

ズームカメラの動的パラメータ推定のための SIFT と SURF の性能比較

須田 新^{*1}, 村上 仁己^{*2}, 小池 淳^{*3}, 三功 浩嗣^{*4}, 内藤 整^{*4}

Performance Comparison between SIFT and SURF for Dynamic Camera Calibration Method of Zoom Camera

Arata SUDA^{*1}, Hitomi MURAKAMI^{*2}, Atsushi KOIKE^{*3}, Hiroshi SANKOH^{*4} and Sei NAITO^{*4}

ABSTRACT : We introduce the parameter estimates when using non-fixed camera. Various methods have been proposed as a free viewpoint video synthesis techniques, although camera parameter(projective relation) is assumed that the space between the camera and shoot the target is known, that was restriction that it can't move the imaging camera. We propose a way that we need camera fixed in position but we can change the angle of it by using the function of pan tilt, and zooming and focusing the camera. The following is our proposition of how to detailed studies on the feature points matching process between frames using the non-fixed camera.

Keywords : Free Viewpoint, SIFT, SURF, Matching feature points

(Received September 21, 2012)

1. はじめに

近年、従来のテレビより更に臨場感を得ることが可能である新たなテレビというものが注目されている。それには3Dテレビ, 超高精細テレビ, 自由視点テレビがある。特に自由視点テレビでは、視聴者が任意に選択する視点にあわせて3次元空間を表示させ、よりインタラクティブな感覚を想起させることが可能になる。

自由視点テレビでは、複数台のカメラで撮影された実写の多視点映像を元に、撮影対象の3次元モデリングを行う。そうすることにより単純にカメラ映像を切り替えるだけでなく、カメラを配置不可能な視点も含め、3次元空間中の任意視点への映像を再現・表示することが可能になる[1]~[5]。これにより人と人との間を通過する事のできる「ウォークスルー」といったことが可能になる。

従来の研究として大空間を対象に、カメラの強校正を前提とせず、画像情報のみを用いて任意視点映像を合成

する手法が存在する [6]。しかしながら、各撮影カメラは固定されており、画角内に撮影対象又はフィールド全体を含めていることを前提としていたため、人物等の被写体領域の解像度が不十分であるという問題があった。また、通常は非固定のカメラで撮影されるサッカー等のスポーツ中継映像では、カメラを動かしてしまう。そのため、カメラパラメータが変化してしまい撮影映像から被写体の3次元位置の検出を行うことができないといった問題があった。故に、既存の映像を適用することが出来ないという課題があった。

上記の課題を踏まえ、本論文では、カメラ本体の3次元位置は固定のまま、パン・チルト及びズーム・フォーカス値の変更を伴う非固定ズームカメラへの応用を目指し、動的に変化するカメラパラメータを推定する方式[6]の精度向上を目標としフレーム間の特徴点マッチング処理に関するより詳細な検討を実施する。

本稿の構成を以下に示す。2章では従来手法の流れ、と問題点の考察を行う。3章では本研究の提案手法についての説明を行う。4章では本研究で行った実験の結果と考察を行う。5章では、本稿のまとめと今後の課題について述べる。

^{*1}: 理工学研究科情報科学専攻

^{*2}: 理工学部情報科学科

^{*3}: 理工学部情報科学科 (koike@st.seikei.ac.jp)

^{*4}: 株式会社KDDI研究所

2. 従来手法

従来手法では、コート競技を対象に、世界座標が既知であるフィールドライン情報を用いて、フィールド平面と撮影カメラ間のカメラパラメータ(平面射影行列)を、フレーム間の特徴点マッチングに基づき動的に推定する手法に注目する。当該手法のSIFT及びSURF特徴量を用いた性能評価を実施することで、両者の特性を明らかにした[7]。従来手法の平面射影行列の推定処理手順は以下の通りである。

- (1) 初期フレーム(フレーム 0)の平面射影行列を計算
- (2) フレームtとフレームt-1 間で対応点候補を抽出 (t=1 から開始する)
- (3) フレームtの平面射影行列を計算
- (4) t=t+1 とし、ステップ(2)へ
(t=Tの時処理を終了)

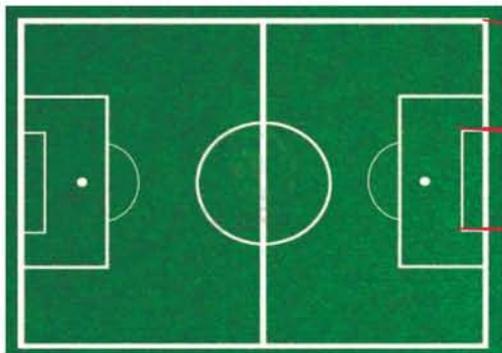
ステップ(1)では、フィールド上のラインの交点など、予め3次元座標が既知の特徴点を元に、フィールド平面上の座標(u, v)と、初期フレームでの画素(u₀, v₀)を式(i)のように対応付ける平面射影行列H₀を推定する。

$$s(u, v, 1) = H_0(u_0, v_0, 1) \quad \dots(i)$$

ステップ(2)では、フレームtとフレームt-1 間で、SIFT特徴量を使用する。

ステップ(3)では、抽出された対応点を使用し、平面射影行列H_tの推定を行う。フレームtで抽出された特徴点の世界座標は、フレームt-1 における対応点の画素座標、及び平面射影行列H_{t-1}を用いて、式(ii)の通り計算することが可能である。

$$s'(u^{(i)}, v^{(i)}, 1) = H_{t-1}(u_{t-1}^{(i)}, v_{t-1}^{(i)}, 1) \quad \dots(ii)$$



ステップ(4)では、t=t+1とし、ステップ2へ戻るといった作業をt=Tとなるまで行う。

2.1 平面射影行列

上記手法ステップ(1)とステップ(3)で平面射影行列を使用した。平面射影行列とは、3次元空間のある平面から別の平面へ投影を行うといったものである。実験ではサッカーの映像を使用しているため、今回においてはフィールド平面から、撮影カメラ映像への投影ということになる。(図1)

2.2 SIFT特徴量

ステップ(2)で使用した対応点候補の抽出方法としてまずSIFT特徴量の説明を行う[8][9]。SIFT(Scale-Invariant Feature Transform)は、特徴点を取得する部分と特徴量を取得する部分に分けることができる。このSIFT特徴量は照明の変化や回転、拡大縮小、移動に不変といった性質を持っている。SIFT特徴量使用時に抽出されるデータの中身としては、特徴点の位置、輝度勾配の方向、スケール、128次元の特徴量が取得される。この128次元の特徴量は特徴点の周囲を16(4*4)ブロックに分割し、各ブロックに対して8方向の勾配ヒストグラムを作成する。結果として128(=4*4*8)の特徴ベクトルが作成される。(図2)

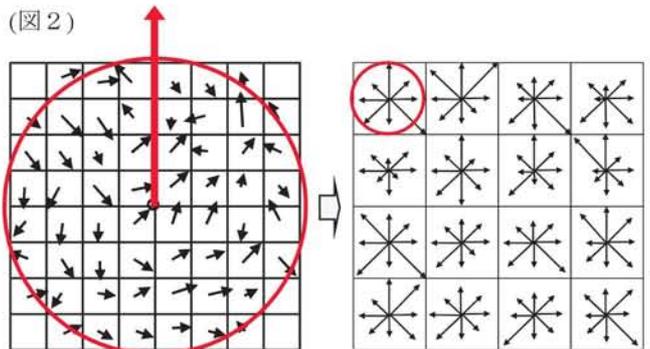


図2 SIFT特徴量

図1 フィールド平面から撮影カメラ映像への投影

2.3 SURF特徴量

SURF(Speeded Up Robust Features)は、SIFTと同じように照明の変化や回転、拡大縮小、移動に不変といった性質を持っている[10]。SURF特徴量を使用した時に抽出されるデータの中身として、特徴点の位置、輝度勾配の方向、スケール、64次元の特徴量が取得される。この64次元の特徴量は、特徴点の周囲を16(4*4)ブロックに分割し、各ブロックに対して4方向の値を作成する。結果として64(=4*4*4)の特徴ベクトルが作成される。(図3)

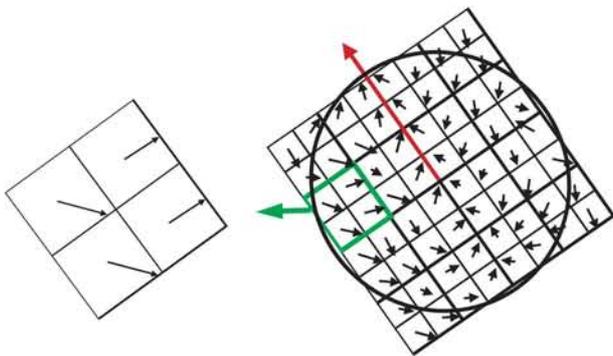


図3 SURF特徴量

3. 提案手法

本研究では、従来手法で説明したステップ(2)とステップ(3)の間に特徴点候補の抽出を行った場合のSIFT特徴量とSURF特徴量の比較を行った。

本研究の特徴点候補の絞込みをフィールド平面内で行う。理由としては、SIFT特徴量とSURF特徴量を取得後に特徴点マッチングを行う。その際、必要な特徴点はフィールド平面である。フィールド平面以外の部分の特徴点を取得すると、特徴点マッチングを行う際にマッチングが正しく行われなかった問題が生じる。他にも、フィールド平面以外の部分の特徴点マッチングを行うとカメラとは逆方向に動いている観客など不要なマッチングも行ってしまう。すると、カメラパラメータ推定には必要でない部分のマッチングも行ってしまうためである。

実行ステップは以下に示す。

- (1) 初期フレーム(フレーム0)の平面射影行列を計算
- (2) フレームtとフレームt-1間で対応点候補を抽出 (t=1から開始する)
- (3) フィールド平面上の特徴点候補のみを流出
- (4) フレームtの平面射影行列を計算
- (5) t=t+1とし、ステップ(2)へ
(t=Tの時処理を終了)

4. 実験

本論文の検討を行う目的で非固定ズームカメラ映像を対象に実験を行った。

実験ではサッカーの映像を使用した。2次元フィールドの大きさ(図4)はゴールが面している部分を横として考え縦52.5(m)、横34(m)とした。

映像はサッカーフィールドを対象に特定の選手、又はボールを追跡するように撮影した映像を、各フレームごとに分割し画像とした。画像の解像度は縦1280画素、横720画素とした。

(図5)

フィールド座標は初期フレーム画像に表示されている点(0,0), (12.42,0), (12.42,2.75), (21.58,2.75), (21.58,0)の5つの点に注目し、撮影画像の画素座標で初期フレームのパラメータ推定(平面射影行列)を行った。

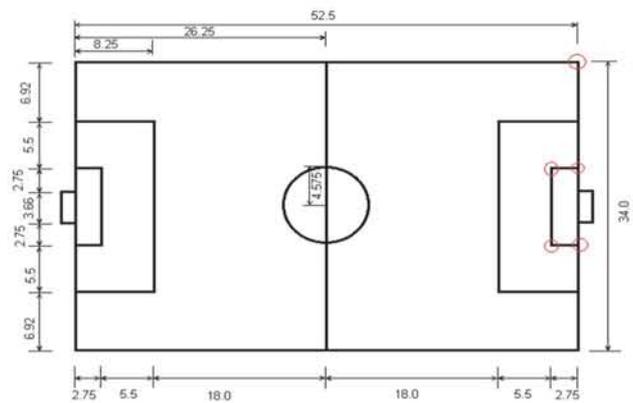


図4 サッカー寸法図と注目点



(1) 初期フレーム



(2) フレーム 30

図5 ズームカメラ映像

4.1 定性評価

図5の映像を対象に、SIFT特徴量及びSURF特徴量の性能比較を行った。まず、定性評価を行った。特徴点をフィールド内に絞り込むために、図4の寸法を利用した。抽出した特徴点を平面射影行列で計算し、図4の寸法内にあるものに絞り込みを行った。特徴点を表した例を図6とし、各特徴量により推定された平面射影行列により、フィールドライン上に推定されたフィールドラインを投

上記手順を行う際の世界座標が既知の画像の特徴点とは、図4で示した5つの注目点に対応する画素を示している。

定量評価の結果は表1に示す。表1の結果をグラフにしたものを図8で示す。

表1, 図8からSURF特徴量を使用した場合でも従来手法で使用されていた、SIFT特徴量と同程度の推定精度が得られることを確認した。

表1 定量評価比較

	5フレーム	10フレーム	15フレーム	20フレーム	25フレーム	30フレーム	35フレーム
SIFT特徴量	0.209323	0.267817	0.209917	0.346667	0.573256	0.90655	1.209435
SURF特徴量	0.208565	0.275068	0.197221	0.32893	0.598717	0.878171	0.938258

(単位 : m)

影した結果を図7に示す。

図7より、各特徴量を用いた場合で推定精度に大きな差はないことが確認できた。

4.2 定量評価

次に、本評価映像に対する定量評価を行った。評価手順については下記に示す。

1. 世界座標が既知の画像上の特徴点を選択する。
2. 上記特徴点の2次元画素座標を取得する。
(本稿ではペイントソフトを使用)
3. 画素座標を平面射影行列により2次元世界座標に射影する。
4. 射影された世界座標と既知の世界座標のユークリッド距離を算出する。

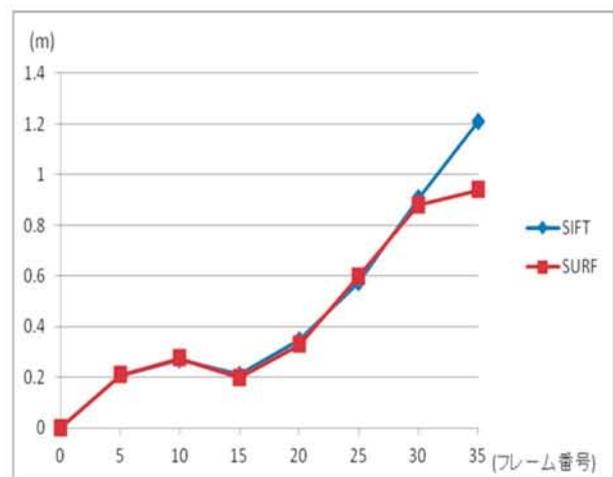


図8 定量評価結果



(1) SURF特徴量



(2) SIFT特徴量

図6 フィールド内特徴点絞り込み後

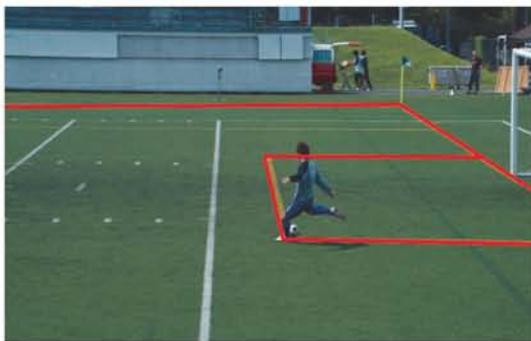


(1) SURF特徴量

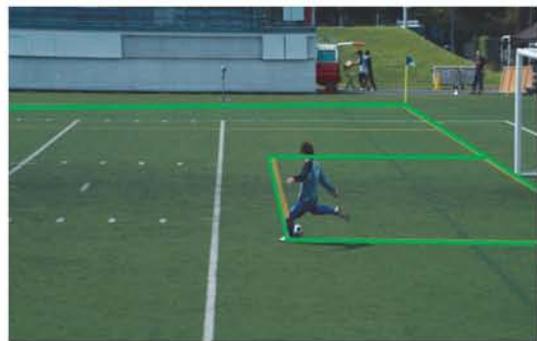


(2) SIFT特徴量

15 フレームの推定精度



(1) SURF特徴量



(2) SIFT特徴量

25 フレームの推定精度



(1) SURF特徴量



(2) SIFT特徴量

35 フレーム目の推定精度

図7 各フレームの推定精度

5. まとめと今後の課題

本稿では、ズームカメラの動的パラメータ推定時のSIFT及びSURF特徴量のより詳細な検討を行った。定性評価の映像からSIFT特徴量とSURF特徴量に推定精度に大きな差がないことが確認できた。定量評価の結果からは、SIFT特徴量とSURF特徴量で同等な推定精度を確保することを確認した。

今後の課題として、撮影対象が動きの遅い人間から動きの速いボールに映ることによりカメラの動きが速くなった場合、特徴点マッチングが正しく行われないうえに、カメラパラメータの推定が正しく行われなかった。この課題を基に、カメラの動きが速くなった場合のカメラパラメータ推定の手法についての検討を行う予定である。

参考文献

- [1] 三功 浩嗣, 石川 彰夫, 内藤 整 "視点間の整合性を考慮した3次元モデルフィードバック型背景分離方式の不要部除去拡張" 電子情報通信学会技術研究報告: 信学技報 109 巻 206 号 75~80 頁
- [2] Akio Ishikawa, Mehrdad Panahpour Tehrani, Sei Naito, Shigeyuki Sakazawa and Atsushi Koike, "Free Viewpoint Video Generation for Walk-through Experience Using Image-based Rendering," ACM Multimedia 2008 Proceedings, pp.1007-1008, 2008.
- [3] 石川 彰夫, 三功 浩嗣, 内藤 整, 酒澤 茂之, "ウォークスルー自由視点映像伝送を効率化するテクスチャの適応的な選択方法"映像情報メディア学会誌 Vol. 65 (2011) No. 11 P 1585-1593
- [4] H.Saito, S.Baba and T.Kanade, "Appearance-Based Virtual View Generation from Multicamera Videos Captured in the 3-D Room", IEEE Trans. On Multimedia, Vol.5, No.3, pp.303-316, September 2003.
- [5] 谷本正幸 "自由視点テレビの開発と国際標準" 名古屋大学情報連携基盤センターニュース Vol.6, No.4 pp.320-334 2007.11
- [6] 三功浩嗣, 内藤整, "動画フレーム間の対応点マッチングに基づくズームカメラの動的パラメータ推定", ITE 年次大会 4-5 (2011).
- [7] 須田新, 村上仁己, 小池淳, 三功浩嗣, 内藤整, "ズームカメラの動的パラメータ推定時の SIFT 特徴量と SURF 特徴量の性能比較", IEICE 総合大会(2012)
- [8] 藤吉 弘亘 " Gradient ベースの特徴抽出: SIFT と HOG(チュートリアル)" 情報処理学会研究報告. CVIM, [コンピュータビジョンとイメージメディア] 2007(87), 211-224.
- [9] 藤吉 弘亘 " 画像局所特徴量 SIFT と最近のアプローチ" 人工知能学会誌 25(6), 753-760, 2010
- [10] 森田 裕之, 木村 真一 " SURF 特徴量による VSLAM フロントエンドの高速化" ロボティクス・メカトロニクス講演会講演概要集 2009, 1A1-F12(1)-1A1-F12(4)