

多視点映像符号化のための 視点合成予測における色補正に関する検討

志水信哉[†], 木全英明[†], 大谷佳光[†]

[†] 日本電信電話株式会社 NTT サイバースペース研究所

〒239-0847 神奈川県横須賀市光の丘 1-1

E-mail : {shimizu.shinya, kimata.hideaki,
ohtani.yoshimitsu}@lab.ntt.co.jp

あらまし

多視点映像の効率的な符号化のために視点合成予測が提案されている。これは既に符号化済みの視点の映像を用いて合成した別の視点の映像を予測信号とすることでカメラ間の冗長性を効率的に取り除くことを目的としている。しかしながら、カメラの個体差や被写体の反射特性の影響で、カメラによって見え方が異なるため、別の視点の映像をワーピングする従来の手法では効率的な映像予測を実現できない。そこで、本報告では適応的フィルタを用いて視点合成画像を補正し、より効率的な多視点映像符号化を実現する方法を提案する。提案する補正手法では、フィルタ係数をブロック毎に復号側で導出することで、補正パラメータを符号化・伝送せずに、被写体に依存した局所的な映像信号の違いを補正する。実験の結果、H.264/AVC Multiview Video Coding に対して最大で約 20%、3 シーケンス平均で約 13%の符号量削減を達成した。

Color Compensation for View Synthesis Prediction in Multiview Video Coding

Shinya SHIMIZU[†], Hideaki KIMATA[†], and Yoshimitsu OHTANI[†]

[†]NTT Cyber Space Laboratories, NTT Corporation

1-1 Hikarino-oka, Yokosuka, Kanagawa 239-0847, JAPAN

E-mail : {shimizu.shinya, kimata.hideaki,
ohtani.yoshimitsu}@lab.ntt.co.jp

Abstract:

View synthesis prediction has been studied to achieve efficient inter-view prediction. View synthesis prediction uses an image synthesized from decoded pictures on the other viewpoints as predicted picture. However, the conventional method has no ability to compensate inter-view mismatches caused by individual specificity of camera and non-Lambert reflection of objects. This paper proposes an adaptive filtering method to perform color compensation in view synthesis prediction. The proposed method estimates an optimal filter block by block to compensate object-dependent local differences. This estimation can be performed at the decoder side, so that no additional information is necessary to encode. The experiments show using proposed method reduces the bitrate by up to 20% and 13% on average for 3 sequences relative to H.264/AVC Multiview Video Coding.

1 はじめに

ユーザが自由にカメラを操作できる自由視点テレビや、シーンの奥行きや立体感の得られる立体映像などの三次元映像が注目を集めている [1, 2]. これは三次元映像を実現するために必要な多視点映像の撮像・処理・表示技術が近年目覚しく発展しているためである. 多視点映像とは複数のカメラを同期させて同じシーンを様々な位置や向きから撮影した映像である.

自由視点テレビを実現する際に、ユーザが指定可能な全ての位置や向きのカメラに対する映像を蓄積、伝送するのは現実的ではない. そのため、限られた数の離散的に配置されたカメラの映像を用いて、任意の位置や向きのカメラの映像を生成する必要がある. また立体映像においても、裸眼で複数の人が同時に鑑賞可能なディスプレイや、ユーザの環境や趣向に合わせてステレオ映像の基線長を適応的に変換できるディスプレイなど、より高度な立体ディスプレイに対応するためには、限られた数のカメラ映像だけを蓄積・伝送し、受信側で必要な映像を生成する必要がある.

多視点映像から任意の視点における映像を生成するためには、非常に密に設置したカメラで撮影した多視点映像を用いるか、シーンのデプス情報を用いる必要があることが知られている [3]. シーンのデプス情報は多視点映像にステレオ法を適用することで推定することができるため、表示側でオンラインで推定しながら任意視点合成を行う方法 [4] と、予めデプスを推定しておき多視点映像と一緒に表示側へ伝送する方法がある [5]. オンラインで推定を行うアプローチでは、表示側で膨大な演算が必要となるだけでなく、ある程度密に配置した多視点映像が必要となってしまうことから、近年では多視点映像と多視点デプスマップを一緒に伝送するアプローチが注目を集めている [6, 7].

多視点デプスマップを追加することによる符号量増大を抑える方法として、視点合成予測を用いた多視点映像符号化がある [8]. 視点合成予測とは、デプスマップによって得られる幾何情報を用いて、別の視点の映像を符号化対象の視点の映像へとワーピングした予測映像を生成することで、多視点映像の符号化効率を向上させる方法である. しかしながら、別の視点の映像信号をワーピングして使用するため、カメラ間で輝度や色の違いがある場合に精度の高い予

測が実現できない.

そこで、本報告では、視点合成予測において色補正を行うことで、多視点映像符号化の効率を向上させる方法を提案する. 第2章で視点合成予測について詳細な説明を行い、第3章で提案する視点合成予測における色補正方法を述べる. 提案手法の性能評価実験は第4章で行い、第5でまとめと今後の課題を述べて本報告を締めくくる.

2 視点合成予測

視点合成予測は多視点映像符号化においてカメラ間予測を効率的に行う方法の1つである. 具体的には、シーンの幾何情報とカメラパラメータを用いて得られる画素ごとのカメラ間の対応関係に従って、別のカメラの映像をワーピングして符号化対象の画像を合成して予測する方法である.

具体的には、まず各画素を式1に従って三次元空間上へ逆投影を行う. この際にカメラから被写体までの距離というシーンの幾何情報を用いることで唯一の3次元点 g が求まる. 次に3次元点 g を式2に従って別のカメラへ再投影することで、別のカメラ上での対応画素位置を得る. そして、得られた画素間の対応関係に応じて画像信号をワーピングして予測画像を生成する. なお、この際に対応画素間の幾何情報が大きく異なる場合はワーピングを行わずに、最後に周辺の背景画素の画像信号を用いてインペインティングを行う.

$$g = R_a^{-1} A_a^{-1} \begin{pmatrix} u_a \\ v_a \\ 1 \end{pmatrix} d - t_a \quad (1)$$

$$k \begin{pmatrix} u_b \\ v_b \\ 1 \end{pmatrix} = A_b R_b (g + t_b) \quad (2)$$

ここで、 A はカメラの内部パラメータ行列、 R と t はカメラの外部パラメータと呼ばれる回転行列と並進ベクトルを表す. ここでの外部パラメータは世界座標系からカメラ座標系への回転及び並進を表している. また、 k はスカラー値であり、 d は対象画素に対して与えられたカメラから被写体までの距離、 (u, v) が画素位置を表す. 添え字 a, b は参照視点と符号化対象視点を表す.

a を参照視点とする方式 [9] も、 b を参照視点とする方式 [8] も提案されているが、本質的には同等な品質の合成映像を生成可能である。なお、本報告では b を参照視点とする視点合成予測を用いる。

このように視点合成予測では、画素ごとに予測信号生成を行うため、ブロック単位の平行予測移動モデルで予測を行う視差補償予測と比較して高精度な予測を行うことが可能である。特に、被写体がカメラの投影面に対して平行ではない場合や各カメラが平行に配置されていない場合など、同じ被写体であっても画素ごとに視差が異なる場合に有効である。

しかしながら、別の視点の画像信号をそのままワーピングして使用するため、視点間で輝度や色の違いがある場合に精度の高い予測を行うことができず、符号化する残差信号が大きくなってしまふ。これまでに我々はこの問題に対して視点合成予測による誤差を空間的または時間的に予測符号化する方法を提案している [10]。しかしこの方法ではどのように視点合成予測の残差を予測したかを示すための情報を符号化する必要が生じ符号量増大を招いてしまふ。

一方、ルックアップテーブルを用いて視点合成画像の色補正を行う方法も提案されている [11]。この方法では視点合成予測の予測残差を小さくすることができるが、フレームごとに補正テーブル符号化を行うため符号量増大を招く。また、フレーム全体に対して1つのテーブルを用いて補正を行うため、局所的な輝度や色の違いが生じている場合には適切な補正を行うことができない。

本報告では、より効率的な視点合成予測を実現するために、適応的フィルタによる色補正、およびエンコーダとデコーダの双方で補正パラメータを推定することで付加情報を発生させない局所的な色補正手法を提案する。

3 適応的フィルタによる色補正視点合成予測

3.1 適応的フィルタ

カメラ間の画像の違いは主にカメラの個体差と被写体が非ランバート反射することによって生じる。カメラの個体差としては、ゲイン、焦点距離、絞り、シャッタースピードなどの違いがあげられる。これらが要

因となって引き起こされるカメラ間の差異は大きく分けて2種類のものが存在する。

1つは色や明るさの違いである。一般的にこの種の違いは1画素ごとの画像信号の関数としてモデル化することができる。ただし、その違いは非線形なものであったり、被写体や空間的な位置によって異なるものであると考えられるためグローバルなモデル化が非常に困難である。

もう1つは空間周波数の違いである。つまりフォーカスが合っているかどうかということである。そのためこの違いは被写体に依存したものである。なお、同じ焦点距離であったとしても、絞りやシャッタースピードによって被写界深度が変化するため、この空間周波数の違いがなくなることはない。

そこで我々はブロック毎に適応的なフィルタを用いた補正を提案する。提案する補正は次の式 3 で表される。

$$Comp[x, y] = \sum_i \sum_j F_{i,j} Syn[x + i, y + j] + o \quad (3)$$

ここで、 $Comp$ は補正後の視点合成画像、 Syn は補正前の視点合成画像、 $F_{i,j}$ はフィルタ係数、 o はオフセット値を表す。

提案手法では、ブロック毎に異なるパラメータ（フィルタ係数 $F_{i,j}$ 、オフセット値 o ）を用いて適応的な補正を行う。局所的に異なるパラメータを用いることで、被写体ごとに異なる補正を行えるだけでなく、本来は非線形な色や明るさの違いをオフセット値によって近似して表現することが可能となる。

3.2 復号側での補正パラメータ導出

提案方式では、ブロック毎にパラメータの異なる補正を行うため、そのパラメータをどのように復号側に伝えるかが問題となる。符号化して伝送する場合、膨大な符号量が必要となるため効率的な符号化を実現することができない。

そこで処理済みの隣接ブロックの情報を用いて、復号側でブロック毎に補正パラメータを導出する方法を提案する。なお、符号化側では復号側と同様の方式で求めたパラメータを使用することで、符号化側と復号側でのミスマッチを防ぐ。

補正は別のカメラの映像を用いて生成された視点合成画像をそのカメラの画像に近づけるために行う。つまり入力画像と補正後の画像の二乗誤差をできる

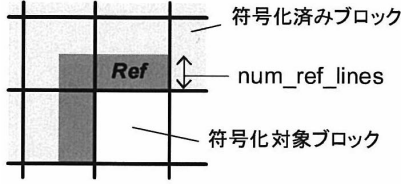


図 1: 補正パラメータ算出のための参照画素

だけ小さくする補正パラメータを導出することが求められる。

ここで、映像符号化とは定められた符号量で表現するという制約下で、入力映像をできるだけ忠実に表現することである。したがって復号画像は入力画像にほぼ等しいと仮定することができる。

そこで提案方式では、処理対象ブロック毎に、隣接する符号化済み画素における視点合成画像と復号画像の誤差が最小になるようにパラメータを導出し、そのパラメータを用いて処理対象ブロックの視点合成画像の補正を行う。

具体的には、補正パラメータを変数として式 4 で表される二乗誤差の最小化問題を、最小二乗法を用いて解くことで補正パラメータを導出する。

$$Err = \sum_{(h,w) \in Ref} (Comp[h,w] - Dec[h,w])^2 \quad (4)$$

なお、 Dec は復号画像を表し、 Ref は処理対象ブロックに隣接する符号化済み画素の集合を表す (図 1)。

4 実験

4.1 実験条件

提案手法を H.264/AVC Amd. Multiview Video Coding (以下 MVC と表す) 検討時の参照ソフトウェアである JMVM (version 8.0) に実装して符号化シミュレーションを行った [12]。

視点合成予測は Inter16x16 モードに 1 ビットのフラグを符号化する方式で実装した。つまり、視点合成予測は 16x16 画素単位でのみ使用可能となっている。なお、視点合成予測が使用される場合は通常の Inter16x16 モードと異なり参照インデックスやベクトルの符号化は行わない。

補正パラメータの導出も 16x16 画素単位で行うものとした。補正におけるフィルタは対象画素を中心

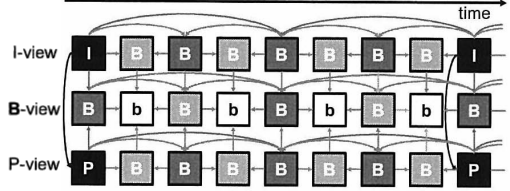


図 2: 参照構造 (GOP Size 8)

とする 5x5 のフィルタを用いた。つまり、補正パラメータ数は 26 個となる。なお、 num_ref_lines は 8 と定義した。

実験には MVC テストシーケンスの breakdancers, akko&kayo, rena の 3 シーケンスを用いた。また、breakdancers のデプスマップに関しては Microsoft が配布しているものを用い [5], akko&kayo と rena に関しては α 拡張による多値グラフカットを用いたステレオアルゴリズムを求めて多視点映像から推定したものを用いた。多視点デプスマップはレートによらず QP36 で符号化した。

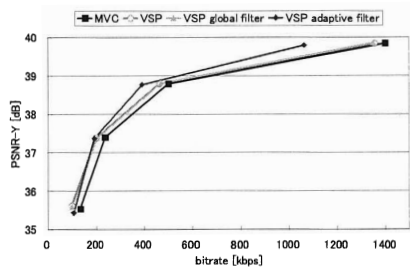
実験におけるカメラ間及び時間方向の参照構造は図 2 に示す GOP サイズが 8 の階層 B ピクチャ構造を使用した。符号化パラメータは表 1 の通りである。

4.2 実験結果

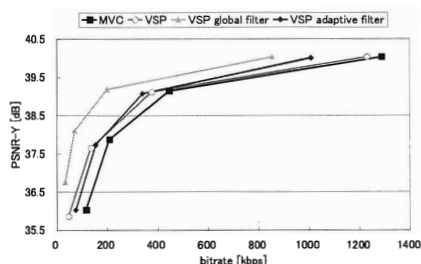
実験結果を図 3,4,5 及び表 2 に示す。比較のために、視点合成予測なし (MVC)、従来の色補正を行わ

表 1: 符号化パラメータ

項目	設定値
GOP size	8
Anchor period	8
Number of reference frames	2
Motion estimation scheme	FME
Entropy coding method	CABAC
Hadamard transform	used
RD-optimized mode decision	used
Layer0 QP	22, 27, 32, 37
Layer1 QP	Layer0 QP + 3
Layer2 QP	Layer1 QP + 1
Layer3 QP	Layer2 QP + 1



(a) P-view



(b) B-view

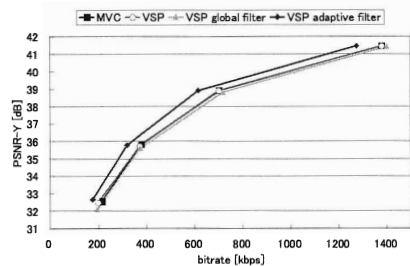
図 3: 実験結果 - RD 曲線 (breakdancers)

ない視点合成予測 (VSP), 式 3 による補正を行うが画像全体に対する補正パラメータを符号化して復号側に通知するもの (VSP global filter), 提案手法 (VSP adaptive filter) の 4 条件の結果を示す。

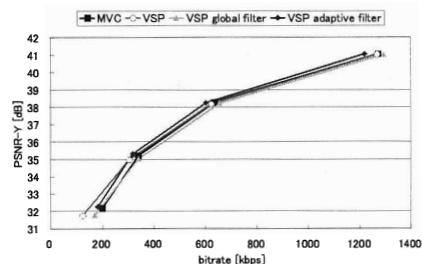
シーケンスによらず, 高レートでの非常に高い性能を発揮し, 最大で約 20% の符号量削減を達成できた。一方, 低レートにおける性能は低く, 特に B-view で

表 2: 実験結果 - Bjøntegaard delta[13]

sequence	filter	BD-Rate	BD-PSNR
breakdancers	no	20.48%	0.29dB
	global	38.40%	0.64dB
	adaptive	20.48%	0.39dB
akko&kayo	no	3.41%	0.12dB
	global	-2.69%	-0.13dB
	adaptive	10.39%	0.50dB
rena	no	5.99%	0.24dB
	global	-8.41%	-0.36dB
	adaptive	8.41%	0.36dB

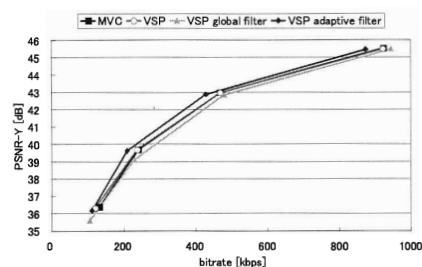


(a) P-view

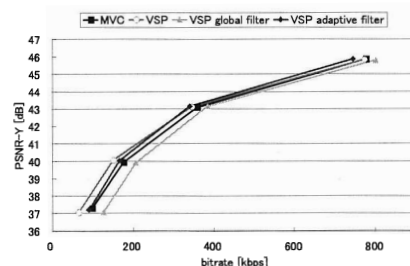


(b) B-view

図 4: 実験結果 - RD 曲線 (akko&kayo)



(a) P-view



(b) B-view

図 5: 実験結果 - RD 曲線 (rena)

は補正を行うことによって性能が劣化してしまっている。そのため全体としては、補正を行わない場合と比較した性能改善は最大で約7%である。

低レートにおける性能低下の原因は、低レートでは復号画像に多くの符号化歪みが生じ、補正パラメータ導出の際の復号画像が入力画像と等しいという仮定が成立しなくなっているからであると考えられる。

また、フレーム全体で同じパラメータを用いて補正する場合と比較して、提案手法のブロック毎にパラメータを変更する方式は、シーケンス平均では高い性能を達成しており有効な手法であると言える。唯一、breakdancersのB-viewにおいてグローバルなパラメータを用いた補正の性能が非常に高い。他のシーケンスにはない傾向であり、更なる調査が必要である。

5 おわりに

本報告では、多視点映像符号化の効率改善のために、ブロック毎に異なるパラメータを用いた補正を行う視点合成予測を提案した。提案手法では、適応的フィルタを用いることでフォーカスの違いに起因する空間周波数の差異も補正する。実験の結果、MVCと比較して最大で約20%、3シーケンス平均で約13%の符号量削減を達成できることが確認できた。

今後は、B-viewや低レートにおいて効率が低下してしまう原因の追求、及びその問題に対処したより効率的な補正方式の検討を行っていく予定である。

参考文献

- [1] Tanimoto, M.: Overview of Free Viewpoint Television, *Signal Processing: Image Communication*, Vol. 21, No. 6, pp. 454–461 (2006).
- [2] Smolic, A., Müller, K., Merkle, P., Fehn, C., Kauff, P., Eisert, P. and Wiegand, T.: 3D Video and Free Viewpoint Video - Technologies, Applications and MPEG Standards, in *Proc. ICME2006*, pp. 2161–2164 (2006).
- [3] Chai, J.-X., Tong, X., Chan, S.-C. and Shum, H.-Y.: Plenoptic Sampling, in *Proc. ACM SIGGRAPH 2000*, pp. 307–318 (2000).
- [4] 高橋桂太, 苗村健: 視点依存奥行きマップの実時間推定に基づく多眼画像からの自由視点画像合成, *映像情報メディア学会誌*, Vol. 60, No. 10, pp. 1611–1622 (2006).
- [5] Zitnick, C. L., Kang, S. B., Uyttendaele, M., Winder, S. and Szeliski, R.: High-quality video view interpolation using a layered representation, *ACM Trans. Graph.*, Vol. 23, No. 3, pp. 600–608 (2004).
- [6] Smolic, A., Müller, K., Dix, K., Merkle, P., Kauff, P. and Wiegand, T.: Intermediate View Interpolation Based on Multiview Video plus Depth for Advanced 3D Video System, in *Proc. ICIP2008*, pp. 2448–2451 (2008).
- [7] ISO/IEC JTC1/SC29/WG11: *Vision on 3D Video Coding* (2009), N10357.
- [8] Yea, S. and Vetro, A.: View Synthesis Prediction for Rate-Overhead Reduction in FTV, in *Proc. 3DTV-Conference*, pp. 145–148 (2008).
- [9] Taguchi, Y. and Naemura, T.: View-dependent Coding of Light Fields Based on Free-viewpoint Image Synthesis, in *Proc. ICIP2006*, pp. 509–512 (2006).
- [10] Shimizu, S., Kitahara, M., Kimata, H., Kamikura, K. and Yashima, Y.: View Scalable Multiview Video Coding using 3-D Warping with Depth Map, *IEEE Trans. Circuits Syst. Video Techn.*, Vol. 17, No. 11, pp. 1485–1495 (2007).
- [11] Yamamoto, K., Kitahara, M., Kimata, H., Yendo, T., Fujii, T., Tanimoto, M., Shimizu, S., Kamikura, K. and Yashima, Y.: Multiview Video Coding Using View Interpolation and Color Correction, *IEEE Trans. Circuits Syst. Video Techn.*, Vol. 17, No. 11, pp. 1436–1449 (2007).
- [12] Pandit, P., Vetro, A. and Chen, Y.: JMVM 8 Software, JVT Doc. JVT-AA208 (2008).
- [13] Bjøntegaard, G.: Calculation of average PSNR differences between RD-curves, VCEG Doc. VCEG-M33 (2001).