

自由視点映像閲覧に適した全身身体動作の検討 A Study of Whole Body Action for Browsing Free Viewpoint Video

山桐 靖史 北原 格 亀田 能成 大田 友一[†]
Yasushi YAMAGIRI Itaru KITAHARA Yoshinari KAMEDA Yuichi OHTA

1. はじめに

本研究では、自由視点映像の撮影において、仮想カメラを身体動作により直感的に操作するインタフェースについて検討する。

多数のカメラから取得した映像を解析し、視聴者が望む任意の視点位置からの映像を生成する自由視点映像に関する研究が長く進められている [1]-[4]。特に、サッカーフィールドのような大規模空間への応用は、スポーツ等のライブ中継の新しい観戦方式として期待されている [6]-[7]。自由視点映像の生成法については Visual Hull を用いた方式 [5]、視点の内挿を用いた方式 [6]、選手を平面で近似する人物ビルボード方式 [7] など、様々な研究が進められている。このような自由視点映像の生成法に関する研究が増えつつあるため、これからはその利用法として仮想カメラの操作方法に関する研究が待たれている。

仮想空間内の歩行を中心とした移動を身体動作によって実現する研究 [8] もされてきているが、自由視点映像における仮想カメラ操作では、歩行を対象とする実現とはまた異なる問題が存在する。

フィールド上にいる選手と同じようにユーザがふるまうだけで仮想カメラを自在に制御できれば、臨場感の高いスポーツ観戦を楽しむことができると期待される。そうした楽しみを実現するために、自由度の大きいユーザの身体動作を用いて、説明がなくても直感的に仮想カメラの操作ができるような仮想カメラインタフェースが理想である。

本研究では、4 種類の身体動作を使った新しい仮想カメラ操作を提案する。自由視点映像を撮影している仮想カメラを操作して好みの映像を取得するためには、典型的なカメラ操作である前進、方向転換(パン操作)、首振り(チルト操作)、垂直方向移動の 4 種類のカメラ操作が最低限必要と考えられる。この 4 種類の操作を行うためのユーザ側の動作を同じく 4 種類用意する。実生活で使用するような自然な動作で、かつ互いに類似していないような動作として、それぞれ「足踏み」、「体の回転」、「体の傾き」、「腕かき」を選択する。各動作は排他的な関係になっており、ユーザにとって区別しやすく、かつ簡単な動作のみの組み合わせで好みの映像を取得して閲覧できる。本研究では、各操作の制御を実現するための技術基盤として、Microsoft 社の Kinect センサから得られる関節情報を用いる。

2. 身体動作を用いたインタフェース

自由視点映像閲覧時に、典型的に用いられる 4 種類の仮

[†] 筑波大学, University of Tsukuba.

想カメラの操作である、前進、パン操作、チルト操作、垂直方向移動を制御対象とする。これらに対応したユーザの動作には実生活において自然な動作と考えられる足踏み、体の回転、体の傾き、腕かきをそれぞれ選択する。

具体的には、ユーザが足踏みすることで自由視点空間の仮想カメラを前方へ平行移動させる操作を行う。ユーザが方向転換したい方向に体をひねるように動作することで仮想カメラのパン操作を行う。ユーザが空を仰ぐときや地面へ俯くときはその方向に体を前後に傾けることで仮想カメラのチルト操作を行う。空間中を垂直に泳ぐように腕かきをすることで仮想カメラを垂直方向に移動する操作を行う。

垂直方向移動に関しては現実で移動する際、ジャンプ、体の伸び、水中の場合腕かきなどが考えられるが、ジャンプは一瞬の上下運動で移動ではないため候補から外す。体の伸びは身体全体を使うので検出すべき姿勢が判別しにくいといった難点がある。よって、今回は腕かきを採用することにした。

上記で示した 4 種類の動作は互いに排反な関係にあり、ユーザにとっては区別が容易である。これらの動作を組み合わせることで仮想カメラの移動を制御でき、自由視点映像を好きなように閲覧することが可能となる。これを Figure 1 で示す。ここで、図中の円錐はユーザの視線方向を示している。

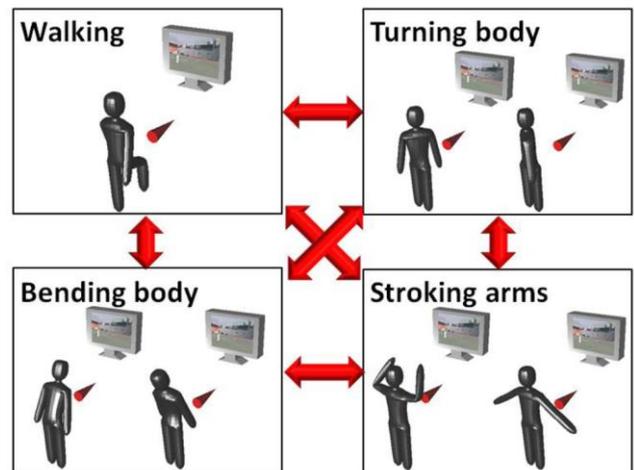


図 1. 4 つの全身身体動作

一般に、仮想カメラの操作対象となるのは、仮想カメラの現在位置を表す「視点」と仮想カメラがどこに注目しているのかを表す「注視点」の 2 点である。スポーツ観戦などでは、仮想カメラのロールは不要であるため、2 点の位置関係により仮想カメラの姿勢が一意に決定されることになる。本提案インタフェースでは、注視点はユーザの向いている方向の先にあるとする。これによって、

ディスプレイ上にはユーザ視点から見たような一人称視点映像が提示できるので、ユーザはあたかも選手と同じフィールド上で観戦しているかのような臨場感のある映像を閲覧できる。

ユーザには、身体動作をしている間、常に好みの映像が閲覧できるような環境が必要となる。一方で、我々の考えるインタフェースではディスプレイはユーザの正面に置かれている。このため、ユーザが全く自由に身体動作するとディスプレイから視線が外れてしまうことになる。そこで、自然な実生活での動作を取り入れつつ、ディスプレイを常に見続けることができるようにするための制限が必要になる。ここで、制限とはディスプレイから視線が離れない範囲でユーザの身体動作を制限することである。動作認識は、ユーザの視線制限を考慮してパラメータ調整を行う。



図 2. Walking



図 3.GOAL1 (CG)

3. 実験

動作確認とパラメータ値の調整、また被験者実験による評価を行うため提案インタフェースを実装した。自由視点映像の生成には、Koyama らの人物ビルボード方式[7]を用い、関節などの計測処理には Microsoft 社の Kinect for Windows SDK(ver.1.5)を用いる。リアルタイムに得られる人体の関節点情報をもとに、4動作の認識を実時間で行うシステムを開発した。

まず、仮想カメラ操作に対応する身体動作の適切性を確認する実験を行った。被験者は自由視点映像と仮想カメラの概念を理解する学生 10 名である。各被験者に 4 種類の仮想カメラ操作を説明し、そのあとですぐに 4 つの質問をした。質問内容は、フィールド上で前進、パン操作(右)、チルト操作(上)、垂直方向移動(上)の各操作をしたとき自分ならどう動作しますか、というもので、質問後すぐに対応する動作を実演してもらった。実験結果を表 1 に示す。前進(forward)とパン操作(pan)に関してはほぼ共通しており、それぞれ足踏み、ひねりが直感的動作として用いられていることがわかる。一方でチルト操作(tilt)と垂直方向移動(vertical)に関しては、被験者によって意見が分かれており、特に共通の有効な動作が見られない。よって、チルト操作と垂直方向移動にはそれぞれ予め用意した上体反らしと腕かきを今後も用いることとする。

次に、仮想カメラ操作実験として、被験者に、画面に表示された 3 次元空間フィールド上を、覚えてもらった動

作を用いて自由に移動してもらい、本人がよいと言うまで動作の練習をしてもらった。これで本実験に入る前にすべての動作を習得した状態となる。その上で、実験として、スタート地点からゴール地点まで移動するタスクを計 3 回行った。スタート地点は毎回同じ(フィールド中央)であり、ゴール地点のみ 3 種類異なる地点を用意した。ここでゴール地点とは、事前に示しておいたゴール画像と同じ見え方となる地点とする。GOAL1 の例を図 3 に示す。なお、用意した地点は 4 種類の動作をすべて使わないとたどり着けないようになっている。ゴール地点への到着の判断は被験者の申告による。実験の結果、3 つのタスクで移動に要した総時間は 103~265 秒であった。分散が大きいのは、今回の仮想カメラ操作に後退がなかったため、一旦思った方向と異なる視点になるとその修正に時間がかかる様子が見られたためである。この実験後、1 から 5 までの 5 段階評価で主観評価を得たところ、「サッカー観戦に使用することに適していると思う」に対しては 2.7 とやや否定的な結果となった。しかしながら、「実際にこれを使用してサッカー観戦したいと思う」に対しては 3.3 とわずかではあるが肯定的な傾向となった。今回の実証実験ではまだ自在な仮想カメラ操作が提案する身体動作によって実現できたとはまでは言えないが、利用者としてはそれでも期待を寄せている傾向がみてとれる。今後は後退などの仮想カメラ操作を導入し、改良を重ねていく必要がある。

表 1. 仮想カメラ操作に対応する動作

被験者ID	forward	pan	tilt	vertical
1	足踏み	ひねり	上体反らし	腕かき
2	足踏み	ひねり	上体反らし	ジャンプ
3	足踏み	ひねり	両手挙げ	両手横上げ
4	足踏み	ひねり	首上向き	背伸び
5	足踏み	ひねり(足有)	上体反らし	ジャンプ
6	足踏み	ひねり	首上向き	ジャンプ
7	足踏み	ひねり(足有)	首上向き	首上向き
8	足踏み	ひねり(足有)	首上向き	前屈→背伸び
9	足踏み	ひねり	片手を上へ	前屈→正立
10	足踏み	片手を横へ	片手を上へ	両手を上へ

参考文献

- [1] T. Kanade, P. Rander, P.J.Narayanan, "Virtualized Reality: Constructing Virtual Worlds from Real Scenes", IEEE Multimedia 1997, Vol.4, No.1, pp.34-47.
- [2] W.Matusik, C.Buehler, R.Rasker, S.J.Gortler, L.McMillan, "Image-Based Visual Hulls", ACM SIGGRAPH, pp.369-374, 2000.
- [3] J. Carranza, C. Theobalt, M. A. Magnor, H. Seidel, "Free-Viewpoint Video of Human Actors", ACM Transaction on Graphics, vol.22, No.3, pp.569-577, 2003.
- [4] 谷本正幸, "自由視点テレビ", 三次元映像フォーラム, vol.15, no.3, pp.17-22, 2001.
- [5] A. Laurentini, "The Visual Hull Concept for Silhouette-Based Image Understanding", IEEE Trans. Pattern Anal. Mach. Intell. vol.16, no.2, pp.150-162, 1994.
- [6] N.Inamoto, H.Saito, "Immersive Observation of Virtualized Soccer Match at Real Stadium Model", ISMAR, pp.188-197, 2003.
- [7] T. Koyama, I. Kitahara, Y. Ohta, "Live Mixed-Reality 3D Video in Soccer Stadium", ISMAR, pp.178-187, 2003.
- [8] 矢野博明, 葛西香里, 齊藤秀之, 岩田洋夫, "ロコモーションインタフェースによる歩行感覚の共有", ヒューマンインタフェース学会論文誌, vol.4, no.2, pp.119-126, 2002.