《4》
評価および標準化動向《4-3》
標準化に向けた取組
～MPEG-FTV標準化の動向～〔正会員〕妹尾孝憲[†]

キーワード 立体映像，自由視点テレビ，多視点映像，視差補償予測，奥行推定

1 ま え が き

自分や人々があたかもその場にいるように感じる超臨場感コミュニケーション¹⁾の実現には，実写でかつ自然な立体映像が不可欠である．立体映像の実現手段には，左右の目に個別の映像を送る2眼立体映像や，これを拡張した多眼立体映像，3次元の被写体から出る光の振幅と位相を再現するホログラフィなどがある²⁾．これらの撮影にはすべて，図1に示すように複数の異なる視点から撮影された多視点映像を利用することができる．多視点映像はまた，受信者が被写体を自由な視点位置から見ることのできる自由視点テレビ (FTV: Free-viewpoint TV)³⁾を実現するためにも利用できるが，多数の視点映像を必要とし，伝送や蓄積の負担が大きいという問題がある．多視点映像を圧縮するため，ISO (International Organization for Standardization) 傘下のMPEG (Moving Picture Expert Group) は，2001年よりFTV符号化方式の標準化を始め，第1フェーズとして2009年にMVC (Multi-view Video Coding) 符号化方式を標準化した⁴⁾．この方式は，従来のMPEG-4 AVCの動き補償フレーム間予測を，視点映像間の視差補償予測にも拡張適用したものである．2007年からは第2フェーズとして，映像の奥行情報を利用して符号化効率を改善する3DV (3-Dimensional Video) 符号化方式の検討を開始した⁵⁾．以下では，これらの標準化取組を紹介するとともに，NICTでの検討について述べる．

2 多視点映像の符号化 (MVC)

2.1 カメラ配置とテスト映像

多視点映像では，図1に見られるように，各カメラには同じ被写体が写っているため，大きな相関がある．多視点映像符号化 (MVC) では，この性質を利用する．テスト映像のカメラ配置には，図2に示すような各種の配置が採用され，テスト映像は標準化参加者から提供された．当初は，カメラのアライメントのずれを画像の射影変換で補正する

図1 多視点映像の撮影

平行化処理を，符号化の中で行うことも検討されたが，コーデックの負担が大きいため，符号化の前に行うことになった．この補正処理には，既知の格子パターン等を撮影してカメラのレンズ歪み補正を行った後，映像中の格子パターンの消失点や無限遠とみなせる被写体点を各視点映像の対応点として求め，この無限遠点の位置が，すべての視点映像で一致するように各視点映像の射影変換行列を求めて，各視点映像を射影変換して，すべてのカメラの方向と内部パラメータを揃え，次に，有限距離にある特徴点の各視点映像での対応点を通るエピポーラ線が平行になるように，共通の射影変換行列を求めて再度射影変換するもの等がある⁶⁾．

2.2 検討された符号化技術

検討された符号化技術を，図3に示す．図3 (a) は，時空間階層双方向予測である．従来の動き補償フレーム予測を，多視点映像間にも適用したもので，視差補償をブロックの平行移動で行い，多視点映像をフレーム間と視点間の両方から予測するものである．相関の高いフレーム間の予測に階層双方向予測を行うことで，符号化効率を上げている．図3 (b) は，照明補償予測方式である．照明の位置やカメラの色感度バラツキによる視点映像間の輝度や色の違いを補償する値を，視点映像の平均値からのオフセット量として加えることにより，予測符号化効率を上げているが，視

[†] 情報通信研究機構 ユニバーサルメディア研究センター
"MPEG-FTV Standardization Activity" by Takatori Senoh (National Institute of Information and Communications Technology, Koganei)

図2 カメラ配置とテスト映像 (口絵カラー参照)

点映像間の色合わせは、符号化の前に行える。図3 (c) は、射影変換予測である。視点映像間で対応点マッチングを行って視差量を奥行きマップ(映像の奥行き)として求め、得られた視差量を用いて射影変換を行って視点映像を予測する。射影変換の方法として、視差量とカメラパラメータから、被写体の3次元座標を求め、得られた3次元の被写体を、予測したい視点映像に射影して予測するものと、予測される視点映像をブロックに分割し、各ブロックは、3次元空間内にある平面の射影であると近似して、Affine変換を一般化した視点映像間の射影行列 (Homography Matrix) を求めて予測するものがあるが、高S/Nを達成するために予測残差を符号化伝送すると、複雑な処理に見合うほど符号

化率は高くない。図3 (d) は、視点映像間の視差ベクトルどうしの予測符号化を行うことで、視差ベクトルのデータ量を減らすものである。視差ベクトルのデータ量は、映像データ量に比べて少なく、符号化効率はあまり上がらない。図3 (e) は、非対称マクロブロック予測である。視差量の異なる被写体毎の形状に合わせてブロック形状を選択することにより、ブロックの予測精度を高める。ブロック形状が被写体に完全には一致せず、予測効率は余り高くない。図3 (f) は、適応予測フィルタである。ブロック単位の視差補償予測を行うと、オブジェクトの境界付近のブロック歪みが大きくなるので、適応的にローパスフィルタを掛けて平滑化するポストフィルタである。さらに、視差ベク

符号化順序: T0の(S0~S7), S0の(T1~T8), S1の(T1~T8), ... S7の(T1~T8)の順. ランダムアクセスには8フレームの遅延あり.

(a) 時空間動き保証予測

(b) 照明補償予測

(c) 射影変換予測

(d) 視差ベクトル予測

(e) 非対称マクロブロック予測

(f) 適応予測フィルタ

図3 MVCに提案された各種の符号化技術(口絵カラー参照)

トルを用いて隣の視点映像の動きベクトルを予測する技術も検討されたが、符号化効率の改善量は0.5dB程度である。

2.3 標準化された方式

多視点映像符号化方式 (MVC) として、最終的に図3 (a) の時空間階層双方向予測技術のみが採用された。異なる視点映像間では、被写体の見える角度が異なっていたり、前景に隠れて背景が見えなくなるオクルージョンがあるので、MVCの符号化効率は、従来のMPEG-4 AVCで全視点映像を並列符号化する場合の約1.5倍程度である。MVCは、高密度光ディスク (Blu-Ray Disc™) 用のステレオ3D映像符号化方式として採用された。

3 立体映像の符号化 (3DV)

立体映像符号化 (3DV) では、符号化効率を上げるため、図4に示すように、すべての視点映像を符号化する代わりに、少数の視点映像とその奥行情報 (奥行マップ) を符号化し、復号側で奥行情報を元に射影変換を行って、符号化されなかった視点映像を合成することでデータ量を削減する。この方法は、カメラのアライメントや色感度が正確に揃えられていて、奥行き情報が正確であれば、高精度に任意の視点からの映像を合成できる。また、奥行情報を用いると、図5に示すように、被写体から出る光の波面を計算可能で、これを用いて究極の立体映像である電子ホログラフィも実現できる⁷⁾。図5では、ホログラムを作るための参照光はホログラム面に垂直に当たっていると仮定した。これらの応用のためには、多視点映像から高精度に奥行情報を抽出する必要がある。

3.1 多視点映像の奥行

多視点映像とその奥行には、図6に示すような関係がある。図6は、同じ焦点距離 f のカメラがベースライン X 軸上に等間隔 L で水平かつ平行に並んでいる場合である。図6で、被写体上の注目点 P までのベースラインからの距離 D と、各カメラ映像上での光軸から対応画素位置までの距離 x_1 , x_2 には次式の関係がある。

$$x_1 + x_2 = \frac{fL}{D} \quad (1)$$

図6では、簡単のために注目点の高さを $Y=0$ としたが、式 (1) は Y の値によらず成立する。対応画素位置のズレ量 ($x_1 + x_2$) を視差量と言い、多視点映像の奥行値として使われることが多い。被写体の奥行を視差量で表すと、あるカメラ映像の画素をその視差量だけずらして、隣のカメラ映像を容易に合成できたり、立体映像として表示する場合の映像の奥行値として使える等の利点がある。以下では、視差量で表された値を奥行値として使う。奥行値の精度を保つ為に、シーン毎に奥行値をフルスケールで表せるように、奥行値を正規化することが行われる。シーン毎の最大視差量 $d_{\max}$ 、最小視差量 $d_{\min}$ が与えられれば、視差量を8ビット (0~255) で表す場合、正規化は次式で行える。

図4 奥行マップによる多視点映像の合成

図5 奥行映像からの電子ホログラフィ再生

図6 多視点映像と奥行の関係

$$depth = \frac{255(d - d_{\min})}{d_{\max} - d_{\min}} \quad (2)$$

3.2 奥行推定

3.2.1 対応点マッチング

多視点映像から奥行を得るには、まず対応点マッチング

図7 対応点マッチング

を行い、マッチング誤差が最小になる画素位置のズレ量を求める。具体的には、図7に示すように、カメラ2の映像（ビュー2）の全画素 $\text{pix2}(x, y)$ に対して、カメラ1の映像（ビュー1）の画素位置 $\text{pix1}(x+d, y)$ を少しずつずらして行き ($d=0, 1, 2, \dots$)、画素値の差分絶対値 $|\text{pix1}(x+d, y) - \text{pix2}(x, y)|$ が最小になるズレ量 d を探す。

$$D = |\text{pix1}(x+d, y) - \text{pix2}(x, y)|$$

$$d = \min_d \{D\} \quad (3)$$

この対応点マッチングは、カメラのズレやノイズに弱いので、信頼性を高めるために、画素値として、輝度値に加えて色差値も用いることや、3×3画素程度のブロックマッチング等が行われる。

3.2.2 平滑化

対応点マッチングのみによる奥行推定法は、カメラ映像の平行化処理の誤差や、カメラ映像間の色合わせ誤差、映像に混入したノイズ等の影響を受けやすい⁸⁾。その結果、得られる奥行マップは多くの誤差を含む。この問題に対しては、信頼伝播 (Belief Propagation)⁹⁾ やグラフカット (Graph Cuts) 理論¹⁰⁾ を用いて、例えば次式に示されるように、近隣画素で求めた奥行値との差分 (奥行値の不連続性成分) を、マッチング誤差 D に重み加算した値のすべての画素 (x, y) での和である評価値 E が最小になるように、推定される奥行値 d を求める平滑化処理が行われる。

$$E = \sum_{x,y} \{D + w|d1(x+1, y) - d2(x, y)|\}$$

$$d = \min_d (E) \quad (4)$$

この平滑化処理は、図8に示すように、すべての画素 (x, y)

図8 平滑化

でのすべての奥行値候補 d の中から、式 (4) に示す評価値 E が最小 (実際には極小) となる奥行値 d を画素 (x, y) 毎に決定するものである。

個々の被写体上では、奥行値ができるだけ同じ値になるようにすることで評価値の合計を下げられるが、被写体の境界部分では、奥行値が不連続になり評価値が上がる。そのため、被写体の境界部分を手動で求めたり、色の違いなどで検出して (セグメンテーション)、その部分だけ奥行きの不連続性の重み w を減らすことも行われる。しかし、色の違いだけで被写体境界と判定すると、同じ被写体上でも色が異なれば奥行の不連続性が許されることになり、奥行マップの品質が下がる。そこで、静止部分の奥行値を手動で求めて、その奥行値でのマッチング誤差 D を0にしたり、さらに、時間方向のフリッカを軽減するために、フレーム間の画素比較で静止している部分を検出し、その部分だけ、前フレームで得られた奥行き値との差分をマッチング誤差 D として、平滑化処理を行うことも行われる。

3.2.3 オクルージョンと偽マッチング

カメラのズレやノイズの他に、図9に示すような、手前の被写体の陰に隠れて対応点がないオクルージョンや、一様なテクスチャでの偽マッチングの問題がある^{11) 12)}。オクルージョンは、複数の視点映像に対するマッチング誤差の中から最小値を選ぶ事で軽減できるが、繰り返しパターンや一様なテクスチャを持つ被写体では、誤った奥行値での偽マッチングを生じやすい。これを軽減する方法として、

図9 オクルージョンと偽マッチング

図10 適応奥行推定

複数の視点映像毎のマッチング誤差の平均値を用いる方法があるが、オクルージョン部ではマッチング誤差に大きなノイズが乗り、最適奥行き値を推定し難い問題がある。

3.2.4 適応奥行推定

本項では、NICTで検討している適応奥行推定について述べる。この方法では、図10に示すように、3台のカメラを使い、被写体の境界部分ではオクルージョンによる奥行の誤推定を防ぐために、左右カメラ映像でのマッチング誤差の最小値を選択し、それ以外の部分では、偽マッチングを防ぐために左右のマッチング誤差の平均値を選択する。オクルージョンの有無は、繰り返し奥行き推定により判定する。1回目に推定された奥行きマップで、注目点の左右近傍の奥行き値の差が小さい場合は、オクルージョンなしと判定して、マッチング誤差の平均値を選択する。差が大きい場合はオクルージョンの可能性があると判定して、左右近傍の奥行き値をそれぞれ、注目点の奥行き値と仮定して、対応点を左右のカメラ映像内で求めて、注目点との一致度を調べ、左右映像でのマッチング誤差の差が大きい場合は、オクルージョンが生じていると判定して、最小マッチング誤差値を選択する。図11に見られるように、常に最小マッチング誤差を選択して得られた奥行きマップを使って合成した映像に比べて、約1dB高いPSNRが得られている。

3.3 多視点映像と奥行きマップの符号化

多視点映像と奥行きマップの符号化には、図12に示すように、従来の映像符号化ツール（AVC、MVCなど）が使用できる。AVCを用いた場合は、各視点映像と奥行きマップはそれぞれ個別の映像ストリームとして符号化される。MVCを用いた場合は、多視点映像と奥行きマップは個別に符号化されるが、多視点映像間または奥行きマップ間では視差補償予測を行いながら符号化されるので、符号化効率が改善される。これらを総称して、多視点映像と奥行きベースの

図11 Champagne Tower 300フレーム合成映像のPSNR (dB)

図12 MVDベースの符号化

符号化（MVD: Multi View Depth）と呼ぶ。MVCを用いる方法は、3DV符号化提案を評価するためのアンカー符号化方式として検討されている。

一方、中間視点映像が奥行きマップを使って射影できたように、符号化すべき視点映像や奥行きマップ自身も、他の視点からの映像や奥行きマップから射影できるので、1視点映像とその奥行きマップと、射影できないオクルージョン部分の映像とその奥行きマップのみを符号化すれば、データ量を大きく削減できる。この方法を、階層化奥行き映像ベースの符号化（LDV: Layered Depth Video）と呼ぶ。MVD、LDVいずれの場合でも、奥行きマップに要するビットレートは、視点映像の約1/5程度である。

3.4 視点映像合成

符号化された視点位置の映像をその奥行きマップで射影して、符号化されなかった視点位置の映像を合成すると、合成映像の画素がすべては埋まらず、射影されない画素が穴となってしまう。これをメディアンフィルタや近傍の画素値をコピーするインペイント法で埋めると、色やテキストが部分的に異なる被写体で歪みが大きくなる。そこで、

図13 合成奥行きマップによる視点映像の合成

図13に示すように、合成映像の視点位置での奥行きマップを符号化された奥行きマップから射影して作り、この奥行きマップの穴を埋めて、対応する画素を符号化された視点位置の映像から射影して求める。このようにすると、同じ被写体であれば奥行きは急激には変化しないので、合成映像の歪みが少なくなる。合成された映像のノイズをさらに減らすために、被写体の境界にフィルタを掛けたり、静止部分にフレーム間フィルタを掛けることも行われる。

多視点映像の符号化効率をさらに改善する目的で開始された3DV符号化方式の標準化は、2011年1月現在、標準化のためのテスト映像として多視点映像とその奥行きマップを準備中であり、2011年中に符号化の提案募集とその評価を行う予定である。

4 む す び

ISO傘下のMPEGグループで標準化が行われている、MPEG-FTVの標準化動向に付いて、その第1、第2フェーズの動向を紹介するとともに、NICTで検討している適応奥行き推定について述べた。現在、第3フェーズは予定されていないが、あるとすれば、第1、第2フェーズの符号化標準を用いた自由視点映像サービスのためのプロトコルなどが考えられる。この標準化活動への国内の取組みに関しては、超臨場感コミュニケーションフォーラムの中に名古屋大学を代表とした作業班が設置され、各社の精力的な活動が行われている。FTVで使われる多視点映像は、立体映像や自由視点映像の基本要素であり、今後その重要性が増して行くと思われる。多視点映像のみを入力とし、画素ブロックの平行移動で多視点映像間の視差補償予測を行うMVCは、すでに、標準化され実用化が始まった。一方、多視点映像から画素毎の視差量を抽出して得られた奥行きマップを用いて、さらに高い符号化効率を目指す3DVは、現在、標準化が検討されている。今後、裸眼立体映像や自由

視点映像の符号化に使われていくと思われる。その中で奥行きマップは、多視点映像の符号化効率改善に有効であるのみならず、ホログラフィなどの空間立体映像の生成にも重要な要素である。これらの応用が成功するためには、高品質な奥行きマップを手軽に得られるようにすることが重要であり、世界中でその検討が行われている。今後の進展に期待したい。

最後に、本報告で用いた多視点映像"Akko&Kayo"と"Rena"は、名古屋大学谷本研究室で撮影されたものです¹³⁾。"Akko&Kayo"の撮影には、筆者が在籍した東京大学安田青木研究室も協力しました。ここに関係各位に深く感謝致します。

(2011年1月11日受付)

〔文 献〕

- 1) 榎並和雅, 奥井誠人, 井ノ上直己: "NICTにおける超臨場感コミュニケーションの研究戦略," 第228回研究会講演予稿, 画電学会06-03, (映像学, 30, 58), (信学会, 106, 338), (電気学会EDD-06-75~85), pp.1-6 (2006)
- 2) 本田捷夫統括: "高度立体動画通信プロジェクト最終成果報告書", 通信・放送機構 (1997)
- 3) M. Tanimoto: "Overview of Free View-point Television", Signal Processing: Image Communication, 21, 6, pp.454-461 (2006)
- 4) ISO/IEC 14496-10, or ITU-T H.264 (2009)
- 5) 谷本正幸: "自由視点映像伝送方式に関する国際標準技術の研究", 戦略的情報通信研究開発推進制度 (SCOPE), 第5回成果発表会 (2009)
- 6) 福島慶繁, 松本健太郎, 圓道知博, 藤井俊彰, 谷本正幸: "特徴点軌跡並行化によるカメラアレイレクティフィケーション", 映像学誌, 62, 4, pp.564-571 (2008)
- 7) T. Senoh, K. Yamamoto, R. Oi, T. Mishina and M. Okui: "Computer Generated Electronic Holography of Natural Scene from 2D Multi-view Images and Depth Map", Proc. of 2nd International Symposium on Universal Communication (ISUC) 2008, pp.126-133 (2008)
- 8) K. Palaniappan, et. al: "Robust Stereo Analysis, Computer Vision", Proc. International Symposium on Digital Object Identifier, pp.175-181 (1995)
- 9) J. Sun, N. Zheng and H. Shun: "Stereo matching Using belief propagation", IE3 Trans. Pattern Analysis and Machine Intelligence, 25, 7, pp.787-800 (2003)
- 10) Y. Boykov, O. Veksler and R. Zabih: "Fast Approximate Energy Minimization via Graph Cuts", IE3 Trans. Pattern Analysis and Machine Intelligence, 23, 11, pp.1222-1239 (2001)
- 11) 苗村健, 原島博: "オクルージョンを考慮した多眼画像からの視差推定", 信学ソ大, SD-7-1 (1996)
- 12) 今泉浩幸, 片山美和, 岩館祐一: "任意視点画像表示を目的とした多眼画像からの奥行き推定", 信学技法, IE2001-59, PRMU2001-79, MVE2001-58, pp.109-116 (2001)
- 13) <http://www.tanimoto.nuee.nagoya-u.ac.jp/>

妹尾 孝憲 1975年、大阪大学大学院修士課程修了。同年、松下電器(株)(現パナソニック(株))入社。1988年～1990年、MIT Media Lab客員研究員。2005年、東京大学国際・産学共同研究センター特任研究員。2006年、同先端科学技術研究センター特任教員。2007年より、情報通信研究機構、専攻研究員。立体映像の研究に従事。九州大学客員教授、工学博士。正会員。