2005－AVM－50（7）
2005／10／7

Disparity Map を用いた多視点動画像符号化に関する一検討

志水信哉†，北原正樹†，上倉一人†，八島由幸†
†日本電信電話株式会社　NTT サイバースペース研究所

本稿では，多視点動画像を基準視点動画像と Disparity Map を用いて符号化する手法を提案する．ここでの Disparity Map とは，通常の視差補償とは逆に，参照フレームとなる基準視点動画像のマクロブロック毎に求めた全ての視点に対する視差補償情報である．これによって視差補償ベクトル間に存在する相関を利用して高効率な符号化を達成する．ある多視点動画像に対して本手法を適用したところ，多視点動画像を独立して符号化するよりも，平均 PSNR が 37dB 付近で約２倍の符号化効率を達成できることを確認した．

A Study on Encoding of Multi-view Video Using Disparity Maps

Shinya SHIMIZU†, Masaki KITAHARA†, Kazuto KAMIKURA†, Yoshiyuki YASHIMA†
†NTT Cyber Space Laboratories, NTT Corporation

In this paper, we propose a method of encoding multi-view video using base viewpoint video and disparity maps. Disparity map is a set of information for disparity compensation. Contrary to general disparity compensation, this disparity information is calculated at every macro block of reference frame. This reduces redundant information existing between compensation vectors, so that this method achieves high performance video compression. As a result of the experiment, this method achieved about half rate compared with the individual encoding when PSNR is around 37dB.

1. はじめに

近年, 映像の高画質化や高解像度化が進む一方で, 演算装置の発展に伴う計算処理能力の向上や, 光ネットワークなどブロードバンド化に伴う通信帯域の拡大と双方向通信の実現により, これらの能力を生かした新しい映像に関する研究開発が盛んである. そのような次世代の映像の１つとして自由視点映像と呼ばれる映像に高い関心が集められている．自由視点映像とは，これまでのように全ての利用者が配信側で定められた唯一の視点からの映像を見るような受動的な映像と異なり, 利用者がそれぞれ自由に視点位置を変えて所望の視点からの映像を楽しむことが出来るような映像である．

このような自由視点映像を作り出す技術に関しては, これまでに数多くの研究が行なわれている[1]．また合成される映像の品質の向上に伴い, これらの技術が実際に利用された例もある. 最も有名なものとしては, 映画 The Matrix のなかで, この種の技術によって時間が静止した状態で視点が動く映像が作り出されている．この場合，合成処理はオフラインで行なわれているが, リアルタイムに合成処理が行なえるようなシステムも研究されている．2001 年には，Eye-Vision を用いることでスーパーボウルの中継映像が自由視点合成されながら実際に放送された．

自由視点映像のデータとしては, 合成した際により高画質な映像が得られるという点から, 同じシーンを複数のカメラで撮影した多視点動画像が用いられることが多い. 視点移動の範囲を広げたり，映像の品質を向上させたりするには，多くのカメラを設置する必要がある[2]．そのため，自由視点映像のための多視点動画像は非常に膨大なデータ量となる．したがって，自由視点映像配信の実現には多視点動画像の効率的な符号化が必要不可欠である．最近では，この多視点動画像の符号化は，国際標準化団体 MPEG で国際標準規格の検討が開始されている[3].

多視点動画像を用いて自由視点映像を実現させる場合, 多視点動画像の符号化には高圧縮という性能のほかに, 任意の視点へのランダムアクセス性や任意の視点の映像を出来るだけ少ない視点の映像だけで部分復号できるような機能が重

志水信哉
〒239-0847 神奈川県横須賀市光の丘 1-1
日本電信電話株式会社　NTT サイバースペース研究所
shimizu.shinya@lab.ntt.co.jp

要であると言われている[4]．高いランダムアクセス機能は表示する視点位置が変わるとその視点の映像を合成するのに必要なカメラ映像が替わるため, 操作への早いレスポンスを実現するために必要不可欠である. 任意の視点の映像を部分復号できる機能は, 視点合成には常に全てのカメラの映像が必要なわけではないため, 利用者に要求する通信帯域を削減したり, デコード負荷を軽減したりするために必要となる. 特にカメラの数が膨大な場合, この部分復号できる機能は自由視点映像を実現するためには欠かすことはできない機能となる.

これまでに様々な多視点画像の符号化方式が提案されている. それらの技術の多くは視差補償と呼ばれる手法を基に形成されている. 視差補償とは，MPEG-2 や MPEG-4 などの通常の動画像符号化において用いられる, ブロック毎にフレーム間の動きを予測して符号化効率を向上させる手法である動き補償を, カメラ間の相関を利用するために, 撮影カメラの異なるフレーム間でブロックマッチングを行い, 符号化効率を向上させる手法である．この手法は MPEG-2 Multiview Profile においてテレオ動画像を符号化するツールとして採用されている.

JeongEun らは，この MPEG-2 Multiview Profile を応用して視点映像間で予測符号化を適用して複数の視点の映像をまとめて符号化する仕組みを提案している[5]．この符号化の枠組は多視点動画像からステレオ動画像に視点数を変更する View Scalability を実現しているが, 全ての視点動画像で同じ GOP を構成するため，低遅延な視点映像間のランダムアクセスを実現できない.

岡らは時間的に近い２フレームと, 空間的に近い２フレームの合計４フレームを参照画像として符号化可能なMピクチャというものを導入して符号化効率を向上させている[6]．しかし，この方法では M ピクチャとして符号化されたフレームを復号するためには, 多くのフレームを復号する必要があるため, 柔軟なランダムアクセスを実現することができない.

木全らは GoGOP(Group of GOP)構造を導入して高いランダムアクセス性と所望の視点映像を得るために必要な映像のみを復号する(Partial Decoding)や, 必要なデータのみを伝送する方法(View Scalability)を実現している[7]．GoGOP の概念は GOP を BaseGOP と InterGOP に分類し, BaseGOP では GOP 内だけで参照を許可し, InterGOP では GOP 内だけでなく他の BaseGOP や InterGOP を参照して符号化するというものである. この手法では InterGOP の数によってランダムアクセス性などの機能と符号化効率が変化する. つまり柔軟なランダムアクセス性を実現するためにはある程度符号化効率を犠牲にする必要がある.

また, これまでに提案されている手法ではブロック毎に動き補償または視差補償のどちらか一方を選択して利用している．このことは，時間方向と空間方向とに同時に存在する冗長性を十分に取り除くこと出来ていないことを示す．もし，多視点動画像に特有な性質を用いることで, 時間方向と空間方向の２つの相関を同時に利用することができれば, 更に効率的な多視点動画像の符号化が可能であると考えられる.

本稿では, 自由視点映像通信のための動画像符号化として, 高いランダムアクセス機能や柔軟に映像を部分復号できる機能を実現し, かつ高い符号化効率を実現する多視点動画像符号化方式として, Disparity Map を用いた符号化方式について検討を行なう．第２章で今回検討を行なった Disparity Map を用いた符号化方式についてまとめ, 第３章ではある多視点動画像に対して行なった実験結果を報告し, 第４章でまとめと今後の課題を述べて本稿を締めくくる.

2. Disparity Map を用いた多視点動画像符号化

本稿で提案する手法は, 基準視点と呼ぶ１つの動画像から非基準視点と呼ぶ基準視点以外の動画像を予測して符号化を行なう．本手法は，１）基準視点の選出及び符号化，２）Disparity Map の生成及び符号化，３）非基準視点動画像の符号化，という３つのプロセスからなる．以下，各プロセスについて詳しく説明する．また，簡単のために各動画像を撮影するカメラは全て同じ焦点距離を持ち，光軸が全て同じ向きであり，焦点が同一平面上に存在し, その位置関係が既知で固定されているものとする.

2.1. 基準視点の選出及び符号化

提案手法では, 基準視点動画像から非基準視点動画像を予測し，その差分を符号化する．そのた

め予測効率が最大となるものを基準視点とすることが望ましい. あるカメラの動画像から別のカメラの動画像を予測する際, 同じカメラ位置でない限り必ず予測不可能な空間が存在する. 予測できないということが最も予測精度が悪いため, 予測できない空間ができるだけ少なくなるように基準視点を選ぶ. つまり, 基準視点のカメラの撮影空間と他のカメラの撮影空間の共通する空間の大きさをカメラ毎に求め, その和が最大となるものを基準視点とする.

基準視点動画像の符号化は, 単体で考えると普通の動画像符号化となるため通常の符号化方式（本稿では H.264）を用いて符号化する.

2.2. Disparity Map の生成及び符号化

提案手法では, 基準視点動画像から非基準視点動画像を予測するための付加情報として視差情報を用いる. 通常の視差情報を用いた視点間予測の視差補償は符号化対象フレームのマクロブロック（以下, MB と呼ぶ）毎に求めるのだが, 本手法ではそれとは逆に, 参照フレームとなる基準視点のフレームの MB 毎に求める. 被写体の表面が完全拡散反射であると仮定すると, 基準視点と各カメラとの視差ベクトルには以下の関係式が成り立つ.

$$L\begin{pmatrix} D_x \\ D_y \end{pmatrix} = f\begin{pmatrix} \delta x \\ \delta y \end{pmatrix} \quad \cdots\cdots \quad (1)$$

この式で(Dx, Dy)は視差ベクトル, (δx, δy)が基準視点と補償対象視点との間の距離ベクトル, f がカメラの焦点距離であるとするとは, L が基準視点の MB 毎に求まる定数である. つまり基準視点のある MB に対して求まる各非基準視点への視差ベクトルは１つのスカラー値で表すことができる. 従って基準視点から非基準視点の予測に必要な情報は Disparity Map（以下, DM と呼ぶ）と呼ぶグレースケール動画像となる. この DM は通常の動画像として符号化を行なう. この符号化による誤差は予測精度に影響を与えるため, 最終的な全体の符号化効率を考えてその符号化精度を変更する必要がある. ただし, 通常は２次元量のベクトルが MB に対して視点毎に存在していたのが, １次元量のスカラー値が MB に対して１つずつしか存在しなくなるため, ロスレスで符号化したとしても従来のベクトルをロスレスで符号化していたものより符号量を削減できる.

2.3. 非基準視点動画像の符号化

非基準視点動画像はフレーム毎に基準視点動画像から DM を用いて視差補償を行なった後, その差分を動画像として符号化する. このとき基準視点動画像や DM を符号化した際の符号化歪みが非基準視点動画像の品質に影響を与えないようにするために, 基準視点動画像や DM は１度符号化した後, 復号したものを使用する.

2.2 節で述べたとおり本手法で用いる視差情報は参照フレームの MB 毎に与えられる. そのため符号化対象フレームのピクセル毎に１つの視差ベクトルが得られるわけではない. ２つ以上の視差ベクトルが得られるピクセルや視差ベクトルのないピクセルが生じる. ２つ以上の視差ベクトルが得られるピクセルに対しては, 視差ベクトルが小さいほど被写体がカメラに近いということから, 最も小さな視差ベクトルをそのピクセルの視差ベクトルとする. 視差ベクトルの存在しないピクセルが存在するということは, 通常の予測と異なりフレーム全体に対して予測画像が生成されないことを意味する. このとこは符号化対象となる残差が大きくなり符号化効率の低下を招く. そこで他のピクセルからフレーム内予測をすることによって, より多くのピクセルに対して予測値が得られるようにする. フレーム内予測は予測値のある近辺のピクセルから行なう. 具体的な方法としては, 左右や上下それぞれの最も近い予測値のあるピクセルの値から線形補完する方法や, 近いピクセル数個の平均を取る方法や, MB ごとに直流成分のみ予測を行なう方法などいくつか存在する. この中からどれか１つの予測方法に固定してしまうことも, 最も符号化効率のよい予測手法を選び, パラメータとして符号に埋め込むことも可能であるが, そのオーバーヘッドとの兼ね合いを考慮する必要がある.

このようにして求められた予測画像と原画との差分画像を符号化するのだが, 通常の動画像符号化のように差分に対して DCT とエントロピ符号化をするのではなく, 動画像として更に動き補償をしてから DCT・エントロピ符号化を行なう. 通常の差分画像のフレーム間には相関が残っておらずフレーム間予測をしても符号化効率がよくなるとはいえない. しかし, 本手法の場合, 予測画像の生成する際に残差が最も小さくなるも

のを予測したのではなく, 多視点動画像の空間的な性質を利用して予測している．そのため，差分画像には時間的な相関が残っていると考えられるので，それを動き補償して符号化する．原画よりビットレンジが２倍の動画像となるため, 本稿では H.264 の High Profile を用いて符号化する．

2.4. 機能と拡張性

提案手法では各視点を符号化する際に, 常に基準視点と DM を使用するため，この２つは常に伝送・復号する必要がある．しかし，非基準視点はこの２つにのみ参照を行なうため, 残りのカメラ映像が必要な場合は, これに加えて必要な視点の差分動画像のみ伝送･復号すればよい. つまり，任意の視点の映像を常に「必要な視点数＋１」個分の符号を取り出し, 伝送または部分復号が可能という機能を実現する．

また，基準視点と DM から予測画像を生成する手順で, 任意の視点の映像をある程度の品質で生成できるため, 様々な視点へのランダムアクセスも可能である．もちろん，それに加えて対応する視点の差分動画像を復号するだけでより品質のよい映像を生成することも可能である．

この章の最初で「カメラは全て同じ焦点距離を持ち，光軸が全て同じ向きであり，焦点が同一平面上に存在し, その位置関係が既知で固定されている」という仮定をしたが，カメラの焦点距離や光軸，焦点位置が仮定と異なっていても Computer Vision(CV)の分野の技術を用いれば(1)の式を変更するだけで対処することができる．もちろん, 視差ベクトルは１つのスカラー量で表されるため，それ以外の部分への変更はない．また，位置関係は固定でなくても，ヘッダーなどにその情報を記載することで対処できる. 位置関係に必要なパラメータは最大 6 個であり, その符号量は全体に比べれば小さい．さらに，位置関係が既知でなくても, ある程度の精度で映像から予測する手法も CV の分野に存在するため, それを用いることで対処できる．

3. 実験及び考察

3.1. 実験条件

提案手法を名古屋大学工学研究科の谷本研究室により提供された Xmas(図 1)[9]のシーケンスに適用した. 使用した多視点動画像のスペックを表 1 にまとめる．

図 1. テスト画像（Xmas）

表 1. テストシーケンススペック

カメラ間隔	水平 3cm
カメラ台数	１１台
解像度	VGA(640x480)
フォーマット	8bit YUV(4:2:0)
フレーム数	100

このシーケンスではカメラが一直線に並んでいるので, 基準視点は全体の中心にある視点となる．本実験では Disparity は 16x16 の MB ごとに求めた．ある時刻における DM は図 2 のように求められた．基準視点画像と DM から生成される予測画像は図 3 のようになり，その差分画像は図 4 のように求まる．

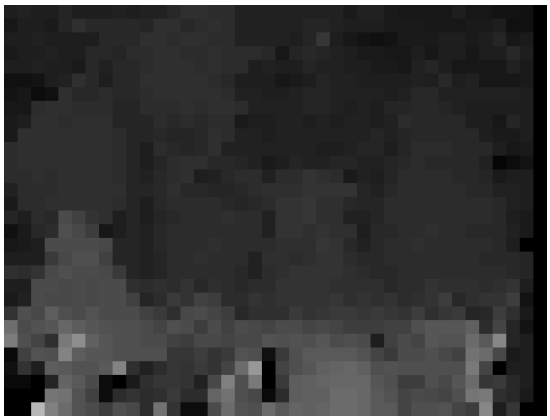

図 2. Disparity Map の例

3.2. 実験結果

基準視点動画像，DM，非基準視点の差分動画像の符号化には H.264 のリファレンスソフトウェア JM9.8 を用いた[8]．また DM は YUV400 の動画像, 差分動画像は９ビットの動画像として

符号化を行なった．提案手法の符号化結果と各カメラ映像を個別に JM9.8 で符号化した場合の結果を図 5 に示す．また図 6 には提案手法で符号化されたある視点の映像を復号した映像を示す．

図 3. 予測画像の例

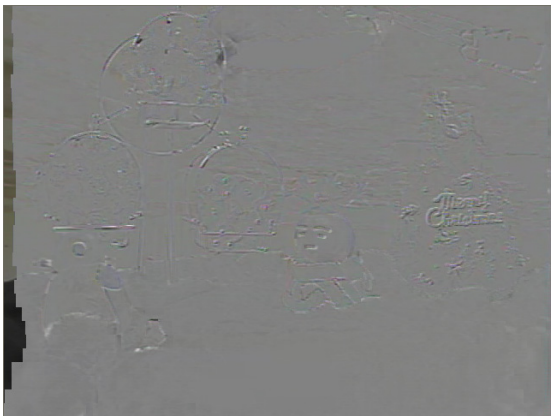

図 4. 差分画像の例

図 5 で示されている符号量は１１個の映像全てを符号化するのに必要な符号量である．つまり，提案手法の符号量は基準視点動画像の符号量，DM の符号量，１０個の差分動画像の符号量の合計である．ただし，DM は全てのレートで同じものを使用し，その符号量は 440704 ビットで，全体の符号量の多くとも１%未満であった．

3.3. 考察

独立に符号化するのに対して提案手法では，PSNR が 37dB 付近で約 50%の符号量を削減することができた．高レートになるに従って，提案手法によるゲインが少なくなっているが，これは基準視点動画像を全ての実験で同じ品質の符号化をしたためであると考えられる．つまり，高レートの部分では予測に使う基準視点動画像の品質が十分でなかったため，差分が大きくなり，その符号量が増加してしまったと考えられる．

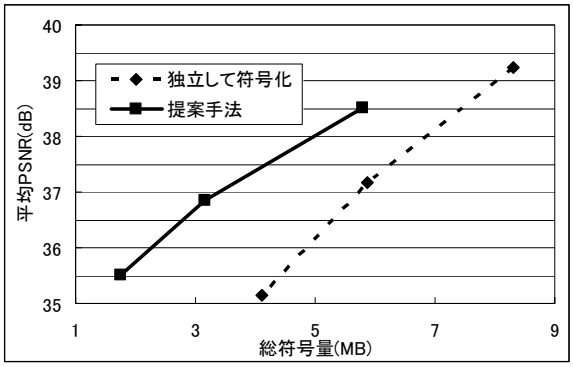

図 5. 実験結果

図 6. デコード結果(PSNR 38.58)

図 7 に各非基準視点の差分動画像に必要であった符号量をまとめる．これによると基準視点に近いほどその符号量が少ないことが分かる．これは基準視点から近いほど DM による予測が当たっていたと言える．15cm はなれた非基準視点での符号量は高レートではほぼ独立して符号化したものと等しくなってしまっているため，提案手法の適用範囲を広げるためには何らかの対処が必要である．

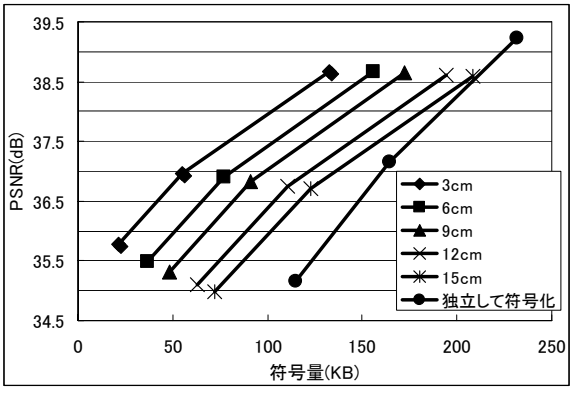

図 7. 基準視点との距離と
差分動画像の符号化効率の関係

4. まとめと今後の課題

本稿では多視点動画像の符号化手法として，Disparity Map を用いた符号化手法を提案した．提案手法によると PSNR が 37dB 付近で各動画像を独立に符号化するのに対して約半分の符号量で符号化が可能であった．また，提案手法では視点間の参照関係が基準視点との間にしかないため，参照関係を増やして符号化効率を向上させる手法と異なり，任意の視点の映像を常に「必要な視点数＋１」個分の符号を取り出し，伝送または部分復号することができるという機能も実現することができている．

なお，本稿で実験として用いたシーケンスはカメラ間隔が密で，被写体との距離が十分あるためあまり多くのオクルージョンがないような映像であった．オクルージョン部分は基準視点の映像からは予測が出来ないため符号化効率が低下してしまうと考えられる．そのため，今後はオクルージョンやカメラ間隔拡大への対処法を考え，提案手法の有効性についてさらなる検討を行なう予定である．また，本稿では Disparity Map や差分動画像の符号化に対してその性質を考慮しないで通常の動画像符号化手法を適用したが，これらには通常の動画像と異なる性質があると考えられるため，それらの符号化手法についても検討を行なっていく予定である．

本稿では簡単に触れただけであるが，基準視点動画像や Disparity Map の符号化精度が全体の符号化効率に影響を与えるため、全体としてのビット配分の方法も検討する必要がある．

謝辞

本研究において実験に利用されたテストシーケンス Xmas[9]を提供していただいた名古屋大学工学研究科の谷本正幸教授ならびに谷本研究室に深く感謝申し上げます．

参考文献

[1] H.-Y.Chum, S.B.Kang, and S.-C.Chan, “Survey of Image-Based Representations and Compression Techniques,” *IEEE Trans, Circ. and Syst. for Video Tech.*, vol. 13, pp. 1020–1037, Nov. 2003.

[2] J.-X.Chai, X.Tong, S.-C.Chan, and H.-Y.Shum, “Plenoptic sampling,” *in Proc. ACM Annu. Computer Graphics Conf.*, pp. 307-318, July 2000.

[3] “Call for Proposals on Multi-view Video Coding,” ISO/IEC JTC1/SC29/WG11 N7327, July 2005.

[4] “Requirements on Multi-view Video Coding v.4,” ISO/IEC JTC1/SC29/WG11 N7282, July 2005.

[5] J.E.Lim, K.N.Ngan, W.Yang, and K.Sohn, “A multiview sequence CODEC with view scalability,” *Signal Processing Image Communication*, vol. 19, pp.239-256, 2004.」

[6] 岡，ナ，藤井，谷本，“自由視点テレビのための動的光線空間の情報圧縮，” *3次元画像コンファレンス 2004*, 5-1, 2004.

[7] 木全，北原，志水，上倉，八島，“自由視点映像通信のための多視点映像符号化の一検討，” *FIT 2004*, pp.225-226, 2004.

[8] JM Software http://iphome.hhi.de/suehring/tml/

[9] Masayuki Tanimoto and Toshiaki Fujii, “Test Sequence for Ray-Space Coding Experiments,” document M10408 MPEG Waikoloa Meeting, 2003.