

単眼カメラとARマーカによる低演算自由視点映像生成

Low-Complexity Free Viewpoint Video Generation with a Single Camera and a Fiducial Marker

中野 学[†] 田治米 純二[†] 仙田 裕三[†]

Gaku NAKANO Junji TAJIME Yuzo SENDA

1. はじめに

自由視点映像とは、ユーザが自由に視点を選択出来る映像のことである。自由視点映像を用いることで映像表現の幅が広がり、コミュニケーションがより活発化すると考えられる。自由視点映像の表現方法は2Dベースと3Dベースの2つに分けられる。3Dベース手法として、視体積交差法によりボリュームを復元する手法[1]や Structure from Motion を利用して点群の3次元復元を行う手法[2]が提案されている。2Dベース手法として、デプスマップや光線空間からカメラ間の中間画像を生成する手法[3][4]が提案されている。これらの手法は、いずれも撮影に多数のカメラやカメラ以外の様々な機材が必要であったり、演算量が膨大であったり、最終的に得られる画質が低いといった問題がある。そのため、実用レベルでは、回転台に被写体を載せて固定カメラで全周画像を撮影し、ユーザの操作に応じて撮影画像を切り替える2Dベースの手法[5]（以下回転台手法）が利用されている。その理由は、撮影画像をそのまま表示するので、他の手法に比べて高画質になるためである。しかし、回転台の利用は撮影環境が制限されるため、ユーザにとって大きな制約となる。このように、(1)高画質(2)撮影環境の制約が少ない(3)低演算量をすべて満たした手法は実現されていない。そこで、本稿ではこれら3つの問題点を克服した自由視点映像生成手法を提案する。

2. 手持ちカメラによる撮影

回転台手法の課題は、回転台やカメラを固定する機材を不要にすることである。手持ちカメラを正確に操作して、被写体の全周映像を撮影できれば、理論的には回転台手法と同じ映像が生成できる。しかし、これほどの正確な操作は、非常に困難である。図1に回転台手法と手持ちカメラによる撮影画像の違いを示す。回転台手法は、回転軸を中心として安定した撮影が可能である。一方、手持ちカメラではフレーム毎に被写体の位置、大きさ、傾きが異なってしまう。そのため、得られた映像は、そのままでは自由視点映像にはならない。

3. 提案手法

手持ちの単眼カメラとARマーカを用いた手法を提案する。提案手法の流れ図を図2に示す。まず、3次元空間中の仮想的な回転軸（以下仮想軸）と仮想軸の画像上の変換先となる軸（以下基準軸）を設定する。ARマーカでフレーム毎のカメラ位置姿勢を検出し、仮想軸を画像へ射影する。射影された仮想軸と基準軸が一致するように相似変換を施す。この処理を全フレームで逐次的に実行すれば、仮

想軸を中心として、回転台を用いずにあたかも被写体を一定距離から周回撮影したかのような画像を得る。

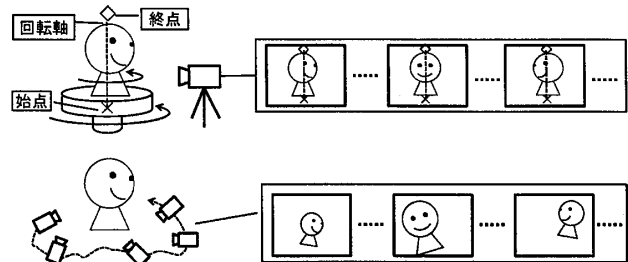


図1: 撮影方法による撮影画像の違い
(上段: 回転台手法, 下段: 手持ちカメラ)

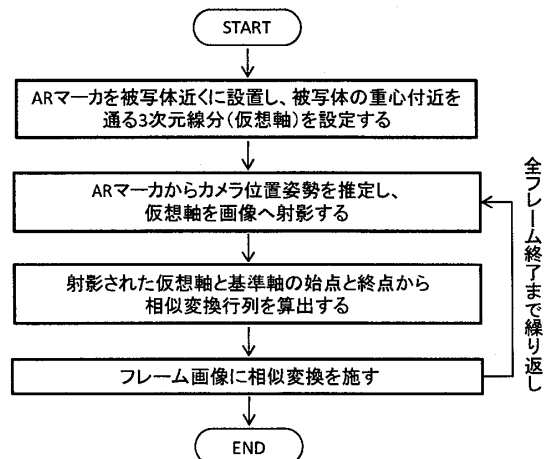


図2: 提案手法の流れ図

3.1 仮想軸の設定および画像への射影

仮想軸を設定するために、ARマーカを利用する。ARマーカを被写体付近に配置し、被写体のおおよそ重心を通るような仮想軸を設定する。仮想軸の画像への射影は式(1)で表される。

$$\begin{bmatrix} \mathbf{u}_{ij} \\ 1 \end{bmatrix} \propto \mathbf{K} [\mathbf{R}_i \ \mathbf{t}_i] \begin{bmatrix} \mathbf{X}_j \\ 1 \end{bmatrix} \quad (1)$$

ここで、添字の i はフレーム番号、 j は仮想軸の始点($j=1$)と終点($j=2$)、 $\mathbf{X}_j = [X_j, Y_j, Z_j]^T$ は仮想軸の始点および終点の3次元座標、 $\mathbf{u}_{ij} = [u_{ij}, v_{ij}]^T$ は i 番目のフレームにおける仮想軸の始点と終点の画像座標、 $[\mathbf{R}_i \ \mathbf{t}_i]$ は i 番目のフレームにおけるカメラ位置姿勢、 \mathbf{K} はカメラの内部パラメータ、 α は定数倍に等しいことを表す。 \mathbf{K} は事前にカメラキャリブレーションを行い、計測しておく。 \mathbf{X}_j をユーザ

[†]日本電気株式会社
情報・メディアプロセッシング研究所
Information and Media Processing Labs, NEC Corporation

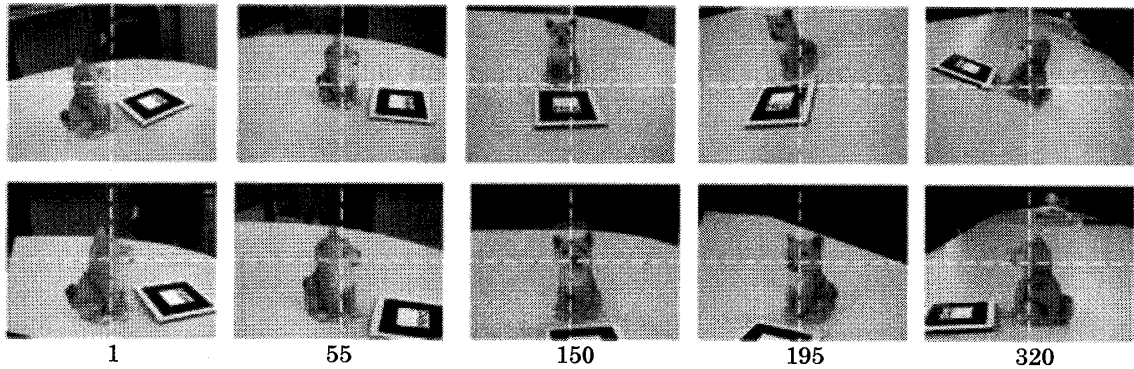


図4: 実験結果

(上段: 入力画像, 下段: 出力画像, 数字はフレーム番号)

が一度設定すれば, $[\mathbf{R}_i \ \mathbf{t}_i]$ は AR マーカより得られるため, 式(1)に従いフレーム毎の \mathbf{u}_{ij} が得られる.

3.2 相似変換

式(1)で得られた仮想軸 \mathbf{u}_{ij} を基準軸へ相似変換する. 相似変換は式(2)で表される.

$$\mathbf{u}'_j = \mathbf{M}_i \begin{bmatrix} \mathbf{u}_{ij} \\ 1 \end{bmatrix} \quad (2)$$

ここで, 添字の i, j および \mathbf{u}_{ij} は式(1)と同様で, $\mathbf{u}'_j = [u'_j, v'_j]^T$ は変換先である基準軸の始点と終点, $\mathbf{M}_i = \begin{bmatrix} a_i & b_i & c_i \\ -b_i & a_i & d_i \end{bmatrix}$ は i 番目のフレームにおける相似変換パラメータである. \mathbf{u}'_j は事前にユーザが設定する.

始点と終点の2点から4つの連立方程式を得るため, これを解いて \mathbf{M}_i が求まる. 得られた \mathbf{M}_i をフレームの全画素に適用すると, 図3に示すように仮想軸が基準軸に一致した画像を得る.

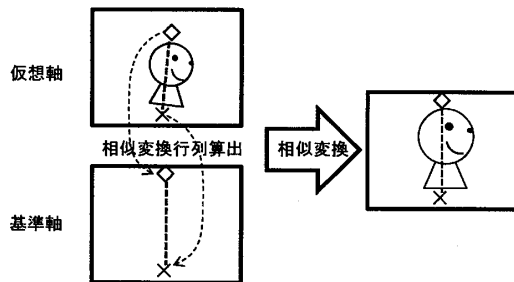


図3: 仮想軸から基準軸への相似変換

4. 実験

提案手法を実装し, 実映像に対して実験を行った. 計算機の構成は, CPU Intel Core2 Duo 3.16GHz, メモリ 2GB, OS Windows XP SP3 である. 実装には C++言語, AR ライブラリとして ARtoolkit[6]を用いた. 入力画像はサイズ 640×480 , 345 フレームである. 仮想軸の3次元座標(単位: ミリメートル)を $\mathbf{X}_1 = [0.0, 60.0, 0.0]^T$, $\mathbf{X}_2 = [0.0, 60.0, 120.0]^T$ とし, 基準軸の画像座標を $\mathbf{u}'_1 = [320, 512]^T$, $\mathbf{u}'_2 = [320, 128]^T$ とした.

実験結果を図4に示す. 上段が手持ちカメラで撮影した入力画像, 下段が出力画像である. 数字はフレーム番号,

点線は画像の中央を表す. 出力画像における被写体の大きさや傾きが全フレームで一貫しており, 一定距離から周回撮影を行ったかのような画像に変換されていることが確認できる. また, 入力画像に相似変換を適用するだけなので主観的な画質劣化がほとんど生じていない.

1フレームあたりの演算量は, AR マーカ検出から画像の相似変換が平均 5m 秒, フレーム画像の入出力等すべて含めると平均 10m 秒である. これは一般的に入手可能な計算機で高速処理可能な演算量である.

5. おわりに

手持ちの単眼カメラと AR マーカを用いて撮影制約の少ない自由視点画像生成手法を提案した. AR マーカで推定したカメラ位置姿勢を用いて, 被写体の重心を通る3次元空間中の仮想軸をフレーム画像へ射影する. 射影された仮想軸と基準軸が一致するように, フレーム画像に相似変換を施す. その結果, 撮影環境の制約が少なく, 被写体を一定距離から周回撮影を行ったような高画質な画像を低演算量で生成できる. 今後は, 事前のカメラキャリブレーションや AR マーカ等の制約を緩和することが課題である.

参考文献

- [1] J. Starck and A. Hilton, "Surface Capture for Performance-Based Animation," IEEE Computer Graphics and Applications, Vol.27, No.3, pp.21-31 (2007).
- [2] S. Agarwal, N. Snavely, I. Simon, S. M. Seitz, and R. Szeliski, "Building a Rome in a Day," ICCV2009, pp.72-79 (2009).
- [3] 中西敦士, 藤井俊彰, 木本伊彦, 谷本正幸, "EPI 上の対応点軌跡を用いた適応フィルタによる光線空間データ補間", 映像情報メディア学会誌, Vol. 56, No. 8, pp.1321-1327 (2002).
- [4] C. Zitnick, S.B. Kang, M. Uyttendaele, S. Winder, and R. Szeliski, "High-quality video view interpolation using a layered representation," ACM Transactions on Graphics, Vol23, No.3, pp600-608 (2004).
- [5] 有限会社カフェデジタル <http://www.cafedigital.jp/>
- [6] H. Kato and M. Billinghurst, "Marker Tracking and HMD Calibration for a video-based Augmented Reality Conferencing System," Proc. IWAR '99, pp.85-94 (1999).