

研究速報

移動カメラアレーを用いた時空間の自由視点画像生成

横井 孝紀<sup>†</sup> 福嶋 慶繁<sup>††</sup> (正員)

圓道 知博<sup>†</sup> (正員)

メヒルダド パナヒブル テヘラニ<sup>†</sup> (正員)

藤井 俊彰<sup>†††</sup> (正員) 谷本 正幸<sup>†</sup> (正員：フェロー)

Spatio-Temporal Free-Viewpoint Image Generation Using Moving Camera Array

Takanori YOKOI<sup>†</sup>, Nonmember, Norishige FUKUSHIMA<sup>††</sup>, Tomohiro YENDO<sup>†</sup>, Mehrdad PANAHPOUR TEHRANI<sup>†</sup>, Toshiaki FUJII<sup>†††</sup>, Members, and Masayuki TANIMOTO<sup>†</sup>, Fellow

<sup>†</sup>名古屋大学大学院工学研究科, 名古屋市  
Graduate School of Engineering, Nagoya University, Furouchi, Chikusa-ku, Nagoya-shi, 464-8603 Japan

<sup>††</sup>名古屋工業大学大学院工学研究科, 名古屋市  
Graduate School of Engineering, Nagoya Institute of Technology, Gokiso-cho, Showa-ku, Nagoya-shi, 466-8555 Japan

<sup>†††</sup>東京工業大学大学院理工学研究科, 東京都  
Graduate School of Science and Engineering, Tokyo Institute of Technology, Ookayama, Meguro-ku, Tokyo, 152-8550 Japan

あらまし 本研究では、従来の固定されたマルチカメラアレーに代わり、被写体が広範囲に渡って動くシーンに対応できる移動カメラアレーを提案する。更に移動しながら撮影をすることにより密に光線を取得できることを生かした自由視点画像生成手法を提案する。

キーワード Free-viewpoint TV (FTV), 自由視点画像生成, 移動カメラアレー, 光線空間

1. まえがき

視聴者が自由に視点を選択することができる自由視点テレビ (FTV) が提案されている [1]。この自由視点テレビを実現するために、Image Based Rendering (IBR) の一種である光線空間法を用いて自由視点画像を生成する研究が行われている [2]。従来の自由視点画像の取得系として、マルチカメラアレーが多く用いられてきた。マルチカメラアレーの例としては、直線上にカメラを配置した直線カメラアレー、円形にカメラを配置した円形カメラアレーなどが挙げられる。これら複数台のカメラを用いるカメラアレーは、固定されているため、撮影を行う際に被写体の動ける範囲を制限してしまう。そこで本論文では、直線カメラアレーを動かす移動カメラアレーを取得系として用いることの提案を目的とする。提案手法では、移動することにより様々な場所で物体が広範囲に移動するシーンを撮影することができる。更に、光線を密に取得できると

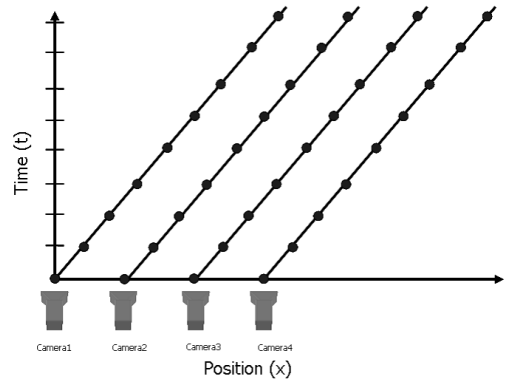


図1 移動カメラアレー  
Fig. 1 Moving camera array.

いう特長を生かし、生成する自由視点画像の画質向上を目指す。

2. 移動カメラアレー

縦軸を時間、横軸を位置とするグラフ上で、移動カメラアレーが一方向に移動しながら撮影するときの関係は図1のように表すことができる。図中の丸はカメラが撮影するときの時間・位置を表しており、一つの丸が一枚の画像に相当する。移動カメラアレーは移動しながら撮影を行うことで、非常に密な撮影を行うことができるのが分かる。これは光線空間で考えると、撮影対象のうち、動く物体の光線は時々刻々と変化するため、密に取得することはできないが、静止している物体の光線は一定のため、移動カメラアレーを使うことで、固定カメラアレーより密に取得可能ということを表している。

3. 移動カメラアレーを用いた自由視点画像生成

前章で説明した特徴を活かした自由視点画像生成手法を提案する。図2は移動カメラアレーが撮影する際の位置、時刻を表している。図2中の三角は左カメラ  $L(t-j), L(t), L(t+i)$ , 四角は右カメラ  $R(t-j), R(t), R(t+i)$ , 丸は仮想カメラ  $V(t)$  をそれぞれ表している。従来の画像生成では仮想視点  $V(t)$  を生成する際には、時刻の同じ、左画像  $L(t)$  と右画像  $R(t)$  のみを用いてきた。本手法では、移動カメラを用いて密に光線を取得できることを生かすため、従来使用されていた2枚の画像に加えて、時刻の異なる、より生成位置に近い画像  $L(t+i)$  と  $R(t-j)$  を画像生成に用いる。つまり、画像生成に計4枚の画像を使用する。提案手法の概略は以下のとおりである。仮想視点の静止した部分  $V_S(t)$  を、時刻が異なる  $L(t+i)$

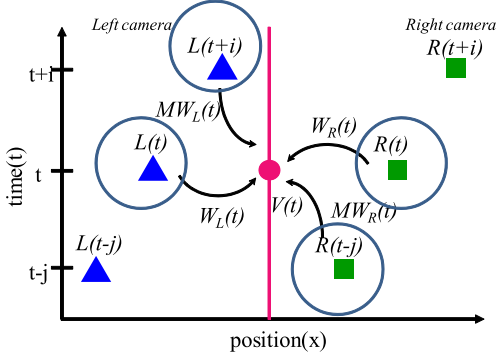


図 2 画像生成に使用する画像  
Fig. 2 Images used in view synthesis.

と  $R(t-j)$  を、オプティカルフローにより仮想視点にワープさせた画像  $MW_L(t)$ ,  $MW_R(t)$  から作る．動いている部分  $V_D(t)$  は、同時刻の左右の画像  $L(t)$ ,  $R(t)$  を奥行画像によりワープさせた画像  $W_L(t)$ ,  $W_R(t)$  から生成し、 $V_S(t)$  と  $V_D(t)$  をマージすることで仮想視点の画像  $V(t)$  を生成する．詳細は以下で説明する．

まず、静止した部分の画像  $V_S(t)$  を生成するため、左視点の時刻  $t+i$  の画像が動いているか否かを示すフラグ  $SD_{L_{t+i}}$  を求める．そのために、左の画像の組合せから二つの密なオプティカルフローを求める．そのオプティカルフローは、時刻  $t+i$  の左画像から時刻  $t$  の左画像へのオプティカルフロー  $M_L^t(t+i)$  と、時刻  $t$  の左画像から時刻  $t-j$  の左画像へのオプティカルフロー  $M_L^{t-j}(t)$  であり全画素で動きベクトルが計算されている．次に、このオプティカルフローから、静・動の判定を行う．本論文では、全てのオプティカルフローは Werlberger らの手法 [3] により求めている．判定の様子を図 3 に示す．オプティカルフロー  $M_L^t(t+i)$  の画像中の画素位置  $\mathbf{p} = (x, y)^T$  のベクトルを、

$$\mathbf{m} = (u, v)^T = M_L^t(t+i)|_{\mathbf{p}} \quad (1)$$

としたとき、その画素が示す時刻  $t$  の動きベクトルは

$$\mathbf{n} = M_L^{t-j}(t)|_{\mathbf{p}+\mathbf{m}} \quad (2)$$

で表される．本論文では、カメラは低速でかつ等速直線運動をしていると仮定する．すると、動きベクトルがフレーム間で等速ならば、カメラの運動だけを表した動きベクトルとなるため、静止した物体である可能性が高い．そこで、ベクトル  $\mathbf{n}$  とベクトル  $\mathbf{m}$  の長

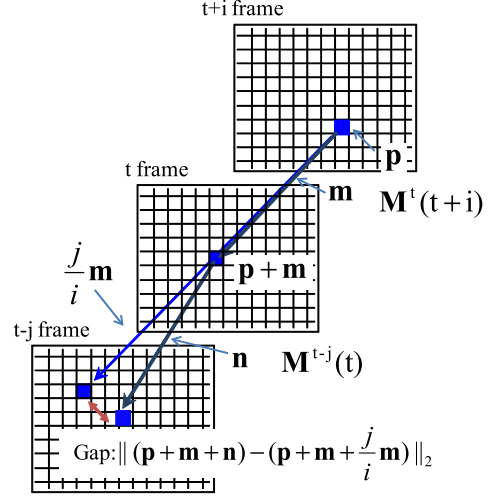


図 3 物体の静・動判定  
Fig. 3 Static or dynamic region detection.

さを  $\mathbf{n}$  に合わせて正規化したものと比較する次式で、この静・動を判定する．

$$SD_{L_{t+i}}|_{\mathbf{p}} = \begin{cases} \text{static} & \|(j/i)\mathbf{m} - \mathbf{n}\|_2 < T \\ \text{dynamic} & \text{else} \end{cases} \quad (3)$$

ここで、 $\|\cdot\|_2$  はベクトルの L2 ノルムを表し、 $T$  はしきい値を表す．しきい値は経験的に 4.0 と設定した．この処理を全ての画素  $\mathbf{p}$  について行い、また右画像の系列も同様に行い、時刻  $t-j$  の右画像のフラグ  $SD_{R_{t-j}}$  を求める．

次に左右の時刻の異なる画像間で、オプティカルフローを求める． $L(t+i)$  から  $R(t-j)$  へのオプティカルフローを  $M_{LR}^{t-j}(t+i)$  とし、その逆を  $M_{RL}^{t+i}(t-j)$  とする．もし画素  $\mathbf{p}$  のフラグ  $SD_{L_{t+i}}|_{\mathbf{p}}$  が static なら、画素位置  $\mathbf{p}$  の動きベクトル  $\mathbf{q}$  は、

$$\mathbf{q} = M_{LR}^{t-j}(t+i)|_{\mathbf{p}} \quad (4)$$

と表される．次に、このベクトルを  $L(t+i)$  から  $V(t)$  までの動きベクトルへと変換するために、等速直線運動を仮定し、動きベクトルの長さを調節する重み  $w$  を

$$w = \frac{i}{i+j} \quad (5)$$

とすれば、仮想視点への動きベクトルは、 $w\mathbf{q}$  で表される．つまり、 $L(t+i)$  の画素位置  $\mathbf{p}$  から仮想視点  $V(t)$

へ、動きベクトルに従って画素の移動を行うと、移動後の画像  $MW_L(t)$  の  $\mathbf{p} + w\mathbf{q}$  の位置に  $L(t+i)|_{\mathbf{p}}$  の画素が入る。

$$MW_L(t)|_{\mathbf{p}+w\mathbf{q}} = L(t+i)|_{\mathbf{p}} \quad (6)$$

$R(t-j)$  も同様に、もし画素  $\mathbf{p}$  のフラグ  $SD_{R(t-j)}|_{\mathbf{p}}$  が static なら、下記の式に従った処理を行えばよい。

$$\mathbf{r} = \mathbf{M}_{RL}^{t+i}(t-j)|_{\mathbf{p}} \quad (7)$$

$$\bar{w} = \frac{j}{i+j} \quad (8)$$

$$MW_R(t)|_{\mathbf{p}+\bar{w}\mathbf{r}} = R(t-j)|_{\mathbf{p}} \quad (9)$$

次に、 $MW_L(t)$ ,  $MW_R(t)$  の全ての画素  $\mathbf{p}$  において下記の演算を行い、静止している部分の仮想視点上の画像  $V_S(t)$  を生成する。

$$V_S(t) = \begin{cases} \bar{w}MW_L(t) + wMW_R(t) & \text{case LR} \\ MW_L(t) & \text{case L} \\ MW_R(t) & \text{case R} \\ \text{indefinite} & \text{else} \end{cases} \quad (10)$$

case LR のときは、左右の画素が同じ画像座標に投影された場合である。この場合は、補間する視点の時間に近い方を優遇してブレンドする。また case L, case R はそれぞれ、左、右しか同じ画像座標に投影されない場合であり、この場合は、投影された画素を優先する。これらに当てはまらないとき、不定 (indefinite) とする。

次に、動いている部分の画像  $V_D(t)$  は、Mori らの手法 [5] の手法を用いて合成する。簡単に記述するとこの手法は、まず、左右の視点の奥行画像  $D_L(t)$   $D_R(t)$  と画像  $L(t)$ ,  $R(t)$  を目的の視点  $V(t)$  にワーピングさせる。

$$W_L(t) = \text{warp}_L(L(t), D_L(t)) \quad (11)$$

$$W_R(t) = \text{warp}_R(R(t), D_R(t)) \quad (12)$$

そのワーピングさせた画像を  $W_L(t)$ ,  $W_R(t)$  とする。次に、そのワーピングさせた画像の小さなホールを埋めながら左右の画像を仮想視点からの距離の重みに応じてブレンドする。

$$V_D(t) = \text{blend}(W_L(t), W_R(t)) \quad (13)$$

この手法の場合、全ての画素で値が決定することが保証できない。そのため文献 [5] では、不定部分に対して周囲の画素値から補間する手法であるインペイント [6] を用いていた。しかし、提案手法では、インペイントは行わず、不定部分はそのままとする。なお、本論文では事前に、グラフィカット [7] により奥行画像を求めた。

$V_D(t)$  は、不定部分を除けば、動領域を含む全ての画素が決定されている。しかし、次の画像合成の演算により、静止している部分  $V_S(t)$  により上書きされる。

$$V'(t)|_{\mathbf{p}} = \begin{cases} V_D(t)|_{\mathbf{p}} & V_S(t)|_{\mathbf{p}} = \text{indefinite} \\ V_S(t)|_{\mathbf{p}} & \text{else} \end{cases} \quad (14)$$

最後に、 $V'(t)$  の不定部分をインペイントし、全ての画素を決定した画像が  $V(t)$  となる。

#### 4. 実験

移動カメラアレーのシーケンス *kendo* [4] を使い、提案手法と左右画像のみを使用する従来手法の比較を行った。実験では、図 4 のカメラアレーを動かしながら撮影した。使用したカメラはカメラアレーの中心に位置する 3 台である。両端 2 台を画像合成の左右の参照カメラとして用い、中心のカメラ位置に相当する視点の画像を合成した。その中心の 1 台は、ground truth として PSNR を計算するためだけに用い、画像合成には使用しない。また、各カメラの位置は Structure from Motion [8] により求めることで、合成する仮想視点までの距離が近い、時系列の映像を選択した。従来手法としては、文献 [5] の奥行画像のみを用いる手法 (式 (11)~(13) に相当) を用いた。こちらの場合は最後に不定部分をインペイントしている。

提案手法の実験結果を図 5 に、画像中右の女性を拡

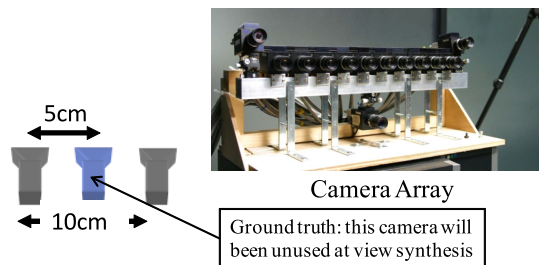


図 4 カメラアレー  
Fig. 4 Camera array.



図5 提案手法により生成した画像  
Fig. 5 Synthesis view of proposed method.



(a) 従来手法

(b) 提案手法



(c) ground truth

図6 静止している領域を拡大  
Fig. 6 Enlarging static objects.

大したものを図6にそれぞれ示す。図6より、女性の周りで見られるアーチファクトを低減することができた。また、客観評価にはPSNRを用い、従来手法

33.23 dB, 提案手法 34.19 dB となり、約 1.0 dB の向上が確認できた。

## 5. むすび

本論文では、広い空間での撮影など、従来用いられてきた固定カメラアレーが不向きな環境での代替手段として、移動カメラアレーを使用することと、その特徴を活かした画像生成手法を提案した。本方式が有効に働くのは、カメラアレーの移動が等速直線運動で仮定できることや、移動速度が、カメラ間の光線を取得可能な程度に緩やかである場合である。また、実験により画質が向上したことを確認し、その有効性を示した。

今後の課題としては、動いている領域・静止している領域を分けるしきい値決定の自動化が考えられる。

謝辞 本研究の一部は、総務省戦略的情報通信研究開発推進制度(SCOPE)及び、科学研究費補助金若手研究(B)(課題番号22700174)によって行われた。

## 文 献

- [1] M. Tanimoto, "Free-viewpoint TV," The Forum for Advancement of Three Dimensional Image Technology and Arts, vol.13, no.3, pp.17-22, 2001.
- [2] T. Fujii, T. Kimoto, and M. Tanimoto, "Ray space coding for 3D visual communication," Proc. PCS'96, vol.2, pp.447-451, March 1996.
- [3] M. Werlberger, T. Pock, and H. Bischof, "Motion estimation with non-local total variation regularization," Proc. IEEE Computer Vision and Pattern Recognition, June 2010.
- [4] <http://www.tanimoto.nuee.nagoya-u.ac.jp/MPEG-FTVProject.html>
- [5] Y. Mori, N. Fukushima, T. Yendo, T. Fujii, and M. Tanimoto, "View generation with 3D warping using depth information for FTV," Image Communication, vol.24, no.1-2, pp.65-72, Jan. 2009.
- [6] A. Telea, "An image inpainting technique based on the fast marching method," J. Graphics Tools, vol.9, no.11, pp.23-34, 2004.
- [7] Y. Boykov, O. Veksler, and R. Zabih, "Fast approximate energy minimization via graph cuts," IEEE Trans. Pattern Anal. Mach. Intell., vol.23, no.11, pp.1222-1239, Nov. 2001.
- [8] N. Snavely, S. Seitz, and R. Szeliski, "Photo tourism: Exploring photo collections in 3D," ACM Trans. Graphics, vol.25, no.3, pp.835-846, July 2006.

(平成23年2月28日受付, 7月6日再受付)