3-7　MPEG多視点映像符号化の標準化活動

3-7 MPEG Multi-View Image Coding Standardization

妹尾孝憲　山本健詞　大井隆太朗　栗田泰市郎
SENOH Takanori, YAMAMOTO Kenji, OI Ryutaro, and KURITA Taiichiro

要旨

国際標準化機構（ISO：International Organization for Standardization）傘下の動画像符号化グループ（MPEG：Moving Picture Expert Group）では、立体映像や自由視点映像の元となる多視点映像の符号化標準の策定活動が行われている。NICTもこの活動に参加し、技術検討に寄与している。符号化方法には、多視点映像間にもフレーム間の動き補償予測を適用して圧縮するMVC（Multi-view Video Coding）や、多視点映像から映像の奥行を抽出して視点数を間引いて圧縮し、復号化時に必要な視点映像を合成する3DV/FTV（3Dimensional Video/Free-viewpoint TV）などがある。本稿では、これらの技術内容全般を紹介すると共に、NICTで行った検討内容を報告する。

Standardization of multi-view image coding, which is the seed of 3-dimensional Videos or free-viewpoint TVs is going on in the Moving Picture Expert Group (MPEG) under International Organization for Standardization (ISO). MVC (Multi-view Video Coding) compresses the multi-view images by view prediction estimating disparities between views as well as frame prediction estimating the motion vectors between frames. Another new movement is an investigation on a higher coding efficiency by estimating the image depths. In this paper, these multi-view image coding technologies are introduced together with the research performed in NICT.

[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
3-dimensional video, Free-viewpoint TV, Multi-view image, Disparity-compensated prediction, Depth estimation

1　まえがき

人々のコミュニケーション手段を高度化する為に、超臨場感コミュニケーションが研究されている[1]。超臨場感コミュニケーションでは、奥行きを持った立体映像を用いた高臨場感コミュニケーションを目指している。

その為の立体映像実現手段には、左右の目に個別の映像を送る2眼立体映像方式や、これを拡張した多眼立体映像方式、3次元の被写体映像空間を再現する体積表示方式、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方式などがある[2]。これらは全て、複数の異なる視点から撮影された多視点映像を利用している。

多視点映像は又、受信者が被写体を自由な視点位置から見る事の出来る自由視点テレビFTV[3]を実現する為にも利用出来るなど用途は広いが、多数の視点映像を伝送する必要があり、伝送や蓄積の負担が大きい。

自由視点映像や立体映像に使われる多視点映像を圧縮する為、国際標準化機構（ISO: International Organization for Standardization）傘下の動画像符号化専門家グループ（MPEG: Moving Picture Expert Group）は、2001年より自由点映像FTVの符号化標準の策定を進めており、FTVの第1フェーズとして、2009年にMVC（Multi-view Video Coding）を、携帯電話向けTV放送や高密度光ディスク等に使われているMPEG−4 Video Part10（Advanced Video Coding）符号化標準（ISO/IEC14496−10/ITU−T H.264）のAnnex H

(Multiview video coding)として策定した[4]。

この方式は、従来のMPEG−4 AVC(ITU−T H.264と共通規格)の動き保証フレーム間予測を、視点映像間の視差補償予測にも拡張適用したものであり、高密度光ディスクの3D映像符号化方式として採用された。又、FTVの第2フェーズとして2007年からは、映像の奥行情報を利用して、符号化効率の更なる改善を目指す3DV/FTV(3-Dimensional Video/Free-viewpoint TV)符号化標準の検討を開始した[5]。

以下では、これら符号化方式の技術内容を紹介すると共に、NICTで検討した適応的奥行推定について述べる。

2 多視点映像の符号化(MVC)

2.1 多視点映像

多視点映像は、図1に示す様に1つのシーンを多数のカメラで撮影したものであり、カメラの数だけ視点位置の異なった映像から成る。カメラ間隔を人の瞳間距離(約6.5 cm)にしておけば、任意の2カメラ映像をそのまま、2眼立体映像として利用可能である。2眼立体映像では、左右の目にそれぞれ別の映像を届ける為に、偏光眼鏡やシャッターグラスを用いて左右映像を分離したり、ディスプレイに視差バリヤーを設けて、左右眼がそれぞれ別の映像を見られるようにした、裸眼ステレオ立体ディスプレイなどが実用化されている。又、より多数のカメラ映像を用いて、それぞれの映像の視域を連続的につなげて表示する裸眼立体映像にも利用可能である。

2.2 カメラ配置とテスト映像

図1に見られるように、各カメラから得られる多視点映像には、同じ被写体が写っているので、大きな相関がある。そこでこの視点映像間の相関を利用して多視点映像符号化方式(MVC)の標準化が行われた。標準化を行うに当たっては、図2に示す様な各種のカメラ配置が提案され、夫々のテスト映像が標準化参加者から提供された。当初は、各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平行化処理を、符

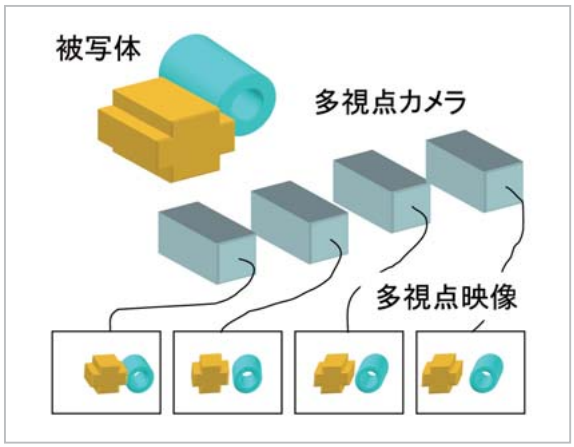

図1 多視点映像の撮影

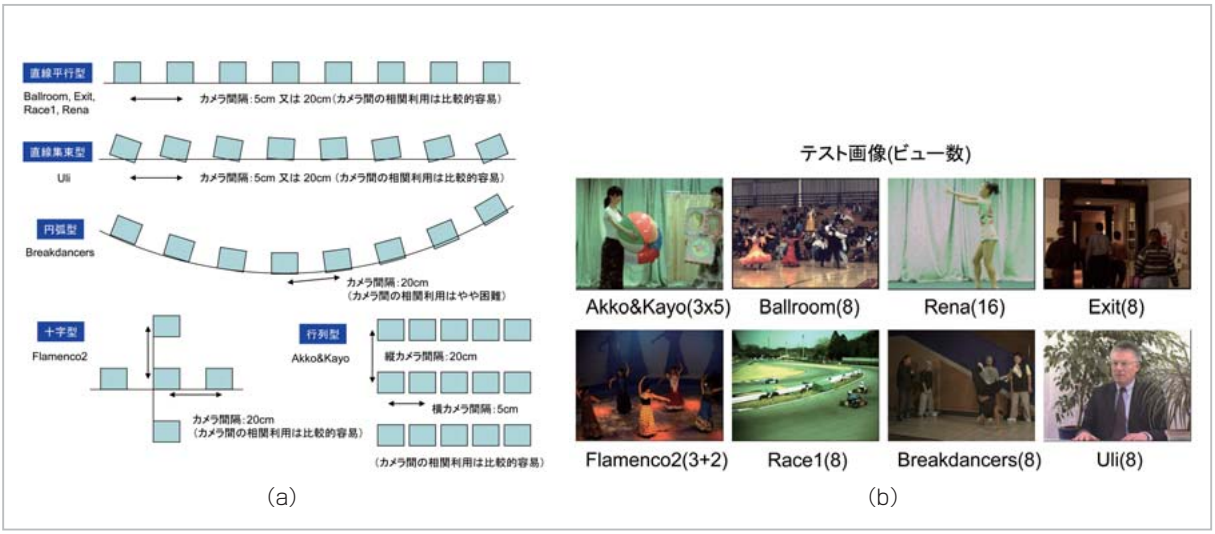

図2 カメラ配置とテスト映像
(a)カメラ配置の例、(b)テスト映像の例

号化の中で行う方法も提案されたが、コーデックの負担が大きい為、カメラ側で行う事になった。

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ひずみ補正を行った後、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、全ての視点映像で一致する様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する。これで全てのカメラの方向と内部パラメータが揃った状態になる。次に、有限

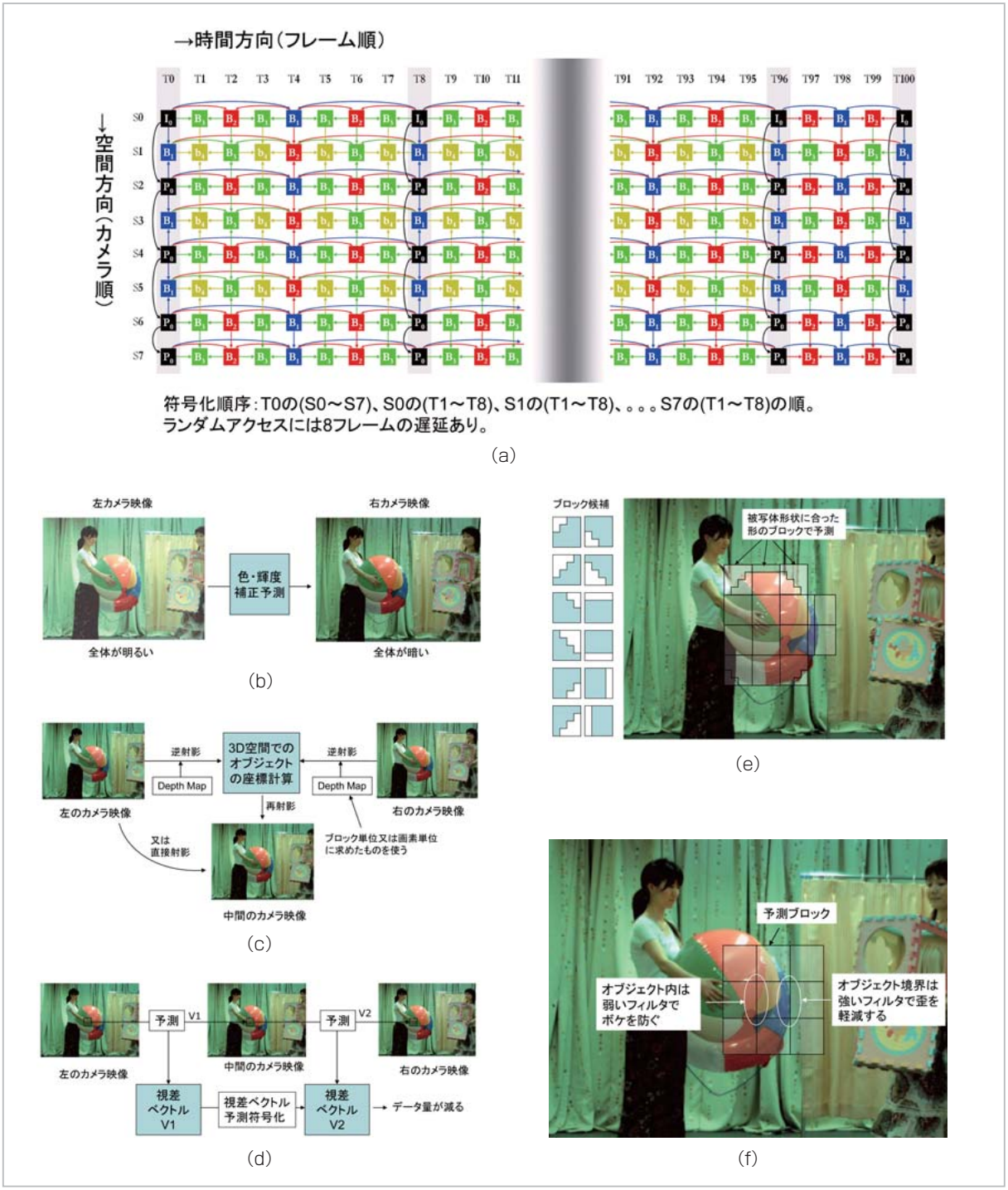

図3　多視点映像符号化標準（MPEG-4 MVC）に提案された各種の符号化方式

(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

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様に、共通の射影変換行列を求めて再度射影変換するものがある[6]。

2.3 提案された符号化方式

これらのテスト映像を用いて、各種の符号化方式が提案された。その主なものを、図3に示す。図3(a)は、時空間階層双方向予測方式で、従来のフレーム間予測に用いられていた、フレーム間で動いたブロックの方向と距離を、ブロックマッチングによる動きベクトル探索で求め、基準となるフレームのブロックをこの動きベクトル量だけ移動して予測されるフレームから差し引き、残った差分を変換符号化と量子化、及び可変長符号化して送る方式を、視点毎の映像にも適用したもので、ブロックの平行移動探索を視点映像間で行い、得られたベクトルで視差補償されたブロックを、フレーム予測候補に加え、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。異なる視点映像間では、被写体の見える角度が異なったり、前景に隠れて背景が見えなくなるオクルージョンの問題があり、予測効率はあまり高くないが、相関の高いフレーム間の予測に階層双方向予測を使う事で、符号化効率を上げている。

図3(b)は、照明補償予測方式で、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ずれを補償する値を、視点映像間の平均値からのオフセット量として加える事により、予測符号化効率を上げるものである。この視点映像間の色合わせは、カメラ側で行う事によりコーデックの負担を減らせるので採用されなかった。

図3(c)は、射影変換予測方式であるが、視点映像間で対応点マッチングを行って視差量(映像の奥行き)を求め、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方式である。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、予測するものがあったが、複雑な処理に見合うほど予測効率が上がらず採用されなかった。

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事で、視差ベクトルのデータ量を減らす方式であるが、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ず、採用されなかった。

図3(e)は、非対称マクロブロック予測方式である。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事により、ブロックの予測精度を高めるものであるが、ブロック選択の為の情報が増え、予測効率は余りあがらず、採用されなかった。

図3(f)は、適応予測フィルタ方式で、ブロック単位の視差補償予測を行うと、ブロックの境界で視差の違いが大きいと、映像が不連続になりブロック歪が出るので、視差の違いの大きい部分のみに適応的にローパスフィルタを掛けて平滑化するものである。主観的には歪を軽減出来るが、映像がぼける為、採用されなかった。

その後、視差ベクトルを用いて隣の視点映像の動きベクトルを予測する方式が提案されたが、符号化効率の改善量が約0.5 dBと少ないとの理由で採用されなかった。

2.4 標準化された符号化方式

結局、多視点符号化方式(MVC)は、図3(a)の時空間階層双方向予測方式のみを採用して標準化された。この方式は、静止している被写体ではフレーム間差分が0となり、大きな圧縮効果が得られる双方向フレーム間予測(図中のBフレーム)を、更に階層化する事によって符号化効率を上げている。一方、視点映像間の予測では、被写体が静止していても常に視差が生じるので、映像全体の視差補償予測が必須となる。視差が生じると、隣の視点映像では前景の陰に隠れていた背景が見えて来たり、隣の視点映像で見えていても、視点が変わると見える角度が変わり、対応するブロック形状が歪むので、画素ブロックの単純な平行移動による予測では予測効率が上がらず、大きな圧縮率は望めない。その結果、MVCでの符号化効率の改善量は、従来のMPEG−4 AVCなどで全視点映像を符号化伝送する場合の約1.5倍程度であったが、高密度光ディスク用の3D映像符号化方式に採用された。

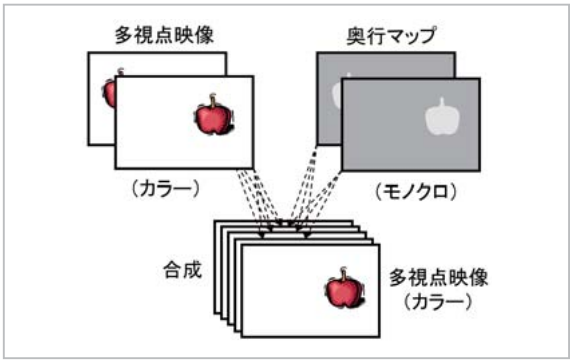

図４　奥行マップによる多視点映像の合成

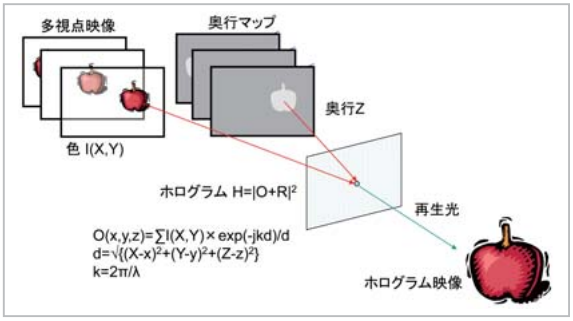

図５　奥行映像からの電子ホログラフィ再生

3　立体映像/自由視点映像の符号化（3DV/FTV）

立体映像や自由視点映像の元となる多視点映像の符号化効率を更に改善する為に、MPEGでは2007年から、多視点映像符号化の第2フェーズとして、3DV/FTV符号化方式の検討を始めた。そこでは、符号化効率を上げる為、図4に示す様に、全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（デプスマップ）を符号化し、復号側で奥行情報を元に射影変換を行って、必要な視点映像を合成する事でデータ量を削減する方法を検討している。

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映像を合成できるので、多視点映像を基本とする立体映像の符号化の他、視聴者が自分の見たい位置から自由にシーンを見ることの出来る自由視点映像サービスにも利用できる。又、奥行情報を用いると、図5に示す様に被写体から出る光の波面を計算可能で、これを用いて理想的な立体映像方式である電子ホログラフィも実現出来るので[7]、将来の立体映像/自由視点映像符号化方式として期待されている。その為には、多視点映像から奥行きを抽出する必要がある。

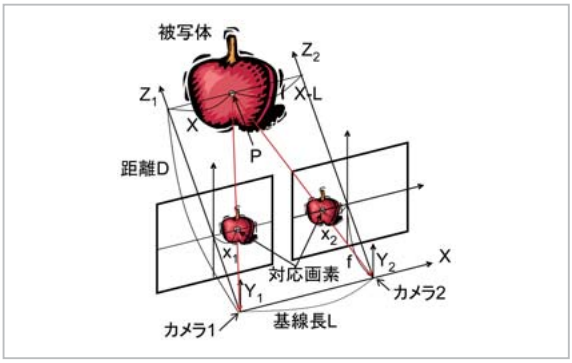

図６　多視点映像と奥行の関係

3.1　多視点映像の奥行

多視点映像とその奥行には図6に示す様な関係がある。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・平行に並んでいる場合であるが、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対応画素位置x_1、x_2には次式の関係がある。

$$\frac{f}{D}=\frac{x_1-x_2}{L} \tag{1}$$

図では簡単の為に注目点の高さをY=0としたが、(1)式は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)を視差と言い、多視点映像の奥行値として使われる事が多い。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出来たり、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。

奥行値の精度を出来るだけ落とさない為に、シーン毎に奥行値がフルスケールで表せる様に、奥行値を正規化する。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、正規化は次式で行われる。

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

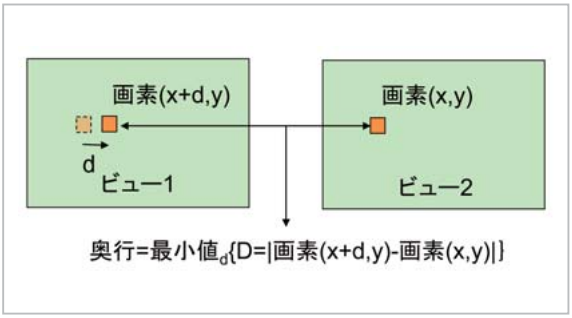

図 7 対応点マッチング

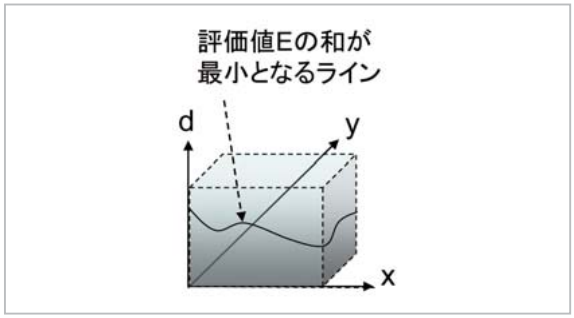

図 8 平滑化

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 に示す様に、カメラ 2 の映像(ビュー 2)の全画素 $pix(x,\ y)$ に対して、カメラ 1 の映像(ビュー 1)の画素位置 $pix(x,\ y)$ を少しづつずらして行き(d=0, 1, 2, …)、画素値の差分絶対値 | 画素 $(x+d,\ y)$ − 画素 $(x,\ y)$ | が最小になるズレ量 d を探す。

$$D = \left| pix(x+d, y) - pix(x, y) \right|$$
$$d = \min_d \{D\} \tag{3}$$

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、信頼性を高める為には、画素値として、輝度値に加えて色差値も用いる事や、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)やグラフカット(Graph Cuts)理論を用いて、近隣画素で求まった奥行値との差分(奥行値の連続性成分)をマッチング誤差に重み加算した評価値が最小になる様に、推定される奥行値を平滑化する事が行われる[9][10]。

$$E = \sum_{x,y} \{D + w\left| d(x+1, y) - d(x, y) \right|\}$$
$$d = \min_d (E) \tag{4}$$

この平滑化処理では、図 8 に示す様に、対応点マッチングで得られた全ての画素での全ての奥行値候補に対するマッチング誤差の中から、被写体境界部での奥行値の不連続性を考慮して、画像全体での評価値が最小(実際には極小)となる様に奥行値を決定する。

個々の被写体上では、奥行値が出来るだけ同じ値になる様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。その為、被写体の境界部分を色の違いなどで検出して(セグメンテーション)、奥行きの連続性の重みを減らす事も行われる。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事になり、奥行マップの品質が下がる。

又、時間方向のフリッカを軽減する為に、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マッチング誤差 D を加算平均したり、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 にしてから、信頼伝播やグラフカット処理を適用する事も行われる。

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 に示す様な、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問題や、一様なテキスチャでの偽マッチングの問題がある[11][12]。

オクルージョン問題は、複数のカメラ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出来るが、繰り返しパターンや一様なテキ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、複数のカメラ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マッチング誤差に大きなノイズが乗り、最適奥行

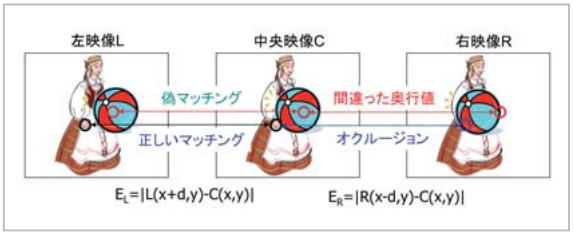

図９ オクルージョンと偽マッチング

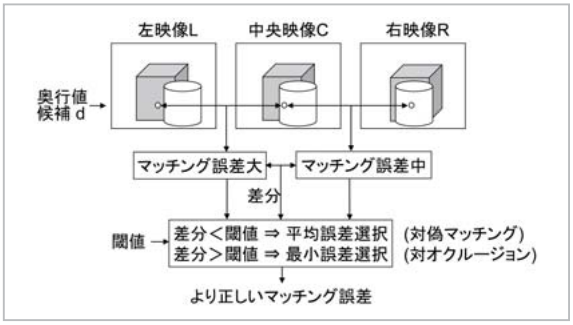

図 10 適応的奥行推定

値を推定し難い問題がある。

3.2.4 適応的奥行推定

以下では、NICTで検討した適応的奥行推定について述べる。この奥行推定法は、複数のマッチング誤差の最小値と平均値を適応的に切り替えて、オクルージョンと偽マッチングの両方の課題を軽減する。ここでは、奥行推定の演算量を増さない為に、図10に示す様に水平・平行に配置した3台のカメラ映像のみを使う。中央カメラ映像Cの画素毎に、奥行き値dを0から最大視差量まで変えながら、両隣のカメラ映像L、Rで対応する画素値と、中央映像の画素値との差分絶対値 D_L、D_R を、左右それぞれで求める。左右のマッチング誤差の差が閾値より大きい場合は、片方の映像でオクルージョンが発生している可能性が高いと判断し、小さい方のマッチング誤差を用いて奥行値を推定する。両者の差が小さい場合は、オクルージョンは発生していないと判断して、偽マッチングの可能性を軽減する為に左右のマッチング誤差の平均値から奥行き値を推定する。閾値は、被写体によって経験的に定めるが、最大マッチング誤差量の1/10程度の値が適する。

以下に、適応的マッチング誤差選択の効果を示す。実験には、図11に示す3DV/FTV標準化検討に用いられている名古屋大学提供のテスト映像(Champagne Tower)の一部を用いた[13]。カメラ間隔は5 cm、映像は平行化と色補正済みである。

実験では、先ずビュー38、39、40を用いて、図10に示したアルゴリズムに基づいて、ビュー39の奥行マップを求め、同様にしてビュー40、41、42からビュー41の奥行マップを求め、これら2つの奥行マップとそのビューから、中間のビュー40'を合成して、カメラ映像ビュー40と比較した。マッチング誤差の平均値と最小値の選択基準は、外部から与えられるパラメータとし、予備実験により、最大マッチング誤差量＝255の場合、th＝33を用いた。奥行推定には、MPEG 3DVグループで開発されている参照ソフトウエア(DERS5)を修正して用いた[14]。修正前のソフトウエアは、3ビューの映像を使って左右ビューでのステレオマッチング誤差を全ての仮定奥行値で求め、最小マッチング誤差に、グラフカット理論に基づく平滑化処理を行って各画素の奥行値を決めるものである。図12に、修正したソフトウエアで、最小マッチング誤差を選択した場合と、マッチング誤差の平均値を選択した場合と、閾値 th≦33で最小マッチング誤差選択、それ以外でマッチング誤差の平均値選択を行う適応マッチング誤差選択を行った場合のデプスマップを示す。

最小誤差選択では、テーブルのエッジ部分の奥行値はシャープに求まっており、オクルージョンに強い事が確認されるが、グラスの奥行値がやや平坦になっている。これは、偽マッチングの影響と思われる。平均誤差選択では、グラスの奥行値は細かく求まっており、偽マッチングに強い事が確認されるが、テーブルのエッジ部分の奥行値にボケが見られ、オクルージョンに弱い事が分かる。尚、背景の奥行値が左側と右側で大きく異なるのは、背景が黒一色でテキスチャが殆どない為、どの奥行値でもマッチング誤差が大きくならず、奥行値を決定する平滑化処理で、近隣のテキスチャを持つ部分の奥行値に強く影響された結果である。テキスチャのない部分では、誤った奥行値で中間ビューを合成しても画質の劣化はほとんど検知されないが、シーン中のスピーカポールの脚の様に暗い被写体の奥行値を誤ると、背景が消えたり2重像になる問題がある。

適応的に最小誤差又は平均誤差を選択した場合は、グラスの奥行値は細かく求まっており、偽

マッチングに強く、テーブルのエッジ部分でも、背景にスピーカポールの脚などのテキスチャがある部分では、テーブルの奥行値の広がりは見られず、オクルージョンにも強い事が確認される。

同様にして求めたビュー 41 の奥行マップと、ビュー 39 の奥行きマップ及び、それらのカメラ映像から中央のカメラ映像（ビュー 40）を合成した結果を図 13 に示す。ビュー合成には、MPEG 3 DV の参照ソフトウエア（VSRS3.5）を用いた。このソフトウエアは、左右ビューのデプスマップとそのカメラパラメータから、中間ビューのデプスマップを射影して作り、その奥行値に対応する画素を左右のビューから射影して中間ビューを合成する。

(a) 最小誤差選択、(b) 平均誤差選択、(c) 適応誤差選択の夫々で推定された奥行マップから合成されたビューのピーク信号対雑音比（PSNR）は、夫々 31.92 dB、33.76 dB、32.65 dB であり、平均誤差選択による奥行マップからの合成映像の PSNR が最も高かった。この SN の差は、主にグラス等の細かな部分の奥行値の精度の差から来ている。最小誤差選択による奥行値から合成された映像の主観画質は良いが、カメラで撮影された画像と比べると、図 14 に示す様に誤差が大きい。

この誤差は、図 15 に示す様に、グラスで鏡面反射される照明光の位置が、グラス毎で異なる事に起因する。最小マッチング誤差選択では、図 15 (a) に示す様に、カメラ毎に異なる輝点位置がマッチ

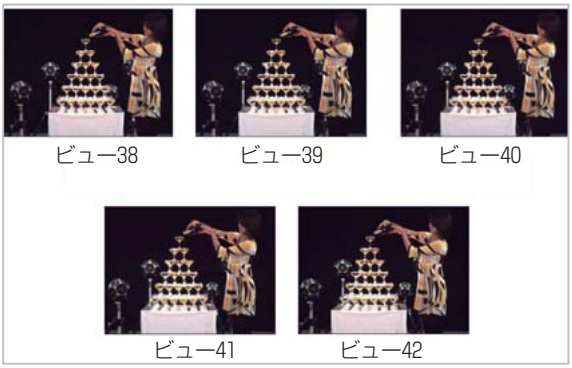

図 11 実験映像（Champagne Tower）

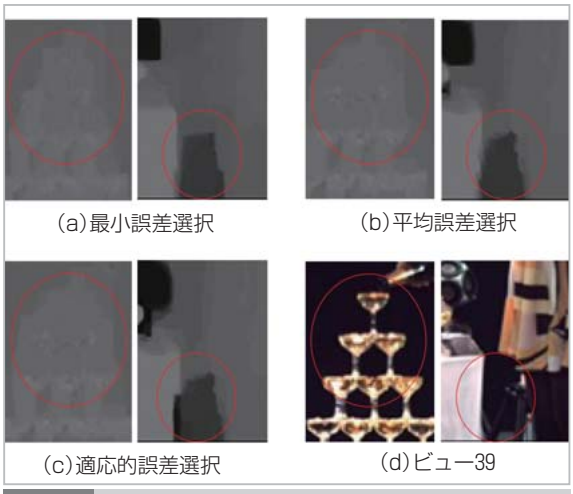

図 12 デプスマップとビュー 39

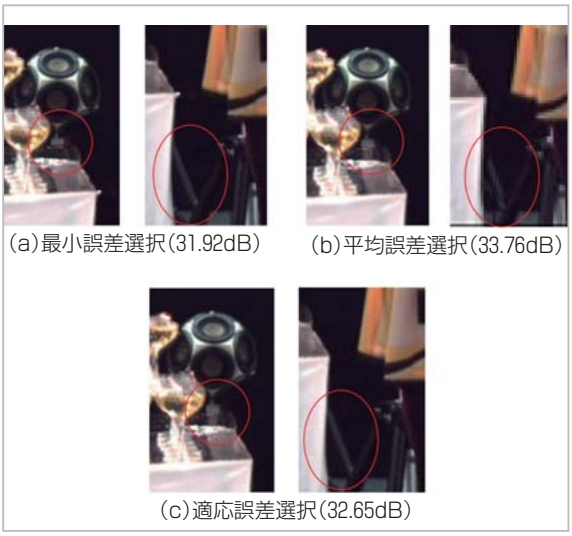

図 13 ビュー 39、41 から合成されたビュー 40

図 14 合成ビューとカメラビューの差分

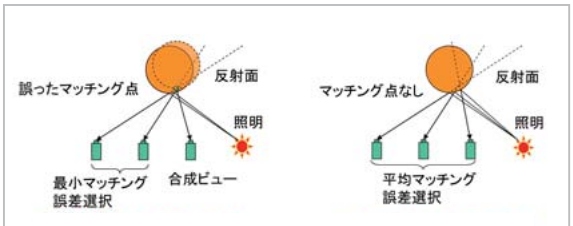

図 15 鏡面反射のマッチング

ングする様に奥行値が推定される為、正しい対応点マッチングにならず、推定された奥行値は誤った値となり、これを使って合成された映像は、カメラで撮影された映像から僅かにズレる為、SNが劣化する。

平均マッチング誤差選択では、3つのカメラ映像の輝点が同時にマッチングする奥行値はなく、推定される奥行値は、どの値でも同じ様なマッチング誤差を持つ。このマッチング誤差をグラフカットなどで平滑化すると、近隣の正しくマッチングされた画素の奥行値との差が小さくなる様に、奥行値が決められる為、輝点部分の奥行値もほぼ正しい値となる。

一方、平均誤差選択による奥行きマップから合成された映像では、テーブルの右端のスピーカの脚が2重像になっている事が認められるが、被写体が暗い為、PSNRには大きく影響しない。2重像になったのは、脚が暗いテキスチャの為、不正な奥行値でもマッチング誤差が大きくならず、奥行値の平滑化処理で隣のテーブルの奥行値の影響を強く受け、スピーカの脚部分の奥行値を誤った為である。

又、テーブルの周囲に巻かれた半透明フィルムを通して見える右側のスピーカポールが、最小誤差選択法では消え掛けているが、平均誤差選択や適応誤差選択法では消えていない。これは図16に示す様に、合成ビューでスピーカポールと半透明フィルムが重なっている部分の奥行値として、最小誤差選択法では半透明フィルムの奥行値が優先されるのに対して、平均誤差選択法では半透明フィルムの奥行値が正しく求まらず、3台のカメラ映像で背景のポールが重ならない様な奥行値に収束した為である。

適応誤差選択の奥行マップから合成されたビューでは、PSNRは若干下がったが、半透明フィルムの後ろのスピーカーポールは正しく合成されており、テーブルの右側のスピーカポールの脚の2重像も解消されている。

図17に、Champagne Towerシーケンス300フレームのビュー39、41の奥行きを適応マッチング誤差選択で求めて合成したビュー40のPSNRの変化をグラフで示す。NICTで検討した適応的奥行推定法は、最小誤差選択法に比べて約1dB高いPSNRが得られており、その効果が確認された。

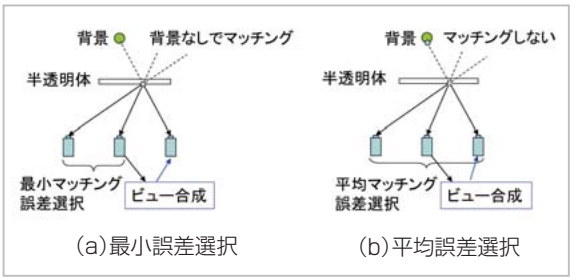

図16 半透明被写体の奥行推定

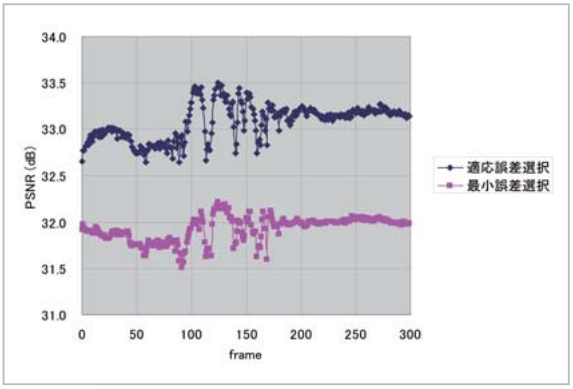

図17 Champagne Tower 300フレームの合成映像のPSNR(dB)

た。

3.3 符号化

以下では、推定された奥行マップと視点映像を符号化・復号化する方法、及び復号化された視点映像と奥行マップから中間の視点映像を合成する方法について、MPEGで検討されている内容を述べる。

3.3.1 多視点映像と奥行マップの符号化

多視点映像の奥行マップが得られると、図18に示す様に、中間視点映像を射影変換で合成出来るので、全ての視点映像を符号化する必要はなく、少数の視点映像とその奥行マップのみを符号化すれば良い。奥行マップの精度が十分高ければ、合成される視点映像の品質も高いが、復号側で得られる奥行マップの精度が不十分な場合は、合成される視点映像の誤差分も符号化する必要がある。

この視点映像と奥行マップの符号化には、図19に示す様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、各視点映像と奥行マップは夫々個別の映像ス

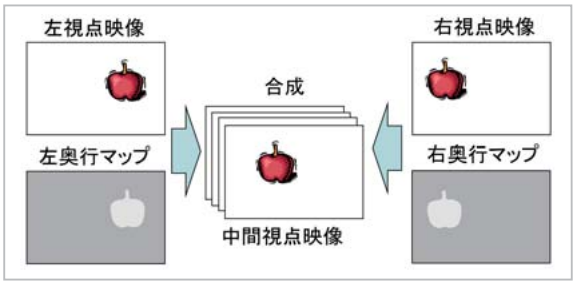

図 18 奥行映像からの多視点映像合成

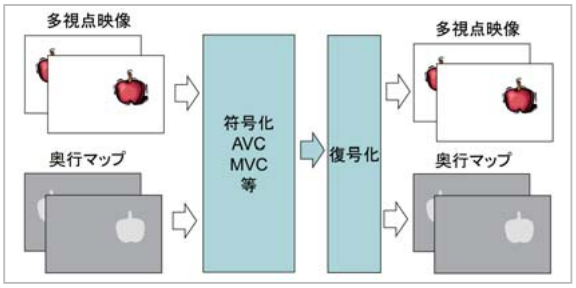

図 19 MVD ベースの符号化

トリームとして符号化される。MVC を用いた場合は、視点映像間又は奥行きマップ間で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が、視点映像と奥行マップは個別に符号化される。これらを総称して多視点映像と奥行ベースの符号化と呼ぶ(MVD: Multi View Depth)。

3.3.2 階層化奥行映像の符号化

中間視点映像が奥行マップを使って射影出来た様に、符号化すべき視点映像や奥行マップ自身も、図 20 に示す様に、1 つの視点映像や奥行マップから射影出来る。射影できないのは、オクルージョン部分のみであるので、オクルージョン部分のみを追加して符号化すれば、符号化すべきデータ量を大きく削減できる。この方法を、階層化奥行映像ベースの符号化と呼ぶ(LDV: Layered Depth Video)。LDV データの符号化は、それぞれの視点映像や奥行マップ及びオクルージョン部のデータを個別に AVC 又は MVC で符号化すれば良い。

この方法は、多眼ディスプレイ用映像などの、視点映像間の視差量があまり大きくない場合には有効であるが、視点間距離が大きい場合や、視差量が大きい場合には、符号化効率が下がる。

いずれの場合でも、奥行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。したがって奥行マップを用いると、2 眼映像の符号化では従来方式(MVC など)と余り符号化効率の差はないが、ある視点数までは符号量は増えないので、視点数が増える程、符号化効率が上がる特徴がある。

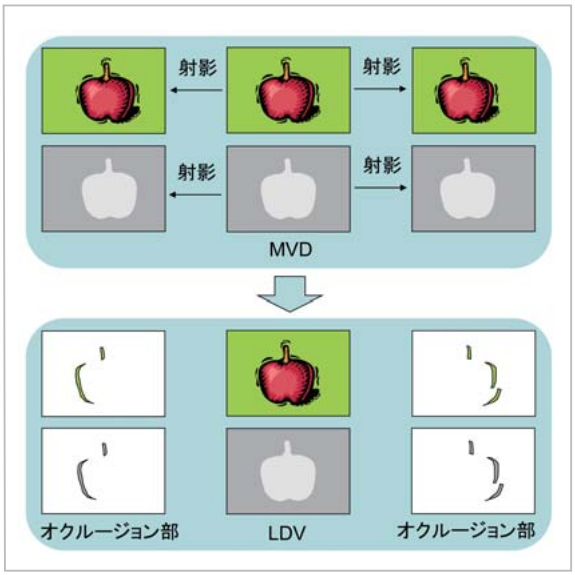

図 20 MVD の LDV 化

3.4 視点映像合成

復号側での中間視点映像の合成は、両側の視点映像をそのデプスマップを使って直接射影すると、合成映像の画素が全ては埋まらず、射影されない画素が穴となって残る。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、色やテキスチャが部分的に異な

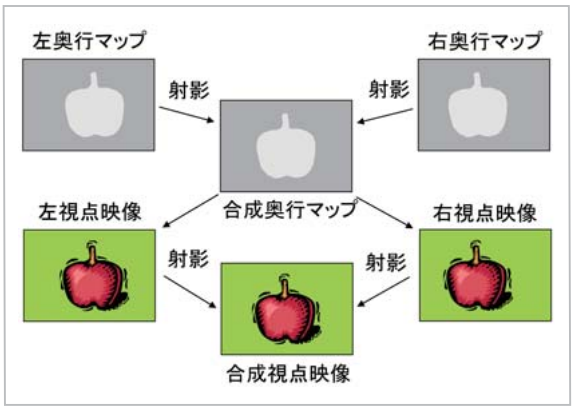

図 21 合成奥行マップによる視点映像の合成

る被写体で歪が大きくなる場合がある。別の方法として、図21に示す様に合成映像の視点位置でのデプスマップを両側のデプスマップから射影して作り、このデプスマップの穴を同様に埋めてから、そのデプスに対応する画素を両側の視点映像から射影すると、同じ被写体であればデプスは急激には変化しないので、合成映像の歪が少なくなる。合成された映像のノイズを更に減らす為に、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事も行われる。

4 むすび

ISO傘下のMPEGグループで標準化が行われている、多視点映像の符号化方式に付いて紹介すると共に、NICTで検討した適応的奥行推定について述べた。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視点映像間の視差補償予測を行うMVCは既に標準化され、実用化が始まった。多視点映像から画素毎の視差量を奥行マップとして抽出して、更に高い符号化効率を目指す3DV/FTVは、これから標準化され、裸眼立体映像や自由視点映像に使われて行くと思われる。その中で奥行マップは、多視点映像の符号化効率改善に有効であるのみでなく、電子ホログラフィなどの空間立体映像や自由視点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為には、高品質な奥行マップを手軽に得られる様にする事が重要であり、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。

謝辞

本報告で用いた多視点映像“Akko & Kayo”と“Champagne Tower”は、名古屋大学谷本研究室で撮影されたものです。“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。

参考文献

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（映情学会Vol. 30，No. 58）（信学会Vol. 106，No. 338）（電気学会EDD-06–75～85），pp. 1–6，2006.

2 本田捷夫統括，“高度立体動画像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, Vol. 21, No. 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，”映情学誌，Vol. 62，No. 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," IE^3 Trans. Pattern Analysis and Machine Intelligence, Vol. 25, No. 7, pp. 787–800, 2003.

10 Y. Boykov, O. Veksler, and R. Zabih, "Fast Approximate Energy Minimization via Graph Cuts," IE^3 Trans. Pattern Analysis and Machine Intelligence, Vol. 23, No. 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/
14 ISO/IEC JTC1/SC29/WG11，"Draft Report on Experimental Framework for 3D Video Coding," N11273，2010.

妹尾孝憲（せのおたかのり）
ユニバーサルメディア研究センター
超臨場感基盤グループ専攻研究員
工学博士
電子ホログラフィ、立体映像

山本健詞（やまもとけんじ）
ユニバーサルメディア研究センター
超臨場感基盤グループ主任研究員
博士（工学）
電子ホログラフィ、３次元画像工学

大井隆太朗（おおいりゅうたろう）
ユニバーサルメディア研究センター
超臨場感基盤グループ主任研究員
博士（科学）
光波伝播解析、ホログラフィ、３次元撮像技術、イメージセンサー

栗田泰市郎（くりたたいいちろう）
ユニバーサルメディア研究センター
超臨場感基盤グループグループリーダー　博士（工学）
テレビシステム、動画信号処理、ディスプレイ表示方式と画質、フレキシブルディスプレイ、立体映像