

3次元形状復元と人物追跡を用いた 単眼映像コンテンツのカメラワーク推定に関する検討

坂井 甚太[†] 宋戸 英彦[‡] 北原 格[‡]

筑波大学理工学群工学システム学類[†] 筑波大学計算科学研究センター[‡]

1. はじめに

シーンに応じて映画監督が考案するカメラワークは、映像作品の個性を生み出す要因の一つである。カメラワークとは、映画やテレビで被写体を撮影する際のカメラの動かし方のことであり、その考案は効果的な場面表現を行う上で欠かせない技能である。一方で映像作品制作におけるカメラワークの考案は、映画監督などの製作者が自身の知識や経験に基づいて行うため、その考案プロセス（撮影術）の保存・伝承が課題となっている。我々は、コンピュータビジョン技術を用いて映像作品のカメラワークを推定することで、映画制作の撮影術の獲得を目的とする研究に取り組んでいる。

本稿では、映像作品に含まれるシーンの3次元情報から映像を撮影したカメラワークを推定する方法について検討する。具体的には、Structure from Motion (SfM) を用いて撮影シーンとカメラの位置情報を取得し、3次元骨格推定を用いて被写体の位置姿勢を推定する。

2. 映像作品の撮影シーンの3次元情報取得

映像作品（単眼映像）からコンピュータビジョン技術によって取得できる情報としては、SfM [1] による3次元点群、Semantic Segmentation [2] による領域の属性、骨格推定 [3] による被写体（人物）のポーズが考えられる。これらの情報から各フレームにおける被写体とカメラの位置姿勢情報を求め、カメラワークを推定する。なお本手法では、人物サイズとカメラの内部パラメータは事前に取得可能な状況を対象とする。

カメラワークは大きく以下4種類に分類される。

- (1)固定撮影：カメラを固定したまま撮影する技法
- (2)パン・チルト：カメラを水平方向または垂直方向に回転させる撮影技法
- (3)ズーム：カメラの焦点距離の変化により被写体の見た目を拡大/縮小する撮影技法
- (4)移動撮影：シーン中を動く被写体に合わせてカメラを移動させながら撮影する技法

(4)で撮影した映像には運動視差が生じるが、(1)(2)(3)は運動視差がほとんど生じないため SfM などステレオ視に基づく3次元形状復元処理の適用が困難である。本稿では (4)で撮影された映像

に対象を絞り、コンピュータビジョン技術による被写体とカメラの位置姿勢情報の取得法について述べる。

3. カメラワークの推定法

図1に示す処理の流れに沿って提案手法について説明する。

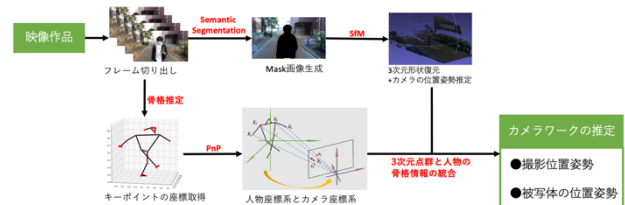


図1 提案手法の処理の流れ

3.1 動的な物体を含む単眼映像からの3次元点群生成

移動撮影で得られた5-10秒程度の動画を1フレーム毎に分割した画像群にSfMを適用し、撮影シーンの3次元情報（3次元点群）と撮影カメラの位置姿勢を推定する。本稿ではこの時の座標系を世界座標系とする。本手法では、Instance Segmentation というタスクを解くことにより画素単位で物体識別するMask R-CNN [4]を画像群に適用し人物領域を切り出す。人物領域を対応点探索領域から除外することで3次元復元精度を向上させる。

3.2 被写体の3次元骨格情報取得と撮影カメラとの相対的位置関係の推定

深層学習により人体の33か所のキーポイントを推測するBlazePose [5]を用いて、フレーム毎に動的な被写体（人物）の3次元骨格を推定する。2次元画像情報から3次元骨格の実スケールを推定するのは困難であるため、本手法では、事前に得たサイズ情報を用いてBlazePoseが出力する人物座標系をスケーリング変換することで、世界座標系と同スケールの3次元骨格を得る。

n 点の3次元座標とその画像上での観測座標からなるペアおよび内部カメラパラメータからPnP問題 [6]を解くことで、その座標系におけるカメラの位置姿勢を推定する。本手法では、前段で推

定した人物の骨格座標 33 点を用いて人物座標系におけるカメラの位置姿勢を推定する。

3.3 3次元点群と人物の骨格情報の統合

図2に示すように、前節で取得した人物座標系におけるカメラの位置姿勢と SfM で求めた世界座標系におけるカメラの位置姿勢の対応関係から、カメラ座標系を介して人物座標系と世界座標系の間の剛体変換を求め、被写体の骨格座標を世界座標系に変換する。

世界座標系とカメラ座標系の剛体変換を T_{wc} 、人物座標系とカメラ座標系の剛体変換を T_{hc} とする。人物座標系での人物の座標を X_h 、世界座標系での人物の座標を X_w とすると、 $T_{wc}X_w = T_{hc}X_h$ の式が成り立つ。この式を変形すると $X_w = T_{wc}^{-1}T_{hc}X_h$ となり、世界座標系における被写体（人物）の位置姿勢が求まる。

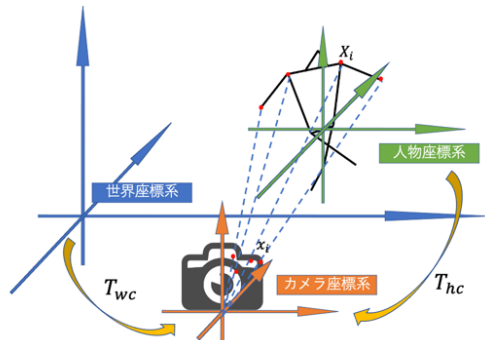


図2 各座標系の関係

4. 評価実験

カメラと人物の位置姿勢推定の精度評価実験の流れを図3に示す。正解データとして、撮影映像の各フレームにおける人物とカメラの位置を手動で計測した情報を用いる。マーカ位置に人物が立った状態でコマ撮り撮影し、本手法によって推定されたカメラと人物の位置姿勢と正解データの絶対誤差を算出する。

4.1 カメラワークの種類による人物位置姿勢の推定精度

移動撮影は以下の4種類に分類される。各カメラワークを適用した映像から人物の位置姿勢を推定し、その精度を求める。

- (a) ドリーイン：カメラが被写体に近づく
- (b) ドリーアウト：カメラが被写体から遠のく
- (c) トラック：動く被写体に対してカメラ自体を追従させる
- (d) アーク：カメラが被写体を中心に回り込む

4.2 ポーズによる人物位置姿勢の推定精度

被写体のポーズ（直立、両手上げ、歩行、着席の4種類）が人物の位置姿勢の推定精度に与える影響について調査する。

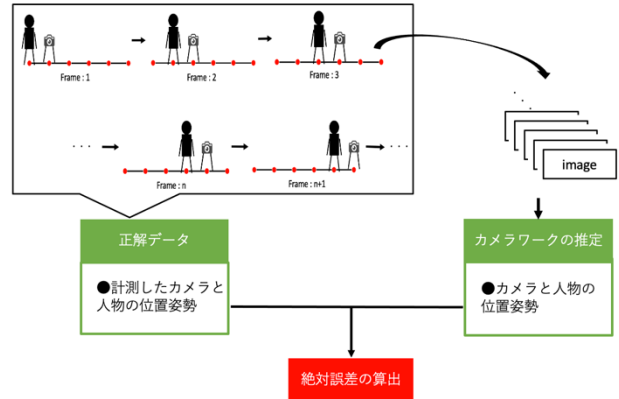


図3 評価実験の流れ

5. おわりに

SfMで復元した撮影シーンに人物の3次元骨格を統合する方法を考案し、それらによって得られる情報からカメラワークを推定する手法について検討した。今後は、運動視差が含まれないカメラワーク(1)(2)(3)で撮影された映像からシーンの3次元情報を推定する手法について検討を進める。

参考文献

- [1] J. L. Schönberger and J. Frahm, “Structure-from-Motion Revisited,” in 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 4104–4113, doi: 10.1109/CVPR.2016.445.
- [2] J. Long, et. al. Fully convolutional networks for semantic segmentation. CVPR 2015.
- [3] B. Tekin, I. Katircioglu, M. Salzmann, V. Lepetit, P. Fua, “Structured Prediction of 3D Human Pose with Deep Neural Networks,” In BMVC, 2016.
- [4] He, K., Gkioxari, G., Dollár, P., and Girshick, R. : Mask R-CNN, IEEE Conference on Computer Vision and Pattern Recognition, pp.2980–2998, 2017.
- [5] Valentin Bazarevsky Ivan Grishchenko Karthik Raveendran Tyler Zhu Fan Zhang Matthias Grundmann: BlazePose: On-device Real-time Body Pose tracking
- [6] Gao, X.-S., X.-R. Hou, J. Tang, and H.F. Cheng. “Complete Solution Classification for the Perspective-Three-Point Problem.” IEEE Transactions on Pattern Analysis and Machine Intelligence. Volume 25, Issue 8, pp. 930–943, August 2003.