

曲面形状の動的更新背景バッファの導入による 自由視点 VoD システムの改良

石川 彰夫 松村 篤志 川田 亮一 小池 淳

株式会社 K D D I 研究所 〒356-8502 埼玉県上福岡市大原 2-1-15

E-mail: ao-ishikawa@kddilabs.jp, matsumura@kddilabs.jp, kawada@kddilabs.jp, koike@kddilabs.jp

あらまし 参照映像と奥行き情報を用いて生成される自由視点映像では、同一フレーム内において隠蔽領域の情報が得られないことによる画質の低下が課題であった。そこで、筆者らは先に、この問題を解決するため、動的に更新される背景バッファを用いて出力映像を補完する手法を提案した。しかし、この手法では、前景から背景まで連続的に存在する地面などの上の隠蔽領域がうまく補完されないという問題があった。そこで、本検討では、このような場合でも高精度な補完を可能とするため、従来手法で用いていた平面の背景バッファの代わりに、曲面などの多様な背景の表現を可能とする背景バッファを用いた手法を提案する。実験により提案手法の有効性を示す。

キーワード 自由視点映像、奥行きマップ、背景バッファ、画像補完

Improvement of Free Viewpoint VoD System by Introducing Curved Dynamic Background Buffer

Akio ISHIKAWA Atsushi MATSUMURA Ryoichi KAWADA and Atsushi KOIKE

KDDI R&D Laboratories 2-1-15 Ohara, Kamifukuoka-shi, Saitama, 356-8502 Japan

E-mail: ao-ishikawa@kddilabs.jp, matsumura@kddilabs.jp, kawada@kddilabs.jp, koike@kddilabs.jp

Abstract It is important problem that the quality of the output images can be degraded because of lack of information on occluded regions in the same frame, with free viewpoint images synthesized using reference images with their depth maps. Then, previously, we proposed the method of interpolating an output image using the background buffer updated dynamically in order to solve this problem. However, in this method, there is the problem which the occluded regions on the ground which exists continuously from a view to a background a front is not interpolated well. Then, this paper proposes the new method using the background buffers, which can express various shapes of backgrounds such as curved surfaces instead of the background buffer of the plane used by the conventional method, in order to enable a highly precise interpolation. The effectiveness of the proposed method was evaluated by several experimental results.

Keyword Free Viewpoint Video, Depth Map, Background Buffer, Image Interpolation

1. はじめに

近年、映像処理/通信技術の発達に伴い、ユーザがカメラ付き携帯電話で撮影した映像をやり取りするなど、インタラクティブな映像アプリケーションを実現する技術が普及してきている。それに伴い、次世代のインタラクティブな映像処理/通信技術で扱うコンテンツの1つとして、3次元自由視点映像が注目されている。3次元自由視点映像とは、ユーザが、実際に撮影を行った際のカメラの位置から見た場合の映像のみならず、ユーザが任意に選択した視点からの映像を再構成することも可能とするオンデマンド型の映像である。

完全な3次元自由視点映像を実現するためには、ユーザが所望する視点の全てについて対応する映像データを用意しなければならないが、ユーザに任意の視点

から鑑賞することを許す自由視点映像では、この方法は非現実的である。この問題に対処するため、一般に、複数の代表的な視点に対応する映像データのみをそれぞれ用意し、それ以外の視点に対応する映像を他の映像の情報から補完もしくは推定する手法を用いる。しかしながら、この手法を用いても、3次元自由視点映像コンテンツのデータ量は通常の映像コンテンツに比べて非常に多く、現状の携帯電話ではデータの送受信に時間がかかる。また、データを取得するために多眼カメラが必要になるなど、一般ユーザには実現が困難な点がある。

そこで、筆者らは、1視点から撮影された映像（以下、参照映像）とその奥行き値（以下、奥行きマップ）を用いて、ユーザが任意の視点を選べる3次元自由視

点映像を実現する手法を提案した[1][2]. 同手法を用いると、奥行きマップを参照画像に付随する情報として取り扱うことができるため、MPEG-4 などの付随情報の伝送に対応した符号化フォーマットとの整合性がよい[3].

本検討では、筆者らがこれまで提案した手法を改良し、より効率的に3次元自由視点映像を実現する手法を提案する.

2. 従来手法

筆者らは、先に、参照映像の各フレーム毎に、そのフレームの画像（以下、参照画像）とそれに対応する奥行きマップのみから背景バッファを動的に生成/更新し、これを参照することで出力画像を補完し、高精度な自由視点映像を生成する手法を提案した.

以下で、この従来手法の詳細を述べる.

2.1. 自由視点画像の仮生成

ここでは、対象画像を O と表した際の、 O 内の座標 (x, y) における画素値を $O(x, y)$ と表すことにする.

参照画像と奥行きマップから各画素の3次元的位置が決まるので、視点の回転移動および平行移動に基づき射影変換を施すことで、任意の視点から見た画像を再構成することができる. ここで、視点の回転移動を 3×3 の行列 \mathbf{R}' で定義し、視点の平行移動を 1×3 のベクトル \mathbf{t}' として定義する. 参照画像を I 、奥行きマップを D として、 I における画素の位置を (u, v) で表し、 I における座標 (u, v) の奥行きを $D_{I(u,v)}$ で表した場合、自由視点画像 A において同一の位置を指し示す画素（以下、対応点） (u'', v'') との関係は、下記の式(1)によって表される.

$$\begin{pmatrix} D_{I(u,v)} \\ v \\ 1 \end{pmatrix} - \mathbf{t}' \times \mathbf{R}' \begin{pmatrix} u'' \\ v'' \\ 1 \end{pmatrix} = 0 \quad \dots(1)$$

A と I との間の画素の対応を求めるには、式(1)を $(u'', v'', 1)$ について解けばよい. (u'', v'') が求められるので、 A の対応点に I の対応する画素の画素値を代入することにより（下記の式(2)）、 A を仮生成することができる.

$$A(u'', v'') = I(u, v) \quad \dots(2)$$

2.2. 背景バッファの動的更新

前述の処理だけでは、参照画像において手前の物体により隠蔽されていた領域は画素値がそもそも存在しないので（図 1(a)）、視点を移動した際に、再構成された画像内に空隙が残ってしまう（図 2）.

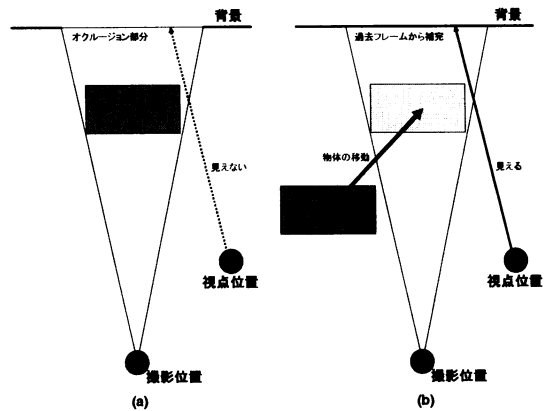


図1 隠蔽領域の補完



図2 空隙を補完しない場合

この問題を解決するため、筆者らは、背景バッファを用いて空隙を補完する手法を提案した. すなわち、過去のフレームにおける背景の画素値を背景バッファに保持しておき（図 1(b)）、この画素値を用いて現在の仮生成画像の空隙を補完する手法である（図 3）. 背景バッファに保持されている背景と現在のフレームの背景とを合成することで、背景バッファを更新し続けていれば、常に最新の画素値で補完ができることになる.

ただし、背景バッファの更新に用いられる射影変換は、奥行きを考慮に入れることができないため、分割した参照画像を用いる[4]. 参照画像を分割するために、まず I の奥行き分布の統計をとる. ここで、 n を整数とし、統計のステップ幅を S 、 I における奥行きが i 以上 j 未満である画素の個数を $v(I)$ とすると、 I の奥行き分布の統計 $V(n)$ は、下記の式(3)で表される.

$$V(n) = v(I_{(n+1)S}) \quad \dots(3)$$

$V(n)$ にガウシフィルタを施して平滑化し、極小値をとる際の奥行き n の値を求め、 nS を分割指標として値が小さい順にラベル $min_1, min_2, \dots, min_M$ をふる。この時、分割された画像 I_m ($m = 0, 1, \dots, M$) の画素値は下記の式(4)で表される。

$$I_m(x, y) = \begin{cases} I(x, y) & \text{if } min_1 \leq D_{I(x,y)} \leq min_{m+1} \\ \text{null} & \text{otherwise} \end{cases} \quad \dots(4)$$

ただし、 $min_0 = -\infty$, $min_{M+1} = \infty$ とし、null は画素が存在しないことを表す。



図3 空隙を補完した場合

背景バッファを U_m で表す。先頭フレームでは、 $m \geq 1$ の場合に I_m をそのまま U_m に投影する。その他のフレームでは、 I_m と U_m から 8 点以上の対応点を探索し、下記の式(5)を用いて、射影変換行列 B_m を算出する[5]。ここで、 I_m 内の座標 (u_i, v_i) と U_m 内の座標 (u'_i, v'_i) (u'_i, v'_i) は対応点であるとする。

$$\begin{pmatrix} u_i \\ v_i \\ 1 \end{pmatrix} \times B_m \begin{pmatrix} u'_i \\ v'_i \\ 1 \end{pmatrix} = 0 \quad \dots(5)$$

次に、求められた (u_i, v_i) と (u'_i, v'_i) の対応を下記の式(6)に代入することにより、 I_m と U_m との合成を行い、背景バッファの更新を行う。

$$U_m(u'_i, v'_i) = \begin{cases} U_m(u'_i, v'_i) & \text{if } I_m(u_i, v_i) = \text{null} \\ I_m(u_i, v_i) & \text{otherwise} \end{cases} \quad \dots(6)$$

2.3. 出力画像の生成

仮の自由視点画像の空隙部分(対応点が存在しない画素)を背景バッファに格納された画素値で補完することにより、出力画像を生成する。

具体的には、まず A と U_1, U_2, \dots, U_M から 8 点以上の対応点を探索し、下記の式(7)を用いて、射影変換行列 B'_1, B'_2, \dots, B'_M を算出する。ここで、 A における座標 (u'', v'') と U_m における座標 (u'_m, v'_m) は対応点であるとする。

$$\begin{pmatrix} u'_m \\ v'_m \\ 1 \end{pmatrix} \times B'_m \begin{pmatrix} u'' \\ v'' \\ 1 \end{pmatrix} = 0 \quad \dots(7)$$

次に、 $m = 1, 2, \dots, M$ の順に従って、 (u'', v'') と (u'_m, v'_m) の対応を、下記の式(8)に代入することにより、 A に対する補完を行った出力画像が得られる。

$$A(u'', v'') = \begin{cases} U_m(u'_m, v'_m) & \text{if } A(u'', v'') = \text{null} \\ A(u'', v'') & \text{otherwise} \end{cases} \quad \dots(8)$$

2.4. 処理の流れ

図4に、各フレーム毎の処理の流れを示す。

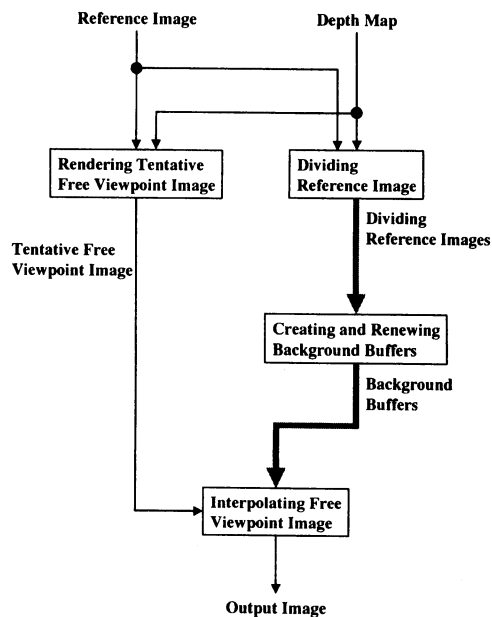


図4 処理の流れ

3. 提案手法

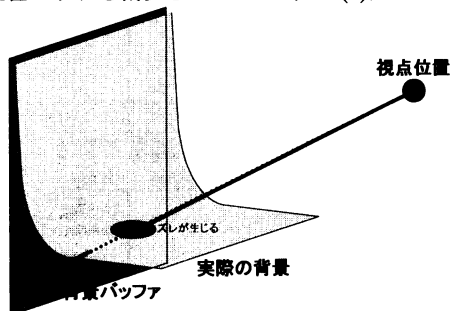
3.1. 従来手法の課題

従来手法では、背景部分に動きがある場合を考慮して、各フレーム毎に背景バッファの更新を行っていた。背景バッファを更新することにより、過去のフレームの背景が現在のフレームの背景に残像として残ってしまう事態を避けることができた。その反面、背景バッファに格納される背景画像は、奥行きを考慮に入れることができず、常に平面に近似される必要があった。

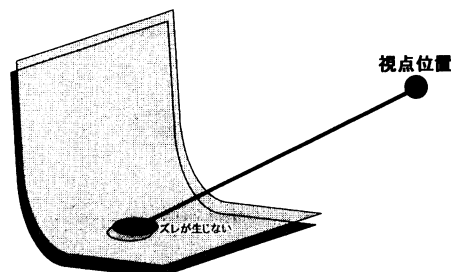
そのため、例えば、前景から背景まで連続的に奥行きが変化する地面や床などの上の隠蔽領域では、背景バッファに格納された背景画像の形状が平面であるため、背景バッファに格納された背景と実際の背景との奥行きの違いによって、両者の位置にずれが生じてしまい（図 5(a)）、仮の自由視点画像の隠蔽領域が正常に補完されず、画像によっては出力画像の中の相当する領域が不自然になってしまう。

3.2. 曲面背景バッファ

前節で述べた課題を解決するため、提案手法では、背景に動きが少ない映像に限り、背景バッファに格納された背景の形状を奥行きマップに基づく実際の形状と一致させることにより、背景バッファ（以下、提案手法の背景バッファを「曲面背景バッファ」と呼び、従来手法の背景バッファを「平面背景バッファ」と呼んで区別する）に格納された背景と実際の背景との間の位置のずれを減少させている（図 5(b)）。



(a) 背景バッファの形状が平面である場合



(b) 実際の背景と背景バッファの形状が一致した場合

図5 曲面背景バッファ

平面背景バッファでは背景の画素値のみを保持していたが、曲面背景バッファは奥行きマップと同じ形状を保つため、奥行き値も保持しておく必要がある。そこで、曲面背景バッファには、画素値に加えて各画素に対する奥行き値も格納する。具体的には、式(5)を用いて求めた (u_m, v_m) と (u'_m, v'_m) の対応を奥行きマップにも適用する。厳密には、背景に動きがある場合や視点が運動している場合は、 U_m の座標 (u'_m, v'_m) における奥行き値は I_m の座標 (u_m, v_m) における奥行き値と一致しないが、本検討では、背景の動きや視点の運動が十分に小さい場合を想定している。

ここで、 I_m の座標 (u_m, v_m) における奥行き値を $D_I(u_m, v_m)$ 、 U_m の座標 (u'_m, v'_m) における奥行き値を $D_U(u'_m, v'_m)$ とおく。 (u_m, v_m) と (u'_m, v'_m) の対応を下記の式(9)に代入することにより、 D_m と D_U との合成を行い、曲面背景バッファの奥行き値の更新を行う。

$$D_U(u'_m, v'_m) = \left. \begin{cases} D_I(u_m, v_m) & \text{if } \min_m \leq D_I(u_m, v_m) \leq \min_{m+1} \\ D_U(u'_m, v'_m) & \text{elseif } D_I(u_m, v_m) = \text{null} \\ \min_m & \text{elseif } D_I(u_m, v_m) \leq \min_m \\ \min_{m+1} & \text{elseif } \min_{m+1} \leq D_I(u_m, v_m) \\ \text{median}(D_I(u_m, v_m)) & \text{otherwise} \end{cases} \right\} \quad \dots(9)$$

一般に、背景バッファ I_m の奥行き値は、式(4)から明らかな通り \min_m から \min_{m+1} までの間に存在する。すなわち、 $\min_m \leq D_I(u_m, v_m) \leq \min_{m+1}$ が成立する（式(4)第1式）

ので、この範囲外の奥行き値が得られてしまった場合は \min_m もしくは \min_{m+1} を指定しておく。

曲面背景バッファでは、各画素の3次元的位置が決まっているので、Aと同様に、視点の回転移動および平行移動に基づき射影変換を施すことで、任意の視点から見た背景を再構成することができる。具体的には、Aにおける対応点を (u'_A, v'_A) とおくと、 (u'_m, v'_m) と (u'_A, v'_A) の関係は、下記の式(10)によって表すことができる。

$$\begin{pmatrix} D_{U_m} \\ u_{i_m} \\ v_{i_m} \\ 1 \end{pmatrix} \begin{pmatrix} u'_m \\ v'_m \\ 1 \end{pmatrix} - \mathbf{t}' \times \mathbf{R}' \begin{pmatrix} u'_A \\ v'_A \\ 1 \end{pmatrix} = 0 \quad \dots(10)$$

式(10)を (u'_A, v'_A) について解くことによって、 A と U_m との間の画素の対応が求まるので、 A の対応点に U_m の対応する画素値を代入することにより、 A の背景を再構成することができる。

最後に、 $m=1,2,\dots,M$ の順に従って、 (u'_A, v'_A) と (u'_m, v'_m) の対応を、下記の式(11)に代入することにより、 A に対する補完を行った出力画像が得られる。

$$A(u'_A, v'_A) = \begin{cases} U_m(u'_m, v'_m) & \text{if } A(u'_A, v'_A) = \text{null} \\ A(u'_A, v'_A) & \text{otherwise} \end{cases} \quad \dots(11)$$

4. シミュレーション実験

曲面背景バッファに関する以上の検討を踏まえて、提案手法の有効性を評価するために、シミュレーション実験を行った。以下に、シミュレーション実験に用いたコンテンツとその結果の例を示す。

本実験で使用した、参照画像と奥行きマップの例を、それぞれ図6と図7に示す。



図6 参照画像



図7 奥行きマップ

図8に、従来手法で用いる平面背景バッファの例を示す。参照画像の本来のカメラ位置から右に 30° 回転した位置に視点があると想定した場合である。

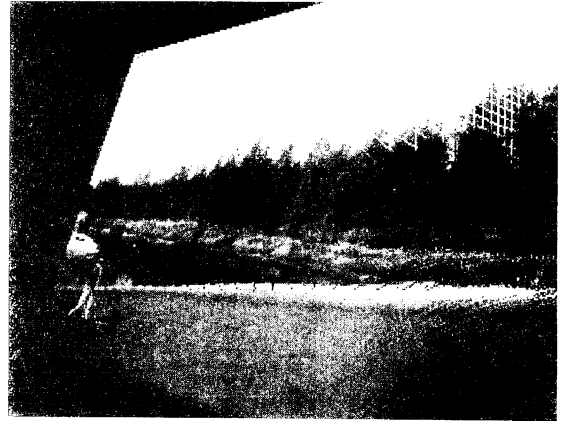


図8 平面背景バッファの例

図9に、同じ位置に視点を想定した場合の曲面背景バッファの例を示す。表示部分の輪郭から、実際の奥行きに近い形状であることが判る。

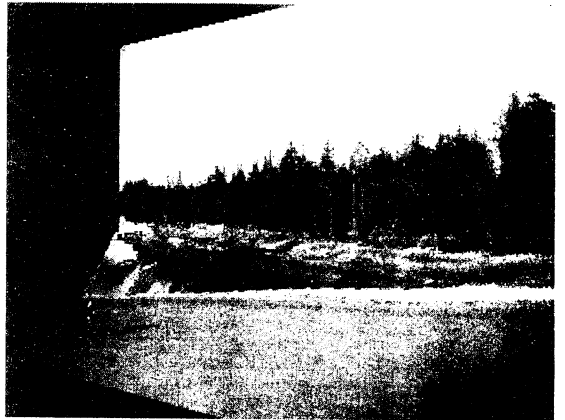


図9 曲面背景バッファの例

背景に動きがある場合は、多層背景バッファ[6]を用いて、背景の中の動きのある部分を別の層の背景バッファに格納することで対処できる。

実験では、ゴルフ場でゴルフを楽しむプレイヤーの映像を用いた。画像サイズは水平方向が320pixel、垂直方向が240pixelで、全体で624フレームのシーケンスである。各フレーム毎に奥行きマップが求められており、256階調（奥が0、手前が255）で奥行きが表現されている。簡単のため、画像の分割は前景と背景の2つのみとし、 $min_1=195$ で分割した。

図10に、補完を行わない場合の出力画像例(a)と、従来の平面背景バッファを用いて補完を行った場合の出

力画像例(b)と、曲面背景バッファを用いた補完を行った場合の出力画像例(c)を示す。比較のため、実際の出力結果から、背景バッファによる補完が最も多く行われている、プレイヤー周辺の領域を切り出し、拡大している。いずれの画像も、視点位置は右に15°、上に15°回転移動させてある。

補完を行わない場合は、図10(a)に示すように、隠蔽領域(参照画像では、人物の背後の背景であった領域)が空隙となる。ここでは、黒く塗りつぶして示している。

従来の手法で補完を行った場合は、図10(b)に示すように、空隙は補完されて無くなるものの、本来の背景と補完された背景との間にズレ(池の縁など)が生じてしまう。

提案手法で補完を行った場合を、図10(c)に示す。本来の背景と補完された背景との間のズレはほぼ無くなり、一目では補完された範囲が分からない程度に品質が改善されている。



(a) 補完なし



(b) 平面背景バッファ



(c) 曲面背景バッファ

図10 シミュレーション結果

5. おわりに

本検討では、参照映像と奥行きマップのみを用いて3次元自由視点映像を生成する手法において、従来手法で用いていた平面背景バッファの代わりに、奥行きマップと形状を一致させた曲面背景バッファを用いることにより、再構成映像の品質を改善する手法を提案

した。曲面背景バッファを用いることにより、従来手法で課題となっていた、地面上など奥行きが連続的に変化する部分の隠蔽領域の画質を大幅に改善することができる。

今後の課題としては、多層背景バッファの曲面化など再構成映像の更なる品質改善や、処理の高速化手法の検討が挙げられる。また、今回の検討では、再構成画像の品質の改善度合いは定性的な評価しか行わなかったが、今後は、定量的/客観的な品質評価の基準とその方法を検討する必要がある。

謝 辞

日頃より熱心なご指導・ご鞭撻をいただき KDDI 研究所の浅見所長に深く感謝の意を表する。

文 献

- [1] 松村, 内藤, 川田, 小池, 松本, "任意視点動画画像の高圧縮伝送を目的とした隠蔽領域補完方式の提案," 信学技報, Vol.IE2003-66, pp.63-68, 2003.
- [2] 松村, 内藤, 川田, 小池, 松本, "複数枚の背景バッファを用いた自由視点動画画像に対する高精度な補完方式," 映像メ冬季大会, No.8-8, 2003.
- [3] C. Fehn, K. Schuur, P. Kauff and A. Smolic, "Meta-Data Requirements for EE4 in MPEG 3DAV," ISO/IEC JTC1/SC29/WG11, MPEG02/M9559, Pattaya, Thailand, Mar. 2003.
- [4] R. Szeliski and P. Golland, "Stereo Matching with Transparency and Matting," in Sixth International Conference on Computer Vision, Bombay, India, pp.517-524, 1998.
- [5] Richard Szeliski, "Video Mosaics for Virtual Environments", IEEE Computer Graphics and Applications, 16, 2, pp.22-30, Mar. 1996.
- [6] 松村, 内藤, 川田, 小池, 松本, "動的に更新される多層背景バッファを用いた高精度な自由視点映像生成方式," 映像情報メディア学会誌, Vol.58, No.6, pp.850-853, 2004.