

多視点映像中の同時観測軌跡を用いた カメラ協調に基づく対象追跡手法の検討

丸谷 宜史^{†1} 間瀬 健二^{†1}

多視点映像は映像シーン中の様子を自由な視点から観賞することが可能である一方、視点操作に起因する急激な視点位置変化により人物等の映像中の注目対象を見失いやすい問題がある。そこで本研究では各視点における映像中の注目対象の位置を視聴者に明示することで上記の問題を解決することを目指し、その実現のために多視点映像の各視点における注目対象位置を自動で追跡する手法を検討する。一般に単一視点から得られる映像における対象追跡では、隠蔽による特徴点検出の失敗、特徴点対応付けの失敗は回避できない。そこで、本研究では各カメラから同時に観測できる同一対象の軌跡情報を利用してカメラ間の基礎行列を求め、この基礎行列と周辺視点のカメラ情報を基に追跡結果を補完することで、各視点における注目対象を追跡する手法を提案する。実験では周辺の視点から補完した追跡結果と手動で追跡した結果を比較することで提案手法の有用性を確認した。

collaborative target tracking based on trajectories acquired by multiple camera videos

TAKAFUMI MARUTANI^{†1} and KENJI MASE^{†1}

Multiple camera videos are useful for users in watching video contents from free view points. In opposite, the users often lose their subjects when switching their viewpoints of the video. Then we aim to support users to watch their focusing subject by setting hints about each subjects' positions. To achieve this goal, we investigate a method for tracking subjects in multiple camera videos. Generally, subject tracking is done by extracting feature points from a single frame image and connect the feature points acquired from sequential frame images. However, various problems (feature point extraction failure, misconnect of feature points, and so on) are caused by occlusions. To overcome this problem, we make use of cameras neighboring each camera. Then we present a method for collaborative subject tracking by using trajectories acquired by multiple camera video. As a result of experiment using real multiple camera videos, we can show the efficiency of our method.

1. はじめに

近年のデジタル放送の普及によりインタラクティブな映像コンテンツや3次元映像への需要が高まってきている。そのため、映像を自由視点やステレオビジョンで見ることを可能とする多視点映像技術は有用である。一方で多視点映像は視点の自由度が高いため、視点選択時にユーザが注目したい被写体に集中して視聴を継続するには、操作が煩雑であるという欠点を持っている。そのため、多視点映像中において視聴者の希望する被写体の発見を容易にするための手掛かりを提供する必要がある。そこで、我々は視聴者にとって希望する被写体の発見を支援することを目的とし、その実現のため撮影された多視点映像中から被写体を検出・追跡する手法について検討した。

本研究ではその中でも多視点映像を用いることで、周辺視点における追跡結果を利用した特定視点における被写体の追跡を行う方法について検討する。以降では本研究で想定する多視点映像視聴支援方法、および各視点で観測される被写体の追跡結果を利用した多視点映像における対象追跡手法について説明する。

2. 多視点映像視聴支援

近年ではスポーツ映像のような映像コンテンツを様々な視点から視聴できるように、多くの視点から映像を撮影、視点を提供する試みが増えている。Kanadeらは多視点映像を撮影し自由な視点を撮影する枠組みを考案し、フットボール映像の提供に利用している^{1),2)}。林ら³⁾やKamedaら⁴⁾はサッカーを対象に、多視点からプレイの様子を撮影し自由な視点を再現・提示する手法を提案した。このようにして獲得される多視点映像は、視聴者だけでなく選手にとっても自身の動作を様々な視点から振りかえることができるため有用である。

このように視点の自由度が上がる一方で、視聴者には視点選択操作の負担も増えるようになってきた。そこで我々の研究グループは多視点映像の視聴支援を目的とした釘付け視聴方式を提案し⁵⁾、その方式を実装した視聴ツールとしてPeg Scope Viewerを開発した。Peg Scope Viewerのインタフェースを図1に示す。釘付け視聴方式とは、映像中の被写体に注目点を設定し、視点を変更してもつねに選択した注目点を中心にとらえた映像提示を可能と

^{†1} 名古屋大学大学院 情報科学研究科
Graduate School of Information Science, Nagoya University



図 1 Peg Scope Viewer のインターフェイス
Fig.1 Interface of Peg Scope Viewer

する方式であり、特定の被写体に集中して視聴することができる。Peg Scope Viewer は映像中の注目対象のリストを画面左部に表示し、視聴者はそのリストから見たい注目対象を選択することで、視聴者による視点操作の負担を軽減することが可能である。ただし、この実現には、映像における対象の位置情報が必須であるため、映像中の対象追跡技術が必要となってくる。

単一視点映像中の対象追跡手法としては、Louisら⁶⁾の手法に代表されるように、映像中の対象の見えや移動速度等をもとに追跡を行う手法が一般的であるが、単一視点映像における追跡ではオクルージョンに起因した追跡失敗を回避することは難しい。しかし、本研究では多視点映像の使用を前提としているため、現在注目している視点以外の視点で観測される対象の位置をもとに、注目している視点における対象の位置を推測することが可能である。

上記の実現には、任意のカメラ間における基礎行列を獲得する必要がある。このような研究としては、多視点カメラ全体から同時に1次元のキャリブレーション用物体を観測し、全視点のカメラの外部行列を推定する宮川らの手法⁷⁾があるが、これらは全カメラからキャリブレーション用物体を観測する必要があり、また映像コンテンツ撮影前にあらかじめ撮影対象となる空間を網羅するように1次元物体を配置する必要がある。また、映像中における人の移動を観測することで、カメラの内部行列・外部行列を獲得するFengjunらの手法⁸⁾等の手法もあるが、この手法も同様にあらかじめ撮影対象となる空間中で人物が歩き

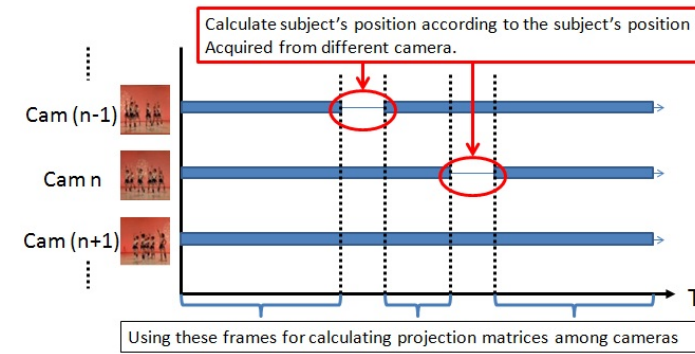


図 2 カメラ協調による対象追跡の概念

Fig.2 collaborative target tracking by multiple camera

回る必要があり、かつ人物の頭から足元までが常に観測される必要がある。これらの手法では多視点映像の映像コンテンツ取得前にあらかじめキャリブレーション用物体や人物が正しく観測される空間を確保する必要があるため、例えばサッカーの撮影でグラウンドに立ち入ることができない等の諸々の理由でそれらの物体を配置できない場合には適用が難しい。

そこで、本研究ではあらかじめキャリブレーション用物体を用いることはせず、各カメラで同時に観測される対象の軌跡をもとにカメラ間の基礎行列を求め、この基礎行列と現在注目している視点以外の視点で観測される対象の位置を利用することで、注目している視点における対象の位置を推測・追跡する手法について検討する。

3. 同時観測軌跡を用いた多視点映像からの対象追跡

3.1 本研究で想定する多視点映像からの対象追跡手法

本研究で想定する対象追跡の方法の概念を図2に示す。図の横軸は時間軸、縦軸はカメラの種類を表しており、また、各カメラにおいて太棒で示されている区間は対象を追跡できている時間区間を、細線で表されている区間は対象を追跡できていない時間区間を表している。

提案手法は対象の位置が観測できないカメラが存在したとき、周辺のカメラと協調して対象の追跡を行う。具体的には、任意の二つのカメラ間で同時に対象の軌跡を観測可能な時間区間が存在するときにはそれぞれのカメラにおける対象の座標値の組み合わせをもとにカメラ間の基礎行列を求めておき、対象が観測できないカメラが存在した場合には対象カメラ



図 3 提案手法による周辺視点の位置情報を用いた対象追跡手法のイメージ
Fig. 3 example of proposed method's result

以外で観測された対象の位置と上記の基礎行列から対象の位置を推測することで追跡を行うものである。その実現イメージを図 3 に示す。

3.2 同時観測軌跡に基づいた対象追跡

3.1 に示したように、本手法はカメラ間の基礎行列を求める処理、注目している視点以外の視点から得られた対象位置と基礎行列を基に対象の位置を推測する処理の二つの処理を行うため、それぞれの処理の詳細を次節に示す。

3.2.1 任意カメラ間の基礎行列導出

今 n_c 台のカメラ、 n_s 個の対象が存在するときに、カメラ C_i ($i = 1, 2, \dots, n_c$) において時刻 t で撮影できる対象 j ($j = 1, 2, \dots, n_s$) の位置を $\mathbf{p}_i^j(t) = (x_i^j(t), y_i^j(t), 1)^T$ とする。(なお、 $\mathbf{p}_i^j(t)$ については、従来手法等により獲得できるものとして、本稿では取り上げないものとする。) ここで、ある二つのカメラ a, b 間の基礎行列を求める場合を考え、ある対象 j について両カメラから同時に観測される時刻の集合を T_0^j とする。このとき観測されるカメラ a, b における対象の位置の集合 $\{p_a^j(t)\}, \{p_b^j(t)\}$ (ただし $t \in T_0^j$) を得ることができる。これらをすべての j について求め、一列にならべると行列 $\mathbf{D}_a, \mathbf{D}_b$ ($\mathbf{D}_a, \mathbf{D}_b$ は $3 \times \sum_j \#|T_0^j|$ 行列) を得ることができる。このとき基礎行列行列 H_{ab} は $\mathbf{D}_b^T \cdot H_{ab} \cdot \mathbf{D}_a = 0$ と表されるので、この式を解くことで基礎行列を求めることができる(基礎行列の計算については既存ライブラリ等を利用するものとする。) 同様に、すべてのカメラ間に対して基礎行列を求めておく。

3.2.2 周辺視点における位置情報を用いた対象位置推定

次にある時刻 t において、カメラ a で対象 j が観測できず、カメラ b で対象が観測することができたとする。このとき、 $\mathbf{p}_b^j(t)$ が得られるので、 $\mathbf{p}_a^j(t)$ は直線 $l_a^b = H_{ab}^T \cdot \mathbf{p}_b^j(t)$ に

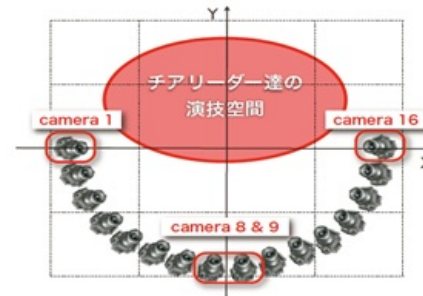


図 4 実験時のカメラ配置
Fig. 4 camera setting



図 5 獲得される多視点映像の例
Fig. 5 example of multiple camera video

存在することがわかる。同様にして対象 j が観測できる他のカメラ i から直線 l_a^i を求めることができるので、最終的にそれらの直線の交点から $\mathbf{p}_a^j(t)$ を求めることが可能になる。上記の処理を全カメラ n_c 、全対象 n_s に対して行うことで、対象追跡を行うことができる。

4. 対象追跡実験

4.1 実験環境

提案手法の有効性検証のため、7人で構成されるチアリーディングの多視点映像を対象として対象の追跡実験を行った。撮影はハイスピードカメラ(CASIO EX-F1, 300fps, 512×384pixels)を16台使って行い、各カメラは原点を中心とした半径537cmの半円状に、原点側を向くように約12度間隔に設置した。実験に用いたカメラ配置の様子を図4に示す。また、各視点のカメラから得られた映像の一部を図5に示す。なお、それぞれのカメラ映像におけるフレーム間での同期については映像を確認して手動で揃えており、各カメラ毎に33000フレーム分取得してある。また、映像中における7人の被写体の胸部中心を追跡点とし、全てのカメラ・全てのフレームにおける全被写体の追跡点を正解データとして手動で検出しておいた。

4.2 実験結果

実験の手順を以下に示す。

- (1) 取得データから任意の n フレーム分の全被写体の追跡点のデータを取り出し、それらを学習データとして各カメラ間の基礎行列を求める。なお、基礎行列の計算はopenCVライブラリ中の関数にアルゴリズムとしてRANSACを指定して行った。

表 1 実験結果
Table 1 experimental result.

学習するフレーム数 n	最小誤差 [pixel]	誤差範囲小の 割合 [%]	誤差範囲中の 割合 [%]	誤差範囲大の 割合 [%]
5000	9.1	32.2	49.8	18.0
10000	10.9	24.0	53.1	22.9
15000	7.3	23.7	48.0	28.3
20000	12.8	19.4	55.4	25.2
25000	17.8	15.3	46.9	37.8
30000	19.6	10.2	56.1	33.7

- (2) ある視点 i のある被写体 j について、隣接する二つのカメラにおける同一被写体の位置情報と基礎行列を基に推定追跡点を計算する。
- (3) 手動で求めた追跡点と推定追跡点との全フレームを通じての誤差平均 e_i^j を評価する。
- (4) 上記手順を i, j, n の組み合わせを変えて行う

上記実験の結果を表 1 に示す。表中の最小誤差は、 n フレーム学習した際の最小の e_i^j を表す。また、誤差範囲小、誤差範囲中、誤差範囲大はそれぞれ、すべての e_i^j のうち、 e_i^j が人間一人分の誤差範囲に収まった場合の割合、 e_i^j が画面内に収まる範囲であった場合の割合、 e_i^j が画面外に出る程度の大きさであった場合の割合を示す。

表 1 から推測されるように、どのような学習フレーム数においても誤差が最小になる場合については人から外れるような追跡が行われないことが確認された。また、学習フレーム数が少なくても誤差範囲小となる追跡が多く行われていることも確認できた。

一方で、多くても 3 割程度の場合しか誤差範囲小である追跡が成功していないことも確認できる。このような実験結果の具体例として、追跡結果の一部を図 6,7 に示す。図中の青・緑線は隣接カメラ上の観測位置から求められたエピポーラ線、白点は手動で抽出した追跡点、赤点は推定追跡点、基線は白点・赤点を結んだ線を表している。追跡の失敗は、図 7 で観測されるような、両エピポーラ線がほぼ平行となる場合によく起きていた。

これは今回の実験では追跡点の大半が画像を上下に分割する中心線付近に集中したためであり、基礎行列の推定が精度よく出来なかったためだと考えられる。また学習するフレーム数の増加に伴い、誤差範囲が増える傾向も、上述のように追跡点の大半が画像の中心線付近に集中したことが原因であると考えられる。実際に学習に使われたシーンの例を図 8 に示す。図中の白点は画像中の追跡点の位置を表している。図 8(a) は多くのシーンで見られた画像の例であり、追跡点がほぼ一直線上に存在することがわかる。一方で図 8(b) は追跡



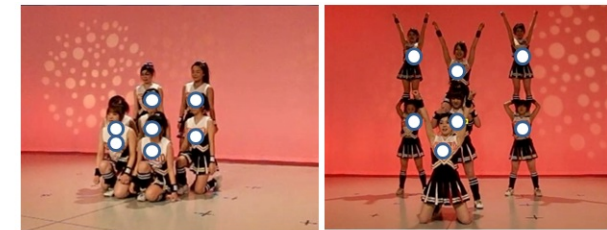
図 6 追跡成功例
Fig. 6 success example



図 7 追跡失敗例
Fig. 7 failure example



(a) An example of Usual scenes



(b) Examples of Special scenes

図 8 学習に使われた映像シーンの例
Fig. 8 examples of scenes

結果が良好であったときの学習データに多く使われていたシーンの例であり、追跡点が画像上で上下に分散していることが確認できる。このように、学習に使う追跡点についてある程度位置にばらつきができるよう、特徴点の選択を行うか、カメラ配置を行えば追跡精度の向上が予想されるため、今後それらの方法について検討を行っていく必要がある。

5. おわりに

本研究では多視点映像に対し被写体の位置情報を付加することを目的とし、多視点映像中の対象追跡の手法について検討を行った。その実現のため、多視点映像間で同時に観測され

る軌跡のデータをもとに視点間の基礎行列を求め、周囲のカメラから得られる位置情報を基に対象追跡を行う手法の検討を行い、実験により提案手法の検証を行った。

今後の課題としては、他の被写体を用いた場合の検証のほか、対象追跡の自動化や、より信頼できる追跡区間のみを利用することによる基礎行列の推定精度の向上、等を通じて手法の有効性を確認することが挙げられる。

謝辞 この研究の一部は独立行政法人情報通信研究機構（NICT）の高度通信・放送研究開発委託研究／三次元映像通信・放送のための中核的要素技術の一環としてなされたものである。また、本研究の遂行にあたり、多くの実験機会と有益な助言を頂いた中京テレビ放送株式会社の川本哲也氏らに深く感謝いたします。

参 考 文 献

- 1) T.Kanade,P.J.Narayanan,P.W.Rander : Virtualized reality:concepts and early results, Proc. of IEEE Workshop on Representation of Visual Scenes, pp.69-76,1995
- 2) Eye Vision,
available from (<http://www.ri.cmu.edu/events/sb35/tksuperbowl.html>)
(accessed 2011-4-20)
- 3) 林邦彦, 斎藤英雄: 多視点サッカー映像からの自由視点映像生成, 情報処理学会研究報告. CVIM, 2006(51), pp.173-180, 2006
- 4) Kameda, Y. Koyama, T. Mukaigawa, Y. Yoshikawa, F. Ohta, Y. : Free Viewpoint Browsing of Live Soccer Games, IEEE International Conference on Multimedia and Expo, 2004. ICME '04.Vol.1, pp.747- 750, 2004
- 5) 間瀬健二, 東海彰吾, 川本哲也, 藤井俊彰: 多視点画像の釘付け視聴方式と操作インタフェースのデザインに関する考察, Human Interface Society, Vol.11, No.1, pp.7-12,2009.
- 6) Louis Kratz , Ko Nishino: : Tracking with Local Spatio-Temporal Motion Patterns in Extremely Crowded Scenes, IEEE conf, CVPR2010, pp.693-700 ,2010.
- 7) 宮川勲, 富山仁博, 岩館祐一: 1次元物体を使った多視点カメラキャリブレーション. 信学技報. PRMU2006-177, pp.37-42, 2006.
- 8) Fengjun Lv, Tao Zhao, Ramakant Nevatia : Camera Calibration from Video of a Walking Human, IEEE trans. PATTERN ANALYSIS AND MACHINE INTELLIGENCE, VOL. 28, NO. 9,pp.1513-1518, 2006.