

## 多視点映像による現実シーンの仮想化 —カメラ間の射影的關係の利用による中間視点映像生成—

斎藤 英雄<sup>†\*</sup>      木村 誠<sup>†</sup>      矢口 悟志<sup>†</sup>      稲本 奈穂<sup>†</sup>

<sup>†</sup> 慶應義塾大学理工学部情報工学科

\* 科学技術振興事業団, さきがけ研究21, 「情報と知」領域

多視点映像を用いて共通の実空間を記録し, 新しい映像として提示する研究が最近盛んに行われている. 本稿では, 多カメラ間の射影幾何学的關係を利用して, 現実シーンを仮想化するために筆者らのグループで行っている研究について述べる. ここで紹介する手法は, カメラ間の射影幾何学的關係を与える弱校正に基づくものである. 弱校正は簡単な手続きで行うことができるので, 多カメラシステムを利用する場合は特に有用となる. 本稿では, このような弱校正により得られるカメラ間の射影幾何学的關係を利用した, 1) 射影グリッド空間における自由視点画像生成法, 2) 3眼ステレオマッチングのための射影幾何学的關係の利用, 3) サッカー映像のための中間視点画像生成法, の3つの手法について述べる.

## Virtualizing Real Scene from Multiple View Images

- View Interpolation of Multiple Cameras Based on Projective Geometry Among Cameras -

Hideo Saito<sup>†\*</sup>, Makoto Kimura<sup>†</sup>, Satoshi Yaguchi<sup>†</sup>, Naho Inamoto<sup>†</sup>

<sup>†</sup> Department of Information and Computer Science, Keio University

\* "Information and Human Activity", PRESTO, JST

In this presentation, we will show methods for interpolation of viewpoint from multiple cameras based on projective geometry. Projective relationship of multiple cameras can be obtained from weak calibration information, that is easily be corrected from the multiple view images. Such projective geometry provides sufficient information to reconstruct the 3D shape of the object with scale and projective transformation ambiguity. Since such ambiguity does not affect to 2D correspondence relationship between the multiple images, we can generate new view point images from multiple cameras by the projective geometry. We will show following approach for interpolating the view point of multiple cameras : 1) View interpolation from 3D shape reconstruction with projective ambiguity in Projective Grid Space (PGS), 2) View interpolation from trinocular stereo based on projective geometry, 3) View interpolation of actual soccer scene taken with multiple cameras based on projective geometry.

### 1 はじめに

近年, 動きのあるイベントを仮想化するという目的で, 現実的に利用可能な数台から10数台程度のCCDカメラを用いて, そこから入力される画像列から任意視点画像を生成する研究 [3, 4, 5, 6, 8, 9, 10, 13, 14] が

盛んである. その応用例として, サッカーや, ボクシング, 相撲などスポーツのTV中継などに利用し, ユーザが自由に自分の見たい視点からの映像を表示させることのできるシステムなどが考えられ, 将来性の高いシステムとなることが予想される.

このような多数のカメラを利用したシステムで3次

元復元を行う場合、通常は、対象空間のユークリッド座標系とカメラ座標系の射影関係を推定（強校正）しておくことが必要であった。カメラを強校正するには、空間上の数点に対し、ワールド座標系における3次元座標と画像座標系における2次元座標を正確に測定し計算する方法[12]が一般的である。このように、空間中に正確に数点の校正用ポイントを設置して測定する必要があるため、特に多数のカメラの校正は非常に手間のかかるものであった。しかも、屋外の場合のように広い撮影空間を対象とする場合には、校正用ポイントを広範囲にわたって正確に設置することは非常に困難であった。

一方、各カメラに共通に見えるマーカ点から検出したカメラ間の対応関係から、カメラ間の射影幾何学的関係を基礎行列（F行列）として得られることが知られている。この基礎行列の推定に利用する対応関係は、3次元点の未知な幾つかの自然特徴点を検出することにより得られるため、強校正に必要な3次元座標の既知な人工的マーカを対象空間に配置する必要が無く、求める労力が格段に削減される。このような、カメラ間の射影的関係の推定による校正を弱校正と呼ぶ。

本稿では、多視点カメラシステムにおいて、強校正を使わずに弱校正のみを行うことにより得られるカメラ間の射影的關係を表すF行列を利用して、視点を補間[1, 11]した中間視点画像を生成するために、筆者らがを行っている3つの研究例について述べる。一つは、多数のカメラの中から2台のカメラを基底カメラとして利用し、対象空間に射影グリッド空間と名づけられたグリッド座標を設定し、この座標と各カメラ間の関係がF行列のみにより得られることを利用した仮想視点画像生成法である。次に、カメラを3台に限定した場合に、3つのカメラの画像が、それぞれ対象とする3次元空間からの平行投影と見なせるように、F行列を利用した平行化を行うことによるステレオ法を紹介する。そして、3つめの手法として、サッカーシーンを対象に撮影された多視点ビデオ画像から、カメラ間のF行列を利用して中間視点画像を生成する手法を紹介する。

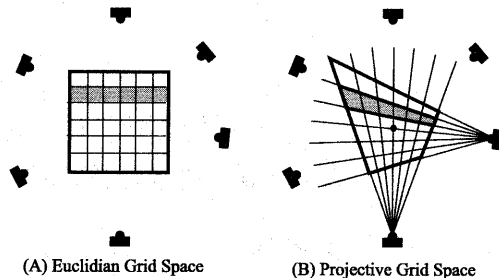


図 1: ユークリッド空間と射影グリッド空間

## 2 射影グリッド空間の利用

筆者らは、この多数のカメラからの射影幾何学に基づいた3次元復元を行うために、カメラ間のエピポラ幾何から“射影グリッド空間”を構成し、この射影グリッド空間において3次元復元する枠組を提案した[8]。

一般に、多数のカメラからの3次元復元のためには、対象を再構成しようとする3次元空間の各点と、多数のカメラ各々についての画像上に投影される位置との関係が必要となる。そこで、通常は強校正を行い、各カメラ毎に3次元空間とカメラの画像座標とを関連づける。これに対し、我々の提案した“射影グリッド空間”では、このグリッド空間と画像上の点との関係を、カメラ間のエピポラ幾何を表すFundamental行列のみを用いて記述することができるため、強校正を行うことなしに、多数のカメラからの3次元復元を行うことが可能になる。

射影グリッド空間は次のように定義される。数台あるカメラシステムのうちの任意の2台を基底カメラ1, 2とする。この2台のカメラそれぞれの視点からの中心射影によって3次元空間を定義する。つまり、空間を定義する3軸として、基底カメラ1の画像のX軸, Y軸, そして基底カメラ2の画像のX軸を用いるわけである。そして、これらの3軸を各 $P, Q, R$ 軸として、射影グリッド空間を定めるのである。

このように、任意の対象シーンに対して、カメラ配置とは無関係に通常の座標系（ユークリッド座標系）を考えるのではなく、任意に配置された多視点カメラのうち2台を用いて、その対象シーン内の3次元位置を示すための射影グリッド空間を定義することにより、

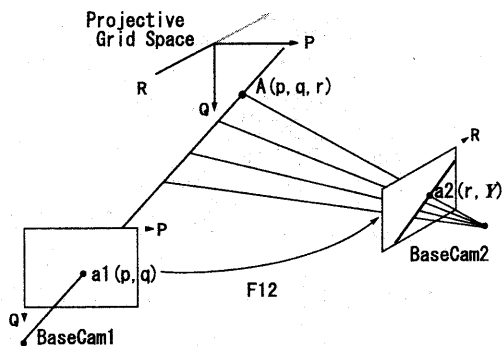


図 2: 射影グリッド点の基底カメラ画像への投影

任意カメラ配置の多視点カメラシステムでも対象の3次元形状復元が可能になるわけである(図1)。そこで、例えば、射影グリッド空間内における各グリッド点を、各カメラ画像に投影し、各シルエット画像において内部か外部かを判定することにより、各シルエットの共通体積として対象物体形状の3次元復元が行えることになる[14]。

射影グリッド空間内の点  $A(p, q, r)$  と各カメラの関係は以下になる。基底カメラ1, 2から得られる各画像を画像1, 画像2, 基底カメラ以外のカメラから得られる画像を画像  $i$  とする。またそれぞれの画像上の  $A$  の投影点を  $a_1, a_2, a_i$ , 画像  $h$  の画像  $k$  に対する  $F$  行列を  $F_{hk}$  と表すものとする。このとき、点  $A(p, q, r)$  は射影グリッド空間の定義より、画像1では  $a_1(p, q)$  に投影される(図2)。また画像2の画像1に対する  $F$  行列  $F_{21}$  を用いて  $a_1$  を画像2に直線  $l$  として投影すると直線  $l$  は(1)式のように表される。

$$l = F_{21} \begin{bmatrix} p \\ q \\ 1 \end{bmatrix} \quad (1)$$

射影グリッド空間の定義より点  $A(p, q, r)$  の画像2における投影点  $a_2$  の  $x$  座標は  $r$  であるから、 $a_2$  はこの直線  $l$  上の  $x$  座標が  $r$  の点として定めることができる。

また、点  $A(p, q, r)$  の画像  $i$  への投影点  $a_i$  を決定するには、まず  $F_{i1}$  を用いて  $a_1$  を画像  $i$  に直線  $l_1$  として投影する。さらに  $a_2$  を  $F_{i2}$  で画像  $i$  上に直線  $l_2$  として投影する。そして、これらの2直線  $l_1$  と  $l_2$  の交点が画像  $i$  の投影点  $a_i$  となる(図3)。

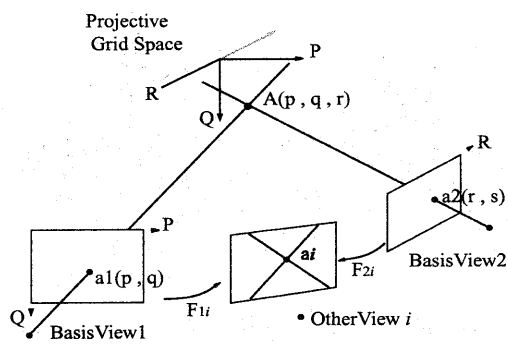


図 3: 射影グリッド点の一般のカメラ画像への投影法

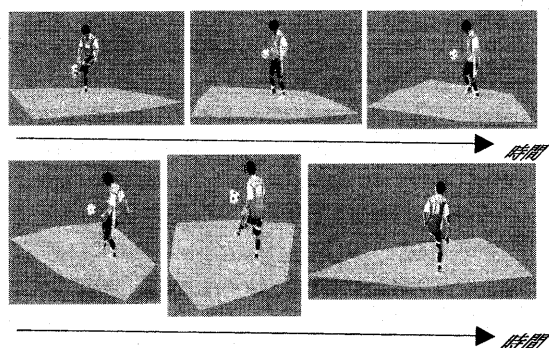


図 4: 未校正カメラ画像からの自由視点映像生成例

このように射影グリッド空間では、多数のカメラの中から2つの基底カメラを選び、これらの基底カメラ間での  $F$  行列により、3次元のグリッド位置を定義する。そして、このグリッド位置と基底カメラ以外のカメラの画像位置との関係は、基底カメラとの  $F$  行列により記述される。このように、射影グリッド空間におけるシルエット法による3次元再構成では、各カメラと基底カメラとの  $F$  行列から、射影グリッド空間の任意の点が各カメラに投影される2次元位置の関係を求め、この位置がシルエットの内部か外部かの判定を行うことによって、多数のカメラからの3次元再構成が行われる。この3次元再構成には、全カメラ数を  $N$  とすると、2つの基底カメラ間の  $F$  行列、それぞれの基底カメラと、それ以外の各カメラ間の  $F$  行列の合計  $1 + (N - 2) \times 2$  組の  $F$  行列が必要とな

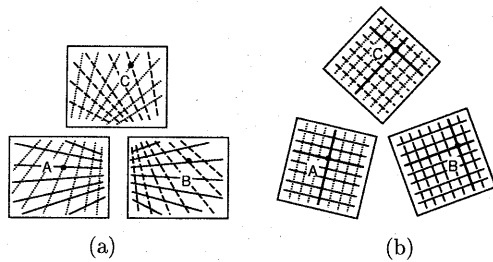


図 5: 3眼カメラにおける画像の平行化 (a) 原画像のエピポーラ線 (b) 3カメラ間でお互いに平行化されたエピポーラ線

るわけである。

図 4 にその一例を示す。この画像は、9 台のカメラを用いて撮影された多視点画像から上記のようにして射影グリッド空間を構築して 3 次元モデルを復元し、この 3 次元モデルから各カメラで撮影された原画像間での画素毎の対応関係を求め、これに基づいて原画像間の中間視点画像を生成することにより合成したものである。また、このとき、射影グリッド空間において復元された 3 次元モデルに、仮想的な芝生のテクスチャを持つ床を合成している。

### 3 3 眼ステレオにおける射影幾何の利用

前章で述べた射影グリッド空間は、多数のカメラ間の射影幾何的關係を利用して、2 台の基底カメラにより共通の 3 次元座標系を定義することに基づいた手法であった。これに対して、カメラが 3 台に限定されるときは、それぞれのカメラ間の射影幾何的關係を利用して、3 眼ステレオに適した座標系を定義することができる。

3 台のカメラの場合は、3 組の F 行列として各カメラの射影幾何的關係を弱校正により推定することができる。このとき、図 5-(a) に示すようにエピポーラ線が描けるとする。これに対して、図 5-(b) に示すようにエピポーラ線が描けるように、3 カメラ間でお互いに平行化することができる。このように平行化すると、3 つの画像の關係は、図 6 に示すように、共通の 3 次元座標空間を通して關係づけることができる。つまり、

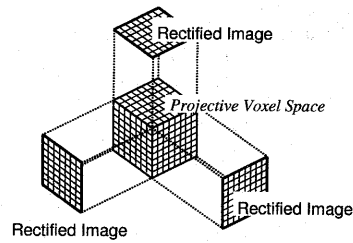


図 6: 3 眼ステレオ画像の統一的に扱う 3 次元空間

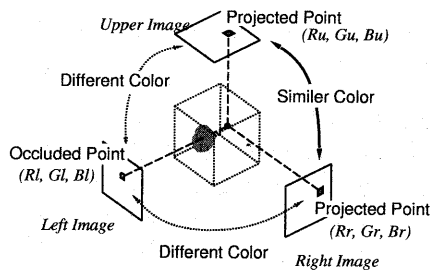


図 7: 3 眼ステレオ画像を統一的に扱う 3 次元空間における遮蔽關係の検出。遮蔽による対応画素間の画素値に違いがあっても、別の組で画素値が一致していることを発見することが容易となっている。

各画像座標は、共通の 3 次元座標系からの 3 組の平行射影により決定される座標系となる。このような共通の座標系で扱えるように平行化することによって、3 つのカメラの対応点關係を統一的に考えることが容易になる。この利点を利用して、3 眼ステレオにおいて、遮蔽關係を判定しながら、対応点を決定することにより、より妥当な任意視点画像を生成することができる。例えば、図 7 に示すように、3 つのカメラ間の關係を統一的に扱えるため、遮蔽による対応画素間の画素値の不一致があっても、別の組で画素値が一致していることを発見することが容易になるために、より安定に遮蔽關係を検出しながら対応点を求めることが可能になる [7]。

図 9 に、図 8 に示した 3 眼ステレオカメラにより撮影された画像から生成した中間視点画像を示す。ここで説明した手法を利用して対応点の検出を行っているため、遮蔽關係を容易に検出することができており、その結果得られる中間視点画像も遮蔽關係を良好に反

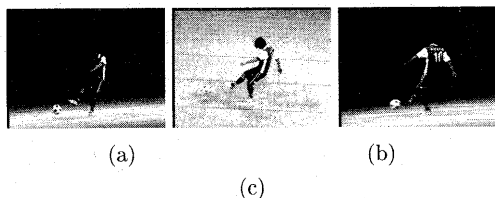


図 8: 3 眼ステレオ入力画像の一例 (a) left image; (b) upper image; (c) right image

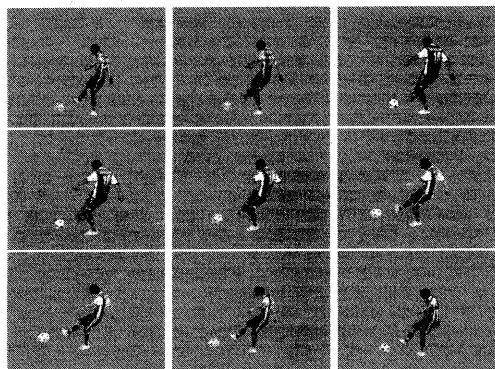


図 9: 3 眼のステレオマッチングに基づいた中間視点画像の生成例。入力 of 3 画像間のベースラインが広い  
ため、かなりの範囲で遮蔽が生じているが、良好に中間視点画像が合成されている。

映したものとなっている様子がわかる。

## 4 サッカー映像の中間視点映像生成

サッカーシーンを対象として、多数のカメラで撮影した映像をもとに、仮想視点を撮影視点間に置き、そこでの見え方を提示する手法 [2] を紹介する。本手法では図 10 に示すように、まず、全体を時間によって形状や位置の変化する移動領域とそうでない静止領域に分け、更に後者を遠景領域とそれ以外の領域に分ける。本稿ではサッカーシーンを対象としているので、移動領域には選手とボールが相当し、遠景領域には観客席などの背景部分が、それ以外の領域にはグラウンドとゴールが相当する。このように 3 領域に分割して処理を行う。そして、各領域の特徴を活かした手法を適用

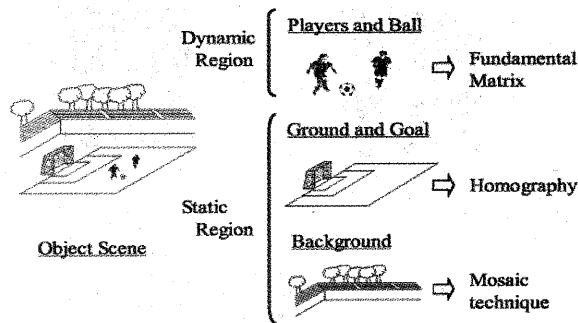


図 10: 対象シーン

することで、精度の良い中間画像の生成を目指す。選手とボールに関してはカメラ間のエピポーラ幾何、グラウンドとゴールに関しては平面射影行列を用いて、異なる視点で得られた画像間の対応点を算出し、補間画像を生成する。一方、遠景に関してはモザイク処理を行い、パノラマ画像から対応する仮想視点での背景を切り出す。最後にそれらを合成して対象全体の中間画像を得る。更に連続フレームに対してユーザーが自由に視点を与えることによって中間映像を生成する。

下記に、それぞれの領域に対する処理を説明する。

**選手及びボール** まず、図 11 に示すように、背景差分によって選手及びボールのシルエットを検出し、検出された複数の選手、ボールのシルエット領域を、異なるカメラ間で対応付けを行う。この対応付けには、選手の足元がグラウンドに接しているという前提を利用して、シルエット領域の最下部の点の座標が、異なるカメラ間でグラウンド面の平面射影行列により関係付

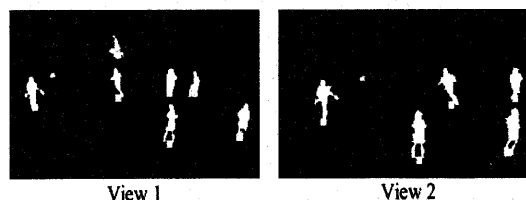


図 11: 検出された選手およびボールのシルエット。選手については、地面に足が接しているという前提を利用して、地面の平面射影行列により与えられる射影関係から別のカメラで撮影されたシルエットとの対応を自動判定する。

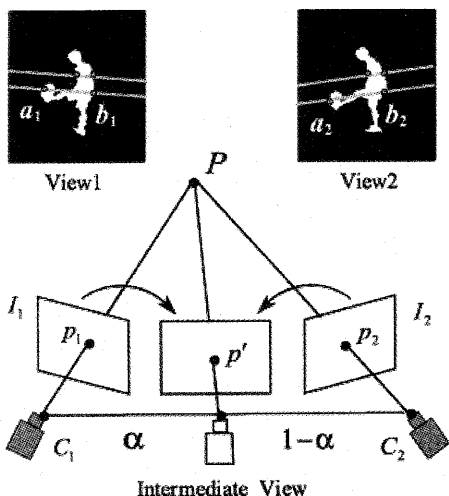


図 12: 中間視点画像における位置の算出

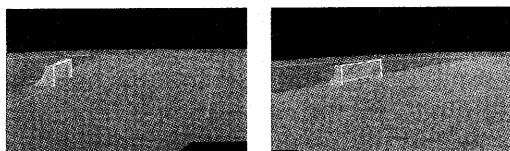


図 13: グラウンドとゴールに対する中間視点画像

けられることを用いている。

次に、異なるカメラ間で対応付けられたシルエット領域内部について、密な対応点関係を求める。これには、図 12 に示すように異なるカメラ間でエビポーラ線を引き、各エビポーラ線とシルエット領域の両端との交点を、 $a_1$  と  $a_2$ 、 $b_1$  と  $b_2$ 、というようにして対応付けを行う。続いてシルエット内部に関して線形補間によって画素毎の対応付けを行う。シルエットの上部から下部にかけて、エビポーラ線を順に投影していき、その間隔を狭くすることで密な対応点情報を得ることができる。これは、選手やボール自身の凹凸は、視点から対象となる選手やボールまでの距離に比べて十分に小さいため、領域内部での視差変化が一定であることを前提としている。

このようにして対応付けされた各点の位置関係から、次の式に示すようにして、その点の中間視点画像における位置を算出する。

$$\hat{p} = (1 - \alpha)p_1 + \alpha p_2 \quad (2)$$



図 14: 遠景に対するパノラマ画像

ここでは、 $p_1$  と  $p_2$  は、それぞれ、原画像  $I_1$  と  $I_2$  における対応点の位置であり、 $\alpha$  は 2 つの原画像の視点位置に対する中間視点位置を、それらの内挿比率で表したものであり、0 の場合はカメラ 1 と同一視点、1 の場合はカメラ 2 と同一視点を表す。この式によって決定される新しい位置に原画像の画素を移動させて画像を合成することにより、各原画像に対するワーブ画像が生成される。各ワーブ画像をさらにブレンド合成することによって、中間視点画像における選手やボールの画像を合成することができる。

**グラウンドとゴール** グラウンドとゴールは平面と見なし、この平面に対する平面射影行列を利用して、異なる視点間での対応関係を求める。つまり、次の式によって、画像間の対応関係を決定する。

$$p_2 \cong H p_1 \quad (3)$$

ここで、 $H$  は対象とする平面の平面射影行列を示し、 $p_1$  と  $p_2$  は、それぞれ原画像  $I_1$ 、 $I_2$  における同次座標である。この平面射影行列が、グラウンドとゴールにおける画素毎の対応関係を与えるので、上記の選手とボールの場合と同様にして、中間視点における画像を合成することができる。図 13 に、グラウンドとゴールに対して生成された中間視点画像の例を示す。ここでは、2 枚の原画像の視点の中央に仮想視点を設定した中間視点画像となっている。

**遠景** スタンドのような遠景領域については、これを無限遠に存在する平面とみなし、隣接するカメラ間の遠景領域についてモザイクングによりパノラマ画像を生成する。そして、このパノラマ画像から、合成する中間視点画像に対応する領域を切り出し、合成するものとする。具体的には、隣接するカメラ 1 と 2 間で、それらの遠景領域を平面とみなし、この平面射影行列  $H_{21}$  を決定する。そして、この平面射影行列を用いることにより、2 カメラ間で遠景領域が連結される。そ

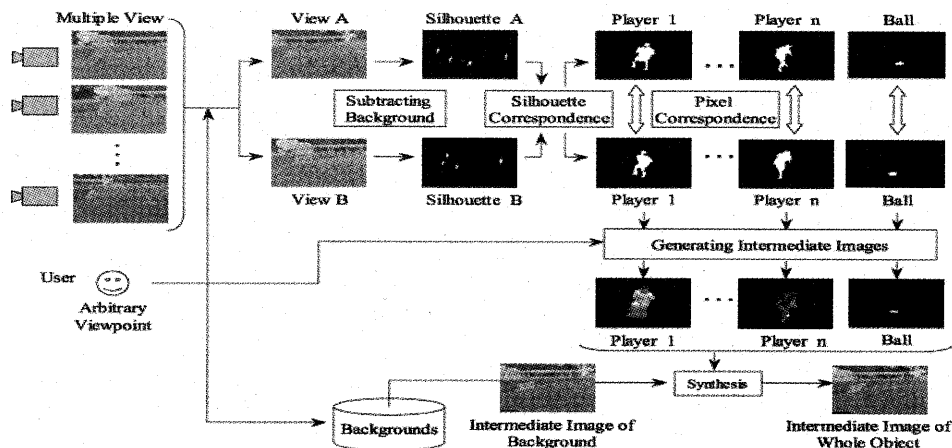


図 16: サッカーシーンの中間映像生成のための流れ図

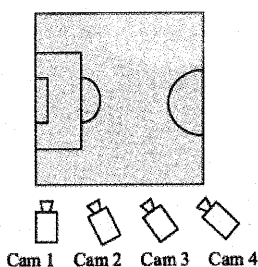


図 15: サッカーシーンの撮影のための多視点カメラの配置

して、中間視点画像の視点を指定するための内挿比率パラメータ  $\alpha$  に応じて、この中間視点画像に合成する遠景領域を切り出すために、次式で表される平面射影行列を利用する。

$$\hat{H} = (1 - \alpha)E + \alpha H_{21}^{-1} \quad (4)$$

ただし、 $E$  は単位ベクトルである。図 14 は、遠景領域に対してモザイクングにより合成したパノラマ画像の一例を示す。

図 16 に、本手法の全体の流れをまとめる。

2000 年秋の関東大学サッカーリーグ戦の数試合について、数台のカメラを用いて実際にスタジアムで撮影した多視点ビデオ画像を用いて、本手法の有効性を確認する実験を行った。図 15 にカメラ配置の概略を示す。ここに示すように、ゴール前のシーンを 4 つのカメラで撮影した多視点画像を入力として、それらの中

間視点画像を合成した。各入力カメラ画像は、720 × 480 画素のカラー画像である。

このような通常のスタジアムでの撮影では、3 次元位置が既知の点を用いた強校正を行うことは殆ど不可能であり、画像間の対応点を自然特徴点を用いて検出することによる弱校正しか事実上行うことができない。この実験では、20 点程度の自然特徴点を手入力により検出することにより各カメラ間の F 行列を推定した。図 17 に、合成した中間視点画像の一例を示す。(a) と (e) がカメラ 1 とカメラ 2 で撮影された原画像、(b)(c)(d) はカメラ 1 とカメラ 2 により合成した中間視点画像、(f)(g)(h) はカメラ 2 とカメラ 3 (原画像は示していない) から合成した中間視点画像である。さらに、内挿比率が 0.5 の場合とカメラ 2 の原画像 (e) の拡大を示す。原画像と比べても、画質劣化が目立たない良好な中間視点画像が合成できていることがわかる。

## 5 おわりに

本稿では、多視点カメラシステムにおいて、弱校正のみを行うことにより得られる射影幾何学的関係を利用してカメラ視点を補間した中間視点画像を生成するための手法として、1) 射影グリッド空間における自由視点画像生成法、2) 3 眼ステレオマッチングのための射影幾何学的関係の利用、3) サッカー映像のための中間視点画像生成法、の 3 つを紹介した。

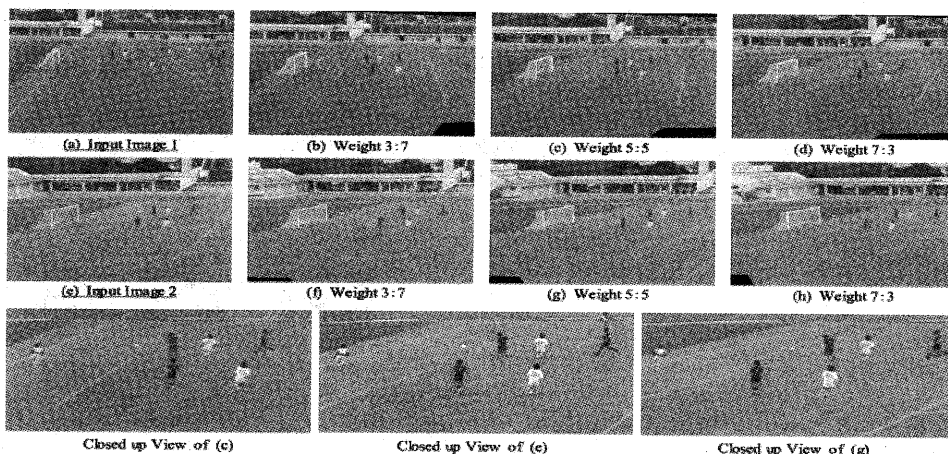


図 17: Generated intermediate views

## 参考文献

- [1] T.Beier, S.Neely : "Feature-Based Image Metamorphosis", Proc. of SIGGRAPH '92, pp.35-42, 1992.
- [2] 稲本奈穂, 斎藤英雄, "サッカーシーンにおける多視点映像からの中間映像生成", 電子情報通信学会PRMU研究会, PRMU2001-151, 2001年11月
- [3] T.Kanade, P.W.Rander, S.Vedula, H.Saito : "Virtualized Reality:Digitizing a 3D Time-Varying Event As Is and in Real Time", International Symposium on Mixed Reality(ISMR99), pp41-57, Yokohama, Japan, March 1999.
- [4] 北原格, 大田友一, 斎藤英雄, 秋道慎志, 尾野徹, 金出武雄, "大規模空間における多視点映像の撮影と自由視点映像生成", 3次元画像工学コンファレンス, July 2000.
- [5] 北原格, 大田友一, 金出武雄, "多視点映像の融合によるスポーツシーンの自由視点映像生成", 画像の認識・理解シンポジウム(MIRU'2000)論文集, vol.I, pp.I3-I8, July 2000.
- [6] I.Kitahara, H.Saito, Y.Ohta, S.Akimichi, T.Ono, T.Kanade, "Large Scale Virtualized Reality", Technical Scheches, IEEE CVPR2001, 2001.
- [7] M.Kimura, H.Saito, "3D Reconstruction Based on Epipolar Geometry", Trans. IEICE on Information and Systems, Vol.E84-D, No.12, pp.1690-1697, 2001.
- [8] H.Saito,T.Kanade : "Shape Reconstruction in Projective Grid Space from Large Number of Images", IEEE Proc. Computer Vision and Pattern Recognition, Vol.2, pp.49-54, 1999.
- [9] H.Saito, S.Baba, M.Kimura, S.Vedula, T.Kanade : "Apperance-Based Virtual View Generation of Temporally-Varying Events from Multi-Camera Images in 3D Room", Second International Conference on 3-D Digital Imaging and Modeling (3DIM99), pp.516-525, 1999.
- [10] 斎藤英雄, 金出武雄 : "多数のカメラによるダイナミックイベントの仮想化". 情報処理学会コンピュータビジョンとイメージメディア研究会, 119-13, Vol.99, No.93, pp.117-124, 1999. 11
- [11] S.M.Seitz, and C.R.Dyer : "View Morphing",proc. of SIGGRAPH '96, pp.21-30, 1996.
- [12] R.Tsai : "A Versatile Camera Caribration Technique for High- Accuracy 3D Machine Vision Metrology Using Off-the-Shelf Tv Cameras and Lenses", IEEE Journal of Robotics and Auto mation RA-3,4, pp323-344, 1987.
- [13] S.Vedula, P.W.Rander, H.Saito, T.Kanade : "Modeling, Combining, and Rendering Dynamic Real-World Events From Image Sequences", Proc. 4th Conf. Virtual Systems and Multimedia, Vol.1, pp.326-322, 1998.
- [14] 矢口悟志, 木村誠, 斎藤英雄, 金出武雄, "未校正多視点カメラシステムを用いた任意視点画像生成", 情報処理学会コンピュータビジョンとイメージメディア研究会論文誌, Vol.42, No.SIG6(CVIM2), pp.9-21, 2001年6月