレートで同じ映像品質を確保することが多い。しかし、通常の利用において、両チャンネルとも正常な時間が多いため、片チャンネルの伝送容量が無駄になってしまう問題がある。

#### ②MPEG-4 などのスケーラビリティ機能

2 つのビデオストリームをそれぞれに基本レイヤと拡張レイヤと定義し。基本レイヤでは、低品質の画像が復号できるもので、拡張レイヤでは、基本レイヤの品質を補強するデータのみを送信する。つまり、基本レイヤと拡張レイヤが受信できれば、より品質の高い映像が復号できるメリットがある。しかし、拡張レイヤだけでは、復号できない制限があるため、仮に基本レイヤにパケットエラーが発生した場合は、送信側に通知して基本レイヤと拡張レイヤのリンクを入れ替える必要があり、既存の伝送構成

に大幅な変更を必要とする。なお、時間解像度が高い復号画像を得る機能を有する時間スケーラビリティ機能、空間解像度が高い復号画像を得る機能を有する空間スケーラビリティ機能そして量子化の工夫により画品質が高い復号画像を得る機能を有するSNR(signal-noise ratio)スケーラビリティ機能がある。

③多重記述符号化(Multiple Description coding 以下, MD)

MD は、符号化方式の工夫により片チャンネルだけでも復号を可能とし、両チャンネルが揃えば高品質な映像が復号できる方式である。これは、②で述べた時間の工夫、空間の工夫、量子化の工夫がされていて、時間の工夫であれば、例えば偶数フレームと奇数フレームをそれぞれ符号化したり、空間の工夫であれば、解像度の違う映像を符号化したり、量子化の工夫であれば、量子化パラメータを変えた映像を符号化して、両チャンネル受信できれば高品質な映像が復号できる方法である。しかし、スケーラビリティ機能と比べて符号量が多くなる問題がある。

1.2.2 複数通信経路に適した多重記述ビデオ符号化の概要

MD は、複数チャンネルが存在するネットワーク上で使われることを想定しており、1つの映像ソースを2つのエンコーダによってそれぞれ符号化し、2つのチャンネルにそれぞれを転送し、それぞれのデコーダが独立して復号できるため、例えば1つのチャンネルがパケットロスでデータが欠落しても、もう片方のデコーダで再生できるメリットがある。MD に適した冗長構成なビデオ符号化方式として多重記述ビデオ符号化(Multiple Description Video Coding 以下, MDVC)がある。

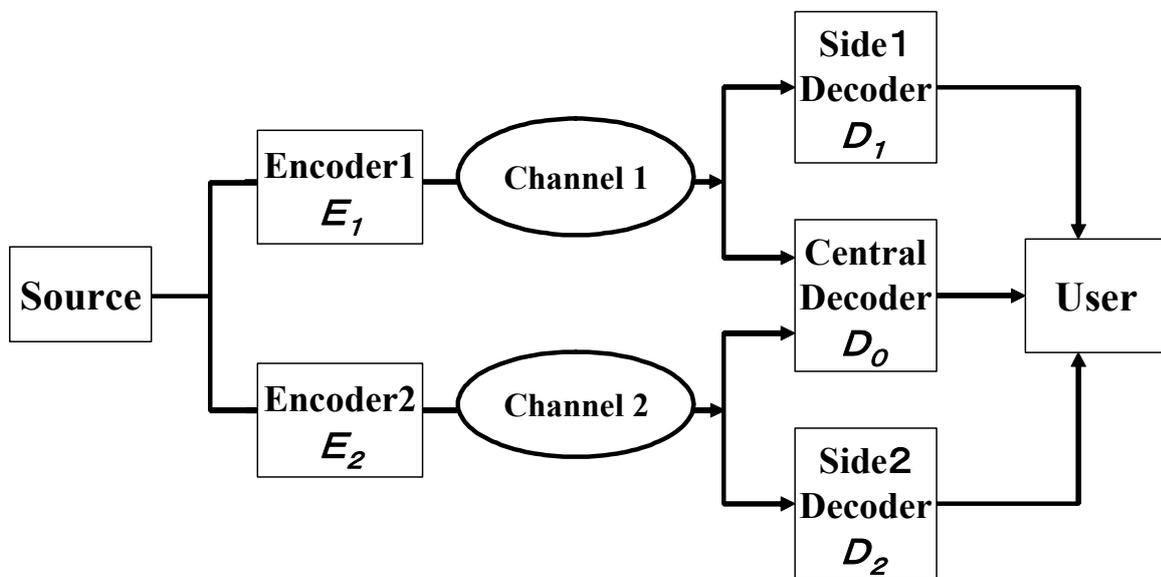


図 1-3 Multiple Description Video Coding の構成

MDVC は、1980 年代に考案された方式[7]で、その基本構成を図 1-3 に示す。情報源から入力した信号は、Encoder1 (E<sub>1</sub>)および Encoder2(E<sub>2</sub>)でそれぞれ符号化し、異なるチャンネルを使って伝送するものである。例えば、チャンネル 1 として無線 LAN、チャンネル 2 として 3G セルラーを使う。受信側では Side1 Decoder (D<sub>1</sub>)および Side2 Decoder (D<sub>2</sub>)それぞれで復号し、相互補完することで途切れの無い通信が行われ、さらに 2 つのチャンネルが受信できれば、Central Decoder (D<sub>0</sub>)で高品質なデータ通信が行われる。このことは、無線区間や各マルチメディアゲートウェイ等で発生する帯域制限やパケットロスに対しても頑健で映像品質の向上に貢献する方式と考えられている。

MDVC の代表的な方式には、スカラー量子化やベクトル量子化を使った方法があり、前者は、両チャンネルで有効なデータが受信できた時は、量子化誤差を平均して最小にするような符号化器の検討[8]、可変長符号を使った符号化器によりエントロピーを制限する方法の検討[9]などが報告されている。また、後者を使った方法としては、符号帳のラベリングを最適化した方法[10][11]やチャンネル毎に歪みの重み付けを行う方法[12][13]などが報告されている。さらに、ベクトル量子化をツリー構造で検討[14]したものがある。一般的には、映像配信の高信頼性確保のため、予備チャンネルなどのバックアップ回線を使って、2チャンネルとも同じビットレート、同じ映像品質で送信する完全二重化方式は、画質向上に効果がある。しかし、本章は、セルラーや無線 LAN など異種無線アクセス網を使ったマルチホーミングを想定しており、通常、各無線アクセス網のビットレートがそれぞれ異なる環境であるため、同じ符号量を送信する完全二重化を想定する手法の適用は難しい。また、異なるビデオストリームを想定した MDVC にスカラー量子化を用いた研究は、2つのビデオストリームを比較して、量子化誤差を平均して最小とする設計で高画質化が行われてきた。しかし、平均化されたビデオストリームの量子化ステップは、符号化された時の量子化誤差を含んだままなので、復号しても量子化誤差は改善されない課題があった。

### 1.2.3 2章で説明する提案内容

第2章では、多重記述に関連する研究として、無線区間での映像配信サービスに適した MPEG-4 を使って、複合的な無線サービスに適した多重記述のビデオ符号化の映像品質の向上を明らかにする。異なるビデオストリームの量子化後の DCT 係数から、量子化前の DCT 係数値(以下、真値)が取り得る範囲を絞り込み、それらの DCT 係数を利用できるように量子化パラメータを操作する手法を提案する。量子化パラメータの操作を DCT 係数領域内で行うため、復号時に参照画像の不一致によるドリフト誤差が生じ、画質劣化を引き起こす恐れがある。そこで、量子化パラメータを操作するにあたっては、ドリフトによる影響も考慮している。

本章は、セルラーや無線 LAN などの無線アクセス網のビットレートがそれぞれ異なる環境であることを想定し、異なるビットレート、異なる映像品質である不均質ネットワーク環境にスコープを当て、携帯電話で一般的なビデオ符号化方式である MPEG-4 を用いて、従来の符号化アルゴリズムを変更することなく、異なる量子化操作を合成して PSNR(Peak signal-to-noise ratio)を向上させる MDVC 方式を提案する。実際に MPEG-4 のシミュレーションを行って、PSNR が向上することが確認できたので、2章で詳細な結果を述べる。なお、量子化マトリクス(低周波数成分に比べて高周波数成分を粗い量子化で行い画質を改善するビット割り当ての重み付けマップ)を事前に作成し、送受信器で共有する方法も考えられるが、MPEG-4 のアルゴリズムを変更することなく、リアルタイム通信の実現を目的にしているため、この方式は対象外とする。

### 1.3 歩行者数の推定に使われるビデオ符号化の概要

#### 1.3.1 歩行者や自動車の追跡技術の現状

従来、交差点の改良や交通流の改善などのため複数の歩行者や自動車を個別に追跡する技術が開発されてきた。最近では、大きな人の流れの制御や混雑状況の監視などのため、多数の歩行者が存在する環境下でその概数を推定する技術の開発が注目されている。このような追跡・推定方法として、一般的に、画面内の移動物体と背景領域を分離する背景差分法[15]-[18]とオプティカルフローを用いて歩行者の動きを検知・追跡し歩行者数を推定する手法[19]-[21]がある。背景差分法は、予め移動物体の存在しない背景画像を取得しておき、入力画像と背景画像との差分処理結果に応じた 2 値化を行なうことによって移動物体領域を抽出する手法である。個々の物体全体をかたまりとして抽出することができるので物体抽出性は高い。しかしその反面、物体の中で背景濃度に近い部分が欠落する等の問題がある。移動物体の存在しない背景画像を取得するのが難しい状況や背景自体が変化する状況においては背景画像を画像処理によって生成する必要がある。この背景画像を精度良く生成できれば背景差分の精度も向上すると考えられる。オプティカルフローは、フレーム間のピクセルレベルでの対応を求めるものであり、オプティカルフローが得られれば、各部分の詳細な動きがわかる。オプティカルフローを求める手法をまとめた。

##### ①ブロッキングマッチング法

ブロックマッチング法では、画像中のあるブロックをテンプレートとして、次時間の画像中からマッチする個所を探索することでフローを求める。適切な特徴を持つ画像においては、確実にフローを計算できるが、比較的計算時間がかかる。また、回転、スケーリングに対するロバスト性は、テンプレートとなるブロックのサイズや画像の特徴に依存する。

##### ②勾配法

勾配法では、「オプティカルフロー拘束方程式」と呼ばれる、輝度の時間や勾配の拘束方程式を用いて、これに制約条件を付加することでフローを求める。比較的高速に全画素について空間の変化を計算できるが、前提条件に合わない個所(急激な輝度変化、ノイズ)では、著しい誤差が発生する事がある。

#### 1.3.2 ビデオ符号化を用いた移動物体の推定の概要

近年、監視カメラの映像は MPEG-4 や H.264/AVC のような符号化方式を使って圧縮し、蓄積されることが多くなってきた。これら圧縮されたビデオストリームを用いて移動物体を解析することで、解析データの前処理が不要となり、処理量が大幅に削減され、リアルタイムに解析できる特徴がある。そのビデオストリームから得られるパラメータを用いて移動物体を追跡・推定する方法も提案されている。H.264/AVC の動きベクトル(Motion Vector 以下, MV)をクラスタリングして移動物体をリアルタイムに追跡する例[22], 複数フレームにわたる類似した MV を識別して移動物体を追跡する例[23][24], MPEG の MV と DCT 係数の成分比較から移動物体を追跡する例[25]-[28], MV と画像の色情報からオブジェクトを追跡する例[29]が報告されている。追跡については、モデル化した歩行者の外観と画像の相関関係から歩行者を抽出する例[30][31][32]も多い。そのほか、ある領域の色成分の割合を事前に設定したテンプレートと比較して追跡対象を抽出し、その位置や動きを推定する例[33], 空間の色情報分布からオクルージョンの影響を予測し、追跡精度を向上する例[34]がある。歩行者数の推定では、移動するオブジェクトやそのエッジから抽出した画素をクラスタリングし、得られたクラスタ数から歩行者数を推定する例[35]-[39]がある。コーナー検出法[40]に基づいて集計された MV を用いて移動物体を識別し、それらから統計的に推定する方

法[41]、近傍の輝度勾配情報から抽出した特徴がある点[42]を用いて推定する例[43]が報告されている。また、統計手法としては、回帰分析を用いて行われることが多く、テンプレート画像とのマッチングにより抽出された情報源を用いて人数推定する方法[44]や歩行者の特徴量に基づいてトレーニングデータを切り出し、ガウス過程により推定している方法[45]がある。これらは、異なるコンテンツに対応したテンプレート画像の精度やトレーニングデータの特徴量を抽出する精度に大きく依存する。また、歩行者の領域を抽出し、重みパラメータによって特徴量を効率的に選出した情報源を用いて人数を推定する方法[46]がある。これまで紹介した手法は、一旦別処理で特徴点を抽出後、歩行者数の推定を行う手法や特徴点を示す学習データが増えるコンテンツを扱う手法は多くの計算量が必要となり、リアルタイムに歩行者数の推定が難しい技術的な課題があった。最近では、複数のステレオカメラを用いて多視点から人物位置や色情報から類似度を考慮したクラスタリングを適用する手法[47]やステレオカメラの差分情報から人物を特定する手法[48][49]が報告されている。しかし、これらはカメラ位置を特定したり、画像処理に必要な実空間内の3次元情報を獲得するためキャリブレーションを行ったり、オクルージョンを回避するために、歩行者の上部から撮影するなどカメラの設置や機能に工夫が求められる。

### 1.3.3 3章で説明する提案内容

第3章では、H.264/AVCによって符号化されたビデオストリームを復号する際に抽出される予測残差信号からフレーム毎に回帰分析を用いて歩行者数を推定し、フレーム内の可変ブロックを参照して歩行者数をリアルタイムに補正する手法を提案している。これまでの手法ではリアルタイムに歩行者数の推定が難しい課題に対して、予測残差信号と可変ブロックを組み合わせることで、オクルージョンにも対応でき、従来の手法と比べて精度は維持しつつ、リアルタイムに歩行者数が推定できることを明らかにする。

## 1.4 歩行者数の推定に使われるネットワークカメラ

ネットワークカメラと画像解析を使った顔認証アプリ、人流計測、異常行動検知といった多様なアプリが開発されてきている。それらはカメラメーカーばかりではなく、ITやソフトウェア業界のプレーヤによって開発されることが多く、ネットワークカメラ市場は、世界のICT産業全体を巻き込んだ巨大な市場となっている。代表的な画像解析ツールとしては、輝度勾配方向共起ヒストグラム(CoHOG)特徴量[50]の抽出や線形サポートベクタマシン(SVM)[51]を搭載しているものが多い。これらは、簡易に画像解析結果が得られるが、用途を限定して利用することが多い。例えば、セキュリティ用途では、顔から個人を識別する必要があるため、近距離に合わせて画像処理を調整する必要があり、マーケティング用途では、人物の動線情報を得るために、オクルージョンが発生しないように、歩行者の上部から撮影する必要がある。

本章は、以下の4章から構成する。第1章は、序論であり、筆者が研究した2つの成果に関連しているビデオストリームのビデオ符号化の概要、多重記述に使われているビデオ符号化の概要および歩行者数の推計に使われる一般的な手法の概要について説明する。第2章は、複数の通信経路に適したMDVC情報から映像品質を向上させる手法を提案し、その性能を明らかにしている。第3章は、ビデオ符号化から統計処理によって歩行者数を推定する手法を提案し、その性能を明らかにしている。第4章は、今後のビデオ符号化を用いた研究の課題と展望について述べるとともに本研究成果の寄与について考察する。

## 第2章 多重記述のビデオ符号化を使った映像品質の向上

### 2.1 序言

3GPP(3<sup>rd</sup> Generation Partnership Project)の ALL-IP 作業委員会において、移動通信区間の ALL-IP が検討されている[52]. これが実現するとモバイルネットワークにおいてパケット通信を使った映像配信サービスが可能となる. しかし、無線区間や各マルチメディアゲートウェイ等で発生する帯域制限やパケットロスに対して、ビデオ符号化方式の誤り耐性技術だけでは映像品質の向上には限界がある. 特に従来から規格化されているスケーラブル映像符号化 (Scalable Video Coding 以下, SVC) 方式を用いて映像配信サービスを行っている場合、基本レイヤの映像データが受信機で正常に受信できないと拡張レイヤの映像データも復号できない仕様となっているため映像配信サービスが停止する場合も想定される.

そこで、本章では、MDVC を拡張する提案を行う. MDVC は、1 つのチャンネルがパケットロスでデータが欠落しても、もう片方のデコーダで再生できるメリットがある. 本章では、時間スケーラビリティ機能と SNR スケーラビリティ機能について、MDVC を用いた 2 つの拡張方式を提案し、いずれの方式も既存の MPEG-4 で規定されているスケーラビリティ機能と比較することで優位性を検証する. なお、空間スケーラビリティ機能は、高解像を実現するために拡張レイヤに多くのビットレートを割り当てることが多く、そのため利用シーンとしては、デジタルシネマ等の高解像度向けに適用されることが多い[53]. 本章は、モバイルネットワークを想定しているため、高解像度向けの拡張レイヤのビットレートは確保できないと思われる. そのため、空間スケーラビリティ機能は本章の対象外とした.

## 2.2 時間スケーラビリティ機能を用いた方式(Temporal MD)の評価

### 2.2.1 序言

MD(Multiple Description)は、1つの映像ソースを2つのエンコーダによってそれぞれ符号化し、2つのチャンネルにそれぞれを転送し、それぞれのデコーダが独立して復号できるため、例えば1つのチャンネルがパケットロスでデータが欠落しても、もう片方のデコーダで再生できるメリットがある。MDに適した冗長構成なビデオ符号化方式として多重記述ビデオ符号化(Multiple Description Video Coding 以下、MDVC)がある。

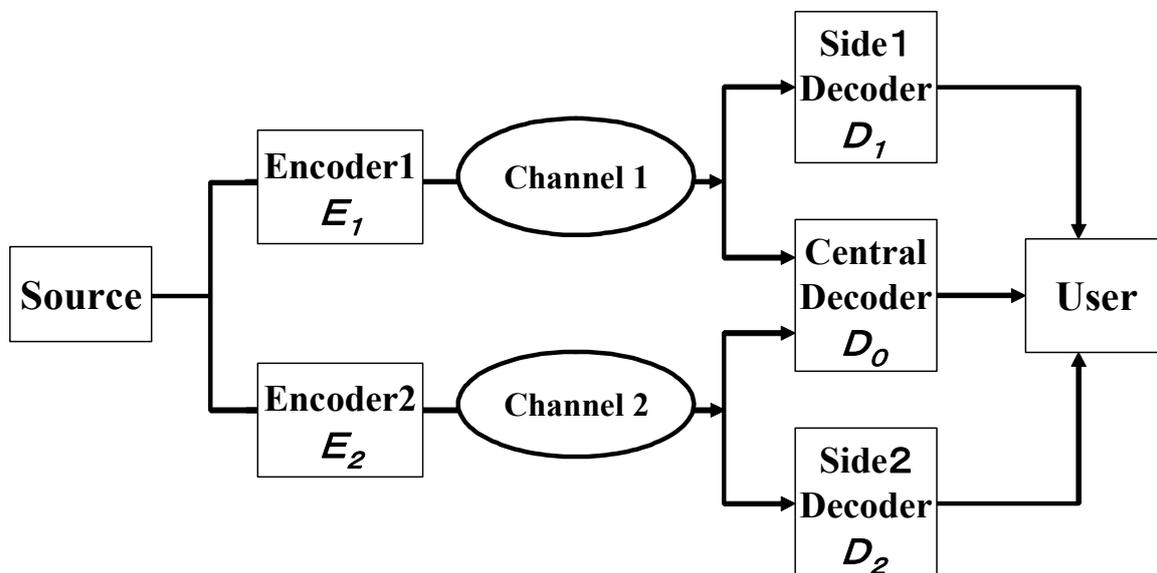


図 2-1 Multiple Description Video Coding の構成

MDVC の基本構成を図 2-1 に示す。情報源から入力した信号は、Encoder1( $E_1$ )および Encoder2 ( $E_2$ )でそれぞれ符号化し、異なるチャンネルを使って伝送するものである。例えば、チャンネル 1 として無線 LAN、チャンネル 2 として 3G セルラーを使う。受信側では Side1 Decoder ( $D_1$ )および Side2 Decoder ( $D_2$ )それぞれで復号し、相互補完することで途切れの無い通信が行われ、さらに2つのチャンネルが受信できれば、Central Decoder ( $D_0$ )で高品質なデータ通信が行われる。

図 2-2 に MPEG-4 の時間スケーラビリティ機能の構成図を示す。

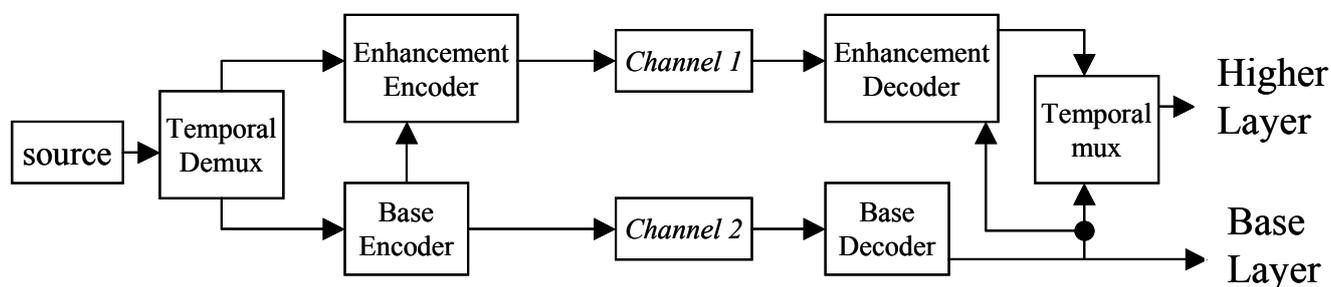


図 2-2 MPEG-4 Scalable Coding の構成

図 2-2 は、情報源から入力した信号は、Temporal Demux で2つに分岐され、基本レイヤと拡張レイヤとして、それぞれ Base Encoder と Enhancement Encoder でエンコードされる。この時、拡張レイヤでは、基本レイヤか拡張レイヤのうち、符号化対象フレームの時刻と最も近いフレームを参照画像として選択する。その後、伝送路をそれぞ

れの Channel で転送し、基本レイヤは、Base Decoder、拡張レイヤは Enhancement Decoder でデコードされ、拡張レイヤが正常に受信できれば、Temporal mux で高品質な復号化が行われる。しかし、MPEG-4 の時間スケラビリティ機能は、拡張レイヤの参照画像だけでなく、基本レイヤの参照画像も帯域制限やパケットロスで復号できないと拡張レイヤの復号ができない問題がある。一方、MDVC は、図 2-1 で説明したように、複数のチャンネル上で使われることを想定していて、1 つの映像ソースを  $E_1$  と  $E_2$  によってそれぞれ符号化し、2 つのチャンネルにそれぞれを転送し、それぞれのデコーダが独立して復号できるため、例えば 1 つのチャンネルがパケットロスで欠落しても、もう片方のデコーダで再生できるメリットがある。

本節は、パケットロスが発生するネットワーク環境において、MDVC の時間スケラビリティ機能を拡張する手法(以下、Temporal MD)を提案する。この Temporal MD の評価は、MPEG-4 Simple Profile と MPEG-4 Simple Scalable Profile とで比較を行った。

## 2.2.2 提案方法

### (1) 参照画像変更の概念

図 2-1 の  $E_1$  と  $E_2$  にそれぞれに時間のずれた映像ソースを入力する。例えば  $E_1$  に奇数フレーム、 $E_2$  に偶数フレームを入力する。受信側では、図 2-1 に示したように、両信号は  $D_1$  および  $D_2$  で復号され、それぞれで表示することが可能で、さらに  $D_1$  と  $D_2$  において正常に映像データが受信できれば、 $D_0$  においてフレームレートの向上が可能となる。この時、パケットロスが発生するネットワーク環境において、図 2-2 のような Temporal MD を配信した場合、パケットロスしたデコーダの映像は、パケットロスによりエラーが伝播するので復号できたデコーダの映像のみを表示することが一般的である。

そこで、 $E_1$  と  $E_2$  の画像が時系列的に似通っていることに着目すると、パケットロスが発生したデコーダの画像を復号する際、正常に復号できたデコーダの画像を参照することでパケットロスしたデコーダの画像を正常に復号させ、エラーの伝播を防止することが可能となる。図 2-3 にその概念を示す。上段は  $E_1$  でエンコードした後の受信側のイメージ。下段は、 $E_2$  でエンコードした後の受信側のイメージ。I はフレーム内符号化画像、P はフレーム間符号化画像を示す。数字はフレーム番号を表す。例えば、フレーム数を 6 とした場合、仮に 4 がエラーで復号できないときは、時系列的に近い 5 を参照画像に切り替えて、6 を復号する。

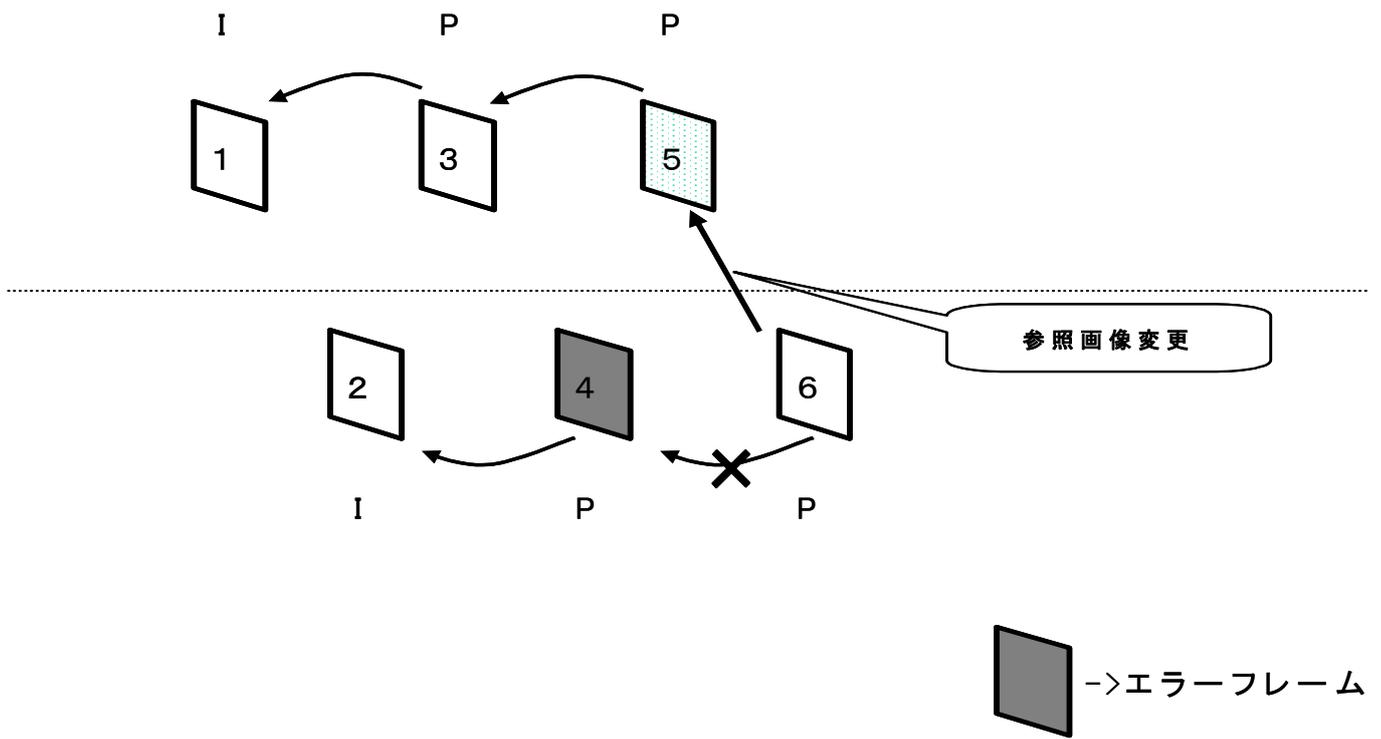


図 2-3 参照画像変更の概念

## (2) 参照画像の変更処理

図 2-4 に参照画像変更処理について示す。

フレーム番号 P3 と P5 は  $E_1$  でエンコードした後の受信側のイメージで、P5 は P3 を参照して画像を生成している。P2, P4, P6 は、 $E_2$  でエンコードした後の受信側のイメージで、P6 は、P4 を参照して画像を生成している。図 2-4 の P4 フレーム内にある MB(Macro Block)が、パケットロスにより消滅した場合、P6 フレームのブロックは参照画像が無いために映像が復号できない。また、そのブロックが復号できないだけでなく、その後そのブロックが参照されていれば、その後の画像も復号できないためエラーが伝播して品質劣化の要因となる。

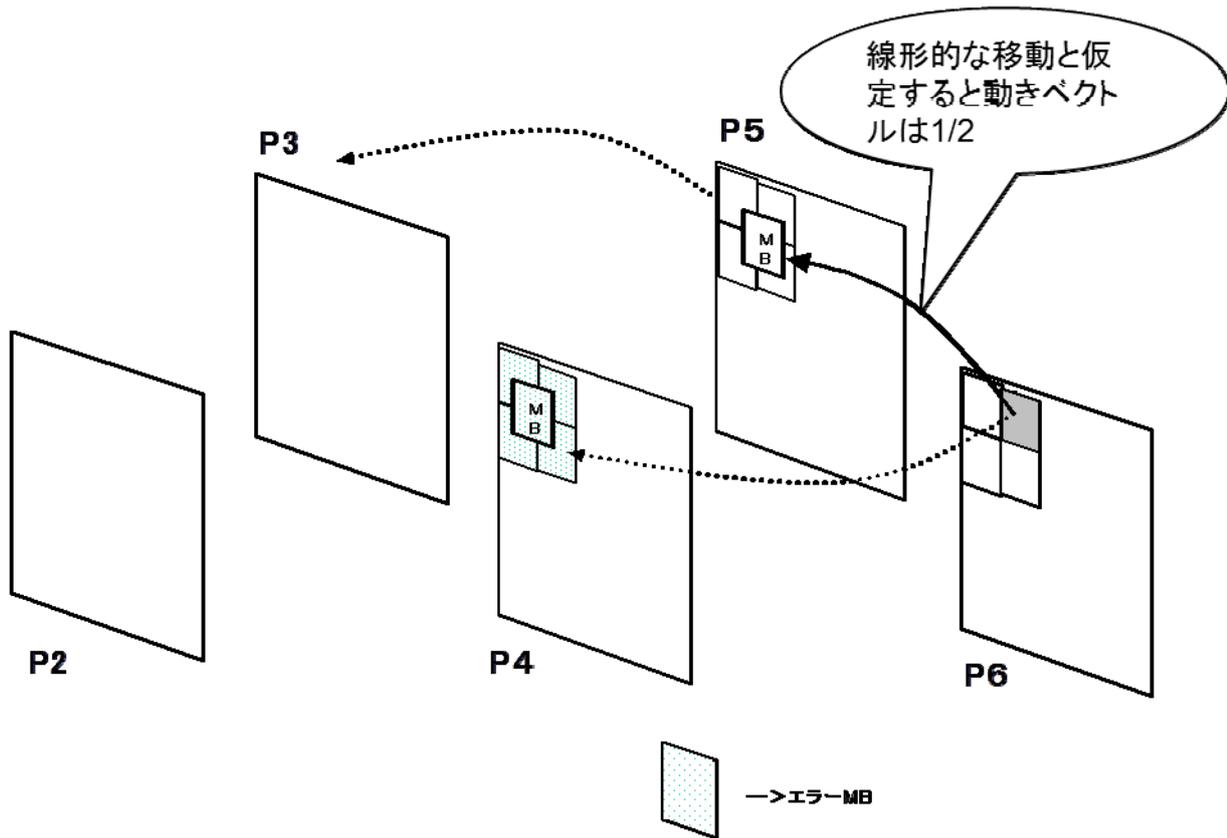


図 2-4 参照画像変更処理

本提案は、例えば、フレーム数が 6 の場合、P6 フレームの参照画像として P4 フレームより P6 フレームに時間的に近い P5 フレームを参照することで、P6 フレームの映像を復号しエラーの伝播を防止する方法である。また、参照フレーム変更に伴って、動きベクトル(Motion Vector 以下、MV)および予測残差信号(Sum of Absolute Difference 以下、SAD)を変換する。変換は、符号化対象フレームと、変換前後の参照フレームとのフレーム間隔比によるスケーリングとする。例えば、図 2-4 のように変更前のフレーム間隔が 2、変更後のフレーム間隔が 1 の場合、フレーム間隔比は 1/2 となり、MV および SAD とともに 1/2 倍する。また、フレーム間隔が 4、変更後のフレーム間隔が 1 の場合、フレーム間隔比は 1/4 となり、MV および SAD とともに 1/4 倍する。

### (3) ポストフィルタ

Temporal MD を用いて正常に受信できた映像を  $D_0$  で表示した場合，“ちらつき”が発生する．理由として，独立した各デコーダに，時間的に隣り合った映像ソースを入力した場合においても，各デコーダは独立しているため，実質は違う映像を交互に表示していることと等価である．そのためそれらを表示した場合，“ちらつき”となって見える．一般的に“ちらつき”の発生している画像には，隣同士の画素は似た画素であることに着目したポストフィルタを用いることが多い．本節もポストフィルタを時間的に前後の画素から重み付け平均を行うことで“ちらつき”の軽減を行う．図 2-5 は概念図を示す．例えば， $E_2$  で復号された画素を  $I_i$ ，その  $I_i$  よりも 1 フレーム前の画素を  $I_{i-1}$ ， $I_i$  よりも 1 フレーム後の画素を  $I_{i+1}$ ，とすると  $D_0$  で表示する画素を  $I_i'$  とした算出方法を式(2-1)に示す．

$$I_i' = I_{i-1} + 2 * I_i + I_{i+1} \tag{2-1}$$

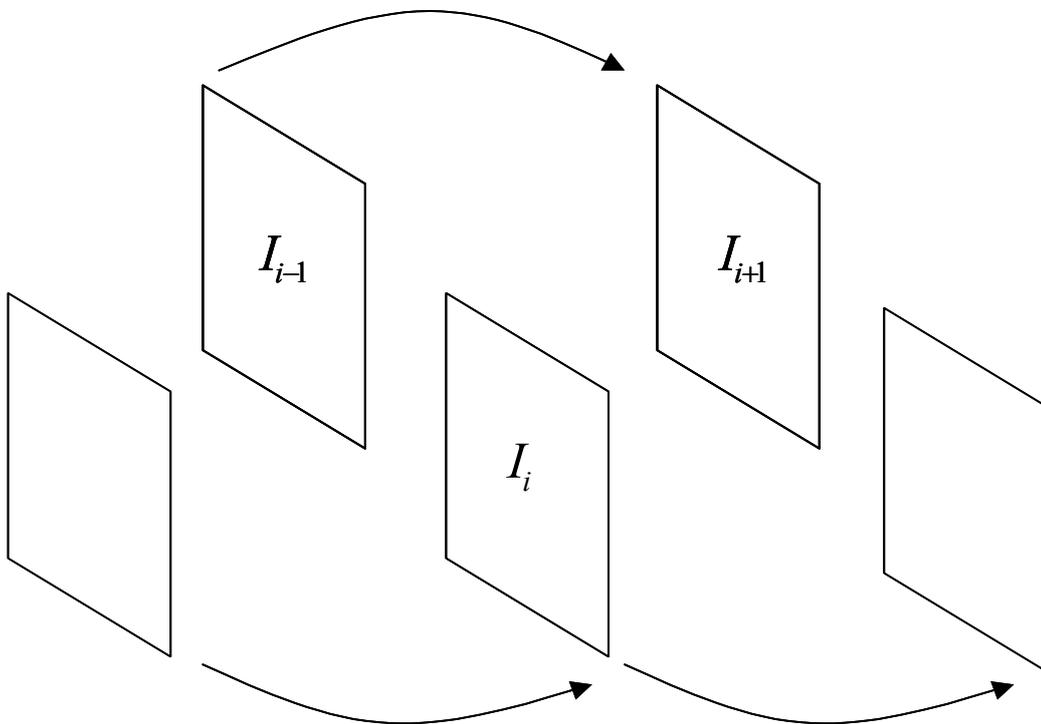


図 2-5 ポストフィルタの概念

## 2.2.3 シミュレーションおよび結果

### (1) 評価諸元

提案手法をシミュレーションにより評価を行った。評価の諸元を以下の表 2-1 に示す。映像ソースは、図 2-6 に映像ソースのスナップショットを示す。MPEG-4 の標準化で使用されている 3 つのパターンを使って評価した。特に、この 3 つの映像ソースは、誤り耐性の評価に使われることが多い。理由としては、Overtime は、全体的に動きの少ないバスタショットである。誤り耐性の評価には比較的優しい映像ソースである。News は、常に真中付近で大きく動いていながら、全体的には動きが少ない評価ができる。これは画面内で動きの違う 2 つの評価ができるためである。Foreman について、前半は、動きが少ないバスタショットであるが、後半は、カメラがパンするほど大きな動きがあるため、1 つの映像ソースの中で 2 つの評価ができる。

表 2-1 評価の諸元

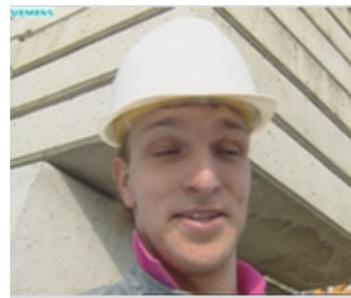
映像ソース	
種類	MPEG-4 評価用の Overtime, News, Foreman
フレーム長	300 フレーム
画像サイズ	CIF(352×288)
フレームレート	30 フレーム/秒
プログラム	
MPEG-4 符号化器	標準化活動用 Microsoft, MPEG-4 Ver1.0 FDIS 準拠
RTP 通信部 / MD decoder 部	独自製作
実行環境	
OS	Windows 2000 Professional
CPU	Server : Pentium III 850MHz Client : Pentium III 800MHz
メモリ	640 MB



Overtime



News



Foreman

図 2-6 映像ソースのスナップショット

$D_1$  と  $D_2$  そして  $D_0$  の各フレームにおける PSNR の評価を行った。実験は、300 フレームの各ビデオストリームにおいてデータ構造は IPPPP・・・とし、Rate Control は OFF, PSNR の測定は 5 回の平均値とする。

## (2) 評価構成

評価に使った実験ネットワークの構成を図 2-7 に示す。

予めサーバに下記の①②③のデータを蓄積し、RTP パケット通信で、パケットエラー発生装置経由でクライアントに送信する。なお、パケットエラー発生装置では、パケットロス率を 1%、3%、5%で実施した。つまり、3 種類のビデオストリームに対して 3 つのパターンのパケットロスの実験を行い、それぞれにおいて各ビデオストリーム全体の平均として PSNR を測定した。PSNR は、原画像と復号した画像との差分を平均二乗誤差で測定した。なお、パケットロス率は、1%がインターネット利用時、3%が無線 LAN 利用時、5%は 3G 利用時のパケットロス率を想定している。

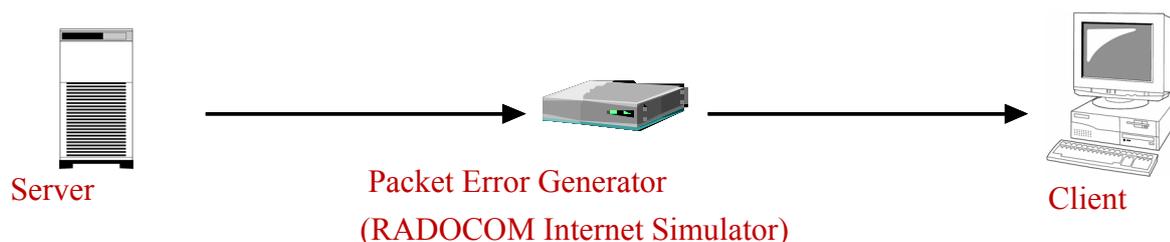


図 2-7 評価に使った実験ネットワークの構成

- ① 通常の MPEG-4 ビデオストリーム
- ② MPEG-4 の SVC ビデオストリーム
- ③ 提案方法のビデオストリーム

## (3) 実験結果

提案手法である Temporal MD, 及び比較手法である MPEG4 Simple Profile および MPEG4 Simple Scalable Profile について R-D(Rate-Distortion)曲線を用いて比較した結果を図 2-8 から図 2-19 に示す。

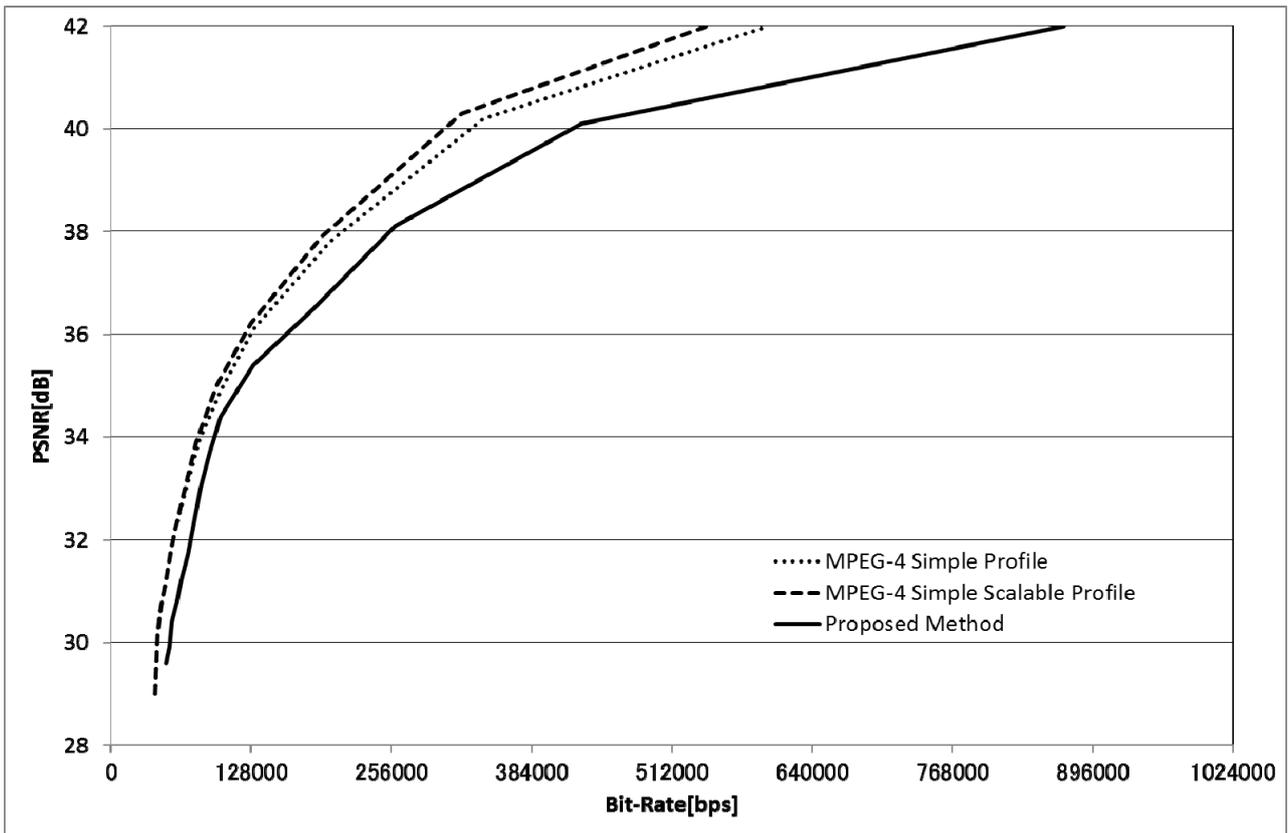


図 2-8 Overtime の R-D 曲線(Packet Loss 0%)

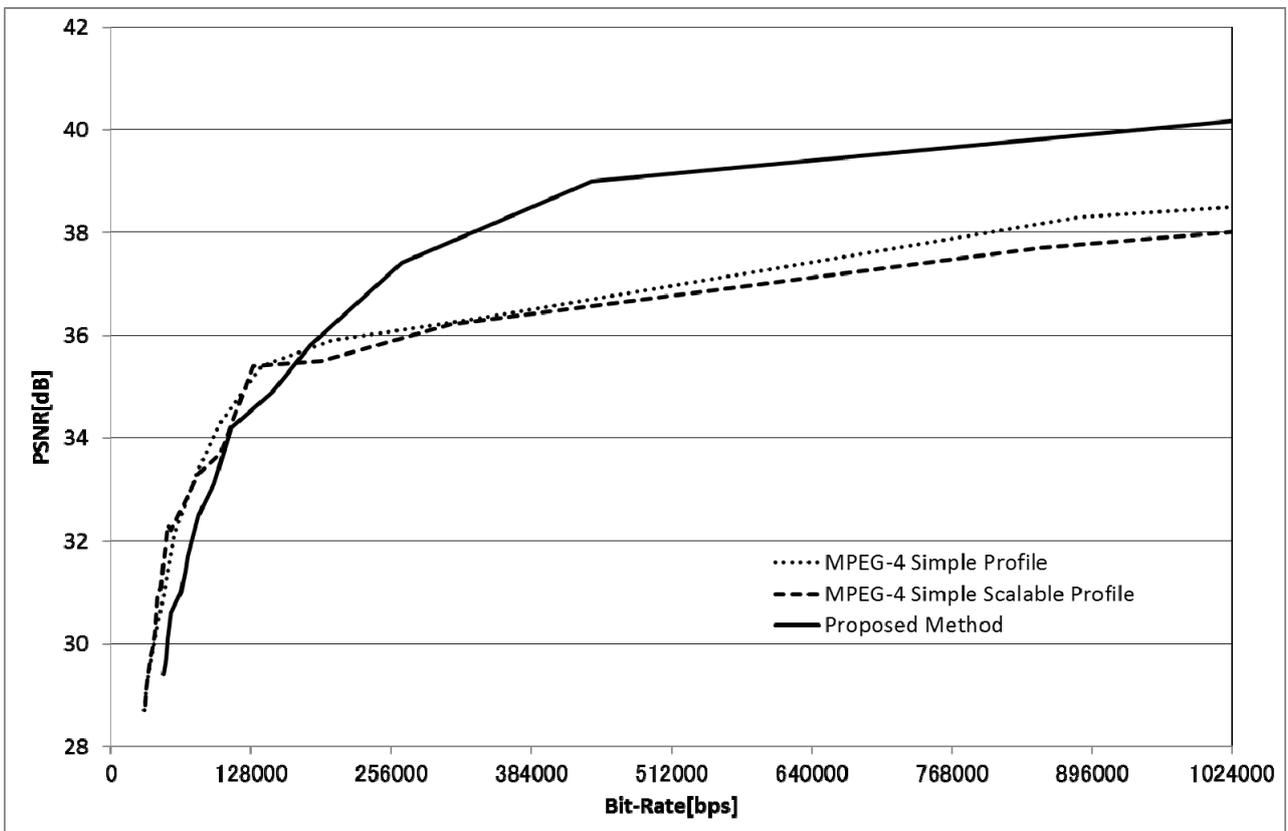


図 2-9 Overtime の R-D 曲線(Packet Loss 1%)

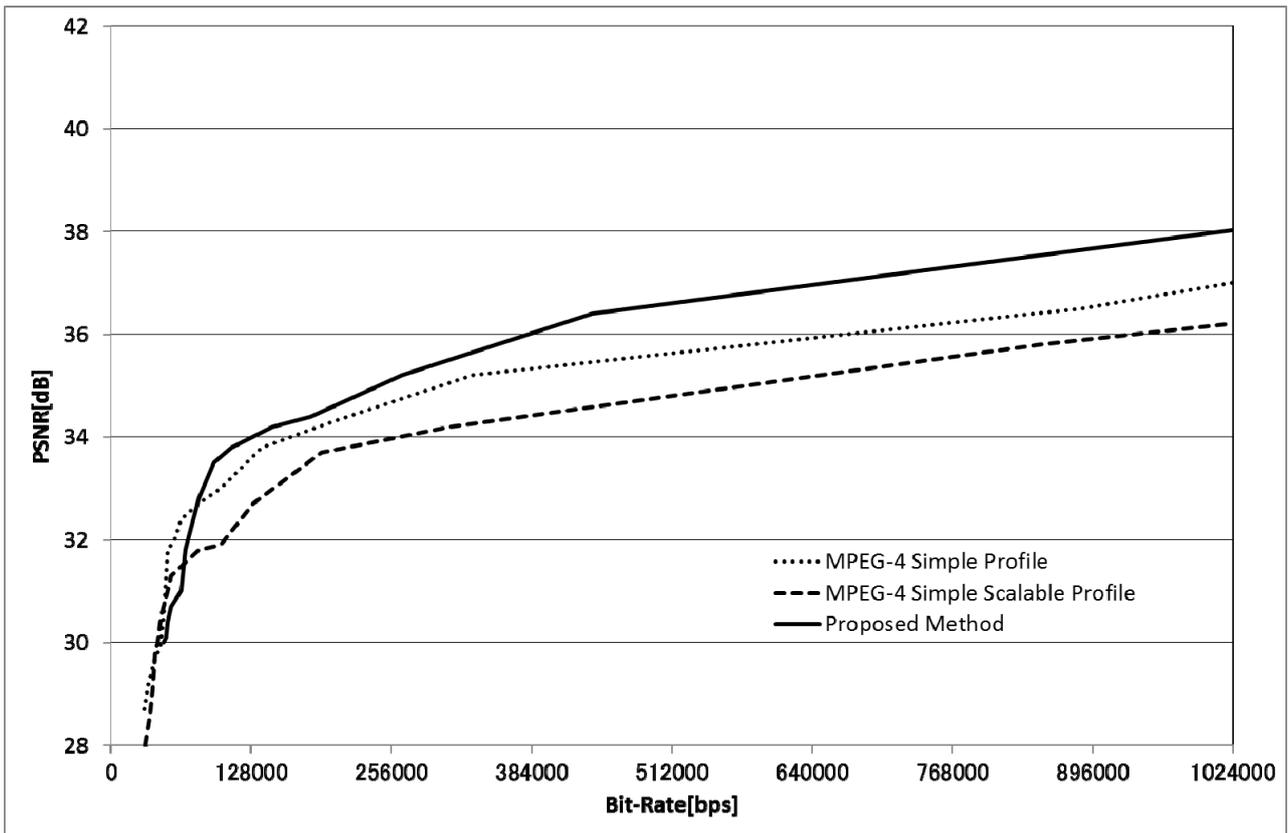


図 2-10 Overtime の R-D 曲線(Packet Loss 3%)

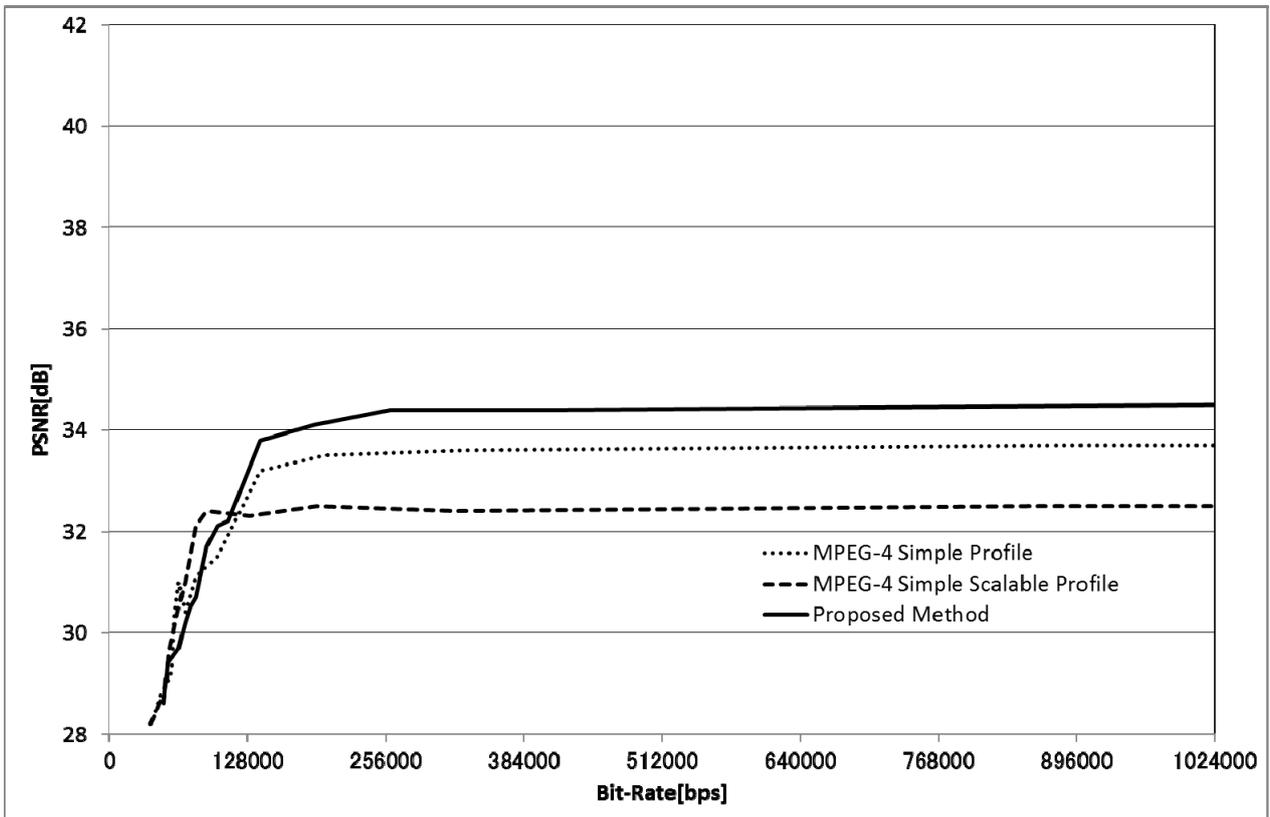


図 2-11 Overtime の R-D 曲線(Packet Loss 5%)

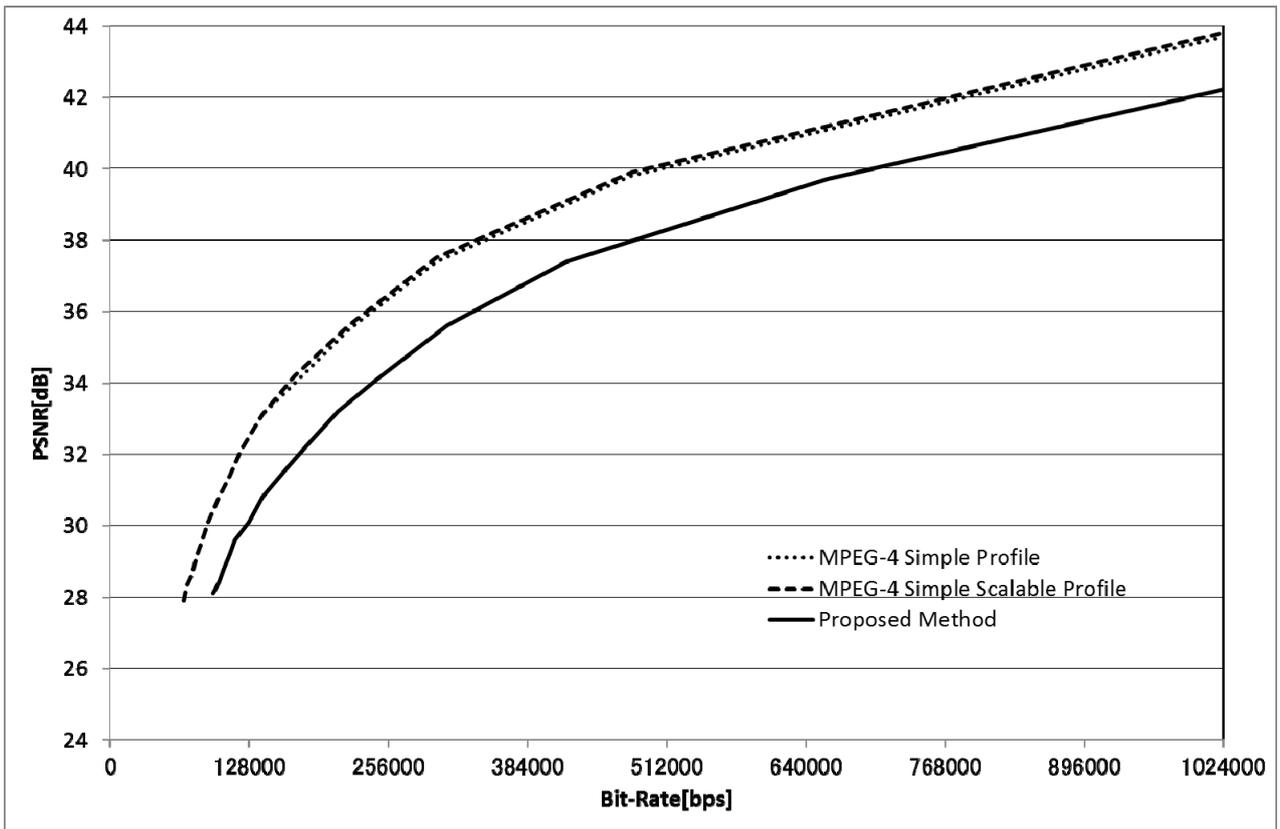


図 2-12 News の R-D 曲線(Packet Loss 0%)

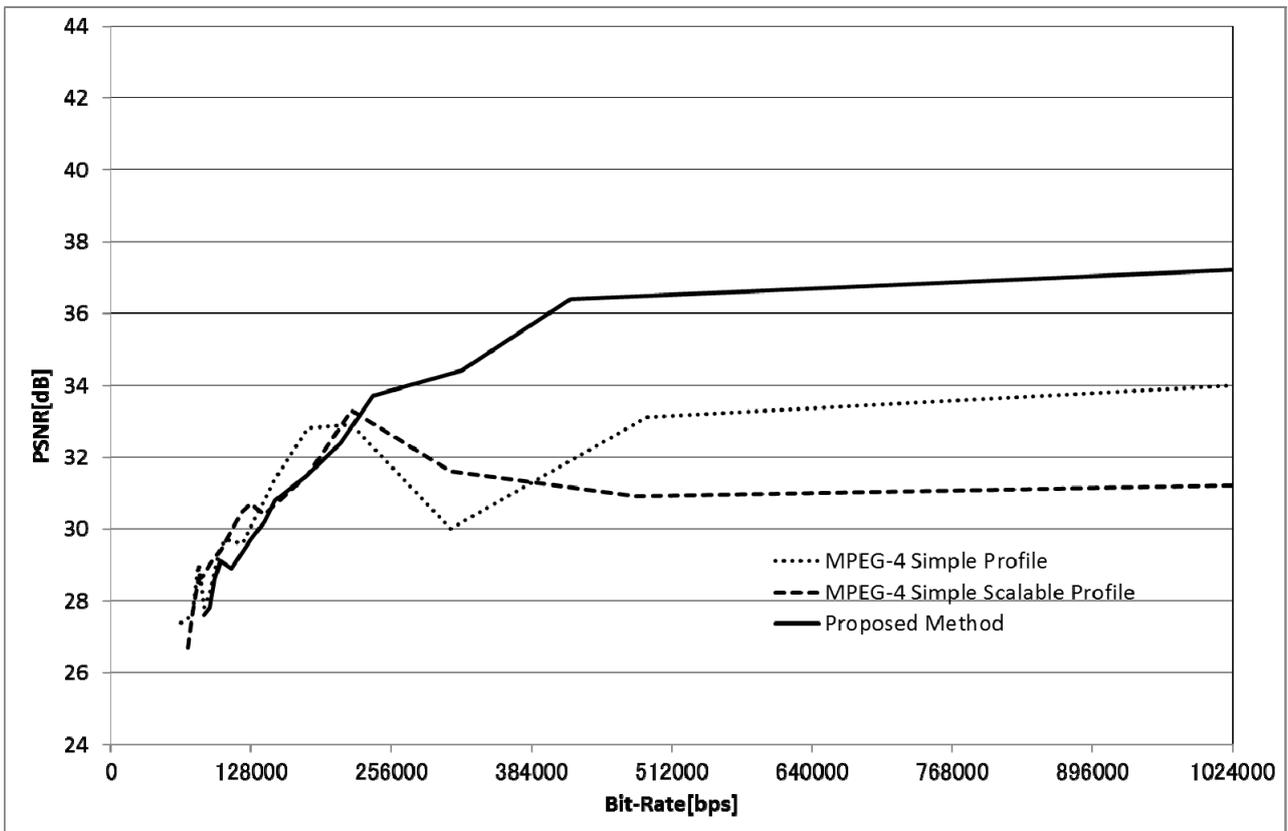


図 2-13 News の R-D 曲線(Packet Loss 1%)

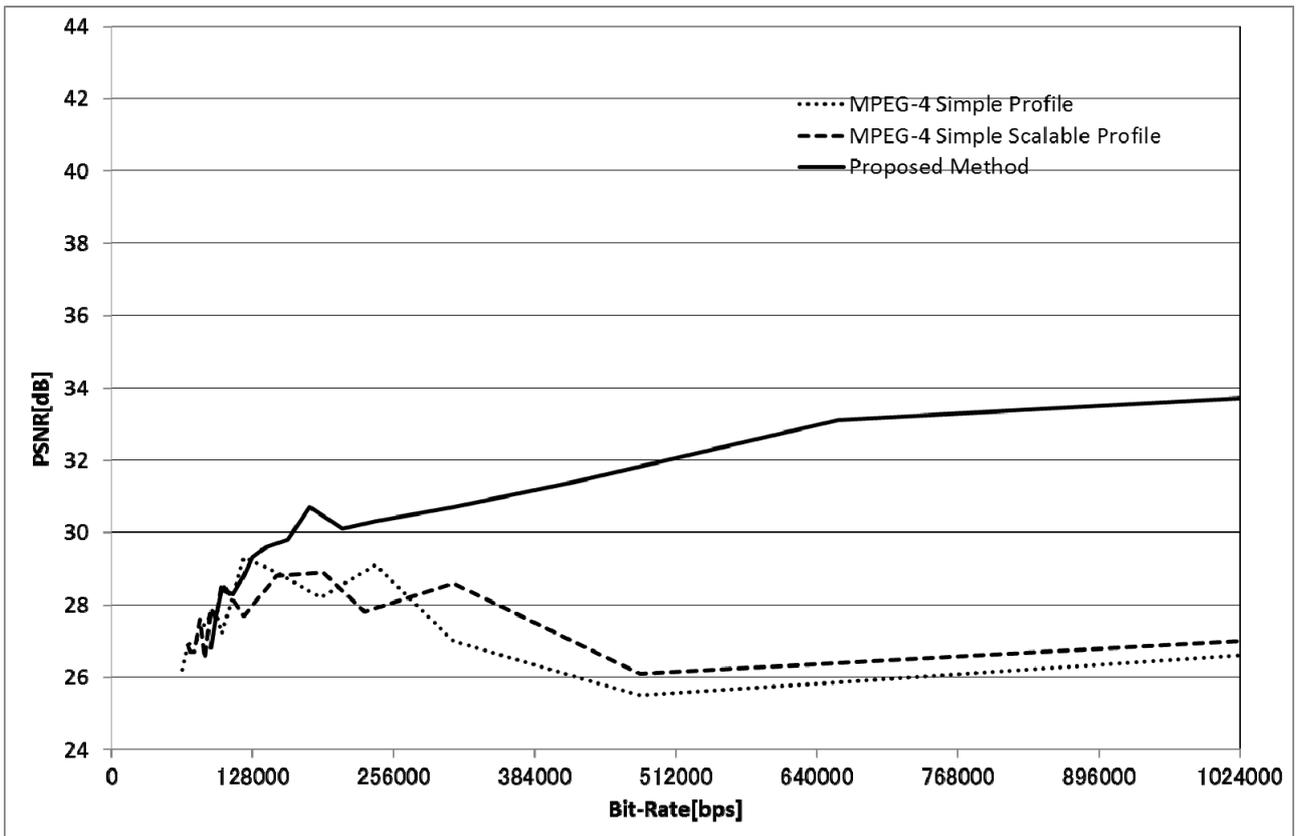


図 2-14 News の R-D 曲線(Packet Loss 3%)

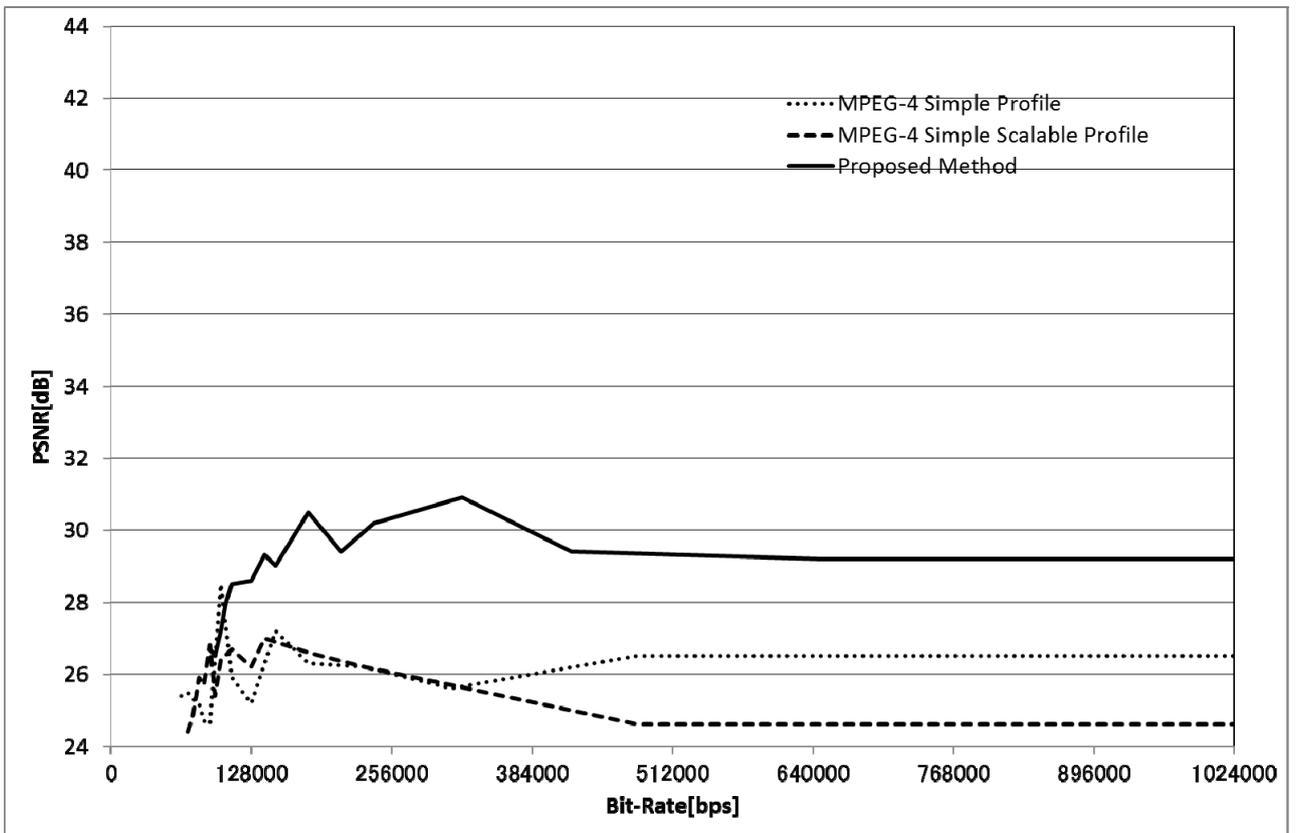


図 2-15 News の R-D 曲線(Packet Loss 5%)

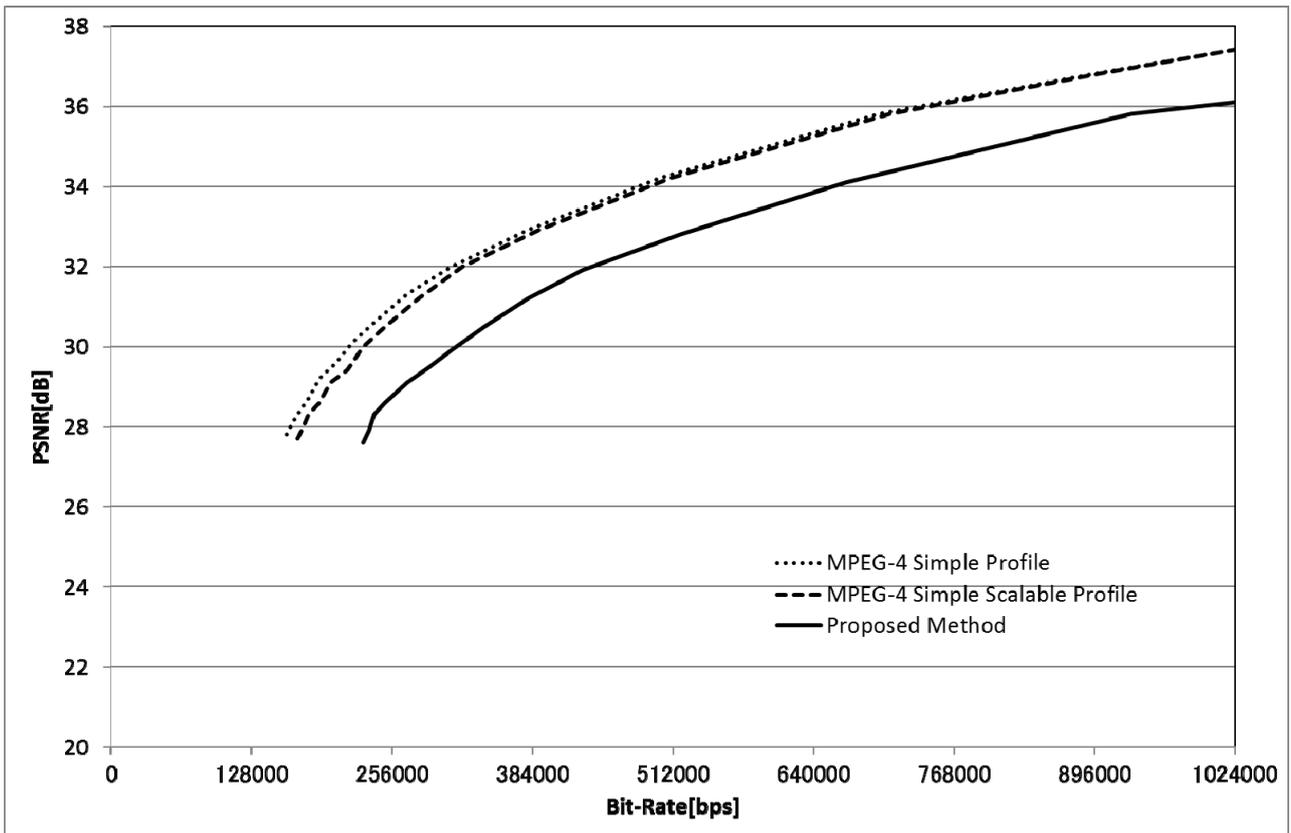


図 2-16 Foreman の R-D 曲線(Packet Loss 0%)

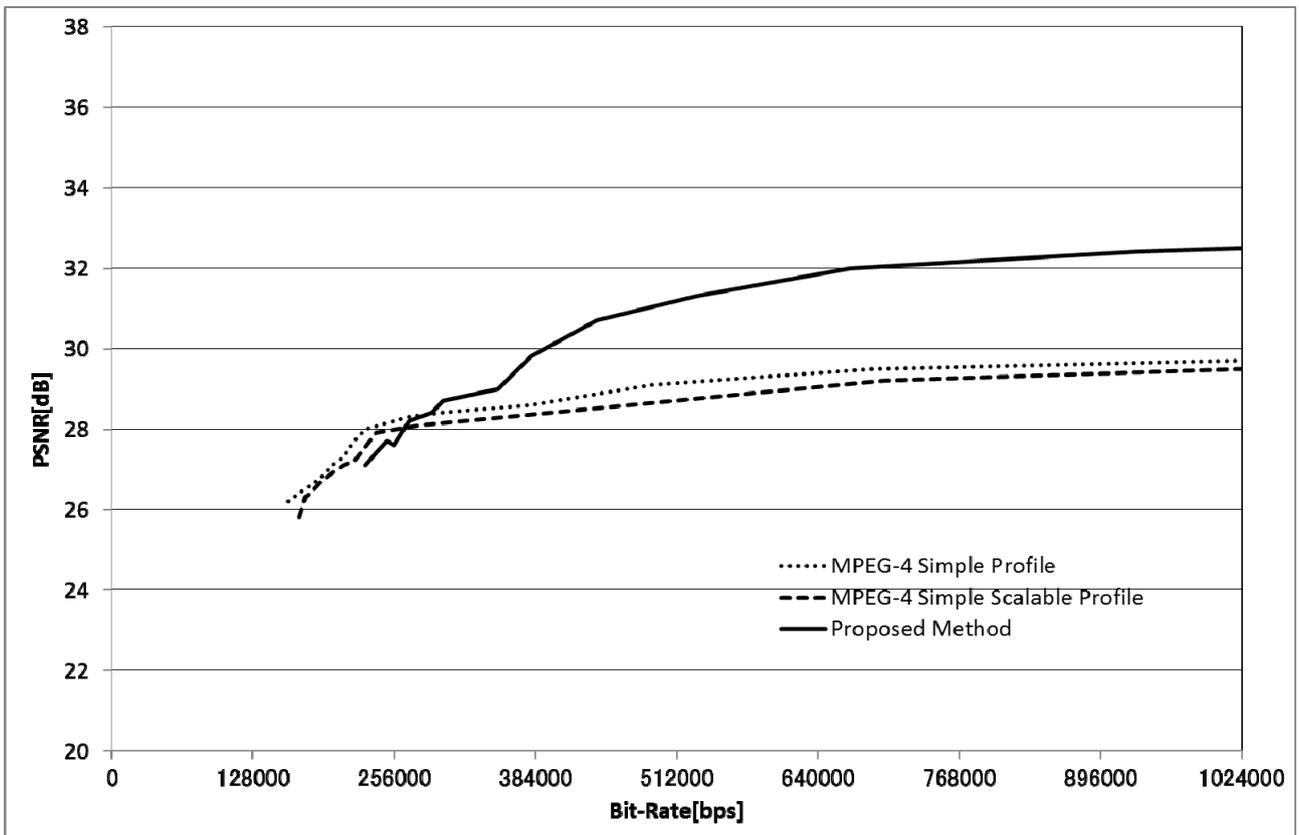


図 2-17 Foreman の R-D 曲線(Packet Loss 1%)

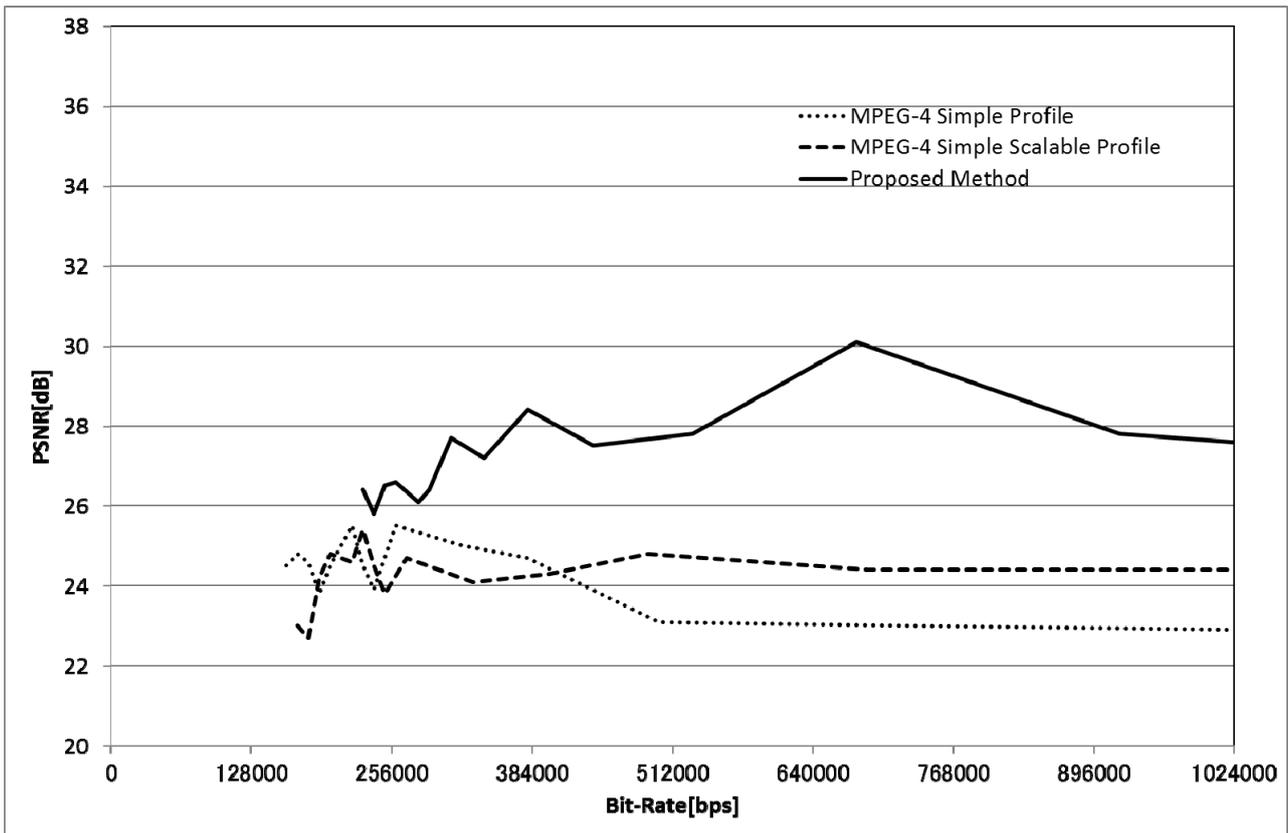


図 2-18 Foreman の R-D 曲線(Packet Loss 3%)

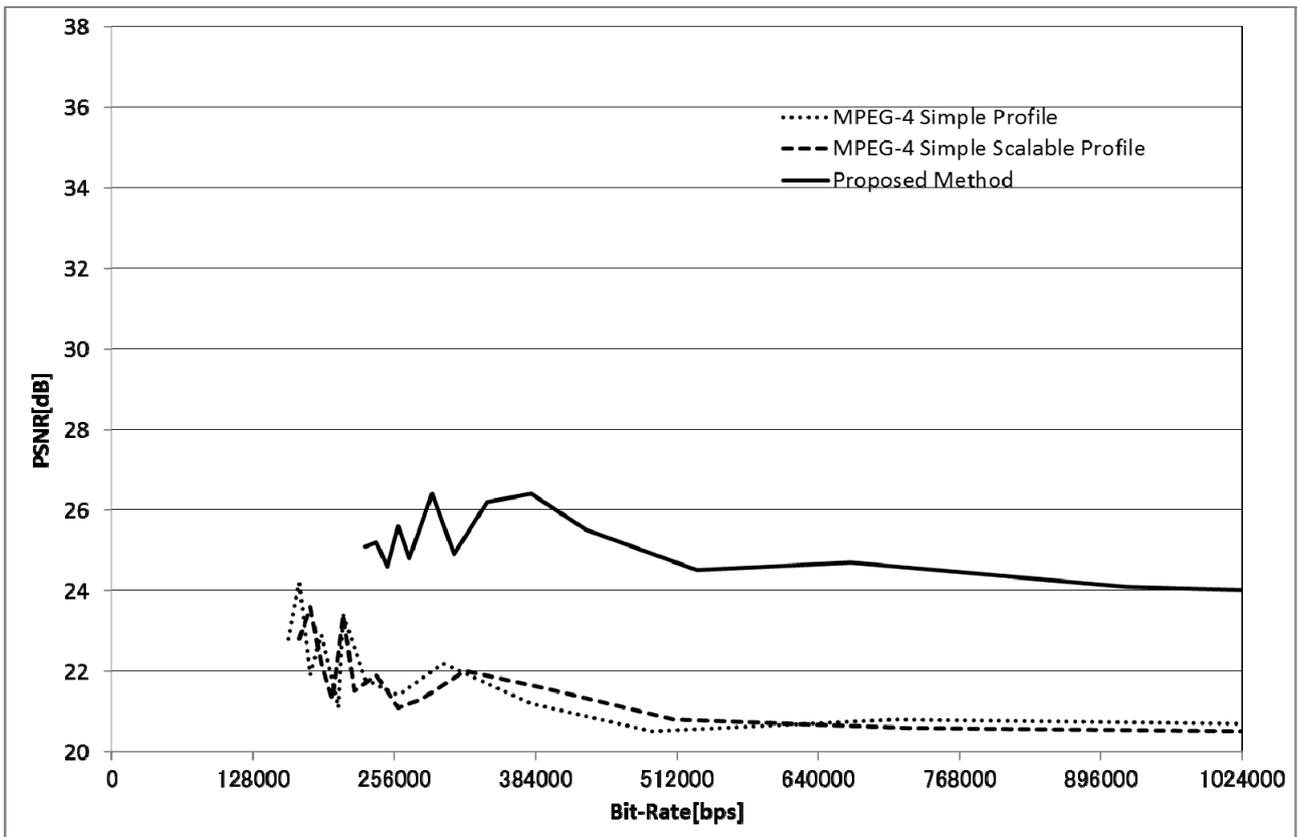


図 2-19 Foreman の R-D 曲線(Packet Loss 5%)

図 2-8 から図 2-11 は、映像ソースに Overtime を使った場合の評価結果を示しており、順にパケットロスが 0%、1%、3%、5%の場合を示している。

図 2-8 より提案方法は、他の方式に比べて約 2dB 程度劣っている。これは、MDVC が冗長構成となっているので、他の方式に比べて劣っていると思われる。しかし、図 2-9 のようにパケットロスが 1%発生する環境では、ビットレートが 160kbps を超えた辺りから Temporal MD の方の精度が徐々に上がっており 384kbps 程度以降では、2dB 程度の品質の向上が確認できた。これは、他の方式でもビットレートが 160kbps 程度までであれば、パケットロスが 1%程度であっても、MPEG-4 の誤り耐性機能で修復できたが、ビットレートが上がると、画像に対するエラーの伝搬が大きくなり修復できないためと思われる。

なお、図 2-10 と図 2-11 を見るとパケットロスが 3%や 5%くらいに発生すると、Temporal MD も含めて他の方式もビットレートが 128kbps に届く前に MPEG-4 の誤り耐性機能で補えなくため劣化する。しかし、Temporal MD の方が、画像に対するエラーの伝搬を抑えることができているため、他の方式よりも 2dB の品質向上が確認できた。

図 2-12 から図 2-15 は、映像ソースに News を使った場合の評価結果を示しており、順にパケットロスが 0%、1%、3%、5%の場合を示している。

図 2-12 のように提案方法は、他の方式に比べて約 2dB 程度劣っている。これは、Overtime と同じで MDVC が冗長構成となっているので、他の方式に比べて劣っていると思われる。しかし、図 2-13 を見るとパケットロスが 1%発生する環境では、ビットレートが 200kbps を超えた辺りから Temporal MD の方の精度があがっており 384kbps 程度では、4-6dB 程度の品質の向上が確認できた。Overtime と違い、他の方式の劣化が激しい。これは、同じフレーム内に動きの激しい映像があるために、その部分の画像に対するエラー蓄積しているためと思われる。

なお、図 2-14 と図 2-15 を見るとパケットロスが 3%や 5%くらいに発生すると、Temporal MD も含めて他の方式もビットレートが 128kbps に届く前に MPEG-4 の誤り耐性機能で補えなくため、劣化するが、Temporal MD の方が、画像に対するエラーの伝搬を抑えることができているため、他の方式よりも 4-7dB の品質向上が確認できた。

図 2-16 から図 2-19 は、映像ソースに Foreman を使った場合の評価結果を示しており、順にパケットロスが 0%、1%、3%、5%の場合を示している。

図 2-16 のように提案方法は、他の方式に比べて約 2dB 程度劣っている。これは、Overtime や News と同じで MDVC が冗長構成となっているので、他の方式に比べて劣っていると思われる。しかし、図 2-17 を見るとパケットロスが 1%発生する環境では、ビットレートが 256kbps を超えた辺りから Temporal MD の方の精度があがっており 384kbps 程度では、2-3dB 程度の品質の向上が確認できた。

なお、図 2-18 と図 2-19 を見るとパケットロスが 3%や 5%の時では、すべての方式が小刻みに変動している。これは、Foreman の映像ソースが、動きの少ない時と激しい時が混在しているため、パケットロスによる画像に対するエラーから回復する時間が動きの少ない時に行われたと思われる。

今回のシミュレーションの結果、パケットロスが発生しない環境では、Temporal MD は他の方式に比べて冗長構成であるために約 2dB 程度劣っていることがわかった。また、低ビットレート（概ね 200kbps 未満）であれば、パケットロスが 1%程度でも、MPEG-4 の誤り耐性機能で修復できることがわかった。しかし、高ビットレート（概ね 200kbps 以上）でパケットロスが発生すると、すべての映像ソースにおいて、Temporal MD の方が、他の方式に比べて 2-6dB 程度優れていることが確認できた。

#### (4) 考察

上記シミュレーションの結果で得られたように、低ビットレートであれば、パケットロスが1%程度であっても MPEG-4 の誤り耐性機能で修復できることがわかった。また、Temporal MD の参照画像変更に伴う参照画像不一致から生じる“ちらつき”による映像の乱れが懸念されたが、実行上さほど問題にならないことが確認できた。“ちらつき”は、蓄積してくると大きな誤りとなって目立ってしまうが、今回の場合、画像に対するエラーが発生した MB だけ参照画像変更を行っているため、全体のフレーム画像としては蓄積しにくいと思われる。

最後に、高ビットレートでは、パケットロスが発生しても、すべての映像ソースにおいて、Temporal MD の参照画像変更が有効であることがわかった。

#### 2.2.4 まとめ

本節では、パケットロスが発生するネットワーク環境において、MDVC の時間スケラビリティ機能を拡張する手法を提案し、MPEG-4 Simple Profile および MPEG-4 Simple Scalable Profile と比較した。MDVC は冗長構成であるため転送レートが高くなり、MPEG-4 Simple Scalable Profile だけでなく、MPEG-4 Simple Profile と比べても 2dB 程度劣化しており非効率であることがわかった。

しかし、パケットロスが発生する環境においては、低ビットレートであれば、パケットロスが1%程度であっても MPEG-4 の誤り耐性機能で修復できるが、高ビットレートが必要な場合は、Temporal MD を用いることで、他の方法に比べて、2-6dB 程度優れていることがわかった。つまり、高ビットレートでパケットロスが発生する可能性がある無線区間では、Temporal MD を使うことで高品質な映像配信サービスの提供に有効であると思われる。

現状では、ネットワークの帯域変動に対しては追従できない課題があるので、PSNR の向上を考慮した参照画像変更の方法を検討しながら、可変ビットスケラビリティを実現し、ネットワーク上での帯域変動に対して高い適応力を持った MPEG4 FGS(Fine Granularity Scalability)の Temporal Scalability に対しても比較検討していく予定である。

## 2.3 SNR スケーラビリティ機能を用いた方式(Quantizer MD)の評価

### 2.3.1 序言

MD(Multiple Description)は、1つの映像ソースを2つのエンコーダによってそれぞれ符号化し、2つのチャンネルにそれぞれを転送し、それぞれのデコーダが独立して復号できるため、例えば1つのチャンネルがパケットロスでデータが欠落しても、もう片方のデコーダで再生できるメリットがある。MDに適した冗長構成なビデオ符号化方式として多重記述ビデオ符号化(Multiple Description Video Coding 以下、MDVC)がある。

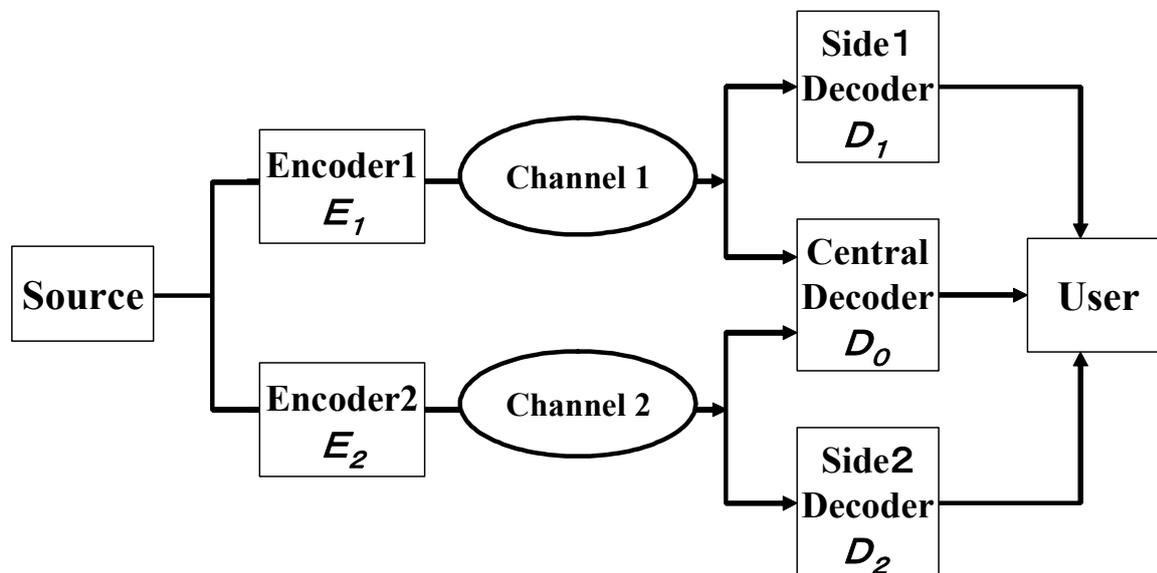


図 2-20 Multiple Description Video Coding の構成

MDVC の基本構成を図 2-20 に示す。情報源から入力した信号は、Encoder1( $E_1$ )および Encoder2( $E_2$ )でそれぞれ符号化し、異なるチャンネルを使って伝送するものである。例えば、チャンネル 1 として無線 LAN、チャンネル 2 として 3G セルラーを使う。受信側では Side1 Decoder ( $D_1$ )および Side2 Decoder ( $D_2$ )それぞれで復号し、相互補完することで途切れの無い通信が行われ、さらに 2 つのチャンネルが受信できれば、Central Decoder ( $D_0$ )で高品質なデータ通信が行われる。

本節は、MPEG-4 を代表とする従来の符号化アルゴリズムを変更することなく、異なる量子化操作を合成して PSNR を向上させる MDVC 方式を Quantizer MD と定義し評価する。本手法は、無線 LAN やセルラーを併用するような複合的無線サービスでの動画伝送に利用することができる。一般に、2 つの異なる量子化パラメータで量子化された同一ブロックの DCT 係数は、逆量子化すると異なる DCT 係数となる。提案方式は、この異なる量子化後の DCT 係数から量子化前の DCT 係数(真値)が取り得る範囲を絞り込み、絞り込んだ DCT 係数を利用できるように量子化パラメータを操作する手法を提案する。

## 2.3.2 提案方法

### (1) 量子化パラメータの操作方法

MDVC の代表的な方式としては、スカラー量子化やベクトル量子化を使った方法があり、前者は、両チャンネルで有効なデータが受信できた時は、量子化誤差を平均して最小にするような符号化器の検討や可変長符号を使った符号化器によりエントロピーを制限する方法の検討などが報告されている。また、後者を使った方法としては、符号帳のラベリングを最適化した方法やチャンネル毎に歪みの重み付けを行う方法などが報告されている。

MDVC を使って量子化パラメータを操作することは、重要な位置づけと思われる。そこで、本節では、異なるビデオストリームの量子化後の DCT 係数から、量子化前の DCT 係数値（以下、真値）が取り得る範囲を絞り込み、絞り込んだ DCT 係数を利用できるように量子化パラメータを操作する方法を提案する。量子化パラメータの操作を DCT 係数領域内で行うため、復号時に参照画像の不一致によるドリフト誤差が生じ、画質劣化を引き起こす恐れがある。そこで、量子化パラメータを操作するにあたっては、ドリフトによる影響も考慮している。

図 2-21 に Quantizer MD の基本概念を示す。縦軸に DCT 係数の値(Level), 横軸は、量子化パラメータとする。E<sub>1</sub> と E<sub>2</sub> それぞれに同じ映像信号を入力するが、E<sub>1</sub> と E<sub>2</sub> では異なる量子化パラメータ(Qp)で DCT 係数値を量子化する。この時 E<sub>1</sub> の量子化パラメータ(Qp1)に対応した量子化ステップ(Qstep1)や E<sub>2</sub> の量子化パラメータ(Qp2)に対応した量子化ステップ(Qstep2)によって丸められてビデオストリームを生成する。つまり、真値は、Qstep1 や Qstep2 の範囲の中に存在する。そこで Quantizer MD は、Qstep1 と Qstep2 が重なり合うところは、量子化によって丸められる前の真値に近似することに着目し、その量子化パラメータを操作する方法を提案する。

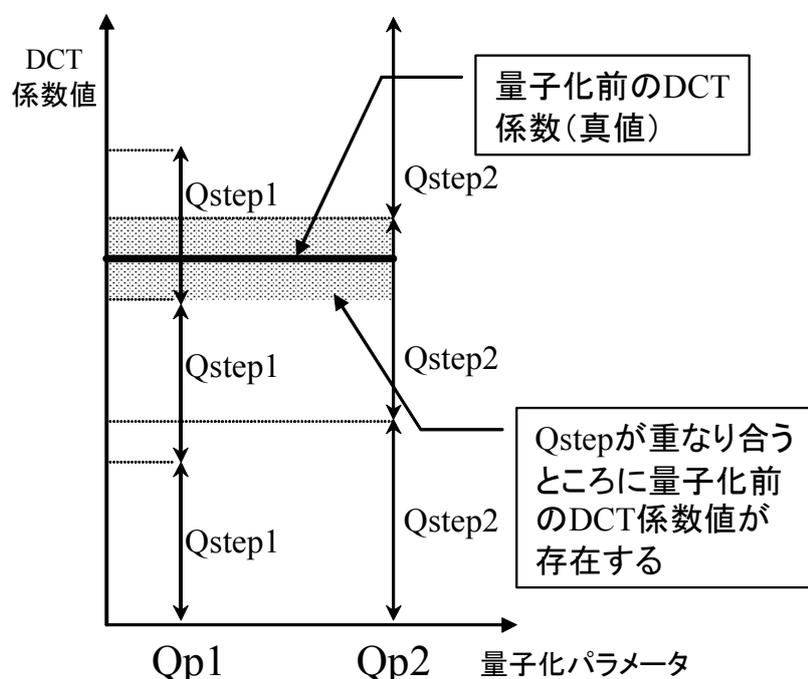


図 2-21 Quantizer MD の基本概念

(2) 量子化パラメータの算出方法

真値を  $D_1$  と  $D_2$  から近似する方法を図 2-22 に示す. 縦軸は DCT 係数値で, 横軸は量子化パラメータである. それぞれの Side Decoder で  $Qp1$  および  $Qp2$  で量子化した後の DCT 係数値を  $qcoef1$  および  $qcoef2$  とし, 逆量子化によって真値が取り得る範囲を求めると.  $qcoef1$  について量子化前の最小値を  $dct\_coef\_min1$  最大値を  $dct\_coef\_max1$  とし, この範囲を **range1** とする. また  $qcoef2$  について量子化前の最小値を  $dct\_coef\_min2$  最大値を  $dct\_coef\_max2$  とし, この範囲を **range2** とする. 真値は, 量子化後の  $qcoef1$  や  $qcoef2$  には誤差が生じ, これが量子化誤差となって映像品質の低下となる. そこで, 真値が取り得る範囲 **range1** と **range2** が重複するところに必ず真値があるため, 新しい真値  $dct\_coef$  とを式(2-2)の通りに定義する.

$min\_dct > 0$  の場合

$$dct\_coef = min\_dct + (max\_dct - min\_dct) / 4$$

$max\_dct < 0$  の場合

$$dct\_coef = max\_dct + (min\_dct - max\_dct) / 4$$

(2-2)

ただし,

$$min\_dct = \text{maximum}(dct\_coef\_min1, dct\_coef\_min2)$$

$$max\_dct = \text{minimum}(dct\_coef\_max1, dct\_coef\_max2)$$

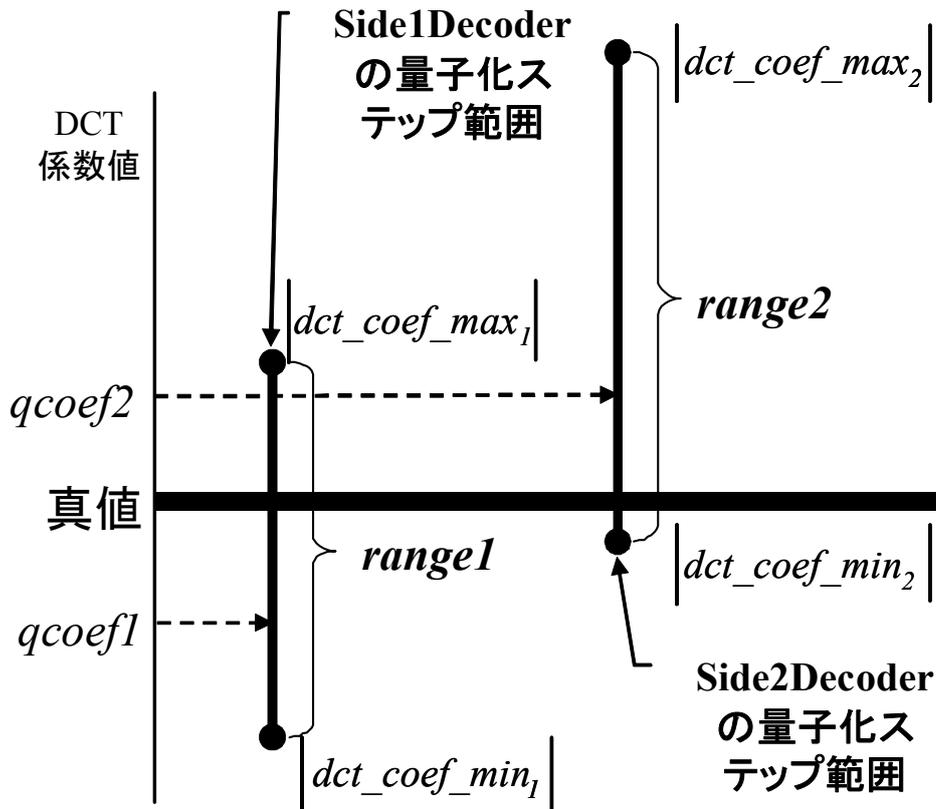


図 2-22 真値の近似方法(**range1** と **range2** が重なり合った場合)

図 2-23 は、 $D_1$  が  $D_2$  に包含された場合について、真値の近似方法を示す。縦軸は DCT 係数の値(Level), 横軸は量子化パラメータである。それぞれの Side Decoder で  $Qp1$  および  $Qp2$  で量子化した後の DCT 係数値を、 $qcoef1$  および  $qcoef2$  とし、それぞれの真値が取り得る範囲を  $range1$  および  $range2$  とする。

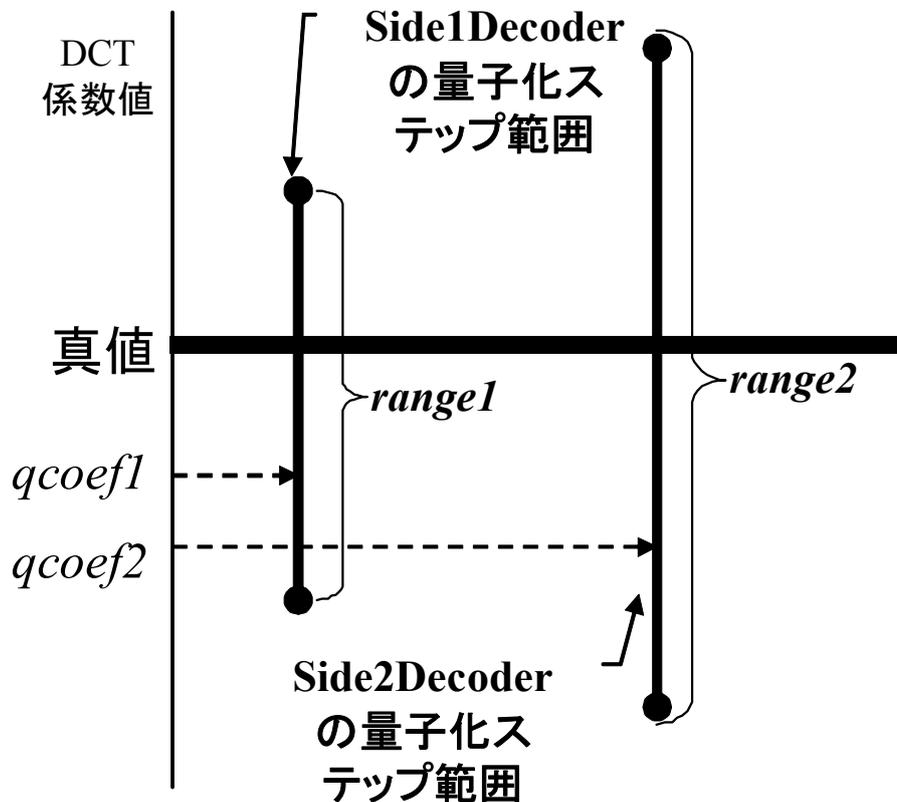


図 2-23 真値の近似方法( $range1$  が  $range2$  に包含された場合)

真値が取り得る範囲は、 $range1$  を  $range2$  が包含しているため必ず真値は、 $range1$  にある。そこで真値である  $dct\_coef$  は下記の式(2-3)と定義する。

$$dct\_coef = dct\_coef1 \quad (2-3)$$

ここで、 $dct\_coef1$  は  $qcoef1$  を逆量子化した値である。

### (3) 参照画像の変更に起因する画質劣化の低減(ドリフト補償)

Side1 Encoder である  $E_1$  や Side2 Encoder である  $E_2$  では、それぞれ独立に動き補償 (Motion Compensation 以下, MC) 予測を行っている。このため Central Decoder である  $D_0$  で、2.2.1 で提案した量子化パラメータの操作を行うと、参照画像が異なることがあり、予測誤差がずれることによるドリフト誤差が発生する。そこで、Side1 Decoder である  $D_1$  と Side2 Decoder である  $D_2$  の参照画像を一致させ、 $D_0$  において再復号化する方法を提案する。具体的な内容は、次の通りである。

原画像と参照画像の差分を算出し、離散コサイン変換後、量子化する式を定義する。つまり、 $D_1$  および  $D_2$  の参照画像データを  $ref1$ ,  $ref2$  とし、原画像データを  $original$  と

すると、 $D_1$  および  $D_2$  に入力されるビデオストリームの DCT 係数を  $qcoef1$  および  $qcoef2$  を式(2-4)式(2-5)で与えた。

$$qcoef1 = Qp1[\text{DCT}(\text{original} - ref1)] \quad (2-4)$$

$$qcoef2 = Qp2[\text{DCT}(\text{original} - ref2)] \quad (2-5)$$

ここで、式(2-5)の  $ref2$  を  $ref1$  として、参照画像データを一致させるために、式(2-5)の両辺に  $Qp2[\text{DCT}(ref2 - ref1)]$  を加算すると式(2-6)を得る。

$$Qp2[\text{DCT}(\text{original} - ref1)] \doteq qcoef2 + Qp2[\text{DCT}(ref2 - ref1)] \quad (2-6)$$

つまり、 $D_1$  および  $D_2$  とも予測誤差が  $(\text{original} - ref1)$  で一致するため、式(2-4)と式(2-6)を  $D_0$  において、再復号することで予測誤差が不一致によるドリフトが低減される。式(2-6)で参照画像を一致させる処理をドリフト補償と称することとする。

### 2.3.3 シミュレーションおよび結果

#### (1) 評価諸元

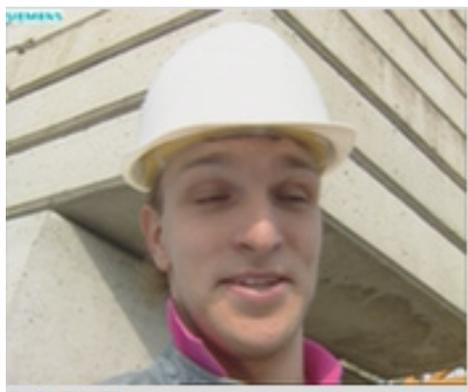
提案手法をシミュレーションにより評価を行った。評価の諸元を以下の表 2-2 に示す。映像ソースとしては、図 2-24 に示すように MPEG-4 の標準化で使用されている 2 つのパターンを使って評価した。Foreman について、前半は、動きが少ないバストショットであるが、後半は、カメラがパンするほど大きな動きがあるため、1 つの映像ソースの中で 2 つの評価ができる。

Mobile&Calendar は、最初からズームアウトしながら全体的に動きが激しく、評価画像としては、符号量が多く発生するため、映像の品質評価ができるためである。

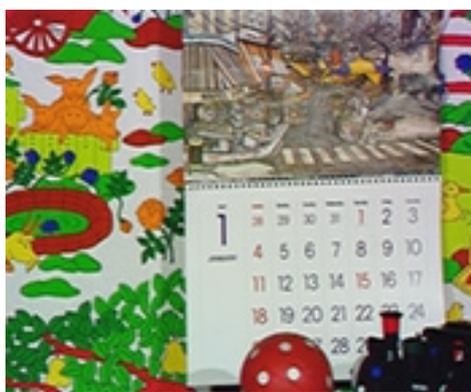
実験は、300 フレームのビデオストリームにおいて最初のフレーム(0frame)は Intra Frame とし、その後はすべて Inter Frame とする。Qp の設定は、画質の異なる 3 パターンの符号化映像を準備する。パターン(a)は Qp1 = 4, Qp2 = 5, パターン (b)は Qp1 = 10, Qp2 = 11, パターン(c)は Qp1 = 24, Qp2 = 25 を、それぞれ設定する。なお、PSNR は、原画像と復号した画像との差分を平均二乗誤差で測定した。

表 2-2 評価ビデオストリームも諸元

映像ソース	Foreman, Mobile&Calendar
画像サイズ	QCIF(176×144)
フレームレート	10 フレーム/秒
フレーム数	300 フレーム



(a)Foreman



(b)Mobile&Calendar

図 2-24 映像ソースのスナップショット

(2) 評価構成

提案した量子化パラメータの操作方法をシミュレーションによって評価する。評価において、Side1 Decoder および Side2 Decoder は、携帯電話で一般的なビデオ符号化方式である MPEG-4 Visual Simple Profile[2]を使用する。Central Decoder の復号化は独自に作成した。図 2-25 にシミュレーションで使用する復号化のブロック図を示す。Side1 Decoder および Side2 Decoder は、MPEG-4 の標準化で規定されている構成である。それぞれビデオストリームが入力されると量子化パラメータ、DCT 係数をそれぞれ逆量子化(IQ)および逆 DCT(IDCT)実施後、動きベクトルから復号画像を生成し、フレームメモリ(FM)に蓄積させ、次のフレームを復号するため動き補償(MC)を行う。Central Decoder では、Side1 Decoder と Side2 Decoder から DCT 係数( $qcoef1/qcoef2$ )および参照画像( $ref1/ref2$ )を分岐させ Central Decoder の量子化操作およびドリフト補償処理部に入力し、ドリフト補償を実施した後に高品質復号画像を生成する。図 2-26 は、復号化の動作を記述したフローチャートである。

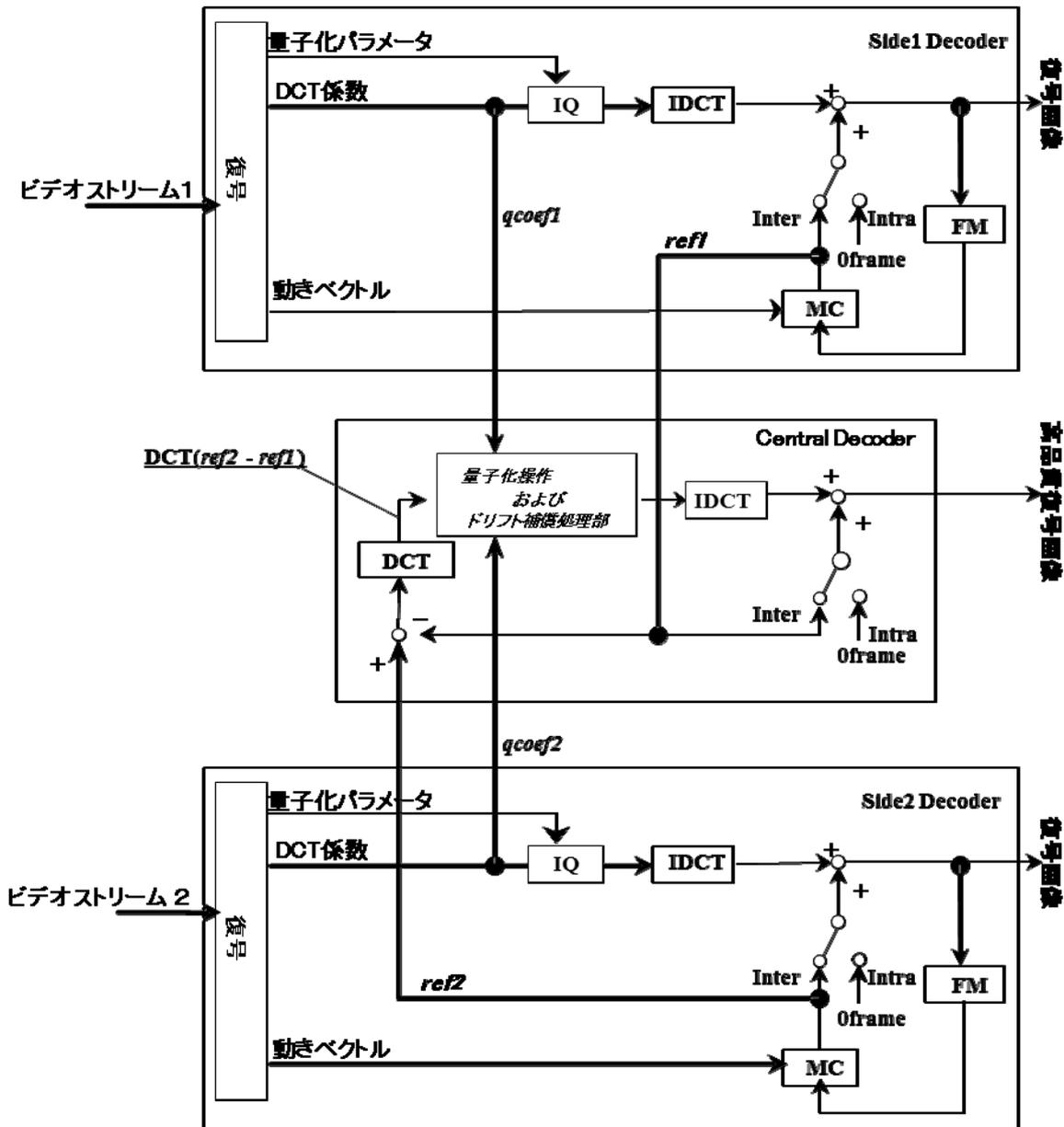
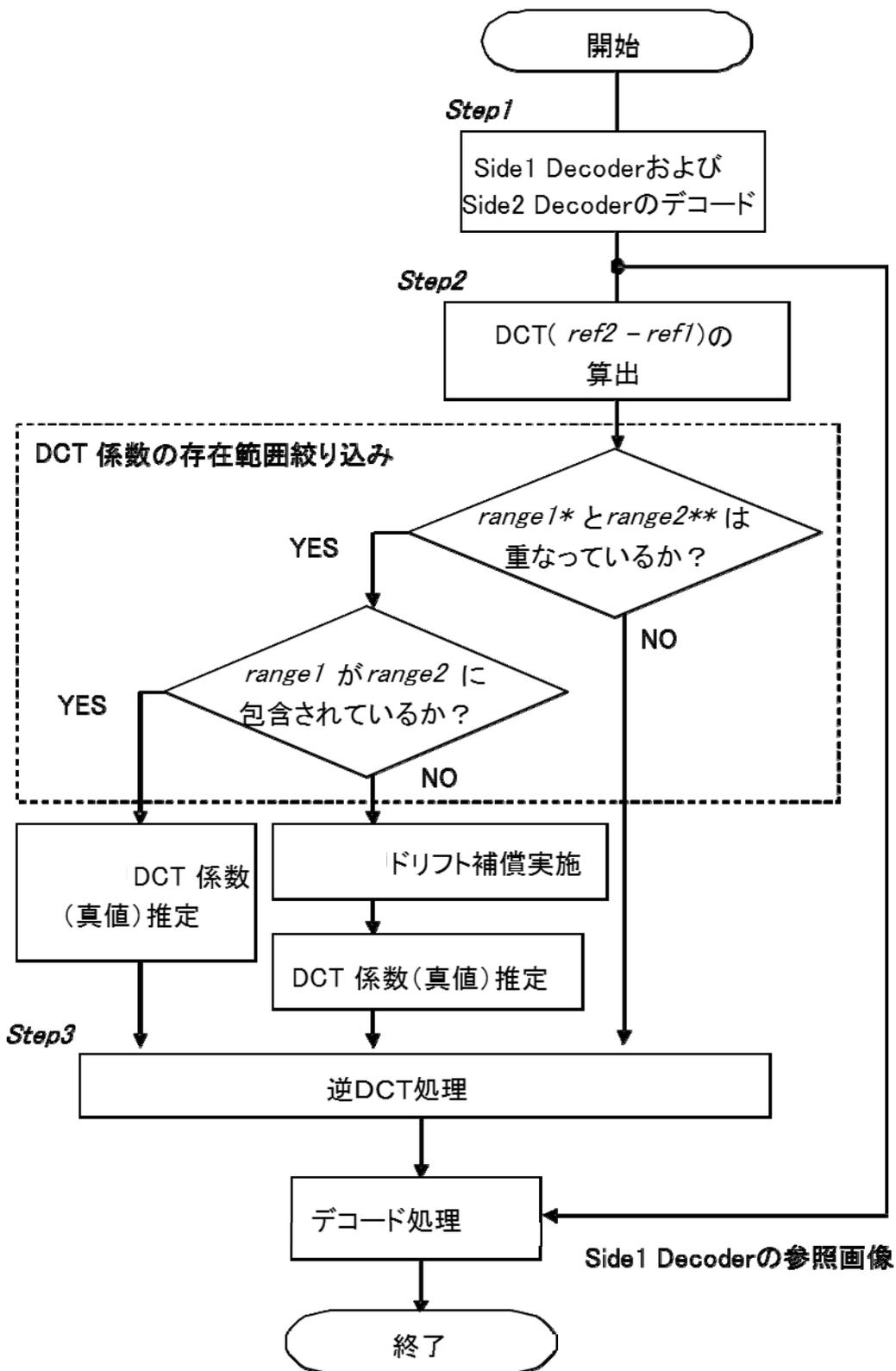


図 2-25 復号化のブロック図



- (\*) *range1* : Side1 DecoderでのDCT 係数存在範囲  
 (\*\*) *range2* : Side2 DecoderでのDCT 係数存在範囲

図 2-26 復号化のフローチャート

図 2-26 の復号化フローチャートより、

Step1 : Side1 Decoder では、ビデオストリーム 1(小さい量子化パラメータで量子化したビデオストリーム)から DCT 係数  $qcoef1$  を Central Decoder に出力すると共に、復号画像  $ref1$  を生成する。同じく Side2 Decoder では、ビデオストリーム 2(大きい量子化パラメータで量子化したビデオストリーム)から DCT 係数  $qcoef2$  を Central Decoder に出力すると共に、復号画像データ  $ref2$  を生成する。

Step2 : Central Decoder では、 $ref1$  および  $ref2$  から、 $DCT(ref2 - ref1)$  を算出し、量子化操作およびドリフト補償処理部に入力する。同処理部で、真値の絞り込みは、各 Side Decoder の取り得る範囲  $range1$  および  $range2$  の重なり具合に応じて処理する。図 2-22 のように部分的に重なり合っている場合は、量子化パラメータを操作することにより発生するドリフト誤差の補償は式(2-6)で算出する。さらに、その値を  $range2$  に用いて、量子化操作パラメータの操作を行い、式(2-2)より真値を推定し、逆量子化を行う。また、図 2-23 のように  $range1$  が  $range2$  に包含されている場合は、式(2-3)を算出し、真値を推定し、逆量子化を行う。もし、 $range1$  および  $range2$  の重なりがなかった場合は、量子化パラメータの操作は行わず、 $qcoef1$  を使って、逆量子化を実施する。

Step3 : 量子化パラメータの操作およびドリフト補償処理部から出力された量子化後の DCT 係数に対して、逆 DCT を行い、 $ref1$  に加算して、復号画像を生成する。

### (3) 実験結果

図 2-27 および図 2-28 は、Side1 Decoder( $D_1$ )、Side2 Decoder( $D_2$ )および、量子化パラメータの操作を適用した Central Decoder( $D_0$ )の PSNR の測定結果である。なお、ドリフト補償を行わなかった時の PSNR の結果は、Central Decoder(w/o drift compensation)で示す。横軸は、ビデオストリームのフレーム数、縦軸は、PSNR を示す。ビデオストリームの各フレームに対して PSNR を測定した結果である。

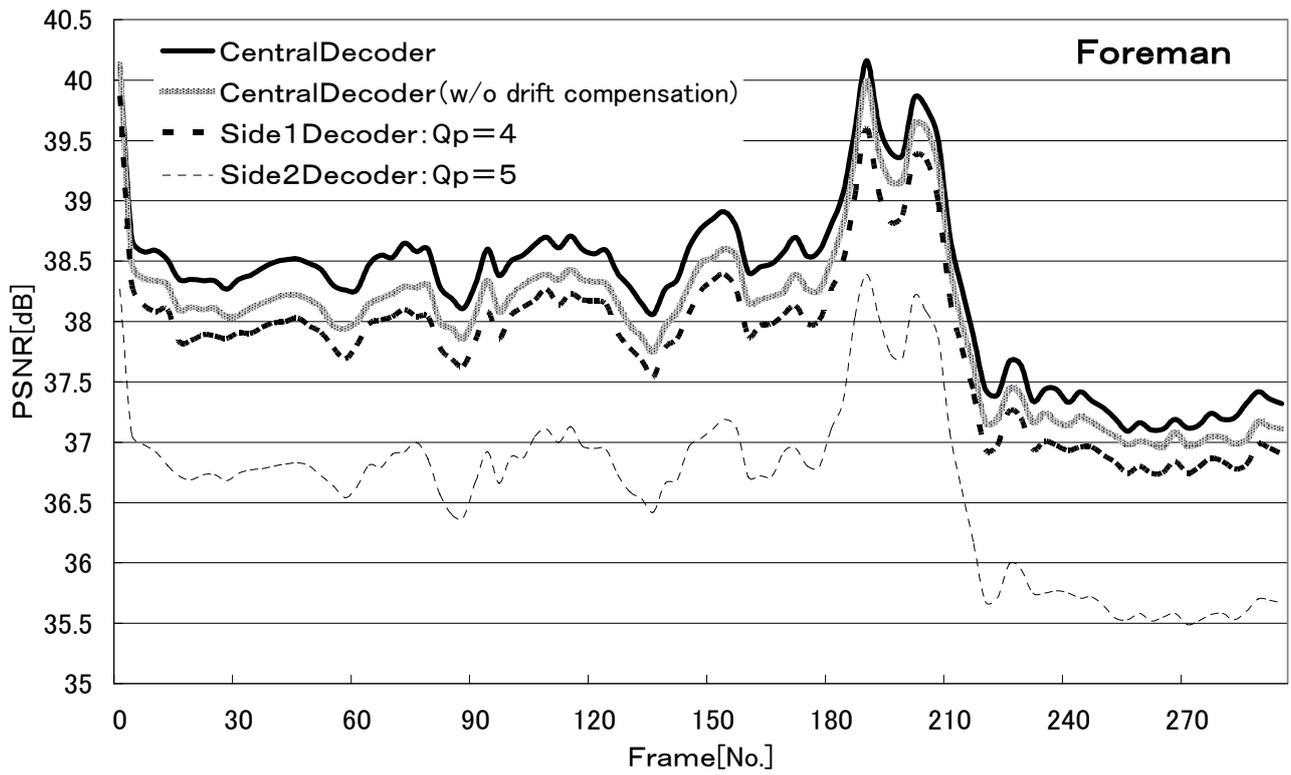
図 2-27 は、評価ビデオストリーム Foreman、図 2-28 は、評価ビデオストリーム Mobile&Calendar を用いて、下記の 3 つのパターンで実験を行った。

(a)  $D_1$  に  $Qp=4$ 、 $D_2$  に  $Qp=5$  を設定し、 $D_0$ 、 $D_1$ 、 $D_2$ 、ドリフト補償無しの PSNR 測定

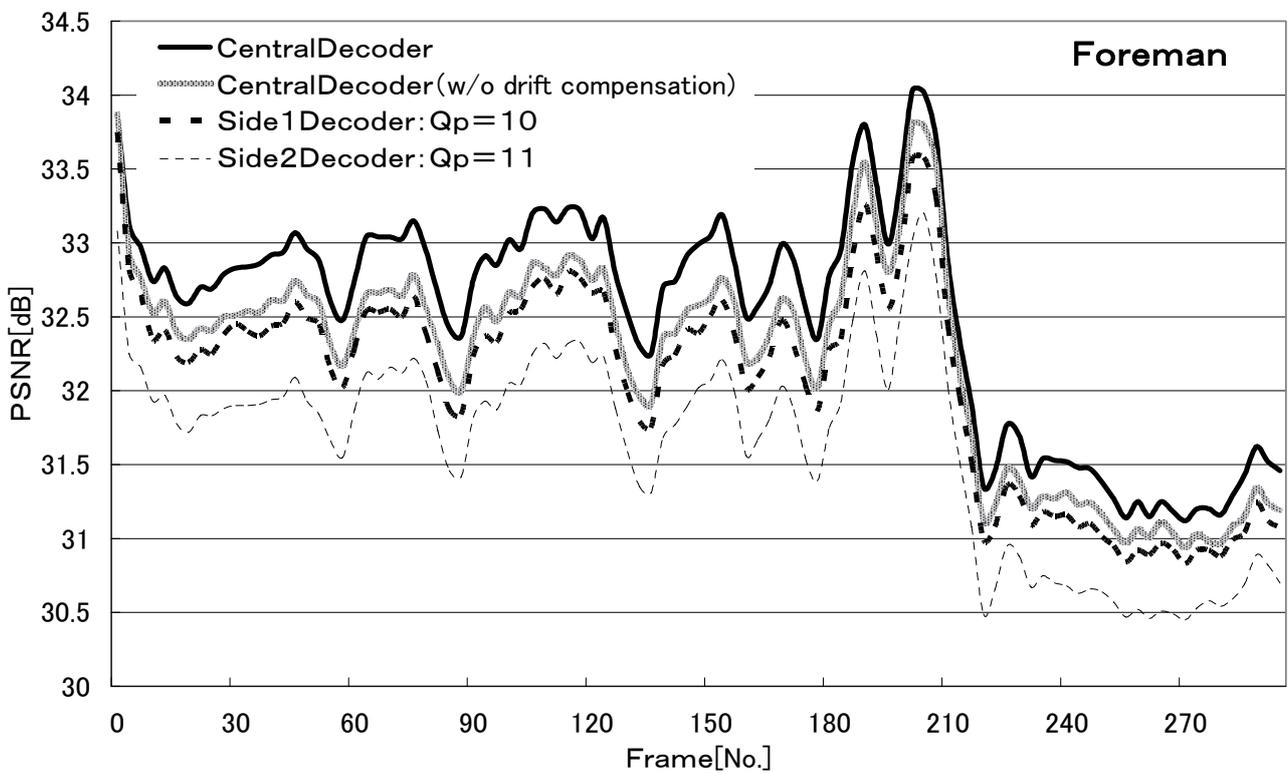
(b)  $D_1$  に  $Qp=10$ 、 $D_2$  に  $Qp=11$  を設定し、 $D_0$ 、 $D_1$ 、 $D_2$ 、ドリフト補償無しの PSNR 測定

(c)  $D_1$  に  $Qp=24$ 、 $D_2$  に  $Qp=25$  を設定し、 $D_0$ 、 $D_1$ 、 $D_2$ 、ドリフト補償無しの PSNR 測定

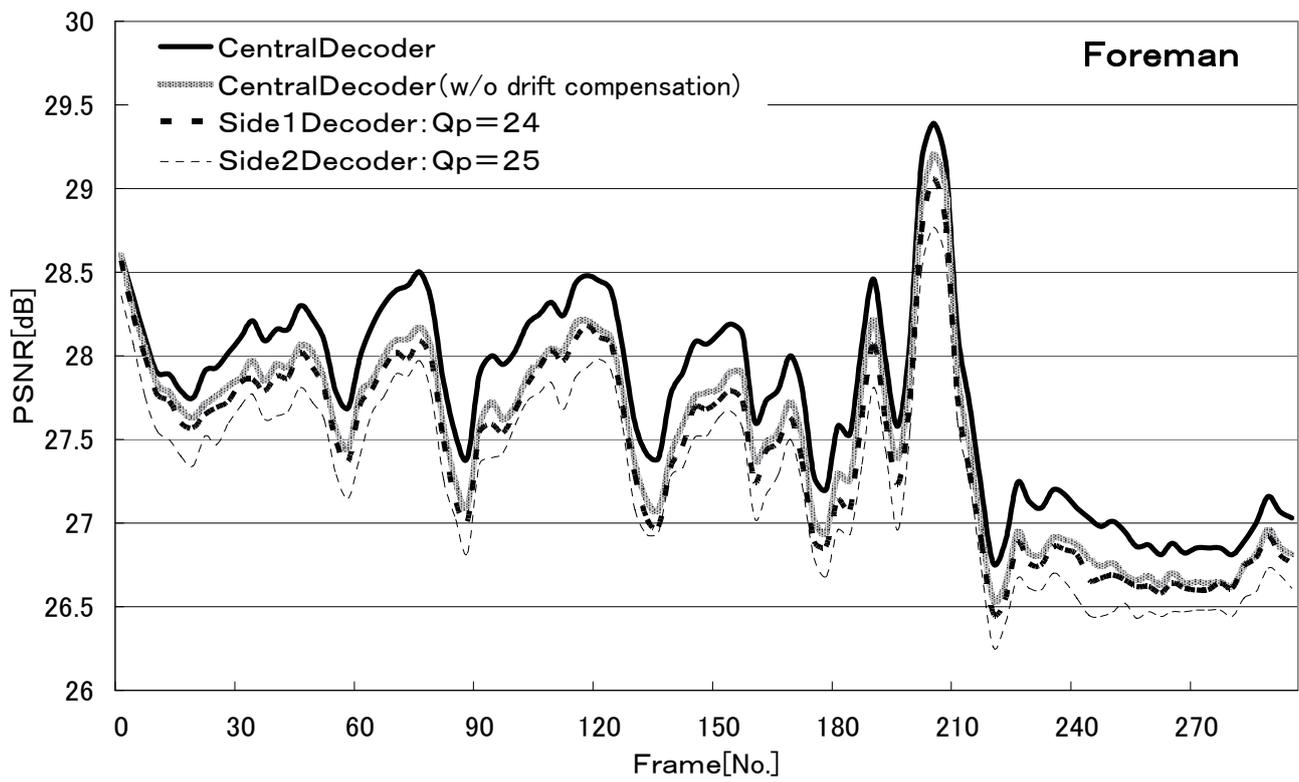
なお、PSNR は、各フレーム単位に原画像と復号した画像との差分を平均二乗誤差で測定する。



(a)  $D_1$  ( $Q_p=4$ ),  $D_2$  ( $Q_p=5$ ) と  $D_0$  の PSNR

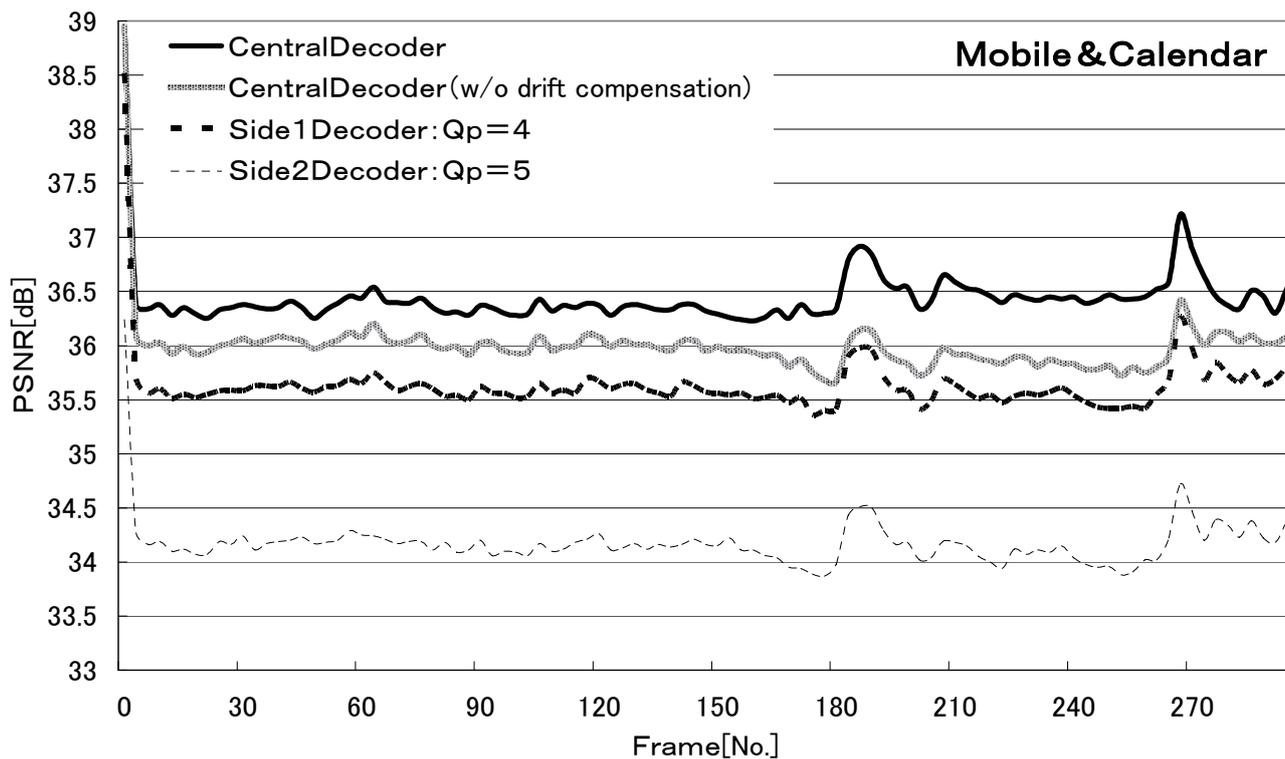


(b)  $D_1$  ( $Q_p=10$ ),  $D_2$  ( $Q_p=11$ ) と  $D_0$  の PSNR

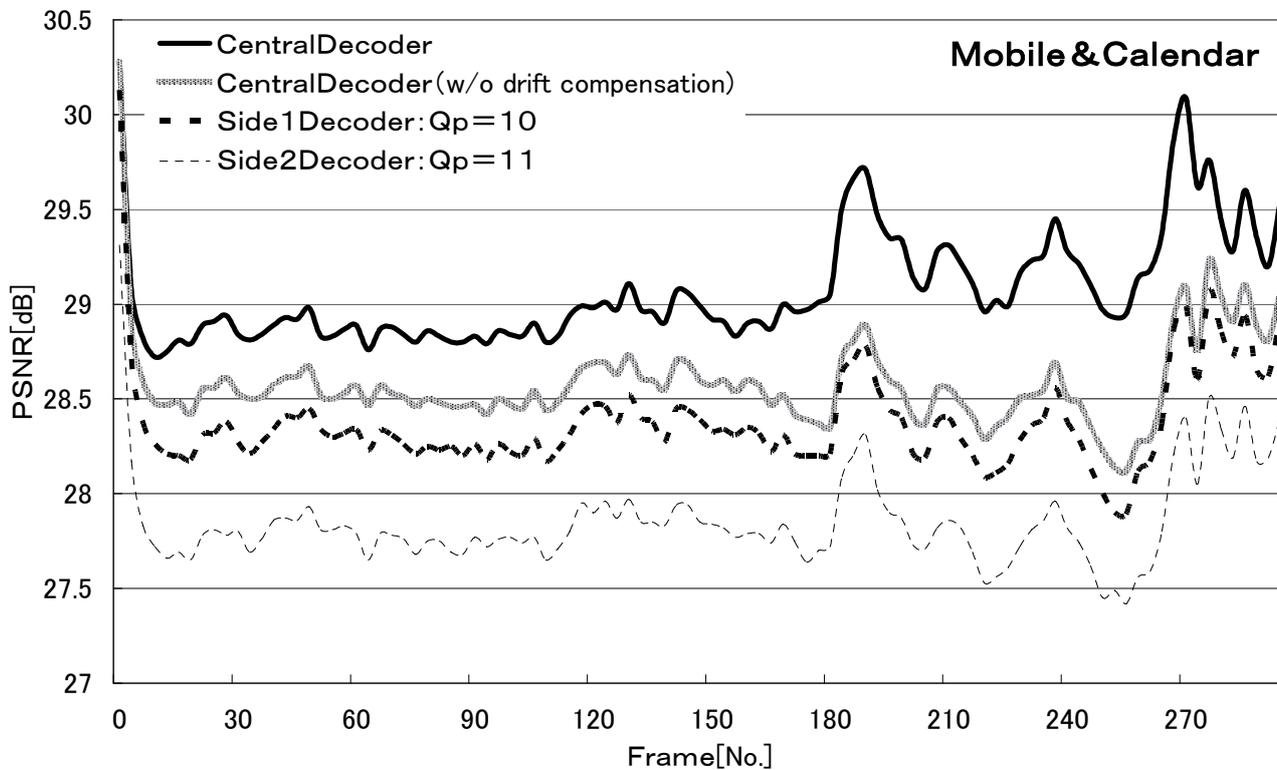


(c)  $D_1$  ( $Q_p=24$ ),  $D_2$  ( $Q_p=25$ )と  $D_0$  の PSNR

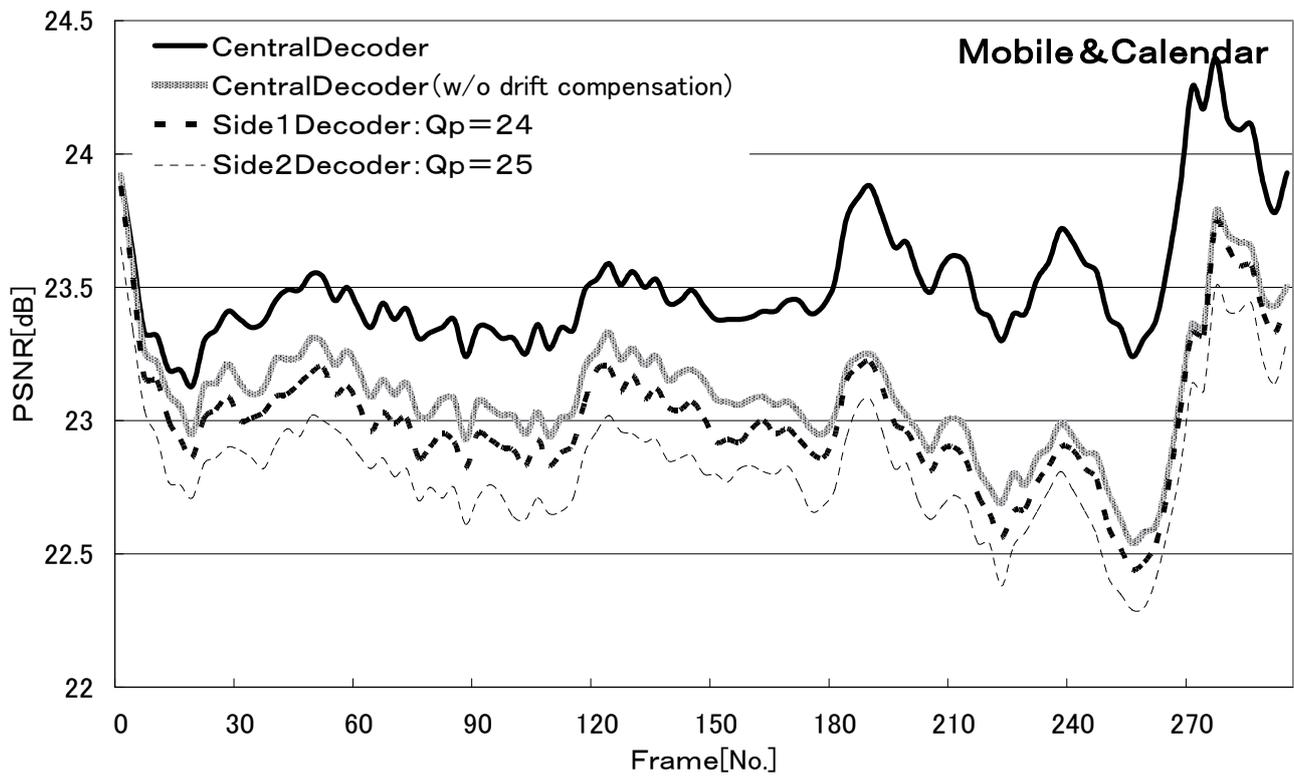
図 2-27  $D_1$  と  $D_2$  および  $D_0$  との PSNR 比較(評価ビデオストリームは Foreman)



(a)  $D_1$  ( $Q_p=4$ ),  $D_2$  ( $Q_p=5$ ) と  $D_0$  の PSNR



(b)  $D_1$  ( $Q_p=10$ ),  $D_2$  ( $Q_p=11$ ) と  $D_0$  の PSNR



(c)  $D_1$  ( $Q_p=24$ ),  $D_2$  ( $Q_p=25$ )と  $D_0$  の PSNR

図 2-28  $D_1$  と  $D_2$  および  $D_0$  との PSNR 比較(評価ビデオストリームは Mobile&Calendar)

図 2-27(a)では、 $D_1(Qp=4)$ と $D_2(Qp=5)$ を比較すると約 1.7dB の差で $D_1$ の品質が良い。これは、MPEG-4 の規格によって  $Qp$  から導き出される  $Qstep$  が決められており、それに基づいて量子化を行っているためである。なお、通常、0.2dB の差があれば主観的に違いがわかる評価値[54]である。次に、 $D_0$  とドリフト補償なしとの比較では、約 0.2-0.3dB の差で品質が良い結果となり、 $D_0$  と $D_1$ との比較では、約 0.5-0.6dB の差で品質の良い結果となった。この結果、ドリフト補償は有効であり、提案手法の Quantizer MD が有効であることがわかった。ただし、180 フレームから 200 フレームの間では、 $D_0$  とドリフト補償なしで差がなくなっている。このシーンは、映像ソースとして、ほとんど動かない映像なので、符号量が少なくなり、一時的にドリフトが発生しなかったと思われる。

図 2-27(b)では、 $D_1(Qp=10)$ と $D_2(Qp=11)$ を比較すると約 0.4dB の差で $D_1$ の品質が良い。これは、図 2-27(a)と同様の理由である。次に、 $D_0$  とドリフト補償なしとの比較では、約 0.3-0.4dB の差で品質が良い結果となり、 $D_0$  と $D_1$ との比較では、約 0.4-0.5dB の差で品質の良い結果となった。この結果、ドリフト補償は有効であり、提案手法の Quantizer MD が有効であることがわかった。

ただし、図 2-27(a)と比べて、Quantizer MD の優位差が少なくなっていることがわかる。これは、 $Qp$  値が大きくなると量子化が粗くなるため、*range* の重なり合いが少なくなり、Quantizer MD の有効性が低くなっていると思われる。

図 2-27(c)では、 $D_1(Qp=24)$ と $D_2(Qp=25)$ を比較すると約 0.2dB の差で $D_1$ の品質が良い。これは、図 2-27(a)と同様の理由である。次に、 $D_0$  とドリフト補償なしとの比較では、約 0.2-0.3dB の差で品質が良い結果となり、 $D_0$  と $D_1$ との比較では、約 0.3-0.4dB の差で品質の良い結果となった。この結果、ドリフト補償は有効であり、提案手法の Quantizer MD が有効であることがわかった。

ただし、図 2-27(b)と同じように、Quantizer MD の優位差が少なくなっていることがわかる。これも、 $Qp$  値が大きくなると量子化が粗くなるため、*range* の重なり合いが少なくなり、Quantizer MD の有効性が低くなっていると思われる。

図 2-28(a)では、 $D_1(Qp=4)$ と $D_2(Qp=5)$ を比較すると約 1.7dB の差で $D_1$ の品質が良い。これは、図 2-27(a)と同様の理由である。次に、 $D_0$  とドリフト補償なしとの比較では、約 0.4-0.5dB の差で品質が良い結果となり、 $D_0$  と $D_1$ との比較では、約 0.8-1.0dB の差で品質の良い結果となった。この結果、ドリフト補償は有効であり、提案手法の Quantizer MD が有効であることがわかった。この映像ソース Mobile&Calendar が Foreman より良い結果となったのは、符号量の多い映像であるため、*range* の重なり合いが多く発生し、Quantizer MD の有効性が高くなっていると思われる。

図 2-28(b)では、 $D_1(Qp=10)$ と $D_2(Qp=11)$ を比較すると約 0.4dB の差で $D_1$ の品質が良い。これは、図 2-27(a)と同様の理由である。次に、 $D_0$  とドリフト補償なしとの比較では、約 0.4-0.7dB の差で品質が良い結果となり、 $D_0$  と $D_1$ との比較では、約 0.7-0.9dB の差で品質の良い結果となった。この結果、ドリフト補償は有効であり、提案手法の Quantizer MD が有効であることがわかった。これも図 2-28(a)と同様に符号量の多い映像であるため、*range* の重なり合いが多く発生し、Quantizer MD の有効性が高くなっていると思われる。

図 2-28(c)では、 $D_1(Qp=24)$ と $D_2(Qp=25)$ を比較すると約 0.2dB の差で $D_1$ の品質が良い。これは、図 2-27(a)と同様の理由である。次に、 $D_0$  とドリフト補償なしとの比較では、約 0.1dB の差で品質が良い結果となり、 $D_0$  と $D_1$ との比較では、約 0.3-0.7dB の差で品質の良い結果となった。この結果、ドリフト補償は有効であり、提案手法の Quantizer MD が有効であることがわかった。

ただし、図 2-28(b)と同じように、Quantizer MD の優位差が少なくなっていることがわかる。これも、Qp 値が大きくなると量子化が粗くなるため、range の重なり合いが少なくなり、Quantizer MD の有効性が低くなっていると思われる。

#### (4) 考察

上記シミュレーションの結果で得られたように、2 つの映像ソースにおいて、3 つのパターンで評価を行った。すべてのパターンにおいて、ドリフト補償の有効性が確認でき、提案手法の Quantizer MD が有効であることがわかった。また、符号量が多くなれば、range の重なり合いが多く発生し、Quantizer MD の有効性が高くなっていると思われる。

ただし、Qp 値が大きくなると量子化が粗くなるため、range の重なり合いが少なくなり、Quantizer MD の有効性が低くなっていると思われる。

考察より、Qp 値が大きくなるに従って量子化が粗くなるため、range の重なり合いが少なくなると思われるので検証を加えた。

図 2-29 に映像ソース foreman、図 2.30 に Mobile&Calendar について、それぞれの range1 と range2 の包含関係（重なり合い、包含、重ならない）を比率で示す。横軸は、Qp1 および Qp2 の設定値。縦軸は、range1 と range2 の包含関係を比率で示す。

図 2-29 を見ると、Qp1=4 と Qp2=5 の時の range1 と range2 の重なる比率は約 28%、包含される比率は約 10%、重ならない比率は 62%であった。Qp1=10 と Qp2=11 の時の range1 と range2 の重なる比率は約 9%、range1 が range2 に包含される比率は約 2%、重ならない比率は 89%であった。Qp1=24 と Qp2=25 の時の range1 と range2 の重なる比率は約 2%、range1 が range2 に包含される比率は約 0.5%、重ならない比率は 97.5%であった。

図 2-30 を見ると、Qp1=4 と Qp2=5 の時の range1 と range2 の重なる比率は約 38%、range1 が range2 に包含される比率は約 13%、重ならない比率は 49%であった。Qp1=10 と Qp2=11 の時の range1 と range2 の重なる比率は約 15%、range1 が range2 に包含される比率は約 3%、重ならない比率は 82%であった。Qp1=24 と Qp2=25 の時の range1 と range2 の重なる比率は約 4%、range1 が range2 に包含される比率は約 1%、重ならない比率は 95%であった。つまり、図 2-29 と図 2-30 を見ても確認できるように、Qp 値が大きくなると量子化が粗くなるため、range の重なり合う比率が少なくなり、Quantizer MD の有効性が低くなることがわかった。

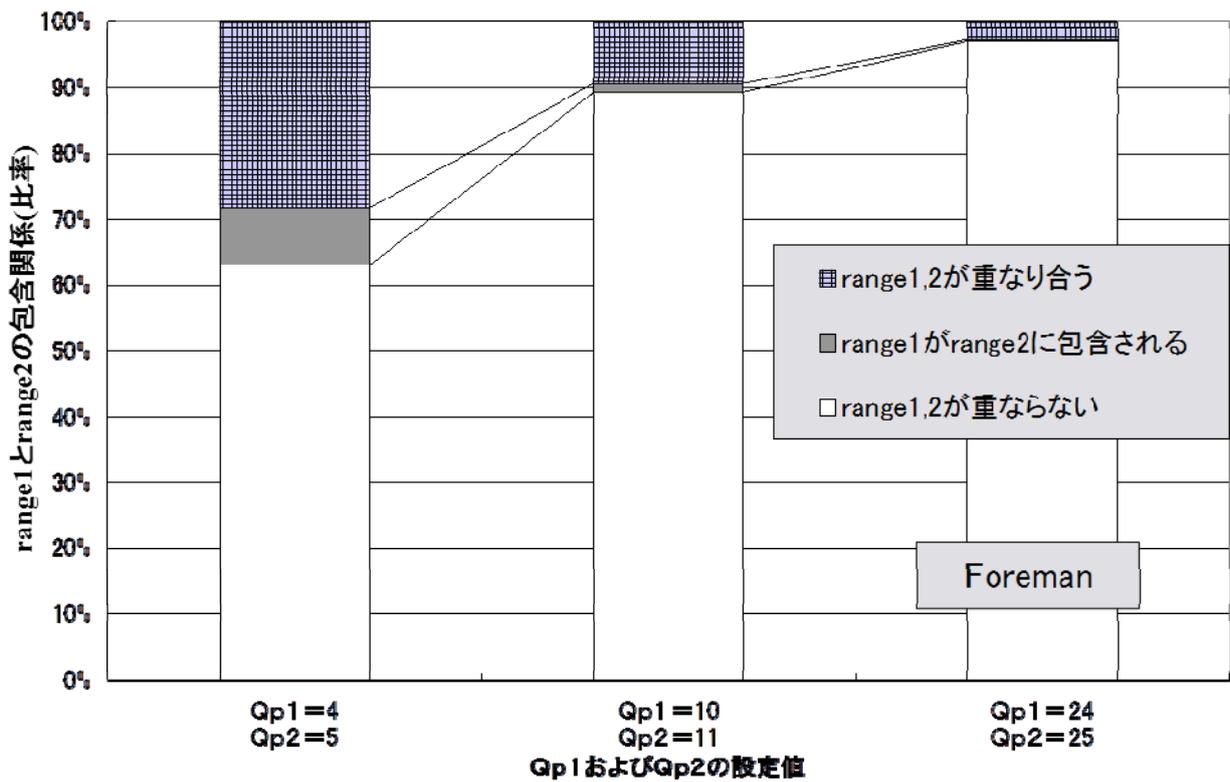


図 2-29 D<sub>0</sub>における *range1* と *range2* の包含関係(Foreman)

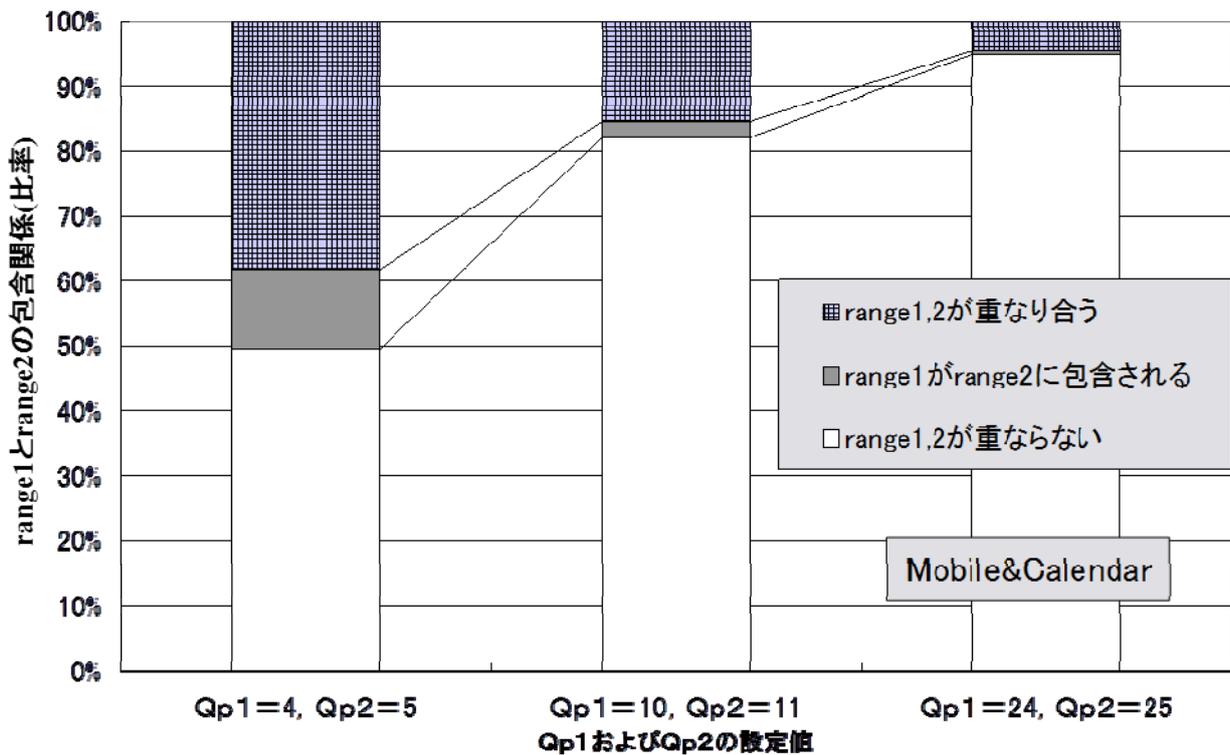


図 2-30 D<sub>0</sub>における *range1* と *range2* の包含関係(Mobile&Calendar)

考察の最後として、 $D_1$  および  $D_2$  における  $Qp$  値の設定指針を得るために *range1* と *range2* の  $Qp$  値の間隔を広げて行く検証を加えた。

図 2-31 に映像ソース *foreman*、図 2-32 に *Mobile&Calendar* について、それぞれの *range1* と *range2* に設定する値を変化させた。横軸は、 $Qp1$  および  $Qp2$  の設定値。縦軸は、*range1* と *range2* の包含関係を比率で示す。

図 2-31 に、 $Qp1=2$  を固定し、 $Qp2$  の値を  $Qp2=3$  から  $Qp2=10$  までを設定した場合について *range1* と *range2* での包含関係を示した。

図 2-31 を見ると、 $Qp1=2$  と  $Qp2=3$  の時の *range1* と *range2* の重なる比率は約 41%，包含される比率は約 30%，重ならない比率は 29%であった。 $Qp1=2$  と  $Qp2=4$  の時の *range1* と *range2* の重なる比率は約 26%，包含される比率は約 65%，重ならない比率は 9%であった。

$Qp1=2$  と  $Qp2=5$  の時の *range1* と *range2* の重なる比率は約 20%，包含される比率は約 74%，重ならない比率は 6%であった。 $Qp1=2$  と  $Qp2=6$  の時の *range1* と *range2* の重なる比率は約 13%，包含される比率は約 84%，重ならない比率は 3%であった。 $Qp1=2$  と  $Qp2=7$  の時の *range1* と *range2* の重なる比率は約 10%，包含される比率は約 88%，重ならない比率は 2%であった。 $Qp1=2$  と  $Qp2=8$  の時の *range1* と *range2* の重なる比率は約 7%，包含される比率は約 91.5%，重ならない比率は 1.5%であった。 $Qp1=2$  と  $Qp2=9$  の時の *range1* と *range2* の重なる比率は約 5%，包含される比率は約 94%，重ならない比率は 1%であった。

$Qp1=2$  と  $Qp2=10$  の時の *range1* と *range2* の重なる比率は約 4%，包含される比率は約 95.5%，重ならない比率は 0.5%であった。

図 2-32 を見ると、 $Qp1=2$  と  $Qp2=3$  の時の *range1* と *range2* の重なる比率は約 44%，包含される比率は約 36%，重ならない比率は 20%であった。 $Qp1=2$  と  $Qp2=4$  の時の *range1* と *range2* の重なる比率は約 30%，包含される比率は約 62%，重ならない比率は 8%であった。 $Qp1=2$  と  $Qp2=5$  の時の *range1* と *range2* の重なる比率は約 23%，包含される比率は約 71%，重ならない比率は 6%であった。 $Qp1=2$  と  $Qp2=6$  の時の *range1* と *range2* の重なる比率は約 15%，包含される比率は約 81%，重ならない比率は 4%であった。 $Qp1=2$  と  $Qp2=7$  の時の *range1* と *range2* の重なる比率は約 12%，包含される比率は約 85%，重ならない比率は 3%であった。 $Qp1=2$  と  $Qp2=8$  の時の *range1* と *range2* の重なる比率は約 9%，包含される比率は約 89%，重ならない比率は 2%であった。 $Qp1=2$  と  $Qp2=9$  の時の *range1* と *range2* の重なる比率は約 8%，包含される比率は約 91%，重ならない比率は 1%であった。 $Qp1=2$  と  $Qp2=10$  の時の *range1* と *range2* の重なる比率は約 7%，包含される比率は約 92.5%，重ならない比率は 0.5%であった。

つまり、図 2-31 と図 2-32 を見ても確認できるように、*range1* と *range2* の  $Qp$  値の間隔を広げて行くと重なる比率が小さくなるので、隣り合う量子化パラメータを  $Qp$  値とした方が、Quantizer MD の有効性が高くなることが確認できた。

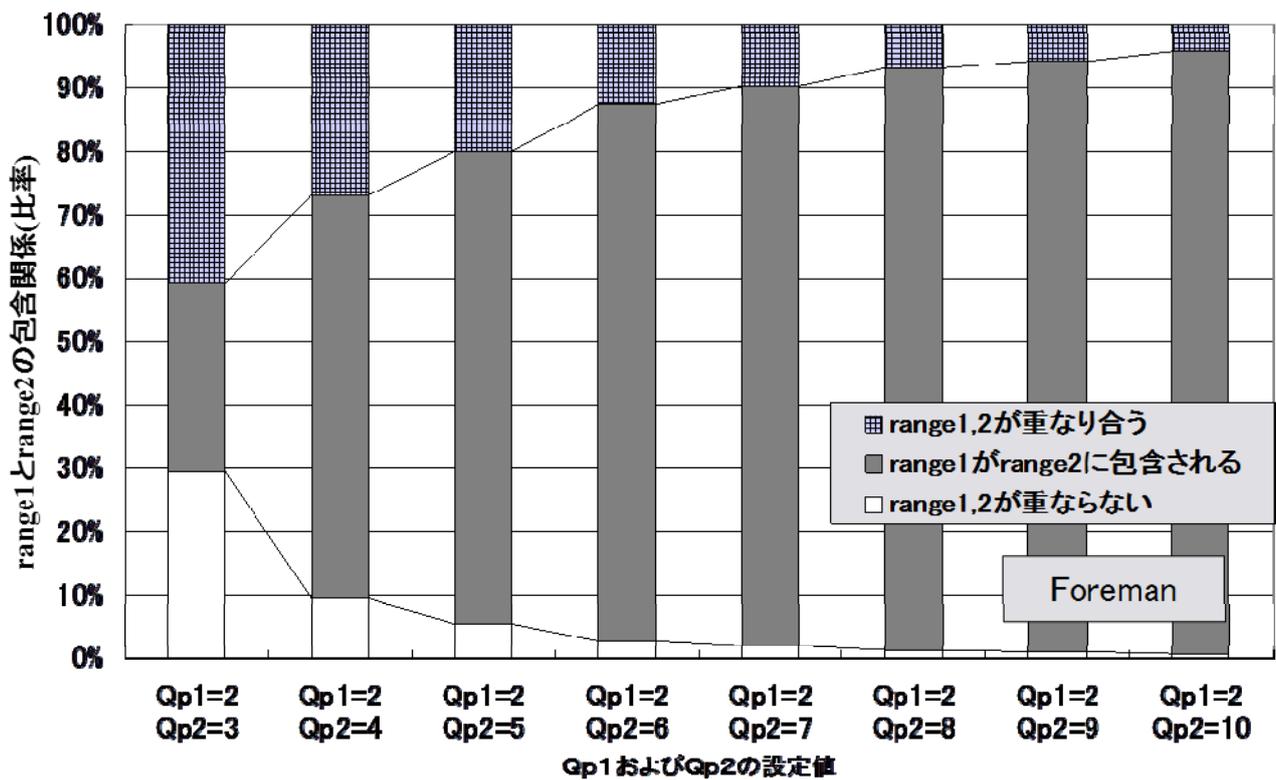


図 2-31 量子化パラメータを設定した場合の包含関係(Foreman)

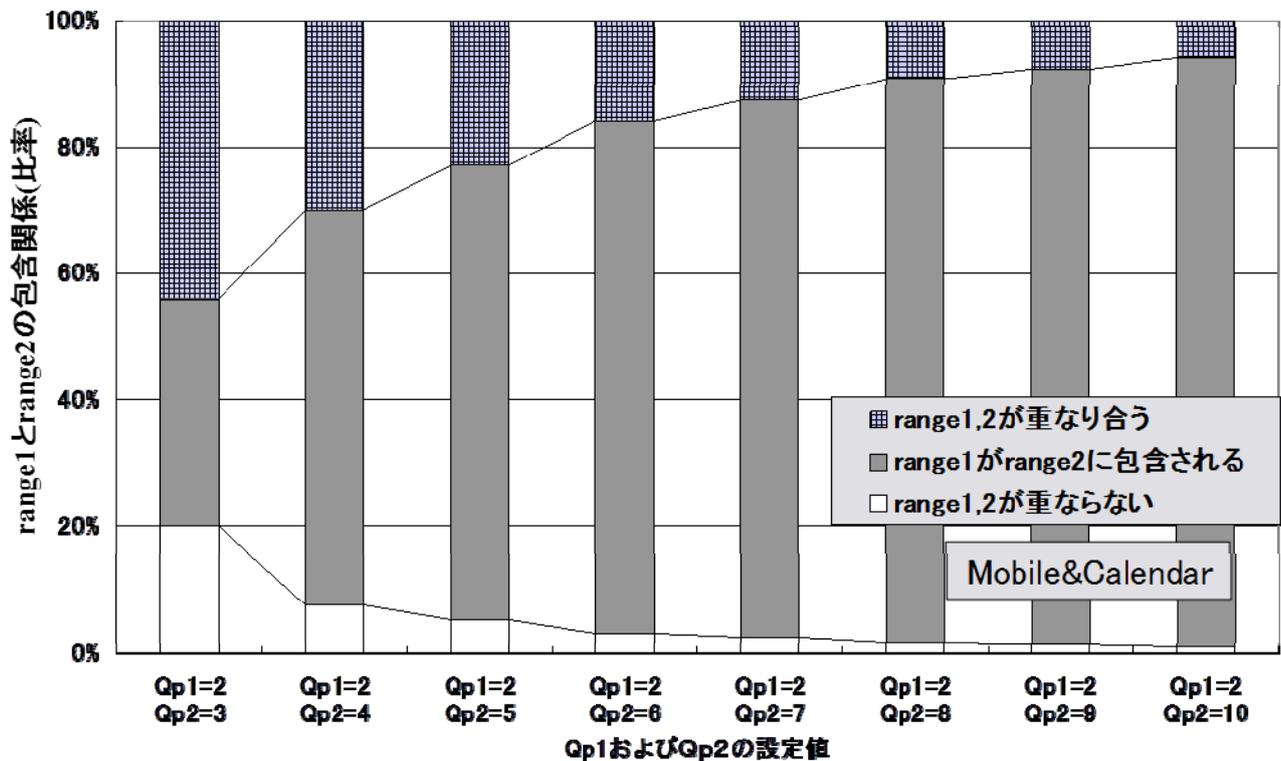


図 2-32 量子化パラメータを設定した場合の包含関係(Mobile&Calendar)

#### 2.3.4 まとめ

シミュレーションにより，映像ソースが Foreman を用いた実験では，

- (a)  $D_1$  に  $Qp=4$ ，  $D_2$  に  $Qp=5$  を設定した環境で，  $D_1$  よりも約 0.5-0.6dB 優れていた．
- (b)  $D_1$  に  $Qp=10$ ，  $D_2$  に  $Qp=11$  を設定した環境で，  $D_1$  よりも約 0.4-0.5dB 優れていた．
- (c)  $D_1$  に  $Qp=24$ ，  $D_2$  に  $Qp=25$  を設定した環境で，  $D_1$  よりも約 0.3-0.4dB 優れていた．

映像ソースが Mobile&Calendar を用いた実験では，

- (a)  $D_1$  に  $Qp=4$ ，  $D_2$  に  $Qp=5$  を設定した環境で，  $D_1$  よりも約 0.8-1.0dB 優れていた．
- (b)  $D_1$  に  $Qp=10$ ，  $D_2$  に  $Qp=11$  を設定した環境で，  $D_1$  よりも約 0.7-0.9dB 優れていた．
- (c)  $D_1$  に  $Qp=24$ ，  $D_2$  に  $Qp=25$  を設定した環境で，  $D_1$  よりも約 0.3-0.7dB 優れていた．

さらに，実験結果より，各 Side Decoder において，真値が取り得る範囲の重なる比率が多くなれば，PSNR が向上することも確認できた．

最後に，MDVC における量子化パラメータの設定要件をまとめ，隣り合う量子化パラメータを設定することで PSNR の向上できることを示した．

なお，本節における産業上で想定される利用イメージは，異種無線アクセス網を想定しており，MDVC を用いた映像配信サービスを行う場合，隣り合う量子化パラメータを設定すると効果的であるため，無線アクセス網の速度が高ビットレートに対応していれば，例えば，チャンネル 1 に  $Qp=4$ ，チャンネル 2 に  $Qp=5$  を設定する．低ビットレートならば，チャンネル 1 に  $Qp=24$ ，チャンネル 2 に  $Qp=25$  を設定すると効果がある．もし仮にチャンネル間の転送レートに大きな差が生じている場合は，遅い方のチャンネルは，送受信側のバッファ管理で解決できるが，この場合，再生画像に遅延が生じることになるので蓄積メディアからの映像配信サービスに限定される．

## 第3章 ビデオ符号化に基づく歩行者数の推定評価

### 3.1 序言

近年、監視カメラの映像は MPEG-4 や H.264/AVC のような符号化方式を使って圧縮し、蓄積されることが多くなってきた。これら圧縮されたビデオストリームを用いて移動物体を解析することで、解析データの前処理が不要となり、処理量が大幅に削減され、リアルタイムに解析できる特徴がある。そのビデオストリームから得られるパラメータを用いて移動物体を追跡・推定する方法も提案されている。H.264/AVC の動きベクトル(Motion Vector 以下, MV)をクラスタリングして移動物体をリアルタイムに追跡する例[22], 複数フレームにわたる類似した MV を識別して移動物体を追跡する例[23][24], MPEG の MV と DCT 係数の成分比較から移動物体を追跡する例[25]-[28], MV と画像の色情報からオブジェクトを追跡する例[29]が報告されている。追跡については、モデル化した歩行者の外観と画像の相関関係から歩行者を抽出する例[30][31][32]も多い。そのほか、ある領域の色成分の割合を事前に設定したテンプレートと比較して追跡対象を抽出し、その位置や動きを推定する例[33], 空間の色情報分布からオクルージョンの影響を予測し、追跡精度を向上する例[34]がある。歩行者数の推定では、移動するオブジェクトやそのエッジから抽出した画素をクラスタリングし、得られたクラスタ数から歩行者数を推定する例[35]-[39]がある。コーナー検出法[40]に基づいて集計された MV を用いて移動物体を識別し、それらから統計的に推定する方法[41], 近傍の輝度勾配情報から抽出した特徴がある点[42]を用いて推定する例[43]が報告されている。また、統計手法としては、回帰分析を用いて行われることが多く、テンプレート画像とのマッチングにより抽出された情報源を用いて人数推定する方法[44]や歩行者の特徴量に基づいてトレーニングデータを切り出し、ガウス過程により推定している方法[45]がある。これらは、異なるコンテンツに対応したテンプレート画像の精度やトレーニングデータの特徴量を抽出する精度に大きく依存する。また、歩行者の領域を抽出し、重みパラメータによって特徴量を効率的に選出した情報源を用いて人数を推定する方法[46]がある。これまで紹介した手法は、一旦別処理で特徴点を抽出後、歩行者数の推定を行う手法や特徴点を示す学習データが増えるコンテンツを扱う手法は多くの計算量が必要となり、リアルタイムに歩行者数の推定が難しい技術的な課題があった。

本章の目的は、H.264/AVC ビデオ符号化器（以下、H.264/AVC）によって符号化されたビデオストリームを復号する際に抽出される予測残差信号からフレーム毎に回帰分析を用いて歩行者数を推定し、フレーム内の可変ブロックを参照して歩行者数をリアルタイムに補正する手法を提案する。

### 3.2 ビデオ符号化に基づく歩行者数のリアルタイム推定方式

#### 3.2.1 H.264/AVC の動き補償と歩行者数の相関

##### (1) H.264/AVC の可変ブロック構成

H.264/AVC で採用されている動き補償(Motion Compensation. 以下, MC)ブロックサイズは、 $16 \times 16$  画素、 $16 \times 8$  画素、 $8 \times 16$  画素、 $8 \times 8$  画素で構成するマクロブロック(Macro Block. 以下, MB)単位のタイプと、 $8 \times 8$  画素をさらに分割した  $8 \times 4$  画素、 $4 \times 8$  画素、 $4 \times 4$  画素のサブマクロブロック(Sub Macro Block. 以下, SubMB)単位の計 7 種類のブロックサイズを指定した MC が行える。これは  $16 \times 16$  ブロック内で動きが激しい部分と平坦な部分がある場合には、柔軟にブロックサイズが変更できるためフレーム間で予測残差信号(Sum of Absolute Difference. 以下, SAD)が削減できる。

## (2) 量子化パラメータによる予測残差信号の変化

H.264/AVC で符号化される時に予測残差信号 SAD が生成される。一般に、符号化アルゴリズムにおいては、移動オブジェクトの動きが大きいところでは、予測が外れる傾向にあり、SAD は移動オブジェクトのエッジ情報として使える可能性がある[55]。量子化パラメータ(Quantization Parameter 以下, QP)によって、SAD とエッジ情報の対応の様子が異なる。H.264/AVC は、QP を 0~51 まで設定(数字が大きいほど復号した画像は荒くなる)できる。

## (3) 予測残差信号と歩行者数の相関

各フレームの画素が持つ SAD の総和(以下, SAD 総和値)を求め、歩行者数との相関を図 3-1 に示す。縦軸に SAD 総和値, 横軸は歩行者数である。図 3-2 に使用したビデオストリームのスナップショットを示す。図 3-1 より SAD 総和値と歩行者数との相関は高い傾向にあることがわかる。なお、統計的にはデータの数によって異なるが、一般的には相関係数  $R$  の二乗(以下,  $R^2$ )は、0.7 より大きいと相関が高い[56]とされている。また、図 3-3 に歩行者 1 人当たりの SAD 総和値を示す。縦軸に歩行者 1 人当たりの SAD 総和値, 横軸は歩行者数である。図 3-3 より歩行者数が多くなる程、1 人当たりの SAD 総和値は、対数曲線的に小さくなる。これは、オクルージョンの影響により歩行者数が多くなると 1 人当たりの SAD 総和値が少なくなるためと思われる。そこで、H.264/AVC によって符号化されたビデオストリームを復号する際に抽出される SAD からフレーム毎に回帰分析を用いて歩行者数を推定し、フレーム内の可変ブロックを参照してオクルージョンによって生じた誤差を補正する手法を提案する。

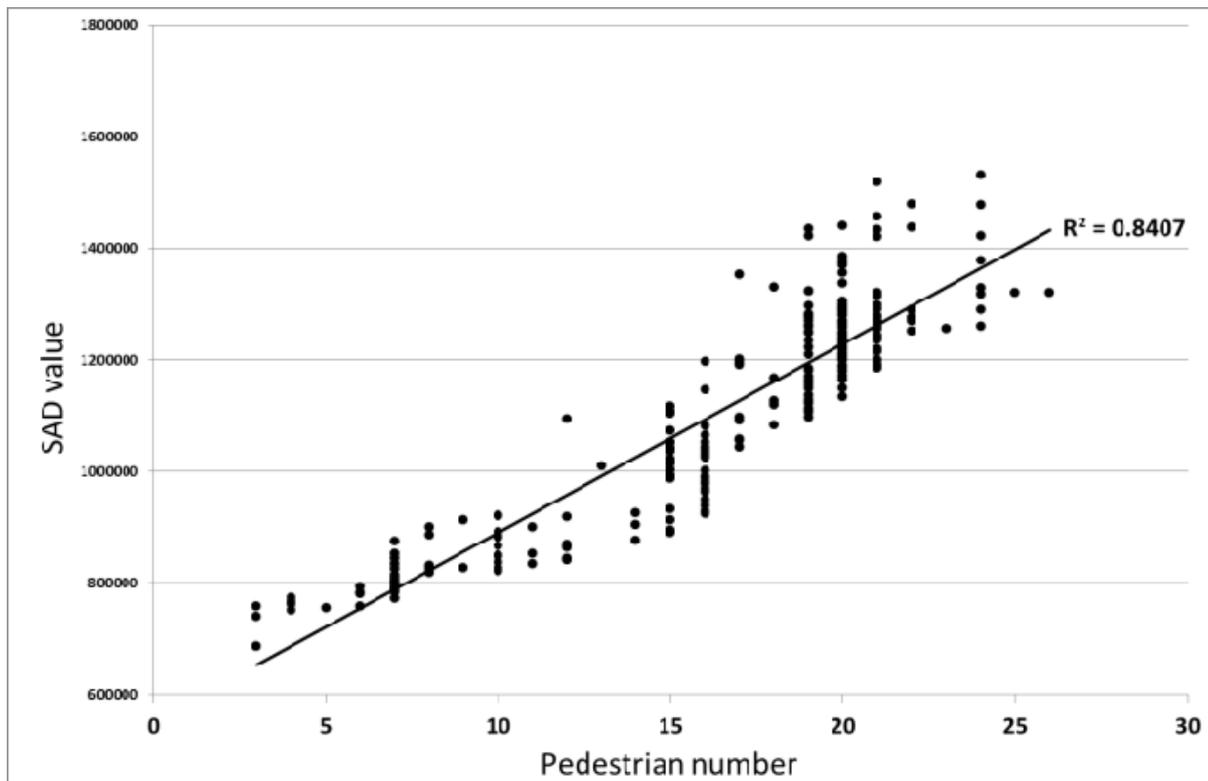


図 3-1 SAD 総和値と歩行者数の相関



図 3-2 ビデオストリームのスナップショット

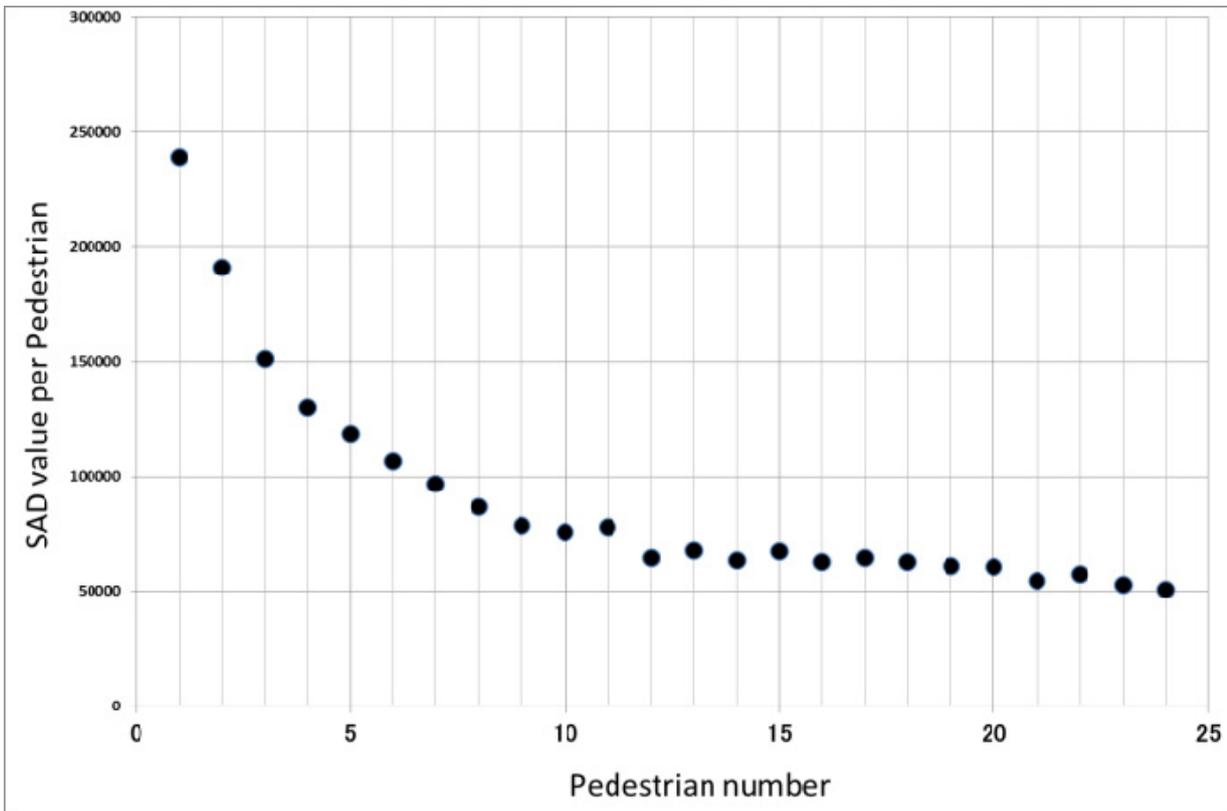


図 3-3 歩行者 1 人当たりの SAD 総和値

### 3.2.2 提案手法

#### (1) 全体構成

図 3-4 に歩行者数を推定する全体構成を示す。前処理にて評価ビデオストリームからトレーニングストリームを切り出し、そのトレーニングストリームから回帰式を算出する。次の歩行者推定処理で、前処理で算出した回帰式を用いて、評価ビデオストリームの各フレームの SAD 総和値から歩行者数を推定する。その歩行者数に対して、各フレームの可変ブロックを用いて補正し、最終的に歩行者を推定する。

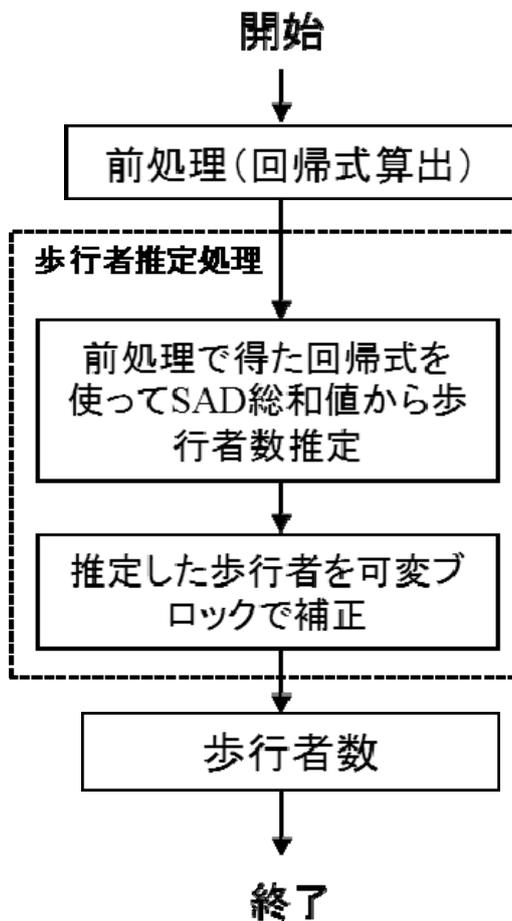


図 3-4 全体構成のフローチャート

(2) 前処理

歩行者数を推定するビデオストリームから数フレームずつ切り出してトレーニングストリームを作成し、SAD 総和値と各フレーム内の実際の歩行者数(以下、真値)との相関が最も高いトレーニングストリームより回帰式を求める。図 3-5 に前処理のフローチャートを示す。各フレームの画素が持つ SAD 総和値を求め(図 3-5-(1))、各フレーム内の実際の歩行者数との相関係数を算出し(図 3-5-(2))、SAD 総和値と真値の  $R^2$  が最大となった  $R^2$  の SAD 総和値と真値から回帰分析により回帰式を求める(図 3-5-(3))。

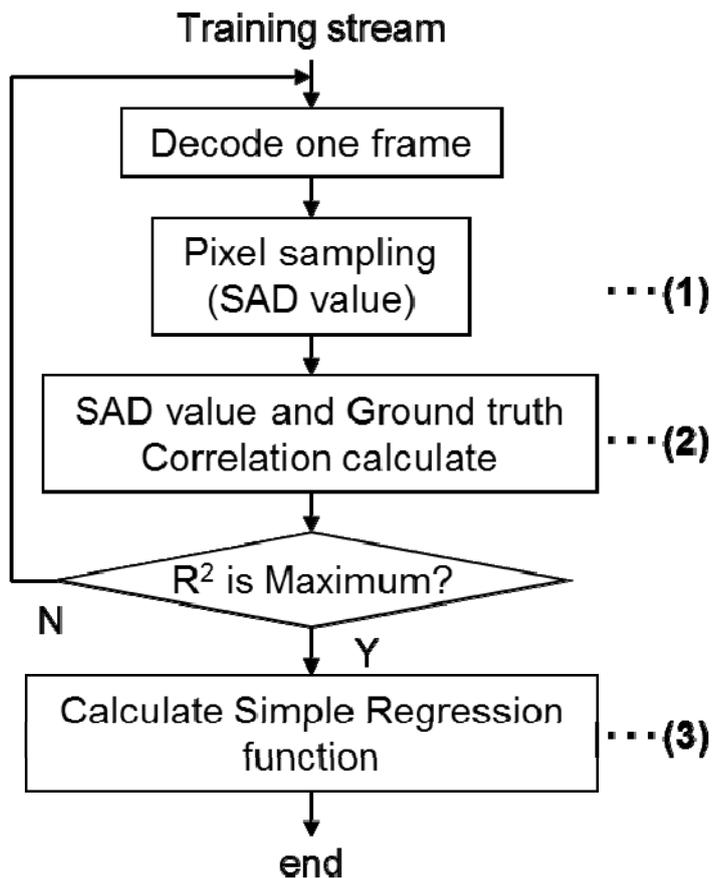


図 3-5 前処理のフローチャート

### (3) 歩行者数の推定処理

回帰式から歩行者数を推定する処理である。図 3-6 に処理フローを示す。ビデオストリームの各フレームを復号後、SAD 総和値を求める(図 3-6-(1))。図 3-5-(3)で求めた回帰式に図 3-6-(1)で求めた SAD 総和値を説明変数として代入し、予測された目的変数を各フレーム内の歩行者数として推定する(図 3-6-(2))。ここで、前処理に使うトレーニングストリームのフレーム数は、ビデオストリームに対して少なくするようにして回帰式を求めるため、図 3-3 で示すようにオクルージョンの影響により推定した歩行者数に誤差が生じるとされる。そこで図 3-6 で推定した各フレームにおける歩行者数に対して MB や SubMB を用いて補正する。MB は、差分画像の動きが少ない時や平坦な部分に選択される傾向にある。これは背景等のゆっくりした変化では効率的に符号化されるためである。一方、SubMB は、動きの激しい部分で選択されるので、歩行者の動き等に選択される確率が高い。そこで、フレーム内の MB と SubMB の割合から歩行者数の補正を行う(図 3-6-(3))。

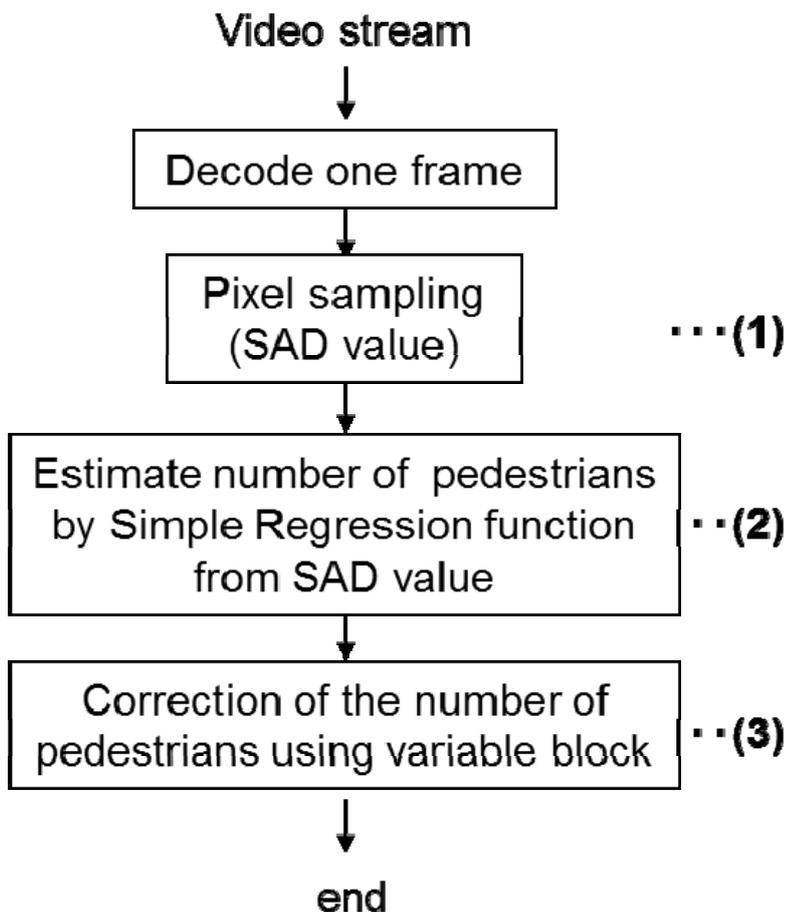


図 3-6 歩行者数の推定

具体的な補正手法は、各フレームの SAD 総和値を説明変数として図 3-5-(3)の回帰式を用いて予測した目的変数である歩行者数を  $N_i$ ，求めるフレームに含まれる可変ブロック (Variable block size number 以下，VBS)における SubMB の比率を  $r_1$ ，図 3-5 のトレーニングストリームで得られた歩行者数に含まれる VBS における SubMB の比率を  $r_2$  とする。ただし，トレーニングストリームで得られた歩行者数の最小値( $\alpha$ )，最大値( $\beta$ )とし，最小値の  $r_2$  を  $r_{2(\alpha)}$ ，最大値の  $r_2$  を  $r_{2(\beta)}$ とする。最終的な歩行者数  $N$  を式(3-1)で求める。

$N_i < \alpha$  の場合

$$N = N_i \frac{r_1}{r_{2(\alpha)}} \quad (3-1)$$

$N_i > \beta$  の場合

$$N = N_i \frac{r_1}{r_{2(\beta)}}$$

#### (4) 符号化に依存した抽出画総数の補正

H.264/AVC で符号化されたビデオストリームは，符号量の割り当てを GOP(Group of picture)単位で行うため，歩行者数を測定するようなビデオストリームを用いる場合，I(Intra-coded)フレームと P(Predicted)フレームで符号量に大きな差が生じる[57]。図 3-7 は，I フレームと P フレームの符号量推移を示す。横軸は，フレーム数，縦軸は符号量を示す。符号化した情報は， $I_n$  フレームおよび  $P_n(1) \sim P_n(14)$  ( $n=1, 2, 3 \dots$ ) フレームから構成される。図 3-7 を見ると各ピークが  $I_n$  フレームであることがわかる。提案手法では，MC が行われない  $I_n$  フレームは使用しない。そこで，歩行者数の連続性を維持する必要があるため  $I_n$  フレームの代わりに， $P_{n-1}(14)$  フレームと  $P_n(1)$  フレームの SAD 総和値の平均値を代入して算出する。

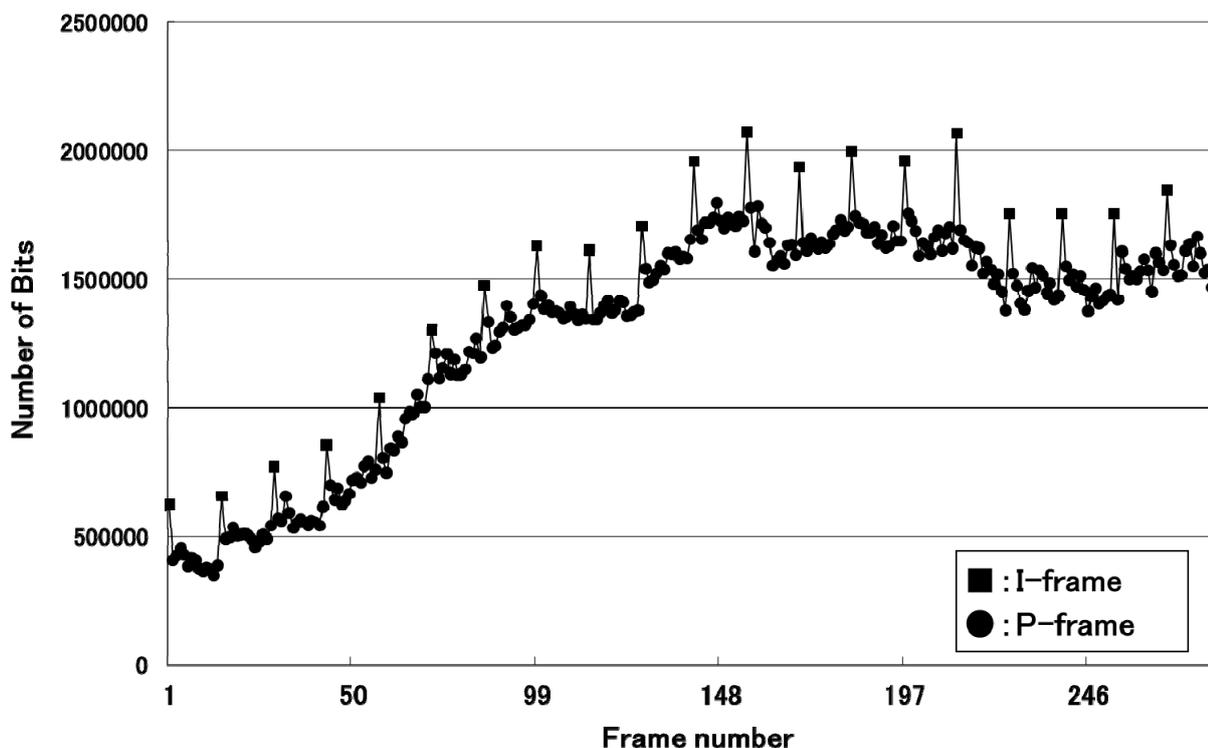


図 3-7 I フレームと P フレームの符号量の推移

### 3.2.3 シミュレーションおよび結果

#### (1) 評価諸元

評価ビデオストリームには、IEEE の workshop で背景差分や人物追跡で使われる PETS2009[58]のデータセット(表 3-1) を用いて、3章で説明した前処理(図 3-5)および歩行者数の推定(図 3-6)を実施する。なお、図 3-8 に評価ビデオストリームのスナップショットを示す。すべての評価ビデオストリームの解像度は 320×240 画素である。画像符号化方式の条件を表 3-2 に示す。画像符号化方式は、I, P フレームのみに対応しているシンプルな Base Profile を用いた。GOB 構造は、一般的な I フレームが 1 枚に対して P フレームは 14 枚の構成とした。

表 3-1 PETS2009 データセット

Video sequence	View	Flames	Conditions	Number of Pedestrian		
				Min	AVG	Max
S1.L1.13-57	1	221	medium density crowd, overcast	5	22.61	34
S1.L1.13-59	1	241	medium density crowd, overcast	3	15.81	26
S1.L2.14-06	1	201	high density crowd, overcast	0	26.28	43
S1.L3.14-17	1	91	medium density crowd, bright sunshine and shadows	6	24.34	41
S1.L1.13-57	2	221	medium density crowd, overcast	8	34.19	46
S1.L2.14-06	2	201	high density crowd, overcast	3	37.10	46
S1.L2.14-31	2	131	high density crowd, bright sunshine and shadows	10	35.19	43
S3.MF.12-43	2	108	very low density crowd, overcast	1	4.99	7

表 3-1 の View1 は、測定しているカメラの光軸と歩行者がほぼ平行な位置にある。つまり、歩行者数が多くなっても、オクルージョンによって人数のカウントが可能なビデオストリーム。一方 View2 は、測定しているカメラの光軸から歩行者が遠ざかるため、フレームの最後は、歩行者同士が完全に重なることを想定したビデオストリームである。

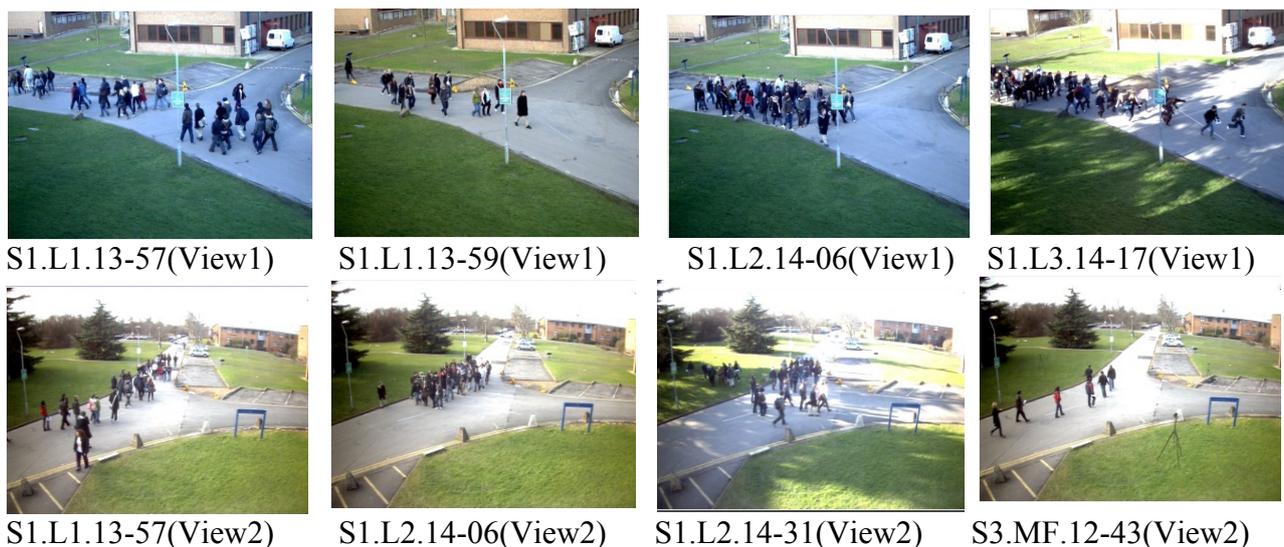


図 3-8 評価ビデオストリームのスナップショット

表 3-2 ビデオ符号化方式の条件

符号化方式	H.264 baseline profile
フレームレート	30 frames/s
GOP 構造	I フレーム x1 P フレーム x14
参照画像フレーム	直前のフレーム
QP 値	10
MV 探索範囲	±32
レート制御	OFF

評価指標としては、PETS2009 でも採用されている平均絶対誤差(Mean Absolute Error 以下, MAE)と平均相対誤差(Mean Relative Error 以下, MRE)を用いた. フレーム総数を  $M$  とすると, 1 フレームにおける推定歩行者数を  $E(i)$ , 真値を  $T(i)$  とすると MAE と MRE は式(3-2)(3-3)で定義される.

$$MAE = \frac{1}{M} \sum_{i=1}^n |E(i) - T(i)| \quad (3-2)$$

$$MRE = \frac{1}{M} \sum_{i=1}^n \frac{|E(i) - T(i)|}{T(i)} \quad (3-3)$$

比較する方式には、PETS2009 で良い評価を得ている Albiol[34]および Conte[36]の手法を選んだ. 両手法とも特徴点を示す学習データから回帰分析により歩行者数を推定しているため, 本手法との構成に合致している.

## (2) 評価構成

評価ビデオストリームを H.264/AVC 標準化テストモデルソフトウェア JM10.2[59]でエンコードし, その符号化データを使って, リアルタイムにデコードした結果からデータを取得し, 回帰分析を行った.

### (3) 実験結果

表 3-3 に Albiol, Conte および提案手法において, いくつかのフレームに対して評価を行った MAE/MRE の比較結果を示す. すべてのビデオストリームにおける MAE(人) および MRE(括弧内%数字)を表す. なお, 提案手法において可変ブロックによる補正を行っていない結果を VBS off, 補正を行っている結果を VBS on としている.

表 3-3 MAE/MRE 比較結果

Video sequence	View	Albiol	Conte	Proposed method	
				VBS off	VBS on
S1.L1.13-57	1	2.80(12.6%)	1.92(8.7%)	2.97(17.2%)	2.75(13.2%)
S1.L1.13-59	1	3.86(24.9%)	2.24(17.3%)	1.82(19.2%)	1.37(11.5%)
S1.L2.14-06	1	5.14(26.1%)	4.66(20.5%)	3.90(18.1%)	2.85(15.5%)
S1.L3.14-17	1	2.64(14.0%)	1.75(9.2%)	7.40(31.5%)	7.3(28.2%)
S1.L1.13-57	2	29.45(106.0%)	11.76(30.0%)	11.25(74.1%)	3.95(24.4%)
S1.L2.14-06	2	32.24(122.5%)	18.03(43.0%)	28.13(70.0%)	27.2(66.9%)
S1.L2.14-31	2	34.09(99.7%)	5.64(18.8%)	18.16(47.8%)	16.12(42.4%)
S3.MF.12-43	2	12.34(311.9%)	0.63(18.8%)	1.40(27.6%)	1.32(27.0%)

表 3-3 より, 提案手法 VBS on は, VBS off と比較(MAE および MRE 共に小さい値の方が推定精度が高い)して, すべての評価ビデオストリームにおいて, 良い結果が得られ, 可変ブロックによる補正の有効性が確認できた.

次に, 提案手法の 1 フレームに要する処理時間をいくつかのビデオストリームに対して測定した結果を表 3-4 に整理した. すべての評価ビデオストリームにおいて処理に掛かった処理時間の幅(秒)と平均時間(秒)を示す.

表 3-4 提案手法の処理時間

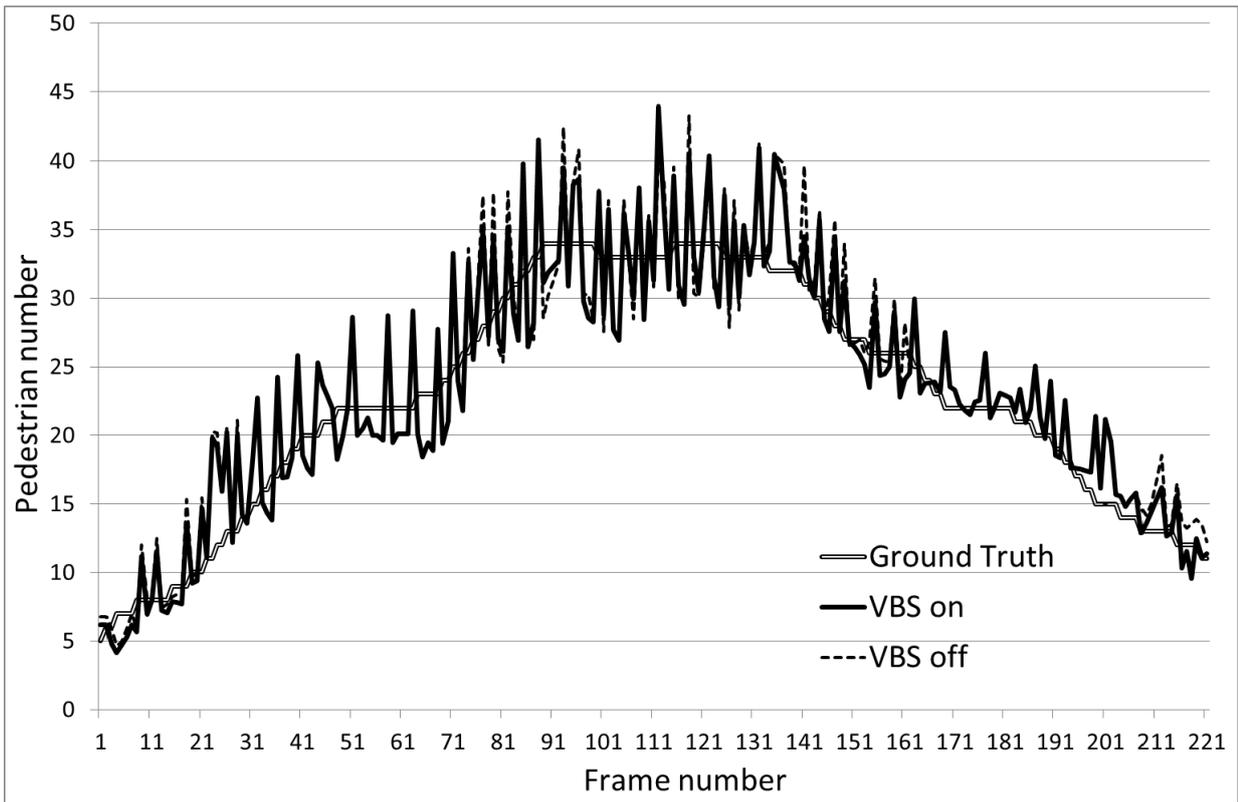
Video sequence	View	Processingtime/each frame(sec)	
		Width measurement	average
S1.L1.13-57	1	0.002-0.005	0.003
S1.L1.13-59	1	0.002-0.004	0.003
S1.L2.14-06	1	0.001-0.005	0.003
S1.L3.14-17	1	0.002-0.005	0.003
S1.L1.13-57	2	0.002-0.004	0.003
S1.L2.14-06	2	0.001-0.005	0.003
S1.L2.14-31	2	0.001-0.005	0.003
S3.MF.12-43	2	0.002-0.005	0.003

提案方法は, 表 3-4 で示すように, すべてのビデオストリームにおいて, 1 フレーム処理時間は, 0.001-0.005 秒の間にあり, 平均で 0.003 秒と計測された. 通常, 1 秒間に 30 フレーム再生するので, 1 秒間に 0.09 秒必要となり, この時間であれば, リアルタイム通信に十分適用できると思われる.

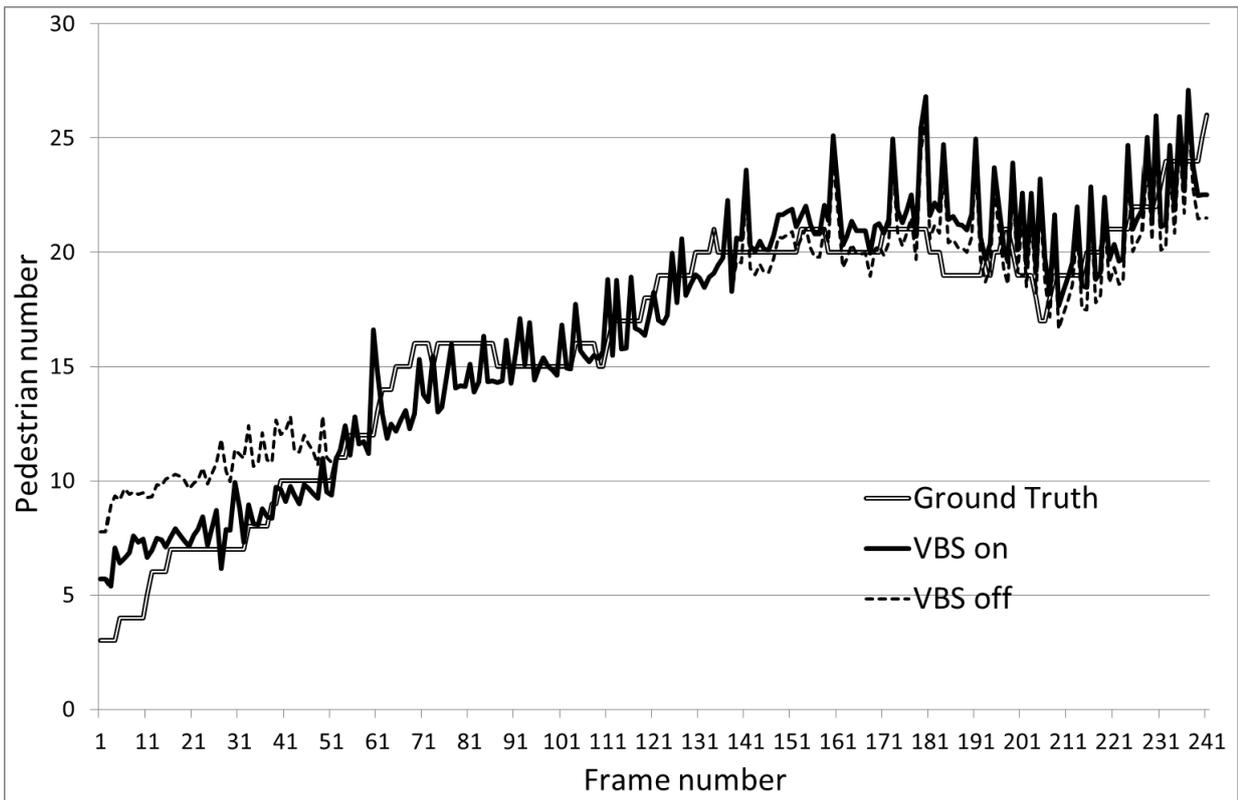
#### (4)-1 可変ブロックによる補正に関する考察

可変ブロックによる補正に関してビデオストリーム毎に考察した。

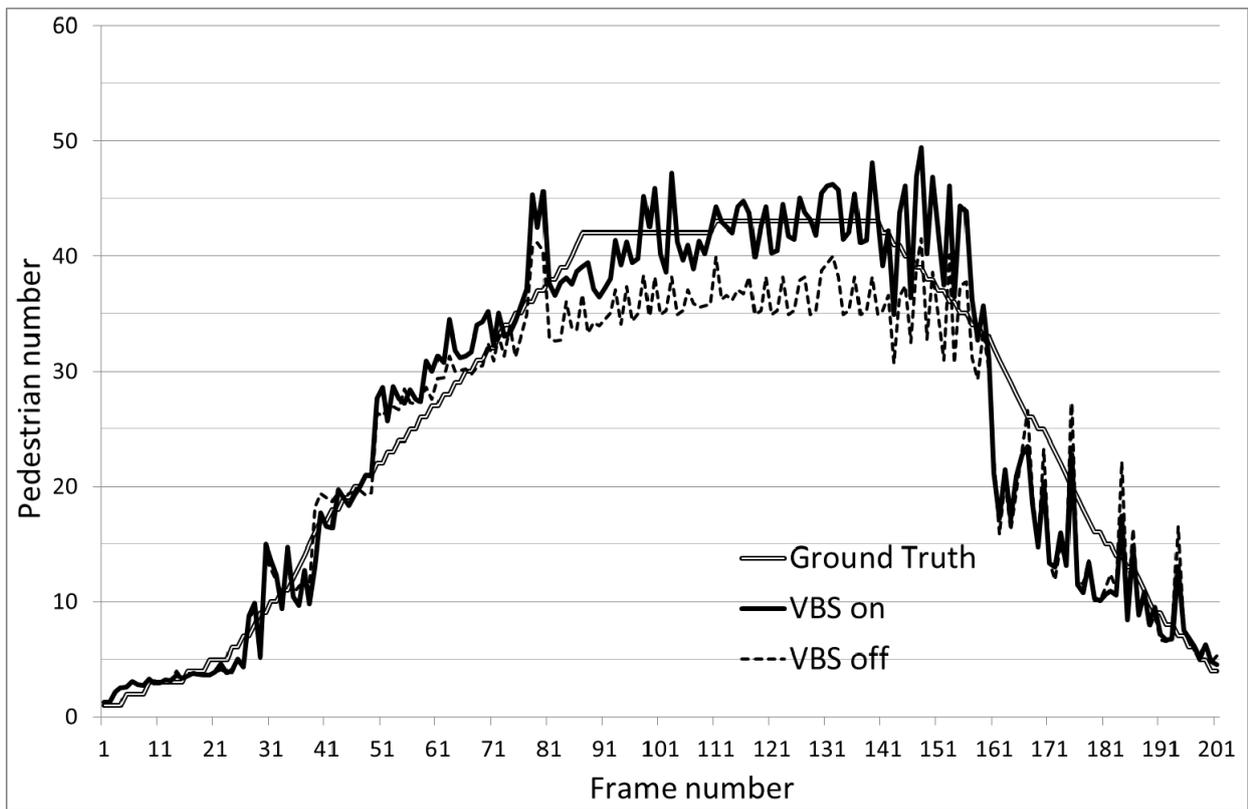
図 3-9(a)～(h)に各フレームにおける VBS on と VBS off を用いて推定した歩行者数と Ground Truth との比較を示す。横軸はフレーム数、縦軸は 1 フレームにおける歩行者数を示す。



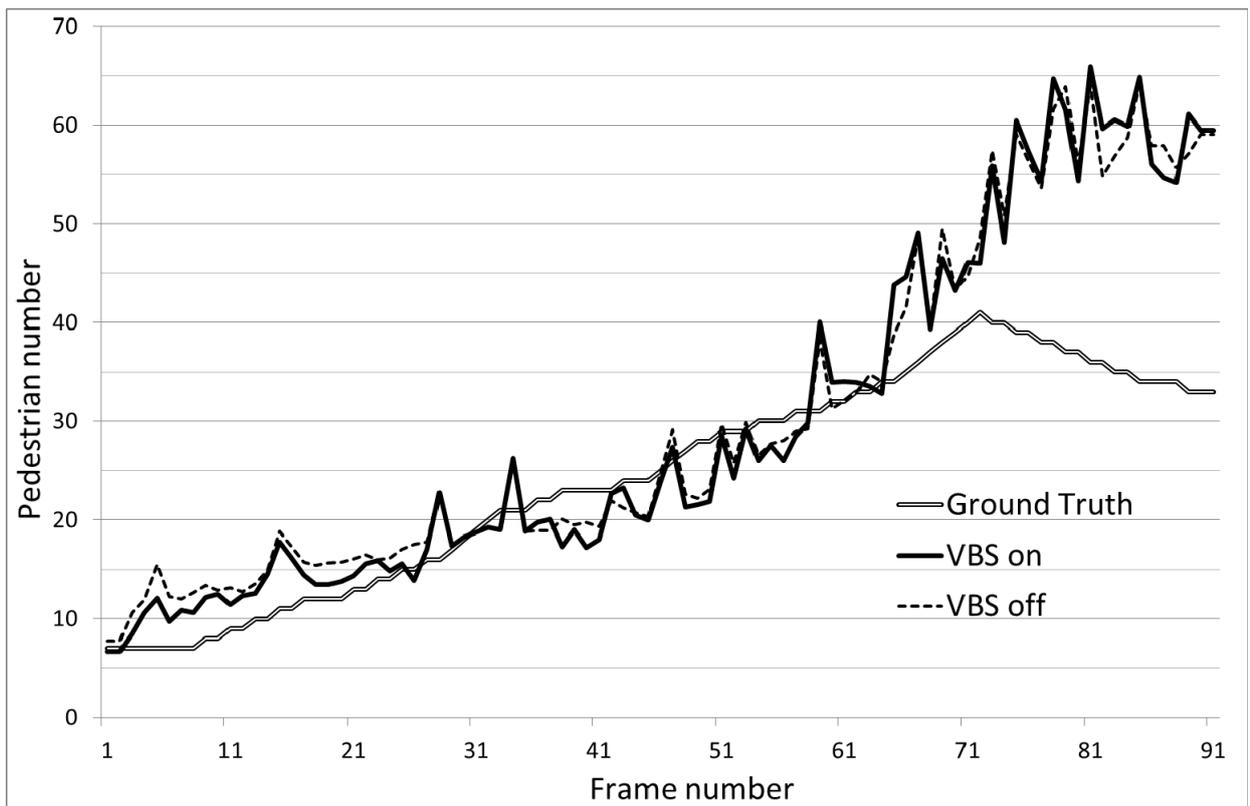
(a) S1.L1.13-57(View1)



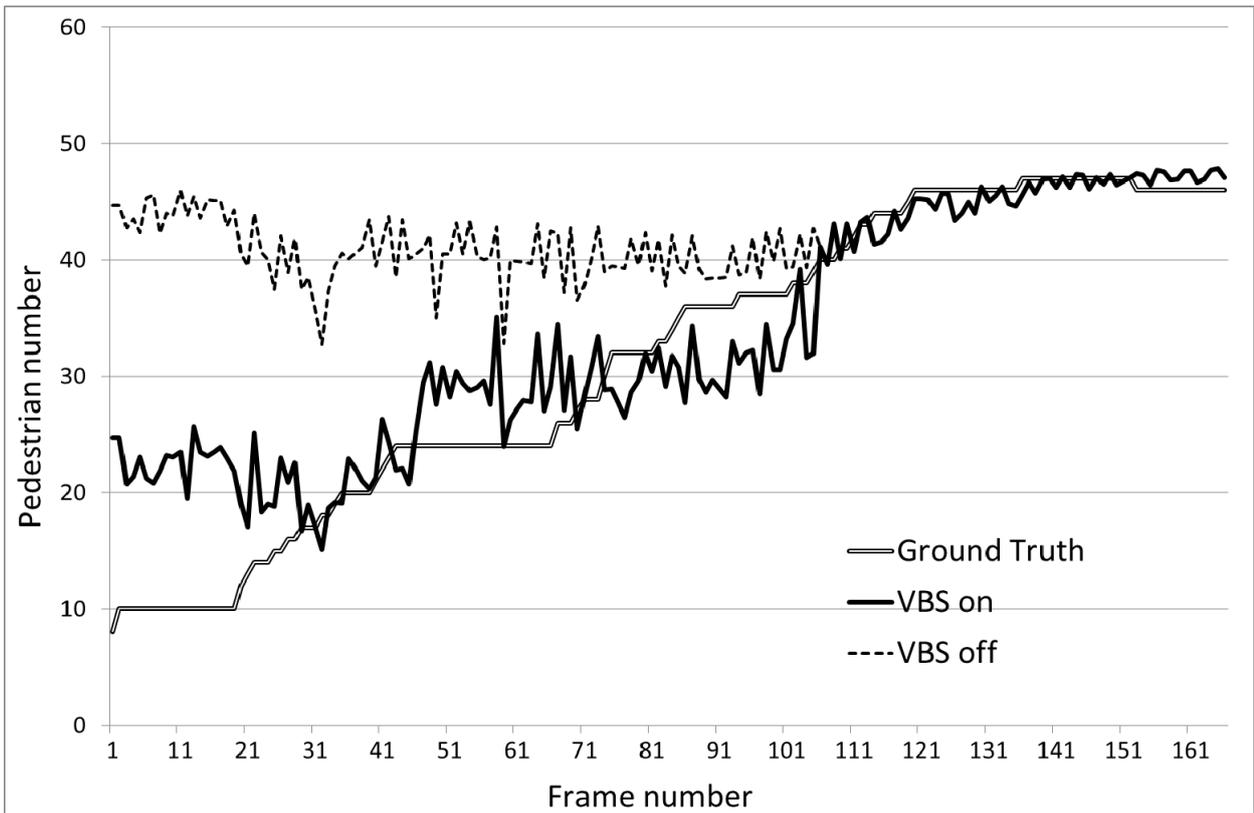
(b) S1.L1.13-59(View1)



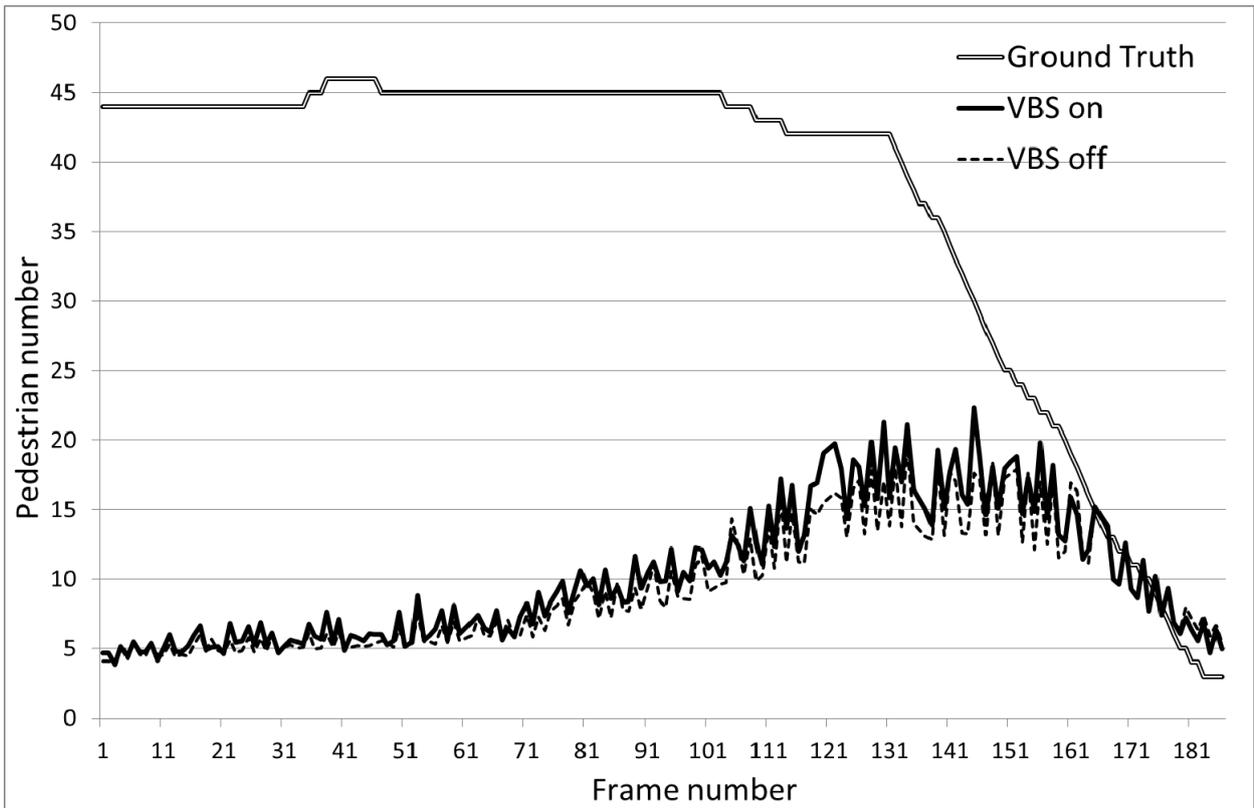
(c) S1.L2.14-06(View1)



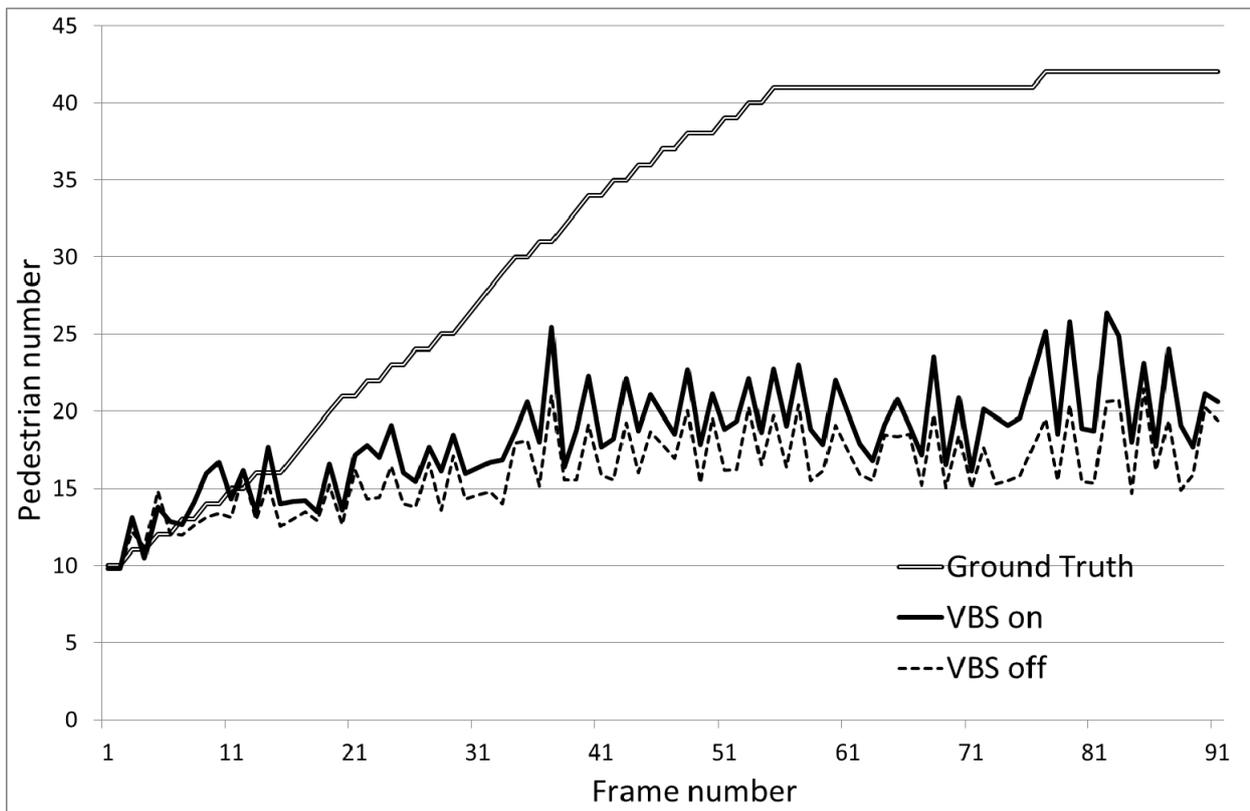
(d) S1.L3.14-17(View1)



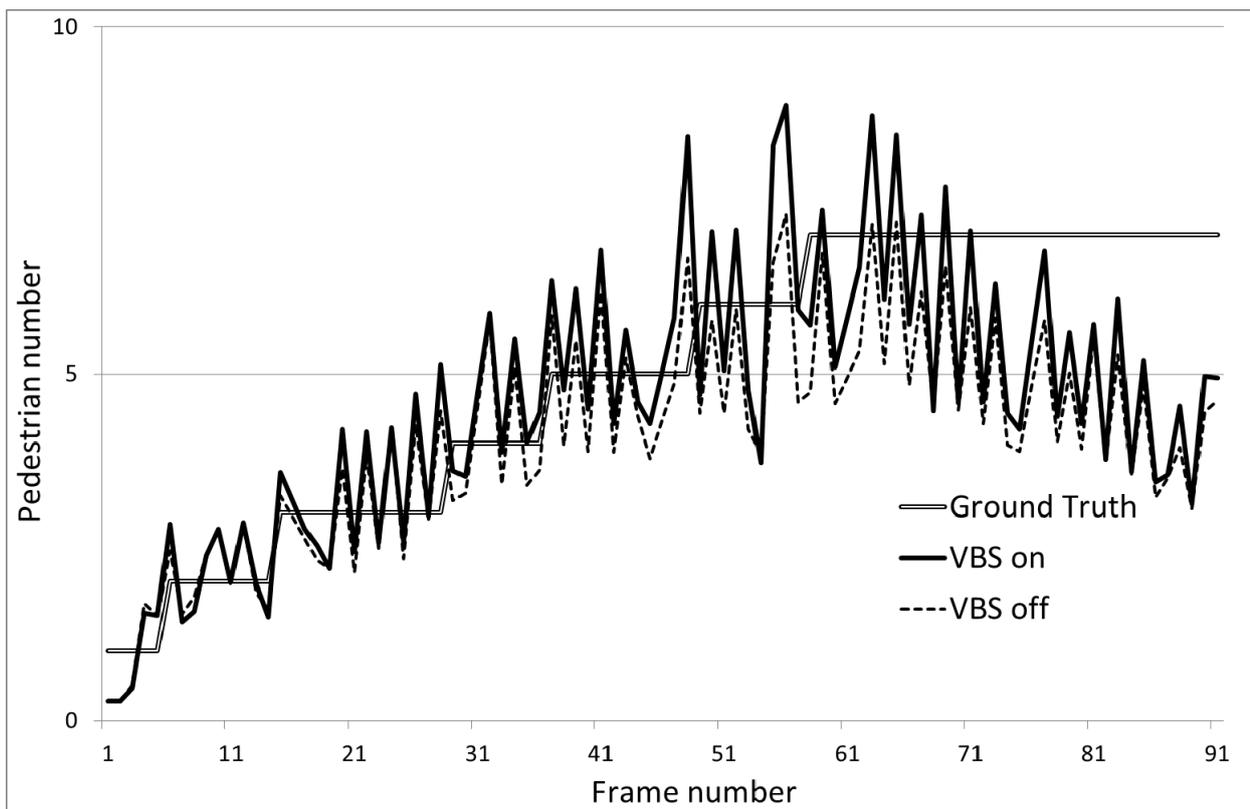
(e) S1.L1.13-57(View2)



(f) S1.L2.14-06(View2)



(g) S1.L2.14-31(View2)



(h) S3.MF.12-43(View2)

図 3-9 各ビデオストリームの VBS on, VBS off と真値の比較

S1.L1.13-57(view1)の特徴は、最初 5 人程度の歩行者が徐々に増えて最大 34 人程度の集団となり、その後は、徐々に減って行くビデオストリームである。その真値に対して、提案手法の VBS on と VBS off は、図 3-9(a)を見ると小刻みに変動している。これは、3.2.2-(4)で示したビデオ符号器の I フレームを使わないため、その代わりに前後のフレームの SAD 総和値の平均値を代入しているために発生している。通常、I フレームの直後の P フレームは符号量が多くなる傾向にあるため、前後を使った SAD 総和値の平均は大きくなる。従って、GOP 分の周期 (15 フレーム) で小刻みに変動する (図 3-7 参照)。以降の結果についても小刻みな変動は、上記の処理のために発生している。また、このビデオストリームにおいて、可変ブロックの補正による精度向上効果は、人数の多いところでは確認できた。これは、測定しているカメラの光軸と歩行者が平行な位置にあるために、人数が多くなるところのオクルージョンが、可変ブロックの補正により効果があったと思われる。人数が少ないところでは、SAD 総和値だけでも良いことも確認できた。

S1.L1.13-59(view1)の特徴は、最初 3 人程度の歩行者が徐々に増えて最大 26 人となるビデオストリームである。その真値に対して、提案手法の VBS on と VBS off は、図 3-9(b)を見ると、ほぼ追従していることがわかる。特に VBS on については、トレーニングストリームで利用した歩行者数(10 人~20 人)以外の歩行者数を推定する場合、式 3.1 を用いて場合分けすることで、可変ブロックによる補正の効果が見られる。

S1.L2.14-06(view1)の特徴は、最初は 0 人から歩行者が集団で増えて行き、最大 43 人になったところからその集団が徐々に減って行くビデオストリームである。その真値に対して、図 3-9(c)を見ると VBS on の可変ブロック補正は、オクルージョンが発生している 30 人以上のところでは補正が大きく貢献していることがわかる。この結果からも可変ブロック補正はオクルージョンに効果があることがわかった。

S1.L3.14-17(view1)の特徴は、最初 6 人程度から徐々に歩行者が増え始め、40 人くらいから徐々に減るのでほぼ、S1.L3.13-59(view1)と同じビデオストリームだが、日差しがあり、歩行者の人影が映りこむビデオストリームである。その真値に対して、図 3-9(d)を見ると後半の 70 フレーム以降に出る 30 人以上での歩行者に映る人影を歩行者と誤っているため、歩行者数の推定人数が大きく誤差が生じていることがわかった。

S1.L1.13-57(view2)の特徴は、最初 8 人程度から徐々に歩行者数は増えるが、測定しているカメラの光軸から歩行者が遠ざかるため、120 フレーム以降は、45 人程度の集団が歩行し続けるビデオストリームである。その真値に対して、図 3-9(e)を見ると 120 フレームまでの歩行者数の推定について、VBS off は誤差を多く含んだ結果となったが、可変ブロックの補正により VBS on の歩行者数の推定に大きく貢献していることがわかった。

S1.L2.14-06(view2)の特徴は、最初から 44 人程度の集団で歩行し、測定しているカメラの光軸から歩行者が遠ざかりながら四方に散らばるビデオストリームである。このビデオストリームは、カメラの光軸に対して遠ざかるため歩行者同士が完全に重なることが多く、オクルージョンも確認できない。図 3-9(f)を見ると 160 フレーム以降に四方に散らばるところが辛うじて推定できたが、160 フレームまでは、推定誤差が大きいことがわかった。

S1.L2.14-31(view2)の特徴は、S1.L2.14-06(view2)のビデオストリームと同じように測定するカメラの光軸に対して歩行者が遠ざかり、かつ日差しがあるビデオストリームである。

図 3-9(g)を見ると、最初から 15 フレームまでは、歩行者を推定できているが、そのフレーム以降は、歩行者に映る人影を歩行者と誤っているため、歩行者数の推定人数が大きく誤差が生じていることがわかった。

S3.MF.12-43(view2)の特徴は、これまでのビデオストリームと違って歩行者数の最大人数は 7 人でオクルージョンも少ないが、60 フレーム以降は、カメラの光軸に対して遠ざかるため、歩行者同士が徐々に重なるビデオストリームである。図 3-9(h)を見ると、60 フレームまでは全体的に多めに歩行者数を推定していることがわかる。これは、歩行者の足元に若干人影が出ており、この影を歩行者と認識して、少し多めに歩行者を推定した結果となった。60 フレーム以降は、歩行者同士が徐々に重なって行くため歩行者数の推定に大きな誤差が発生した。

以上 8 つのビデオストリームの評価結果の考察から、VBS on の可変ブロックの補正に関して以下のような結論を得た。

- ・測定しているカメラの光軸と歩行者が平行な位置にあるビデオストリームのオクルージョンに対して効果がある。
- ・オクルージョンが発生しないような少人数での歩行者数の推定は、可変ブロックの補正がなくても、SAD 総和値だけの予測でも十分である。
- ・式(3-1)を用いた場合分けによる歩行者数の推定に効果がある。
- ・歩行者の人影が映りこむビデオストリームの場合は、動画像に適したフィルタリング処理や色空間補正などの適用が必要である。

#### (4)-2 他手法との比較に関する考察

VBS on と Albiol や Conte の手法との精度比較である。表 3-3 の結果から以下のような性質が確認できた。

S1.L1.13-57(view1)について、VBS on は、MAE に関しては、Albiol より良い結果となったが、MRE の比較においては、少し劣った結果となった。これは、真値計測の誤差が計算結果に影響したと思われるので、精度の比較としては、VBS on と Albiol は、ほぼ同等の精度と思われる。一方、VBS on と Conte とでは、MAE と MRE の両方において Conte の方が良い結果となった。そこで、各フレームの画素が持つ SAD 総和値と真値との相関係数を算出すると  $R^2=0.677$  であった。一般的に  $R^2>0.7$  であれば相関が高いとされているので、回帰分析より導いた回帰式に誤差を含む結果となったと思われる。

S1.L1.13-59(view1)について、VBS on が他の手法と比較して精度の向上が確認できた。これは、測定しているカメラの光軸と歩行者が平行な位置にあるため、画面内の歩行者の大きさが一定であることが高い精度に繋がったと思われる。

S1.L2.14-06(view1)について、VBS on が他の手法と比較して精度の向上が確認できた。これは、歩行者が集団となって動いているビデオストリームであるため、Albiol や Conte のような特徴点を抽出して、歩行者数を推定する手法では、特徴点が重なり合い、誤差を多く発生し、精度が低くなったと思われる。

S1.L3.14-17(view1)について、VBS on の精度が最も低かった。表 3-1 にもあるように、日差しがあるビデオストリームであるため、歩行者の影が映りこむビデオストリームである。VBS on は、歩行者の影も歩行者としてカウントしたため誤差が生じたと思われる。人影の除去には、一般的にフィルタリング処理や色空間補正などがあるが、いずれにおいても、静止画で実証されているだけで動画像に適したフィルタは今後の実験で明らかにする必要がある。

S1.L1.13-57(view2)について、VBS on が他の手法と比較して精度の向上が確認できた。測定しているカメラの光軸から歩行者が遠ざかるため、どの手法も精度は低くなる傾向にあるが、可変ブロックの補正により遠ざかる歩行者を捕捉できたことが、精度向上に繋がったと思われる。

S1.L2.14-06(view2)について、すべての手法において精度が低い結果となった。これは、歩行者が集団となって動いていて、かつ、測定しているカメラの光軸から歩行者が遠ざかるため、どの手法も精度が低くなる傾向であった。いずれの手法も MAE と MRE とも誤差が多い結果となった。つまり、歩行者を推定するには、難しいビデオストリームと思われる。

S1.L2.14-31(view2)について、S1.L3.14-17(view1)と同じように日差しがあるビデオストリームである。しかも、歩行者が集団となって動いていて、かつ、測定しているカメラの光軸から歩行者が遠ざかるため、いずれの手法も MAE と MRE とも誤差が多い結果となった。このビデオストリームも、歩行者を推定するには、難しいビデオストリームと思われる。

S3.MF.12-43(view2)について、VBS on と比較して、Conte の手法が若干良い結果となった。このビデオストリームは、歩行者数も疎らで、オクルージョンも少なく、歩行者数の推定精度が高くなる傾向にあるが、若干足元に人影が写っているために、Conte と比べて精度が低くなったと思われる。

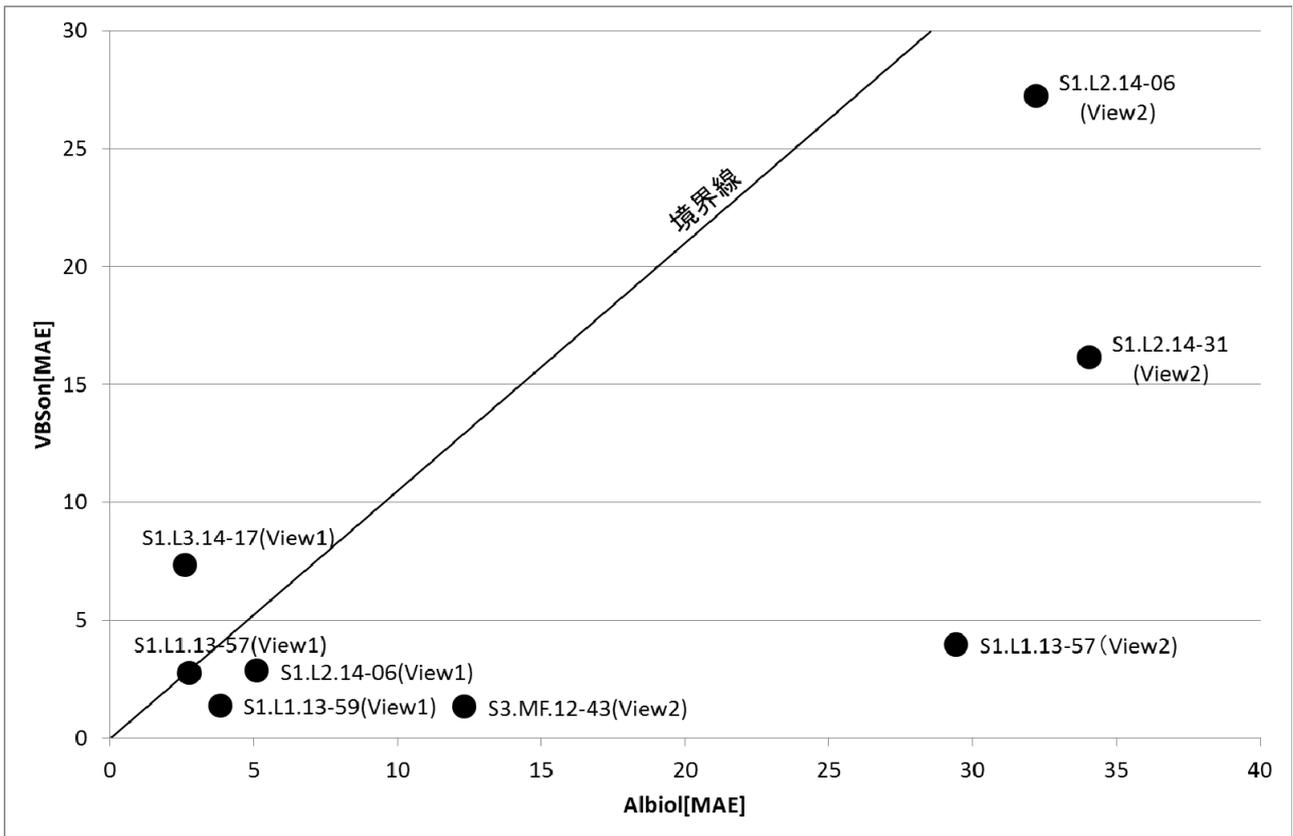
以上 8 つのビデオストリームの評価結果から以下のような結論が得られた。VBS on の歩行者数の推定精度は、Albiol と比べると概ね歩行者数の推定精度が高くなっており、Conte と比べても同等である。ただし、歩行者の影が映りこむようなビデオ

ストリームの場合は、上記でも記載したように、動画像に適したフィルタリング処理や色空間補正などの適用で解決する必要がある。

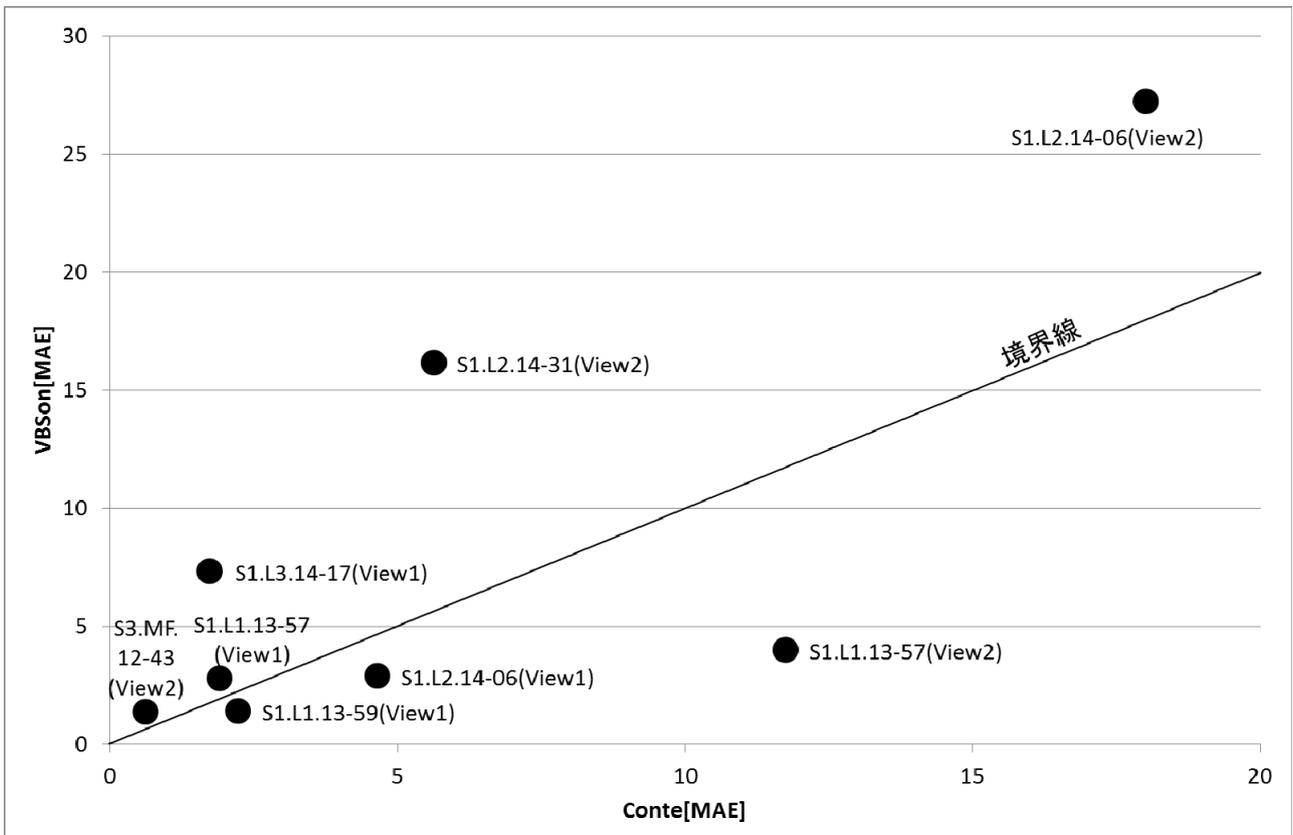
さらに VBS on と Albiol や Conte の手法との精度比較を整理するために、各ビデオストリームに対する Albiol と VBS on の比較を MAE について行った結果を図 3-10(a)、Conte と VBS on の比較を MAE について行った結果を図 3-10(b)に整理した。図中の境界線は、MAE の推定精度の境界を示す。つまり、各ビデオストリームの MAE において、どちらの手法が精度高いか示している。なお、数字の少ない MAE の方が、歩行者の推定精度は高い結果となっている。

図 3-10(a)では、唯一、S1.L3.14-17(view1)だけが、Albiol の MAE に比べて VBS on が劣っている。これは、実験結果で述べたように、ビデオストリームに影が映りこんでいるため VBS on の歩行者の推定精度が低下したことが理由である。その他のビデオストリームでは、歩行者の推定精度が高いことがわかる。

次に、図 3-10(b)では、若干、Conte の MAE に比べて VBS on が劣っているように見える。しかし、境界線から遠ざかる程、推定精度が低いことを表しているが、Conte の MAE に比べて VBS on は境界線に近いので、ほぼ同等の精度であると思われる。



(a) Albiol と VBson の MAE 比較



(b) Conte と VBson の MAE 比較

図 3-10 Albiol, Conte と VBson との MAE 比較

次に、VBS on と Albiol や Conte の手法との処理時間と歩行者数の推定精度の比較をグラフにして評価した。評価したビデオストリームのスナップショットを図 3-11 に示す。



S1.L1.13-59(View1)



S1.L1.13-57(View2)

図 3-11 評価したビデオストリームのスナップショット

図 3-12(a)(b)は、歩行者数の推定に必要な各フレームにおける処理時間を示す。横軸はフレーム数、縦軸は 1 フレームにおける処理時間(秒)を示す。図 3-12(a)(b)ともに 0.003 秒を中心に処理時間が発生していることがわかる。なお、Albiol と Conte は特徴点を抽出するために別処理を行っているため、リアルタイムに処理することは難しいと思われる。

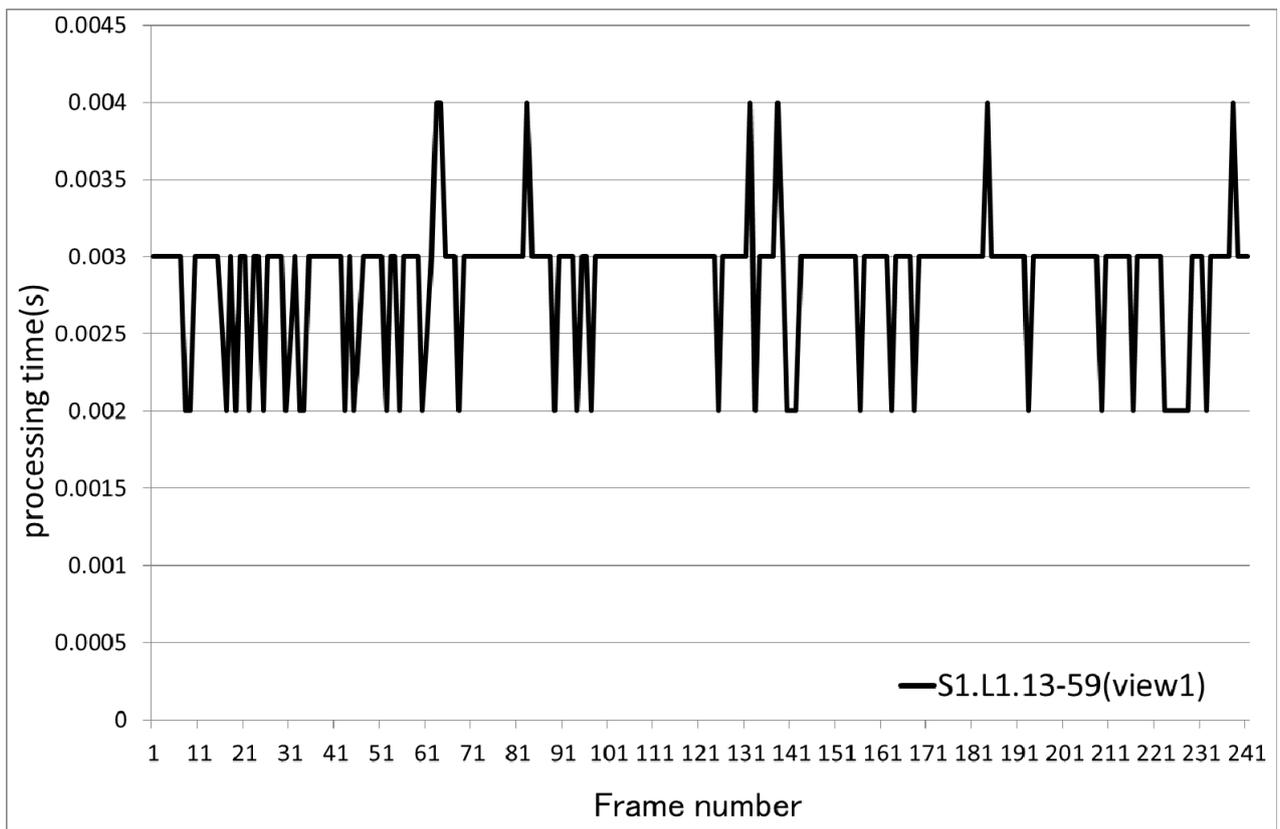
次に推定精度について評価した。

図 3-13(a)(b)に各フレームにおける Albiol と Conte を用いて推定した歩行者数と Ground Truth との比較を示す。横軸はフレーム数、縦軸は 1 フレームにおける歩行者数を示す。

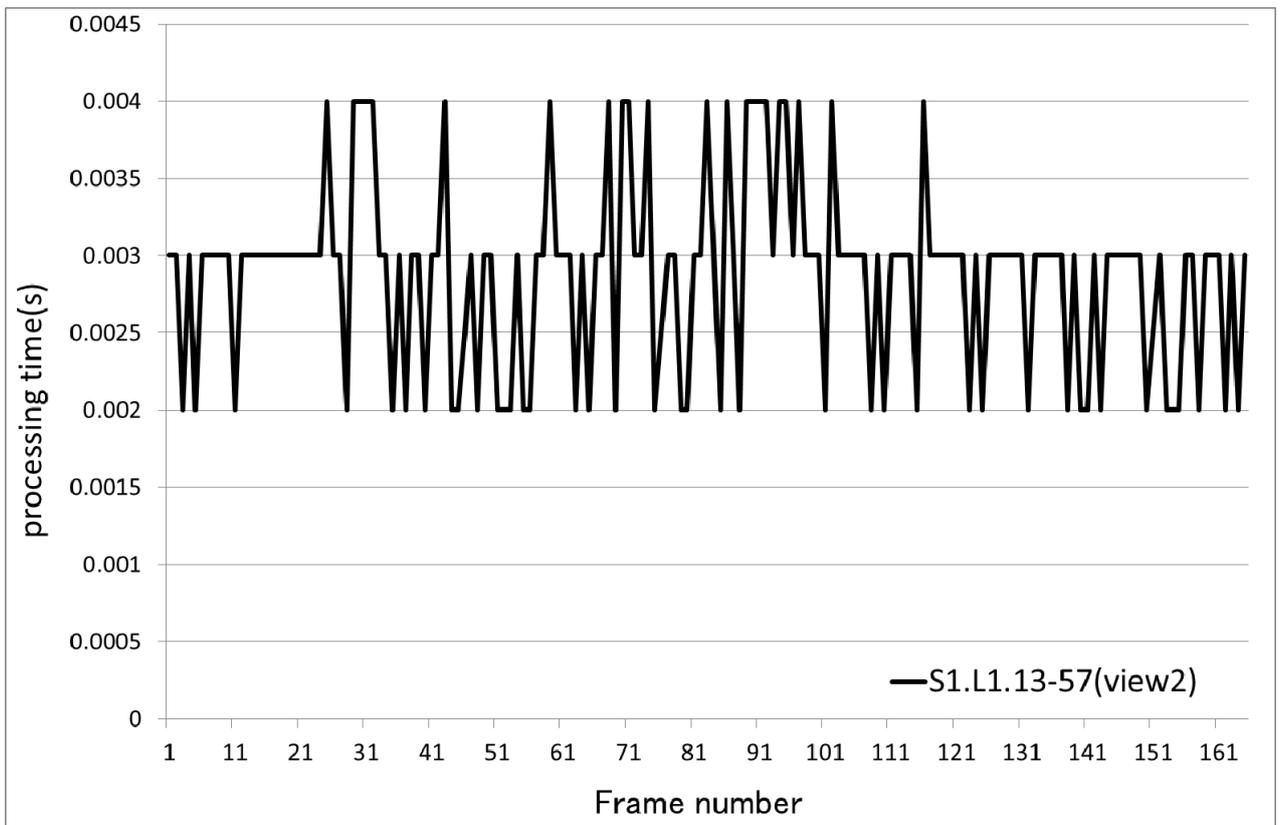
Albiol は、コーナー検出法[40]に基づいて集計された MV を用いて移動物体を識別し、それらから統計的に推定する方法である。S1.L1.13-59(View1)の方は、測定しているカメラの光軸と歩行者が平行な位置にあるビデオストリームのため、比較的、コーナー検出法による特徴点が得られ歩行者数の推定はできている。しかし、S1.L1.13-57(View2)の方は、測定するカメラの光軸から遠ざかるビデオストリームのため、全体を通じて、コーナー検出法の特徴点が得られず、歩行者の推定誤差が大きくなっていることがわかる。また、Conte は、画素境界近傍の輝度勾配情報から抽出した特徴点[42]を用いて推定する手法である。S1.L1.13-59(View1)の方は、コーナー検出方法と同等であるが、S1.L1.13-57(View2)の方は、Albiol と比べて精度が良いことがわかる。これは、輝度勾配情報により遠ざかるビデオストリームでも画素境界が識別できるため、歩行者数の推定も良好に取得できることがわかった。

最後に提案手法については、(4)-1 の考察で述べたように、GOP 分の周期で小刻みに変動しているが、概ね真値に近似していることがわかる。特に S1.L1.13-57(View2)の 110 フレーム以降の遠ざかるシーンにおいて、可変ブロックの補正により効果があることがわかった。

図 3-14(a)(b)に、実際の歩行者数に対して各手法の推定した歩行者数を比較した。横軸は実際の歩行者数、縦軸は推定した歩行者数を示す。図 3-14(a)(b)ともに提案手法の方が、Ground Truth に近い歩行者数の推定を行っていることが確認できた。

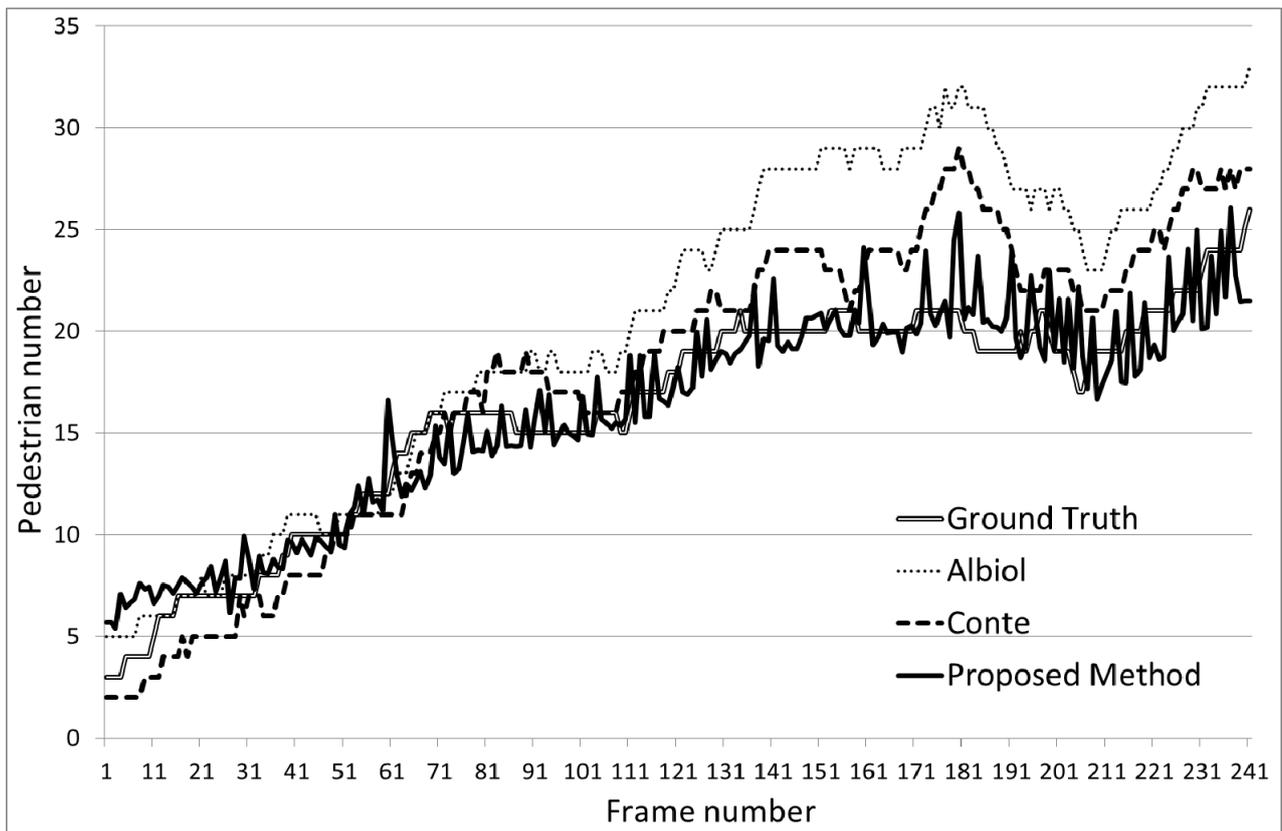


(a) S1.L1.13-59(view1)

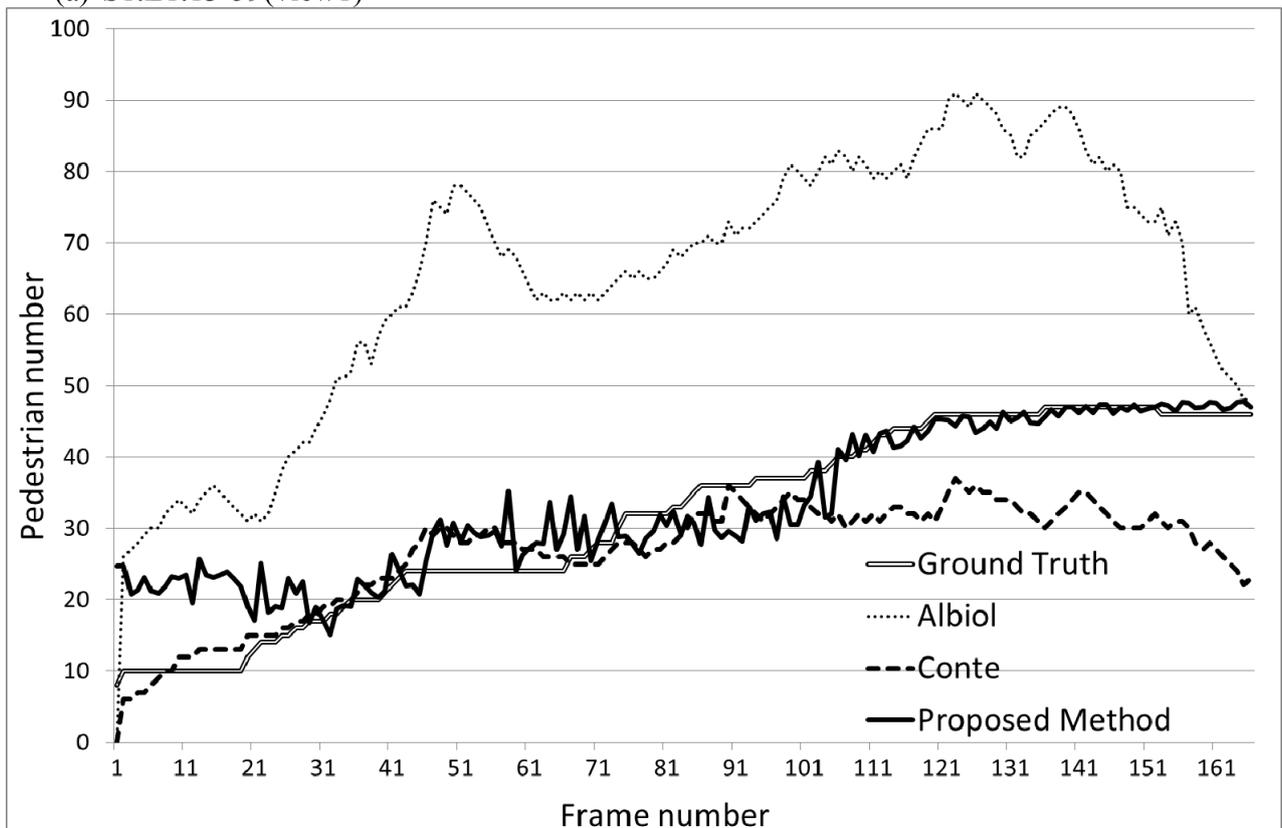


(b) S1.L1.13-57(view2)

図 3-12 各フレームの処理時間

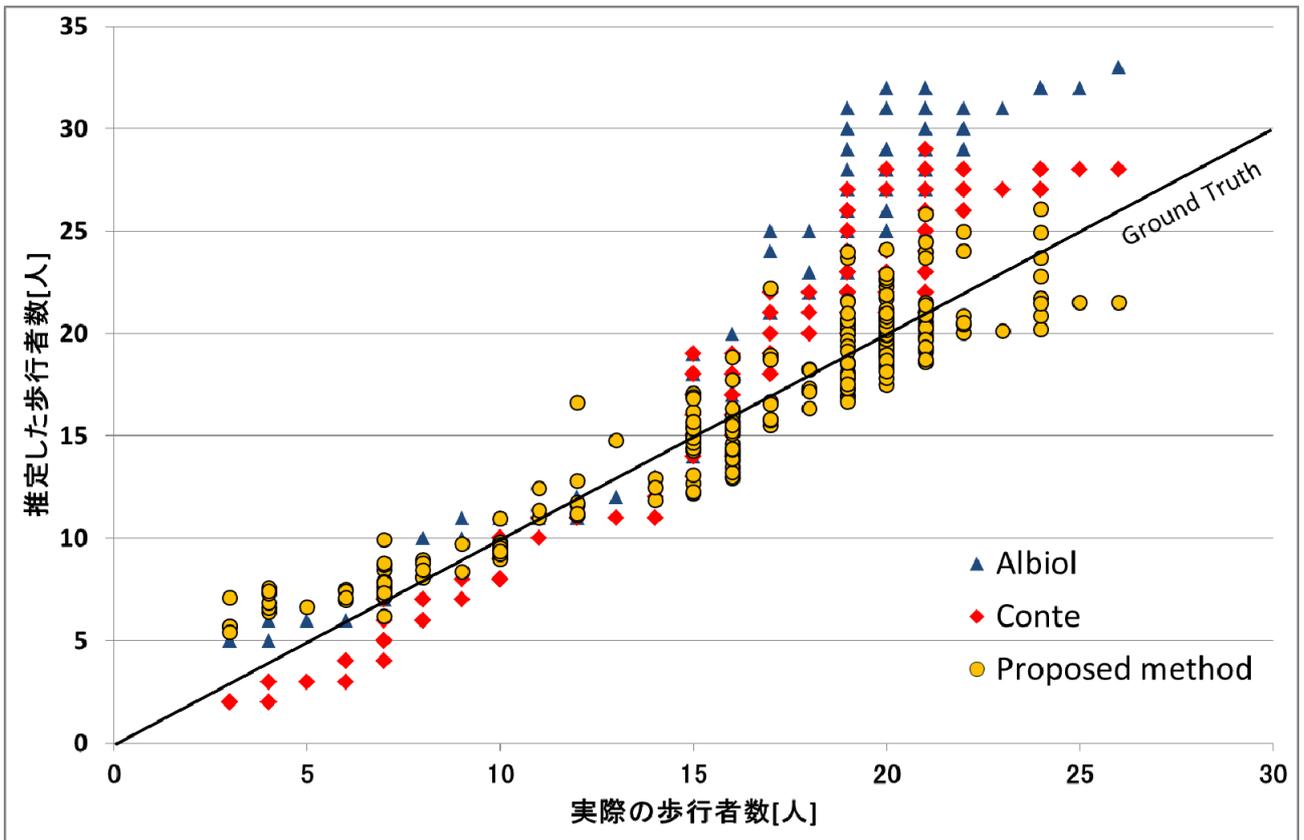


(a) S1.L1.13-59(view1)

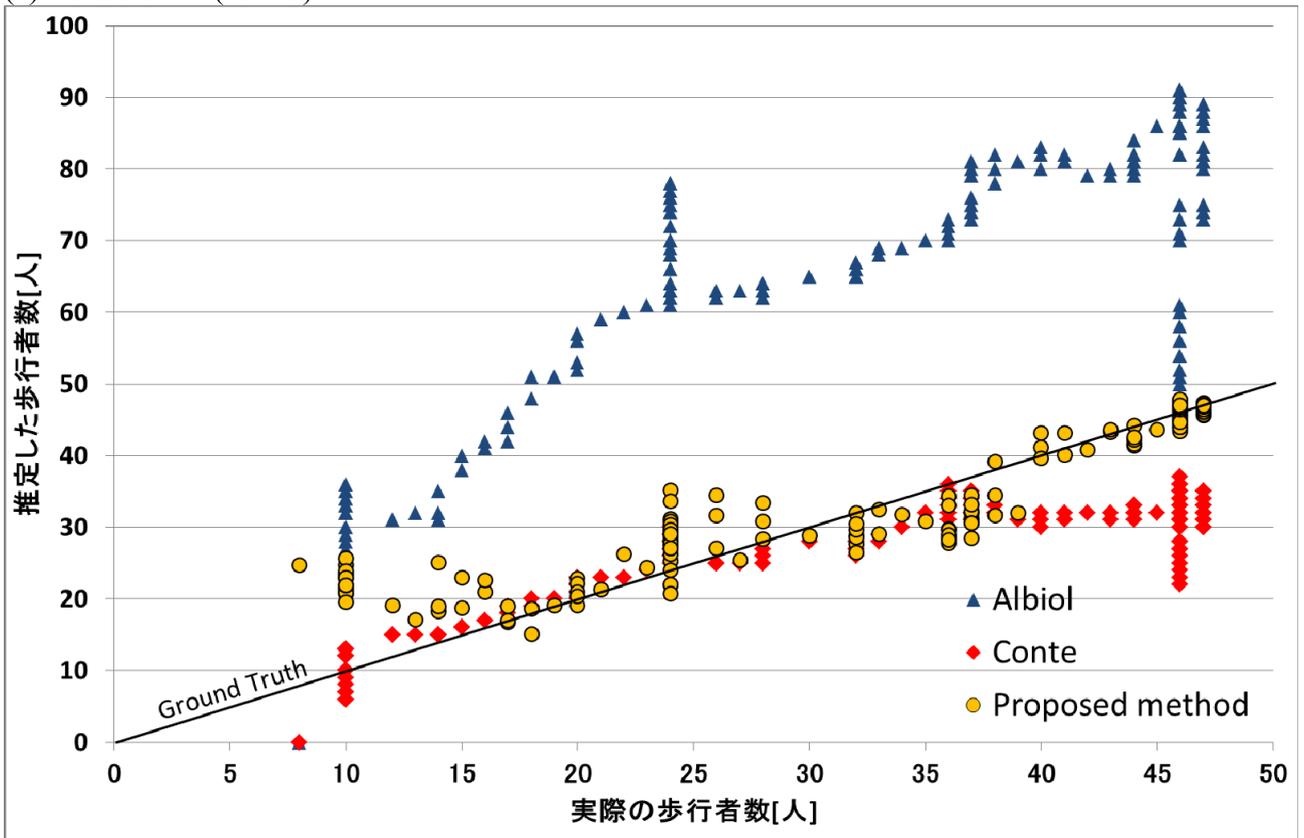


(b) S1.L1.13-57(view2)

図 3-13 ビデオストリーム(a)(b)の各手法における歩行者数の推定と真値の比較



(a) S1.L1.13-59(view1)



(b) S1.L1.13-57(view2)

図 3-14 実際の歩行者数に対する推定した歩行者数の比較

### 3.3 まとめ

評価の結果，提案方式を用いることで，推定精度は同等に維持したままリアルタイムに歩行者数を推定できることを明らかにした．本章における提案方式は，標準化研究で多用されている H.264/AVC を利用してその性能を評価しているが，評価結果は H.264/AVC の特性に依存する．このような H.264/AVC の特性に対する依存性や人影がある場合の改善手法は今後の課題である．また，本章は，PETS2009 のデータベースを用いて実験を行ったため，画像特徴に対する本章の適用範囲が，測定するカメラの光軸と歩行者の位置づけが大きく変化しないビデオストリームに限っている．今後の課題として，対象となるビデオストリームを幅広く，例えば，画面内で歩行者の大きさが大きく変化する場合は含まれるようなビデオストリームを含めた評価を行いたい．

## 第4章 結論

本章では、今後のビデオ符号化を用いた研究成果について述べるとともに本研究成果の寄与について考察する。

サーバからデータを受信しながら映像を再生するビデオストリームには、ビデオ符号化が用いられており、ビデオストリームの低ビットレート化および高品質化の重要な技術要素となっている。一方、ネットワーク環境は、セルラーや無線 LAN などの異種無線アクセス網を組み入れた複合的無線サービスが普及してきている。そのような複合的無線サービスでは、複数の通信経路を持つことが可能となるので、ネットワークの負荷分散、通信障害耐性向上などに効果が期待されている。このような複数チャネルを有するネットワーク環境によって、ビデオストリーム品質が向上してきている。一般的には、同じ品質の映像を異なるネットワークで 2 重化して配信することで、映像配信の信頼性や品質を確保する完全冗長 2 重化方式が使われていることが多い。しかし、ネットワークが異なると帯域などのネットワーク特性が異なるため、それらのネットワークを用いて同じ品質の映像を配信するには、遅延が発生して、リアルタイム性を損なう問題がある。

そこで、複合的無線サービスに適した多重記述のビデオ符号化を用いた映像品質の向上を目的とした。ビデオストリームに対して多重記述ビデオ符号化を行うことで、1 つの映像ソースを 2 つのエンコーダによってそれぞれ送信し、相互補完することで途切れの無いデータ通信が行われ、さらに 2 つのチャネルが受信できれば、高品質なデータ通信が行われる。次に、高品質なビデオストリームの受信ができるようになると、受信したビデオデータを解析することによって様々なアプリケーションが実現すると考えられる。本研究では、その中でも、ビデオ中の歩行者数のリアルタイム推定に着目する。これまで、センサや専用ハードウェアを用いてリアルタイムに歩行者数を推定していたが、設置場所に限りがあった。また、ビデオストリームからデコード後の映像を処理して、歩行者数を推定する技術はあるが、いずれも一旦別処理で歩行者数の推定処理を行うので、リアルタイム処理には難しい技術的な課題があった。

そこで、ビデオストリームのビデオ符号化から歩行者数をリアルタイムに推定する。ビデオストリームを復号する際に抽出される予測残差信号からフレーム毎に回帰分析を用いて歩行者数を推定し、フレーム内の可変ブロックを参照して歩行者数を補正するところに特徴がある。従来手法と同等の性能が得られ、かつリアルタイムに推定できることを確認した。

2 章では、複合的無線サービスに適した多重記述のビデオ符号化から参照画像データを抽出してパケットロスによる画像エラーの修復や量子化パラメータを抽出して、量子化前の DCT 係数を予測することで、映像品質の向上を実現させ、リアルタイムに高品質なビデオストリームが受信できる環境を整えた。

2.2 節では、時間スケラビリティ機能における参照画像の変換処理を行う、Temporal MD と呼ぶ手法を提案した。性能評価の結果、特にネットワークでパケットロスが発生する環境において、3 つの評価ビデオストリームを用いた性能評価の結果、従来の手法よりも映像品質が 0.2dB-0.6dB 良好であることを確認した。

2.3 節では、異なる量子化後の DCT 係数から量子化前の DCT 係数を絞り込み量子化パラメータを操作して、映像品質を向上させる手法である Quantizer MD 手法を提案した。2 つの評価ビデオストリームを用いた性能評価の結果、従来の手法よりも映像品質が 0.3dB-1.0dB 向上したことを確認した。

3章では、H.264/AVC で符号化されたビデオストリームのビデオ符号化を統計処理によって歩行者数をリアルタイムに推定する手法を提案し評価した。提案方法は、ビデオ符号化の予測残差信号から回帰分析を用いて歩行者数を推定後、MB や SubMB を用いて補正し、歩行者数を推定した。8 つの評価用ビデオストリームを用いた性能評価の結果、提案手法は、リアルタイム処理が可能であり、かつ、従来的手法と同等の推定性能が得られたることを確認した。

本章による産業上で想定される利用イメージを図 4-1 に示す。これら 2 つの研究の成果は、例えば、商業ビルなどに設置されたカメラで撮影されている映像をネットワークを介して収集し、それを用いて監視やマーケティングに用いることが考えられる。地理的に離れた様々な箇所に設置されたカメラから映像を受信するためには、無線通信を含んだ様々なネットワークが用いられると考えられる。本論文の成果を用いることで、ネットワーク環境によらず高品質な映像を得ることができるようになる。さらに、得られた映像中の歩行者数に関する情報をリアルタイムに得ることができるため、イベント等における人流把握や警備計画の立案、また災害時の避難経路の確保などの道路環境の改善に活用することができると考えられる。

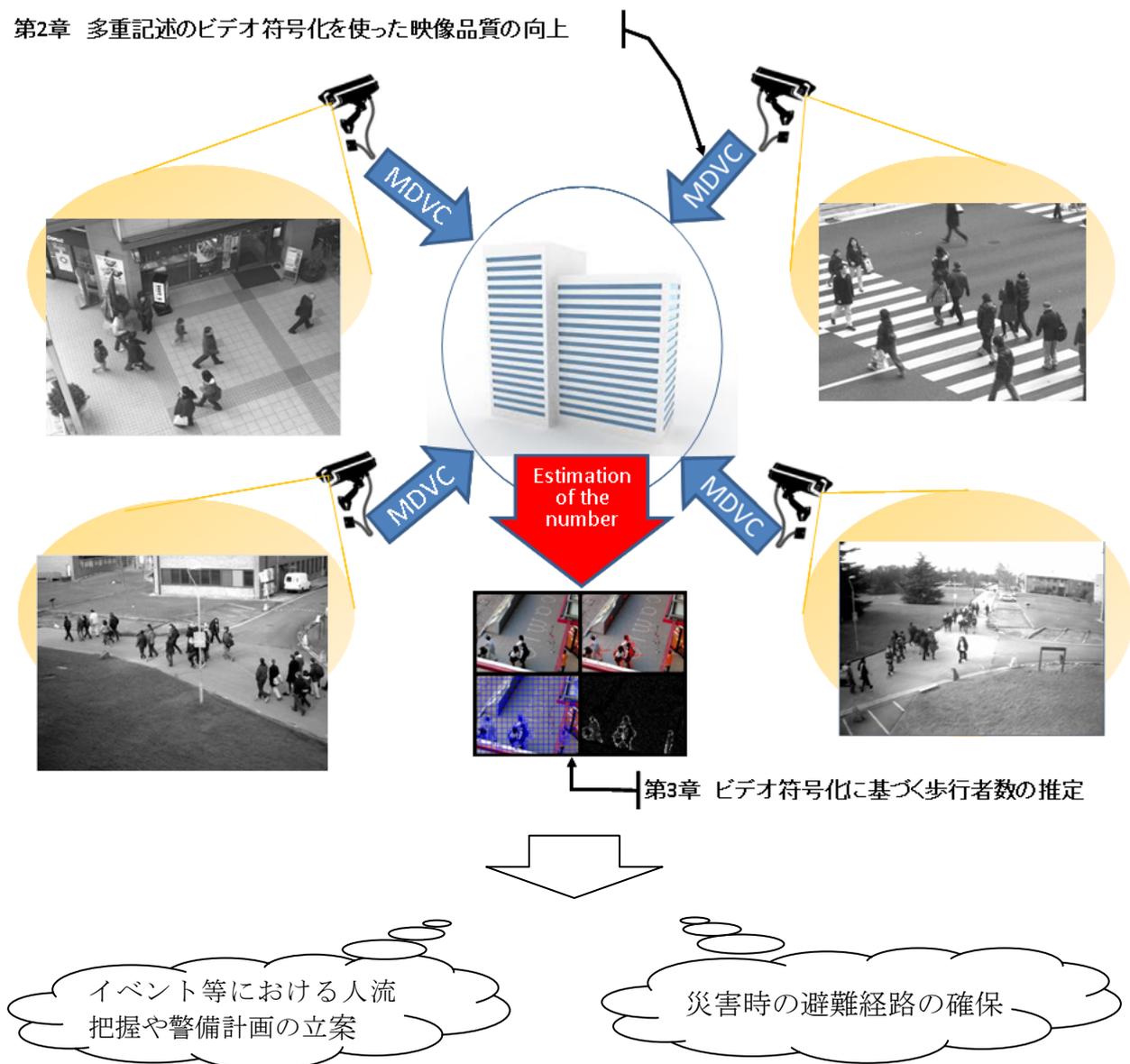


図 4-1 研究成果の利用イメージ

## 参考文献

- [1] 川田亮一,松本修一, “フラットマルチスケラブル高能率高信頼度映像伝送方式”, 電子情報通信学会論文誌, vol.J84-B,No.3, pp.621-631, 2001
- [2] “Coding of Audio-Visual Objects—Part 2: Visual”, ISO/IEC 14496-2(MPEG-4 Visual version 1), ISO/IEC JTC 1, Apr. 1999 (and subsequent editions)
- [3] “Advanced Video Coding for Generic Audio-Visual Services, ITU-T Rec.H.264 and ISO/IEC 14496-10 (AVC)”, ITU-T and ISO/IEC JTC 1, 2003
- [4] “Coding of Moving Pictures and Associated Audio for Digital Storage Media at up to About 1.5 Mbit/s Part 2: Video”, ISO/IEC 11172-2(MPEG-1), ISO/IEC JTC 1, 1993
- [5] “Generic Coding of Moving Pictures and Associated Audio Information Part 2 Video”, ITU-T Rec. H.262 and ISO/IEC 13818-2 (MPEG-2 Video), ITU-T and ISO/IEC JTC 1, Nov. 1994
- [6] “Video Coding for Low Bit Rate Communication”, ITU-T Rec. H.263, Nov.1995 (and subsequent editions)
- [7] L. Ozarow, “On a source-coding problem with two channels and three receivers”, Bell Syst. Tech. J., vol. 59, no. 10, pp. 1909-1921, Dec. 1980
- [8] V.A.Vaishampayan, “Design of multiple description scalar quantizers”, IEEE Trans. Inform. Theory, vol. 39, pp. 821-834, May 1993
- [9] V.A. Vaishampayan and J. Domaszewicz, “Design of entropy-constrained multiple-description scalar quantizers”, IEEE Trans. Inform. Theory, vol. 40, pp. 245-250, Jan. 1994
- [10] S.D. Servetto, V.A. Vaishampayan, and N.J.A. Sloane, “Multiple description lattice vector quantization”, In Proc. IEEE Data Compression Conf., Snowbird, UT, pp.13-22, 1999
- [11] V. Vaishampayan, N. J. A. Sloane, and S. D. Servetto, “Multiple description vector quantization with lattice codebooks: Design and analysis”, IEEE Trans. Inf. Theory, vol.47, pp.1718-1734, 2001
- [12] V. Goyal , J. Kelnr and J. Kovacevic, “Multiple description vector quantization with a coarse lattice”, IEEE Trans. Inf. Theory, vol.48, no.2, pp.781-788, 2002
- [13] J.A. Kelner, V.K. Goyal, and J. Kovacevic, “Multiple description lattice vector quantization: Variations and extensions”, In Proc. IEEE Data Compression Conf., Snowbird, UT, pp. 480-489, 2000
- [14] M. Fleming and M. Effros, “Generalized multiple description vector quantization”, In Proc. IEEE Data Compression Conf., Snowbird, UT, pp. 3-12, 1999
- [15]内村圭一,松島宏典, “オクルージョンを考慮した交通流計測”, 電学論 C, Vol.122-C, No.12, pp.2120–2127, 2002
- [16]T. Zhao and R. Nevatia, “Car detection in low resolution aerial images”, IVC, Vol.21, No8, pp.693-703, 2003

- [17]三谷重知,呉 海元,塩山忠義, “ガボール変換による自動車検出”, 信学論, Vol.J83-D-II, No.12, pp.2641-2651, 2000
- [18]中上友宏,松尾啓志, “人物モデルと体軸方向フィルタリングを用いた実時間人物検出・計数システム”, 電学論 C, Vol.122-C, No.12, pp.2011-2019, 2002
- [19]B.A. Boghossian and S.A. Velastin, “Motion-based machine vision techniques for the management of large crowds”, Proceedings of ICECS '99. The 6th IEEE International Conference on Electronics, Circuits and Systems, vol.2, pp.961-964, 1999
- [20]A.C. Davies, Jia Hong Yin, and S.A. Velastin, “Crowd monitoring using image processing”, Electronics & Communication Engineering Journal, vol.7, no.1, pp.37-47, 1995
- [21]Ernesto L. Andrade, Scott Blunsden, and Robert B. Fisher, “Hidden markov models for optical flow analysis in crowds”, 18th International Conference on Pattern Recognition. ICPR'06, vol.1, pp. 460-463, 2006
- [22] Vimal Thilak and Charles D. Creusere; “Tracking of extended size targets in H.264 compressed video using the probabilistic data association filter”, EUSIPCO2004, pp.6-10, Sep 2004
- [23] 岩崎敏紀,横山貴紀,渡辺俊典,古賀久志, “MPEG ビデオデータの動きベクトルを用いた圧縮領域における移動物体の検出と追跡”, 電子情報通信学会論文誌 D, Vol.J91-D No.6, pp.1592-1603, 2008
- [24] Benzougar, A.,Bouthemy, P.,Fablet, R., “MRF-based moving object detection from MPEG coded video”, Image Processing, 2001. Proceedings. 2001 International Conference on, Volume 3, pp.402-405, Oct. 2001
- [25] Ming-Sui Lee,Shen, M.,Kuo, C.-C.J, “A DCT-Domain Video Alignment Technique for MPEG Sequences”, 2005 IEEE 7th Workshop on Multimedia Signal Processing, pp.1-4, Nov. 2005
- [26] Liu, Long,Feng, Xingle, Ji, Ruirui,Deng, Y, “A Moving Object Segmentation in MPEG Compressed Domain Based on Motion Vectors and DCT Coefficients”, Image and Signal Processing, 2008. CISP '08. Congress on, Volume 3, pp.605-609, May 2008
- [27] Zen, H, Hasegawa, T.,Ozawa, S, “Moving object detection from MPEG coded picture”, Image Processing, 2001. Proceedings. 2001 International Conference on, Volume 3, pp.402-405 vol.3, Oct. 2001
- [28] 米山暁夫,中島康之,柳原広昌,菅野勝, “MPEG ビデオストリームからの移動物体の検出”, 電子情報通信学会論文誌 D, Vol.J81-D2, No.8, pp.1776-1786, 1998
- [29] Lan Dong, Schwartz,S.C, “DCT-Based Object Tracking in Compressed Video”, IEEE proc. ICASSP, Volume 2, pp.665-668, May 2006
- [30] A. Senior, “Tracking people with probabilistic appearance models”, In Proc. IEEE International Workshop on Performance Evaluation of Tracking and Surveillance, pp.48-55, 2002

- [31] T. Zhao and R. Nevatia. "Tracking multiple humans in complex situations", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp.1208-1221, 2004
- [32] A. Senior, A. Hampapur, Y.L. Tian, L. Brown, S. Pankanti, and R. Bolle. "Appearance models for occlusions handling", *Journal of Image and Vision Computing*, Volume 24, Issue 11, pp 1233-1243, November 2006
- [33] J. Yang, Z. Shi, P. Vela, and J. Teizer. "Probabilistic multiple people tracking through complex situations", In *Eleventh IEEE International Workshop on Performance Evaluation of Tracking and Surveillance*, pp79-86, 2009
- [34] A.M. Elgammal and L.S. Davis. "Probabilistic framework for segmenting people under occlusion", In *IEEE International Conference on Computer Vision*, volume 2, pp145-152, 2001
- [35] Hasan C, elik, Alan Hanjali'c, and Emile A. Hendriks, "Towards a robust solution to people counting", In *Interational Conference on Image Processing (ICIP)*, IEEE, Ed., pp.2401-2404, October 2006
- [36] Wusheng Wen, Mengfen Ho, and Chunglin Huang, "People tracking and counting for applications in video surveillance system", In *International Conference on Audio, Language and Image Procesing*, IEEE, Ed., pp.1677-1682, July 2008
- [37] Chao-Ho (Thou-Ho) Chen, Yin-Chan Chang, Tsong-Yi Chen, and Da-JinnWang, "People counting system for getting in/out of a bus based on video processing", In *International Conference on Intelligent Systems Design and Applications*, IEEE, Ed., pp.565- 569, November 2008
- [38] Massimiliano Bozzoli, Luigi Cinque, and Enver Sangineto, "A statistical method for people counting in crowded environments", In *International Conference on Image Analysis and Processing*, IEEE, Ed., pp.506-511, September 2007
- [39] Antonio Albiol, Inmaculada Mora, and Valery Naranjo, "Real-time high density people counter using morphological tools", *IEEE Trans. on Intelligent Transportation Systems*, vol. 2, no. 4, pp.204-218, December 2001
- [40] C. Harris and M. Stephens, "A combined corner and edge detector", In *Proceedings of the 4th Alvey Vision Conference*, pp. 147-151, 1988
- [41] A. Albiol, M. J. Silla, A. Albiol, and J. M. Mossi. "Video analysis using corners motion analysis", In *Eleventh IEEE International Workshop on Performance Evaluation of Tracking and Surveillance*, pp31-37, 2009
- [42] Bay.H, Ess.A, Tuytelaars.T, and Gool.L.V, "Surf: Speeded up robust features", *Computer Vision and Image Understanding*, 110(3) pp.346-359, 2008
- [43] D. Conte, P. Foggia, G. Percannella, ,F.Tufano,M. Vento, "A Method for Counting People in Crowded Scenes", *7th IEEE International Conference on Advanced Video and Signal Based Surveillance*, pp.225-232, 2010
- [44] Jingwen Li, Lei Huang, Changping Liu, "Robust people counting in video surveillance: Dataset and system", *Advanced Video and Signal-Based Surveillance (AVSS)*, 2011 8th IEEE International Conference, pp.54-59, 2011

- [45] Tsung-Yi Lin, Yen-Yu Lin, Ming-Fang Weng, Yu-Chiang Wang, Yu-Feng Hsu, Hong-Yuan Mark Liao, "Cross Camera People Counting with Perspective Estimation and Occlusion Handling", Information Forensics and Security (WIFS), 2011 IEEE International Workshop, pp.1-6, 2011
- [46] Masamichi Shimosaka, Shinya Masuda, Rui Fukui, Taketoshi Mori and Tomomasa Sato, "Counting pedestrians in crowded scenes with efficient sparse learning", Pattern Recognition (ACPR), 2011 First Asian Conference, pp.27-31, 2011
- [47] 原口真悟, 福司謙一郎, 熊澤逸夫, "多数のステレオカメラと色情報を用いた人物追跡手法", 電子情報通信学会技術研究報告. HIP, ヒューマン情報処理 109(471), 229-234, 2010
- [48] 柴田雅聡, 生形徹, 有江誠, モロアレサンドロ, 寺林 賢司, 梅田 和昇, "差分ステレオを用いた人流計測手法の屋外実環境における実証", 日本機械学会論文集 C 編, Vol.79, No.800, pp.1036-1045, 2013
- [49] 榎原孝明, 馬場賢二, 大村明弘, 永野正明, "平面投影ステレオ法を用いたカメラキャリブレーション手法", 電子情報通信学会技術研究報告 コミュニケーションクオリティ 105(281), pp.45-48, 2005
- [50] T. Watanabe, S. Ito and K. Yokoi, "Co-occurrence histograms of oriented gradients for pedestrian detection", Proc. 3rd IEEE Pacific-Rim Symp. Image Video Technol., pp.37-47, 2009
- [51] C. Cortes, and V. Vapnik, "Support-vector network", Machine Learning, vol. 20, pp. 273-297, 1995
- [52] 3GPP TS 22.258, "Service Requirements for the ALL-IP Network (AIPN):Stage1", v2.0.0, Oct, 2005
- [53] 石川孝明, 渡辺裕, "空間スケラビリティを用いた画像解像度変換法", 画像電子学会誌第 34 巻第 5 号, pp.548-557, 2005
- [54] 平田 弘毅, 池原 雅章, "誤り訂正符号を用いた電子透かし", 電子情報通信学会技術研究報告. DSP, デジタル信号処理 100(719), pp.95-102, 2001
- [55] 谷田隆一, 中村健, 清水淳, 八島由幸, "予測残差信号のエッジ情報を利用した動きベクトル符号量削減手法に関する一検討", 電子情報通信学会論文誌 D, Vol.J92-D, No.10, pp.1723-1725, 2009
- [56] 菅民郎, "Excel で学ぶ多変量解析入門", オーム社, 2007
- [57] 宮地悟史, 滝嶋康弘, 羽鳥好律, "画像性質とバッファ制約を考慮した H.264 レート制御方式", 電子情報通信学会論文誌 D, Vol.J88-D2, No.7, pp.1114-1125, 2005
- [58] 11th IEEE International workshop on Performance Evaluation of Tracking and Surveillance, <http://www.cvg.rdg.ac.uk/PETS2009/>

[59]JointVideoTeam(JVT), “ReferenceSoftware”, <http://iphome.hhi.de/suehring/tml/>