

奥行き情報を用いた多視点映像符号化の符号量削減

Rate Reduction for Multi-view Video Coding Using Depth Information

後藤 崇行† 内藤 整‡ 渡辺 裕†
Takayuki GOTO Sei NAITO Hiroshi WATANABE

†早稲田大学大学院 国際情報通信研究科 ‡株式会社 KDDI 研究所
†Graduate School of GITS, Waseda Univ. ‡KDDI R&D Laboratories Inc.

Abstract: Multi-view video coding methods are studied extensively and one of them utilizes camera parameters and depth information. In this paper, we aim at improving this compression method and propose the conversion of disparity vector and depth information each other. Using the converted disparity vectors or depth informations, we can reduce the coding rate of them. Experimental results show that the gain of our proposed method is about 0.5 (dB) in B picture.

1 はじめに

高臨場感を与える映像表現として多視点映像が注目を浴びている。多視点映像は、あるシーンを複数のカメラを使って同時に撮影することで得られるが、一般に、カメラの台数に比例してデータ量が多くなるため、効率的な圧縮が課題となっている。

多視点映像符号化は、従来の MPEG の動き推定の枠組みを視差方向に拡張した方式を中心に数多く検討されている。しかし、視差ベクトルはシーンの奥行きに依存するため、カメラパラメータが与えられている場合には奥行き情報を探索することにより予測効率が向上し、更なる符号化効率の向上が見込めると考えられる。カメラパラメータを使う手法として、参照画像間で視差ベクトルを求め補間する手法 [1] や奥行き情報の探索を従来の予測符号化の枠組みに組み込んだ手法 [2] などが提案されている。

カメラ間の回転や内部パラメータの違いなど、従来の平行移動モデルではモデル化できない領域においては奥行き情報の利用が有効であるが、平行移動モデルが成り立つ領域には従来の平行移動モデルを適用すれば十分であると考えられ、視差ベクトルと奥行き情報をブロック毎に適応的に選択して符号化する方式（以下、視差ベクトル/奥行き情報適応選択符号化）が有効であると考えられる。しかしこの方式では、周囲のブロックからの予測によって視差ベクトル又は奥行き情報を予測符号化する場合、周囲のブロックには視差ベクトルと奥行き情報が混在することとなり、それぞれの予測効率が低下するという問題点がある。そこで、本検討では、視差ベクトルと奥行き情報を相互に変換したものを予測に使用することで予測効率を改善させ、符号量の削減を目指す。

2 視差ベクトル/奥行き情報適応選択符号化

視差ベクトル/奥行き情報適応選択符号化として基本的には [2] の枠組みを用いる。まず従来の動き探索処理を視差方向に適用した視差ベクトル探索を行い、次にエピソード拘束に基づいた奥行き情報の探索を行う。エピソード拘

束による画像座標間の対応は以下のように表現される [3]。

$$\mathbf{X}_{world} = (\mathbf{K}_{cur}\mathbf{R}_{cur})^{-1}(d\mathbf{m} - \mathbf{K}_{cur}\mathbf{t}_{cur}) \quad (1)$$

$$d'\mathbf{m}' = \mathbf{K}_{ref}(\mathbf{R}_{ref}\mathbf{X}_{world} + \mathbf{t}_{ref}) \quad (2)$$

\mathbf{K} , \mathbf{R} , \mathbf{t} はそれぞれカメラの内部行列、回転行列、並進行列を表しており、cur, ref はそれぞれ符号化対象カメラ、参照カメラを表している。 $\mathbf{m} = [u \ v \ 1]^T$, d は符号化対象カメラの画像座標、奥行き情報、 \mathbf{X}_{world} は \mathbf{m} と d から算出される世界座標、 $\mathbf{m}' = [u' \ v' \ 1]^T$ は \mathbf{X}_{world} に対応する参照カメラの画像座標である。 d を決める尺度として一般に以下の SSD (Sum of Square Difference) が用いられる。

$$d = \underset{d}{\operatorname{argmin}} \sum_{u=a}^{a+S-1} \sum_{v=b}^{b+T-1} \{I_{cur}(u, v, d) - I_{ref}(u', v', d')\}^2 \quad (3)$$

I_{cur} , I_{ref} はそれぞれ符号化対象画像の画素値、参照画像の画素値であり、 (S, T) はブロックサイズである。また、奥行き情報探索後にカメラパラメータの誤差を補正する目的で補正ベクトル (c_x, c_y) を探索し、 $\mathbf{m} = [u - c_x \ v - c_y \ 1]^T$ として作用する。

それぞれの探索結果から得られるレート歪みコストから、各ブロック毎に視差ベクトルを使用するか奥行き情報を使用するか決定する。奥行き情報と補正ベクトルの符号化は、基本的に視差ベクトルの符号化と同じ方法でエントロピー符号化する。

3 提案方式

視差ベクトルと奥行き情報が混在する状況においてこれらの発生符号量を低減させるために、視差ベクトルと奥行き情報を相互に変換したものを予測に使用する手法を提案する。視差ベクトルと奥行き情報はエピソード拘束により変換可能であるため、提案方式はこの性質を利用する。奥行き情報から視差ベクトルへの変換は、式 (1) 及び (2) を適用すれば良い。一方、視差ベクトルから奥行き情報への変換は以下の式を用いる。

$$d = \frac{(\mathbf{A}_3 - \mathbf{A}_1) \cdot \mathbf{B} + C_1 - u'C_3}{(u'\mathbf{A}_3 - \mathbf{A}_1) \cdot \mathbf{m}} \quad (4)$$

Table 1: Coding Condition

JM version	11.0
Sequence	Breakdancers
Number of Cameras	5
DC Search Range	± 32 [pixel]
QP (Fixed)	28,32,36,40
Entropy Coding	CABAC
Direct Mode	Spatial



Figure 1: Prediction Structure

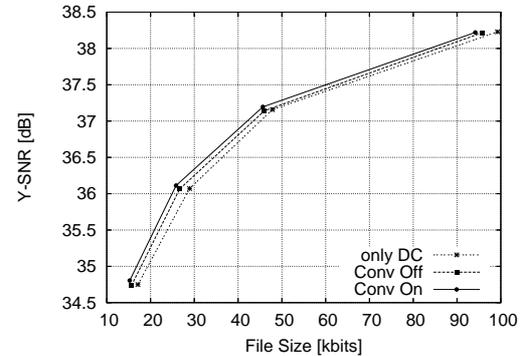
但し, A_i は $K_{ref}R_{ref}(K_{cur}R_{cur})^{-1}$ の i 行成分, $B = K_{cur}t_{cur}$, $C = [C_1 C_2 C_3]^T = K_{ref}t_{ref}$ である. すなわち, 参照画像内のエピポーラ線上にある対応点の x 座標を u' として近似することで奥行き情報を算出する. この理由は, カメラは水平方向に配置されることが多く, この場合のエピポーラ線は画像座標の x 軸とほぼ平行になるからである. 周囲がイントラ等で予測が行えない場合には, 奥行き情報の探索範囲の最大値と最小値の中間値を予測値とする. 画素毎に変換された視差ベクトル又は奥行き情報が算出されるため, 最小の予測ブロック単位である 4×4 毎に平均化したものを格納する.

4 実験と考察

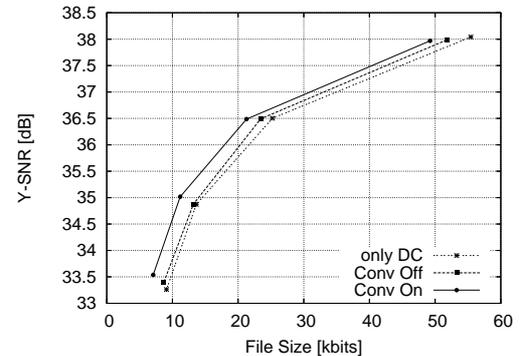
提案方式を H.264 の参照ソフトウェア JM (Joint Model) [4] に組み込み評価を行った. 符号化条件を Table 1 に示す. シーケンスは Breakdancers を用い, カメラ配置は水平方向にアーク状であるためカメラ間で回転が存在している. 奥行き情報は式 (1) における d を 5 から 135 まで 1 間隔で, 補正ベクトルは ± 2 [pixel] の範囲で $1/4$ 画素精度で探索を行った. 予測構造は, 片方向予測と双方向予測の効果を調査するために, Figure 1 の構造を用いた¹.

視点間予測のみを行う P ピクチャと B ピクチャにおける結果を Figure 2 示す. "only DC" が視差ベクトル探索のみ, "Conv Off" が奥行き情報探索は行わぬが視差ベクトル・奥行き情報変換を行わない場合, "Conv On" が奥行き情報探索を行い視差ベクトル・奥行き情報変換を行う場合である. 図から分かるように, P ピクチャではほとんど改善が見受けられないが, B ピクチャにおいて約 0.5[dB] の改善が見受けられる. この理由として, P ピクチャは, 片方向予測によるオクルージョン及び参照カメラとの距離が大きいため疎らに視差ベクトル又は奥行き情報が選択され, 変換の効果が薄かったからであると考えられる. 一方 B ピクチャは, 双方向からの予測によって視差ベクトル又は奥行き情報が数多く選択され, 変換が効果的に行

¹ 図中例えば "P0" とは, P ピクチャ符号化で, カメラ番号が 0 であることを意味している.



(a) P picture (P0)



(b) B picture (B1)

Figure 2: Experimental Results

われて予測効率が向上し, 符号量の削減に寄与したからであると考えられる.

5 おわりに

視差ベクトルと奥行き情報を適応的に用いる多視点映像符号化において, 視差ベクトルと奥行き情報を変換しながら予測を行うことでこれらの発生符号量の削減を行った. 今後は, B ピクチャ符号化時の合成方法や重み付け予測と併せた検討など画質改善手法について検討を行っていく予定である.

参考文献

- [1] M.Kitahara *et al.*: "Multi-view Video Coding using View Interpolation and Reference Picture Selection," IEEE ICME, Jul. 2006.
- [2] S.Yea *et al.*: "Report on Core Experiment CE3 of Multi-view Coding," M13695 MPEG Document, Jul. 2006.
- [3] R.Hartley *et al.*: "Multiple View Geometry in Computer Vision," CAMBRIDGE UNIVERSITY PRESS, 2000.
- [4] <http://iphone.hhi.de/suehring/tml/>

早稲田大学大学院 国際情報通信研究科
〒 367-0035 埼玉県本庄市西富田大久保山 1011
Phone: 0495-24-6143, Fax: 0495-24-6645
E-mail: goto@tom.comm.waseda.ac.jp