

I-054

簡易カメラ校正に基づく大空間自由視点映像の高精度合成方式 Highly Precise Generation Scheme of Wide Space Free Viewpoint Video with Simplified Camera Calibration

三功 浩嗣
Hiroshi SANKOH

石川 彰夫
Akio ISHIKAWA

内藤 整
Sei NAITO

酒澤 茂之
and Shigeyuki SAKAZAWA

(株) KDDI 研究所
KDDI R&D Laboratories Inc.

1. はじめに

近年、視聴者が任意に選択する視点に合わせて3次元の空間を表示することが可能な自由視点映像の合成技術が注目されている[1]。代表的な合成方法として、多視点画像から被写体3次元モデルを復元し、実写画像のテクスチャをマッピングすることで合成を行う手法が提案されている[2]。多視点のカメラ情報をもとに3次元モデルを復元する場合、対象空間の3次元座標と各カメラの画素の対応関係を表す空間射影行列が正確に推定されていることが前提となる。しかしながら、サッカーフィールド等の直径20メートル以上の大空間を対象とした場合、空間射影行列を正確に推定すること(カメラの強校正)は極めて難しい。その結果、推定誤差の影響により、3次元モデルの復元精度が低下するため、最終的に合成される自由視点映像の精度が不十分になるという課題がある。また、生成される3次元モデルのデータ量が極めて大きくなるため、非常に大きな計算量を要する課題がある。

一方、3次元形状モデルを陽に復元しない画像ベース方式の手法として、カメラの強校正を前提とせず、多視点画像間での射影幾何的關係のみを用いる視点内挿法[3]が提案されている。文献[3]では、シーンを人物等の移動領域、フィールド領域、遠景領域の3領域に分割し、領域ごとにエピポラ幾何による拘束条件や平面射影等を適用することで、仮想的に固定カメラの中間に視点を移動させたような見え方を実現している。しかしながら、3次元空間中での位置情報は考慮されず、画像間の対応点探索に基づくモーフィングによる合成処理が基本となるため、画像中で被写体領域が重なるオクルージョン等の影響により対応点が正しく取得できない場合、合成精度が低下するという課題がある。

上記の問題を踏まえ、本稿では、カメラの強校正を前提としない一方、大空間を対象に撮影された多視点画像に含まれる幾何情報をもとに、対象領域の3次元位置情報を考慮した自由視点映像の合成手法

を提案する。具体的には、多視点画像のみから対象空間をモデル化し、モデル座標空間とカメラ間で成立する射影関係を推定することで、3次元空間中の任意視点における合成を可能とする方式を提案する。

2. 関連研究

本章では、大空間を対象とした自由視点映像の合成方式のうち、カメラの強校正を前提としない画像ベース方式に限定して関連研究を述べる。

画像ベース方式に基づく手法の大部分は、文献[3]と同様に、多視点画像間の対応点により推定されるエピポラ幾何やモーフィングに基づくものであり[4][5]、3次元空間中での位置情報を考慮したものは少ない。カメラ間で成立するF行列(Fundamental Matrix)をもとに射影グリッド空間を構築し、当該空間において3次元形状を復元する手法なども提案されている[6]。しかしながら、射影グリッド空間もまた、3次元空間中の位置情報を反映したものではないため、合成できる仮想視点位置は多視点カメラ間の内挿箇所限定されるため、フィールド内や上方からの視点再現は困難であるという課題がある。

一方、上記の課題に対して、単純化3次元モデルを用いた手法[7]が提案されている。文献[7]では、3次元空間中の各被写体を一枚の長方形ポリゴンで近似し、当該ポリゴンに多視点画像から獲得したテクスチャ情報を適切にマッピングする人物ビルボード方式を提案しており、計算コストを抑えた合成を実現している。また、被写体の3次元空間における位置情報が算出されるため、カメラ間の内挿箇所だけでなく、フィールド内の仮想視点合成も違和感なく再現可能である。しかしながら、予め各カメラが強校正済みであることを前提としているため、多視点画像のみが与えられる場合などは、文献[7]の手法をそのまま適用することはできない。

3. 提案手法

上記で述べた課題を踏まえ、カメラの強校正を前提とせず、対象空間の幾何情報をもとに3次元位置情報を考慮した自由視点映像の合成法を提案する。

具体的には、多視点の撮影カメラ画像間での整合性を考慮した上で、対象領域のモデル座標空間を構築し、モデル座標空間と各カメラ間の射影行列を推定する手法を提案する。さらに、構築されたモデル座標空間および推定された射影行列をもとに、自由視点画像を合成する方法について述べる。ただし、被写体領域の合成に関しては、文献[7]の手法をもとに合成を行う。

3.1 モデル座標空間の構築

対象空間のモデル座標を設定する方法について述べる。サッカーフィールドに代表されるように、フィールド平面上ラインやゴールポストの長さに関する情報が既知の場合、各特徴点に適切な座標を設定することで、対象領域の3次元モデル座標を設定可能である。

長さが既知でない場合は、各カメラ画像の幾何情報として、フィールド平面やゴールポスト上の特徴点をもとに、特徴点間の距離比に基づき3次元モデル座標空間を構築する。ただし、各カメラ画像中での距離比ではなく、3次元空間中での距離比を推定する必要がある。そこで、以下のような一連の処理を導入する。

- 各カメラ画像を基準平面に対して正面化
- 正面化された画像間でのスケーリングを統一し、基準画像を設定
- 基準画像中の特徴点座標を一意に決定し、特徴点間の3次元空間中での距離比を算出

(1) 各カメラ画像を基準平面に対して正面化する

まず、各カメラ画像を基準平面に対して正面化する手法について述べる。フィールド平面上のライン等、本来は平行である2直線が、撮影カメラの位置や向きによっては画像中で交わる場合があり、この交点を消失点と呼ぶ。図1のように、2つの消失点を求め、各消失点の画像座標 $V_1(u_1, v_1), V_2(u_2, v_2)$ を算出する。焦点距離を f とすると、レンズ中心を原点とするカメラ座標系における消失点の位置ベクトルは、画像の中心座標 (u_0, v_0) を用いて、それぞれ $V_1(u_1 - u_0, v_1 - v_0, f)$ 、 $V_2(u_2 - u_0, v_2 - v_0, f)$ と表される。2つのベクトルが直交するという幾何的制約から、カメラの焦点距離を式(1)により推定することができる。

$$f = \sqrt{(u_1 - u_0)(u_2 - u_0) + (v_1 - v_0)(v_2 - v_0)} \quad (1)$$

ただし、消失点を求める基準平面に対して撮影カメラが真正面を向いている場合は、消失点は存在せず、焦点距離を推定することはできない。また、画像平

面上の直線群を指定する際に生じる画素ずれ等の誤差により、推定精度が著しく低下する場合もある。

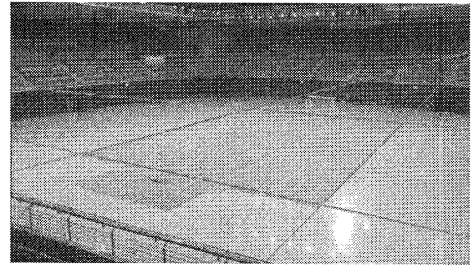


図1. 消失点の算出

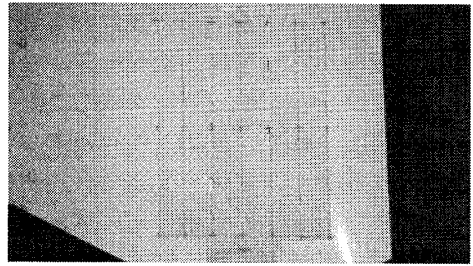


図2. フィールド平面に対する正面化画像

さらに、消失点に対するカメラの方位角 θ および仰角 ϕ は、式(2)および式(3)により推定される。

$$\theta = \arctan \frac{u_0 - u_1}{f} \quad (2)$$

$$\phi = \arctan \frac{v_0 - v_1}{f} \quad (3)$$

以上より、フィールド平面をカメラに対して正面化する変換行列 \mathbf{H}_v は、式(4)により算出される。

$$\mathbf{H}_v = \mathbf{A}_c \mathbf{R}_v \mathbf{A}_c^{-1} \quad (4)$$

ここで、 \mathbf{A}_c はカメラの内部行列であり、式(1)で算出される焦点距離 f および画像の中心座標 (u_0, v_0) を用いて式(5)により決定される。

$$\mathbf{A}_c = \begin{pmatrix} f & 0 & u_0 \\ 0 & f & v_0 \\ 0 & 0 & 1 \end{pmatrix} \quad (5)$$

また、 \mathbf{R}_v は基準面に対する回転行列であり、式(2)および式(3)で算出される方位角 θ 、仰角 ϕ により式(6)により算出される。

$$\mathbf{R}_v = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos(-\pi/2) & -\sin(-\pi/2) \\ 0 & \sin(-\pi/2) & \cos(-\pi/2) \end{pmatrix} \begin{pmatrix} \cos\theta & 0 & -\sin\theta \\ 0 & 1 & 0 \\ \sin\theta & 0 & \cos\theta \end{pmatrix}^{-1} \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\phi & -\sin\phi \\ 0 & \sin\phi & \cos\phi \end{pmatrix}^{-1} \quad (6)$$

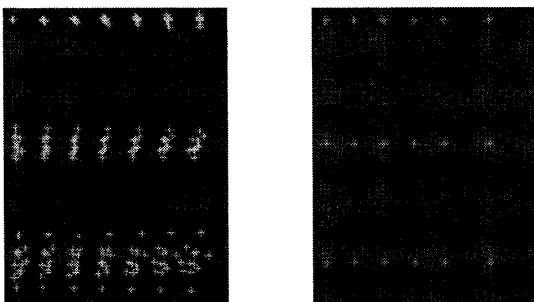
変換行列 \mathbf{H}_c を用いてフィールド平面が正面に来るようにした後、画像内に来るように適宜平行移動することで、図2のような正面化画像を獲得することができる。ここでは、フィールド平面上の特徴点を明示するために画像中に格子点を付与した。

(2) 正面化された画像間でのスケージングを統一し、基準画像を設定する

各カメラ画像について、フィールド平面の消失点を算出し、式(4)により正面化画像を獲得する。この際、カメラごとの推定誤差の影響により、各正面化画像は完全には一致しない。そこで、各カメラの正面化画像ごとにフィールド平面上の特徴点を抽出し、特徴点間の距離比に基づきスケージングの統一を行う。さらに、各カメラでスケージングを統一した特徴点群をモデル空間中の同一平面上に重ね合わせることで、図3(1)のように、特徴点群の分布が得られる。

(3) 基準画像中の特徴点座標を一意に決定し、特徴点間の3次元空間中での距離比を算出する

各視点から抽出された特徴点について、同一直線上に存在するという拘束条件を用いて、クラスタリングすることで、図3(2)のように各特徴点をモデル平面上で一意に定めることができる。さらに、モデル平面上で一意に定められた各特徴点に対して、座標を設定する。同様に、ゴールポスト等を用いて、フィールド平面に直行する平面に対して正面化した画像を、図4のように獲得し、ゴールポストのフィールド平面およびクロスバーとの交点を取得する。さらに、フィールド両端の特徴点を抽出することで、ゴールポストの高さとフィールド領域の幅に関する比率を算出する。以上より、フィールド平面のモデル座標とスケージングを合わせて統合することで、高さ方向の情報を含む、多視点の撮影カメラ画像間で整合性の確保された3次元のモデル座標空間を構築する。



(1) 各カメラの特徴点 (2) クラスタリング後
図3. フィールド平面のモデル座標設定

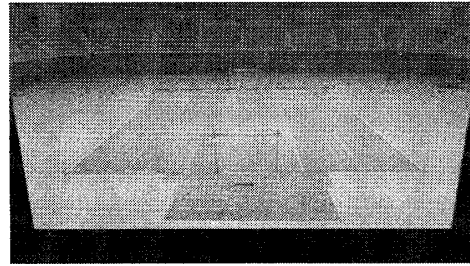


図4. ゴールポストの正面化画像

3.2 射影行列の推定

前節で構築したモデル座標空間中の特徴点座標をもとに、各カメラの射影行列を推定する。以下、フィールド平面と各カメラ間で成立する平面射影行列および、モデル座標空間の3次元座標と各カメラ間で成立する空間射影行列を推定する。

カメラ画像中の画素 (u, v) と、図3(2)のようなモデル座標空間中フィールド平面上の座標 $(u', v', 0)$ の対応関係は、スカラー値 s および平面射影行列 \mathbf{H}_c を用いて、式(7)により算出される。

$$s(u', v', 1)^T = \mathbf{H}_c(u, v, 1)^T \quad (7)$$

さらに、カメラ画像中の画素 (u, v) とモデル座標空間中の座標 (X, Y, Z) の対応関係は、スカラー値 s' および空間射影行列 \mathbf{P}_c を用いて、式(8)により算出される。

$$s'(u, v, 1)^T = \mathbf{P}_c(X, Y, Z, 1)^T \quad (8)$$

平面射影行列 \mathbf{H}_c および空間射影行列 \mathbf{P}_c の自由度はそれぞれ8および11であり、各行列を推定するには、各々4点以上および6点以上の対応関係から、最小2乗法により導出可能である。

式(4)で算出される正面化行列 \mathbf{H}_c は単一カメラの画像情報のみをもとに推定されているのに対して、平面射影行列 \mathbf{H}_c および空間射影行列 \mathbf{P}_c は多視点の撮影カメラ画像間で整合性の確保された3次元モデル座標空間と各カメラ間の射影関係を表している。この結果、画像中の特徴領域の実空間中の3次元位置をより正確に推定することが可能となる。

3.3 背景領域の合成

モデル座標空間に、遠景領域を構築する場合、図5のように、フィールド領域を囲うように、4つの長方形ポリゴンを設定することが考えられる。各長方形ポリゴンを設定する座標は、長方形ごとに4点を複数のカメラ画像にマッピングし、各カメラ画像

におけるフィールド領域と遠景領域の境界が不適切にならないように調整する。

フィールド領域および遠景領域を細かな三角形ポリゴンに分割し、各カメラの空間射影行列をもとに、各頂点座標と各カメラ画像中の画素の対応を算出する。さらに、ポリゴンごとに、当該ポリゴンを捉えているカメラ画像の中からテクスチャ情報を取得しておく。仮想視点における射影行列 \mathbf{P}_i は、内部パラメータ \mathbf{A}_i 、モデル座標空間中の位置座標に基づく並進ベクトル \mathbf{t}_i 、さらにカメラの方位角および仰角に基づく回転行列 \mathbf{R}_i を用いて、式(9)により算出される。

$$\mathbf{P}_i = \mathbf{A}_i [\mathbf{R}_i | \mathbf{t}_i] \quad (9)$$

仮想視点の射影行列をもとに、各ポリゴンの投影領域を算出し、視点位置に応じて最適なテクスチャをカメラ画像から取得し、レンダリングを行う。

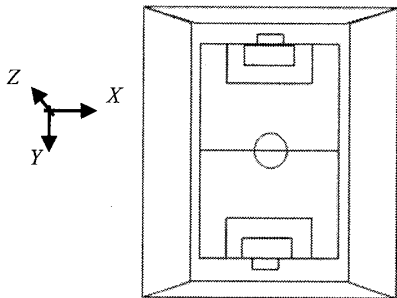


図5. 遠景領域を含めたモデル座標空間

3.4 被写体領域の合成

はじめに、各カメラの平面射影行列 \mathbf{H}_c をもとに、各被写体のモデル座標空間中フィールド平面上的位置推定を行う。各カメラ画像で背景差分を適用し、図6のように人物領域を抽出し、文献[7]と同様に、外接長方形領域を取得し、足下の画素をフィールド平面上に射影する。この際、各カメラの平面射影行列 \mathbf{H}_c の推定誤差の影響により、図7(1)のようにばらつきが生じるため、座標間の距離をもとにクラスタリングを行い、フィールド平面上的被写体位置を決定する。具体的には、フィールド平面上にマッピングされた N 個の点 $p_i(x_i, y_i, 0)$ ($i=0, \dots, N$) について、全ての点の組合せ (i, j) 間で式(10)によりユークリッド2乗距離を算出し、一定の閾値以下の点を同一人物の足下点であるとみなし、グループ分けを行う。

$$d_{i,j}^2 = (x_i - x_j)^2 + (y_i - y_j)^2 \quad (i, j = 1, 2, \dots, n) \quad (10)$$

さらに、各グループ内の点について重心を算出し、図7(2)のように、被写体のモデル座標空間中の位置を決定する。

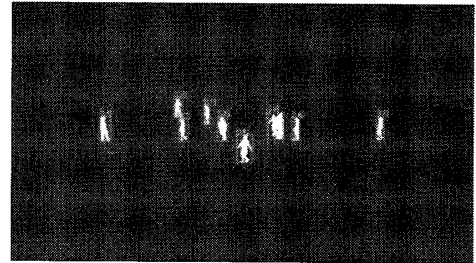
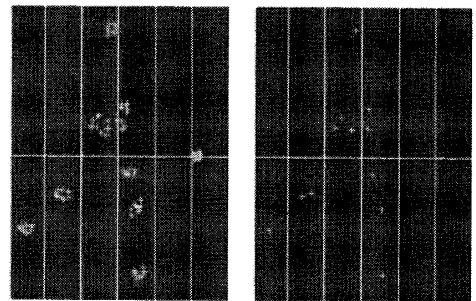


図6. 各視点での背景差分結果



(1) 各カメラからの射影点 (2) クラスタリング後
図7. モデル座標空間における被写体位置の特定

次に、モデル座標空間において、フィールド平面上的被写体位置を中心に適切な高さを持つ長方形ポリゴンを被写体領域として設定する。このとき、基本的には文献[7]と同様に、高さ2メートル、幅1メートルの固定サイズ長方形領域を設定するが、画像中でオクルージョンが生じていない被写体については、各カメラの外接長方形の大きさを考慮してポリゴンサイズを決定する。まず、各外接長方形の底辺中心および両端点の3点を、平面射影行列によりモデル座標空間に射影する。外接長方形の高さと幅の比率をもとに、モデル座標空間中で長方形ポリゴンを構築する。

さらに、モデル座標空間中の各被写体ポリゴンを、各カメラの空間射影行列により、正面化して投影することでテクスチャを取得する。背景領域の合成と同様に、仮想視点における空間射影行列を算出し、被写体領域の投影領域を算出し、撮影カメラ画像の

中から、最適なテクスチャを取得することでレンダリングを行う。

4. 実験結果

大空間を対象に提案手法を適用した場合の有効性を検証する目的で、モデル座標空間の構築精度検証ならびに、自由視点画像の合成精度検証に関する実験を行った。実験用データとして、12台のカメラを直径20メートル程度の円周上に配置してフットサルコートを撮影した多視点映像のある1フレームを対象に実験を行った。多視点画像のうちの隣接する2視点の撮影画像例を図8および図9に示す。



図8. 多視点画像の例 (カメラ1)



図9. 多視点画像の例 (カメラ2)

4.1 射影行列の推定精度検証

提案手法で導入した、画像中の幾何情報をもとにモデル座標空間を構築し、各カメラの射影行列を推定する手法について、推定精度を検証する実験を行った。12枚のカメラ画像から抽出された特徴点をもとにモデル座標空間を構築し、さらに座標空間中の特徴点を各カメラに射影した結果を図10および図11に示す。

本結果により、モデル座標空間中の特徴点群が各カメラに正しく射影されており、平面射影行列および空間射影行列が高精度に推定されていることが確認できた。



図10. 特徴点を射影した結果 (カメラ1)

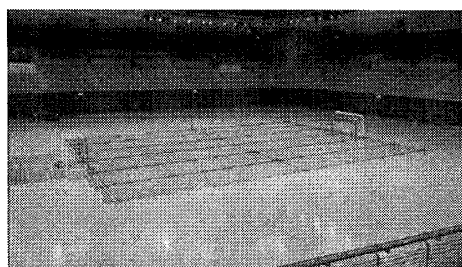
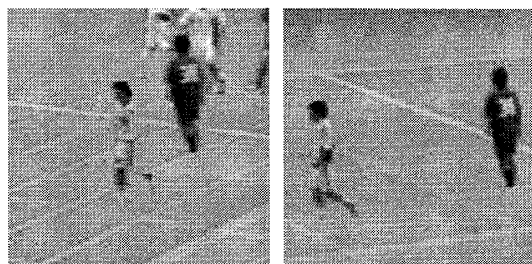


図11. 特徴点を射影した結果 (カメラ2)

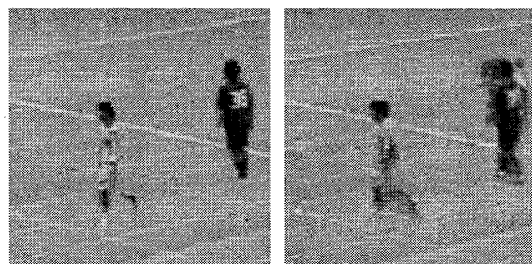
4.2 自由視点画像の合成精度比較

多視点カメラが強校正済であることを前提としない方式のうち、カメラ間の射影関係のみを用いた手法[3]を比較方式として選択し、カメラ間等の仮想視点における画像を合成し、精度を比較した。提案手法および比較手法を用いて、図8および図9の2カメラ間の中間視点画像を合成した。図8および図9におけるフィールド中の特定領域を拡大した図をまとめて図12に示す。さらに、合成結果の特定領域を拡大した画像を図13に示す。



(1) カメラ1 (2) カメラ2

図12. カメラ画像の特定領域拡大図



(1) 提案手法 (2) 従来手法

図13. 中間視点画像の特定領域拡大図

従来手法では右側の人物領域がぼやけているが、提案手法では忠実に合成できていることがわかる。従来手法では、オクルージョンの影響により対応点探索時に誤検出が生じた結果、精度が低下したと考えられる。一方、提案手法では、画像情報のみを用いて推定した射影行列の推定精度が比較的高く、人物位置を適切に推定できた結果、高精度に合成できたものと考えられる。

4.3 フィールド内の仮想視点画像合成

さらに、カメラ間だけでなく、フィールド内に存在する被写体の視点およびフィールド上空からの視点における合成結果を図14、図15に示す。従来手法では、カメラの内挿箇所のみでしか仮想視点の合成を行うことができない一方、提案手法では、3次元空間中の様々な視点における合成が可能であることが確認できた。

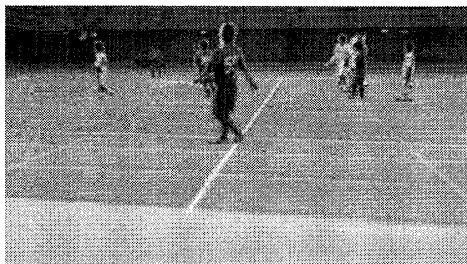


図14. フィールド内仮想視点の合成結果



図15. フィールド上方仮想視点の合成結果

5. まとめ

本稿では、大空間を対象とした自由視点映像の合成精度向上を目的に、カメラの強校正を前提としない画像ベース方式に基づく手法を提案した。従来の画像ベース方式では、3次元空間中の位置情報を考慮していないため、被写体が重なるオクルージョン領域での精度低下や、合成可能な仮想視点位置がカメラを内挿する箇所に限定されるという課題があった。本稿では、上記課題を解決するため、画像情報のみを用いてモデル座標空間を構築し、各カメラの

射影行列を推定することで、3次元空間中の位置情報を考慮した自由視点画像の合成手法を提案した。

多視点画像を対象とした実験により、従来の画像ベース方式に対する有効性を確認した。今後はズームカメラ等の動的カメラへの適用を目指し、画像中から抽出される特徴点が限られる場合にも、フレーム毎に射影行列を推定する手法を導入する。

参考文献

- [1] 石川彰夫, メヒルダド パナヒプル テヘラニ, 内藤 整, 酒澤茂之, 小池 淳: "ウォークスルーを実現するための自由視点映像合成方式", 信学論, vol.J92-D, no.6, pp.854-867 (2009).
- [2] T. Kanade, P. W. Rander, and P. J. Narayanan: "Virtualized Reality: Constructing virtual worlds from real scenes", IEEE Multimedia, vol. 4, no. 1, pp. 34-37 (1997).
- [3] N. Inamoto and H. Saito: "Virtual Viewpoint Replay for a Soccer Match by View Interpolation from Multiple Cameras", IEEE trans. Multimedia, vol.9, no.6, pp.1155-1166 (2007).
- [4] S. M. Seitz and C. R. Dyer, "View Morphing", in Proc. SIGGRAPH '96, pp. 21-30 (1996).
- [5] S. Pollard, M. Pilu, S. Hayes, and A. Lorusso, "View Interpolation for Image Synthesis", in Proc. SIGGRAPH '93 pp. 279-288 (1993).
- [6] S. Yaguchi and H. Saito, "Arbitrary Viewpoint Video Synthesis from Multiple Uncalibrated Cameras", IEEE Trans. Syst., Man, Cybern, B, Cybern., vol. 34, no. 1, pp. 430-439 (2004).
- [7] T. Koyama et al., "Live mixed-Reality 3D Video in Soccer Stadium", in Proc. ISMAR2003, pp.178-187 (2003).