RI-002

高解像度デプスマップによる 超解像を用いた自由視点画像のデータ量削減
Small data volume representation of free viewpoint image based on super resolution with high resolution depth map

杉本 志織† 志水 信哉† 木全 英明† 松浦 宣彦†
Shiori Sugimoto Shinya Shimizu Hideaki Kimata Norihiko Matsuura

1. 序論

次世代の映像メディアの一つとして，視聴者が自由に視点を操作することができる自由視点映像(Free Viewpoint Television, FTV)が注目を集めている[1]．自由視点映像は，対象シーンを多数の撮像装置を用いて様々な位置・角度から撮像してシーンの光線情報を取得し，これを元に任意の視点における光線情報を復元することによって生成できる．しかしながら，シーン内全ての光線情報を撮像によって取得するには膨大な数の撮像装置を密に設置しなければならないため，実現は容易ではない．実際には，疎に配置した少数の撮像装置から得られる光線情報から，何らかの補間手法を用いて未取得の光線情報を合成する必要がある．

この補間合成の手法のひとつとして，多視点映像とそこから推定されるシーンの奥行情報を用いて仮想視点映像を合成する Depth Image based Rendering(DIBR)がある[2]．奥行情報は多視点映像の各画素における被写体までの距離である．自由視点映像を伝送することを考えた場合，奥行情報を送信側で推定し多視点のグレースケール映像(デプスマップ)として記述して伝送することが有効であると考えられる．このアプローチは受信側の演算量を削減すると共に，符号化歪みが重畳する前の多視点映像を用いて，奥行情報を推定することでより精度の高い推定を可能にする．現在，国際標準化団体 MPEG において，自由視点映像を含む新しい三次元映像の符号化方式として，このような奥行き情報を符号化して伝送するフレームワークの検討が進められている[3].

このような多視点映像と多視点デプスマップからなる映像データは膨大な情報量を持つため，より効率のいい符号化方式が必須であり，様々な方式が検討されている．しかしながら符号量の削減と共にもう一つ達成しなければならないこととして，デコーダのスループットとメモリ容量の上限から，映像データの総ピクセル数をある程度低減する必要があるといったことが挙げられる． 2011 年発行の MPEG-3DV の要求文書では，多視点映像と多視点デプスマップからなる映像データの総ピクセル数を単一視点映像の 4 倍以下に抑えるべきとされている[4]

こうした背景を踏まえて，本研究では，自由視点画像の品質を維持しつつ総ピクセル数の少ないデータ表現を提案する．

2. 従来手法

総ピクセル数を削減するためには，多視点映像か多視点デプスマップのどちらか，あるいは両方に対して何らかのダウンサンプリングが必要となる．従来は，デプスマップをダウンサンプリングしピクセル数の削減を行うといったアプローチが多く取られてきた[5]．こうした手法が多く取られる理由として，一般にデプスマップは画面内で連続性が高くダウンサンプリングによる損失が少なく，映像をダウンサンプリングする場合と比べて合成後の映像品質に影響が少ないと考えられていることが挙げられる．特にステレオ画像から中間視点の画像を生成する場合では， 1/2 程度のダウンサンプリングであれば中間視点映像の品質を維持することができるとの報告がある[6]．しかし FTV の場合，中間視点以外のさらに多数の位置・姿勢の仮想視点における映像を合成することが想定される．その場合には視点間の画素対応にはより高い厳密性が求められるため，デプスマップのダウンサンプリングによって三次元情報が欠損し，視点間画素対応の正確性が損なわれた場合に映像品質が大きく低下すると考えられる[7].

図 1 に元解像度の多視点画像と縮小した多視点デプスマップ，縮小した多視点画像と元解像度の多視点デプスマップのそれぞれの組み合わせで生成した仮想視点画像の例を示す．どちらの場合においてもノイズは発生しているが，そのノイズの性質が大きく異なる。多視点デプスマップを縮小した場合は，テクスチャの画質が維持されるため，三次元情報の歪みが少ない被写体内部の画質が高い．しかし，被写体の境界部分においては三次元情報の歪みが大きくなるため，被写体の形状が正しく表現できなくなるというノイズが発生する。一方，多視点画像を縮小した場合は，テクスチャの画質が低下するため，全体的にボケの生じた画像となる。三次元情報は正しく表現されているため，被写体の形状は正しく表現されているが，画像を縮小した際に，被写体間でテクスチャが滲む影響を受け，背景に被写体の一部が現れてしまっている。

デプスマップの欠損に対しては、映像の色成分を考慮したデプスマップのアップサンプリング手法や復元手法の研究が行われている[8][9]．これらの手法では，画像情報や空間的な位置とデプスとの間に相関があると仮定した復元を行うため，正しい三次元情報を復元できる保証はない．また，デプスマップの欠損により視点間の対応付けが行えないため，視点間で欠損した情報を補完し合う事ができない．一方，画像情報の欠損では，視点間の正しい対応関係が分かるため，多視点データであることの特徴を利用して，視点間で欠損しているデータを補完できる可能性が高い．その場合，視点数が多いほど復元性能の向上が期待できるため，非常に多くの視点数を持つ FTV において有効な手法であると考えられる．

そこで本研究では，高解像度のデプスマップを保持し，画像情報においてピクセル数を削減した自由視点画像のデータ表現を提案する．画像のピクセル数を削減したことによる自由視点画像の品質低下を防ぐために，提案手法では，多視点画像を用いた超解像を用いて，視点間で欠損した画

†日本電信電話株式会社 NTT サイバースペース研究所
NTT Cyber Space Laboratories, NTT Corporation

像情報の復元を行う．つまり，提案フレームワークでは，従来のアプローチとは異なり，送信側で多視点画像をダウンサンプリングすることで総ピクセル数の削減を行い，高品質な多視点デプスマップと共に符号化・伝送し，受信側でデプスマップを用いた超解像により画像を復元する．提案データ表現では，一般的な自由視点超解像と異なり，伝送する画像よりも高解像なデプスマップを持つことで，高い画像復元性能を達成する．

図 1　縮小によるデータ量削減で発生するノイズ
左：(a)デプスマップを縮小
右：(b)多視点画像を縮小

3. 提案手法

超解像を利用した自由視点画像表現の概要を図 2 に示す．提案手法では，送信側で多視点画像をダウンサンプリングし，低解像度の多視点画像として多視点デプスマップと共に符号化し伝送する．受信側では多視点デプスマップから視点間のサブピクセルレベルでの対応関係を求め，多視点画像の超解像を行う．こうして復元された多視点画像と多視点デプスマップを用いて自由視点画像を合成する．

提案手法では三次元情報の精度維持による合成精度の向上と，一般にデプスマップの符号化効率が自然画像に比べて良いことから，従来法と比較して総ピクセル数あたりの符号量削減も期待できる．

3.1 多視点デプスマップを用いた多視点画像超解像

デプスマップにより得られる奥行き情報を用いて，各視点の画像の画素値をサブピクセルレベルで対応付けることにより超解像処理を行う．概要を図 3 に示す．

超解像対象視点の高解像度画像を X としたとき，他の視点の高解像度画像 X_k は視点変換行列 F_k を X にかけることにより各画素を移動・合成することで生成できると仮定する．このとき，低解像度の多視点画像 Y_k は X_k をダウンサンプラ D で縮小したものとして表現できる (式 1).

$$Y_k = D_k F_k X \qquad \text{(式 1)}$$

これを(式 2)の最小化問題として解くことで，超解像後の画像 X を求めることが可能となる．

$$\overline{X} = \underset{\underline{X}}{ArgMin}\left\{\left\| D_k F_k \underline{X} - Y_k \right\|_1\right\} \qquad \text{(式 2)}$$

しかしながら，画像にノイズが含まれる場合や対応点にミスマッチがある場合を想定すると，(式 2)の解を安定して収束させることは実際には困難である．そこで本稿における超解像処理は，Bilateral Total Variation (BTV)による正則化項を取り入れたモデルに基づいて行う[10]．この項によって画像全体の変動量に制限を与えることで，安定した原画像の推定が可能となる．BTV による正則化項を取り入れた超解像モデルを(式 3)に示す．

$$\hat{X} = \underset{\underline{X}}{ArgMin}\left[\sum_{k=1}^{N}\left\| D_k F_k \underline{X} - Y_k \right\|_1 + \lambda \sum_{l=-P}^{P}\sum_{m=0}^{P} \alpha^{|m|+|l|}\left\| \underline{X} - S_x^l S_y^m \underline{X} \right\|_1\right] \qquad \text{(式 3)}$$
$$l+m\geq 0$$

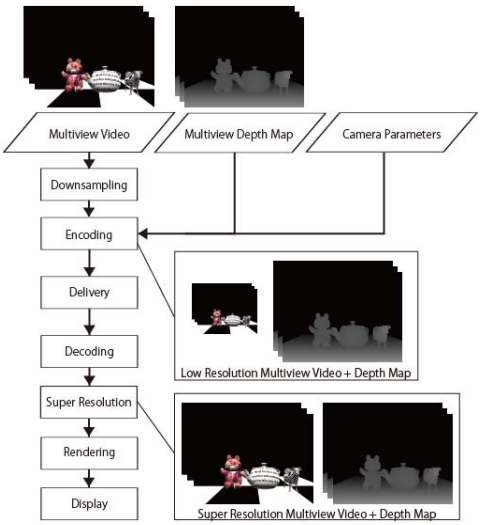


図 2　超解像を用いた自由視点映像表現の概要

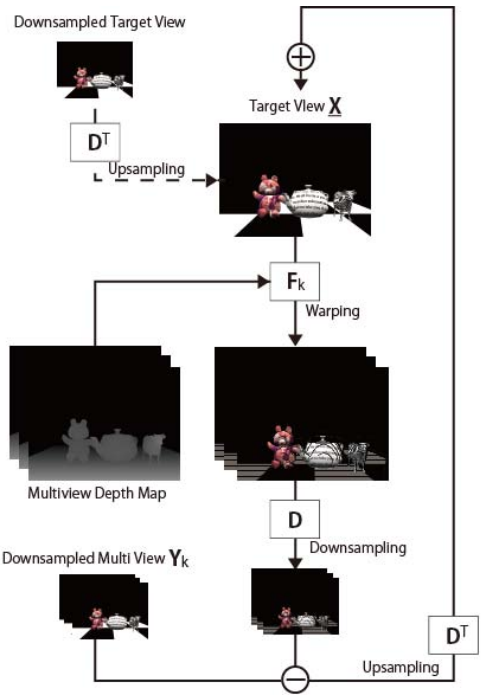


図 3　多視点デプスマップを用いた超解像

S^l_x及びS^m_yは，画像を x 方向に l ピクセル，y 方向に m ピクセルずつ平行移動させる演算子である． BTV による正則化は鋭い勾配に対してペナルティを与えないため，ノイズによる影響を除去しつつエッジを保存する効果が得られる．この問題を最急降下法を利用して解くために展開したものを(式 4)に示す．

$$\underline{\hat{X}}_{n+1} = \underline{\hat{X}}_n - \beta\left\{\sum_{k=1}^{N} F_k^T D_k^T sign(D_k F_k \underline{\hat{X}}_n - Y_k) + \lambda \sum_{l=-P}^{P}\sum_{m=0}^{P} \alpha^{|m|+|l|}[I - S_x^{-l} S_y^{-m}]sign(\underline{\hat{X}}_n - S_x^l S_y^m \underline{\hat{X}}_n)\right\}$$

$l+m\geq 0$

(式 4)

3.2 視点変換行列

視点変換行列 F_k による処理は，図 4 に示すように，X_k の各画素に対して X におけるサブピクセル単位での対応点の近傍画素値の重み付平均を X_k の画素の値とするような処理である．

ブロックマッチング等を用いて対応点を求める一般的な超解像とは異なり，提案データ表現では視点 k におけるデプスマップが存在するため，3D Warping を用いて対応点を求める．特に提案データ表現では，超解像後の解像度に対応する高解像度のデプスマップを有するため，X_k の画素ごとに，対応点を高精度に求めることが可能となる．同定した対応点に基づき，その近傍画素の重みを決定し，図 5 のような疎行列として F_k を定義する．また，重みを決定する際に，対応点間でのデプス値を比較することでオクルージョンを判定し，オクルージョンが発生している場合には全ての重みを 0 にする．

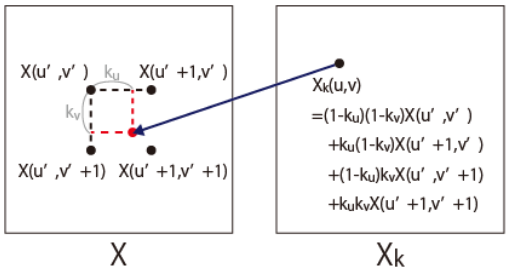
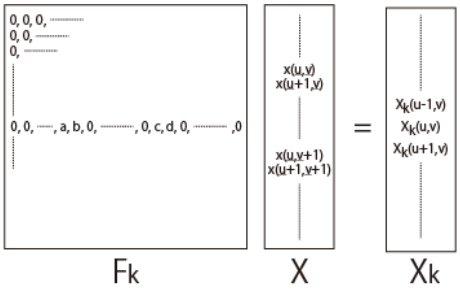


図 4　画素対応関係に基づいた X_k の生成



$X_k(u,v)=aX(\underline{u},\underline{v}) + bX(\underline{u}+1,\underline{v}) + cX(\underline{u},\underline{v}+1) + dX(\underline{u}+1,\underline{v}+1)$

図 5　疎行列 F_k による X_k の生成

3.3 ダウンサンプリングフィルタ

ダウンサンプラは図 6 に示すようにフィルタとサブサンプラに分離できる．ここでのサブサンプラは画素の間引き処理を行うのみであるため，フィルタ特性によってダウンサンプリング性能が決定される．

通常の超解像処理では，入力である低解像度の画像に対して，これを未知の高解像度の画像をダウンサンプリングしたものと仮定する．この場合に用いたフィルタは未知であるため，使用したカメラの撮像プロセス等に基づいてフィルタを仮定し超解像処理を行う．

提案手法では，ダウンサンプリングに用いたフィルタを付加情報として画像データと共に伝送することで，既知のフィルタを用いた超解像処理を行う．このため，画像ごとに復号効率が最大になるように動的にフィルタを設計することができる．

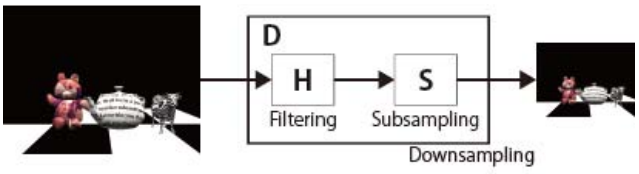


図 6　フィルタとサブサンプラによる縮小

4. 実験

提案手法の有効性を検証するため，シミュレーションにより多視点画像と多視点デプスマップを生成し，本手法を用いて多視点画像のダウンサンプリングと超解像を行った．超解像後の多視点画像と多視点デプスマップを用いて仮想視点画像を生成し，元画像から生成した仮想視点画像と比較して画質評価を行った．

4.1 実験条件

本実験で用いた多視点画像及び多視点デプスマップは，図 7 に示すカメラ配置で，解像度 1024x768[pixel]のものを 8 視点分取得した．この 8 視点は全て被写体に正対する同一平面上に存在している．また，平面の中心から奥行き方向に前進した点を仮想視点位置として設定した．被写体は Teddy，Teapot，Cow にそれぞれテクスチャとして Lena，Text，Border を使用した．取得した多視点画像の一部を図 8 に示す．また，この多視点画像を用いて生成した仮想視点画像の一部を図 9 に示す．多視点画像と多視点デプスマップから仮想視点画像を合成するに当たって 3D Warping を用いる手法を採用した[11]．ただし，正確な比較を行うため合成の前後にノイズ除去などの処理は加えず，また In-painting などの処理も行っていない．

比較のために，表 1 に示す組み合わせで仮想視点画像を生成し，1x1x を正解値として PSNR 値を算出した．縮小率は図 10 のように一辺当たりの率とする．

また，合成に用いる多視点画像・多視点デプスマップをそれぞれ圧縮率[2.0, 1.0, 0.5, 0.4, 0.3, 0.25, 0.2, 0.15, 0.1, 0.05](bit/pixel)で JPEG2000 にて圧縮し，全パターンの組み合わせについて仮想視点画像を生成し，画質評価を行った．エンコーダ・デコーダは kakadu ver6.4 を使用した．

予備実験において，ダウンサンプリングに用いるフィルタとして，バイリニアフィルタとニアレストネイバーフィルタ（フィルタなし）の両方を比較して実験に使用するフィルタを決定した。本稿では，各条件で品質が最大であった多視点画像にはバイリニアフィルタを、デプスマップにはニアレストネイバーフィルタを適用することとした．今回は最もシンプルな 2 種類のフィルタのみを比較して決定したが，3.3 節で述べた通り，提案手法では多視点画像のダウンサンプリングフィルタとして，性能が最大となるものを選択できるため，最適なフィルタの設計に関してはまだ改良の余地がある．

超解像処理に用いたパラメータは，P=2，α=0.7，β=2.0，λ=0.05 である．仮想視点画像超解像処理・仮想視点画像合成処理共にすべての視点画像を用いて行った．

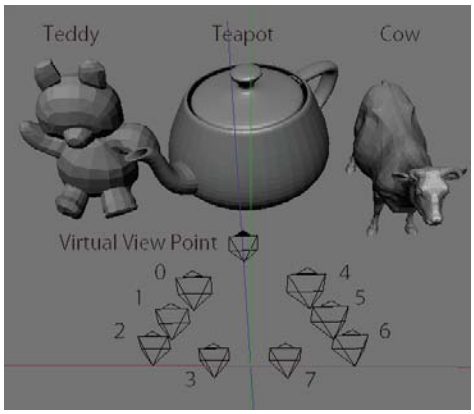


図 7 実験に用いたカメラ配置

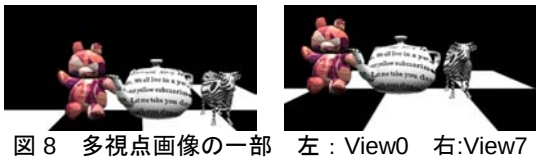


図 8　多視点画像の一部　左：View0　右:View7

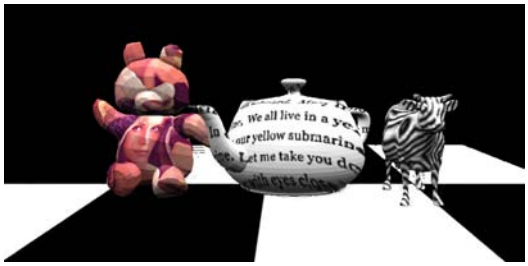


図 9　仮想視点画像

表 1　実験項目

	多視点画像	デプスマップ	超解像処理
1x1x	等倍	等倍	
1x2x	等倍	1/2 縮小	
1x4x	等倍	1/4 縮小	
2x1x	1/2,縮小	等倍	無
4x1x	1/4 縮小	等倍	無
2x1x SR	1/2 縮小	等倍	有
4x1x SR	1/4 縮小	等倍	有

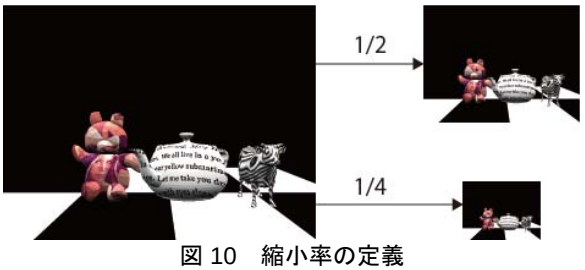


図 10　縮小率の定義

4.2 実験結果と考察

4.2.1 非圧縮画像による比較

非圧縮画像を用いて生成した仮想視点画像の PSNR 値と，生成に用いた多視点画像と多視点デプスマップの合計ピクセル数を 1x1x の場合を N として表 2 に示す．また，生成した仮想視点画像を図 11 に示す．

1/2 縮小の場合，表 2 から，多視点デプスマップを縮小する従来方式では，高い PSNR を達成できないことが分かる．これはデプスマップ上の歪みによる位置ずれが発生するためである．一方，多視点画像を縮小する提案方式では，超解像を行わない場合であっても，幾何的な歪みが少ないため，高い PSNR を達成できているが，テクスチャの解像度が落ちているため，主観品質が低下してしまっている．しかし，超解像を行うことによって，テクスチャを復元し，幾何的に正しく主観品質も高い自由視点画像を生成できた．

1/4 縮小の場合は，超解像を用いた提案データ表現で高い PSNR を達成できたが，テクスチャを十分に復元することができず，従来方式に比べて主観品質が悪い．しかし，提案手法では幾何的に正しい被写体を表現するため，視点を移動させた際の歪みが少ない．つまり，視点位置をインタラクティブに変更するような場合では，本手法により，より高い主観品質の画像を提供できると考える．

表 2　各手法の総ピクセル数と画質

	Pixels	PSNR [dB]
1x1x	N	
1x2x	5N/8	32.0637
1x4x	17N/32	30.4497
2x1x	5N/8	34.4277
4x1x	17N/32	28.6632
2x1x SR	5N/8	41.5107
4x1x SR	17N/32	36.5444

4.2.2 圧縮画像による比較

各手法について，仮想視点画像の PSNR 値と生成に用いた画像・デプスマップの合計符号量から，符号化効率の上界を図 12 に示す．

また，生成した仮想視点画像を図 13，14 に示す．

図 12 から，提案手法の 1/2 縮小の場合での符号化性能は，多視点画像と多視点デプスマップをダウンサンプリングせずに符号化した場合と同等程度である．しかしながら，総ピクセル数は表 2 の通り 5/8 に削減することができた．

従来手法との符号化性能の比較では，高レート部での性能が大きく上回った．これは超解像による復元性能の他に，

デプスマップが同じ解像度の画像に比べて符号化効率が高いことに起因する．

低レート部での性能では従来手法に比べて大きく下回る結果となった．ある程度以上の圧縮率ではデプスマップに対する符号化歪みが縮小による歪みを上回ること，画像に符号化歪みが重畳されることにより超解像の性能が大幅に低下したことが原因として考えられる．

表 2 と図 12 を比較すると，従来手法では符号量が一定以上の組み合わせでは非圧縮画像による PSNR 値にかなり近い値に収束しているのに対し，提案手法では非圧縮画像で得られる値を大きく下回る値で収束している．今回用いた超解像手法では JPEG2000 等一般的な符号化方式を用いて符号化した場合に失われる画像の高周波成分を復元することができないためであると考えられる．また，デプスマップの符号化の際に加わるモスキートノイズによって画素対応に齟齬が生じ，超解像性能が低下した可能性が考えられる．

5. 結論

提案手法を用いて自由視点画像生成に必要なデータを表現することにより，ダウンサンプリングを加えない一般的なデータ表現に比べて同等の符号化効率を得ながらデータ量を大幅に削減することができた．また，デプスマップを縮小する従来手法に比べて，低レート部を除いて，高い画像品質を得ることができた．

今回行った実験では，超解像処理の後に仮想視点画像合成処理を独立に行ったためデコーダ側全体での処理コストは増加したものの，実際には合成映像の品質を復元するために多視点映像をすべて超解像処理する必要はなく，低解像度多視点画像から合成された低解像度仮想視点画像だけを超解像処理する方法を取ることでデコーダ側全体の処理コストを低減することが可能である．

今後の展望としてはまず，オクルージョン等を考慮したより柔軟なダウンサンプリングフィルタの設計により，三次元空間における領域当たりの情報量を平均化することが重要である．例えばオクルージョンの発生により超解像の効果が期待できない領域については元の品質を維持し，全ての視点から観測される領域については超解像に必要な程度までダウンサンプリングする等のアプローチが考えられる．

また，画像の符号化歪みを考慮した超解像手法の検討や，デプスマップに適した符号化手法の検討なども今後の展望としてあげられる．

参考文献

[1] Smolic A, Mueller K, Merkle P, “3D VIDEO AND FREE VIEWPOINT VIDEO – TECHNOLOGIES , APPLICATIONS AND MPEG STANDARDS”, Multimedia and Expo, 2006 IEEE International Conference on (2006).

[2] Zitnick CL, Kang SB, Uyttendaele M, Winder S, Szeliski R, “High-quality video view interpolation using a layered representation”, ACM Transactions on Graphics, Vol.23, No.3 (2004).

[3] ISO/ITEC JTC1/SC29/WG11, “Call for Proposal on 3D Video Coding Technology”, N12036, Geneva (2011).

[4] ISO/ITEC JTC1/SC29/WG11, “Applications and Requirements on 3D Video Coding”, N12035, Geneva (2011).

[5] Yea S, Vetro A, “VIEW SYNTHESIS PREDICTION FOR RATE-OVERHEAD REDUCTION IN FTV”, 3DTV Conference: The True Vision - Capture, Transmission and Display of 3D Video(2008).

[6] Klimaszewski K, Wegner K, Domański M, “Influence of views and depth compression onto quality of synthesized views”, M16758, London (2009).

[7] 福島 慶繁,石橋 豊, “リアルタイム通信のための DIBR による自由視点映像合成”, 電子情報通信学会技術研究報告, Vol.110, No.296 (2008).

[8] Li Y, Sun L, “A NOVEL UPSAMPLING SCHEME FOR DEPTH MAP COMPRESSION IN 3DTV SYSTEM”, Picture Coding Symposium (2010).

[9] Wildeboer MO, Yendo T, Panahpour Tehrani M, Fujii T, Tanimoto M, “COLOR BASED DEPTH UP-SAMPLING FOR DEPTH COMPRESSION”, Picture Coding Symposium (2010).

[10] Farsiu S, Robinson MD, Elad M, Milanfar, “Fast and robust multiframe super resolution”, IEEE TRANSACTIONS ON IMAGE PROCESSING,Vol.13,No10 (2010).

[11] Mori Y, Fukushima N, Fujii T, Tanimoto M, “View generation with 3D warping using depth information for FTV”, Image Communication., Vol. 24, No. 1-2 (2009)

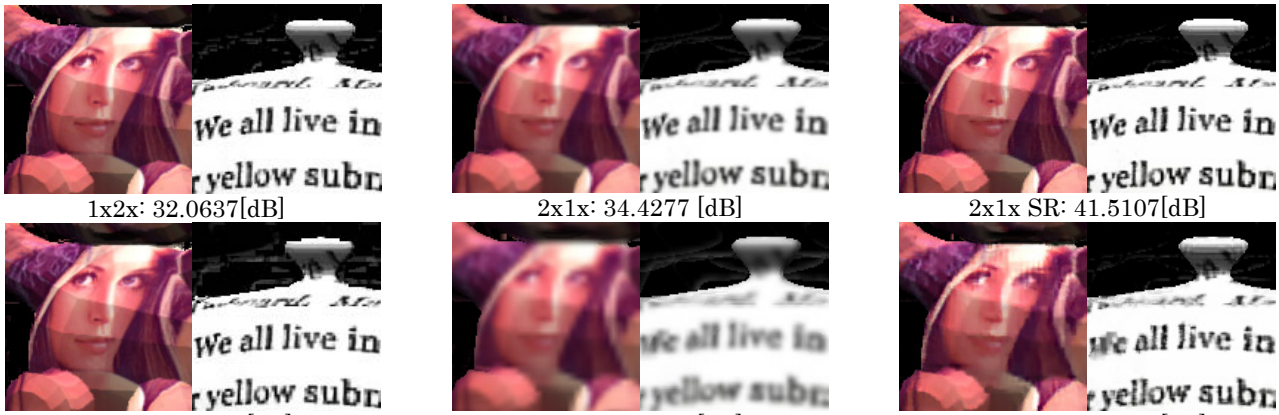


図 11　非圧縮データを用いて生成した仮想視点画像

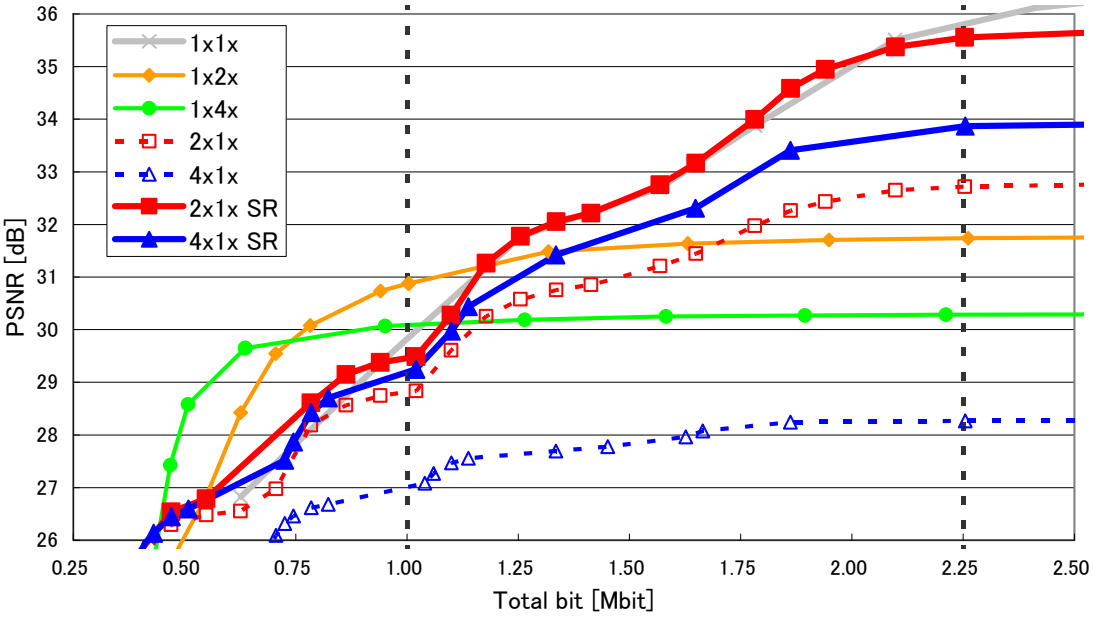
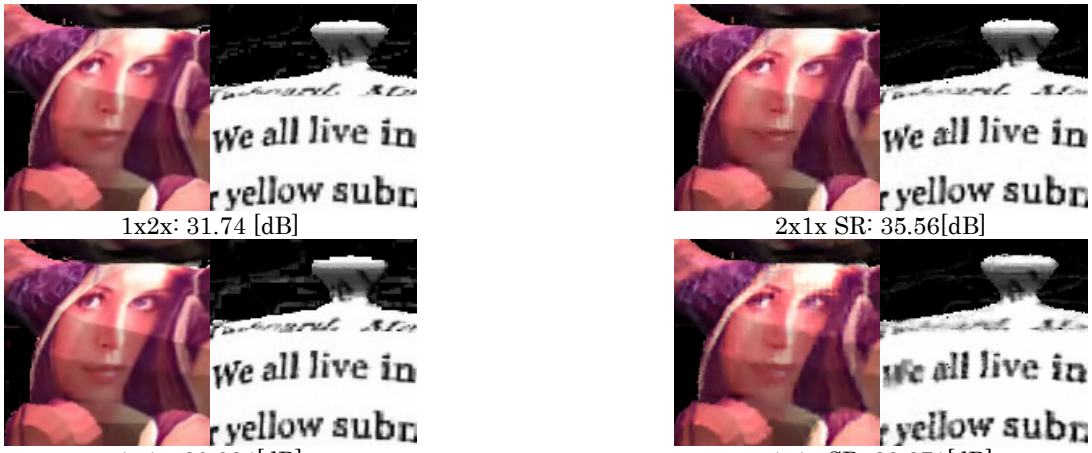


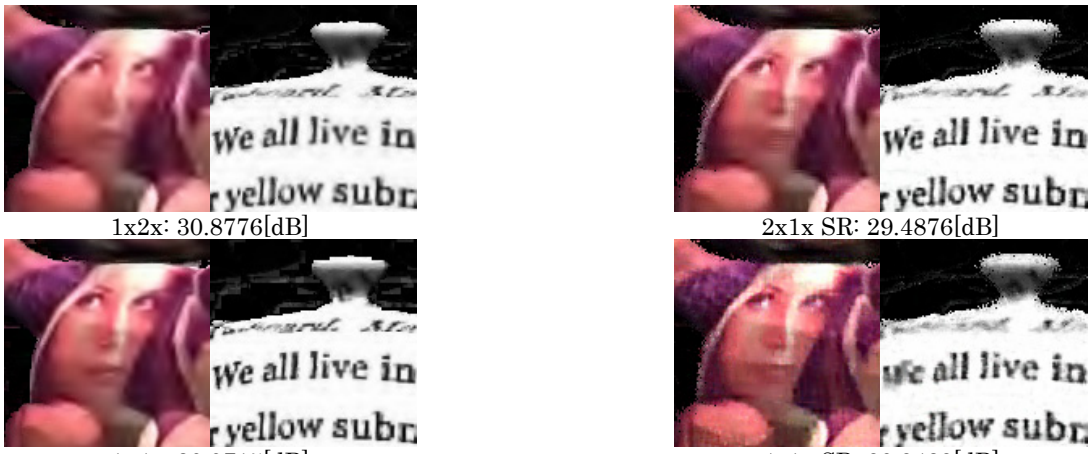
図 12　各手法における符号化効率



1x2x: 31.74 [dB]　　2x1x SR: 35.56[dB]

1x4x: 30.284[dB]　　4x1x SR: 33.871[dB]

図 13　圧縮データを用いて生成した仮想視点画像（Total Size 2.25Mbit）



1x2x: 30.8776[dB]　　2x1x SR: 29.4876[dB]

1x4x: 30.0715[dB]　　4x1x SR: 29.2482[dB]

図 14　圧縮データを用いて生成した仮想視点画像（Total Size 1.0Mbit）