

自由視点テレビのための光線空間情報圧縮

岡慎也† ナ・バンチャンプリム† 藤井俊彰† 谷本正幸†

†名古屋大学 工学研究科 〒464-8603 名古屋市千種区不老町

E-mail: †{oka.purim}@tanimoto.nuee.nagoya-u.ac.jp †{fujii.tanimoto}@nuee.nagoya-u.ac.jp

あらまし 本研究室では、ユーザーが視点を自由に移しながら様々な場所から情景を見ることのできる「自由視点テレビ」を提案している。我々はこのテレビの実現に光線空間法を用いているが、非常に多くの情報量を有するため光線空間情報の圧縮は必要不可欠な技術となる。本研究では時間的に変化する光線空間情報は空間領域、時間領域において非常に大きな冗長性を有することに着目して両軸に対してフレーム間予測を行い、従来法と同じく予測誤差をDCTで符号化するハイブリッド符号化を適応し情報量の削減を試み、圧縮率の改善を図ることができた。

キーワード 自由視点テレビ、光線空間法、ハイブリッド符号化

Ray-Space Coding for FTV

Shinya Oka Purim Na Bangchang Toshiaki Fujii and Masayuki Tanimoto

Graduate School of Engineering, Nagoya University Furo-cho, Chigusa-ku, Nagoya, 464-8603 Japan

E-mail: †{oka.purim}@tanimoto.nuee.nagoya-u.ac.jp †{fujii.tanimoto}@nuee.nagoya-u.ac.jp

Abstract We proposed “Free Viewpoint Television (FTV)” that user can watch the scene from arbitrary viewpoints. For implementation of this Television, we use Ray-Space Method, but this method requires much data, so the compression is essential technology for FTV. In this study, we proposed hybrid coding using inter-image prediction in temporal and spatial domain and then applying DCT coding of residual error, focusing the redundancy in both domains, then improved compression ratio.

Keyword Free Viewpoint Television Ray Space Method Hybrid Coding

1. はじめに

近年、コンピュータやネットワークの発展により3次元空間の情報を利用した様々なコミュニケーションが可能となり、バーチャルリアリティ等の分野において様々な応用が実現することが可能となった。

そのような応用のひとつとして本研究室では従来のテレビを拡張しユーザーが見たい視点で様々な場所から遠隔地のシーンを見ることが出来る自由視点テレビを提案してきた[1]。この自由視点テレビの実現の技術として光線空間法を用いている。この手法を用いることによって従来のポリゴンなどを用いたモデルベースによる3次元空間よりもリアリティのある3次元空間を表現することが可能になる。

FTVの基本的なシステム構成を図1に示す。FTVはリアリティのある画像を表示するために光線空間法という3次元の表現法を用いている。そのためこのシステムは入力として複数台のカメラ

から画像を取得しそこから光線空間を構築しインターネットなど介して伝送を行い、受信側の機器はそれらの画像とユーザーの入力から任意の視点の画像を生成する。

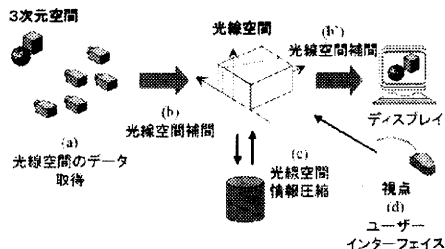


図1 自由視点テレビシステム

このようなシステムにおいて複数のユーザーからのランダムな入力に対して任意の画像を生成するためにはカメラから入力された全ての画像情報

が受信側の機器に伝送されなければならない。このことは動画像の情報量にカメラの台数を乗じた大きさのデータを受信側に送ることに相当し、現在や近い将来のネットワークの技術を用いてもリアルタイムに受信側の機器に伝送することは非常に困難である。そのため、そのままの光線空間データを受信側に伝送するのではなく圧縮を行いインターネットに配信可能なデータ量にすることによってこの問題を解決する必要がある。

実際の光線空間データは図2に示すように一つのシーンを複数のカメラから撮影した画像からなるため空間領域において非常に相関性をもつ。またある一台のカメラについて着目したときに時間領域における変化は動画像における画像の変化と同様であるため時間領域においても非常に大きな相関性をもつことが考えられる。そのため、本研究においては時間的に変化する光線空間のこのような性質に着目し、圧縮を試みる。

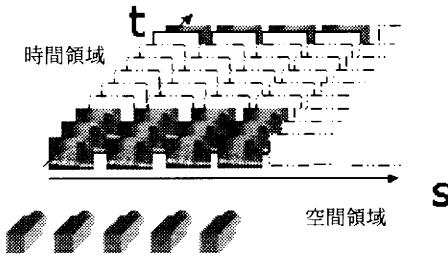


図2 光線空間データの取得

2, 光線空間法

光線空間法とは、空間中を伝搬する光線の情報を用いて3次元の空間情報を記述する方法である。図3に示すような実空間中の基準面を通過する光線は通過位置 (x, y) と通過方向 (θ, ϕ) の4つのパラメータによって一意に表現することができる[2]。すなわち、この空間中の光線は4次元空間 (x, y, θ, ϕ) の内の点に対応させることができる。次にこの3次元空間内の各点にそれぞれの光線の輝度 f を対応させる。これにより、この3次元空間内の光線情報を $f(x, y, \theta, \phi)$ と表すことができる。

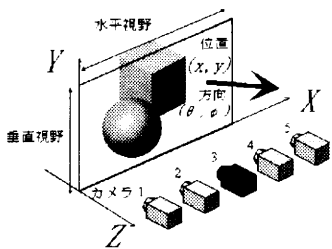


図3 光線空間の定義

このように実空間をパラメータ化すると基準面に水平に配置されたカメラから撮影された画像は $x, y, \tan \theta$ の軸を持つ3次元空間において図4の点線で表される断面に位置していることになる。この空間から任意の面を切り取ることによって任意の画像を生成することが可能となる。加えて生成画像はカメラから撮影された画像を用いているためポリゴンなどのモデルベースの画像に比べて人工的な不自然さのない画像をレンダリングすることが可能である。このような利点から自由視点テレビに光線空間法を用いている。

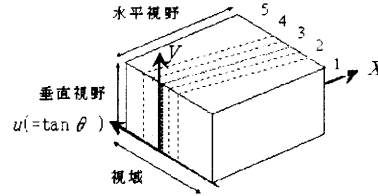


図4 撮影された画像の位置

3 従来の光線空間情報圧縮

近年、光線空間の情報圧縮に対する様々な研究がなされてきた。多次元DCTを用いた圧縮[3]や、 $x - \tan \theta$ 面に対してフラクタル符号化を適応したもの[4]や光線空間とはパラメータのとりからが異なる“light field”[5][6]において画像面にハイブリッド符号化を適応したもの[7]などがある。本研究室においてもベクトル量子化を用いた圧縮[8]やサブバンド符号化[9]を用いた圧縮方法を提案してきたが、いずれにおいても時間的に変化する光線空間情報の時間軸における冗長性に着目したものでないといえる。

4 提案手法

4.1 M (Multi directional) ピクチャ

図2のように時間的に変化する光線空間を考えたときにそれぞれ時間方向に動き補償[10]、空間方向にフレーム間予測[11]のみを行った結果は図6のようになる。

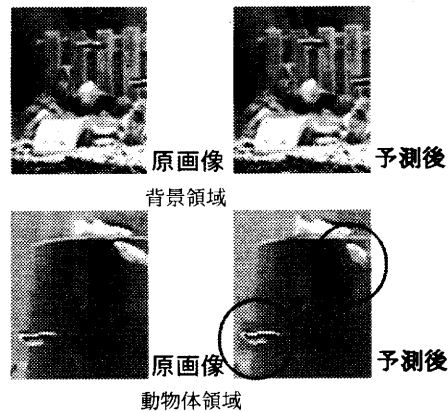


図6(a) 動き補償の結果

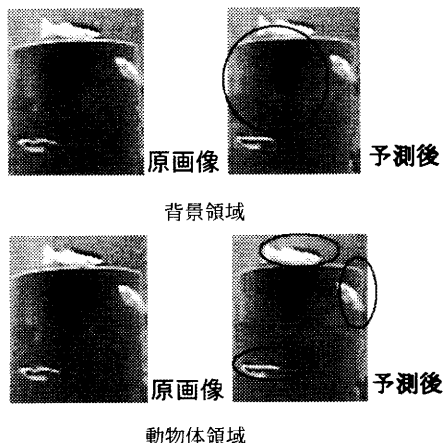


図 6 (b) 視差補償の結果

図 6 (a) のように時間軸にフレーム間予測を行った場合、背景などの変化の少ないところは前後の時刻の画像から効率的に予測が可能となる。しかし、動物体領域における予測後は動物体の前後の時刻における背景がそのまま移動しブロック単位でノイズが発生する。一方、図 6 (b) のように空間領域においてのみフレーム間予測を行うと背景領域はカメラの位置の移動よりブロック状のノイズが発生するが、動物体は図 6 (a) の予測後画像のものよりもブロック状のノイズが少ないものが得られることがわかる。

以上のことから光線空間は時間軸にも空間軸にも画像同士が非常に類似しているため、動き（視差）予測を両軸に適応することによって高い圧縮率を得ることが可能であると考えられる。また、両軸からの予測を行うことによって一方の軸における各々の画像の変化が激しい状況においても、もう一方の軸における画像の変化が小さいものであれば、片方の軸のみの予測よりも効率的な予測が可能であると考えられる。そこで本研究においてはこのように両軸から予測をおこなうピクチャを複数のピクチャを用いるので図 7 に示す M (Multi picture) ピクチャを提案し、その圧縮効果を調べる。

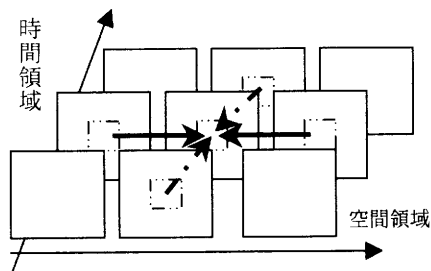


図 7 M (Multi directional) フレーム

4.2 符号器の構成

動き補償を用いて動画を効率的に圧縮する符号化手法としてハイブリッド符号化がある。この手法はフレーム内の画像は直交変換である DCT によって空間的な冗長性を削減し時間的な冗長性は動き補償によって削減する。この手法を用いた符号器は図 8 のような構成からなる。

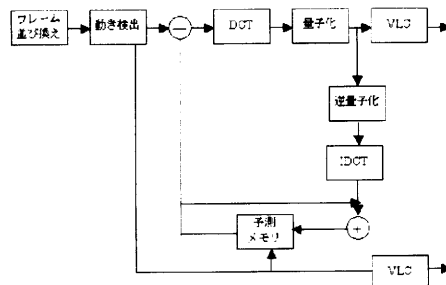


図 8 符号器の構成

この符号器において動きの推定による動ベクトルの値は符号器内に設置されたローカルデコーダによって復元された画像が予測メモリに蓄積されその画像と入力画像との比較を行うことによって算出される。MPEG 1, 2 に使用されている B フレームの場合、動き補償にはこの予測メモリと過去に時刻と将来の時刻の画像を符号化したものをローカルデコーダで復元し蓄積したものと比較して動き推定を行う。本研究における M フレームを採用した符号器の場合、この予測メモリに過去、将来の時刻の画像、右のカメラ、左のカメラの画像の復元画像を予測メモリに蓄積し、その他の機能を実現する構成は同じものを使用する。

4.3 ブロック符号化モード

MPEG などのアルゴリズムにおいては画像を 16×16 画素のマクロブロックに分割し、そのマクロブロック単位でハイブリッド符号化を適応している [12]。本研究において提案する符号器も同様に分割をおこない、ハイブリッド符号化を適応するが、各々のピクチャタイプによって有する符号化モードが異なる。以下に提案する符号化器の符号化モードとそれを有するピクチャタイプとの対応を示す。(I, P, B, はそれぞれ I ピクチャ, P ピクチャ, B ピクチャ。M は M ピクチャ)

- 1、イントラモード (I, P, B, M)
全てのマクロブロックを DCT で符号化する。
- 2、前向き予測 (P, B, M)
時間軸が前のフレームから動き補償
- 3、前向き予測 + DCT (P, B, M)
前向き動き補償後の誤差を DCT で符号化
- 4、後ろ向き予測 (B, M)
時間軸が後ろのフレームから動き補償
- 5、後ろ向き予測 + DCT (B, M)
後ろ向き動き補償後、誤差を DCT で符号化
- 6、右向き予測 (M)

空間軸において右の画像から視差補償
7、右向き予測+DCT (M)

右向き視差補償後、誤差を DCT で符号化

8、左向き予測 (M)

空間軸において右の画像から視差補償

9、左向き予測+DCT (M)

左向き視差補償後、誤差を DCT で符号化

10、補間 (B, M)

B ピクチャの場合、時間軸に前向きと後ろ向きの動き補償後の画素を足して2で割る。

M ピクチャの場合、時間、空間の4つの動き視差ベクトルで補償後足して4で割る。

11、補間+DCT (B, M)

補間後、予測誤差を DCT で符号化

これらのモードのいずれを用いるかの基準として本研究においては圧縮後のビットレートと画像の歪み度の一つのパラメータで制御することが可能である(1)式に示すラグランジュの乗算器を用いた[13].

$$\min_{i=1..11} \{D_i + \lambda R_i\}. \quad (1)$$

式(1)における R はビットレート、D は画像の歪み度を示しているが、提案した符号器においては R はマクロブロック1つを i 番目のモード (i は 1 ~ 11 までのモードのいずれか) で符号化したときのビット数、D は i 番目の符号化モードによって生じた(2)式で計算されるマクロブロック内の SSD (Summed Square Difference) である。

$$SSD = \sum_{j=0}^{15} \sum_{i=0}^{15} (I(i, j) - I'(i, j))^2 \quad (2)$$

(i,j はそれぞれ画像内の 16×16 ブロックのインデックス)

また本研究においてはスライスの先頭に付加する量子化ステップは式(2)で表現される

$$Q = \sqrt{\frac{\lambda}{0.85}} \quad (3)$$

の Q の値を整数にしたものを用いた。

4.4 マクロブロックタイプ

マクロブロックに前向きや後ろ向き、イントラ、補間などいずれの処理を行ったかを判断するためのマクロブロックタイプを付加しなければならない。本研究において提案する符号器も I, P, B ピクチャは MPEG 2 と同じマクロブロックタイプを使用した。M ピクチャは右向き、左向き予測が加わるため 11 種類の符号化モードを識別するためビットを書き込まなければならない。本研究においては符号化モードが等確率で発生するとして 4 ビットを割り当て 1 から 11 までの情報を書き込んだ。

4.5 マクロブロックアドレス

符号化された同じ水平列内のマクロブロックは

図9のような細長い帯上のスライスという領域に接続される。

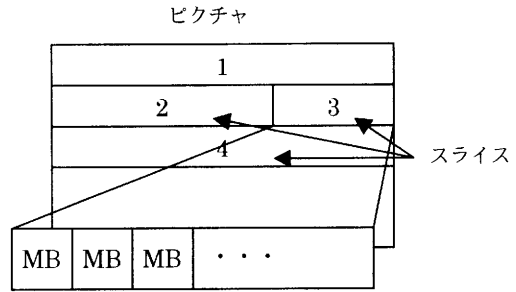


図9 スライス構造

その際、同じスライス上で連続するマクロブロックの情報を効率的にスキップする為、マクロブロックアドレスが設けられている。P は NonMC, B は予測方向と動ベクトルが一致したときにマクロブロックをスキップするが M ピクチャにおいても B ピクチャ同様、予測方向と動ベクトルの一致したときにスキップをおこなった。

4.6 符号化順序

提案する符号器の圧縮順序を示す。図10は B ピクチャの数が2の場合の GOP 示すが B ピクチャはユーザーが任意の数に指定できるものとする。

ピクチャの位置を(空間(s)軸の位置、時間(t)軸の位置)で表現するとすると各々のピクチャを符号化する順序は

I(0,0), I(3,0), P(0,3), P(3,3), B(0,1), B(0,2), B(3,1), B(3,2), B(1,0), B(2,0), B(1,3), B(2,3), M(1,1), M(1,2), M(2,1), M(2,2) となる。t

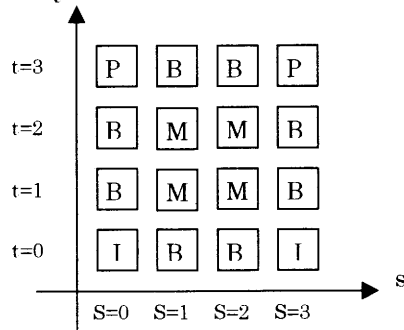


図10 符号化順序

まず、図10に示すように(0,0)と(3,0)の画像を I ピクチャとして符号化し、(0,3)と(3,3)の画像を P ピクチャとして符号化する。その後、(0,0)と(0,3)に対して(0,1), (0,2), また(3,0), (3,3)の画像に対して(3,1), (3,2)の画像を B ピクチャとして符号化する。符号化順をこのようにしたのは復号器が s=1, 2 の画像を規定の時間内に復号できなかった場合にユーザーの入力に対して先に復号が完了した s=0, 3 を表示することができるためである。また

得られたピクチャは M ピクチャの予測のためにローカルデコーダされ符号器内の予測メモリに書き込まれる。

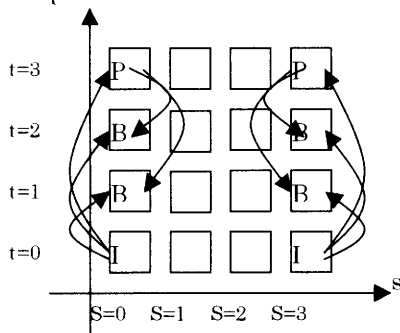


図 11 符号化順序 1

次に先にローカルデコーダで復号が済んだ(0,0)(3,0)を予測フレームとして(1,0)(2,0)の画像を B ピクチャとして予測する。同様に(1,3)(2,3)の画像を(3,0)(3,3)を予測フレームとして B ピクチャの予測を行う。同様に B ピクチャは M ピクチャの予測のために予測メモリに置いておく。

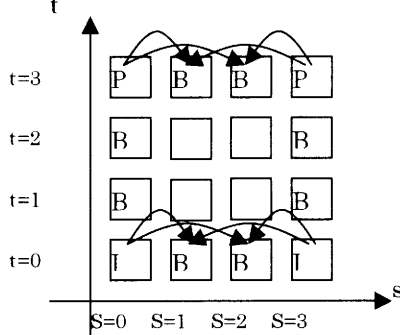


図 12 符号化順序 2

最後に現時点で復号化された B フレームを用いて M フレームの予測を行う。(1,1)は(0,1)(3,1)(1,0)(1,3)(2,1)は(0,1)(3,1)(2,0)(2,3)。(1,2)は(0,2)(3,2)(1,0)(1,3)。(2,2)は(0,2)(3,2)(2,0)(2,3)から各々の予測を行う。

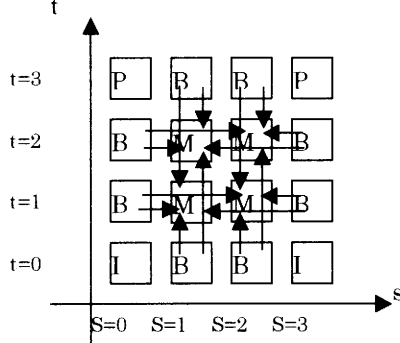


図 13 符号化順序 3

これから得た多くのマクロブロックはスライス状に分け、符号化順に送信する。

5.1 実験

以上のことに基づいた符号器を作成しその性能を評価した。

実験に使用した 3 つの符号器のいずれも予測誤差を DCT で符号化するハイブリッド符号化を適応する。その上で、対比する符号器として時間領域のみのフレーム間予測、空間領域のみにフレーム間予測をもちいた符号器、今回提案した M ピクチャを用いた時間、空間にフレーム間予測を行う符号器である。入力画像としては図 14 に示すような画像を撮影した。同一のカメラを使用しているため、カメラパラメータは同じである。(1枚当たりの空間領域の視差(2.5mm)フレームレート 30fps)

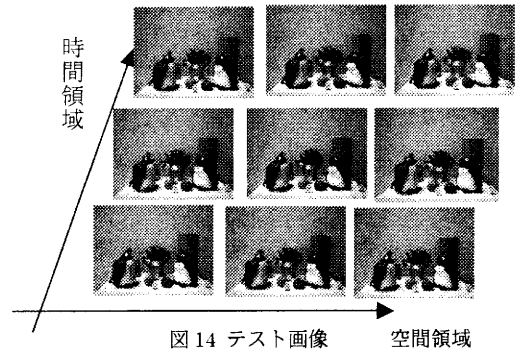


図 14 テスト画像 空間領域

このテスト画像を用いて(3×3の時間、空間軸に画像でなる光線空間)符号化を行った結果を図 15 に示す。

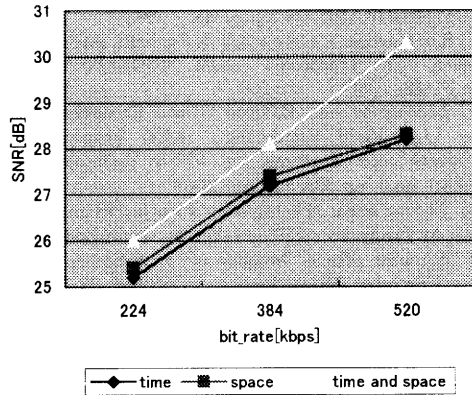


図 15 3 つの符号器の符号化特性

図 15 において time は時間領域のみに予測を行った場合, space は空間領域のみに予測を行った場合, time and space は今回の提案した量軸の予測をお

行った符号器の特性を示す。結果より空間、時間
のみの補償を用いた符号器よりも高い圧縮率が得
られていることがわかる。このグラフから提案手
法は高い復元品質に対しても低いビットレートで
予測可能であることがわかる。

6 まとめ

本研究では時間的に変化する光線空間が時間領
域と空間領域に大きな相関性を持つことに着目し、
時間、空間軸に対するフレーム間予測符号化を行
うことにより、片方のみの軸のみを用いて予測符
号化したものよりも効率的な圧縮を実現できるこ
とを確かめた。

今後は符号化モードの発生確率を求めてマクロ
ブロックタイプに適切な可変長符号を割り当てる
ことや、光線空間の時間軸、空間軸の解像度によ
り適応的に選択するフレームを最適化するなどの
改善を行っていく。

- [1] P. Na Bangchang, 藤井, 谷本, “自由視点テレビ
のためのリアルタイム処理システム”, 信学技
報, cs2002-132, pp. 1-4, Dec. 2002
- [2] 藤井, 金子, 原島, “光線群による3次元空間情報
の表現とその応用”, テレビ誌, vol. 50, no. 9,
pp. 1312-1318, Sept. 1996
- [3] 高野, 苗村, 金子, 原島, “光線情報に基づく空間符
号化 - 光線情報の局所的展開 -”, 映像情報メデ
ィア学会誌, vol. 52, No. 9, pp. 1321-1327, Sept. 1998
- [4] 苗村, 原島, “自己相似モデリングによる多眼三次
元画像の補間と情報圧縮”, vol. 48, No. 10,
pp. 1215-1221, Oct. 1994
- [5] M. Levoy and P. Hanrahan, “Light field rendering”,
ACM SIGGRAPH '96, pp. 31-42, 1996
- [6] S. J. Gortler, R. Grzeszczuk, R. Szeliski, and M. F.
Cohen, “The Lumigraph”, ACM SIGGRAPH '96,
pp. 31-42, 1996
- [7] M. Magnor and B. Girod, “Data Compression
for Light-Field Rendering”, IEEE Trans on
Circuits and Systems for Video Technology,
vol. 10, no. 3, pp. 338-343, Apr. 2000
- [8] M. Kawaura, T. Ishigami, T. Fujii, T. Kimoto
and M. Tanimoto, “Efficient Vector Quantization
of Epipolar Plane Images of Ray Space by
Dividing into Oblique Blocks”, PCS 2001, TP1-11,
pp. 203-206
- [9] Yutaka Hayashi, T. Fujii, and M. Tanimoto,
“Coding for Ray Space based on SPECK”, PSC
Nov. 2003
- [10] 片山, “MPEG Video 技術” <http://home.catv.ne.jp/dd/pub/book/mpeg.html#3>
- [11] M. Magnor “Geometry-Adaptive MultiView
Coding Techniques for Image-based Rendering
Shaker Verlag Aachen, Germany 2001
- [12] ISO/IEC/CD 13818-2 “Generic coding of

moving pictures and associated audio,
ISO/IEC/JTC1/SC29/WG11

[13] G. Sullivan and T. Wiegand “Rate-distortion
optimization for video compression”, IEEE
Signal Processing Mag., pp. 74-90, Nov. 1997