

平成18年度 修士論文

View Interpolationを用いた多視点映像符号化の
効率改善に関する研究
Research on Improvement of Multi-View Video Coding
Using View Interpolation

指導教授 渡辺 裕 教授

2007年3月

早稲田大学大学院 国際情報通信研究科
国際情報通信学専攻

4205A127-7

後 藤 崇 行

Takayuki GOTO

目次

第1章	序論	1
1.1	研究の背景	1
1.2	本研究の目的	3
1.3	本論文の構成	4
第2章	多視点映像システム	5
2.1	まえがき	5
2.2	多視点幾何	5
2.2.1	ピンホールカメラモデル	5
2.2.2	エピポーラ幾何	8
2.3	撮影方法	9
2.3.1	Plenoptic Function と Plenoptic Sampling	9
2.3.2	カメラ配置の種類	9
2.4	補正方法	10
2.4.1	カメラキャリブレーション	10
2.4.2	rectification	11
2.5	符号化方式	13
2.5.1	動画像符号化標準方式	13
2.5.2	多視点符号化の標準化	16
2.5.3	多視点符号化の従来研究	18
2.6	イメージベーストレンダリング	21
2.6.1	形状情報を使わない表現	21
2.6.2	形状情報を使用する表現	22
2.7	表示方法	23
2.7.1	両眼視差方式	23
2.7.2	裸眼立体視方式	24
2.8	むすび	26
第3章	多視点映像符号化の予備実験	27
3.1	まえがき	27
3.2	テストシーケンス	27
3.3	時間相関と視点相関	28

3.3.1	隣接フレーム間の相関	28
3.3.2	片方のみの相関を用いた符号化	30
3.4	視点間予測による符号化	31
3.4.1	予測構造に関する検討	33
3.4.2	重み付け予測の効果	33
3.5	サイマルキャスト符号化と適応予測符号化	35
3.5.1	使用する予測構造及び符号化条件	35
3.5.2	全体での比較	37
3.5.3	時間間隔の違いによる比較	39
3.6	むすび	39
第4章	View Interpolation を用いた多視点映像符号化方式	41
4.1	まえがき	41
4.2	基本とする符号化方式	41
4.3	提案手法	42
4.3.1	レート歪みに基づく奥行き情報探索	43
4.3.2	奥行き情報・視差ベクトル変換	44
4.3.3	奥行き情報の一貫性の利用	48
4.3.4	スキップ/ダイレクトモードの改良	49
4.4	むすび	51
第5章	評価実験	52
5.1	まえがき	52
5.2	実験条件	52
5.3	結果及び考察	53
5.3.1	レート歪みに基づく奥行き情報探索	53
5.3.2	奥行き情報・視差ベクトル変換	57
5.3.3	奥行き情報の一貫性の利用	63
5.3.4	スキップ/ダイレクトモードの改良	65
5.3.5	統合結果	68
5.4	むすび	68
第6章	結論	69
6.1	総括	69
6.2	今後の課題	70
	謝辞	72
	参考文献	73

圖一覽	77
表一覽	78
研究業績	79

第1章

序論

1.1 研究の背景

近年、次世代の映像表現技術として3次元映像または自由視点映像が注目を浴びてきている^(1, 2, 3)。3次元映像とは、映像を立体的に視聴することができたり、視聴者の思うまま自由に視点を変更することができる映像表現であり、あたかもその場にいるような高臨場感を体感することが可能である。現在我々が日常で目にする2次元映像は1台のカメラで撮影された映像であり視点は1つである。2次元映像を画像処理によって立体的に提示する技術もあるが、視点を望むように移動させることはできない。3次元映像の素材は、シーンを複数台のカメラを同期させて撮影することで得られる多視点映像を基本とする。3次元映像が注目を浴びてきている主な要因として、デバイス技術の進歩、コンピュータ・通信技術の発展および社会的需要の増加が挙げられる。

まず、デバイス技術の進歩についてである。3次元映像に関連するデバイス技術は、撮像系、座標入力系、表示系の3つの分野に分けられる。撮像系の分野では、LSI (Large Scale Integration) に代表される半導体技術やナノテクノロジー、微細光学技術の進歩により、CCD (Charge Coupled Device) に代表される撮像素子は既に数百万画素に達するまでとなった。今後も更なる高画素化と共に低価格化が進むことが予測され、より現実感のある、複数のカメラによる撮影が可能になってきている。座標入力系の分野では、シーンの形状情報を得る装置が発展してきており、例えば、光切断法とラインセンサを組み合わせた非接触のものなど、より簡易に形状情報を取得することができるようになってきている。また戸外環境においても、レーザーを利用して大型の建造物の測定も可能になり、観測点の座標はGPS (Global Positioning System) で得られるようになった。表示系の分野では、液晶、プラズマなどを利用したフラットパネルディスプレイの大画面化が急速に進んでいる。液晶は画素単位でフィルタなどを装着でき、元々偏光を利用したデバイスであるため、3次元映像の偏光方式とも相性が良く、またレンチキュラ板の貼り付けも容易である。液晶による3Dディスプレイは既に市場に出回り始めた。

次に、コンピュータ・通信技術の発展についてである。上記デバイスを駆使して得られる3次元映像のデータはデジタルデータであり、コンピュータに取り込んで容易に処理を行うことができる。しかし課題はそのデータ量であり、特に形状情報を取得しない光線ベースの

場合には、多数のカメラを用いて空間を密度濃く撮影する必要があり、得られるデータはギガバイトからテラバイトのオーダーになる。近年HDD (Hard Disk Drive) 容量が増加しているにもかかわらず、テラバイト級の容量が普及段階にあるとは言えず圧縮は必要不可欠であると言える。取得されたデータに対する処理技術も進歩してきている。デバイスによって形状情報が得られなかった場合の画像情報のみから形状情報を推定するCV (Computer Vision) の技術、得られた形状データをコンピュータ内に構築し表示するCG (Computer Graphics) の技術である。CGやCVの技術を融合することで、人間につけた座標入力センサなどを組み合わせて、仮想空間内を歩き回るような体験も可能になってきた。また、3次元画像処理プログラムを容易に行えるようにするために、Web3DやOpenGLなどの高級言語も整備されつつある。一方通信分野では、一般家庭にも光ケーブルが導入され広帯域化が進んでおり、インターネットを介して高速の画像通信が可能になってきている。現状のネットワークでは、ギガバイト、テラバイトのオーダーになる3次元映像の全データを配信するには未だ現実的ではないが、ステレオ映像や少数視点映像などは圧縮符号化を施し伝送できる。モバイル通信の環境も整いつつあり、有線ほどではないが通信速度は年々向上している。また、最近の携帯電話にはカメラやGPS機能が搭載されており、位置を特定するのが容易になっていることから、座標入力装置として手軽に使用することも可能である。

以上のような技術の進歩により3次元映像の普及がより現実味を帯びてきた今日では、医療、教育・ビジネス、芸術、娯楽といった分野を筆頭に需要が高まってきている。医療分野では、CT (Computer Tomography) , MRI (Magnetic Resonance Imaging) , UT (Ultrasonic Diagnosis) などの画像診断機器は常時複数断面画像を撮っており、必要に応じて3次元としてのデータも必要となり、これを医師が見やすいように表示しなければならない。今後進むことが予想されるコンピュータ支援手術、内視鏡手術なども医師に正確な3次元情報を提示する必要があり、遠隔医療ではインターネットなどを介して患者の容態を3次元的に診ることが必要なことから、3次元映像が大きな役割を果たすことは間違いない。教育分野やビジネス分野などでは、遠隔操作や遠隔学習、遠隔共同作業が非常に有益であり、コスト削減や効率化につながる。教育・ビジネスの分野に留まらず、社会生活の多くの場面で良質な3次元映像を必要とする局面が増えてきている。芸術分野では、歴史的な文化遺産や伝統舞踊を劣化のないデジタル化によって記録保存することが必要とされており、詳細な3次元映像として保存することが望ましいとされている。娯楽の分野としては、スポーツ、ゲーム、映画が挙げられる。スポーツでは、好みの選手を見たりするために視点機能が要求されたり、ゲームは元来3次元データとして作成されることが多い。映画は、更なる臨場感を与えるために、ハリウッドを中心に3次元化が進んでいる。

以上、3次元映像が、関連技術の進歩により現実味を帯びてきており社会的な需要が増加してきている背景について述べたが、課題としては、取得される映像データの膨大さが挙げられ、現状では圧縮符号化が欠かせない。

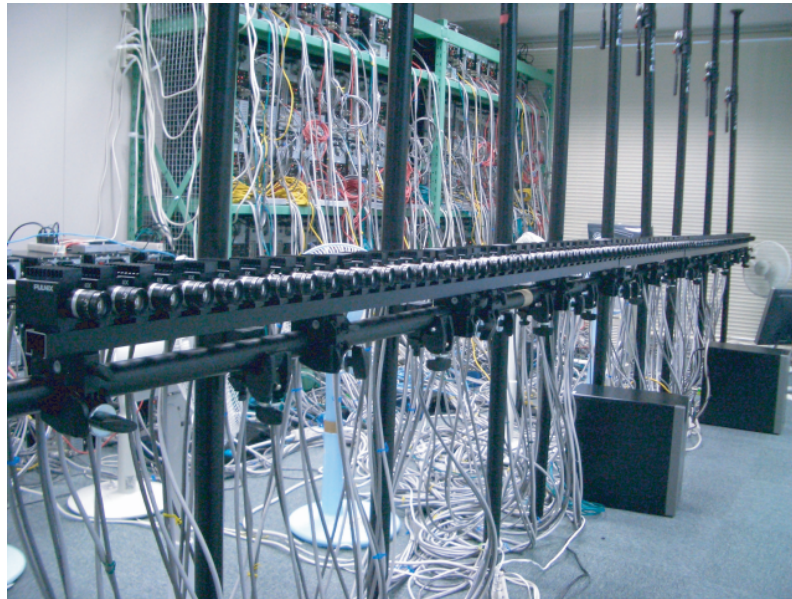


図 1-1: 100 台のカメラによる撮影⁽⁵⁾

1.2 本研究の目的

本研究では、3次元映像の素となる多視点映像の符号化効率改善を目的とする。ネットワークの広帯域化や蓄積装置の大容量化が進んでいとはいえ、3次元映像を非圧縮で扱えるような状況にはなっていない。光線ベースの場合にはカメラを密度高く設置して撮影するため、使用されるカメラ台数は数十台から数百台となるため、特にネットワークを通じた配信においては、高圧縮率を達成できなければ現実的ではない。例えば、Carnegie Mellon 大学では 50 眼近いシステム⁽⁴⁾が、名古屋大学では 100 眼のシステム⁽⁵⁾が、Stanford 大学では 100 眼を超えるシステム⁽⁶⁾が既に開発されている。名古屋大学の 100 眼システムを図 1-1 に示す。将来はますますカメラが安価になり、撮影技術が進歩して多視点映像が容易に取得可能になると、100 台のカメラで撮影するということがより一般的になってくると考えられる。100 台のカメラの場合に、非圧縮の映像データはどの程度になるか試算してみる。カメラの空間解像度を XGA (1024×768)、フレームレートを 30 [fps]、各画素の RGB のビット深度をそれぞれ 8 [bit] とすると、 $1024 \times 768 \text{ [pixel]} \times 30 \text{ [fps]} \times 8 \text{ [bit]} \times 3 \text{ [RGB]} \times 100 \text{ [camera]} \approx 56 \text{ [Gbps]}$ となる。更なる高臨場感を求めるために、空間解像度やフレームレート、ビット深度も向上すると、更にデータレートは増加する。現在のネットワークの帯域がせいぜい 100 [Mbps] であることや流通している HDD の平均的な容量が 300 [GB] 程度であることを考えると、圧縮符号化は欠かせない。

このように、カメラの台数に比例してデータ量が膨大となる多視点映像であるが、従来の 2 次元映像における時間方向の相関の利用に加え、如何に視点間の相関を利用するかが課題であり、時間方向の相関と視点間の相関を適応的に選択する符号化方式を基本に検討されてきた⁽⁴¹⁾。一方 2 章で述べるように、多視点画像間には多視点幾何と呼ばれる幾何的な関

係が存在する．ある視点画像から多視点幾何を使って仮想的な画像を生成する技術を View Interpolation というが，View Interpolation を符号化の枠組みに導入することで符号化効率の向上が見込めると考えられる．View Interpolation を用いた多視点映像符号化方式には，View Interpolation を行う際に使用する奥行き情報をエンコーダ側で生成・符号化する方式⁽³⁴⁾と奥行き情報をデコーダ側で生成する方式⁽³⁵⁾があるが，本研究では前者を基本とし，奥行き情報の探索方法や奥行き情報と視差ベクトルの関係，奥行き情報の性質を利用した符号化効率の改善を目指す．動画像符号化方式としては，2003年3月にISO (International Organization for Standardization) の MPEG (Moving Picture Experts Group) と ITU-T (International Telecommunication Union - Telecommunication Standardization Sector) の共同組織 JVT (Joint Video Team) において H.264/MPEG-4 Part 10 AVC (Advanced Video Coding) (以下，H.264/AVC)⁽¹²⁾ が国際標準化され，現在の汎用符号化方式の主流である MPEG-2⁽⁹⁾ が使用されている領域において，MPEG-2 に取って代わる勢いで普及している．よって今後世の中の動画像符号化は H.264/AVC が中心となっていくことが考えられ，本研究においても H.264/AVC を拡張した多視点映像符号化の効率改善を目指す．

1.3 本論文の構成

以下に本章以降の構成を示す．

第2章 本研究で想定する多視点映像システムについて述べる．まず，多視点画像を扱う上で基本となる多視点幾何について述べ，システムの構成要素である取得，補正，符号化，画像ベースのレンダリング，表示の各要素について説明する．この中で特に符号化は本研究の対象とする領域であり H.264/AVC を中心に説明する．

第3章 多視点映像符号化の特性を調査するために，まず時間相関と視点相関について符号化を交えて調査する．次に，視点間予測のみを行った場合の予測構造およびカメラ間の個体差，照度差を補正するために H.264/AVC の重み付け予測を適用し，それぞれの効果について検討を行う．最後に，時間方向予測と視点間予測の両者が使用可能な場合，どの程度視点間予測が効果的かについて検討を行う．

第4章 奥行き情報を探索して符号化する View Interpolation を用いた多視点映像符号化およびこの方式を改良した符号化効率改善手法を提案する．奥行き情報の探索方法や奥行き情報と視差ベクトルの関係，3次元空間に対する奥行き情報の性質を利用した手法を提案する．

第5章 評価実験を行い，提案手法のそれぞれの手法の効率に関して結果を述べ，考察を行う．

第6章 本研究の総括と今後の課題について述べる．

第2章

多視点映像システム

2.1 まえがき

前章で述べたように、3次元映像、自由視点映像は、実際には多視点映像として得ることができる。すなわち、多視点映像は3次元映像、自由視点映像を包含する関係となっている。そこで本章では、多視点映像の圧縮伝送を考えた場合に必要とされる要素について述べる。多視点映像はあるシーンを複数のカメラを使って同時に撮影することで得られるため、各カメラ映像には同じ被写体がカメラの位置などに応じたずれ（視差）を伴って写る。この視差は、カメラ間で成り立っている多視点幾何によるものであり、多視点幾何は多視点画像を扱う上で全ての要素の基礎となっているため、まずこの多視点幾何について説明する。次に、多視点映像を圧縮伝送する場合の要素は取得、補正、符号化、画像ベースのレンダリング、表示が考えられこれらについて概観する。

2.2 多視点幾何

多視点幾何は、カメラをモデル化したピンホールカメラモデル及びエピポラ幾何によって説明される^(3, 13, 14, 15)。また、多視点映像は複数台のカメラを同期させて得るため、これらの画像間にはエピポラ幾何と呼ばれる多視点特有の制約が働く。以下これらについて述べる。

2.2.1 ピンホールカメラモデル

図 2-1 にピンホールカメラモデルを示す。カメラの光学中心位置を C 、世界座標点を M とする。 M と C を含む線との画像平面との交点为世界座標における画像点であり、 m とする。 C を含む画像平面に直交する線を主軸（principal axis）、主軸と画像平面との交点を主点（principal point）、 C を含む平面を主平面（principal plane）又は焦点平面（focal plane）と呼ぶ。 C と画像平面との距離が焦点距離 f である。

世界座標系における点を $[x \ y \ z]^T$ とすると、三角測量の原理から画像平面には $[fx/z \ fy/z \ f]^T$ としてマッピングされる。世界座標点と画像点が同次座標で表現される場合、以下のように表現される。

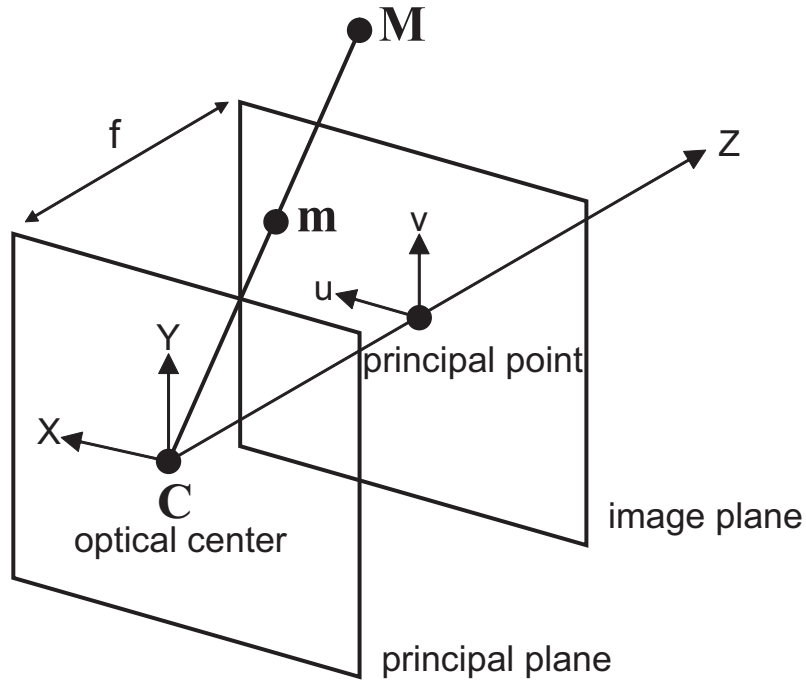


図 2-1: ピンホールカメラモデル

$$\begin{bmatrix} fx \\ fy \\ z \end{bmatrix} = \begin{bmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix} \quad (2.1)$$

式 (2.1) は射影変換行列 P を用い, 同次座標表現を導入することで以下のように表現される.

$$z\mathbf{m} = P\mathbf{M} \quad (2.2)$$

$\mathbf{M} = [x \ y \ z \ 1]^T$ 及び $\mathbf{m} = [fx/z \ fy/z \ 1]^T$ であり, それぞれ同次座標表現である. z はカメラの焦点平面から M への距離, すなわち奥行きを表現している. しかし, 特別な場合でない限り, 式 (2.2) の M の 3 番目の座標値と z の値は一致しない. このように, 3次元空間上の点は射影変換行列 P によって画像座標に変換されるが, これは QR 分解を使って以下のように分解できる.

$$P = K \begin{bmatrix} R|t \end{bmatrix} \quad (2.3)$$

行列 K はカメラの内部パラメータのみを含んでいることから内部行列と呼ばれ, 正則な上三角行列である. R は世界座標系に対する回転行列, t は世界座標系に対する並進ベクトルであり, R と t を外部パラメータという. K はカメラ座標から画像座標への変換を表現しており, 以下のように定義される.

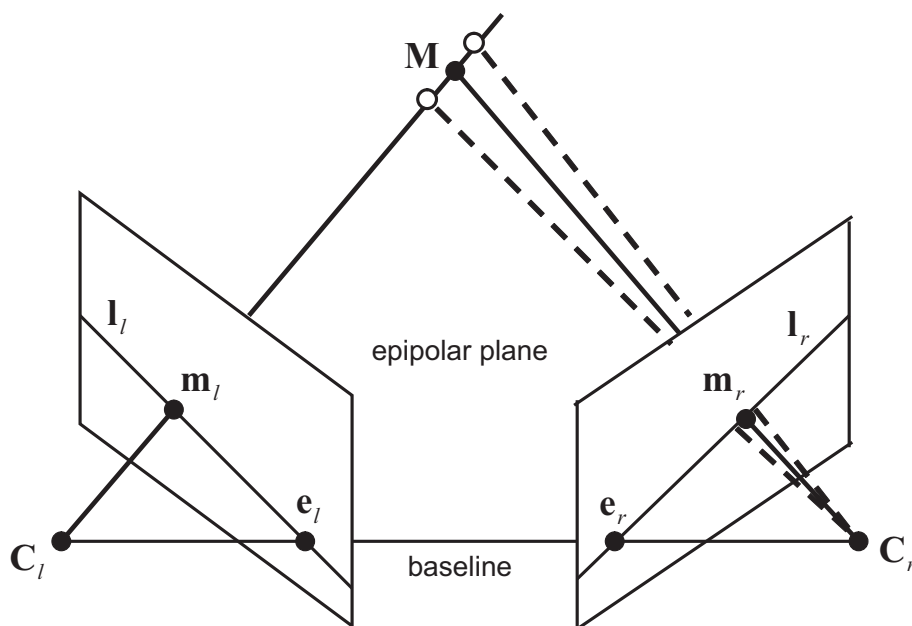


図 2-2: エピポーラ幾何

$$K = \begin{bmatrix} fk_u & \alpha & u_0 \\ 0 & fk_v & v_0 \\ 0 & 0 & 1 \end{bmatrix} \quad (2.4)$$

fk_u, fk_v は、それぞれ水平方向画素及び垂直方向画素における焦点距離を表しており、 (u_0, v_0) は主点の画像座標であり、 α は x 軸と y 軸の非直交性をモデル化する歪みファクターである。

外部パラメータは世界座標系に関するカメラの位置と方向を記述したものである。カメラの光学中心 C は、射影が定義されない唯一の点であるため、以下のように表される。

$$P \begin{bmatrix} C \\ 1 \end{bmatrix} = 0 \quad (2.5)$$

この式を解くと、

$$C = -P_{3 \times 3}^{-1} P_{.,4} \quad (2.6)$$

$P_{3 \times 3}$ とは、 P の最初の 3×3 小行列を表しており、 $P_{.,4}$ は P の 4 番目の列を表している。画像点 $m = [u \ v \ 1]$ の光線は、 m に射影する空間上の点の軌跡であり、 C と m に射影する無限遠にある点を結ぶパラメトリックな線として記述できる。

$$M = \begin{bmatrix} C \\ 1 \end{bmatrix} + z \begin{bmatrix} P_{3 \times 3}^{-1} m \\ 0 \end{bmatrix} \quad (2.7)$$

2.2.2 エピポーラ幾何

エピポーラ幾何は同じ3次元シーンの2つの透視投影間の幾何学的関係を記述したものである。重要な点は、画像の対応点は特定の画像内の線上にあるということである。すなわち、片方の画像内のある一点が与えられると、もう片方の画像内の対応点探索は2次元ではなく、1次元探索となり、高精度の対応点探索が可能となる。

2台のカメラの例を図2-2に示す。今、3次元空間上の点 M が左右のカメラで見えているとし、左のカメラの射影変換行列を P_l 、右のカメラの射影変換行列を P_r とする。すると、式(2.2)から以下のように表現できる。

$$z_l \mathbf{m}_l = P_l M \quad (2.8)$$

$$z_r \mathbf{m}_r = P_r M \quad (2.9)$$

左カメラにおける対応点が m_l 、右カメラにおける対応点が m_r である。この式を幾何的に説明すると、左のカメラの光学中心 C_l と M を結ぶ光線と左のカメラ画像との交点が、 m_l である。同様に、右のカメラの光学中心 C_r と M を結ぶ光線と右のカメラ画像との交点が m_r である。このように、 M と C_l, C_r の3点は平面を定義することがわかり、この平面はエピポーラ平面と呼ばれる。当然 m_l, m_r もエピポーラ平面内にあることがわかる。図2-2内の l_l, l_r はエピポーラ線と呼ばれ、エピポーラ平面と画像平面との交点の集合である。 C_l と C_r を結ぶ線をベースラインといい、画像平面とはエピポールと呼ばれる点 e_l, e_r において交差する。左のカメラ画像内のエピポール e_l は、 C_r の左画像への投影であり、右のカメラ画像内のエピポール e_r は、 C_l の右画像への投影となっている。左の画像内の全てのエピポーラ線は e_l を通過し、右の画像内の全てのエピポーラ線は e_r を通過する。

以上から、点 m_l が与えられると、対応点 m_r が存在しなければならないエピポーラ線が決定することがわかる。以下、 m_l が与えられた場合の、右画像におけるエピポーラ線 l_r の方程式を導出する。エピポーラ線の方程式は、光線を記述する方程式(2.7)を式(2.9)に代入することから得られる。

$$\begin{aligned} z_r \mathbf{m}_r &= P_r M \\ &= P_r \begin{bmatrix} C_l \\ 1 \end{bmatrix} + z_l P_r \begin{bmatrix} P_{3 \times 3, l}^{-1} \\ 0 \end{bmatrix} \\ &= e_r + z_l P_{3 \times 3, r} P_{3 \times 3, l}^{-1} \mathbf{m}_l \end{aligned} \quad (2.10)$$

式(2.10)は、画像点 m_l から m_r への Warping と呼ばれる。左の画像内のエピポーラ線に対する方程式も同様に導出できる。

2.3 撮影方法

多視点映像とは、2台以上のカメラで同時に撮影することで得られるが、これは3次元空間の2次元カメラ画像へのマッピングである。3次元空間と言っても、空間は3次元で表現できるわけではなく Plenoptic Function として表現される。また、ある空間をカメラで撮影することは、Plenoptic Function のサンプリングであり、Plenoptic Sampling として知られている。以下、これらについて説明する。また多視点映像は複数のカメラで撮影を行うため、カメラの配置方法には様々な自由度が存在する。現在広く使用されているカメラ配置について説明する。

2.3.1 Plenoptic Function と Plenoptic Sampling

Plenoptic Function という概念が、Adelson らによって導入された⁽¹⁶⁾。これは、各方向から発せられる、3次元シーン内の各点で観測される放射輝度の強度を記述したものであり、任意の動的シーンに対して、Plenoptic Function は以下のように7次元関数となる。

$$P = P(V_x, V_y, V_z, \theta, \phi, t, \lambda) \quad (2.11)$$

P は「3次元空間内位置 (V_x, V_y, V_z) における、時刻 t の波長 λ に対する角度 (θ, ϕ) の光線強度」という意味である。すなわち7次元の内訳は、空間座標軸 (3D)、空間方位軸 (2D)、時間軸 (1D)、波長 (1D) である。現在、この7次元関数を完全に取得できる装置はないため、それぞれのシステムの目的に応じた制約条件を科して次元数を減らす必要がある。

7次元で表現される Plenoptic Function を2次元の画像上にマッピングする装置がカメラであると考えられ、これは、Plenoptic Function のサンプリング問題と見なすことができる。Chai らは、この空間のサンプリング問題を一般化させる目的で、Plenoptic Sampling を提案した⁽¹⁷⁾。BRDF (Bidirectional Reflectance Distribution Function: 双方向反射率分布関数) は Lambertian であるという仮定の下で解析を行っており、サンプリング間隔は、シーンの奥行きを最小値と最大値によって決定されることが示されている。

2.3.2 カメラ配置の種類

一般に用いられるカメラ配置の種類について図2-3に示す。図2-3(a)の直線配置は、あるシーンの一部のみにおいて視点を動かす場合に相当する。同図(b)の発散配置は、視点はほとんど動かさないが、全周360°を見渡すことができる。これは、ウォークスルーなどの没入アプリケーション向けである。同図(c)の収束配置は、被写体に向かって撮影されたものであり、被写体を全周360°から見渡すことができる。(a)、(b)、(c)を組み合わせたものもあり、(a)と(c)を組み合わせることで、アーク(円弧)状配置とすることができる。カメラ台数の制限などもあり、一般に、(a)の配置ではカメラ間隔は狭く、(b)、(c)の配置ではカメラ間隔は広くなる。

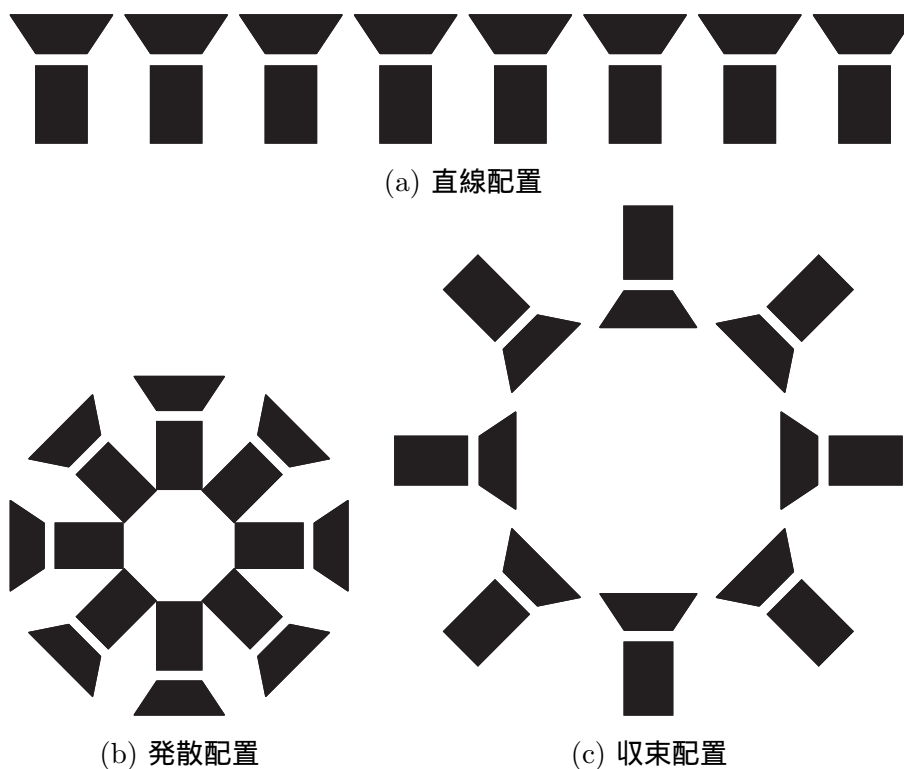


図 2-3: カメラ配置の種類

2.4 補正方法

カメラ間には、ピンホールカメラモデルによるエピポラ幾何が存在することは2.2節で述べた。しかし、エピポラ幾何を用いて対象シーンを復元するには、カメラキャリブレーションを行ってカメラパラメータを求める必要がある*。また、任意の視点を生成するレンダリングにおいてもカメラパラメータは必要となってくる。そこで、まずカメラパラメータを求める方法について述べる。カメラパラメータを画像のみから求めるのは難しい問題であり、必ずしも正確なパラメータが得られるとは限らない。そこで、得られたカメラパラメータを基に、撮影された画像自体を所望のカメラパラメータに合わせて変換する rectification を行うことで都合がよくなる場合があり、これについて述べる。

2.4.1 カメラキャリブレーション

カメラの外部パラメータと内部パラメータを求めることをカメラキャリブレーションと言う[†]。Tsai の手法⁽¹⁹⁾ やその他文献^(13, 14, 20) を中心に様々な手法が存在するが、基本的には以下の2種類に分類される。

*内部パラメータのみを求めれば基本行列 E によって、内部パラメータと外部パラメータが求められていなくとも基礎行列 F によって、エピポラ幾何を記述することは可能である。しかし、それぞれに対して復元できるシーンには制限がある。

[†]内部パラメータを求めるのみを指すこともあるが、ここでは外部と内部の両方を意味する。

- 校正器具を使った校正法

3次元空間における幾何情報が非常に正確にわかっているキャリブレーションオブジェクトを撮影することで行われる。3次元空間における点と画像座標点わかっているという前提から、射影変換行列 P を求める。3次元空間における点と画像座標点の1組からは2つの拘束式が得られ、 P の自由度は11(内部5, 回転3, 並進3)なので、6点わかれば P を算出できる。得られた P を内部行列 K および回転行列 R 、並進ベクトル t のパラメータとして表し、それぞれ求めることができる。

- 自校正法

絶対円錐曲線概念を導入する。絶対円錐曲線とは、3次元の回転および並進のもとで不変な曲線である。射影変換式を絶対円錐曲線の式に代入するが、円錐曲線の投影像は直接求められないため、双対とその直線 l の表現にすることができる。エピポールは既知であり、画像上の点は無限遠(同次座標表現における3番目の値がゼロ)であるので、 l について表現可能である。双対の式に l を代入して無限遠の画像座標点に関する2次方程式が2つ得られ、Kruppa方程式から K の要素を求められる。

これら手法は状況によってはしばしば不安定になることが知られており、平均的に安定してキャリブレーションを行うことができる代表的な手法として、Zhangらの手法⁽¹⁸⁾が知られている。この手法は、上記校正器具を使った校正法と自校正法の間位置する手法であり、最尤推定に基づく非線形反復処理による閉形式解となるため、非常によい結果が得られる。手順は以下のとおりである。

1. 寸法の既知なパターンを平面に貼り付ける。
2. 平面を動かすかカメラを動かすかして、異なる方向から撮影されたようになるように画像を取得する。
3. 画像内で特徴点を検出する。
4. 校正器具を使った校正法で説明した手順で内部パラメータと外部パラメータを算出する。
5. 最小二乗法を使って放射状の歪みモデルを用いたレンズ歪み係数を算出する。
6. 最尤推定法を使って全てのパラメータを精錬する。

2.4.2 rectification

rectificationとは、エピポラ線が画像の片方の軸(通常はx軸)に平行になるように、画像を位置合わせをする変換処理を指す。rectificationされた画像は元のカメラを回転させることによって得られる新しいステレオ装置によって取得された画像と見なすことができる。これにより、視差の探索はエピポラ線上の1次元探索となるため、対応点算出の高速化が実現される。

代表的な手法として、⁽²¹⁾を取り上げる。この手法は、2台のカメラがキャリブレーションされており、それぞれの射影変換行列 P_{o1} , P_{o2} が既知であるとしている。この手法による rectification の背景にある考え方は、それぞれの光学中心の回りに沿ってそれぞれのカメラを回転させ、2台のカメラの焦点平面を共平面にするというものである。よって rectification を行うことで新しい射影変換行列 P_{n1} , P_{n2} が得られ、rectification 後のカメラ画像は、単一のカメラが x 軸に沿って平行移動したものとしてみられるようになる。

rectification 前の射影変換行列は、式(2.3)から以下のように表現される。

$$P_{o1} = K_{o1} \left[R_{o1} | -R_{o1}C_1 \right] \quad (2.12)$$

$$P_{o2} = K_{o2} \left[R_{o2} | -R_{o2}C_2 \right] \quad (2.13)$$

よって新しい射影変換行列は、以下のように表現できる。

$$P_{n1} = K \left[R | -RC_1 \right] \quad (2.14)$$

$$P_{n2} = K \left[R | -RC_2 \right] \quad (2.15)$$

カメラの光学中心のみ rectification 前の光学中心を使用し、回転行列と内部行列は両カメラで合わせ、内部行列 K は任意のものを使用すればよい。回転行列 R は行ベクトルを使って以下の通り表現される。

$$R = \begin{bmatrix} r_1^T & r_2^T & r_3^T \end{bmatrix}^T \quad (2.16)$$

これら行ベクトルは、それぞれカメラ座標系の x , y , z 軸の世界座標を示している。この列ベクトルを以下の手順で求めていくこととする。

1. 新しい x 軸はベースラインに平行でなければならない： $r_1 = (C_1 - C_2) / \|C_1 - C_2\|$
2. 新しい y 軸は x 軸に直交していなければならない： $r_2 = k \times r_1$
3. 新しい z 軸は xy 軸に直交していなければならない： $r_3 = r_1 \times r_2$

k は、任意の単位行列であり、 x 軸に直交する平面内の y 軸の位置を決めるものである。元の z 軸に設定すると、新しい y 軸は、 x 軸と z 軸の両方に直交する軸となる。

rectification を行うには、 P_{o1} の画像平面を P_{n1} の画像平面にマッピングする変換を計算する必要がある。この変換は、 $P_{3 \times 3, n1} P_{3 \times 3, o1}^{-1}$ によって与えられる共線性である。光線の方程式(2.7)から以下のように表現できる。

$$M = \begin{bmatrix} C_1 \\ 1 \end{bmatrix} + z_o \begin{bmatrix} P_{3 \times 3, o1}^{-1} m_{o1} \\ 0 \end{bmatrix} \quad (2.17)$$

$$M = \begin{bmatrix} C_1 \\ 1 \end{bmatrix} + z_n \begin{bmatrix} P_{3 \times 3, n1}^{-1} m_{n1} \\ 0 \end{bmatrix} \quad (2.18)$$

上式をまとめ、以下のように対応点が算出できる。

$$m_{n1} = \frac{z_o}{z_n} P_{3 \times 3, n1} P_{3 \times 3, o1}^{-1} m_{o1} \quad (2.19)$$

rectification 後の画像座標は、元の画像の非整数画素位置に相当するため、この場合、双一次補間によって計算される値を使用する。

2.5 符号化方式

多視点映像のデータサイズはカメラの台数に比例して大きくなるため、圧縮符号化が欠かせない。符号化は標準方式と関連が深く、現在の映像符号化方式は MPEG が中心である。MPEG は 2 次元映像符号化を目的としたものであるが、多視点映像符号化にも容易に適用可能であるため、これら標準方式について説明する。

多視点映像に対する標準方式としては MPEG-2 Multi-View Profile があり、現在標準化に向けて活動が行われているグループとしては 3DAV および JVT がありそれぞれ説明する。また、多視点符号化の従来研究についてまとめる。

2.5.1 動画像符号化標準方式

現在動画像符号化標準方式には、H.261⁽⁷⁾、H.263⁽¹⁰⁾、MPEG-1⁽⁸⁾、MPEG-2⁽⁹⁾、MPEG-4⁽¹¹⁾、H.264/AVC⁽¹²⁾ がある。まずこれらの方式間の比較を行い、次に本研究で中心となる H.264/AVC について概観する。

各種標準方式の比較

1990 年、CCITT (Comite Consultatif International Telegraphique et Telephonique) は ISDN (Integrated Services Digital Network) 上の画像音声通信に画像符号化方式 H.261 を勧告した。H.261 は 64[kbps] を基本とする ISDN に対応し、64[kbps]×n (n=1,2,...,30) を想定した符号化方式である。H.261 はハイブリッド符号化方式を採用した標準であり、現在広く活用されている動画像符号化の多くが H.261 の符号化方式を採用している。ハイブリッド符号化を図 2-4 に示す。符号化対象画像と参照画像間で動き推定を行って予測画像を生成し、差分信号を周波数変換、量子化した信号をエントロピー符号化するというものである。符号化ストリームを所望のデータサイズになるよう制御する部分もある。MPEG は方式が固まりつつあった H.261 を参照しつつ、通信ではなく CD (Compact Disc) などの蓄積メディアへの記録のための符号化方式を目標にした。MPEG は三つのフェーズに分けて標準方式の計画を立てた。まず、CD を想定した 1.5[Mbps] 程度までのデジタルビデオ、次いで標準テレビジョンのデジタル符号化、更に HDTV (High Definition TeleVision) のデジタル符号化である。まず、1993 年には MPEG-1 を規格化し、ビデオ CD や CD カラオケに採用された。MPEG-1 は多くの点で H.261 との共通点を持っているが、最も特徴的なことは時間的に前後するフレームから予測を行う B ピクチャの導入である。次いで着手した MPEG-2 は

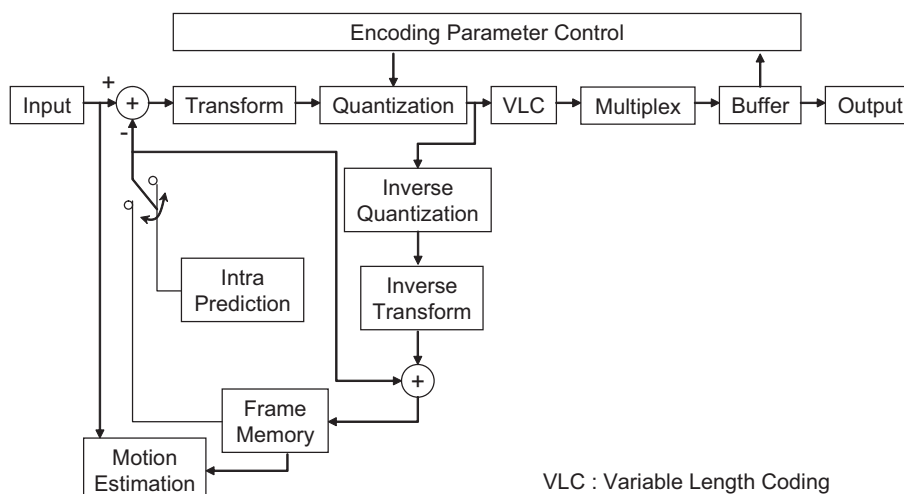


図 2-4: ハイブリッド符号化

ITU-T との共同作業となり、当初は第3フェーズと考えられていた HDTV も吸収し、1994年に MPEG-2/H.262 として規格化された。インタレース符号化などが特徴として挙げられる。MPEG-2 は 4～5[Mbps] 程度の DVD ビデオや 20[Mbps] 程度の HDTV デジタル放送に採用され、現在のデジタル家電の中心となっている。その後 1996 年に H.263 が勧告され、その流れを汲んで 1998 年に MPEG-4 が標準化された。MPEG-4 は 64[kbps] の第3世代携帯電話や 2～3[Mbps] 程度までの携帯用ビデオプレーヤ、デジタルカメラの動画機能などに活用されている。MPEG-4 の特徴的な符号化ツールとして、オブジェクト符号化やグローバル動き補償などが挙げられる。H.264/AVC は H.263 の半分のビットレートで同等の画質を得ることを目標にして検討が始まった ITU-T のプロジェクトである H.26L が前身である。2001 年に MPEG が加わり MPEG と ITU-T の共同組織 JVT が発足し、2003 年に標準化された。ITU-T では H.264、MPEG では MPEG-4 Part10 Advanced Video Coding という名称である。詳細な符号化ツールに関しては、次項で述べる。

これら各種標準方式の変換符号化のブロックサイズ、動き補償サイズ、動きベクトルの精度を表 2-1 に示す。H.261 から MPEG-4/AVC までハイブリッド符号化を用いた方式であり、変換には DCT (Discrete Cosine Transform) が用いられている。独立した処理の最小単位であるマクロブロックは 16×16 で共通であり、動きベクトルを定義するブロックサイズが小さくなっていき、また動きベクトルの精度が高くなってきている。

H.264/AVC

H.264/AVC は、汎用符号化方式 MPEG-2 と比べて約 2 倍の圧縮性能を実現でき、処理量削減技術も成熟してきたため、今現在広く普及してきている。H.264/AVC の特徴的な符号化ツールについて以下で説明する⁽²³⁾。

表 2-1: 各種標準方式の比較

Name	MPEG-1	MPEG-2 H.262	MPEG-4	H.261	H.263	MPEG-4/AVC H.264
Year	1991	1994	1998	1990	1996	2003
Transform	8×8	8×8	8×8	8×8	8×8	4×4
MC block size	16×16	16×16	8×8, 16×16	16×16	8×8 16×16	16×16 ~ 4×4
MV accuracy	1/2 pel	1/2 pel	1/2 pel	1 pel	1/2 pel	1/4 pel

可変ブロックサイズ動き補償 従来の標準方式の動き補償の単位は16×16画素単位が基本であり、H.263およびMPEG-4においては8×8画素単位の動き補償も利用可能となった。H.264/AVCでは、16×16、16×8、8×16、8×8、8×4、4×8、4×4の計7種類が使用可能となっている。8×8未満のサイズは、8×8ブロックに対するサブブロック分割となっている。7種類のブロックサイズが使用可能となり、形状や動きに適した予測を行うことができる。

複数枚参照フレームによる動き補償 従来の標準方式における動き補償フレーム間予測において、参照フレームとして使用できるフレームは、Pピクチャは直前のIピクチャ又はPピクチャであり、Bピクチャは直前および直後のIピクチャ又はPピクチャに制限されている。H.264/AVCでは、複数の参照フレームを参照フレームバッファに格納して保持し、直前又は直後以外のフレームを使用することができる。Bピクチャにおいては、過去の2フレームを参照フレームとして用いたり、Bピクチャを参照することが可能である。複数参照フレームにより、Iピクチャより前のフレームも参照可能となっており、H.264/AVCでは、必ずしもIピクチャから復号を開始することができるとは限らない。そこで、IDR (Instantaneous Decoder Refresh) ピクチャが導入されており、Pピクチャ及びBピクチャは、IDRピクチャをまたいだ参照フレームの使用ができないようになっている。

1/4画素精度動き補償 動き補償の精度は、輝度信号は1/4画素精度、色差信号は1/8画素精度である。1/4画素精度は、MPEG-4 ASP (Advanced Simple Profile) で先に導入されている。1/2画素精度は係数 $[1/32, -5/32, 20/32, 20/32, -5/32, 1/32]$ である6タップFIR (Finite Impulse Response) フィルターを用いて生成され、特に高周波係数に対する再現性が高くなっている。1/4画素精度は、1/2画素精度の線形補間によって生成される。

画素レベルイントラ予測 従来の標準方式では、イントラ予測(フレーム内予測)符号化は、周波数領域によって行われるが、H.264/AVCでは、符号化対象ブロックの隣接符号化済み画素から補間によって予測画像を生成する。予測画像生成の単位は、輝度信号は4×4画素及び16×16画素の2種類が基本である。色差信号は8×8画素の1種類のみである。予測画像生成の種類は、輝度4×4画素の場合は9種類、輝度16×16画素及び色差8×8の場合は4

種類が利用できる。High Profile 以上のプロファイルでは、 4×4 画素単位のイントラ予測を拡張した 8×8 画素単位も利用可能である。 8×8 画素単位予測の場合には、 8×8 サイズの整数変換が、 4×4 又は 16×16 の場合には 4×4 サイズの整数変換が適用される。

整数変換 従来の標準方式では、周波数変換を 8×8 画素のブロックを単位として、原画像又はフレーム間予測の予測誤差画像に DCT (Discrete Cosine Transform) を適用して DCT 係数を求め、その係数を量子化する。DCT はコサイン関数を用いるため、実数精度の演算が必要となり、エンコーダとデコーダでミスマッチが生じる可能性がある。ミスマッチが発生すると復号が進むにつれ画像にノイズが蓄積してしまうため、定期的にイントラマクロブロックを挿入する必要があった。H.264/AVC では、16 ビット整数精度で演算が可能な 4×4 画素サイズ整数変換を採用しており、整数変換は、加減算とビットシフトのみによって演算が可能のため、実装が非常に容易である。また、演算がすべて整数精度で行われることで、実数演算の実装差によるミスマッチが生じない。High Profile より上のプロファイルでは、 8×8 画素サイズ整数変換も使用可能であり、 4×4 整数変換と 8×8 整数変換を適応的に切り替えて使用することが出来る。

閉ループ内デブロッキングフィルタ H.264/AVC では、H.261 で採用された閉ループ内フィルタと似た、閉ループ内にデブロッキングフィルタが設置されている。図 2-4 におけるフレームメモリ直前に設置される。このフィルタは H.261 のようなブロック全体の平滑化フィルタではなく、整数変換のブロック境界のみを平滑化してブロックノイズの発生を抑制するというものである。デブロッキングフィルタは圧縮率向上や主観画質向上のためには効果的であるが、一方でデコーダ側での処理量が大いという問題点も有している。

コンテキスト適応型エントロピー符号化 H.264/AVC では、シンタックス要素のエントロピー符号化として指数ゴロム符号の他に、ハフマン符号をベースとした CAVLC (Context-Adaptive Variable Length Coding) と算術符号を用いた CABAC (Context-Adaptive Binary Arithmetic Coding) のいずれかが使用可能である。CABAC の方が算術符号化を用いていることからより高圧縮が可能である一方演算量が増加するため、CABAC は Main Profile と High Profile 以上で使用可能である。CA (コンテキスト適応型) とは、周囲の状況に応じて効率の良い符号化方式を適応的に選択する、という意味である。

2.5.2 多視点符号化の標準化

多視点映像符号化の標準方式としては、MPEG-2 Multi-view profile がある。また、H.264/AVC をベースにした多視点映像符号化が MPEG 3DAV において検討されていたが、近年 ITU-T と MPEG の共同組織である JVT に活動が移っており、これらについて概観する。

MPEG-2 Multi-View Profile

立体映像符号化として規定されている唯一の標準方式である。1996年9月のMPEG会合で、MPEG-2ビデオ(ISO/IEC 13818-2)に、Multi-View Profile (MVP)として2眼式立体テレビ画像の符号化方式が Amendment 3 として追加された⁽²⁴⁾。

MVPはその名の通りプロファイルを規定しており、すべてのレベルとの組み合わせが可能である。またMVPでは、立体テレビ映像の撮影条件を示すカメラパラメータとして、カメラの位置、方向、レンズの絞り、焦点距離、画角などを伝送する方法が規定されている。

符号化方法は時間スケラビリティを拡張した方式になっており、左眼用画像は、動き補償予測のみを用いる通常のMPEG-2 Main Profile (MP)で符号化され、右眼用画像は、過去の画像から予測する動き補償予測と、同じ時刻に表示される左眼用画像から予測する視差補償予測を用いて符号化される。この視差補償予測の定義を除くと右画像の符号化はMPの符号化と全く同一であり、予測後の残差をDCT、量子化、エントロピー符号化することで画像データを圧縮したビットストリームを得る。

MVPは主にステレオ映像を想定したものであり、多視点映像に関しては詳しく検討されていない。また、時間スケラビリティをベースにしており処理が複雑であるため実用化には至っていない。

MPEG 3DAV と JVT

MPEGでは3次元映像音声技術の国際標準化を見据えた3DAV活動を行っている⁽²⁵⁾。3DAVとは3D Audio Videoの略であるが、3D Videoの標準化提案が行われたことから開始された活動である。

3DAVにおける具体的な活動としては、まずMPEGの既存の標準との差分を明確にする実験EE (Exploration Experiment) が行われた。

EE1:全方位映像

EE2:自由視点テレビ/自由視点映像

EE3:ステレオ映像の符号化

EE4:3D TV用奥行き情報の符号化

EE1とEE3、EE4はMPEG-4として採用されている技術で実現できることが示され、活動を終了した。EE2に関しては、光線空間ベースとモデルベースのものが提案され、モデルベースのVDTM (View Dependent Texture Mapping) やPCR (Point Cloud Representation) は、SNHC (Synthetic Natural Hybrid Coding) グループで規格作成が行われているAFX (Animation Framework eXtension) に採用され、モデルベースの自由視点映像はMPEGの技術で実現できることとなった。一方、光線空間ベースの表現は、自由視点映像の生成手法も含めて検討すべき課題とされた。次にEEの結果を受け、CfC (Call for Comments) を行った結果、産業界の自由視点テレビへの関心が高いことがわかり、自由視点テレビ/自由

視点映像ならびに多視点映像符号化 (MVC (Multi-view Video Coding)) に焦点を絞り、活動を続けることとなった。

MVC は、2004 年の会合で、自由視点テレビを最終ゴールとして CfE (Call for Evidence) が発行され、2005 年の会合で優位性を示す十分な実験結果が得られたことから CfP (Call for Proposal) の発行がなされた。CfP の結果が 2006 年の会合で審議され、HHI の方式が優れているという結果になった。2006 年 7 月まで 3DAV で議論が行われていた MVC は、それ以降 ISO/IEC と ITU-T の共同組織 JVT で検討が行われることとなった。現在 JVT における Core Experiment としては、以下のようなものが検討されている。

- CE9: Illumination compensation
- CE10: View synthesis prediction
- CE11: Disparity and motion vector coding

2.5.3 多視点符号化の従来研究

多視点符号化としては、従来の動き補償予測を視差補償予測に適用した手法⁽²⁶⁾を基に数多くの手法が検討されている。動画像符号化の場合は、適応的に動き補償予測又は視差補償予測を選択して符号化方式を中心に検討されている。従来研究を参照構造、視差ベクトル削減、View Interpolation の適用、基準視点と拡張視点符号化の 4 つに大きく分類した。以下これらについて述べる。

参照構造

多視点符号化の場合、視点画像をどの順序で符号化を行うかについて、様々な可能性があり、これら参照構造に関して検討が行われている。

Magnor らは、カメラが 2 次元の格子状に配置されて撮影された Light Field 画像を符号化する際の順番について提案している⁽²⁷⁾。まず 4 隅の画像をイントラ符号化し、次にこの 4 枚の画像の 1 枚の中心画像を 4 隅の 4 枚の画像から視差補償予測符号化する。そして、4 隅の画像の間の 4 枚の画像を符号化した 5 枚の画像のうちの近い画像から視差補償予測符号化する。符号化したこれら 9 枚の画像を直線で結んでできる四分領域に同様の符号化を再帰的に繰り返す、という参照構造である。視差が小さい Light Field に特化した構造であり、複数カメラの場合には必ずしも有効ではない。

木全らは、受信側で全多視点映像のうち部分復号によって部分的な情報を得る (partial decoding) ため、及び所望の視点映像を得るために必要なデータのみを伝送する機能 (view scalability) のために、GoGOP (Group of GOP) 構造を提案している⁽²⁸⁾。GoGOP は GOP 構造の概念を多視点に拡張したものであり、GOP 単位に視点映像の符号化データを構成し、GOP 間で予測符号化を適用する。GOP を Base GOP と Inter GOP に分類し、Base GOP は同 GOP に含まれる画像情報のみを、Inter GOP は同 GOP に含まれる画像の他に、他の

Inter GOP や Base GOP に含まれる画像を参照して復号可能な GOP である。GoGOP 構造は、想定するシステムの要求条件に応じて適応的に Base GOP と Inter GOP の参照関係を決定することで、様々なシステムに柔軟に対応できるとしており、符号化効率には触れられていない。

石川らは、視点間予測を用いないサイマル方式と端の視点映像から順番に視差補償予測する一端方式、カメラの配置的に重心に最も近いカメラから視差補償予測する中心方式について比較を行っている⁽³⁰⁾。視点間予測は P ピクチャが用いられている。中心方式が良い場合もあるが、被写体までの奥行きが大きく視差の影響が少ないシーケンスでは特に違いがない、という結論が得られている。また文献⁽³¹⁾において、2次元カメラ配置における視差補償予測構造が検討されている。比較方式としては、サイマル方式、一端方式、カメラ配置の重心に最も近い視点映像から順方向予測のみを行う中心方式、重心に最も近い視点映像から双予測も行う中心方式の4つである。双予測を用いる予測構造が一番符号化効率が高くなることが示されている。

視差ベクトル削減

視差ベクトルはカメラの位置や向きに応じて大きくなり、動きベクトル以上に効率的な圧縮が必要とされる。そこで、視差ベクトルの符号化効率を向上させる手法が提案されている。

志水らは、H.264/AVC における動きベクトル予測を多視点映像符号化用に拡張させた手法を提案している⁽³²⁾。視差ベクトルを周囲から予測する際に、周囲のブロック内の参照フレームが符号化対象フレームと異なる場合に、その異なるブロックの変位ベクトルが指している領域を見て、符号化対象フレームと同じ参照フレームの場合にその領域のベクトルを予測ベクトルとする手法である。P ピクチャのみの構造で検討を行っており、時間間隔の異なる B ピクチャに関しては対処していない。

石川らは、H.264/AVC における2種類のダイレクトモード（空間と時間）を用いて視差補償を行った場合の比較を行っている⁽³⁰⁾。ダイレクトモードは、特に低ビットレート時に符号量を削減する有効な方式である。多視点映像シーケンス符号化においては、カメラの向きや色調の違いなどにより、時間ダイレクトモードが有効に働かないため空間ダイレクトモードが適している、としている。

木全らは、カメラ配置が固定な状況下では時間変換に対する視点映像間の視差の変化が少ない、という仮定に基づいた手法を提案している⁽³³⁾。GOP の先頭フレームの視差ベクトルを蓄積して後のフレームの視差ベクトル予測に利用し、この方式を利用しない場合には H.264/AVC の変位ベクトル予測を行って視差ベクトルを予測符号化するというものである。更に文献⁽²⁹⁾において、階層的な参照フレームメモリ構造及び動きベクトルと視差ベクトルの識別のためのメモリ構造について提案し、これらベクトルを符号化する際に利用することで符号化効率を向上させている。

View Interpolation の適用

カメラパラメータが高精度に求められており、このカメラパラメータを使った View Interpolation を符号化の枠組みに直接組み込むという手法が提案されている。多視点特有の関係を記述したカメラパラメータを利用することで、符号化効率の向上が見込めるとするものである。View Interpolation とは、2枚の画像間を補間すること以外に、1枚の画像から外挿して仮想的な画像を生成するという意味も含んでいるとする。

Yea らは、H.264/AVC をベースに、奥行き情報をエンコーダで生成してエントロピー符号化しビットストリームに挿入する、という枠組みを提案している⁽³⁴⁾。ブロック毎に奥行き情報を探索することで、エピソード幾何に基づく射影変換が補償可能となり、また、奥行き情報を上手く探索することで、符号化効率を向上させる可能性がある。しかし、未だ有効な結果は得られていない。本研究における直接的な従来手法であり、後の4章において詳細に述べる。

Tanimoto らは、H.264/AVC をベースに、B ピクチャの符号化効率の改善を目的とし、参照フレーム間のブロックマッチングによる View Interpolation を行って予測画像を生成し符号化する方式を提案している⁽³⁵⁾。奥行き情報を符号化する必要がないことから、符号化効率を向上させている。補間のための視差ベクトルを推定する際に、Droese らの手法⁽³⁷⁾をベースにした補間を行っているが、視点補間には制御パラメータが多いことや、デコーダでもブロックマッチングを行わなければならないことによる演算量の増加という問題点が挙げられる。Kitahara らは、この手法を拡張させ、カメラパラメータの誤差を補正する1次元の補正ベクトルを探索する手法を提案し、符号化効率を向上させている⁽³⁶⁾。

基準視点と拡張視点符号化

複数の視点映像の中から、全ての視点に対して一番相関が高い一つの視点を基準視点として従来の符号化方式で符号化する。その他の拡張視点に関しては、基準視点の復号画像を使って予測を行い、差分を符号化するという枠組みである。

志水らは、基準視点を通常の H.264/AVC で符号化し、その他の視点は奥行き情報を使った 3D-Warping によって基準視点から各視点に対する予測画像を生成し、得られる差分画像を H.264/AVC 符号化するという手法を提案している⁽³⁸⁾。奥行き推定には複数台カメラを用いる Multiple-Baseline Stereo 法を用い、奥行き情報の符号化にはロスレス符号化を用いている。サイマルキャスト符号化と比較しているが、視点間に視差補償を行った場合とは比較していない。

秋山らは、ある視点映像を基本レイヤとしその他の視点映像をこの基本レイヤの映像から予測を行って符号化する方式において、基本レイヤの映像から他の視点映像を予測するとき射影幾何学的関係を利用した手法を提案している⁽³⁹⁾。前景領域と背景領域が分離できているという仮定において、前景領域同士または背景領域同士で対応点を4組以上算出し、射影変換行列を求めている。基本レイヤを射影変換して生成される予測画像との差分画像を MPEG エンコーダに入力するため、基本レイヤの映像、その他の視点映像の差分及び対応

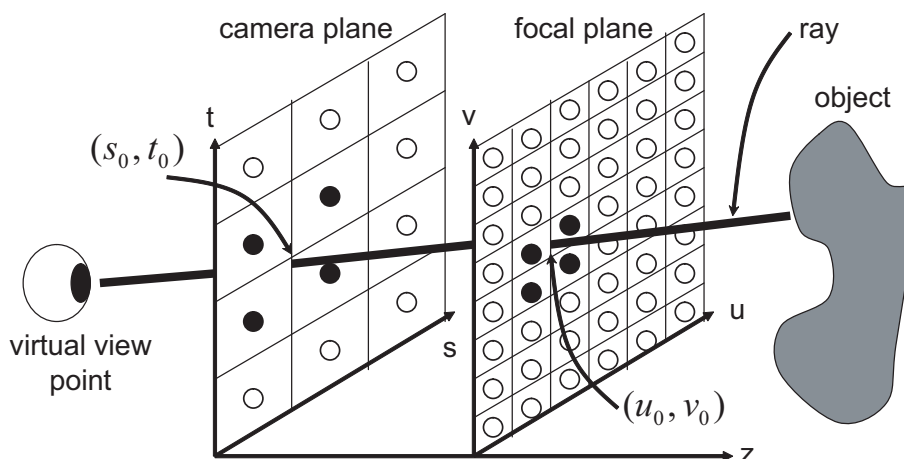


図 2-5: Light Field Rendering/Lumigraph

点の組み合わせデータを符号化する．比較対象がサイマルキャスト符号化であり，視差補償を行う方式との比較が行われていない．

田口らは，被写体空間を俯瞰する代表的な視点画像を代表視点画像とし，この代表視点画像とその他の視点画像間で視点依存幾何モデルを算出して代表視点画像をワーピングし，得られる予測画像との差分画像を符号化する，という枠組みを提案している⁽⁴⁰⁾．低ビットレート時に符号化効率が改善されること，また代表視点からの距離に応じてワーピングによる予測画像の性能が変化し符号化効率が変動することが示されている．

2.6 イメージベースレンダリング

対象とする空間の形状情報をどの程度用いるか，という観点でイメージベースレンダリング技術が分類されている⁽⁴¹⁾．一般に，形状情報を用いるほど空間をモデル化するための計算量が高くなる．逆に用いない場合，大量の画像を必要とするため，データ量が増大する傾向がある．両者はトレードオフの関係にある．これらを統一的に扱う Unstructured Lumigraph Rendering⁽⁴²⁾ も提案されている．以下，この分類方法に基づき，イメージベースレンダリングについて概観する．

2.6.1 形状情報を使わない表現

シーンの形状情報を用いないため，7次元の Plenoptic Function に制約をかけて，次元数を減らすことが主である．式(2-11)における t と λ を減らした 5次元表現の Plenoptic Modeling⁽⁴³⁾，光線をカメラ平面(2次元)と焦点平面(2次元)の4次元で表現する Light Field Rendering⁽⁴⁴⁾ および Lumigraph⁽⁴⁵⁾，カメラを平面の同心円に制約することで回転角(2次元)と半径(1次元)の3次元で表現する Concentric Mosaics⁽⁴⁶⁾ などがある．以下，Light Field Rendering と Lumigraph を取り上げる．

Light Field Rendering は、図 2-5 のように、光線がカメラ平面と焦点平面という平行な 2 つの平面の交点によって表現されるという仮定に基づいている。カメラ平面とはカメラの光学中心の位置を表しており (s, t) のパラメータで、焦点平面とは各カメラの画像座標を表しており (u, v) のパラメータで表される。あるカメラ平面の点と焦点平面の全ての点を結ぶことで、通常の画像が得られる。新しい仮想的な視野の生成は、視野を光線に分解し、光線一本一本を値をすぐ近くの画像座標値を quadrilinear 補間することで得られる。例えば図 2-5 の、光線の例 (u_0, v_0, s_0, t_0) は、2 つの平面の隣接する離散化された点を結ぶことで得られる 16 本の光線から補間されることになる。この処理を仮想視点の解像度分を行うことで仮想視点を得られるため、複雑度はシーンの複雑度ではなく仮想視点の解像度に依存する。

Lumigraph は、データ表現の観点からは Light Field Rendering と同じであるが、取得方法が異なる。Light Field Rendering は規則的なカメラ間隔で撮影するためシーンの形状情報に関する情報は必要がないが、Lumigraph は不規則なカメラ間隔・向きで撮影された画像を対象としているため、画像をカメラ平面と焦点平面にリサンプリングする必要があり、visual hull による近似的な形状情報を用いている。リサンプリングを行い仮想的な視野を生成する手順は Light Field Rendering と同様の手順によって行われるため、形状情報を使用しない表現として分類される。

2.6.2 形状情報を使用する表現

形状情報を用いる表現にも、暗示的か明示的かで分類される。暗示的とは、画像上の対応点のみ用いる表現であり、View Interpolation⁽⁴⁷⁾、View Morphing⁽⁴⁹⁾ がある。一方明示的な形状情報を用いる表現としては、形状情報と画像情報を対応付けたテクスチャマップを生成する View Dependent Texture Mapping⁽⁵⁰⁾、1 画素に複数の奥行き情報を格納する Layered Depth Images⁽⁵¹⁾、形状情報の他に画像を何百枚も撮影して反射マップ (lumisphere) を生成する Surface Light Field⁽⁵²⁾ などがある。以下、暗示的に形状情報を用いる表現について説明する。

2 枚の画像が与えられており、これらの中で密度濃く対応点が求められている場合には、任意の視点を生成することが可能であるとするものが View Interpolation である[‡]。この手法は、2 枚の入力画像がすぐ隣接して撮影されたものであるときに上手く働き、逆に離れて撮影された場合にはオーバーラップ領域が小さくなり生成できない領域が多く発生することとなる。また、全てのカメラが同じ方向を向いている場合にのみ上手く働く。

View Morphing とは、View Interpolation を拡張させ、2 枚の画像間のカメラパラメータが異なっている場合での補間にも対処できるようにしたものである。よって、補間する 2 枚の画像のカメラパラメータが既知であるという前提で 2 枚の画像間の光学中心を結ぶ線において任意視点を生成する。任意視点は、カメラの動きが視線方向に対して垂直な場合には 2 枚の画像の線形補間によって生成される。しかし 2 枚の画像が互いに平行でない場合には、

[‡]本研究では、必ずしも 2 枚の画像間での画像生成を View Interpolation としているわけではなく、1 枚の画像から生成する手法も含めている。

カメラパラメータを使用して pre-warp を行い2枚の画像を平行にした後，平行化した画像を使って任意視点を生成する．生成された視点は pre-warp の影響を受けているため，この影響を無くすために post-warp を行い最終的な仮想視点を得る．

これら対応点を用いる手法において一番困難であるのは，画像間の対応点をどのように算出するかである．対応点算出には，CV の分野において様々な手法が提案されている⁽⁴⁸⁾．しかしこれらは全ての状況に対して安定的に対応点が求まるというわけではなく，現在もこの問題に関して研究が行われている．

2.7 表示方法

3D ディスプレイは，人間の眼が2つあることを利用して，その形状や大きさを正確に表現すべく開発されたものである．3D に見える原理は，実際に人間が見ているものが，左眼と右眼で若干異なっている（視差がある）ことを利用している．

表示方法は主に，メガネを装着する両眼視差方式（メガネ方式）と何も装着しない裸眼立体視方式に分類され，以下これらについて説明する⁽¹⁾．

2.7.1 両眼視差方式

メガネをかけて立体視を可能とする両眼視差方式には，アナグリフ方式，偏光方式，時分割方式，分光視差方式があり，以下これらについて述べる．

アナグリフ方式

ステレオ画像対の片方を赤色に，他方は青色にして，観察者はその補色のメガネ，すなわち片側は赤，他方は青のレンズの入ったものをかければ，片方の目には片方の絵しか入らないというメカニズムを用いた方式である．アナグリフ用メガネは安価であり，また印刷も通常の方式でよいので，現在でも立体写真，図面，絵画，漫画などに広く利用されている．今日でもしばしば映画やテレビで使用されることがある．特に，映画館の上映施設（映写機やスクリーン）の変更を必要とせず，またテレビでも，特別の装置を用意することなく観賞ができるため便利である．

しかし，色フィルタを使用するために，カラー化しにくく，最近のアナグリフの映画では，カラー画像の一部をアナグリフにしてカラー的な表現を試みているケースもある．また，画像の色とフィルタの色的一致が難しいためにクロストークが起こりやすく，更に左右の画像が交互に優位に見える現象（両眼視野闘争）を生じやすく，眼精疲労を起こす原因にもなっている．

偏光方式

アナグリフ方式ではカラー情報が失われるため、この欠点を解決したのが偏光方式である。人間の目は偏光に感じないため、直交した偏光膜を左右のメガネに付け、2台のプロジェクタにも同じようにすれば、左右の目に別々の視差のある画像を提示できるというものである。この方式は構造が簡単であり、同時に多数の人が見ることができるといえるため、現在まで広く用いられている。

その反面、投影装置が2台ないしは特殊な画像を分離する装置を必要としたり、フィルタを二重に必要とするため透過光量が低くなる。また、直線偏光フィルタが使用されている場合は、首を傾けると偏光角がずれ、クロストークが生じ、立体視できなくなる。更に、スクリーンに偏光の乱れが起こらない特別のスクリーンを必要とする。現時点で、解像度や色再現性、多数の観客が同様の状態で鑑賞できる方式としては最も優れているため、広く一般に立体映画に採用されている。

時分割方式

映画は目の残像現象を利用しており、同様に左右の画像を時間的に切り替えて投影し、それと同期して目の前の高速シャッターを切り替えて、左目には左画像だけ、右目には右画像だけに見えるようにしても立体視できる。最近では、先に述べた偏光膜と液晶を組み合わせれば、所要の高速シャッターが容易に実現できる。この方式はプロジェクタやディスプレイは1台でしかも通常のものでよいのが利点で、特に大画面に適している。また、カラー画像を全天周型スクリーンに投影する場合にも、時分割方式が採用される。さらに最近では、テレビやパソコンで立体映像を見る場合にも使われるようになり、シャッタースピードが高速化されて、フリッカのない立体映像を鑑賞することができるようになっている。

分光視差方式

プロジェクタおよびメガネに使用されるフィルタに、RGBのそれぞれを前後の波長に分けるフィルタを用いて、左右画像を分離する方式である。これにより、従来のアナグリフ方式では不可能であったカラーの立体映像を見ることが可能である。投影用のプロジェクタは1台で時分割で交互に左画像と右画像をあらゆるスクリーンに投影でき、ちらつきやクロストークがない立体映像を見ることが可能である。また、ドーム型スクリーンにも投影可能である。この方式では、プロジェクタにフィルタを内蔵し、時分割で表示する必要があるため、特別なプロジェクタを用意する必要がある。

2.7.2 裸眼立体視方式

特別な装置を装着する必要の無い裸眼立体視方式として、パララックスバリア方式、レンチキュラ方式、インテグラルフォトグラフィを取り上げる。

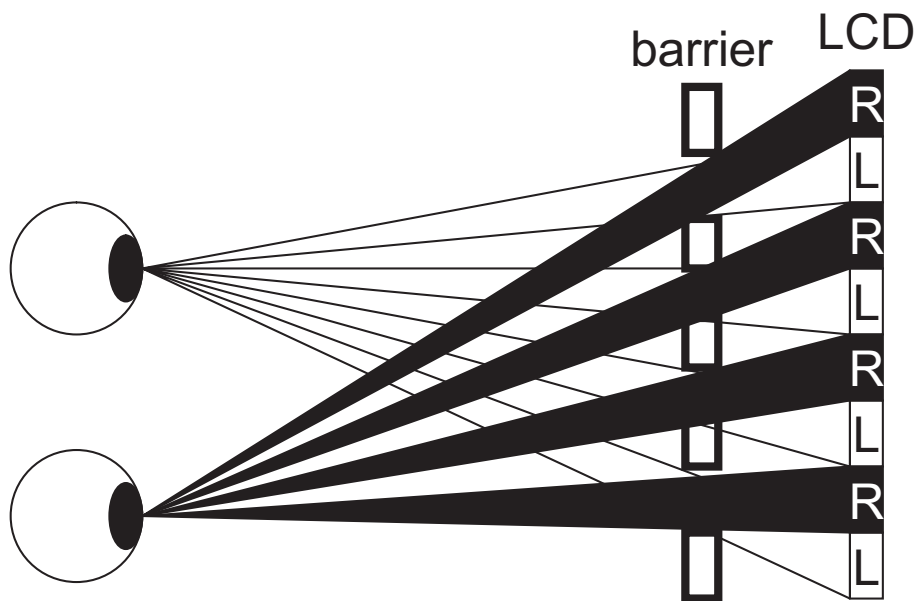


図 2-6: パララックスバリア方式

パララックスバリア方式

目の位置が動いてもステレオ対が提示できるようにしたものがパララックスバリア方式である。ステレオ対は縦長に分割して交互に並べられ、その手前に同じ周期のスリット（縦縞状に開口部のあるアパーチャ）すなわちパララックスバリアを置くという構造になっている。この様子を図 2-6 に示す。これを適当な距離から見るとバリアに遮られて、左右の目は別々の画像を提示されるので立体視が生じる。構造が簡単なため、現在も広告などによく使われている。ステレオ対は 2 枚に限ることはなく、順々に視点を変えた複数の画像を同じように分割して並べ、スリットの幅をそれに応じて狭くすれば、目の移動に伴って順々に異なる立体画像を見ることができる。更にこの方式を発展させ、2 次元の格子状のバリアを設けて、左右方向のみならず、上下方向にも対応した立体像を得ることができる。

最近開発された製品は、表示部分に液晶板を 2 枚重ねて、観察側にバリアを形成する液晶板を置く。この方式の利点は、同一モニタで 2D と 3D を自由に切り替えられる点にある。また、画素毎に 1 ピッチずつずらして、モアレや画像の逆転現象を軽減することができる方式も開発されている。

レンチキュラ方式

パララックスバリア方式の欠点は、バリアで光が遮られるため像が暗くなるということである。これは複数枚になると著しく、バックライトによる照明などが必要となる。この欠点に対し、バリアの代わりにかまぼこ型のレンズを並べた板、すなわちレンチキュラ板を用いたレンチキュラ方式である。視点によってレンズの焦点位置がずれていくので、それに対応した分割画像を用意すればよい。複数枚の画像を写し込むことも可能である。現在はプラス

チック加工技術の進歩により、安価で良質なレンチキュラ板が容易に入手できるようになり、画像処理との組み合わせで、広く実用されている。

最近、裸眼立体視用として脚光を浴びるようになり、パララックスバリア方式と同様に液晶やプラズマディスプレイの表示装置に使われるようになってきた。また、視点追従型にして、画像の反転などの現象を減らす工夫や、できるだけ大勢が観察できるように画面数を増やしたり、クロストークを軽減するためにダブルレンチキュラスクリーンを使うなど、様々な工夫がなされてきている。

インテグラルフォトグラフィ

インテグラルフォトグラフィは、レンチキュラ方式と似ており、かまぼこ型レンズのレンチキュラ板の代わりに、微小な円形レンズを敷き詰めたレンズ板を使用する点が異なる。原理はレンチキュラ方式と同じだが、円形レンズにより上下の視差も取り込める。逆に、レンチキュラ方式はインテグラルフォトグラフィで、上下の視差を省略したものといえる。以前は、インテグラルフォトグラフィに使用される微小な円形レンズを製造することが困難であったために実用に至らなかったが、最近になって精密加工技術が進歩したことによって、インテグラルフォトグラフィも見直されつつある。

2.8 むすび

本章では、多視点映像の圧縮伝送を考えた場合に必要とされる要素について述べた。まず、多視点画像を扱う上で全ての要素の基礎となっている多視点幾何について説明した。次に、多視点映像を圧縮伝送する場合の要素として取得、補正、符号化、画像ベースのレンダリング、表示を挙げ、それぞれについて具体例を交えながら概観した。

第3章

多視点映像符号化の予備実験

3.1 まえがき

2次元映像符号化に対する多視点映像符号化の違いとは、時間方向の相関だけでなく、視点間の相関も利用可能な点であり、如何に視点間の相関を利用するかが課題である。よって本章では、視点間相関が時間相関に対しどの程度有効なのかを調査することを目的とし、予備実験を行う。まず、時間相関と視点相関を比較する。次に視点間予測のみの場合において、符号化における予測構造に関して検討を行い、H.264/AVCの重み付け予測を用いてカメラ間の個体差、照度差を補正して符号化した場合についても検討する。最後に、視点間予測と時間予測の両方を適応的に用いる手法を従来手法に位置づけ、どのような状況下で視点間予測が有効かについて調査する。

3.2 テストシーケンス

本論文において使用する多視点映像テストシーケンスは、現在、JVTにおいて検討が進められているMVCのテストシーケンスのうち、6種類のシーケンスを使用する。これら6つのシーケンスの画像を図3-1に示す。これらのシーケンスは全てカメラキャリブレーションが行われており、回転行列 R と並進ベクトル t 、内部行列 K が算出されている。以下、シーケンスの内容について説明する。

Akko&Kayo 画像サイズ 640×480 [pixel] で、フレームレート 30 [fps] のシーケンスである。映像はカメラは5列 \times 3行の二次元の格子に固定に配置されて取得されたが、本論文では、上段の一段のみを利用する。すなわち水平方向一次元直線配置である。rectification は行われてないが、得られているカメラパラメータは回転行列、内部行列が全カメラで等しい。並進ベクトルは、カメラ間で x 方向の変位があるのみである。背景がカーテンで仕切られた室内で撮影したものである。カメラとカーテンの間を二人の女性が物を持って歩き回るといった内容である。

Rena 画像サイズ 640×480 [pixel] で、フレームレート 30 [fps] のシーケンスである。カメラは水平方向一次元直線に固定配置されている。rectification が行われており各カメラ間で

回転行列は等しいが、内部行列における光学中心の画像内の x 座標 (式 (2.4) における u_0) が異なっている。並進ベクトルは、カメラ間で x 方向の変位があるのみである。背景がカーテンで仕切られた室内で撮影されたものである。カメラとカーテンの間を一人の女子がバトンを使って踊るを行うという内容である。

Ballroom 画像サイズ 640×480 [pixel] で、フレームレート 25 [fps] のシーケンスである。カメラは水平方向一次元直線に固定配置されている。rectification が行われており、各カメラ間の回転行列と内部行列はすべて等しい値になっている。並進ベクトルは x 成分のみ値を持っている。ダンスルームで複数の人が社交ダンスをしており、各ダンスペアが入れ代わり立ち代り撮影されている。

Race1 画像サイズ 640×480 [pixel] で、フレームレート 30 [fps] のシーケンスである。カメラは水平方向一次元直線に配置されている。rectification は行われていないため回転行列はカメラ間で異なるが、内部行列は等しい。並進ベクトルは x 成分のみ値をもっている。サーキット場で撮影されたシーケンスであり、レーシングカーが走っている様子をカメラが追跡して撮影している。カメラが追跡しているため、カメラは時間方向で動いている。

Breakdancers 画像サイズは 1024×768 [pixel] で、フレームレート 15 [fps] のシーケンスである。カメラは水平方向一次元アーク状に固定配置されている。アーク状配置なので rectification は行われておらず、回転行列および内部行列がカメラ間で異なり、並進行列もカメラ間でそれぞれの成分が変位を持っている。中心のダンサーが激しくブレイクダンスを踊り、周囲の人がそのダンスに乗りながら観ているという内容である。

Ballet 画像サイズは 1024×768 [pixel] で、フレームレート 15 [fps] のシーケンスである。カメラは水平方向一次元アーク状に固定配置されている。カメラパラメータの様子は Breakdancers と同じである。奥行き感のある一室に人が二人存在し、片方の女性がバレエを行い、もう片方の男性がそれを観ているという内容である。動きの少ないシーケンスである。

3.3 時間相関と視点相関

時間方向の相関と視点間の相関を比較するために、まず隣接フレーム間の差分を測り、比較する。また、時間方向予測のみと視点間予測のみを行って符号化を行い違いを比較する。

3.3.1 隣接フレーム間の相関

時間相関と視点相関の違いを比べるために、時間方向の隣接フレーム間または視点方向の隣接フレーム間で以下の式によって算出される MAD (Mean Absolute Difference: 平均絶対値誤差) を測った結果を表 3-1 に示す。



(a) AkkoKayo



(b) Rena



(c) Ballroom



(d) Race1



(e) Breakdancers



(f) Ballet

図 3-1: 使用するテストシーケンス

表 3-1: 視点方向又は時間方向の MAD

sequence	time	view
Akko&Kayo	5.495732	22.576077
Ballet	2.842927	14.427687
Ballroom	4.442435	23.789694
Breakdancers	7.606853	11.125984
Race1	11.538867	15.236266
Rena	2.375622	5.002139

表 3-2: 符号化条件 (JM)

JM version	11.0
QP (fix)	28, 32, 36, 40 (QPI=QPP=QPB)
Search range	± 128 [pixel]
Entropy coding	CABAC

$$MAD = \frac{\sum_{x=0}^{S-1} \sum_{y=0}^{T-1} |I_1(x, y) - I_2(x, y)|}{S \times T} \quad (3.1)$$

I_1, I_2 は時間方向または視点方向の隣接画像の画像を表しており, (x, y) はフレーム内画素位置, S, T はフレームサイズである.

表 3-1 から明らかなように, 時間方向の MAD の方が視点間の MAD よりも 2 倍から 10 倍近く低いということである. すなわち単純差分においては, 時間方向の画像の方が視点方向の画像よりもより似通っていることを示している. この傾向のカメラ間隔やフレームレートによる違いは特に見受けられなかった.

3.3.2 片方だけの相関を用いた符号化

視点間相関のみ及び時間相関のみを用いて符号化した場合のレート歪み特性について調査を行う. 具体的には図 3-2 に示すような隣接フレームを参照画像として予測を行う構造を用いて符号化を行う. 符号化には H.264/AVC の参照ソフトウェアである JM (Joint Model) ⁽⁵⁴⁾ を使用した. 符号化条件を表 3-2 に示す. レート制御は行わず, I, P, B ピクチャの QP (量子化パラメータ) は同じにしている.

実験結果を図 3-3 に示す. "time correlation" は図 3-2(a) の構造を, "view correlation" は図 3-2(b) の構造を, それぞれ用いて符号化した場合である. 横軸は全フレームの平均発生符号量, 縦軸は平均 SNR である.

図 3-3 からわかることとして, 時間方向の相関のみを用いて符号化した方が常にレート歪み特性が良いということである. これは, 先に行った隣接フレーム間の差分の結果と同じで

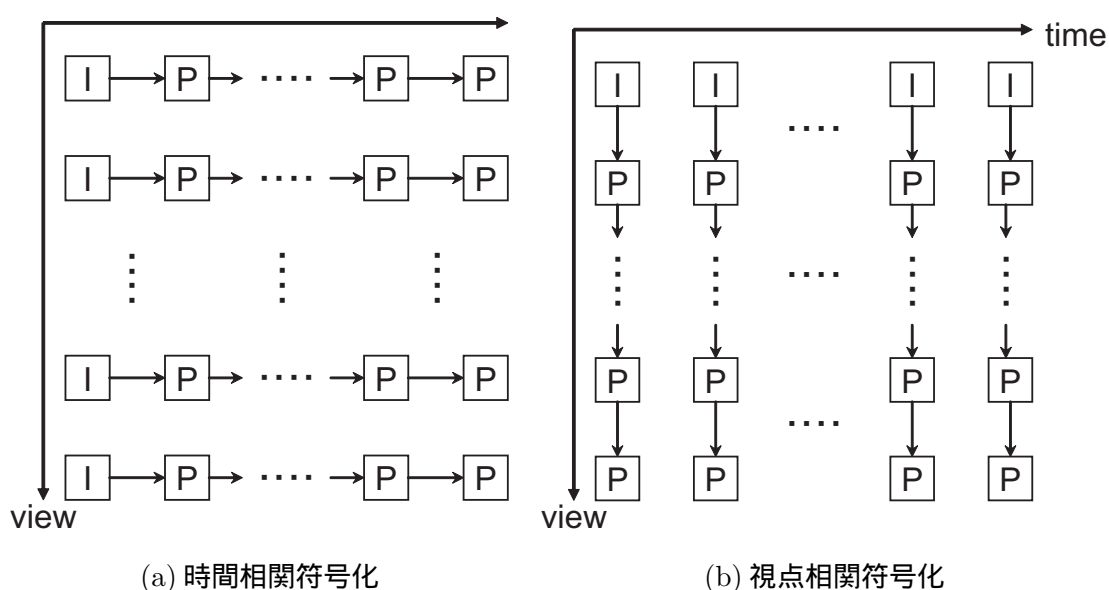
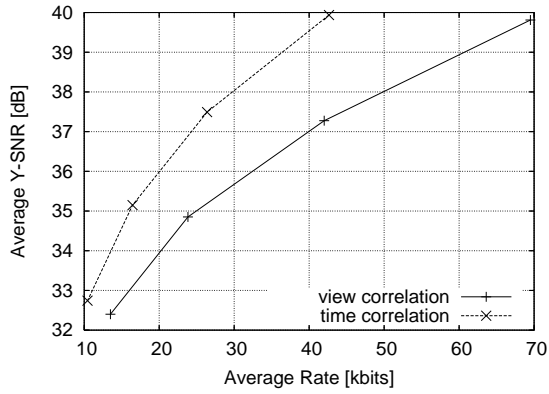


図 3-2: 視点間相関符号化と時間相関符号化の予測構造

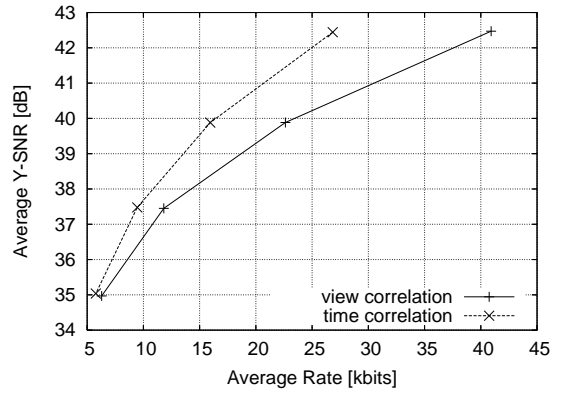
あり、違いは動き（視差）補償処理によって補償するか否かであるが、例えば動き（視差）補償処理を行ったとしても、時間方向の相関の優位性は非常に高いことを示している。どのシーケンスにおいても時間相関の優位性が高いことから、この傾向はカメラ間隔やシーケンスの内容などに依存せず、優位性が成り立つことがわかる。この原因として、視点間予測のみの場合には、カメラ間隔に起因する大きな視差が発生し、これを符号化するためにより多くのビットが発生する。また、カメラ間には個体差、照度差があり、同じシーンを撮影したとしても同じ画像値が得られるとは限らない。たとえ正しい視差ベクトルが算出されたとしても、値の大きな予測誤差信号が発生し、結果としてレート歪み特性を悪化させるからである。時間相関のみを使用する場合には、撮影シーンで大幅な輝度変化などがない限りは、この問題は発生しない。また、高ビットレート時により違いが表れているのは、高レート時には歪みが重視されるが、視点間予測のみでは時間方向予測ほど符号化対象フレームに対する参照フレームにおける共通領域が存在しなかったからである。低レート時の方が差が少なくなっているのは、歪みよりも発生符号量が重視され、視差ベクトルの値が大きいといえども、視差ベクトルも動きベクトルと同様に相関があり、H.264/AVC のベクトル予測の枠組みで、これら相関を利用することができたからであると考えられる。

3.4 視点間予測による符号化

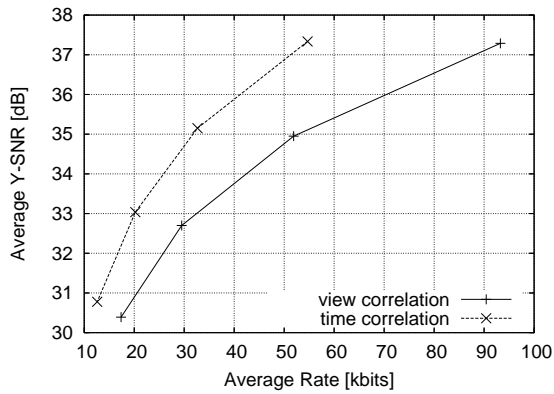
従来の2次元映像は時間順に符号化するより他なかったが、多視点映像になると、カメラ間でどのような参照関係にすべきか、という自由度が発生することとなる。そこで、本節では、Iピクチャ、Pピクチャ、Bピクチャをどのカメラに割り当てることで、符号化効率がどのように変化するかについて調査する。また、カメラ間の個体差や撮影シーンの照明環境などにより生じる信号値の違いを補正する一手法としてH.264/AVCの重み付け予測を適用



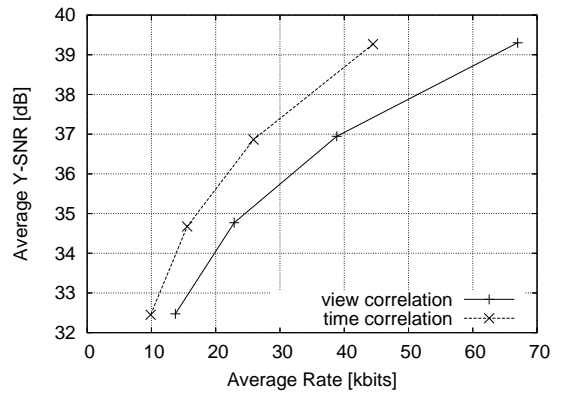
(a) Akko&Kayo



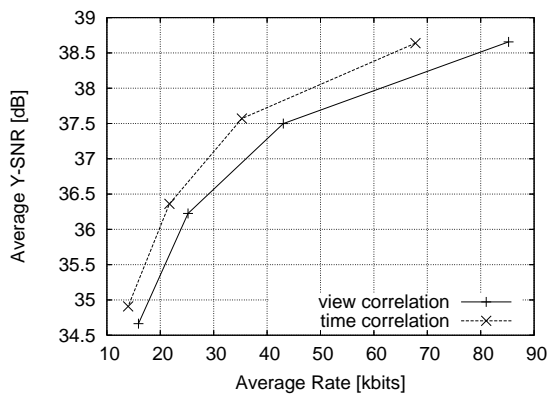
(b) Rena



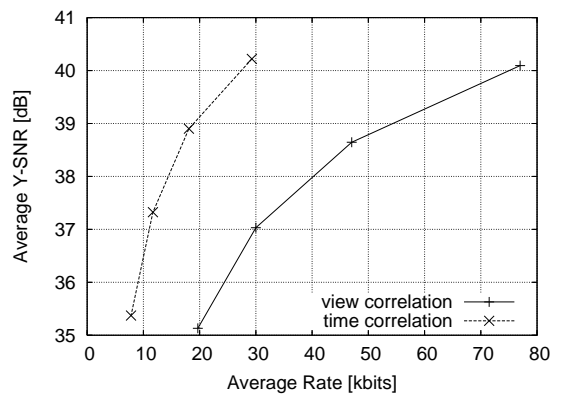
(c) Ballroom



(d) Race1



(e) Breakdancers



(f) Ballet

図 3-3: 視点間予測のみと時間予測のみ

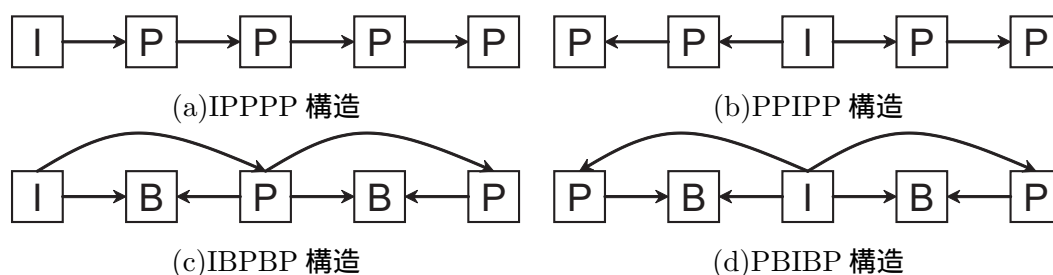


図 3-4: 視点間予測の構造

し、効果について調査する。

3.4.1 予測構造に関する検討

本論文において、検討しているカメラは、水平方向に設置された5台であり、Bピクチャの有無及びIピクチャの位置の違いから、図3-4のようなIPPPP、IBPBP、PPIPP、PBIBPという構造が考えられる。Pピクチャの参照フレームは1台のカメラであり、Bピクチャは両脇のIピクチャ又はPピクチャのカメラから双方向で予測されるという構造である。

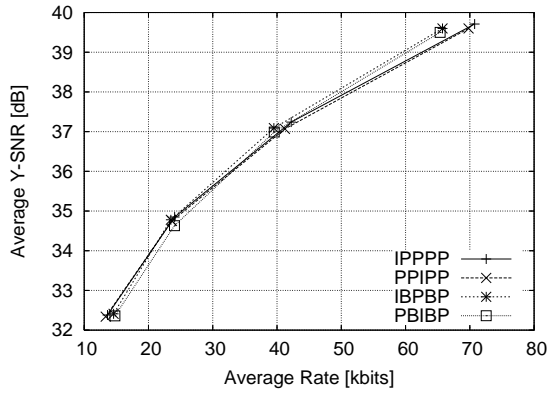
本実験においてもJMを使って符号化を行った。実験条件は表3-2のとおりである。実験結果を図3-5に示す。

結果からわかることとして、中心にIピクチャを置く構造(PPIPPまたはPBIBP)がレート歪み特性が良くなっていることが多いということである。Iピクチャはイントラ符号化され、同じ量子化パラメータで量子化すると、必然的に発生符号量はインターピクチャよりも大きくなる一方、画質が向上する。よって、画質が高いピクチャがより多くのピクチャから参照されることで、全体のレート歪み特性が上がるからであると考えられる。

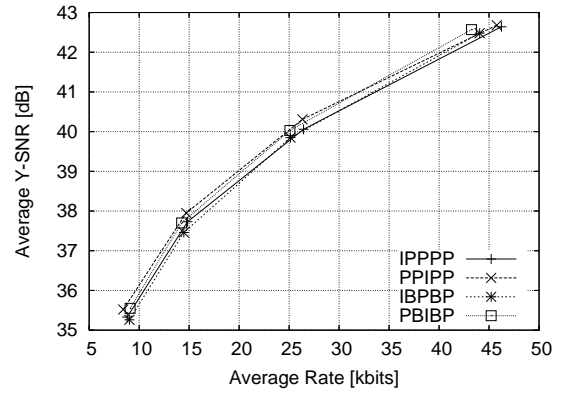
その他わかることとして、カメラ間の距離が小さいAkko&KayoとRenaにおいてはBピクチャを含めることで全体のレート歪み特性が上がる場合があるが、その他のカメラ間の距離が大きいシーケンスでは、Bピクチャを含めない予測構造の方が符号化効率が高くなる場合があるということである。これは、Bピクチャを含める構造の場合、Pピクチャ符号化時の参照フレーム間隔が大きなものとなり、参照フレームとの共通領域が小さくなったり視差ベクトルの符号量が大きくなることでより多くのイントラ符号化が選ばれることとなり、Pピクチャの符号化効率が低下するからである。Bピクチャは双方向予測により符号化効率が向上するが、それ以上にPピクチャの影響が大きいからであるといえる。

3.4.2 重み付け予測の効果

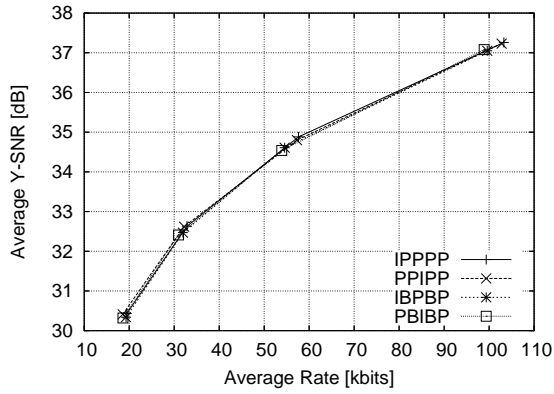
先に視点間予測と時間予測の符号化効率の違いにカメラの個体差を挙げたが、他にも被写体の反射特性、撮影シーンの照明環境などにより、同じシーンを撮影したとしても各カメラで得られる信号値は異なってくる。これら効果を補正する式の一次近似は一般に以下の式で与えられる⁽⁵⁶⁾。



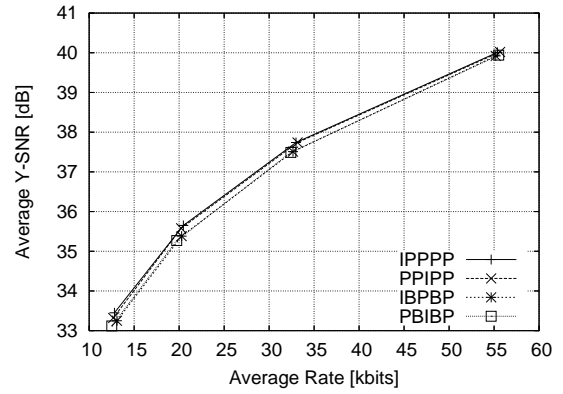
(a) Akko&Kayo



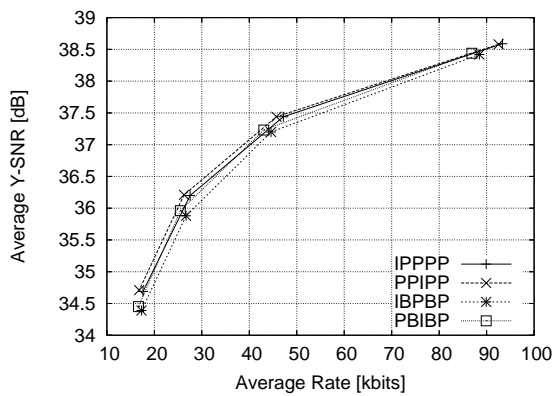
(b) Rena



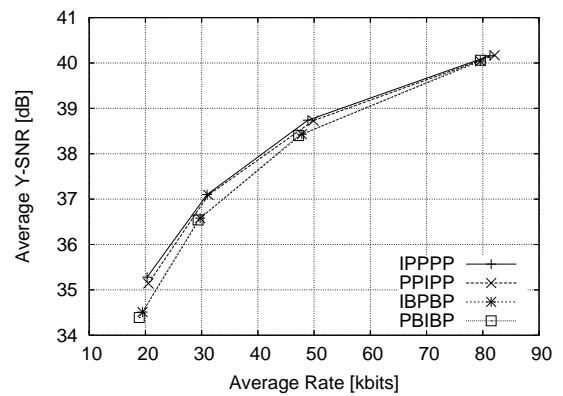
(c) Ballroom



(d) Race1



(e) Breakdancers



(f) Ballet

図 3-5: 予測構造の違い

$$y = a \cdot x + b \quad (3.2)$$

カメラ間の個体差を補償する一手法として、H.264/AVCにおける重み付け予測が有効であることが知られている⁽³³⁾。H.264/AVCにおける重み付け予測は以下の式で表される。

$$I_{wref}(x, y, t) = a \cdot I_{ref}(x + d_x, y + d_y, t - s) + b \quad (3.3)$$

$I_{wref}(x, y, t)$ が重み付け後の画像、 $I_{ref}(x, y, t)$ は既に復号された画像であり、 (d_x, d_y) は変位ベクトル、 s は時間間隔を表している。H.264/AVCのデフォルトの方法では a のみが求められ、 b は算出されない*。よって、係数 a 、 b は以下の式で算出される値を使用し、デコーダ側でも同じ予測画像を生成するために、パラメータ a 、 b は符号化されて伝送される。

$$a = \frac{\sum_{(x,y) \in F} I_{cur}(x, y, t)}{\sum_{(x,y) \in F} I_{ref}(x, y, t - s)} \quad (3.4)$$

$$b = 0 \quad (3.5)$$

F は画像を表している。今回も符号化条件は表3-2のとおりで、JMを使って符号化を行った。参照構造はIPPPPとした。

図3-6からわかる通り、重み付け予測を行うことで基本的には符号化効率が向上することがわかる。しかし、Race1とAkko&Kayoシーケンスでは、符号化効率が低下する結果となった。これらシーケンスには、式(3.4)のみでは補正されない誤差要因が含まれていたからであると考えられる。実験結果として、必ずしも全てのシーケンスにおいて、H.264/AVCデフォルトの重み付け予測が有効であるわけではないが、多視点符号化において符号化効率を向上させる可能性があることがわかった。

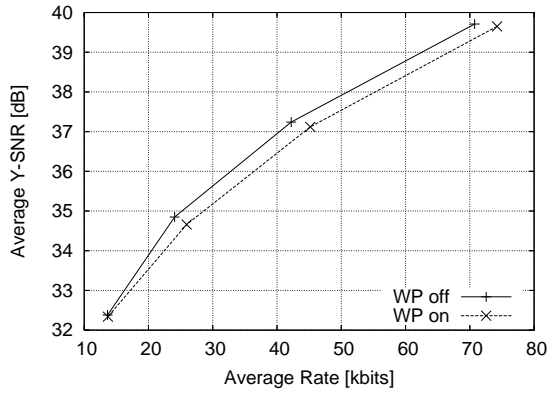
3.5 サイマルキャスト符号化と適応予測符号化

3.3節により、時間方向と視点方向どちらか一方を使用する場合は、時間方向の相関が高いことを確認した。本節では、時間方向予測と視点間予測の両方が使用可能で適応的に使用する場合、どの程度視点間予測が有効かについて実験を行う。

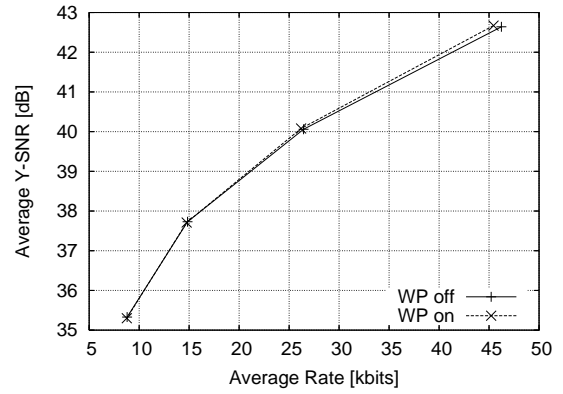
3.5.1 使用する予測構造及び符号化条件

視点・時間適応予測符号化の効率を調べるために使用する予測構造としては、現在MPEGにおいて評価実験が行われている図3-7の構造を用いた⁽⁵³⁾。この構造は、視点間予測の構造は先に述べたIBBP構造であり、時間予測の構造は、時間スケラビリティを実現する階層的Bピクチャ構造である。時間予測には片方向予測は含まれておらず、GOP内のBピクチャは同じGOP内のIピクチャまたはBピクチャから階層的に予測される。よって、各

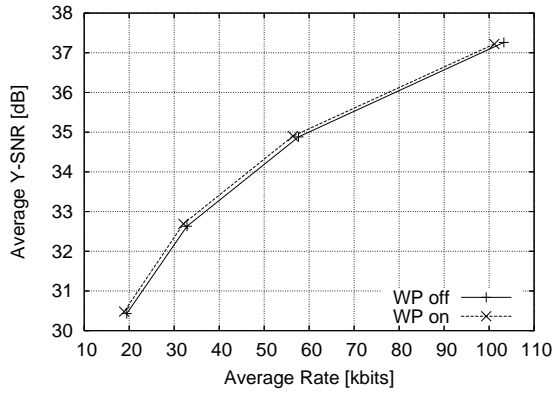
* b の算出方法は照度補償として現在JVTにおいて検討がなされている。



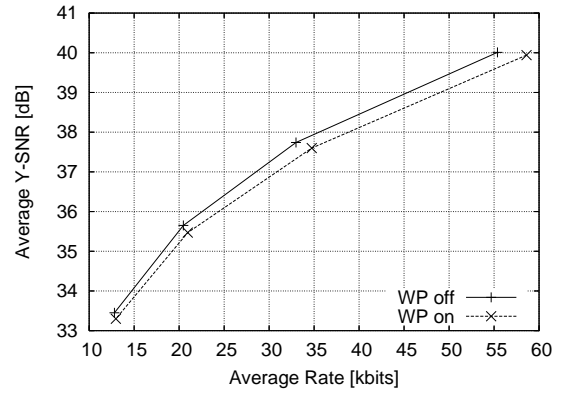
(a) Akko&Kayo



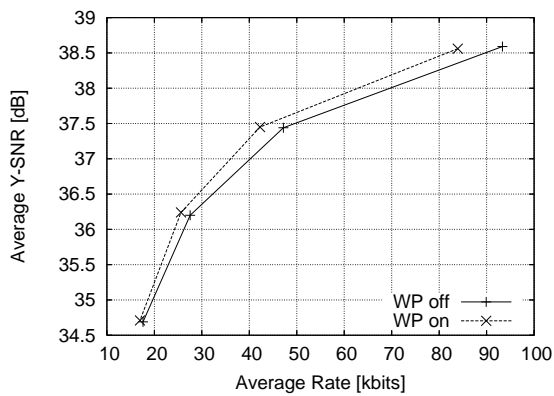
(b) Rena



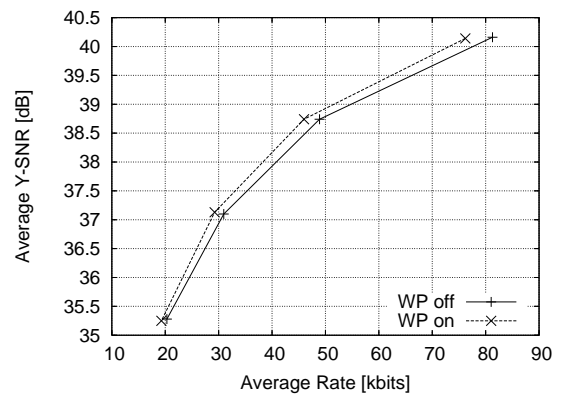
(c) Ballroom



(d) Race1



(e) Breakdancers



(f) Ballet

図 3-6: 重み付け予測の効果

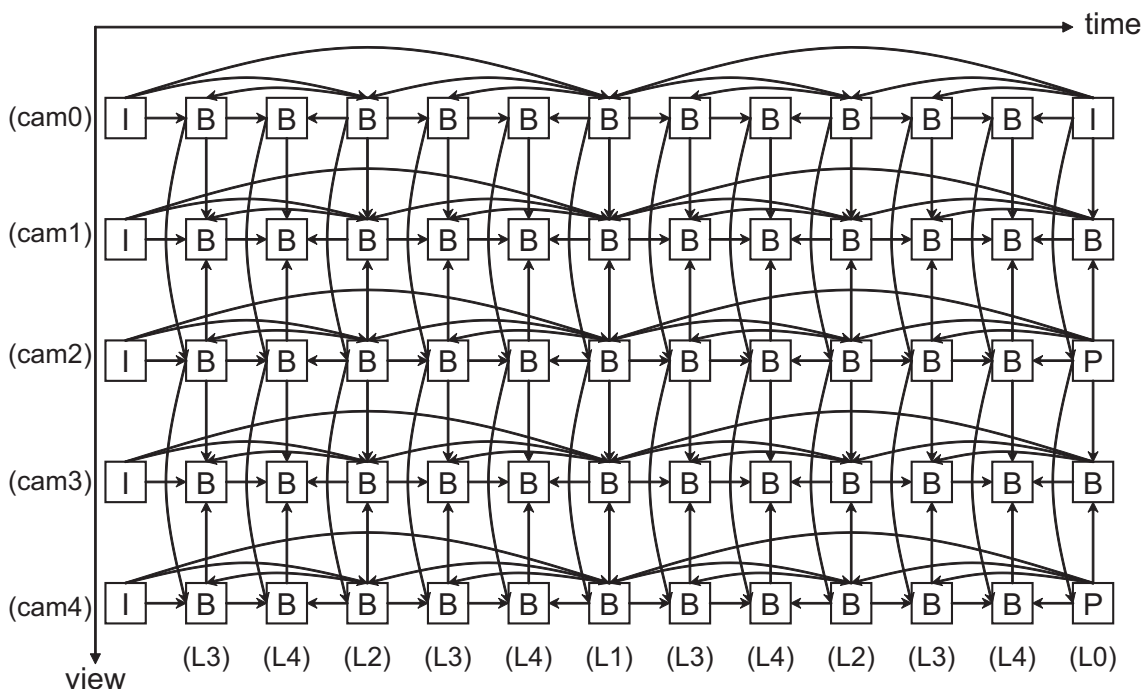


図 3-7: 時間・視点適応予測符号化の予測構造 (GOP=12)

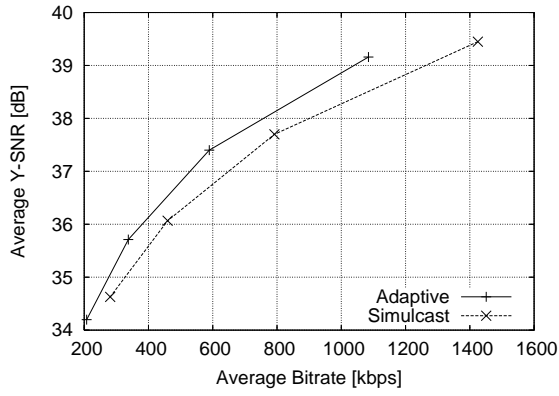
表 3-3: 符号化条件 (JMVM)

JM version	11.0
BaseQP	22, 27, 32, 37
Search range	± 96 [pixel]
Entropy coding	CABAC

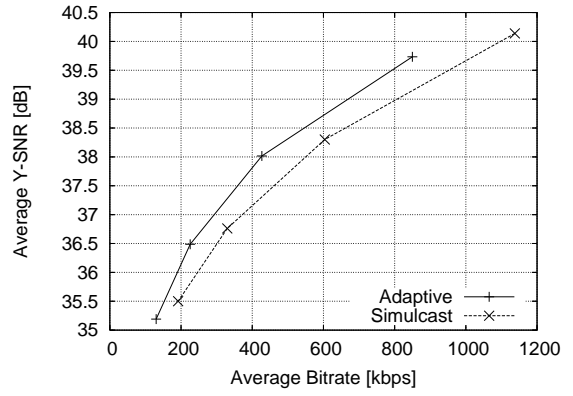
時刻において、参照フレーム間の時間的な間隔が、異なるという特徴になっている。参照ソフトウェアとしては、JVT において配布されている参照ソフトウェア JMVM⁽⁵⁵⁾ を使用した。符号化条件を表 3-3 に示す。“BaseQP”とは、I ピクチャおよび P ピクチャ符号化時に使用される QP 値である。図 3-7 の下部に記載されている“(L0)” ~“(L4)”は時間方向のレベルを表しており、階層が深くなるにつれて大きくなる。このレベルが大きくなるにつれより大きな QP を用いて符号化を行う。

3.5.2 全体での比較

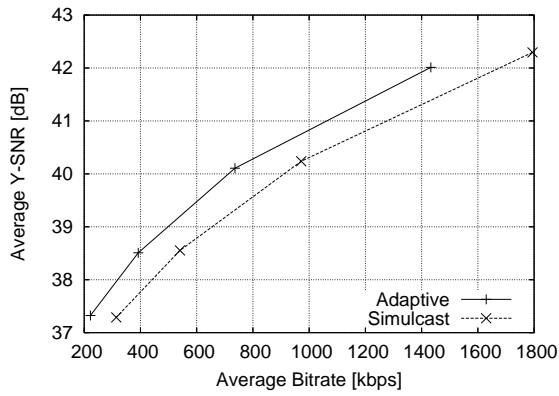
全フレームでの平均した結果を図 3-8 に示す。横軸は、各視点におけるビットレートの平均値、縦軸は平均 SNR である。図からわかる通り、視点間予測を加えることで全体の性能は向上し、向上度合いは 0.5 ~ 1.0[dB] 程度である。図 3-8 の結果は、シーケンスに関係なくほぼ一定の利得が出ていることがわかる。



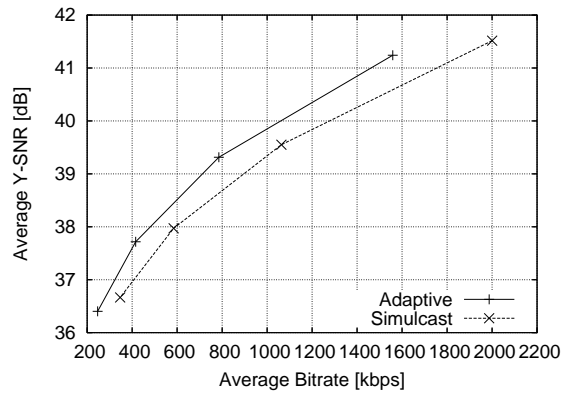
(a) AkkoKayo



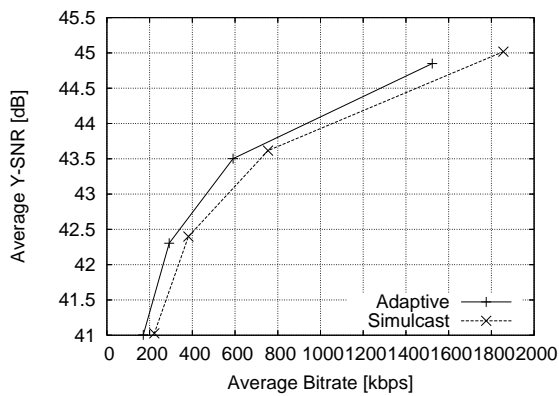
(b) Rena



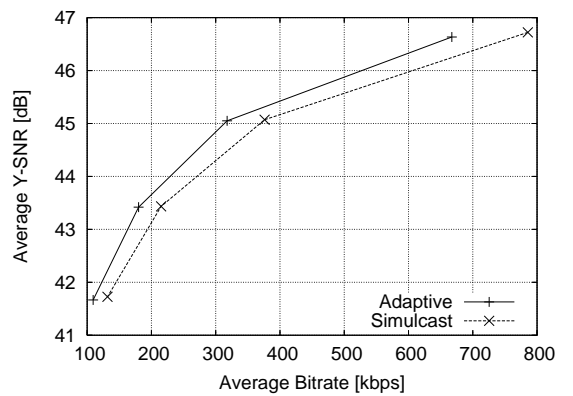
(c) Ballroom



(d) Race1



(e) Breakdancers



(f) Ballet

図 3-8: 時間・視点適応予測符号化とサイマルキャスト符号化

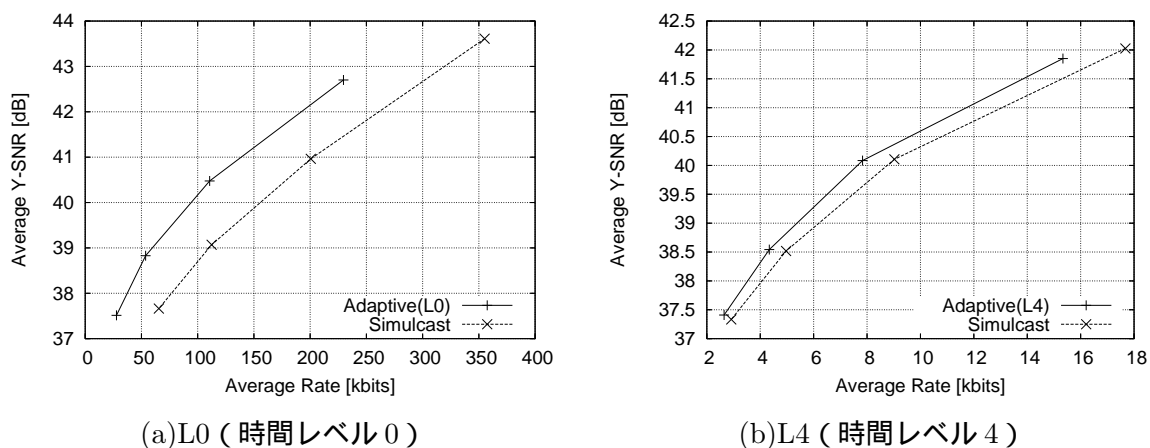


図 3-9: 時間間隔の違いによる視点間予測の効果 (Ballroom)

3.5.3 時間間隔の違いによる比較

時間予測に階層的 B ピクチャを使っているため、時間位置によって時間方向の相関は変化するが、視点間予測の相関は大きく変化しない。そこで、階層的 B ピクチャの時間レベルに対する視点間予測の効果について検討する。図 3-9 に、Ballroom での図 3-7 における L0($t=12$) の場合と L4($t=8$) の場合のレート歪み特性を示す。横軸は各時刻における全視点画像の平均発生符号量、縦軸は平均 SNR である。L0 は時間予測を行わない視点間予測のみを行うフレームであり、L4 は時間方向で直前・直後のフレームによる時間予測及び視点間予測を行うフレームである。L0 における利得は約 2[dB] であり、L4 における利得は約 0.5[dB] であり、L0 における利得が大きいことがわかる。図 3-8 における平均的な利得は約 1[dB] であることから、全体の利得は時間間隔の大きなフレームによる効果であることがわかる。すなわち、時間予測と視点間予測の両者が使用可能な場合では、時間間隔が大きいフレームにおいて視点間予測が有効となり、特に、時間予測を行わないアンカーフレーム (L0) による利得が大きいことがわかる。図 3-9 は Ballroom の結果であるが、その他シーケンスでも同様の傾向が見受けられた。また、図 3-7 の構造からもわかるとおり、L1 は L0 参照フレームを、L2 は L0 または L1 参照フレームを、以下同様に参照するため、より上位のフレームにおいて符号化効率が向上することで全体の符号化効率も向上することがわかる。

3.6 むすび

本章では、時間・視点適応予測符号化の性能評価を行った。まず、視点間予測の特性について調べるために、時間方向、視点方向の隣接フレーム間の差分について調査し、時間方向の相関が視点方向の相関よりも高いことがわかった。この関係が符号化においてもそのまま成り立つかどうか、片方向予測のみを用いて符号化を行い、時間相関を使用する方がレート歪み特性が良いことがわかった。次に、視点間符号化における予測構造について検討を行い、I ピクチャをカメラ列の中心に配置することで、端に配置するよりもレート歪み特性が良く

なることがわかった。また、カメラ間の信号値の違いを符号化において補正する一手法としてH.264/AVCの重み付け予測を取り上げ、重み付け予測を適用することでレート歪み特性が改善されるシーケンスがあることがわかった。最後に、時間予測と視点予測を適応的に選択して用いる符号化と時間相関のみを使うサイマルキャスト符号化による比較を行い、視点間予測を行うことで全体のレート歪み特性は向上し、特に視点間予測のみを行うアンカーフレームにおいて利得が高くなることがわかった。

第4章

View Interpolationを用いた多視点映像符号化方式

4.1 まえがき

本章では View Interpolation を用いた多視点映像符号化について述べる。View Interpolation とは、2章で述べたように、得られているカメラ画像を用いて仮想的な視野画像を生成する技術のことを言う。View Interpolation を用いた多視点映像符号化は JVT においても盛んに議論されている分野であり、視点間予測において従来の2次元映像符号化では使用されないカメラパラメータを使用する点が特徴的である。3章の結果から、時間方向予測と視点間予測を適応的に選択する符号化方式においては、時間方向予測が選ばれる傾向にあるため、提案手法は、視点間予測の符号化効率向上を目指すこととする。提案手法は既に提案されている View Interpolation を利用した奥行き情報を符号化して送信する枠組みにおいて符号化効率の改善を目指すため、まずこの基本とする符号化方式について説明する。次に奥行き情報の特性を利用した符号化効率改善案について提案する。

4.2 基本とする符号化方式

提案手法は、文献⁽³⁴⁾の方式を基本としている。この手法は、H.264/AVC をベースに、奥行き情報をエンコーダで生成してエントロピー符号化を行うという枠組みであり、図4-1に示す。図中、太線で示してあるブロックが H.264/AVC の枠組みに付加された奥行き推定・補償の部分である。奥行き推定は以下の式で算出される。

$$z = \arg \min_z \sum_{u=a}^{a+S-1} \sum_{v=b}^{b+T-1} \{I_{cur}(u, v) - I_{ref}(u', v', z)\}^2 \quad (4.1)$$

$I_{cur}(u, v)$ は符号化対象フレームの画像位置 (u, v) における画素値、 $I_{ref}(u', v', z)$ は符号化対象カメラ座標 (u, v) を奥行き z を用いて式(2.10)による Warming を行って得られる、参照フレームの画像位置 (u', v') における画素値である。 S, T は符号化対象ブロックのブロックサイズ、 a, b はブロック内の開始画素位置である。 z を $minZ$ から $maxZ$ まで $stepZ$ ずつ変化させ、得られる画像座標位置間で SSD (Sum of Square Difference) を最小化する z を符号化対象ブロックの奥行き情報とする。奥行き情報を算出した後更に、カメラパラメータを

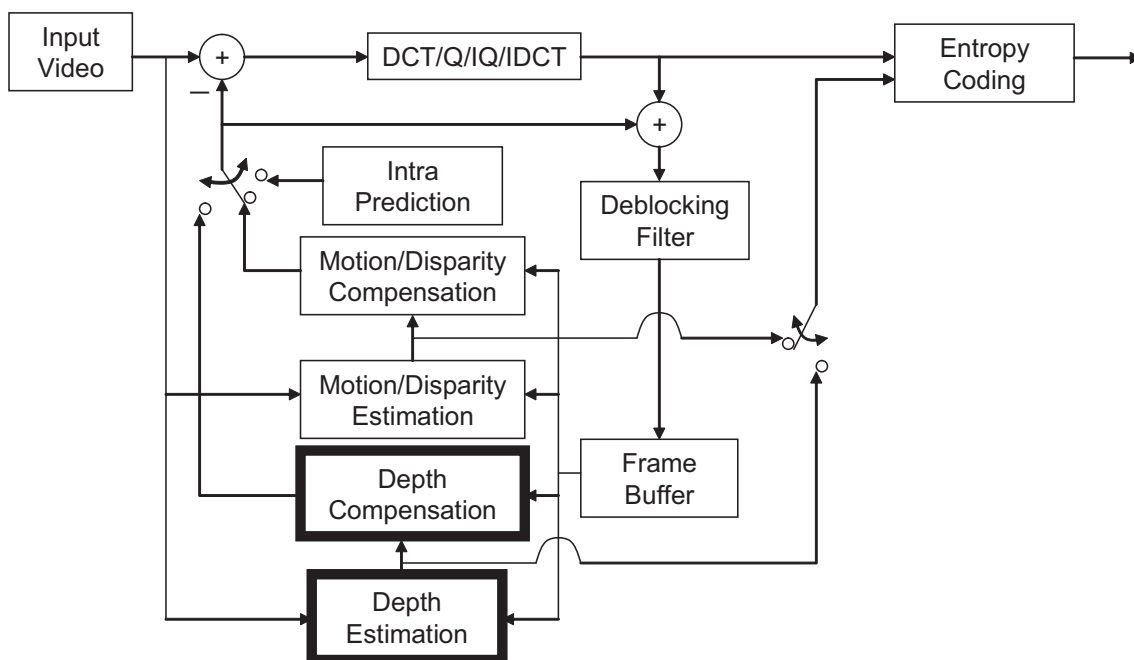


図 4-1: H.264/AVC に奥行き情報探索を付加した符号化器

補正する目的で、補正ベクトルを探索する。これは、式(4.1)において、 z を使って符号化対象カメラの座標を $(u - c_x, v - c_y)$ として (u', v') を求め、この (u', v') と (u, v) 間でコストを最小化する補正ベクトルを求める。補正ベクトルの探索範囲は $1/4$ 画素精度で $\pm 2[\text{pixel}]$ で良いとされている。得られた奥行き情報 z と補正ベクトル (c_x, c_y) を符号化する。

視差ベクトルを、従来の動き補償の枠組みを視差方向に適用して得られる2次元の変位量として定義すると、奥行き情報は1次元の値であり次元数が小さくなるため符号化効率を向上させる可能性がある。また、奥行き情報と視差ベクトルは、画素毎に見れば同じ情報であるが、MPEG エンコーダのようなブロック毎に情報を求めていく場合、ブロック毎という制約条件が科せられるため、ブロックで1つ奥行き情報を符号化するか視差ベクトルを符号化するかで違いが発生する。

4.3 提案手法

提案方式は文献⁽³⁴⁾で提案されている枠組みにおける符号化効率改善を目指すものである。すなわち、奥行き情報を生成して予測画像生成に使用し、奥行き情報を符号化するという枠組みの符号化効率の改善を目指すものである。この手法に対する改善案として以下の4点を提案する。

- 奥行き情報の探索コスト
- 奥行き情報と視差ベクトルの相互変換による予測効率の向上
- 奥行きの一貫性の利用

- スキップ/ダイレクトモードの改良

4.3.1 レート歪みに基づく奥行き情報探索

文献⁽³⁴⁾では、奥行き探索に用いる指標として式(4.1)のように、SSD (Sum of Square Difference) を用いている。しかし、各奥行き情報間で SSD が僅差であるが符号化すべき奥行き情報の値が大きく異なる場合、発生符号量を無視して奥行き情報を算出することによって、全体としてのレート歪み特性が悪くなる場合があるということが考えられる。よって、レート歪みコストをベースとした探索のほうが、特に低ビットレート時にレート歪み特性が向上することが予想される。具体的には、以下の式で表されるコストに基づき奥行き情報を探索する。

$$J = SAD + \lambda_{depth} \times R_{depth} \quad (4.2)$$

$$SAD = \sum_{u=a}^{a+S-1} \sum_{v=b}^{b+T-1} |I_{cur}(u, v) - I_{ref}(u', v', z)| \quad (4.3)$$

$$\lambda_{depth} = \sqrt{\lambda_{mode}} = \begin{cases} \sqrt{0.85 \times 2^{\frac{(QP-12)}{3}}} & \text{(only P picture)} \\ \sqrt{0.65 \times 2^{\frac{(QP-12)}{3}}} & \text{(B picture included)} \end{cases} \quad (4.4)$$

I_{cur} , I_{ref} , (u, v) , (u', v', z) , S , T , a , b は式(4.1)の定義と同じである。 R_{depth} は奥行き情報を符号化した場合の発生符号量であり、算出方法は以下で述べる。ラグランジュの未定乗数 λ_{depth} は、動きベクトル探索時と同じ λ_{mode} の平方根を用いる。 λ_{mode} は符号化モード決定時の実際に符号化したときの歪み (SSD) と発生符号量からマクロブロック毎に最適なモードを選択するときを使用される値であり、経験的に得られた値である⁽⁵⁷⁾。Bピクチャの有無で微妙に値が異なる。16×16 から 4×4 ブロック毎にコスト J を算出し、 J を最小化する時の奥行き情報をそのブロックにおける奥行き情報として決定する。

R_{depth} は奥行き情報の予測値と符号化対象ブロックの奥行きとの差分から得られる発生符号量の見積もり値であり以下のように求められる。

$$R_{depth} = UVLC(diff_depth) \quad (4.5)$$

$$diff_depth = curr_depth - pred_depth \quad (4.6)$$

$$pred_depth = median(depth_A, depth_B, depth_C) \quad (4.7)$$

$curr_depth$ が現符号化対象ブロックにおいて得られた符号化すべき奥行き情報の値であり、 $pred_depth$ が $curr_depth$ を予測符号化するために使用される予測値である。これらの値から $diff_depth$ を求め、この値を UVLC (Universal Variable Length Coding) で符号化したときに得られるビット数が R_{depth} である。すなわち、奥行き情報を実際に符号化したときの発生符号量ではない。 $median(a, b, c)$ は a と b と c の中間値を算出するという意味で

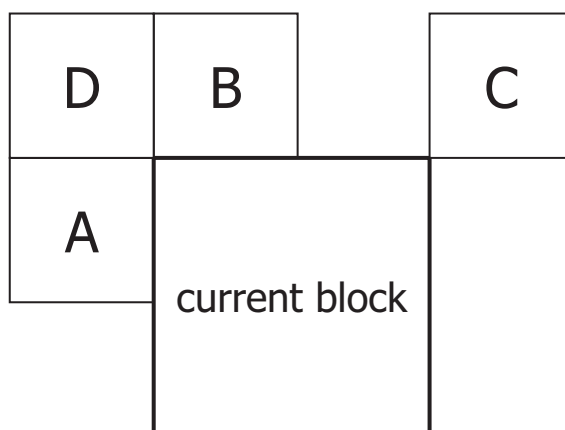


図 4-2: 奥行き情報予測値の生成に使用する周辺ブロック

あり, $depth_A$, $depth_B$, $depth_C$ とは, 図 4-2 の周囲の符号化済みブロック A, B, C の奥行き情報を意味している. 位置の関係上でブロック C が使用できない場合には, ブロック C の代わりにブロック D を使用する. ブロックサイズが 16×8 の場合で下側のブロックの場合には上側ブロックのみの予測を使用したり, ブロックサイズが 8×16 の場合で左側のブロックの場合には左側ブロックのみの予測を使用するという条件処理は, H.264/AVC で使用されている処理と同じである. また, 符号化対象ブロックの参照フレームインデックスがブロック A, B, C のいずれかのみと同じで他とは異なる場合には, 同じブロックからの予測のみを行う.

4.3.2 奥行き情報・視差ベクトル変換

図 4-1 の枠組みは H.264/AVC が基本となっており, 視差ベクトルや奥行き情報といったサイド情報を, 図 4-2 に示すように周囲の符号化済みブロックの値から予測しその予測値との差分値を符号化する. 周囲に奥行き情報と動きベクトル, 視差ベクトルが混在する状況となるが, これらは互いに直接予測に使用し合うことはできない. その結果, 例えば, 奥行き情報を符号化する場合に周囲に動きベクトルまたは視差ベクトルが存在する場合, 予測符号化に使用できる候補が少なくなるため予測効率が上がらず, 結果として符号化効率向上に寄与しないということが考えられる.

しかし, 奥行き情報と視差ベクトル, 動きベクトルのうち奥行き情報と視差ベクトルは, 持っている情報の意味自体は同じで形式が異なるだけであり, これらは数値的に相互に変換が可能であるため, 提案手法はこの性質を利用する. 以上の様子を図 4-3 に示す. 奥行き情報を視差ベクトルに, 視差ベクトルを奥行き情報に変換するため, 視差ベクトルと奥行き情報の両方の符号化効率向上に寄与することが予想される. 以下, 変換式について述べる.

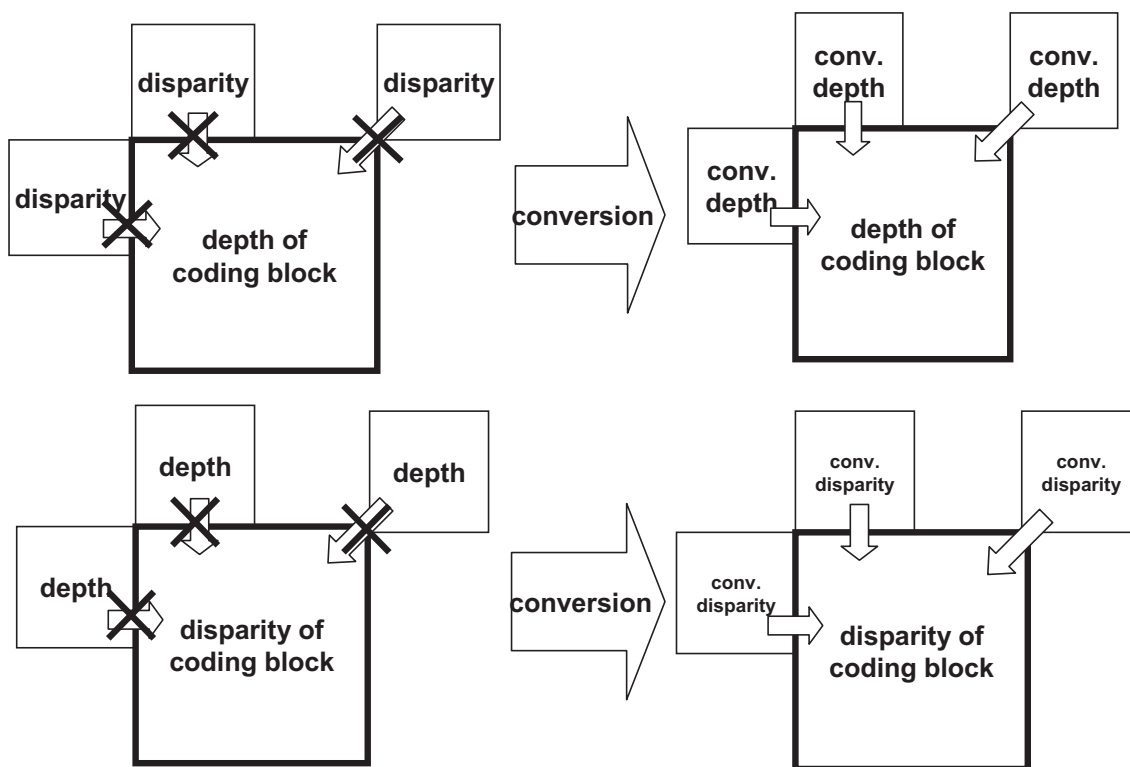


図 4-3: 奥行き情報・視差ベクトル変換の概念

変換式

奥行き情報から視差ベクトル ブロック毎の奥行き情報 z , 補正ベクトル (c_x, c_y) がわかっており, 現在のブロックの画像座標を (u, v) とすると, $m = [u - c_x \ v - c_y \ 1]^T$ として Warping の式 (2-10) に代入することで m' が得られる. 変換された視差ベクトル (dx_{conv}, dy_{conv}) は, 以下のように求められる.

$$\begin{bmatrix} dx_{conv} \\ dy_{conv} \end{bmatrix} = \begin{bmatrix} u \\ v \end{bmatrix} - \begin{bmatrix} u' \\ v' \end{bmatrix} \quad (4-8)$$

奥行き情報は $minZ$ から $maxZ$ までエピポーラ線に沿って次元探索されるので, カメラパラメータが特別誤った値でない限り, 変換された視差ベクトルが誤った値である可能性は低い. よって奥行き情報から視差ベクトルへ変換する場合の誤差対処は行わず, 上式で変換されたそのままの値 (dx_{conv}, dy_{conv}) を後の符号化の参照ベクトルとして用いることとする.

視差ベクトルから奥行き情報 視差ベクトルがわかっているため, 対応点 m と m' が既知であるということである. そこでまず, 式 (2-10) を以下のように表現し直す.

$$z'm' = zAm + e \quad (4.9)$$

$$A = K_{ref}R_{ref}(K_{cur}R_{cur})^{-1} \quad (4.10)$$

$$e = K_{ref}(t_{ref} - R_{ref}R_{cur}^{-1}t_{cur}) \quad (4.11)$$

カメラパラメータとしては、正式に統一されたものがないため、 R 及び t が世界座標系に対する値か、カメラ座標系に対する値かによって、 A の式は変わってくる。例えば、上の式における R と t はカメラ座標系に対するパラメータであり、シーケンスとしては、Breakdancers と Ballet が対応する。一方、Akko&Kayo や Rena シーケンスのような R のみがカメラ座標系に対するものであり、 t は世界座標系に対するものである場合もあり、以下ようになる。

$$A = K_{ref}R_{ref}(K_{cur}R_{cur})^{-1} \quad (4.12)$$

$$e = K_{ref}R_{ref}(t_{cur} - t_{ref}) \quad (4.13)$$

更に、MPEG 3DAV において提案されているカメラパラメータ⁽⁵⁸⁾ では以下のようになり、これらは Ballroom と Race1 が該当する。

$$A = K_{ref}R_{ref}^{-1}R_{cur}K_{cur}^{-1} \quad (4.14)$$

$$e = K_{ref}R_{ref}^{-1}(t_{cur} - t_{ref}) \quad (4.15)$$

このように、カメラパラメータの定義は統一されておらず様々であるが、結局式(4.9)のように表現できる。よって式(4.9)を z について解くことによって、 m と m' から奥行き情報が算出できる。両辺に m' との外積を取ると、

$$0 = zm' \times Am + m' \times e \quad (4.16)$$

更に両辺に $m' \times Am$ との内積を取ると、

$$0 = z\|m' \times Am\|^2 + (m' \times Am, m' \times e) \quad (4.17)$$

$$z = \frac{(m' \times Am, e \times m')}{\|m' \times Am\|^2} \quad (4.18)$$

得られた式(4.18)によって視差ベクトルから奥行き情報が得られる。しかし、視差ベクトルは必ずしもエピポラ幾何に従った値であるわけではなく、カメラキャリブレーションが不十分な場合の誤差、カメラ間の個体差、照度差に起因する誤差が発生する可能性があると考えられ、対処しなければならない。誤差が大きいと、本来存在するはずのない奥行き値が算出される。この誤った奥行き値を後の予測に用いるとブロック毎に誤差が伝搬していくこととなり、結果として性能を悪化させる恐れがある。これを防ぐために式(4.18)に

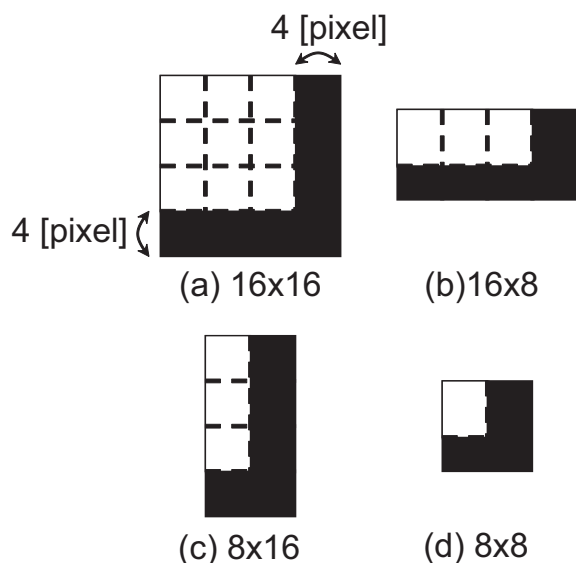


図 4-4: 奥行き情報・視差ベクトル変換の適用ブロック

よって求められる z を $minZ$ と $maxZ$ の間に収まるようにクリッピング処理を行う。これにより、変換される奥行き情報はどんなに誤差の含まれた視差ベクトルが算出されようとも $minZ$ 以上 $maxZ$ 以下であることが保証されるため、誤差の伝搬による性能の悪化を防ぐことが可能である。一方で値が $minZ$ 以上 $maxZ$ 以下になるということは、この $minZ$, $maxZ$ の値に性能が依存してくる可能性もある、ということになる。また、視差ベクトルがゼロの場合は、奥行きが無大であることを意味し無限遠にあると考えられるため、変換される z の値を $maxZ$ とする。

適用箇所

奥行き情報が視差ベクトルか動きベクトルかは、参照フレームインデックスによって識別する。H.264/AVC の枠組みでは、 8×8 から 16×16 ブロック毎に参照フレームインデックスを変更可能であり、このブロック毎に変換を行えばよいが、図 4-2 の参照関係を考慮するとブロックの全てのサイド情報を変換する必要がない。

具体的には図 4-4 の黒の領域で示した各ブロック毎の右下領域のみ変換すれば良い。変換には行列演算が必要となり、デコーダの負荷を増加させることが予想されるが、図の部分のみを変換することで、全てのブロック内画素の情報を変換するよりは、演算量の増加を抑えることが可能である。

図 4-2 の参照関係におけるブロック A, B, C の最小単位が 4×4 画素であることから、奥行き情報・視差ベクトル変換も 4×4 ブロックを基本に行う。ブロック毎に一定値の奥行き情報または視差ベクトルとなっているが、これらを変換して得られる値はそれぞれ異なってくるため、16 通りの変換された値が算出される。これらを 4×4 ブロック単位の予測に合わせるために、以下のように平均化を行う。

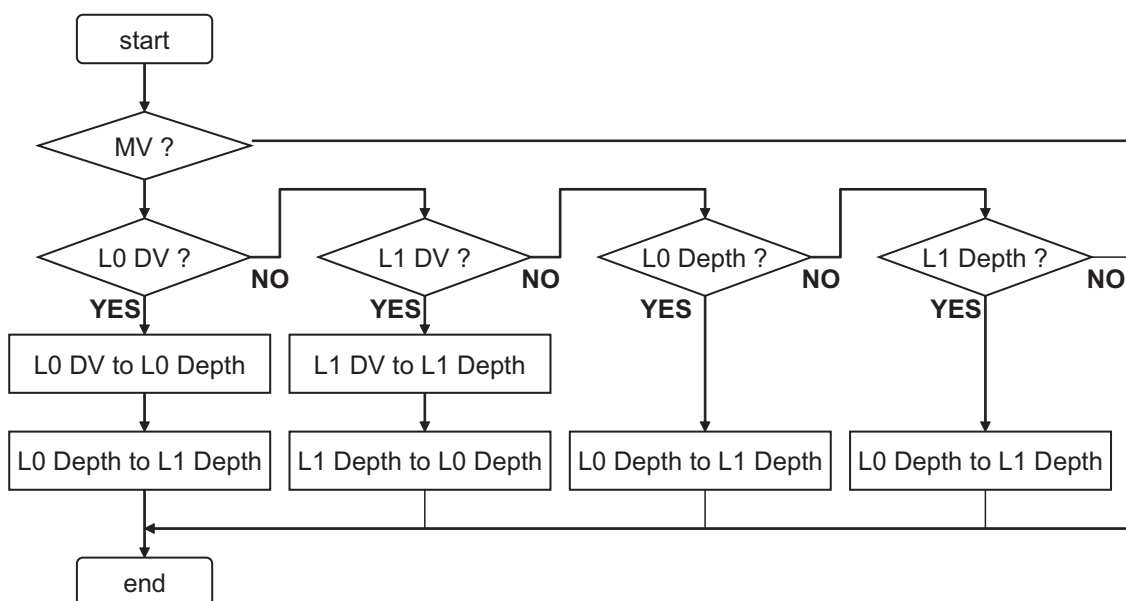


図 4-5: 奥行き情報の一貫性の利用による流れ

$$4 \times 4_value(S/4, T/4) = \frac{\sum_{x=S}^{S+3} \sum_{y=T}^{T+3} conv_value(x, y)}{16} \quad (4.19)$$

$4 \times 4_value(a, b)$ は 4×4 ブロックを 1 つの座標として見たとき位置 (a, b) における，変換された視差ベクトル又は奥行き情報の平均値である． $conv_value(a, b)$ は画素毎の変換された視差ベクトル又は奥行き情報である． (S, T) は 4×4 ブロック内の一番左上にある画素の位置である．

4.3.3 奥行き情報の一貫性の利用

奥行き情報の定義は様々であるが，本研究では符号化対象カメラ座標における奥行き情報を奥行き情報として定義しているため，前方 (L0) の参照フレームバッファに対して求められた奥行き情報であろうと，後方 (L1) の参照フレームバッファに対して求められた奥行き情報であろうと，これら 2 つに違いはないはずである．すなわち，L0(又は L1) の奥行き情報はそのまま L1(又は L0) の奥行き情報でもある．

P ピクチャ符号化には L0 の参照フレームバッファのみを用いた予測となるため，上記の性質は利用できない．一方，B ピクチャの場合は L0 の参照フレームバッファと L1 の参照フレームバッファの両方が使用可能であるため，上記の性質を利用することができる．すなわち，奥行き情報の一貫性の利用は B ピクチャ符号化時のみに適用可能である．B ピクチャ符号化には，L0 参照フレームのみを用いた L0 予測，L1 参照フレームのみを用いた L1 予測，L0 参照フレームと L1 参照フレームの両方を用いた双方向予測がある*．これらの予測のう

*H.264/AVC では，L0 参照フレームを 2 枚又は L1 参照フレームを 2 枚使用する双予測も可能であるが，今回は検討の対象外である．

ち、双方向予測の場合には、L0、L1 それぞれに対するサイド情報が算出されているため片方向の情報をもう片方に利用する必要がない。よって奥行き情報の一貫性が適用できるのは、L0 予測又は L1 予測の場合である。以上の流れをフローチャートとして図 4-5 に示す。

L0(又はL1)の奥行き情報が選択された場合には、上記の通りであるが、L0(又はL1)の視差ベクトルが算出された場合には、前節の視差ベクトルから奥行き情報への変換を用いる。こうすることで、例えばL0の視差ベクトルであったとしても、L0の奥行き情報とL1の奥行き情報がそれぞれ得られ、後の予測に利用されるようになる。片方向の奥行き情報からもう片方へコピーされた奥行き情報を更に視差ベクトルに変換することはしない。なぜなら、視差ベクトルは参照フレームに対する変位として定義されるが、コピーされた奥行き情報に対する参照フレームが定まらないからである。以上の処理を行うことで、更に予測のために使用可能な情報が増え、予測効率の向上に寄与すると考えられる。

4.3.4 スキップ/ダイレクトモードの改良

特に低ビットレート時に符号化効率を改善させるためには、低ビットレート時により多く選ばれるスキップモード(Pピクチャ時)、ダイレクトモード(Bピクチャ時)に対する改良が効果的であると考えられる。以下、スキップモードとダイレクトモードについて述べ、これらを改良する手法について述べる。

スキップモード

ブロックサイズ 16×16 に対してのみ定義され、ヒューリスティックに基づき、周囲のブロックの状況から符号化対象ブロックの変位ベクトルを算出して予測画像を生成するものである。予測残差信号も符号化しないことから、歪みが大きくなるが、発生符号量を格段に抑えることができる。低ビットレート時には発生符号量の方が重要となってくるため、予測の精度が高ければレート歪みの観点から有効な手法である。

具体的な予測方法は、現符号化対象ブロックの右ブロック(図4-2のブロックA)と左(同図のブロックB)の状態を見る。AかBのブロックの参照フレームインデックスがゼロ(すなわち直前に符号化されたフレーム)でかつ変位ベクトルがゼロである場合には、符号化対象ブロックの参照フレームインデックスをゼロとし、更に符号化対象ブロックの変位ベクトルをゼロとする。AとB両方の参照フレームインデックスがゼロでなかったり、変位ベクトルがゼロでない場合には、参照フレームインデックスをゼロとして、周囲ブロックからの予測変位ベクトルが符号化対象ブロックの変位ベクトルにセットされる。

ダイレクトモード

H.264/AVCには空間ダイレクトモードと時間ダイレクトモードの2種類が使用可能であるが、多視点映像符号化においては、空間ダイレクトモードが適していることが知られてい

る⁽³⁰⁾。よって、空間ダイレクトモードを対象とするため、空間ダイレクトモードについて述べる。

図4-2のブロックA, B, Cの参照フレームインデックスの値のうち、最も小さい参照フレームインデックスの値を符号化対象ブロックの参照フレームインデックスとして用いる。得られた参照フレームインデックスの値がイントラ予測符号化を表す-1の場合、ゼロが使用される。ダイレクトモードの参照フレームインデックスは、L0参照フレームとL1参照フレームがありそれぞれ別々に適用される。

次に、L1参照フレームのインデックスがゼロであるフレーム内の同じ位置に属するブロック (Colocated Block) の変位ベクトル情報を調べ、変位ベクトルが値を持たなかった場合には、予測変位ベクトルがゼロとして適用される。そうでない場合には、周囲ブロックからの予測変位ベクトルがセットされる。

参照フレームインデックス算出方法の変更

スキップモードでは、使用される参照フレームインデックスが常にゼロであるため、奥行き情報が選択されることはない。また、ダイレクトモードでは、ブロックA, ブロックB, ブロックCの参照フレームインデックス値の最小値が使用されるが、奥行き情報は使用可能な参照フレームインデックスの後半に位置づけられているため、奥行き情報が予測として選択される可能性は低い。よって、提案手法としては、スキップモードまたはダイレクトモードにおける参照フレームインデックスの算出の仕方に関して以下のような改良を行う。

1. ブロックA, ブロックB, ブロックCの参照フレームインデックスが全て等しい場合、その参照フレームインデックスを符号化対象ブロックに対して用いる。
2. 1. でない場合、ブロックA, ブロックB, ブロックCのうち2つが等しい参照フレームインデックスを用いる。
3. 2. でない場合、ブロックA, ブロックB, ブロックCの参照フレームインデックスの最小値を用いる。
4. 得られた参照フレームインデックスが-1 (すなわちイントラ予測符号化されたブロック) の場合にはゼロとする。

上記処理により、スキップモードにおいて周囲が奥行き情報予測の場合には奥行き情報予測として働くこととなる。一方、ダイレクトモードは元々奥行き情報予測が行われる可能性がないわけではないが、より奥行き情報予測が選ばれるようになると考えられる。提案する参照フレームインデックス算出方法は、特に奥行き情報を重視したものではない。周囲の状態を見て、多数決によって符号化対象ブロックのサイド情報予測を行うというものである。

4.4 むすび

本章では、View Interpolation を用いた多視点映像符号化として、まず提案手法が基本とする従来手法の枠組みについて説明した。基本とする符号化方式はH.264/AVCをベースとした、奥行き情報をエンコーダで生成し奥行き情報による予測を行うという枠組みである。次に提案手法について述べた。提案手法は4つあり、それぞれレート歪みコストに基づく奥行き情報探索、奥行き情報・視差ベクトル変換、奥行き情報の一貫性の利用、スキップ/ダイレクトモードの改良である。

第5章

評価実験

5.1 まえがき

本章ではシミュレーションによる評価実験を行う。4章で述べた提案手法をJMに実装した。3章で示した通り、視点間予測と時間予測が使用可能なフレームにおいては、時間予測が使用される傾向にある。よって時間予測を使用しないアンカーフレームに対して評価を行う。

5.2 実験条件

符号化条件を表5-1に示す。予測構造は3章の予備実験から、Pピクチャのみの場合は、PPIPP構造、Bピクチャを含む場合はPBIBP構造を使用した。必ずしもBピクチャを含めた方が効率が上がるわけではないが、比較検討のため、また奥行き情報の一貫性の利用はBピクチャのみに適用可能であるため、Bピクチャ有りのPBIBP構造のデータについても検討を行う。重み付け予測に関しては、3章の予備実験の結果からAkko&KayoとRace1以外の全てのシーケンスで符号化効率が向上するため使用する。

使用する奥行き情報探索のパラメータを表5-2に示す。 $minZ$ から $stepZ$ ずつ $maxZ$ まで探索するという意味である。 $minZ$, $maxZ$ は、原画像を用いて、隣接カメラに 4×4 ブロックでWarpingを行い、SSDが最小なった場合の奥行き値のヒストグラムから決定した。また $stepZ$ は、得られた $minZ$ と $maxZ$ から手動で決定したものである。現在、これらパ

表 5-1: 符号化条件

Number of cameras	5
Prediction structure	PPIPP, PBIBP
QP (fix)	28, 32, 36, 40 (QPI=QPP=QPB)
Search range	± 128 [pixel]
Weighted prediction	on (except for Akko&Kayo and Race1)
Entropy coding	CABAC

表 5-2: 奥行き情報探索の範囲

sequence	minZ	maxZ	stepZ
Akko&Kayo	2000	10000	50.0
Ballet	5	135	1.0
Breakdancers	5	135	1.0
Ballroom	1	30	0.5
Race1	1	150	1.0
Rena	2000	10000	50.0

ラメータを系統的に決定する手法は存在していない。

5.3 結果及び考察

5.3.1 レート歪みに基づく奥行き情報探索

結果

奥行き情報を探索する場合のコスト値を，SSD の場合とレート歪みコストとした符号化を行った．PPIPP 構造での結果を図 5-1 に，PBIBP 構造での結果を図 5-2 にそれぞれ示す．図中の凡例の意味は，“SSD”が奥行き情報探索時に SSD を用いて符号化を行った場合，“RD”がレート歪みコストを用いて符号化を行った場合である．横軸は，奥行き探索を行うフレームにおける発生符号量の平均であり，縦軸は，輝度信号の平均 SNR である．

考察

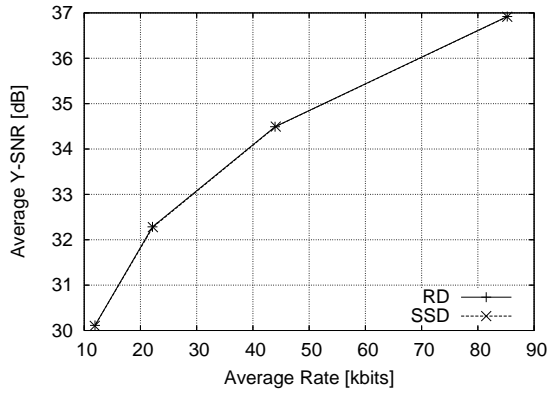
結果からわかることとして，全体的に際立った違いが見受けられないということである．ごくわずかに違いが見受けられたのは，Breakdancers と Ballet のシーケンスである．図 5-1 の PPIPP 構造と図 5-2 の PBIBP 構造では，PBIBP 構造の方がより効果が出ている．これは，P ピクチャは参照カメラとの距離が大きく，インター符号化が選ばれにくいからである．一方 B ピクチャは双方向からの予測が可能で，よりインター符号化が選ばれやすくなっており，更に符号化モード決定時のラグランジュ乗数は，B ピクチャを含む場合にはインター符号化がより選ばれやすいように値が設定されているからである．このように，B ピクチャではより多くのインター符号化が選ばれやすくなり，レート歪みコストによる探索の効果が出たものと考えられる．そこで，違いの表れた PBIBP 構造での Ballet シーケンスにおける，奥行き情報，視差ベクトル，イントラ符号化の選択率を表 5-3 に示す．選択率の算出方法は，そのマクロブロック内で 1 ブロックでも奥行き情報が選択された場合にそのマクロブロックは奥行き情報が選択されたとし，奥行き予測を行うフレームにおける総マクロブロック数に対する割合を示している．よって，視差ベクトルの割合は，そのマクロブロックが視

表 5-3: 奥行き情報探索コストの違いによる各種割合 (PBIBP)

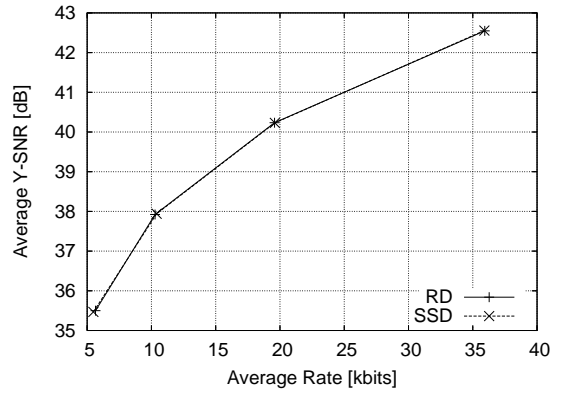
sequence	QP	cost value	depth rate [%]	intra rate [%]	disparity rate[%]
Ballet	28	SSD	20.0	46.6	33.5
		RD	24.7	45.9	29.4
	32	SSD	21.2	40.2	38.6
		RD	24.3	39.5	36.1
	36	SSD	15.0	37.5	47.5
		RD	15.7	34.8	49.5
	40	SSD	12.7	32.5	54.8
		RD	14.2	32.4	53.4
Breakdancers	28	SSD	22.7	34.7	42.6
		RD	25.4	28.1	44.6
	32	SSD	22.7	29.5	47.7
		RD	27.3	28.1	44.6
	36	SSD	19.8	29.8	50.4
		RD	24.1	28.8	47.1
	40	SSD	13.6	28.1	58.3
		RD	21.2	25.9	52.9

差ベクトルのみで符号化された場合の割合となっている。表 5-3 からわかるとおり、コストが SSD の場合よりも RD の場合の方がより奥行き予測が選択されていることがわかる。高ビットレート時にも SSD より RD の方が選択されていることもわかる。

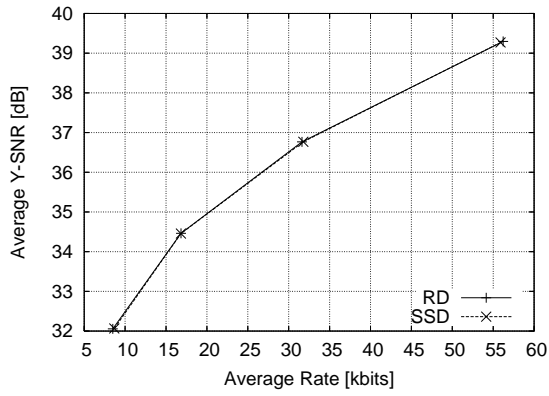
動き/視差ベクトルに対して周囲の値を考慮したレート歪みコストによる探索が有効なのは、これら変位ベクトルは 2 次元の探索領域の探索候補があるからである。例えば、 $\pm 16[\text{pixel}]$ で高速化手法を用いない場合、 $33 \times 33 = 1089$ 点及び小数画素精度の場合はそれら周辺という候補がある。歪みのみを評価値として、特に周期的な模様やテクスチャの無い領域を探索する場合には、歪みの値が僅差にも関わらず発生符号量の大きなベクトルが選ばれる可能性がある。一方奥行き情報の探索は、2 章で述べたエピポーラ制約に基づきエピポーラ線上の探索となるため、2 次元探索ではなく、1 次元探索となる。例えば Breakdancers で今回使用したパラメータは、 $\min Z = 5$ 、 $\max Z = 135$ 、 $\text{step} Z = 1$ であり、計 131 通りの値しか発生しない。一方今回用いた視差ベクトル探索範囲は、 $\pm 128[\text{pixel}]$ であり、 $257 \times 257 = 66049$ 点である。このように、制約によって候補数が格段に減ることで取りうる値の候補が限られ、周囲の状況を考慮したレート歪みコストでなくとも取り得る値そのものが少ないため、レート歪みコストであろうと SSD コストであろうと、差は表れなかったと考えられる。符号量の項の効き目が薄いとすると、式 (4.2) は SAD と SSD の差になるが、レート歪みコストによる奥行き情報探索は、SSD による探索よりも符号化効率は悪くなることはなかった。



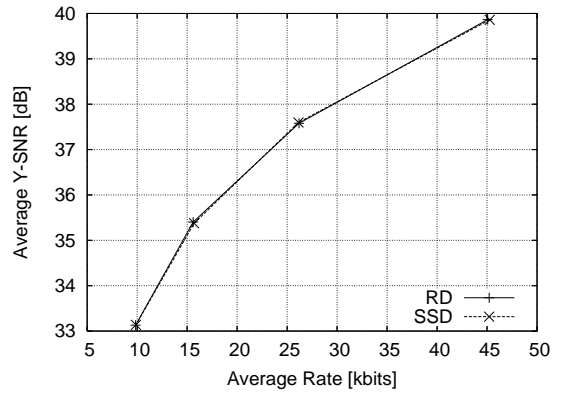
(a) Ballroom



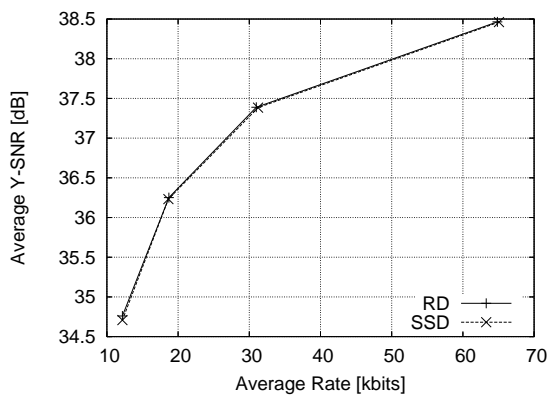
(b) Rena



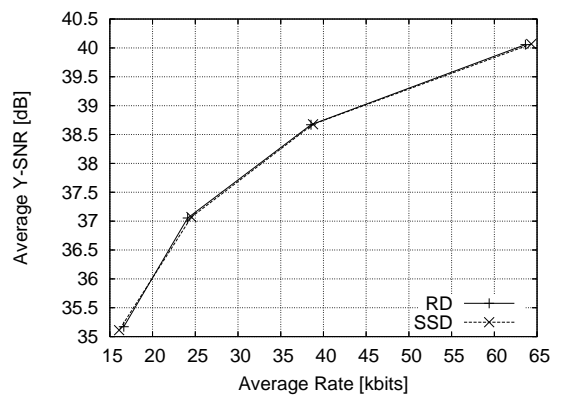
(c) AkkoKayo



(d) Race1

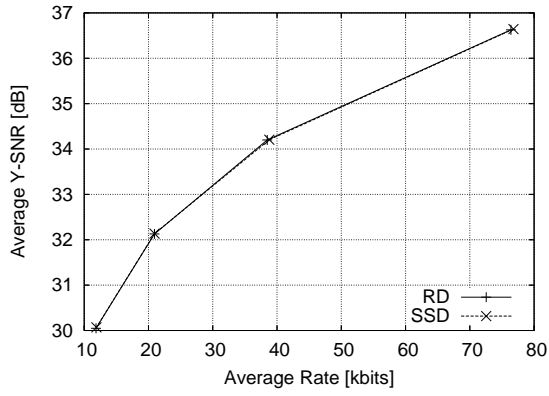


(e) Breakdancers

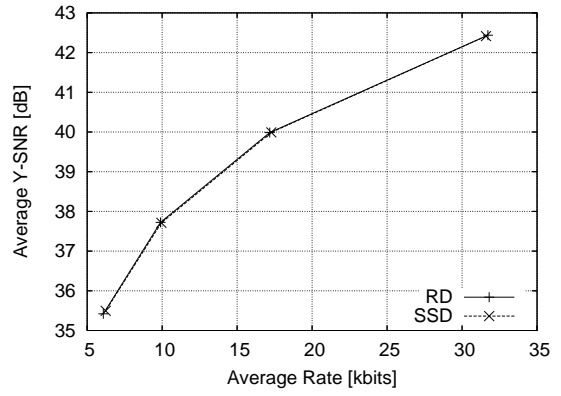


(f) Ballet

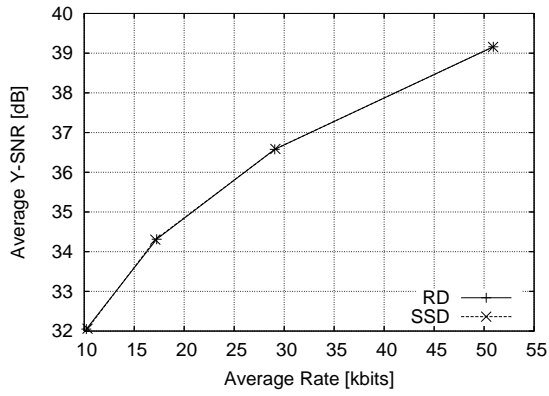
図 5-1: 奥行き情報探索コストの違い (PPIPP)



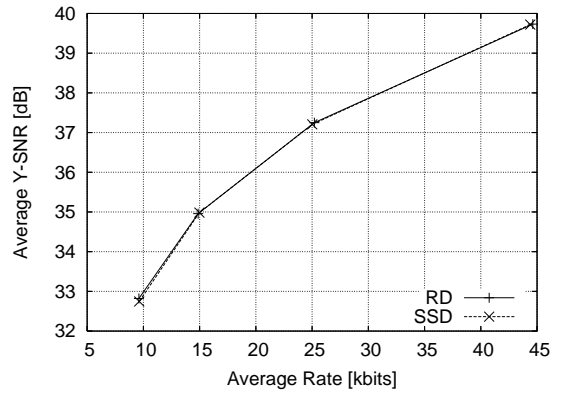
(a) Ballroom



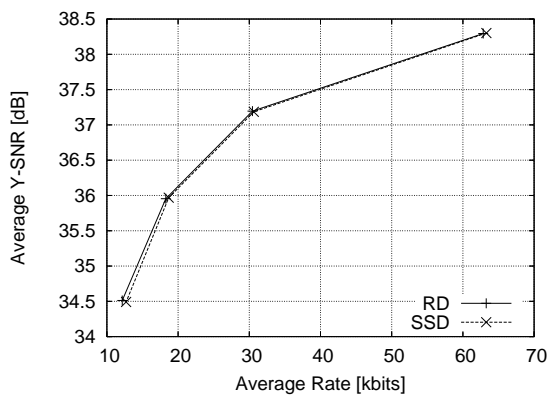
(b) Rena



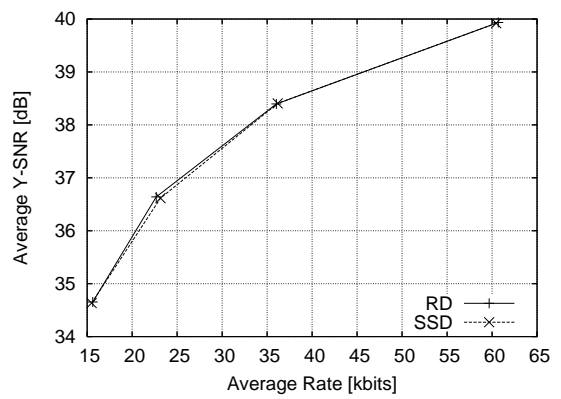
(c) AkkoKayo



(d) Race1



(e) Breakdancers



(f) Ballet

図 5-2: 奥行き情報探索コストの違い (PBIBP)

5.3.2 奥行き情報・視差ベクトル変換

結果

奥行き情報と視差ベクトルの相互変換を行って符号化した場合のPPIPP構造での結果を図5-3に、PBIBP構造での結果を図5-4にそれぞれ示す。図中の凡例の意味は、“CO on”が奥行き情報・視差ベクトル変換を行った場合のグラフ，“CO off”が奥行き情報探索は行いが、奥行き情報・視差ベクトル変換は行わない場合のグラフである。また視差探索のみで符号化した場合のグラフである“only DC”も併せて載せた。横軸及び縦軸は、奥行き情報探索コストの違いの場合と同じである。

考察

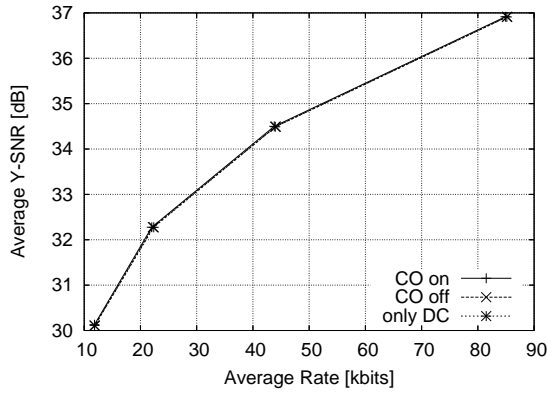
今回は、PPIPP構造でもPBIBP構造でも同じような効果が表れている。この結果から、Bピクチャの有無に関わらず、ほぼ同様の結果が得られることがわかる。奥行き情報探索コストの結果と同様に、BreakdancersとBalletにおいてより際立った違いが表れ、RenaとRace1においても多少の違いが見受けられた。一方、Akko&KayoとBallroomは前節同様差が表れなかった。そこでBreakdancers、Race1、Akko&Kayoの奥行き情報、視差ベクトル、イントラ符号化の選択率の割合を表5-4に、符号化モードの割合を表5-5にそれぞれ示す。

これら効果としてまず以下の2点が考えられる。

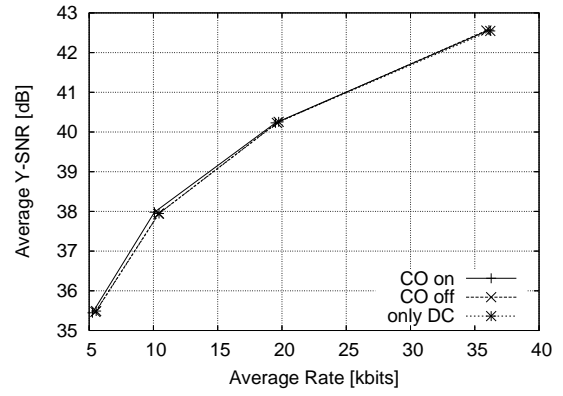
1. 制約された線（エピポーラ）上を探索することにより、コンパクトに被写体の視差を補償することができた。
2. カメラ間のカメラパラメータの違いを補正できた。

1番目の理由に関しては、4章において述べたとおりである。2番目の理由に関しては、BreakdancersとBalletは3章で示した通り、1次元の水平方向にアーク状に配置されたカメラによって得られたシーケンスであり、カメラ画像間で回転が存在している。またカメラの内部行列も異なっており、精度の高いパラメータが得られている。Race1およびRenaは1次元の水平方向に直線のカメラ配置であるが、Race1は回転行列のキャリブレーションが高精度に求められており各カメラ間の微妙な回転の違いが記述されている。Renaシーケンスは内部行列の光学中心位置を表すパラメータがわずかに異なる。このような回転や内部行列すなわち拡大・縮小の違いは、従来の平行移動モデルを仮定したMPEG符号器の苦手とする効果であり、これらを補正することが出来たために利得が得られたと考えられる。

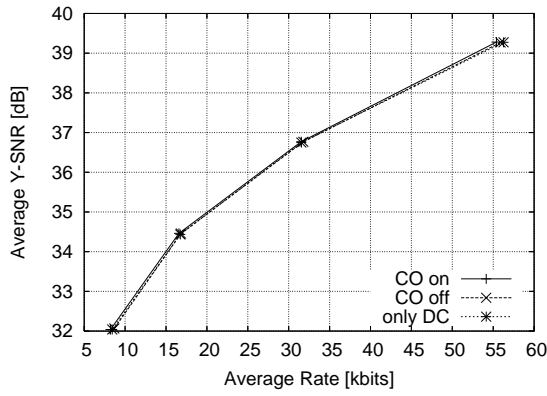
上記理由は“CO off”にも当てはまるが、図5-3、5-4からわかるとおり、“only DC”と比較して“CO off”はあまり違いが表れていない。しかし、表5-4からわかるように、奥行き予測が選ばれていないわけではないため、レート歪みの観点から視差ベクトルよりも良いブロックが存在していることになる。“CO on”では効果が出ていることから、奥行き情報探索自体は有効であるが、周囲からの予測によって奥行き情報や視差ベクトルを符号化する際に



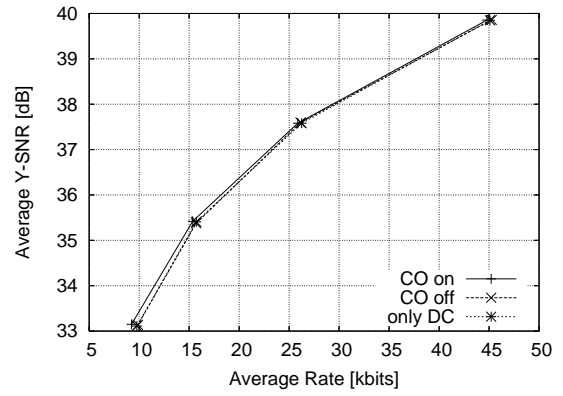
(a) Ballroom



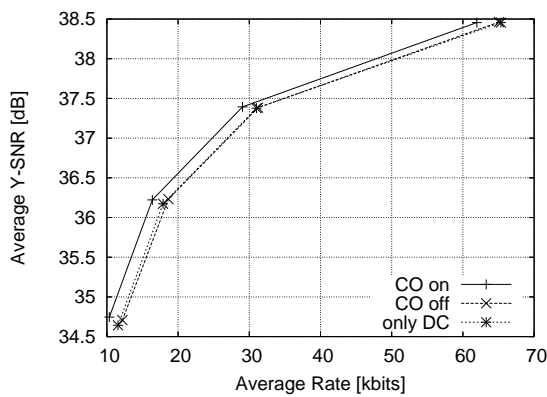
(b) Rena



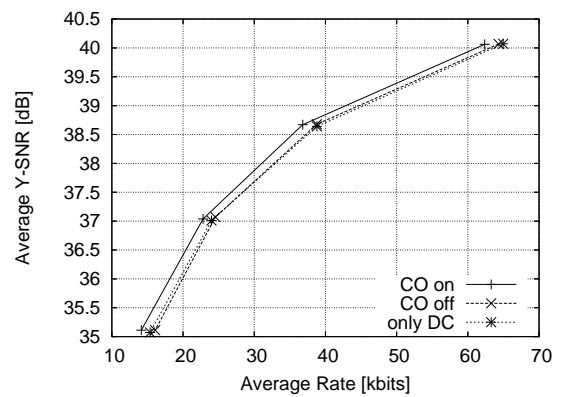
(c) AkkoKayo



(d) Race1

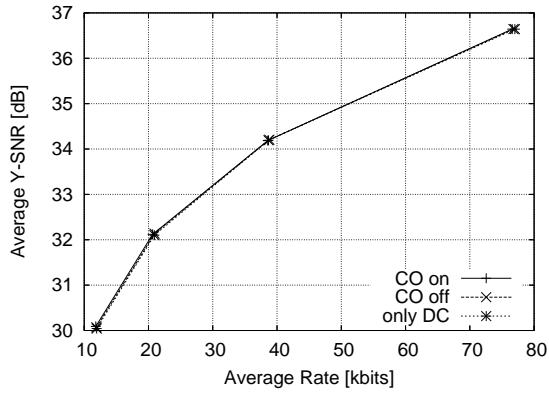


(e) Breakdancers

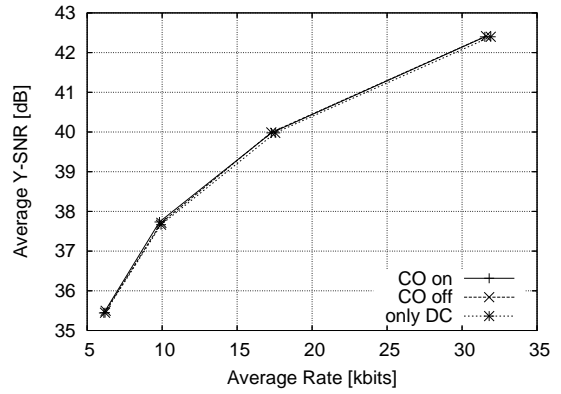


(f) Ballet

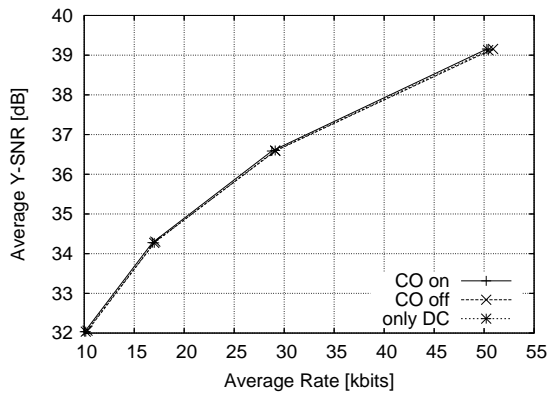
図 5-3: 奥行き情報・視差ベクトル変換 (PIPP)



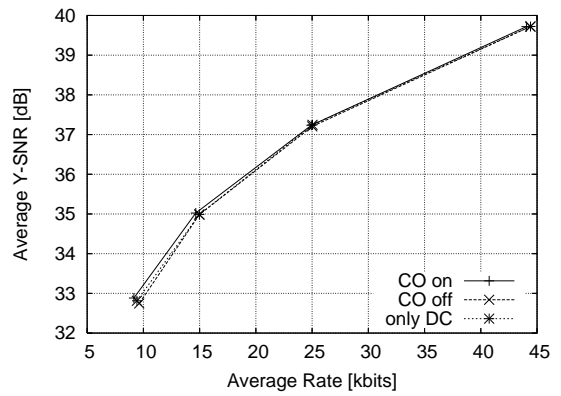
(a) Ballroom



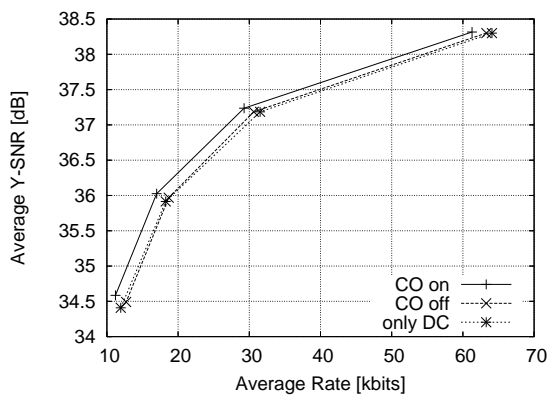
(b) Rena



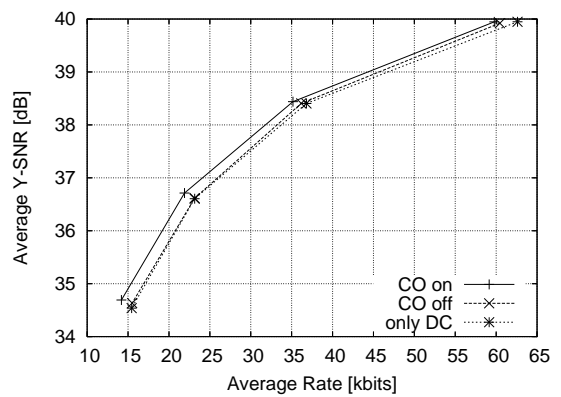
(c) AkkoKayo



(d) Race1



(e) Breakdancers



(f) Ballet

図 5-4: 奥行き情報・視差ベクトル変換 (PBIBP)

表 5-4: 奥行き情報・視差ベクトル変換による各種割合 (PPIPP)

sequence	QP	conversion	depth rate [%]	intra rate [%]	disparity rate[%]
Breakdancers	28	off	14.6	43.6	41.8
		on	17.0	40.3	42.8
	32	off	16.8	38.3	44.9
		on	13.6	33.7	52.7
	36	off	18.2	32.7	49.1
		on	11.0	20.3	68.7
	40	off	17.3	25.6	57.1
		on	8.4	13.8	77.8
Race1	28	off	4.7	46.3	49.0
		on	4.9	45.4	49.7
	32	off	4.0	43.2	52.9
		on	4.9	40.9	54.2
	36	off	5.8	35.4	58.9
		on	4.2	33.3	62.6
	40	off	7.9	31.9	60.2
		on	4.3	28.8	66.9
Akko&Kayo	28	off	13.8	21.3	64.9
		on	10.0	21.2	68.8
	32	off	5.3	21.0	73.7
		on	6.2	20.9	72.9
	36	off	3.7	17.8	78.5
		on	4.0	18.1	77.9
	40	off	1.4	15.0	83.7
		on	2.7	14.2	83.1

表 5-5: 奥行き情報・視差ベクトル変換によるモードの割合 (PIPP)

sequence	QP	conv.	skip[%]	16×16[%]	16×8[%]	8×16[%]	8×8[%]	intra[%]
Breakdancers	28	off	14.8	24.7	7.0	7.2	2.7	43.6
		on	18.6	21.5	5.3	5.0	1.3	40.3
	32	off	21.2	29.2	4.9	5.1	1.3	38.3
		on	34.2	20.5	5.3	5.0	1.3	33.7
	36	off	31.7	28.1	3.2	3.6	0.7	32.7
		on	50.0	21.9	3.4	3.7	0.7	20.3
	40	off	42.5	28.4	1.7	1.4	0.4	25.6
		on	68.3	13.3	2.2	2.2	0.3	13.8
Race1	28	off	12.0	26.5	8.5	5.0	1.7	46.3
		on	13.3	26.2	8.7	4.8	1.6	45.4
	32	off	22.5	25.3	5.5	2.6	1.0	43.2
		on	23.5	25.7	5.7	3.3	0.9	40.9
	36	off	39.3	19.6	3.9	1.4	0.4	35.4
		on	41.7	18.3	4.6	1.7	0.4	33.3
	40	off	46.6	17.5	2.6	1.4	0.1	31.9
		on	53.1	14.4	2.4	1.1	0.1	28.8
Akko&Kayo	28	off	5.1	45.2	7.5	15.4	5.4	21.3
		on	5.6	44.0	8.3	15.2	5.7	21.2
	32	off	16.9	41.0	6.1	11.7	3.3	21.0
		on	17.3	40.7	5.5	12.2	3.5	20.9
	36	off	38.4	31.4	4.1	7.0	1.2	17.8
		on	39.8	30.0	4.0	6.7	1.4	18.1
	40	off	62.7	17.0	2.1	2.6	0.6	15.0
		on	63.6	16.2	2.4	3.0	0.5	14.2

疎らで孤立した予測となり歪みよりも発生符号量の増加の影響が顕著になったからであると考えられる。すなわち、奥行き予測による利得と視差ベクトルの予測効率低下による損失が相殺し合うこととなり、結果としてレート歪み特性は変化しなかったものと考えられる。

表5-4から、奥行き情報・視差ベクトル変換を行うことで、奥行き予測を行うマクロブロック数の割合が小さくなることがわかる。低ビットレートになればなるほど奥行き予測を行うブロックの割合は減り、視差予測を行うブロックの割合が高くなっていることもわかる。この傾向は Breakdancers と Race1 の両シーケンスで成り立っており、記載していない Ballet と Rena でも同様の傾向が見受けられた。そこで表5-5に注目してみると、変換を行うか否かでの大きな違いは、特にスキップモードの割合であることがわかる。スキップモードは探索を行わずに、周囲のブロック情報のみから現ブロックの情報を生成する符号化モードであり、予測がうまくいけば符号化効率の向上に大きく寄与する。奥行き情報を視差ベクトルに変換することで、スキップモード時に使用可能な周囲ブロックの視差ベクトルが増え、より精度の高い予測を行うことが出来たため、符号化効率が向上したことがわかる。すなわち、視差ベクトルから奥行き情報への変換よりも、奥行き情報から視差ベクトルへの変換の方が符号化効率向上への寄与度が大きいことがわかる。

一方で、図5-3, 5-4からわかるとおり、Ballroom と Akko&Kayo のシーケンスではほとんど効果が表れなかった。この原因として、これらシーケンスのカメラパラメータの違いがないことが挙げられる。Ballroom は rectification が行われており回転行列と内部行列は全カメラで全く同じである。Akko&Kayo シーケンスは rectification が行われてないが、与えられているカメラパラメータは回転行列と内部行列が全て同じである。これらシーケンスに共通する回転行列と内部行列が同じであるという場合、Warping の式(2-10)は以下のようになる。

$$z'm' = zm + e \quad (5.1)$$

更にこれらのシーケンスは並進ベクトルの z 成分もゼロであることから $z' = z$ が成り立つため以下のようなになる。

$$u' = u + \frac{e_u}{z} \quad (5.2)$$

$$v' = v + \frac{e_v}{z} \quad (5.3)$$

e_u, e_v は e の第1成分、第2成分である。これは明らかに平行移動成分のみの補償となり、従来の動き/視差補償と全く同じである。奥行き情報で変位がコンパクトに表現され奥行き予測が使用されて符号化されても、得られる利得はわずかであることが考えられる。表5-4, 5-5を見る限り奥行き予測が全く選ばれていないわけではないが、上記理由により符号化効率への寄与度は低かったものと考えられる。

表 5-6: 奥行き情報の一貫性の利用による各種割合 (Ballet)

QP	ZCOPY	depth rate [%]	intra rate [%]	disparity rate[%]
28	off	24.4	4.0	71.6
	on	28.5	3.6	67.9
32	off	17.8	1.2	81.0
	on	20.8	1.2	78.0
36	off	12.8	0.6	86.6
	on	16.6	0.5	82.9
40	off	8.2	0.2	91.6
	on	11.3	0.1	88.6

5.3.3 奥行き情報の一貫性の利用

結果

奥行き情報の一貫性の利用は、B ピクチャに対してのみ適用可能であるため、PBIBP 構造のみで実験を行った。結果を図 5-5 に示す。図の横軸は、B ピクチャのみの発生符号量の平均であり、縦軸も B ピクチャのみの平均 SNR である。図中の凡例の意味は、“ZCOPY on”が奥行き情報の一貫性を利用した方式、“ZCOPY off”が利用しない奥行き情報・視差変換のみを行う方式である。前項の考察から、Akko&Kayo と Ballroom のシーケンスは省いてある。

考察

図 5-5 からわかることとして、違いが表れたのは、Breakdancers、Ballet 及び Rena である。図 5-5 の Rena、Breakdancers、Ballet の図から、わずかであるが高ビットレート時により効果が表れていることがわかる。そこで Ballet の奥行き情報、視差ベクトル、イントラ符号化の選択率の割合を表 5-6 に、符号化モードの割合を表 5-7 に示す。

表 5-6 において consistency の”on”と”off”の奥行き予測の選択割合を比較すると、高ビットレート時のほうが大きいこと、およびどのレート（どの QP）においてもほぼ一定の割合で増えていることがわかる。一方、表 5-6 では、符号化モードの割合はほとんど変化していない。すなわち、符号化モード分布の特性は変化しないが、一貫性を利用することによって符号化すべき奥行き情報の値が減り、特に高ビットレート時により奥行き予測が選択され、図 5-5 のような傾向の利得が得られたことがわかる。奥行き情報の一貫性の利用は、特に高ビットレート時に有効であることがわかった。

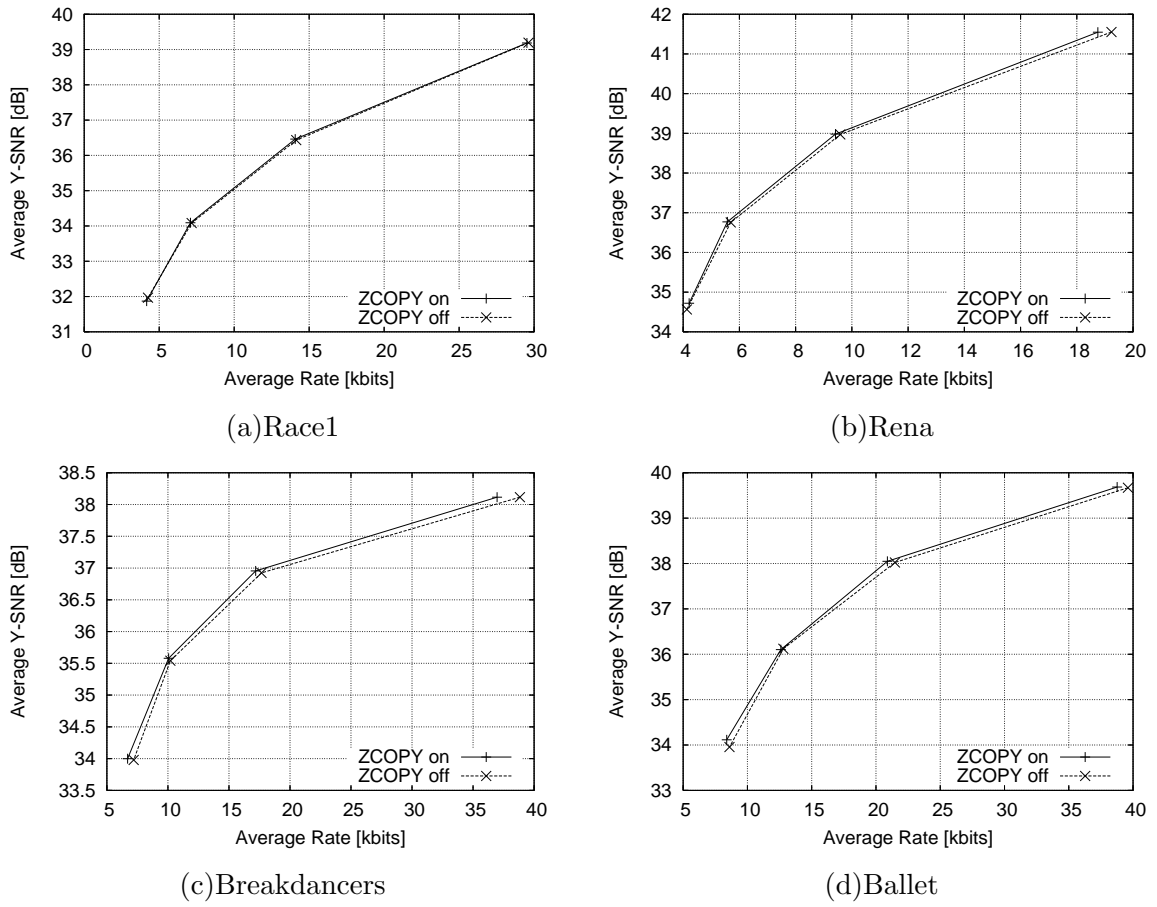


図 5-5: 奥行き情報の一貫性の利用 (PBIBP)

表 5-7: 奥行き情報の一貫性の利用によるモードの割合 (Ballet)

QP	ZCOPY	direct[%]	16×16[%]	16×8[%]	8×16[%]	8×8[%]	intra[%]
28	off	39.8	42.6	5.9	5.2	2.6	4.0
	on	40.2	42.3	5.6	5.6	2.7	3.6
32	off	55.1	38.2	2.5	2.5	0.5	1.2
	on	56.0	36.7	2.7	2.6	0.8	1.2
36	off	67.4	29.0	1.7	1.1	0.2	0.6
	on	68.2	28.5	1.7	1.0	0.1	0.5
40	off	75.7	21.9	1.4	0.8	0.0	0.2
	on	76.6	21.3	1.2	0.6	0.1	0.1

5.3.4 スキップ/ダイレクトモードの改良

結果

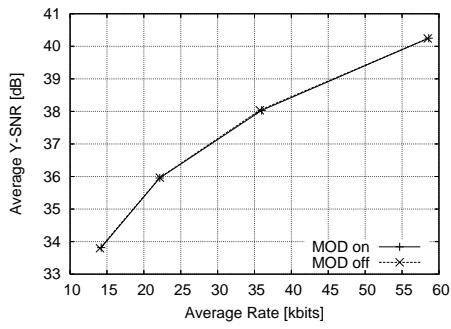
スキップ/ダイレクトモードの改善手法の効果を見るためには、P ピクチャにおいてはスキップモード、B ピクチャにおいてはダイレクトモードとなるため、両方含まれる PBIBP 構造で実験を行った。これらを別々で見えるために、P ピクチャのみの結果を図 5-6 に、B ピクチャのみの結果を図 5-7 にそれぞれ示す。図 5-6 中の凡例の意味は、“MOD on”がスキップモードの参照インデックスを変更した方式、“MOD off”が変更しないが奥行き情報・視差ベクトル変換のみを行う方式である。図 5-7 中の凡例の意味は、“MODZCOPY on”が前項の奥行き情報の一貫性とダイレクトモードの参照フレームインデックスを変更した方式、“ZCOPY on”が参照フレームインデックスの変更は行わないが奥行き情報の一貫性を適用した方式である。

考察

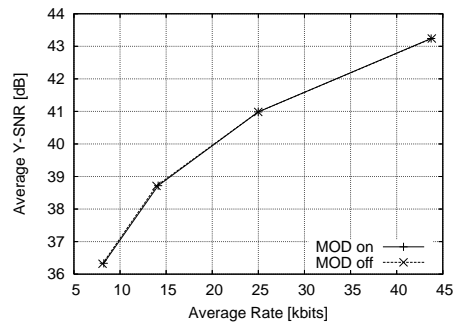
これらの図より、ほとんどのシーケンスで違いが表れていないことがわかる。B ピクチャの Ballet において違いが表れたため、Ballet のスキップ/ダイレクトモードの割合を表 5-8 に、奥行き情報、視差ベクトル、イントラ符号化の選択率の割合を表 5-9 にそれぞれ示す。表中の“MOD on”は、P ピクチャの場合はスキップモードの改良を行った場合、B ピクチャの場合はダイレクトモードの改良を行った場合を示している。一方“MOD off”は、P ピクチャの場合はスキップモードの改良を行わない奥行き情報・視差ベクトル変換のみを行う場合、B ピクチャの場合はダイレクトモードの改良を行わない奥行き情報の一貫性までを行う場合である。

表 5-8 からスキップ/ダイレクトモードの割合自体はあまり変化していないことがわかるが、表 5-9 から奥行き予測の割合が増えていることがわかる。すなわち、従来の H.264/AVC のスキップ/ダイレクトでは奥行き予測が選ばれない、もしくは選ばれにくかったが、提案手法によってより選ばれることとなったことがわかる。また、表 5-9 から、特に B ピクチャにおいて奥行き予測の割合が増えていることから、図 5-7 による違いが表れたと考えられる。

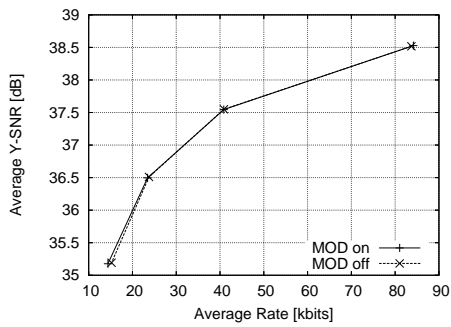
しかし、図 5-6、5-7 を見る限り Ballet に違いが表れたのみで、他のシーケンスでは効果が見受けられていない。参照フレームインデックスを変更して奥行き予測が選ばれたとしても、予測によって算出された奥行き情報による予測は、従来の変位ベクトルと大差ないことがわかった。低ビットレート時にスキップ/ダイレクトモードがより多く選ばれることは確かであるため、参照フレームインデックスの変更ではなく、違ったヒューリスティックを導入することで、改善の余地があると考えられる。



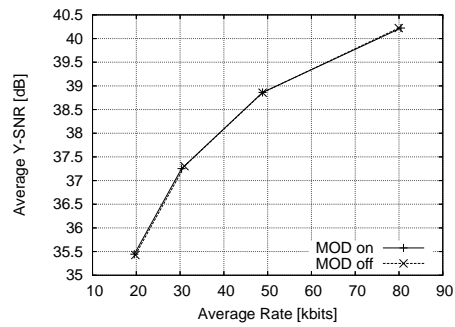
(a)Race1



(b)Rena

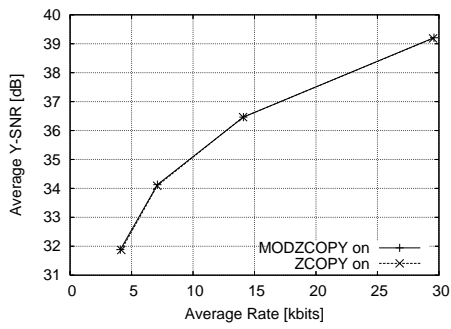


(c)Breakdancers

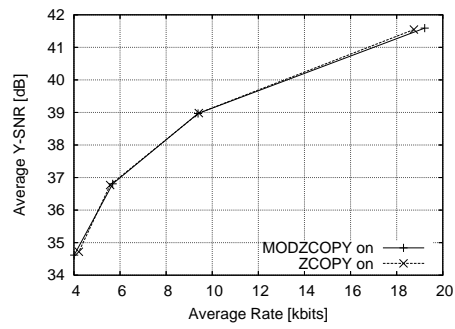


(d)Ballet

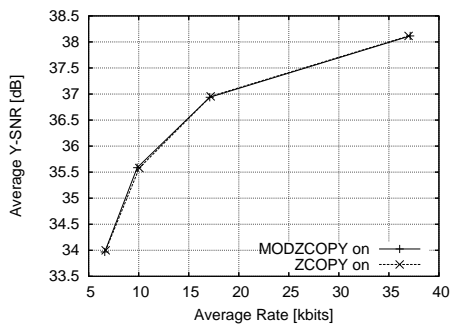
図 5-6: スキップモードの改良 (PBIBP)



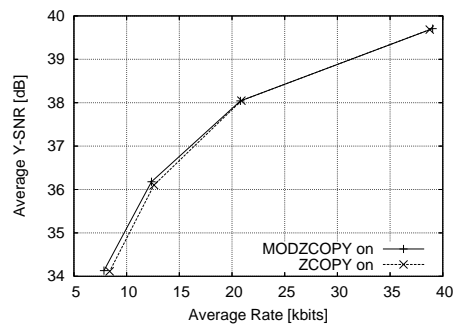
(a)Race1



(b)Rena



(c)Breakdancers



(d)Ballet

図 5-7: ダイレクトモードの改良 (PBIBP)

表 5-8: スキップ/ダイレクトモードの改良によるスキップ/ダイレクトモードの割合 (Ballet)

picture	MOD	QP			
		28	32	36	40
P	off	3.3	10.6	19.4	35.0
	on	3.3	11.1	21.6	37.4
B	off	42.7	57.3	67.4	74.1
	on	41.5	57.9	68.1	75.6

表 5-9: スキップ/ダイレクトモードの改良による各種割合 (Ballet)

picture	QP	consistency	depth rate [%]	intra rate [%]	disparity rate[%]
P (skip)	28	off	7.9	75.8	16.3
		on	8.5	75.5	16.0
	32	off	8.2	69.5	22.2
		on	10.2	69.1	20.7
	36	off	7.1	63.6	29.2
		on	10.4	61.4	28.2
	40	off	5.8	52.1	42.1
		on	14.9	50.3	34.9
B (direct)	28	off	25.5	13.7	60.8
		on	33.0	13.7	53.3
	32	off	19.5	6.3	74.2
		on	34.7	6.1	59.2
	36	off	15.0	3.1	82.0
		on	37.0	3.6	59.4
	40	off	13.4	2.0	84.7
		on	34.6	1.0	64.4

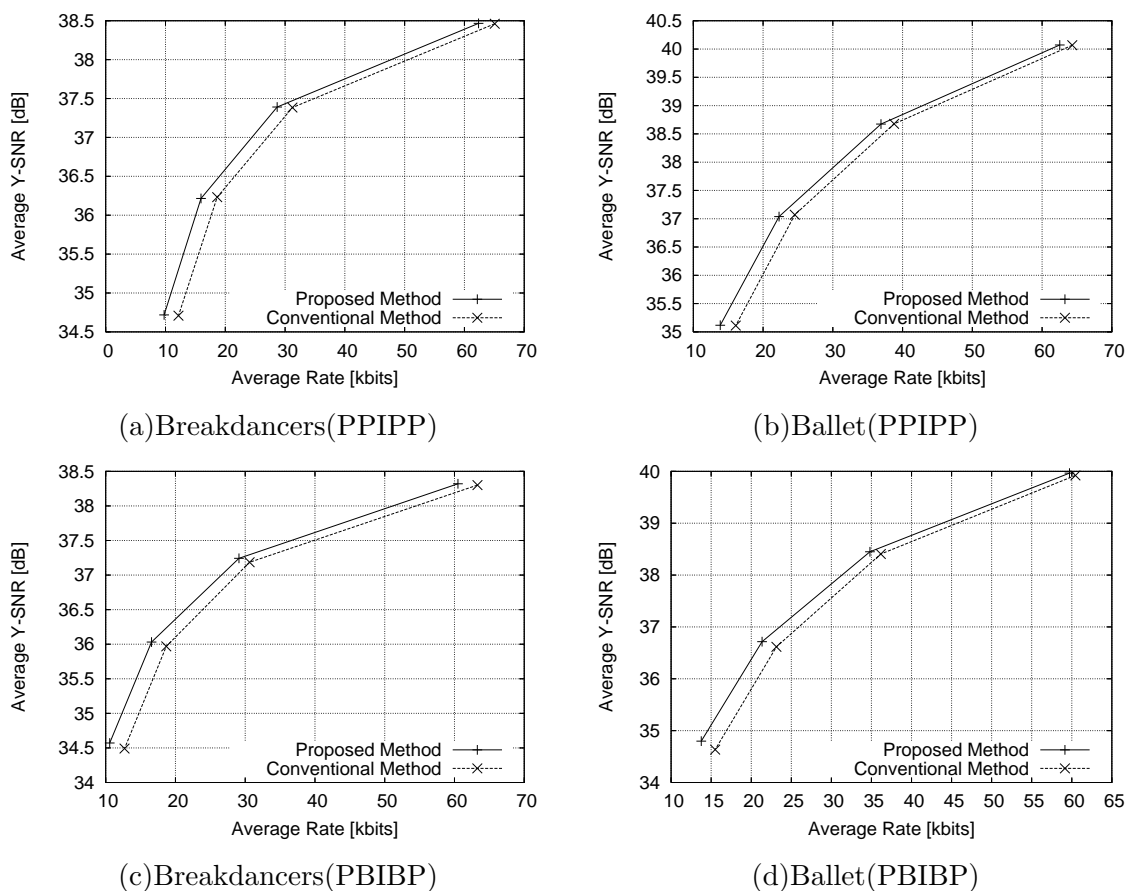


図 5-8: 統合結果

5.3.5 統合結果

まとめとして、従来手法⁽³⁴⁾に対する提案手法の効果を示す。以上の結果から、提案手法は Breakdancers と Ballet においてより効果があることがわかり、これらシーケンスの PPIPP、PBIBP 構造における比較を図 5-8 に示す。横軸は奥行き予測を行うフレームの平均発生符号量、縦軸は平均 SNR である。図から、提案手法を全て併せることで約 0.5[dB] の利得が得られることがわかった。また画質に関しては調査したところ、特に目立ったノイズなどは発生していなかった。

5.4 むすび

本章ではシミュレーションによる提案手法の評価を行った。評価実験は、時間予測を行わないアンカーフレームに対して行った。レート歪みコストに基づく奥行き情報探索、奥行き情報・視差ベクトル変換、奥行き情報の一貫性の利用、スキップ/ダイレクトモードの改良それぞれに対して実験・考察を行った。特に符号化効率改善に寄与した手法は奥行き情報・視差ベクトル変換処理であった。全ての手法を統合した結果は従来手法に対して、最大約 0.5[dB] の利得が得られることがわかった。

第6章

結論

本章では本論文を総括し、今後の検討課題を述べる。

6.1 総括

本研究では、View Interpolation を符号化に取り入れ、奥行き情報をエンコーダ側で生成・符号化する多視点映像符号化の効率改善を目的とし、奥行き情報の特性に着目した符号化効率の改善を行った。時間方向予測と視点間予測の両者が使用可能な場合には、基本的に時間方向予測が選ばれるという予備実験から、時間方向予測を行わないアンカーフレームにおいて提案手法の有効性を示した。これにより、序論で設定した多視点映像符号化の更なる圧縮効率の向上という目的に対し、時間方向の予測を行わないフレームにおいて、より少ない符号量でデータを表現できることとなった。また提案手法は、時間方向のランダムアクセスを必要とするような環境において発生符号量を抑えることができる手法であるといえる。

第1章では本研究の背景と目的、および本論文の構成について述べた。

第2章では本研究で想定する多視点映像システムについて述べた。まず、多視点画像を扱う上で基本となる多視点幾何について述べ、システムの構成要素である取得、補正、符号化、画像ベースのレンダリング、表示の各要素について説明した。この中で特に符号化は本研究の対象とする領域であり H.264/AVC を中心に説明した。

第3章では多視点映像符号化の特性を調査するために、まず時間相関と視点相関について、それぞれの相関のみを利用した符号化を行い、時間相関を用いる場合が符号化効率が高くなることがわかった。次に、視点間予測のみを行う場合において、予測構造変化させて符号化した場合には、Iピクチャをカメラ配置の重心に配置する構造が符号化効率が高くなることがわかった。またカメラ間の輝度・色差を補正するために H.264/AVC の重み付け予測を適用した場合には、必ずしも全てのシーケンスに対して有効ではないが、基本的に符号化効率を向上させることができることがわかった。最後に、時間方向予測と視点間予測の両者が使用可能な場合には、時間方向予測を行わないアンカーフレームにおいて利得が大きく、隣接する時間方向のフレームから予測が可能な場合には、視点間予測の効果は薄くなることがわかった。

第4章では奥行き情報を探索して符号化する View Interpolation を用いた多視点映像符号化における符号化効率改善手法を提案した。3章の結果から、視点間予測の符号化効率向上に焦点を絞り、レート歪みコストに基づく奥行き情報の探索方法、奥行き情報と視差ベクトルの相互変換による予測効率の向上、奥行き情報の一貫性を利用した B ピクチャ符号化の予測効率改善手法、周囲ブロックからの多数決処理によるスキップ/ダイレクトモード時の予測方式の改良といった4種類の符号化効率改善案を提案した。

第5章では評価実験を行い、提案手法のそれぞれの手法の効率に関して結果を述べ、考察を行った。4種類の手法のうち特に効果が表れた手法は、奥行き情報・視差ベクトル変換であり、特に奥行き予測の選択割合が減っていたことから、奥行き情報から視差ベクトルへの変換の効果が大きいことがわかった。4つの手法を組み合わせることで、奥行き予測を行うフレームにおいて、従来手法に比べ最大 0.5[dB] の利得が得られることを示した。

第6章は結論であり、本論文の総括および今後の課題について述べている。

6.2 今後の課題

今後検討すべき課題を以下に示す。

- 可変奥行き情報ステップ

本研究では、奥行き情報 z を $minZ$ から $maxZ$ まで $stepZ$ ずつ増やしながらか探索を行うが、 z と対応する視差ベクトルは反比例の関係にあり、一様に $stepZ$ ずつ増やしたのでは、重複して探索される座標や逆に探索されない座標が発生していると考えられる。そこで、 z の値が小さいときは $stepZ$ を小さく、 z の値が大きいときは $stepZ$ を大きくすることで効率が改善する可能性がある。しかし、 z の値はシーンの中間値で多く発生することから、 $minZ$ と $maxZ$ の中間値から $minZ$ 、 $maxZ$ に向けて徐々に粗くしていく、というような探索方法も考えられる。

- 双奥行き予測

奥行き情報の一貫性の利用は、符号化対象カメラ座標において、どの参照フレームに対して Warping しようと z の値は不変であるという特性を利用したが、B ピクチャ符号化時に、L0 参照フレームと L1 参照フレームの z 値を同じにするという制約を課しながら奥行き情報を探索する、ということも考えられる。この場合、L0 参照フレームか L1 参照フレームのどちらかに対する z のみを符号化すればよいため、奥行き情報量削減につながる場合があると考えられる。

- 照度補償

カメラ間の照度差を補正する目的で H.264/AVC の重み付け予測を適用したが、中には悪化するシーケンスも表れたことから、必ずしも H.264/AVC の重み付け予測が有効であるわけではない。式 (3.4) (3.5) における a 、 b の値をより適応的に求めることで、更なる符号化効率の向上が見込めると考えられる。

- 動画像への対応

本研究では，3章での予備実験の結果を受け，評価実験では視点間予測のみで検討を行った．視点間予測のみで最大 0.5[dB] 程度の利得であるため，動画像へ適用した場合にはこれら利得は小さくなることが考えられる．そこで，動きベクトルとの兼ね合いを念頭に検討する必要があると考えられる．

- 高速化

式(2.10)からわかるとおり，1画素の対応点候補座標を求めるためには， 3×3 の行列と3次元ベクトルとの行列演算，3次元ベクトル同士の加算，同次座標から画像座標への変換のための除算2回が必要である．これをブロックサイズ分行ってある奥行き候補に対する1ブロック分の対応点が求まるため，従来の動き推定処理と比較すると，演算量の増加は非常に大きなものとなり，高速化は避けられない．そこで，式(2.10)自体を高速化したり，探索範囲を狭める手法が有効であると考えられる．探索範囲を狭めるには，周囲ブロックからの奥行きの予測値を中心により狭い探索範囲で探索する手法，階層的に奥行き情報を探索する手法，視差ベクトル探索で得られた値を奥行き情報に変換してその周囲を探索する手法などが考えられる．

謝辞

本研究の機会及び素晴らしい実験環境を与えて下さり、貴重な時間を割いて研究の方向性を御指導頂きました渡辺 裕 教授に心から感謝致します。

学部時代の指導教授であり、研究の方向性について有意義な御助言を賜りました早稲田大学国際情報通信研究科 富永 英義 教授に深く感謝致します。

本研究のきっかけを与えて下さり、研究の進め方や研究の方向性、研究の細部に至るまで、数々の有意義な御意見を賜りました株式会社 KDDI 研究所 内藤 整 博士に深く感謝致します。

貴重な御意見、様々な御提案を頂いた渡辺研究室の皆様に御礼申し上げます。研究の方向性に悩む私を常日頃から励まして頂いた石川 孝明 氏に心から感謝致します。研究の相談に快く応じた確かな助言を頂いた河村 圭 氏に心から感謝致します。気晴らしの遊びに付き合ってくれた伊谷 裕介 氏に心から感謝致します。私の独り言を快く聞いてくれた加藤 幸一 氏に心から感謝致します。行事などの際に率先して引導してくれた山本 勇樹 氏に心から感謝致します。

富永研究室の皆様に、御礼申し上げます。特に、3年間共に励まし合い切磋琢磨して頑張った同期の友人に感謝致します。

最後に、私をここまで育てて下さった家族に深く感謝します。

平成 19 年 3 月 15 日

後藤 崇行

参考文献

- (1) 尾上守夫, 池内克史, 羽倉弘之, 3次元映像ハンドブック, 朝倉書店, 2006.
- (2) 谷本正幸, 藤井俊彰, “自由視点映像技術,” 映像情報メディア学会誌, Vol.60, No.1, pp.29-34, January 2006.
- (3) O.Schreer, P.Kauff and T.Sikora, 3D Video Communication -Algorithms, concepts and real-time systems in human centred communication-, Wiley, 2005.
- (4) <http://www.cs.cmu.edu/virtualized-reality/>
- (5) <http://www.imi.coe.nagoya-u.ac.jp/>
- (6) <http://graphics.stanford.edu/papers/CameraArray/>
- (7) ITU-T Recommendation H.261, International Standard: “Video Codec for Audiovisual Services at px64kbit/s,” 1990.
- (8) ISO/IEC 11172-2, International Standard: “Coding of Moving Pictures and Associated Audio for Digital Storage Media at up to about 1.5 Mbits/s,” 1993.
- (9) ISO/IEC 13818-2, International Standard: “Information Technology- Generic Coding of Moving Pictures and Associated Audio Information:Video,” 1995.
- (10) ITU-T Recommendation H.263, International Standard: “Video Coding for Low Bit Rate Communication,” 1996.
- (11) ISO/IEC 14496-2, International Standard: “Information Technology- Coding of Audiovisual Objects,” 1998.
- (12) ISO/IEC 14496-10, International Standard: “Advanced Video Coding for Generic Audiovisual Services,” (ITU-T Rec.H.264), version 3:2005.
- (13) 佐藤淳, コンピュータビジョン –視覚の幾何学–, コロナ社, 1999.
- (14) 徐剛, 辻三郎, 3次元ビジョン, 共立出版, 1998.
- (15) 出口光一郎, ロボットビジョンの基礎, コロナ社, 2000.
- (16) E.H.Adelson and J.R.Bergen, “The Plenoptic Function and the Elements of Early Vision,” Computational Models of Visual Processing, MIT Press, pp.3-20, 1991.
- (17) J.-X.Chai, X.Tong, S.-C.Chan and H.-Y.Shum, “Plenoptic Sampling,” Proceedings of ACM SIGGRAPH, pp.307-318, 2000.
- (18) Z.Zhang, “A Flexible New Technique for Camera Calibration,” Microsoft Research Technical Report, MSR-TR-98-71, December 1998.

- (19) R.Y.Tsai, "A Versatile Camera Calibration Technique for High-Accuracy 3D Machine Vision Metrology Using Off-the-Shelf TV Cameras and Lenses," *IEEE Journal of Robotics and Automation*, Vol.Ra-3, No.4, pp.323-344, August 1987.
- (20) R.Hartley and A.Zisserman, "Multiple View Geometry in Computer Vision," Cambridge University Press, 2000.
- (21) A.Fusiello, E.Trucco and A.Verri, "A compact algorithm for rectification of stereo pairs," *Machine Vision and Applications* 12, pp.16-22, March 2000.
- (22) 山本健詞, 圓道知博, 藤井俊彰, 谷本正幸, "自由視点テレビのためのマルチカメラの画像変換," *映像情報メディア学会誌*, Vol.60, No.2, pp.219-226, February 2006.
- (23) 大久保榮, 角野眞也, 菊池義浩, 鈴木輝彦, *H.264/AVC 教科書*, インプレス, 2004.
- (24) ISO/IEC 13818-2 Amendment 3: "Information Technology -Generic Coding of Moving Pictures and Accosiated Audio:Video," 1996.
- (25) 木全英明, "MPEG 3DAV 国際標準化の動向," *映像情報メディア学会誌*, Vol.60, No.2, pp.143-149, February 2006.
- (26) M.E.Lukacs, "Predictive Coding of Multi-Viewpoint Image Sets," *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Vol.11, pp.521-524, April 1986.
- (27) M.Magnor and B.Girod, "Hierarchical Coding of Light Fields with Disparity Maps," *IEEE International Conference on Image Processing (ICIP)*, Vol.3, pp.334-338, October 1999.
- (28) H.Kimata, M.Kitahara, K.Kamikura, Y.Yashima, T.Fujii and M.Tanimoto, "System Design of Free Viewpoint Video Communication," *IEEE International Conference on Computer and Information Technology (CIT)*, pp.52-59, September 2004.
- (29) H.Kimata, M.Kitahara, K.Kamikura and Y.Yashima, "Multi-View Video Coding Using Reference Picture Selection for Free-Viewpoint Video Communication," *Picture Coding Symposium (PCS)*, December 2004
- (30) 石川彰夫, 川田亮一, 小池淳, "H.264 の多視点符号化拡張における視差補償予測構造に関する一検討," *電子情報通信学会総合大会*, D-11-43, March 2006 .
- (31) 石川彰男, 酒澤茂之, 小池淳, "2次元カメラ配列の多視点映像符号化における予測構造に関する一検討," *情報科学技術フォーラム (FIT)*, J-045, pp.291-292, September 2006.

- (32) 志水信哉, 木全英明, 上倉一人, 八島由幸, “参照フレームのベクトルを用いた多視点映像符号化のためのベクトル予測,” 画像符号化シンポジウム (PCSJ), P-1.02, November 2006.
- (33) 木全英明, 北原正樹, 上倉一人, 八島由幸, 藤井俊彰, 谷本正幸, “自由視点映像通信のための低遅延多視点映像符号化,” 電子情報通信学会論文誌 D, Vol.J89-D, No.1, pp.40-55, January 2006.
- (34) S.Yea, J.Oh, S.Ince, E.Martinian and A.Vetro, “Report on Core Experiments CE3 of Multiview Coding,” ISO/IEC JTC1/SC29/WG11 Document M13695, July 2006.
- (35) M.Tanimoto, T.Fujii, K.Yamamoto, M.Kitahara, H.Kimata, K.Kamikura and Y.Yashima, “Response to Call for Proposals on Multi-view Video Coding: Overview and Coding Tool,” ISO/IEC JTC1/SC29/WG11 Document M12969, January 2006.
- (36) M.Kitahara, H.Kimata, M.Tanimoto, T.Fujii and K.Yamamoto, “Report of Core Experiment on View Interpolation (Multi-view Video Coding CE3),” ISO/IEC JTC1/SC29/WG11 Document M13509, July 2006.
- (37) M.Droese, T.Fujii and M.Tanimoto, “Ray-Space Interpolation based on Filtering in Disparity Domain,” 3D Image Conference, pp.231-236, June 2004.
- (38) 志水信哉, 北原正樹, 上倉一人, 八島由幸, “奥行き情報を用いた多視点映像符号化に関する一検討,” 画像符号化シンポジウム (PCSJ), P-3.01, November 2005.
- (39) 秋山修, 斎藤英雄, “視点間の射影幾何的關係を利用した多視点動画像のデータ圧縮,” 映像情報メディア学会誌, Vol.59, No.1, pp.136-145, January 2005.
- (40) 田口裕一, 苗村健, “自由視点画像合成に基づく光線空間符号化手法の提案と評価方法の検討,” 情報科学技術フォーラム (FIT), J-009, pp.189-192, September 2005.
- (41) H.-Y.Shum, S.B.Kang and S.-C.Chan, “Survey of Image-Based Representations and Compression Techniques,” IEEE Transactions on Circuits and Systems for Video Technology, Vol.13, No.11, pp.1020-1037, November 2003.
- (42) C.Buehler, M.Bosse, L.McMillan, S.Gortler and M.Cohen, “Unstructured Lumigraph Rendering,” Proceedings of ACM SIGGRAPH, pp.425-432, August 2001.
- (43) L.McMillan and U.Bishop, “Plenoptic Modeling: An Image-based Rendering System,” Proceedings of ACM SIGGRAPH, pp.39-46, August 1995.
- (44) M.Levoy and P.Hanrahan, “Light Field Rendering,” Proceedings of ACM SIGGRAPH, pp.31-42, August 1996.

- (45) S.J.Gortler, R.Grzeszczuk, R.Szeliski and M.F.Cohen, "The Lumigraph," Proceedings of ACM SIGGRAPH, pp.43-52, August 1996.
- (46) H.-Y.Shum and R.Szeliski, "Rendering with Concentric Mosaics," Proceedings of ACM SIGGRAPH, pp.299-306, August 1999.
- (47) S.Chen and L.Williams, "View Interpolation for Image Synthesis," Proceedings of ACM SIGGRAPH, pp.279-288, August 1993.
- (48) D.Scharstein and R.Szeliski and "A Taxonomy and Evaluation of Dense Two-Frame Stereo Correspondence Algorithms," Microsoft Research Technical Report, MSR-TR-2001-81, November 2001.
- (49) S.M.Seitz and C.M.Dyer, "View Morphing," Proceedings of ACM SIGGRAPH, pp.21-30, August 1996.
- (50) P.E.Debevec, C.J.Taylor and J.Malik, "Modeling and Rendering Architecture from Photographs: A Hybrid Geometry- and Image-based Approach," Proceedings of ACM SIGGRAPH, pp.11-20, August 1996.
- (51) J.Shade, S.Gortler, L.-W.He and R.Szeliski, "Layered Depth Images," Proceedings of ACM SIGGRAPH, pp.231-242, August 1998.
- (52) D.N.Wood, D.I.Azuma, K.Aldinger, B.Curless, T.Duchamp, D.H.Salesin and W.Stuetzle, "Surface Light Fields for 3D Photography," Proceedings of ACM SIGGRAPH, pp.287-296, August 2000.
- (53) "Description of Core Experiments in MVC," ISO/IEC JTC1/SC29/WG11 N7798, January 2006.
- (54) Joint Model (H.264/AVC Reference Software), <http://iphone.hhi.de/suehring/tml/>
- (55) Joint Multi-view Video Model, CVS Repository
- (56) K.Kamikura, H.Watanabe, H.Jozawa, H.Kotera and S.Ichinose, "Global Brightness-Variation Compensation for Video Coding," IEEE Transactions on Circuits and Systems for Video Technology, Vol.8, No.8, December 1998.
- (57) G.J.Sullivan and T.Wiegand, "Rate-Distortion Optimization for Video Compression," IEEE Signal Processing Magazine, pp.74-90, November 1998.
- (58) A.Behrens, E.Martinian, H.Sun and A.Vetro, "Comments on Call for Proposals on Multiview Video Coding," ISO/IEC JTC1/SC29/WG11 Document M12230, July 2005.

図一覧

1-1	100 台のカメラによる撮影 ⁽⁵⁾	3
2-1	ピンホールカメラモデル	6
2-2	エピポーラ幾何	7
2-3	カメラ配置の種類	10
2-4	ハイブリッド符号化	14
2-5	Light Field Rendering/Lumigraph	21
2-6	パララックスバリア方式	25
3-1	使用するテストシーケンス	29
3-2	視点間相関符号化と時間相関符号化の予測構造	31
3-3	視点間予測のみと時間予測のみ	32
3-4	視点間予測の構造	33
3-5	予測構造の違い	34
3-6	重み付け予測の効果	36
3-7	時間・視点適応予測符号化の予測構造 (GOP=12)	37
3-8	時間・視点適応予測符号化とサイマルキャスト符号化	38
3-9	時間間隔の違いによる視点間予測の効果 (Ballroom)	39
4-1	H.264/AVC に奥行き情報探索を付加した符号化器	42
4-2	奥行き情報予測値の生成に使用する周辺ブロック	44
4-3	奥行き情報・視差ベクトル変換の概念	45
4-4	奥行き情報・視差ベクトル変換の適用ブロック	47
4-5	奥行き情報の一貫性の利用による流れ	48
5-1	奥行き情報探索コストの違い (PPIPP)	55
5-2	奥行き情報探索コストの違い (PBIBP)	56
5-3	奥行き情報・視差ベクトル変換 (PPIPP)	58
5-4	奥行き情報・視差ベクトル変換 (PBIBP)	59
5-5	奥行き情報の一貫性の利用 (PBIBP)	64
5-6	スキップモードの改良 (PBIBP)	66
5-7	ダイレクトモードの改良 (PBIBP)	66
5-8	統合結果	68

表一覧

2-1	各種標準方式の比較	15
3-1	視点方向又は時間方向の MAD	30
3-2	符号化条件 (JM)	30
3-3	符号化条件 (JMVM)	37
5-1	符号化条件	52
5-2	奥行き情報探索の範囲	53
5-3	奥行き情報探索コストの違いによる各種割合 (PBIBP)	54
5-4	奥行き情報・視差ベクトル変換による各種割合 (PPIPP)	60
5-5	奥行き情報・視差ベクトル変換によるモードの割合 (PPIPP)	61
5-6	奥行き情報の一貫性の利用による各種割合 (Ballet)	63
5-7	奥行き情報の一貫性の利用によるモードの割合 (Ballet)	64
5-8	スキップ/ダイレクトモードの改良によるスキップ/ダイレクトモードの割合 (Ballet)	67
5-9	スキップ/ダイレクトモードの改良による各種割合 (Ballet)	67

研究業績

	題 名	発表年月	発表掲載誌	連名者
国内				
(1)	フィルムグレインを考慮した H.264/AVC モード決定手法に関する一検討	2005 年 3 月	信学春季全大	後藤 崇行 常松 祐一 渡辺 裕 富永 英義
(2)	フィルムグレインを考慮した H.264/AVC 符号化方式に関する一検討	2005 年 7 月	情報処理学会 AVM 研究会	後藤 崇行 常松 祐一 渡辺 裕
(3)	H.264 における動き補償ブロックサイズ高速決定手法に関する一検討	2005 年 7 月	情報処理学会 AVM 研究会	常松 祐一 後藤 崇行 渡辺 裕
(4)	動き補償付時間方向フィルタリングにおける適応的 Update 手法に関する一検討	2005 年 11 月	画像符号化シンポジウム PCSJ 2005	後藤 崇行 石川 孝明 渡辺 裕
(5)	奥行き情報を用いた多視点映像符号化の符号量削減	2006 年 11 月	画像符号化シンポジウム PCSJ 2006	後藤 崇行 内藤 整 渡辺 裕
(6)	View Interpolation を用いた多視点符号化の効率改善に関する一検討 (発表予定)	2007 年 3 月	信学春季総合大会	後藤 崇行 内藤 整 渡辺 裕