This article is a technical report without peer review, and its polished and/or extended version may be published elsewhere



第21回日本バーチャルリアリティ学会大会論文集(2016年9月)

マルチタッチ入力を用いた 多視点映像切り替えインタフェース

A Method to Switch Multiple-View Videos Using Multi-Touch Interface

北村誠之1), 亀田能成2), 北原格2)

1) 筑波大学大学院 システム情報工学研究科 (〒305-8573 茨城県つくば市天王台 1-1-1, s1620773@u.tsukuba.ac.jp)

2) 筑波大学 計算科学研究センター

(〒305-8577 茨城県つくば市天王台 1-1-1, {kameda, kitahara}@iit.tsukuba.ac.jp)

概要:育児支援現場において保育士やセラピストの介入作業の様子を撮影した映像を用いて、セラピスト養成訓練や子供の行動分析を目的とした取り組みが進んでいるが、複数の被写体が自由に動き回る育児支援現場では、大事な場面を撮り逃すケースが発生しやすい。本発表では、多視点映像を用いた介入作業観察方式を提案する。映像閲覧には撮影した映像を連続的に切り替えるBullet-Time 方式を用いることで高品質な映像提示を実現し、マルチタッチ入力による映像閲覧インタフェースを開発することでセラピストのような画像処理の非専門家にとって閲覧操作のしやすいインタフェースの実現を目指す。

キーワード: 多視点映像, マルチタッチインタフェース, 育児支援, 技能養成訓練

1. はじめに

保育士やセラピストの介入作業の様子を撮影した映像を用いて、セラピスト養成訓練や子供の行動分析を目的とした取り組みが広まりつつある。育児現場では、子供や保育士といった複数の被写体が自由に動き回るため、カメラ1台を用いた撮影では撮影範囲外に被写体が出たり、被写体同士によるオクルージョンが発生したりといった問題が考えられる。そこで我々は、複数地点から同時に撮影した多視点映像を用いた育児・介入作業の観察方式に関する研究に取り組んでいる[1].

多視点映像の一般的な閲覧法の一つは、図1に示すように撮影した全映像を一覧表示し、その中から自分の見たい視点を選択するものである。しかし、このような閲覧法では、数多く表示されている映像の中から希望する視点位置の映像を見つけ出す必要があり、台数の増加に伴い作業が煩雑になるという問題が存在する。

希望する視点を見つけ出す手間を削減する閲覧法として、複数のカメラで撮影した映像を連続的に切り替えるBullet-Time 方式がある.被写体の周囲に複数台のカメラを配置し、カメラの位置関係に従って、つまりカメラの並び順に映像を切り替えながら多視点映像を提示する手法である.カメラの並びに沿って見え方が変化するため、視点

Nobuyuki KITAMURA, Yoshinari KAMEDA, and Itaru KITAHARA





図 1: マルチディスプレイを用いた多視点映像の閲覧

移動感を再現することが可能である。しかし、従来のBullet-Timeでは、カメラ配置の際に設定した位置(多視点カメラの光軸が空間中で交わる点)に注視点が固定されてしまう。この問題に対して、明智ら[2]は、多視点映像からカメラ間の相対的な位置・姿勢と被写体の3次元情報を推定することによってBullet-Time方式を拡張し、観察者が注視点を自由に設定可能な手法を提案している。この手

法を育児支援現場で撮影した映像に適用することにより、子供や保育士、セラピストの顔や手を中心とした視点切り替えが可能になり、観察の自由度が増加した.本研究でもこの閲覧手法を用いることとする.

教育の現場で Bullet-Time 映像を閲覧する場合,閲覧者となる保育士やセラピストが 3 次元画像処理の知識に明るいとは言い難い.本研究では,そのようなユーザでも閲覧しやすいインタフェースの実現を目指す.我々はマルチタッチ入力による閲覧インタフェースに着目している.タッチ操作を入力インタフェースとするタブレット端末が急速に普及した結果,多くの人々がその操作に慣れていることが期待できる.またユーザがデバイスを直接ポインティングするため直感的な操作が可能である.

本稿では、直感的な視点操作の実現を目標としたマルチタッチ入力による多視点映像閲覧インタフェースを提案する. 本提案インタフェースを育児支援現場に導入することにより、セラピスト養成訓練の効率化が期待できる.

2. 関連研究

2.1 多視点映像閲覧

サッカーのような大規模空間で行われるスポーツ競技では、選手やボールが常に移動するため、固定カメラ1台を用いて撮影した映像では選手同士が重なるオクルージョンが発生し、選手追跡が困難となる。岩瀬ら[3]は追跡選手の予測位置情報を用いてオクルージョンの有無を毎フレーム全カメラに対して判定し、オクルージョンが発生していないと判断したカメラを用いて選手位置を推定する手法を提案している。1台のカメラで撮影した映像のみを用いた追跡処理と比べ高精度な追跡処理を実現していることから、多視点映像での観測の有効性が確認できる。グラウンドを取り囲むように多数のカメラを配置し、各カメラを連動して制御することで、選手やボールを様々な方向から観察可能なスポーツ中継を実現した Eye Vision[4]も、多視点映像による多面的な観測の有効性を活用した一例である。

2.2 Bullet-Time の多重解像度閲覧方式

明智ら[2]は、多視点画像群に新たな注視点を再設定し、その点に対してズームイン操作を行うことにより、疑似的な前後方向の視点移動が可能な多重解像度閲覧方式を提案している。注視点を再設定する際に利用するカメラパラメータは、Structure from Motion (SfM) [5]を用いて推定する.推定したパラメータに基づいてステレオ処理を行い観察者の注目点の3次元座標を求める。レンズ中心から注目点に向かう直線とカメラの光軸とがなす角を算出し、図2に示すように、その角度が0になる、つまり画像の中心で注目点が観察されるように撮影カメラをレンズ中心周りに回転させる(2次元射影変換行列を推定する)ことにより、Bullet-Time 注視点の再設定を実現している。

上述した処理で生成される多視点画像上では,注目物体の見かけの大きさが不均一であるため,映像切り替え時

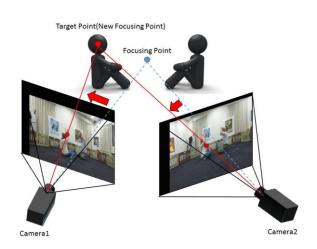


図 2: 注視点再設定処理

の違和感の原因となる. そこで,提示画像を撮影する仮想 カメラの内部パラメータを統一することにより,注目物体 の見かけの大きさを均一にする.

2.3 マルチタッチ入力を用いた仮想カメラ操作

柏熊ら[6]は、任意視点からの映像を生成する自由視点映像において、マルチタッチ操作を用いて仮想カメラを操作する手法を提案した。多様なマルチタッチ操作の中でどの操作が仮想カメラ操作に適しているかを実験的に調査している。本研究では、その実験結果を参考にし、マルチタッチ入力による仮想カメラ操作を実装する。

3. Bullet-Time 閲覧システム

Bullet-Time 映像撮影・生成・提示システムでは、被写体の周囲に複数台のカメラを配置し同期撮影を行う. 特に、本研究で対象とする育児支援現場では、子供の衝突事故を回避するために、保育用サークルパネルに無線操作が可能な小型カメラを取り付ける. 以降、この撮影装置一式をメディアサークルと呼ぶ.

Bullet-Time における注視点は、通常撮影画像の中央に固定される。図3に示すようにユーザが撮影時の注視点と異なる点を注目している場合、その点が注視点と一致するような2次元射影変換を推定し、撮影カメラを仮想的に首ふ

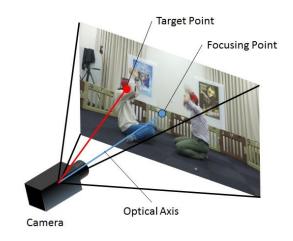


図 3: 注視点 (Focusing Point) と注目点 (Target Point)

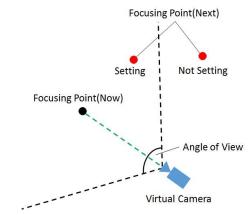


図 4: 注視点と仮想カメラ (Virtual Camera) の関係

りさせる必要がある. 図3は被写体(左)の頭を注目点と している. この2次元射影変換は注目点の3次元位置をス テレオ法で推定することにより算出する.

この2次元射影変換を推定する際に, 誤った注目点の3 次元位置が推定される場合がある. そこで図4に示すよう に, 仮想カメラの位置・姿勢を用いて注目点が存在し得る 範囲を求め、その範囲外の3次元位置が算出された場合に は、位置推定が失敗したと判断して注目点を設定しないこ ととする.

4. マルチタッチ操作

Bullet-Time 映像を閲覧する際に必要となる操作は、主 に視点切り替えと注目点の設定である. そこで我々は、柏 熊ら[6]による研究を参考にし、マルチタッチを用いた操作 インタフェースを実現する. 柏熊らは, 自由視点映像を撮 影する仮想カメラ操作とマルチタッチ入力の関係性につ いて調査を行い、仮想カメラの移動と相性の良いタッチジ エスチャに関する知見を得た. 我々はその知見をベースと しながら Bullet-Time 映像における仮想カメラ操作のタッ チジェスチャについて検討を進める.

視点移動に対応した多視点画像の切り替え操作では, 左 右方向はストローク操作、前後方向(拡大縮小)はピンチ イン・アウト操作を用いる. 仮想カメラを左右方向へ移動, つまり視点切り替えを行う場合には、図 5(b)に示す1本指 のストローク操作を用い,指を動かした方向と同じ方向へ 移動させる. また, 注視点方向への移動は, 図 5(d),5(e)に 示すピンチイン/アウトのジェスチャを採用した. 図 5(d) のピンチインはディスプレイ上に置いた2本の指を近づ ける操作であり、図 5(e)のピンチアウトは2本の指を遠ざ ける操作である. 仮想カメラが注視点へズームインする場 合はピンチイン, ズームアウトする場合はピンチアウトの ジェスチャを行う. 注視点を与えるポインティング操作は 正確な入力が必要と考えるため、図 5(a)に示す1本指での タップ操作を採用する. 映像閲覧中に提示画像をウィンド ウ内で平行移動させる方法として、図 5(c)に示す2本指の ストローク操作を採用した. 指の動きにあわせて画像が平 行移動する. 以上のタッチジェスチャを表1にまとめる.

5. 実証実験

5.1 撮影環境

慶應義塾大学こども支援ラボにおいて多視点映像撮影 データの撮影実験を行い提案手法の実証実験を実施した. 撮影空間は 658[cm]×496[cm] の広さの支援現場であり, 図 6に示すように部屋全体を使いカメラ 20 台を円形に配置 した. 撮影に使用したカメラは GoPro 社 HERO4 ブラック エディションである. 4K 解像度 (3840 画素×2160 画素) の映像を毎秒30枚撮影した.メディアサークルに用いる パネルは幅 71[cm] ×厚み 5[cm] ×高さ 56[cm]の育児用パネ ル (マロンS) を 20 枚用いた. また, SfM によるキャリ ブレーションの精度をあげるために、メディアサークルの 外側にポスターを複数枚設置した. 被写体は大人 2~3 名で, 育児支援の模擬セッション場面を約30分撮影した.撮影 後の SfM には VisualSFM [7] を用いた.

表 1: タッチジェスチャー覧

指の本数	ジェスチャ	仮想カメラ操作
一本指	タップ	注視点設定
	ストローク 左右	仮想カメラ移動 左右
二本指	ストローク 上下左右	画像の平行移動
	ピンチイン・アウト	仮想カメラ移動 注視点方向



(b)







図 5: タッチジェスチャ方法





図 7: 提示画像例(上:ズームアウト 下:ズームイン)



図 6: 撮影環境

5.2 結果及び考察

図7に生成された Bullet-Time 映像の一例を示す.これは男性の手の部分に注視点を設定した場合の画像である.マルチタッチ入力により直感的な操作が可能となり,スムーズな映像閲覧を行うことができた.しかし,注視点が誤推定される場合が存在し,その状態で視点切り替えを行うと提示画像が乱れてしまう問題があった.誤推定時の処理について検討が必要と考える.

6. おわりに

本研究では、マルチタッチ入力を用いて Bullet-Time 映像を閲覧するインタフェースを提案した. 映像閲覧時に必要となる仮想カメラ操作に対応したマルチタッチジェスチャを実装することで、より直感的な操作による映像閲覧を可能にした. また、実際に育児支援現場で撮影実験を行い、Bullet-Time 映像の閲覧を行った. 注視点の誤推定を減らすことでインタフェースの操作性を向上することが今後の課題である.

本研究は、科学技術振興機構(JST)の戦略的創造研究推

進事業(CREST)における研究領域「人間と調和した創造的協働を実現する知的情報処理システムの構築」の研究課題「ソーシャル・イメージング:創造的活動促進と社会性形成支援」により行ったものである.

参考文献

- [1] 志田全弘,明智奈央,北原格,大田友一,松田壮一郎,山本淳一,鈴木健嗣,"簡単に組み立て可能な多視点映像の撮影および閲覧方式",電子情報通信学会 CNR/PMRU 研究会,信学技報,vol.114,no.455, CNR2014-57,pp.143-148,2015
- [2] N.Akechi, I. Kitahara, R. Sakamoto, Y. Ohta, "Multiple Resolution Bullet-Time Effect", ACMSIGGRAPH-Asia, 2014
- [3] 岩瀬幸子,斉藤秀雄, "多視点画像を用いたサッカー 選手の追跡",情報処理学会研究報告コンピュータビ ジョンとイメージメディア(CVIM)2002, pp.119-124, 2002
- [4] T. Kanade et al, "Eye Vision", http://www.ri.cmu.edu/events/sb35/tksuperbowl.html. 2001
- [5] C. Wu, "Towards Linear-Time Incremental Structure from Motion", Proceedings of the 2013 International Conference on 3D Vision, pp.127-134, 2013
- [6] J. Kashiwakuma, I. Kitahara, Y. Kameda, Y. Ohta, "A Virtual Camera Controlling Method Using Multi-Touch Gestures for Capturing Free-viewpoint Video", The 11th edition of the European Interactive TV Conference (EuroITV), pp.67-74, 2013
- [7] C. Wu, "VisualSFM: A Visual Structure from Motion System", http://ccwu.me/vsfm, 2011