

射影幾何に基づく多視点カメラの中間視点映像生成

齋藤 英雄^{†,††} 木村 誠[†]
矢口 悟志[†] 稲本 奈穂[†]

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さかんに行われている。この際、多視点カメラのカメラパラメータを推定するには、対象空間に位置が既知なマーカを複数配置する必要がある、その実施には多大な手間を要する。本論文では、多視点画像間の自然特徴点の対応関係のみから容易に得ることのできる射影幾何学的関係を利用して、多視点画像から中間視点画像を生成する手法について述べる。ここで紹介する手法は、カメラ間の射影幾何学的関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。本論文では、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関係を利用した、1) 射影グリッド空間における視体積交差法による射影的形状復元に基づいた中間視点画像生成法、2) サッカーシーンの持つ射影幾何学的性質を利用した中間視点画像生成法、の2つの手法について述べる。

View Interpolation of Multiple Cameras Based on Projective Geometry

HIDEO SAITO,^{†,††} MAKOTO KIMURA,[†] SATOSHI YAGUCHI[†]
and NAHO INAMOTO[†]

In this paper, 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, which can easily be collected 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), geometry, 2) View interpolation of actual soccer scene taken with multiple cameras based on projective geometry in soccer scene.

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^{1),6)~12),15)~17),20)~23)}がさかんである。その応用例として、サッカーや、ボクシ

ング、相撲等スポーツのTV中継等に利用し、ユーザが自由に自分の見たい仮想視点からの映像を表示させることのできるシステム等が考えられ、将来性の高いシステムとなることが予想される。たとえば、スポーツを撮影した多視点カメラ映像の切替えにより、あたかもカメラが動いているような効果の映像を表示するシステム²⁴⁾はすでに実用化されている。

本論文では、多視点カメラを用いて撮影した対象シーンを任意視点から観察可能なシステムを目的として、実カメラ視点の中間の視点における画像を生成する手法について述べる。異なる視点の画像からそれらの中間の見え方を合成するための手法として、モーフィング²⁾を利用した研究^{3),18)}がある。このような手法においては、異なる視点間で対応点を与える必要があるが、異なる視点の画像間で対応点を自動的に決定することは容易ではない。対応点は、対象シーンの

† 慶應義塾大学理工学部情報工学科

Department of Information and Computer Science, Keio University

†† 科学技術振興事業団、さきがけ研究 21「情報と知」領域

“Information and Human Activity,” PREST, JST

現在、独立行政法人産業技術総合研究所デジタルヒューマン研究ラボ、科学技術振興事業団 CREST

Presently with The National Institute of Advanced Industrial Science and Technology, and CREST, JST.

現在、NTTコムウェア株式会社

Presently with NTT Comware Corp.

3次元構造情報が与えられれば決定できる。

多数のカメラを利用したシステムで3次元形状復元を行う場合、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)する必要がある。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¹⁹⁾が一般的である。しかし、空間中に数点の校正用ポイントを正確に設置して測定することは容易ではなく、特に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となる。

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次元点の未知ないく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なく、求める労力を格段に削減することができる。このような、カメラ間の射影幾何学的関係の推定による校正を弱校正と呼ぶ。

本論文では、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、多視点画像に撮影された対象シーンの3次元幾何情報を射影的に推定することによって、多視点画像間の画素ごとに対応関係を求め、この対応関係を用いて、実カメラ視点の中間の視点における画像を生成するための2つの手法について述べる。

まず、弱校正された多数のカメラの中から選択された2台の基底カメラを用いて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、射影グリッド空間内で対象の3次元形状を復元する手法について述べる。この手法では、基底カメラにより決定される射影グリッド空間のグリッド座標と各カメラ座標との関係が、カメラ間の F 行列のみにより与えられることを利用して、強校正されていない多視点画像から射影グリッド空間内で形状復元を行うものである。多視点画像からの3次元形状復元には、視体積交差法を利用する。これは、各視点の画像から対象物体のシルエットを抽出し、このシルエットにより決定される視体積を、共通の3次元空間で交差させ、その共通部分であるVisual Hull⁵⁾を対象形状として復元するものである。

このような視体積交差法を利用した3次元形状復元

は、対象シーンを全周囲から取り囲むようなカメラ配置で撮影されている必要がある。これに対し、全周囲にカメラが存在せず、ある特定の方向にのみ数台のカメラが存在する場合もある。たとえば、サッカーに対して、メインスタンドにのみカメラが数台設置されるような場合である。このような場合、視体積交差法等を利用して、サッカーシーンの3次元形状を復元することは困難である。しかし、3次元形状を復元しなくても、サッカーシーンの持つ幾何学的性質を利用すれば、中間視点画像を生成するために必要な視点間の対応点を得ることは可能である。本論文では、サッカーシーンを複数の視点で撮影した多視点画像から中間視点画像を生成するために、対象空間を移動領域(選手とボール)、静止領域(グラウンド、サッカーゴール、遠景)に分割し、各領域に関する隣接するカメラ間での射影幾何学的関係を利用して中間視点画像を生成する手法について述べる。

本論文では、これらの2つの手法の原理と、それぞれの手法に基づいて生成された中間視点画像を示すことにより有効性を検証する。

2. 射影グリッド空間での3次元復元にに基づく中間視点画像生成

2.1 射影グリッド空間

筆者らは、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、カメラ間のエピポーラ幾何から“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¹⁵⁾。

一般に、多数のカメラからの3次元復元には、対象を再構成しようとする3次元空間の各点と、その点が各カメラの画像面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、各カメラごとに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、このグリッド空間内の各点と、それが画像に投影される位置との関係を、カメラ間のエピポーラ幾何を表す F 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。図1に示すように、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

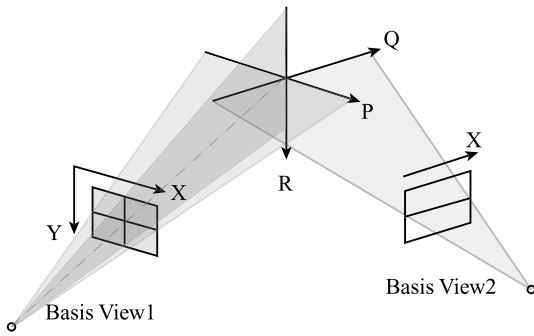


図 1 射影グリッド空間の定義
Fig. 1 Definition of Projective Grid Space.

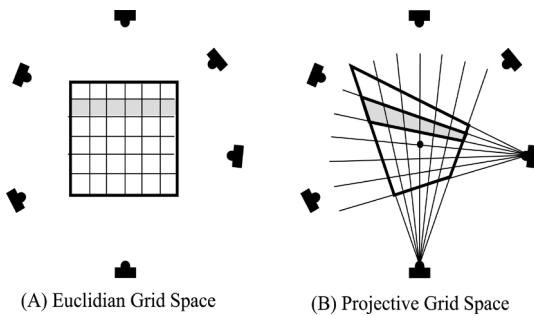


図 2 ユークリッド空間と射影グリッド空間
Fig. 2 Euclid Grid Space and Projective Grid Space.

3 軸を各 P, Q, R 軸として, 射影グリッド空間を定める.

このように, 任意の対象シーンに対して, 図 2 (a) のように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, 図 2 (b) のように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れば, 任意配置の多視点カメラ間の関係を示す F 行列によって, 各カメラと対象の 3 次元空間を関係付けることができる.

2.2 射影グリッド空間における 3 次元形状復元

上記の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を求め (視体積交差法), 対象物体形状の 3 次元復元を行うことができる²³⁾.

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ようになる. 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, 図 3 に

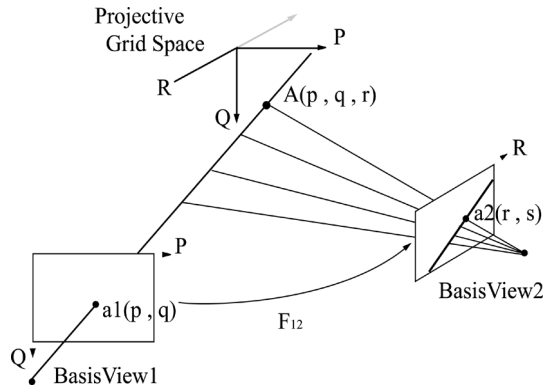


図 3 射影グリッド点の基底カメラ画像への投影
Fig. 3 Projection of a point in Projective Grid Space onto images of base cameras.

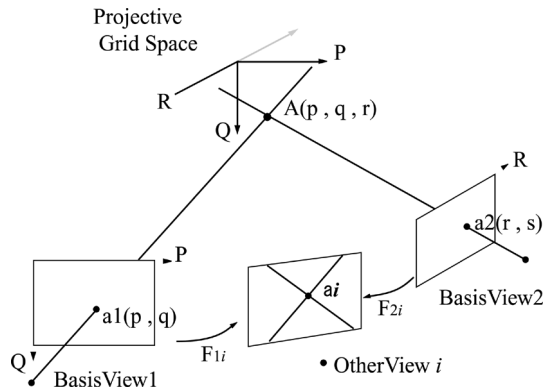


図 4 射影グリッド点の一般のカメラ画像への投影法
Fig. 4 Projection of a point in Projective Grid Space onto images of non-base cameras.

示すように画像 1 では $a_1(p, q)$ に投影される. 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 は式 (1) 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 $A(p, q, r)$ を基底カメラ画像以外の画像 i へ投影した点 a_i を決定するには, 図 4 に示すように, まず F_{1i} 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.

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視体積交差法による 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。

なお、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 行列が必要となる。

2.3 仮想視点画像の作成

以上のようにして射影グリッド空間において復元された 3 次元モデルから仮想視点画像を生成する。射影グリッド空間で復元された 3 次元モデルは、射影変換の曖昧性を含んだ 3 次元形状であり、真の形状を表していないが、実カメラ画像間の画素ごとの対応関係や 3 次元モデル上の各点の実カメラに対する前後関係・遮蔽関係は求めることができる。そこで、射影グリッド空間において復元された 3 次元モデルの各点が、実カメラに見えるか見えないかの可視判定を行い、その関係から、入力画像間の画素ごとの対応関係を求めることができる。画素ごとの対応関係が与えられれば、モーフィング手法²⁾を視点内挿³⁾に利用することにより、入力画像間の中間視点画像を生成することができる。視点内挿では、図 5 に示すように、隣接する 3 つの実カメラ視点位置を任意の重みで平均した位置が仮想視点位置になるものと仮定し、実カメラ画像間の対応する画素位置を同じ重みで平均した位置にその画素を移動することにより、中間視点画像を合成する。

具体的な手続きは次のようになる。仮想視点を内挿合成するために利用する 3 台の実カメラ画像を参照画像とし、それらを内挿する比率を w_1, w_2, w_3 で表すものとする。このとき、射影グリッド空間における点の位置を A 、その参照画像上での投影点を P_1, P_2, P_3 とし、その点の画素値を $I_1(P_1), I_2(P_2), I_3(P_3)$ とする。また、復元された 3 次元形状から判定された、 A が各参照画像に見える (1) か見えない (0) かを、 v_1, v_2, v_3 で表すものとする。視点内挿では、参照カメラ画像間の対応する画素位置 P_1, P_2, P_3 を内挿比率 w_1, w_2, w_3 で平均した位置に、その対応している画素を移動する。したがって、内挿され

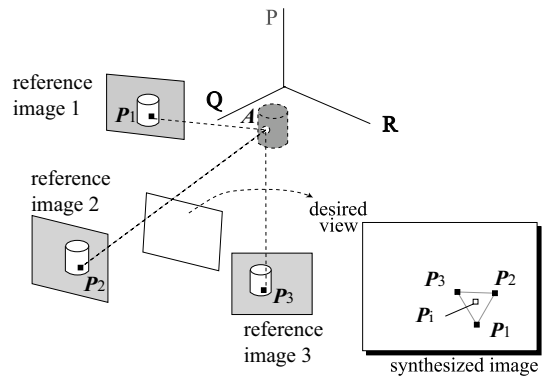


図 5 仮想視点の位置と仮想視点の画像合成
Fig. 5 Position of a virtual viewpoint, where an image is synthesized.

た仮想視点における A の投影点の位置 P_i は、

$$P_i = \frac{w_1 P_1 + w_2 P_2 + w_3 P_3}{w_1 + w_2 + w_3} \quad (2)$$

のように与えられることになる。一方、この位置における画素値 $I(P_i)$ は、参照カメラで対応している画素の画素値 $I_1(P_1), I_2(P_2), I_3(P_3)$ を、同様に内挿比率 w_1, w_2, w_3 で平均した値となるが、復元された 3 次元形状から対応する点が見えないと判定された参照カメラの画素値を除いた平均値とするので、

$$I_1(P_i) = \frac{v_1 w_1 I_1(P_1) + v_2 w_2 I_2(P_2) + v_3 w_3 I_3(P_3)}{v_1 w_1 + v_2 w_2 + v_3 w_3} \quad (3)$$

のように求められる。ただし、 A が 3 つの参照画像のすべてに見えない場合、つまり $v_1 = 0, v_2 = 0, v_3 = 0$ の場合、この視点では画素値は求めない。これを 3 次元表面モデル表面上のすべての点に対して行うことにより、中間視点画像を得る。

図 6 に、対象空間を取り囲むように配置された 18 台のカメラを用いて撮影された多視点画像列のあるフレームについて、提案手法により中間視点画像を生成した例を示す。この図の 3 角形の頂点が合成に用いた実カメラ画像である。また各辺上の画像はそれぞれ 2 枚の画像の中間として合成した画像である。中央の画像はウェイトを 0.33, 0.33, 0.34 とした場合の 3 枚の画像から合成された画像である。ここでは、射影グリッド空間を定義した 2 台の基底カメラに撮影された床平面を用いて、床平面モデルを射影グリッド空間内で決定し、これに芝のテクスチャを与えることにより、仮想的な床を合成している。

2.4 複数対象空間の統合

たとえば、サッカーフィールド等のように非常に広

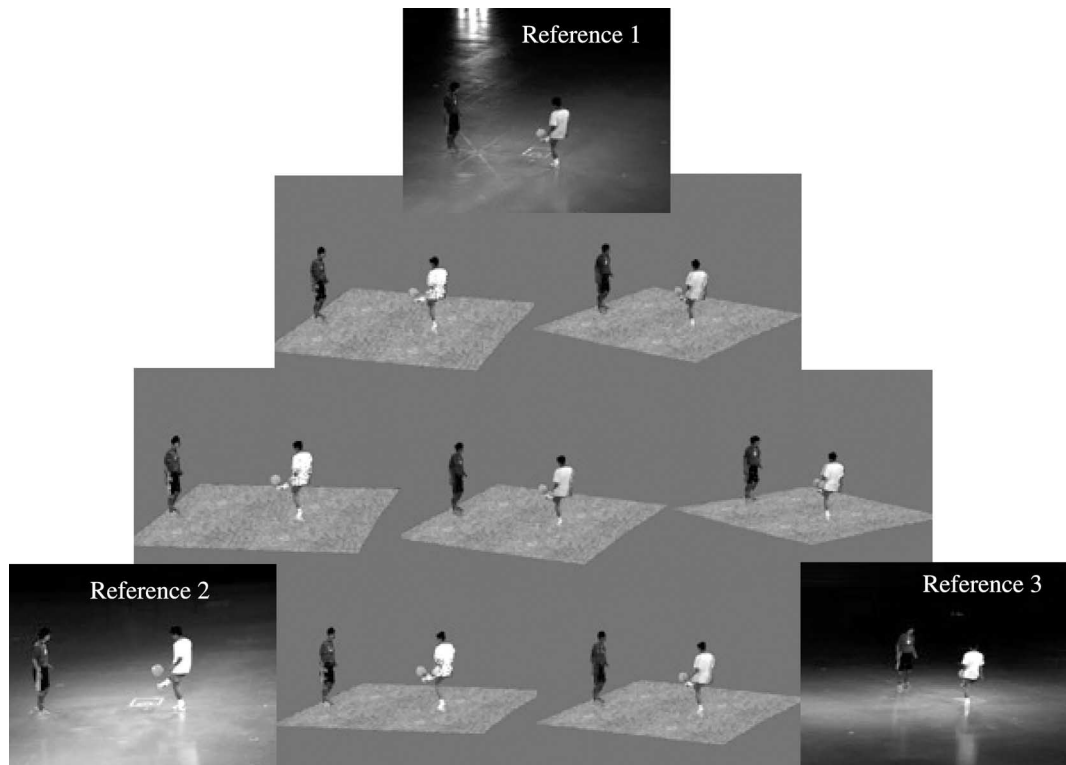


図 6 未校正カメラ画像からの自由視点映像生成例

Fig.6 Example of free viewpoint images which are synthesized from uncalibrated camera images.

大な空間を対象空間とすると、個々のカメラがフィールド全域をとらえていると、フィールドに存在する選手等が画面に占める割合が相対的に小さくなってしまい、このような選手等の物体に対する画像の解像度が不足することになる。そこで、対象空間を複数の小空間(セル)に分割し、セルごとに3次元モデルを構築し、統合することによって全体の3次元モデルを構築し、中間視点画像を生成することは有用である。この節では、射影グリッド空間に基づいて、対象空間を分割し、再統合することにより3次元モデルを構築する手法について述べる。

射影グリッド空間は2台の基底カメラのカメラ座標系によって定義された3次元空間である。そのため用いられる基底カメラが異なる2つの射影グリッド空間を構築した場合、それぞれまったく違う3次元空間軸を持つことになる。しかし、2台の基底カメラが分割された複数のセルで共通ならば、違うセルでも共通の射影グリッド空間を定義することができる。そこで図7に示すように、対象空間全体が画角に収まるような2台のカメラを、各セルのみを画角に収めたカメラグループとは別に用意し、これらの2台のカメラを基底カメラとすることにより、両セル共通の射影グリッ

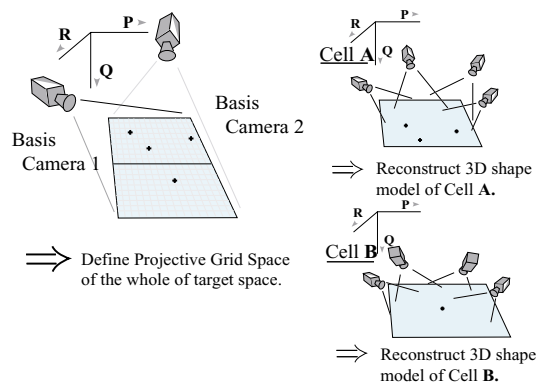


図 7 射影グリッド空間と各セル 3 次元モデルの構築

Fig.7 Construction of Projective Grid Space and 3D shape model for each cell.

ド空間の座標系を設定する。各セルをとらえたすべてのカメラは2台の基底カメラとの間で、F行列によって関係付けられる。そして、各セルに対応するカメラグループに撮影された多視点画像から前述の手法を用いることにより、各セルの3次元モデルを作成する。各セルで作成された3次元モデルは共通の座標系を持つ射影グリッド空間内で作成されているため、図8に示すように、対象空間全体の3次元モデルとして

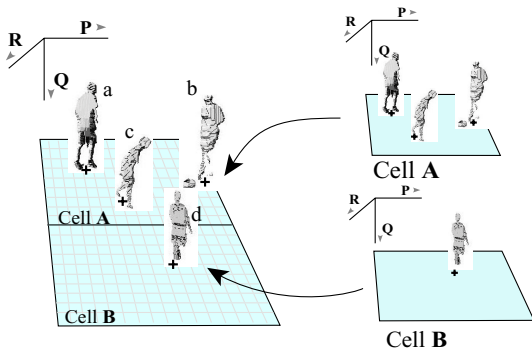


図 8 2つのセルの3次元モデルの統合

Fig. 8 Integration of 3D shape models in two different cells.

統合することが可能である．また前述したように，射影グリッド空間に作成された3次元モデルは，射影グリッド空間を定義する2台の基底カメラとの間がF行列で関係付けされているすべてのカメラの画像に投影することができる．つまり，セルBで作成された3次元モデルであっても，射影グリッド空間を定義する基底カメラが共通であるため，グループAのカメラの画像に投影することができる．つまり，基底カメラが共通であれば，射影グリッド空間内の3次元モデルはすべてのカメラの画像に投影することができるのである．またこのことから，複数のセルで作成した3次元モデルを統合してシーン全体の3次元モデルを構築した場合でも，前述の仮想視点画像合成の枠組みを用いることができる．

以上に述べた方法により，実際にシーンを2つのセルに分けて撮影した画像を用いて仮想視点画像を合成した．対象としたシーンは4人の選手がボール回しをしているシーンである．対象空間を図9のように2つのセル，セルA，セルBに分割し，カメラ1, 3, 5, 7をグループA，カメラ4, 6, 8をグループBとして用いた．さらに，基底カメラ1(カメラ2)，基底カメラ2(天井カメラ)をグループA, B共通に用いた．

入力として用いた各グループの画像と基底カメラ1, 2は，図9に示されている．グループAのカメラ5では，セルAには含まれない選手dが画角から外れ，撮影されていないことが分かる．このため，選手dの3次元形状モデルはグループAのカメラからは形状復元されない．一方，この人物はグループBのカメラのすべてに撮影されている(画像中央)ので，グループBのカメラからは形状復元される．同様に，グループBのカメラ6・カメラ4では，セルBには含まれない選手a・選手bがそれぞれ画角から外れているため，選手a・選手bはグループBのカメラからは形状

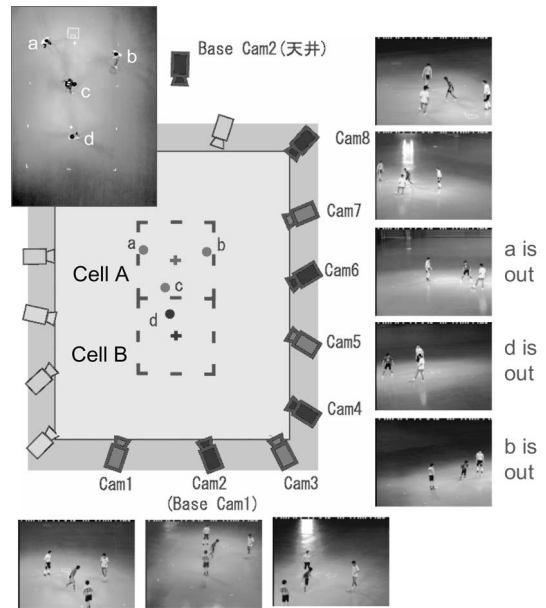


図 9 複数対象空間とカメラグループ

Fig. 9 The multiple object spaces and groups of cameras.

復元されないが，これらの選手はグループAのカメラからは形状復元される．そこで，グループAおよびグループBで構築されたそれぞれの3次元形状モデルを統合し，統合した3次元モデルを用いて生成した仮想視点画像を図10に示す．両端が合成に用いた参照実画像(カメラ1とカメラ3)であり，上の画像がグループAのカメラから構築された3次元モデルを用いて合成された画像，下が統合されたモデルを用いて合成された画像である．これを見てみると，セルAを撮影したグループAのカメラだけでは合成できていなかった選手dが，2つのセルを統合した場合は合成できていることが分かる．

このように，広い対象空間を扱う場合でも，対象空間を複数の部分空間に分け，それらに共通の射影グリッド空間の座標軸を共通の基底カメラにより決定できれば，それぞれの部分空間ごとに3次元モデルを復元して統合することができるので，全体の空間を対象とした仮想視点画像を生成することが可能となる．

なお，ここでは，対象空間を複数のセルに分割して形状復元する手法を用いたが，このような分割をせずに，共通の基底カメラで定義されている射影グリッド空間上の各点で，その点を視野に持つすべてのカメラの画像選択的に利用すれば，より多くのカメラの画像を同時に利用した形状復元が可能になる．本論文では，複数のセルに分割した場合の処理の簡略さ—たとえば，射影グリッド空間のために必要となるメモリ容量が少

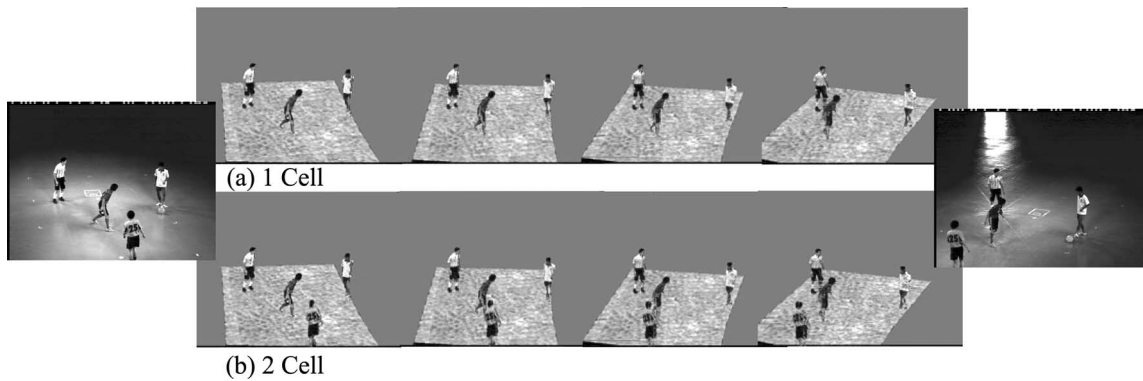


図 10 2 つセルの 3 次元モデルを統合して合成した仮想視点画像
 Fig. 10 Virtual viewpoint images synthesized by integrating 3D models in two different cells.

なくて済む点、また射影グリッド空間上の各点で、それを視野に持つカメラの判定を行う処理を省略可能な点—等から、対象空間を分割する手法を採用したが、カメラをグループに分けずに、より多くのカメラの画像を同時に利用できれば、より精度の高い 3 次元モデル復元が行える可能性があり、今後の検討課題としたい。

3. サッカー映像の中間視点映像生成

3.1 処理概要

本手法では、実際にサッカースタジアムの観客席等の適当な位置に固定した数台のカメラで撮影した多視点映像を入力とする。サッカーの試合の前後に、サッカーフィールド内に 3 次元位置が正確に計測されたマーカを複数配置できれば、固定したカメラのカメラパラメータを推定できるが、実際はそのようなマーカを配置することは事実上不可能に近い。しかし、3 次元位置が未知な自然特徴点を異なるカメラ間で対応付けることにより、カメラ間の射影幾何学的関係を表す F 行列を求めることは容易である。そこで本手法では、まず始めに、あらかじめ入力された多視点映像から適当なフレームを選び、そこから自然特徴点を手入力により抽出することにより F 行列を算出する。

しかし、このようにして推定した F 行列を使って、前章に述べたような射影グリッド空間を設定したとしても、対象シーンを全周囲から取り囲むようなカメラ配置で撮影されていないければ視体積交差法を適用した 3 次元形状復元は困難である。ここでは、たとえば、サッカーのメインスタンドのみに数台のカメラが設置された状況を想定し、対象シーンを全周囲から取り囲まない多視点映像からでも、隣接するカメラ間の射影幾何学的関係を利用して中間視点映像を生成する手法について述べる。

本手法では、サッカー映像を、時間とともに位置や形状の変化する移動領域（選手とボール）と、それら以外の静止領域に分け、領域ごとに隣接するカメラ間で画素ごとに対応点を決定し、2.3 節と同様の原理に基づく視点内挿法により中間視点映像を生成し、最後にこれらの領域を合成する。

静止領域は、対象シーンがサッカー映像であることを利用して、いくつかの平面領域に分割できるものと近似する。平面に近似された各領域に対して、隣接するカメラ間の平面射影行列 (Homography) を用いて画素ごとに対応点を決定し、中間視点映像を生成する。この平面射影行列は、対応する領域内で数点の対応点を手入力により検出することにより算出する。これらの処理では、対応点の抽出等、一部マニュアル作業をとまなうが、静止領域に対するものであるため、入力多視点映像に対して 1 回行うだけでよく、作業は容易である。

一方、移動領域については、あらかじめ算出した F 行列と、静止領域の処理で算出したグラウンド面に関する平面射影行列を用いて、隣接するカメラ間で対応点を密に算出し、中間視点映像を生成する。

最後にそれらを静止領域、移動領域の順に合成して対象シーン全体の中間視点映像を得る。図 11 に、本手法の全体の流れをまとめる。

下記に、それぞれの領域に対する処理を説明する。

3.1.1 静止領域

静止領域に対する処理は、入力される多視点映像から、選手やボールの存在しないフレームをあらかじめ抽出しておき、この多視点映像に対して行う。本手法では、対象がサッカーであることを利用して、静止領域をグラウンド、サッカーゴール、スタンド等の遠景の 3 領域に分割し、それぞれを平面で近似する。サッカーゴールについては、これを構成する複数の平面群

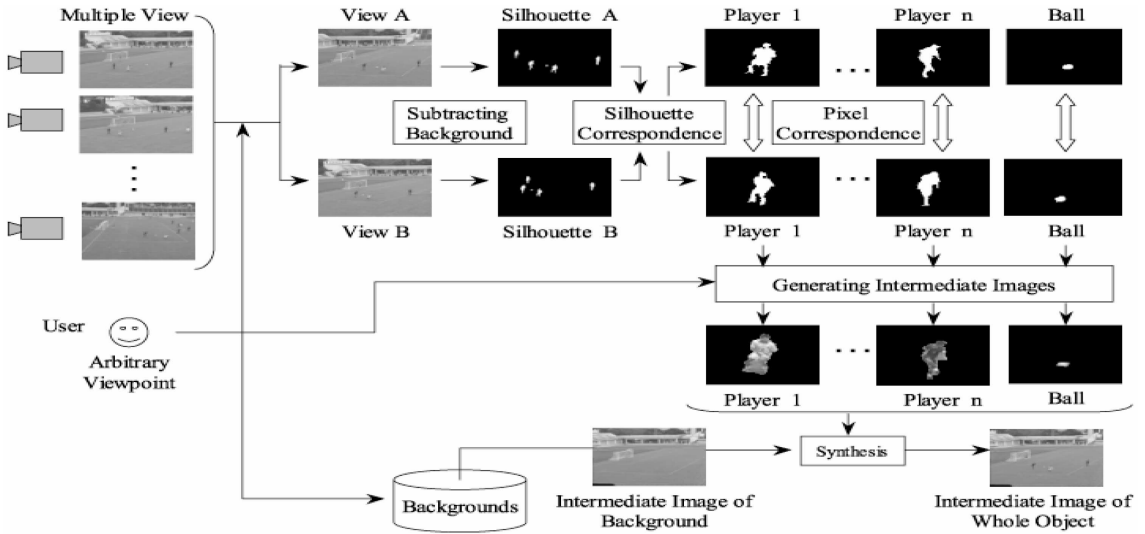


図 11 サッカーシーンの中間映像生成のための流れ図
 Fig. 11 Flow diagram for intermediate view synthesis of soccer scene.

で近似する．遠景については，カメラから遠方であることから，平面近似が可能と考える．

このような近似に基づき，各視点の画像を複数の平面領域に分割し，隣接するカメラ間で対応する領域ごとに平面射影行列を算出する．この平面射影行列は，対応する領域内で数点の対応点を手入力により与えることにより求める．各領域に対する平面射影行列が算出されれば，次式により隣接するカメラ間での対応点を画素ごとに得ることができる．

$$p_2 \cong Hp_1 \tag{4}$$

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，それぞれカメラ 1，カメラ 2 における同次座標である．この平面射影行列が，複数の平面として近似された各平面領域における画素ごとの対応関係を与える．

このように隣接するカメラ間で対応付けされた各点の位置関係から，次式によりその点の中間視点画像における位置を算出することができる．

$$\hat{p} = (1 - \alpha)p_1 + \alpha p_2 \tag{5}$$

ここで， p_1 と p_2 は，それぞれ，カメラ 1，カメラ 2 における対応点の位置であり， α は 2 つのカメラ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，各カメラに対するワーブ画像が生成される．さらに，各ワーブ画像を次式のようにブレンド合成することによって，中間視点画像を合

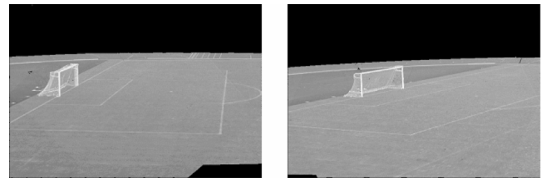


図 12 グラウンドとゴールに対する中間視点画像
 Fig. 12 Intermediate view images for the ground and the soccer goal.

成することができる．

$$I = \begin{cases} I_1, & \text{if } I_2 \text{ is not available in camera 2} \\ I_2, & \text{if } I_1 \text{ is not available in camera 1} \\ (1 - \alpha I_1) + \alpha I_2, & \text{otherwise} \end{cases} \tag{6}$$

図 12 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．

ただし，遠景領域に対しては，隣接するカメラの 2 枚の原画像間で共通の領域が狭く，ワーブ画像の片方にしか画素値が存在しない画素が大半を占めるため，上式のようにワーブ画像を作成し，それをブレンドする，という処理では計算効率が悪い．そこで，まず，隣接するカメラ 1 と 2 間で，平面と見なした遠景領域に関する平面射影行列 H_{21} を用いることにより，カメラ 1 の遠景領域にカメラ 2 の遠景領域を連結したパノラマ画像を生成する．そして，中間視点画像の視



図 13 遠景に対するパノラマ画像

Fig. 13 Panoramic view image for the distant view.



図 14 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

Fig. 14 Silhouettes of the detected players and ball. The silhouettes of the players are corresponded with those in the different camera images using the homography between the cameras by assuming that the players stand on the ground.

点を指定するための内挿比率パラメータ α を用いて、遠景領域に関するカメラ 1 の画像から中間視点画像への平面射影行列を次式のように算出し、これを利用してパノラマ画像から中間視点画像における遠景領域を生成する処理を行う。

$$\hat{H} = (1 - \alpha)E + \alpha H_{21}^{-1} \quad (7)$$

ただし、 E は単位ベクトルである。図 13 に、合成したパノラマ画像の一例を示す。

3.1.2 移動領域

移動領域については、入力された多視点映像から、まず図 14 に示すように、背景差分によって選手およびボールのシルエットを検出する。そして、検出された複数の選手、ボールのシルエット領域を、隣接す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、グラウンド面の平面射影行列により関係付けられることを用いる。

次に、隣接す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 15 に示すように異なるカメラ間でエピソード線を引き、各エピソード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での視差変化は一定と近似し、シルエット両端の対応関係を線形補間

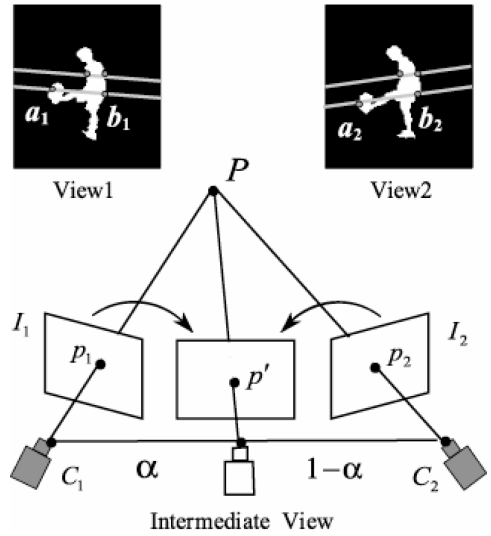


図 15 中間視点画像における位置の算出
Fig. 15 Determination of the corresponding point in the intermediate view image.

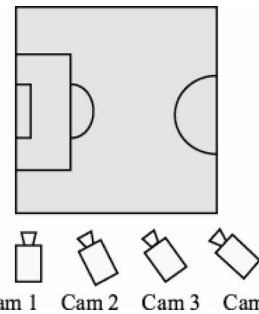


図 16 サッカーシーンの撮影のための多視点カメラの配置
Fig. 16 Arrangement of the multiple cameras for capturing the soccer scene.

することによって画素ごとの対応付けを行う。この近似は、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ことに基づいている。このようにして、シルエットの上部から下部にかけてエピソード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。

そして、この対応関係を用いて、式 (5)、式 (6) のようにして、中間視点画像を生成する。

3.2 中間視点画像の例

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 16 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、

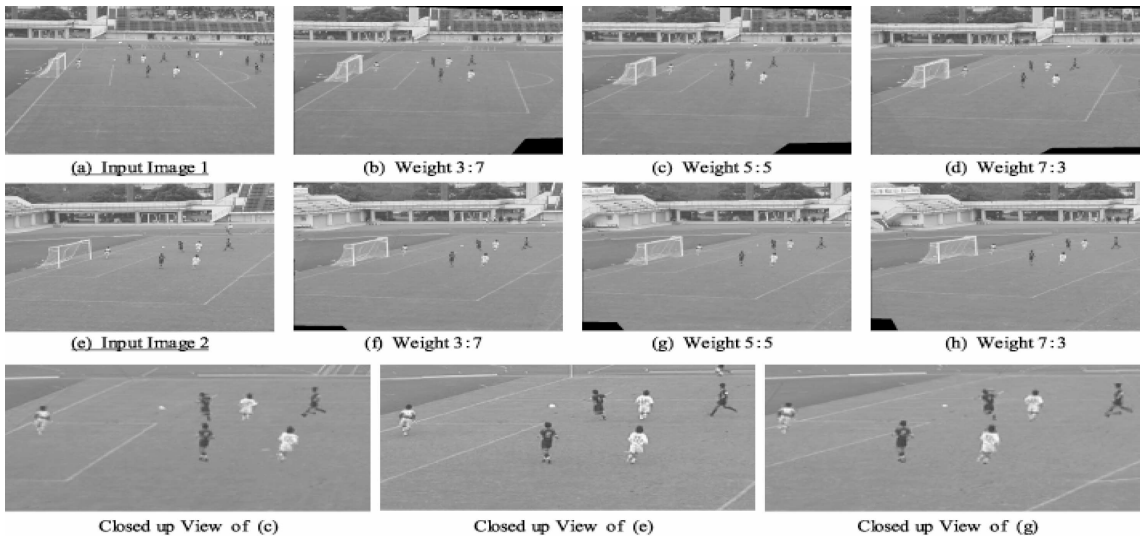


図 17 生成された中間視点画像の例

Fig. 17 Example of the synthesized intermediate viewpoint images for the soccer scene.

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ほとんど不可能であり、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、画質劣化の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分かる。

4. おわりに

本論文では、多視点画像間の自然特徴点の対応関係のみから容易に得ることのできる射影幾何学的関係を利用して、実カメラ視点を内挿した仮想視点における中間視点画像を合成する手法について述べた。まず、弱校正により得られるカメラ間の射影幾何学的関係によって対象空間に座標系を構成するための、射影グリッド空間という枠組みについて説明した。そして、この射影グリッド空間において、視体積交差法を用い

て対象シーンの3次元復元を行い、復元された3次元モデルから実カメラ視点の中間視点画像を生成する方法を示した。次に、視体積交差法等を利用した3次元形状復元が困難となる実際のサッカーシーンを撮影した多視点画像を用いた中間視点画像生成を行うために、移動物体である選手・ボール、静止物体であるグラウンド面とサッカーゴール・遠景に対象シーンを分割し、それぞれの射影幾何学的性質に基づいて中間視点画像を生成し、それらを合成する手法を示した。

本論文では、中間視点画像の仮想視点が、3ないし2つの参照画像の視点を指定した重みで内分した位置になると仮定している。しかし、この仮定が正確に成り立つのは、3つの参照画像の場合、これらの投影中心を通る平面に平行な平面上に3つの参照画像面が存在している状況に限られる。また、2つの参照画像の場合は、これらの投影中心を通る直線に平行な平面上に2つの参照画像が存在している状況に限られる。本論文では、視点内挿に用いる隣接したカメラは大よそ同じ方向を向いており、焦点距離も大よそ等しいと考え、この仮定が近似的に成立するものと見なしているが、参照画像を撮影したカメラの位置関係によっては、この近似が仮想視点画像の見え方に大きな歪を発生させる可能性がある。2つの参照画像からの内挿については、この仮定を正確に満足するように視点内挿する手法がView Morphing¹⁸⁾として提案されているので、今後は、この手法の考え方を適用した内挿法¹⁴⁾を検討する予定である。

また本論文では、視点の内挿に2台か3台のカメラを用いる手法について述べた。この場合、内挿した仮

これらの結果のフルカラー画像および生成した中間視点映像は、<http://www.ozawa.ics.keio.ac.jp/Saito/research/vpod/index.html> から見る事ができる。

想視点移動できる範囲は、カメラが2台の場合にはカメラの投影中心を結んだ線上、カメラが3台の場合にはカメラの投影中心を通る面上に限定されている。これを拡張して視点の内挿に用いるカメラを4台(もしくはそれ以上)にすれば、内挿した仮想視点ができる範囲を、カメラの投影中心を結んだ多面体の体積空間に拡張できると考えられる。このような拡張についても、今後の研究課題としたい。

謝辞 本論文の2章で用いた多視点画像は、通信・放送機構委託研究(地域提案型研究開発制度)「仮想化現実技術による自由視点三次元映像スタジオ通信の実証研究」の一環として撮影されたものである。本研究プロジェクトの関係各位に感謝する。

参考文献

- 1) Avidan, S. and Shashua, A.: Novel View Synthesis by Cascading Trilinear Tensors, *IEEE Trans. Visualization and Computer Graphics*, Vol.4, No.4, pp.293-306 (1998).
- 2) Beier, T. and Neely, S.: Feature-Based Image Metamorphosis, *Proc. SIGGRAPH'92*, pp.35-42 (1992).
- 3) Chen, S. and Williams, L.: View Interpolation for Image Synthesis, *Proc. SIGGRAPH'93*, pp.279-288 (1993).
- 4) 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, 電子情報通信学会技術報告, PRMU2001-151 (2001).
- 5) Laurentini, A.: The Visual Hull Concept for Silhouette Based Image Understanding, *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol.16, No.2, pp.150-162 (1994).
- 6) Matusik, W., Buehler, C., Raskar, R., Gortler, S. and McMillan, L.: Image-Based Visual Hulls, *Proc. SIGGRAPH 2000*, pp.369-374 (2000).
- 7) Moezzi, S., Tai, L.C. and Gerard, P.: Virtual View Generation for 3D Digital Video, *IEEE MultiMedia*, Vol.4, No.1, pp.18-26 (1997).
- 8) Neumann, J. and Aloimonos, Y.: Spatio-Temporal Stereo Using Multi-Resolution Subdivision Surfaces, *International Journal of Computer Vision*, Vol.47, No.1-3, pp.181-193 (2002).
- 9) Kanade, T., Rander, P.W., Vedula, S. and Saito, H.: Virtualized Reality: Digitizing a 3D Time-Varying Event As Is and in Real Time, *International Symposium on Mixed Reality (ISMR99)*, pp.41-57 (1999).
- 10) 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 (2000).
- 11) 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.13-18 (2000).
- 12) Kitahara, I., Saito, H., Ohta, Y., Akimichi, S., Ono, T. and Kanade, T.: Large Scale Virtualized Reality, *IEEE Computer Vision and Pattern Recognition (CVPR2001)*, CD-ROM Vol.2, Technical Sketches (2001).
- 13) Kimura, M. and Saito, H.: 3D Reconstruction Based on Epipolar Geometry, *Trans. IEICE on Information and Systems*, Vol.E84-D, No.12, pp.1690-1697 (2001).
- 14) 木村 誠, 斎藤英雄: 3カメラ間の射影幾何を用いた View Morphing, 情報処理学会研究報告, CVIM-135-16 (2002).
- 15) Saito, H. and Kanade, T.: Shape Reconstruction in Projective Grid Space from Large Number of Image, *IEEE Proc. Computer Vision and Pattern Recognition (CVPR99)*, Vol.2, pp.49-54 (1999).
- 16) Saito, H., Baba, S., Kimura, M., Vedula, S. and Kanade, T.: Appearance-Based Virtual View Generation of Temporally-Varying Events from Multi-Camera Images in the 3D Room, *2nd International Conference on 3-D Digital Imaging and Modeling (3DIM99)*, pp.516-525 (1999).
- 17) 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, 情報処理学会研究報告, CVIM-119-16, Vol.99, No.93, pp.117-124 (1999).
- 18) Seitz, S.M. and Dyer, C.R.: View Morphing, *Proc. SIGGRAPH'96*, pp.21-30 (1996).
- 19) Tsai, R.: A Versatile Camera Calibration Technique for High-Accuracy 3D Machine Vision Metrology Using Off-the-Shelf TV Cameras and Lenses, *IEEE Journal of Robotics and Automation*, RA-3, 4, pp.323-344 (1987).
- 20) Vedula, S., Rander, P.W., Saito, H. and Kanade, T.: Modeling, Combining, and Rendering Dynamic Real-World Events From Image Sequences, *Proc. 4th Conf. Virtual Systems and Multimedia (VSMM'98)*, Vol.1, pp.326-332 (1998).
- 21) ウ 小軍, 延原章平, 和田俊和, 松山隆司: 多視点映像からの実時間3次元形状復元とその高精度化, 情報処理学会研究報告, CVIM-131-9 (2002).
- 22) ウ 小軍, 圓藤康平, 和田俊和, 松山隆司: 3次元ビデオ映像の能動的実時間撮影と対話的編集・表示, 電子情報通信学会技術報告, PRMU2000-187 (2001).
- 23) 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).

24) <http://www.pvi-inc.com/eyevision>

(平成 14 年 4 月 2 日受付)

(平成 14 年 9 月 12 日採録)

(担当編集委員 内海 章)



斎藤 英雄 (正会員)

昭和 62 年慶應義塾大学理工学部電気工学科卒業。平成 4 年同大学院電気工学専攻博士課程修了。博士(工学)。同年慶應義塾助手，専任講師を経て，平成 13 年より慶應義塾

理工学部情報工学科助教授。この間，平成 9 年～11 年まで，学術振興会海外特別研究員として米国カーネギーメロン大学ロボティクス研究所に滞在し，主に Virtualized Reality の研究に従事。平成 12 年より科学技術振興事業団さきがけ 21 研究員兼務。電子情報通信学会，電気学会，映像情報メディア学会，VR 学会，計測自動制御学会，IEEE 各会員。



木村 誠

平成 7 年慶應義塾大学理工学部電気工学科卒業。平成 14 年同大学院開放環境科学専攻博士課程修了。平成 10 年～11 年，米国カーネギーメロン大学ロボティクス研究所訪問研究生。平成 14 年 10 月から，産業技術総合研究所デジタルヒューマン研究ラボ，科学技術振興事業団 CREST 研究員。博士(工学)。コンピュータビジョンに関する研究に従事。電子情報通信学会会員。



矢口 悟志

平成 12 年慶應義塾大学理工学部情報工学科卒業。平成 14 年同大学院開放環境科学専攻修士課程修了。同年，NTT コムウェア株式会社入社。在学中，多視点映像からの任意視点映像生成に関する研究に従事。



稲本 奈穂

平成 14 年慶應義塾大学理工学部情報工学科卒業。現在，同大学院開放環境科学専攻修士課程在学中。多視点サッカー映像からの任意視点映像生成に関する研究に従事。電子情報通信学会学生会員。