

ウェアラブルビジョンのための実環境画像からの視点位置姿勢実時間推定手法
Real-time camera parameter estimation from images for wearable vision system

大隈隆史¹ 蔵田武志¹ 興相正克² 坂上勝彦¹
Takashi Okuma Takeshi Kurata Masakatsu Kouroggi Katsuhiko Sakaue
¹ 電子技術総合研究所 ² 早稲田大学
Electrotechnical Laboratory Waseda University
¹{okuma,kurata,sakaue}@etl.go.jp
²kouroggi@muraoka.info.waseda.ac.jp

Abstract 本稿では、視点位置から撮影された実環境画像からユーザの視点位置姿勢情報を実時間で推定する手法を提案する。従来、実環境画像からの視点情報の実時間推定手法では、実環境に人工的に配置されたマーカを抽出・追跡し、この情報を用いて視点情報を獲得もしくは補正している。本手法では色ヒストグラムマッチングとテンプレートマッチングを組合せて適用することで、単眼の実環境画像からマーカを用いずに既知の特徴点情報を追跡してユーザの視点情報を獲得し、実環境を撮影した画像と3次元CGの合成画像を実時間で提示する。

1. まえがき

計算機の小型化に伴い、いつでもどこでも利用できる着用可能な計算機(ウェアラブルコンピュータ)の実現の可能性が高まり、注目を集めている[1]~[6]。ウェアラブルコンピュータの重要な機能のひとつとして着用者の能力拡張、補助機能が挙げられるが、このような機能は状況把握に基づく情報収集・提示といった機能を用いて実現される。

一方で現在、現実環境にコンピュータグラフィックス(CG)を用いて提示される仮想環境を融合して提示する複合現実感技術に注目が集まっている[7][8]。複合現実感技術は現実環境における作業を補助するための情報を仮想物体としてユーザに提示するためのさまざまな要素技術から成り立つ。本稿では複合現実感技術で用いられる画像からの視点位置姿勢推定手法[9][10]を用いて、現在著者らが構築を進めているウェアラブルビジョンシステム[2]に実装されている実映像への注釈情報の重ね合せ機能[5][6]に3次元情報の提示手法を付加することを目指す。

以下まず2節では、現在までに開発してきたウェアラブルビジョンシステムの概要を紹介し、複合現実感技術を用いた3次元情報提示とウェアラブルビジョンシステムを組合せることによる得失を検討する。3節では、本論文で提案する自然特徴点追跡に基づく視点位置推定手法のアルゴリズムについて述べる。4節では提案手法の動作を検証するために実環境画像を用いた実験とその結果

について述べる。

2. ウェアラブルビジョンシステム

2.1 システムの概要

著者らは常時カメラを着用している状況特有の利点、問題点、ならびに応用について検討するために、実際に稼動するウェアラブルシステムを構築している[2]~[6]。本システムは着用者が身につけるウェアラブル機器と画像処理などを受け持つリモートPCクラスタ、およびそれらを繋ぐ無線LANからなる。図1に本システムのウェアラブル機器を身につけた着用者の様子を示す。着用者が身につけるヘッドセットは超小型カラーCCD(東芝IK-SM43H, レンズ焦点距離4mm)、単眼超小型ディスプレイ(MicroOptical社C1 Clip-On Display)からなり、着用者の目や顔を隠さないスタイルになっている。また、このようなスタイルを取ることで、着用者の視線方向とカメラの視線方向をある程度一致させながら、着用者の視野内に出力画像を提示することが可能になる。

2.2 複合現実感技術

現実環境から取得された情報を用いて作成された仮想環境を提示する技術やその仮想環境をさらに現実環境と融合して提示する技術は複合現実感技術と呼ばれる[7][8]。複合現実感を実現するにあたり、現実環境における任意の位置に任意の姿勢で仮想物体を描画するために、現実環境の基準座標系と仮想環境の基準座標系を一致させる必要が



図 1：ウェアラブル機器着用者の外観

ある。これは現実環境に設定される基準座標系におけるユーザの視点の位置姿勢情報を実時間で獲得し、この視点情報に基づいて描画された CG 画像を現実環境に合成して提示することで実現される。このため、ユーザの視点位置姿勢情報の獲得は複合現実環境を実現するために必要不可欠な問題である。ユーザの視点情報を獲得する手法は、電磁気式、超音波式などの 3 次元位置センサを用いる手法[11]と、現実環境をカメラによって撮影した画像から視点情報を推定する手法[10][12][13]に大別できる。本稿では特に画像からの視点情報推定手法に着目する。なぜならば、画像からの視点情報推定手法ではカメラが視点位置に取り付けられていると仮定し、入力画像内に撮影される 3 次元位置が既知の複数の特徴点からカメラの位置姿勢を推定することで視点情報を得るため、現在のウェアラブルビジョンシステムの構成と親和性が高いと考えられるからである。

従来、人工的なマーカを自然環境に配置して既知の特徴点として利用することで特徴点の追跡を簡略化していた [9] ~ [13][20]。しかしながら、マーカを特徴点として利用する限り、測定可能範囲を拡大するためにはマーカの設置範囲を拡大しなければならない。マーカを用いずに現実環境に存在する自然特徴点を利用することができるようになれば、特徴点の配置や認識方法を工夫することで測定可能範囲を自由に設定できるようになる。測定可能範囲の拡大は「いつでも、どこでも」利用可能な情報提示を目指す上で必要不可欠な処理であると考えられる。しかし、特徴点を認識する処理はノイズや光源環境

の変化などの影響を受けやすい。また、視点の移動にともなって、視点情報の推定に用いる特徴点が隠蔽されたり撮影範囲外へ出たりするため、カメラ位置姿勢の推定に必要な特徴点を安定して追跡するのは難しい。その結果、任意の範囲での実時間カメラ位置姿勢推定は一般に困難となる。

そこで本稿では人工的なマーカを用いずに現実環境に存在する自然特徴点をユーザの 3 次元視点位置姿勢の推定に用いるために、色ヒストグラムマッチング[14][15]とテンプレートマッチングを併用する自然特徴点探索手法を提案する。また、入力画像中で追跡に成功した特徴点の個数に応じて、撮影されたマーカの画面座標値からカメラの位置姿勢を求める複数の計算方法を切替えて用いる手法[11]を用い、実際に 3 次元情報を合成する実験システムを試作した。次節以降で提案手法について詳細に述べる。

3. 3 次元情報提示機能

現在のウェアラブルビジョンシステムに複合現実感技術を用いた 3 次元情報提示機能を付加することで、現在までに実現してきた注釈情報の提示機能[3][4]の表現力を向上させることができる。また、3 次元形状情報入力機能[5][6]により取得された 3 次元形状データの自然な視点移動による観察などが可能になる。

3.1 処理の概要

画像が入力されてからカメラの位置姿勢情報が取得されるまでの処理は以下を繰り返すことで進行する。

1. 入力画像から 3 次元位置が既知である複数の自然特徴点を探索・追跡し、画面座標値を求める。
2. 抽出された特徴点の個数に応じた PnP 問題処理ルーチンでカメラの位置姿勢を推定する。
3. 推定したカメラの位置姿勢をもとに、撮影されていない特徴点の画面座標を推定する。

以降では、各段階での処理について述べる。

3.2 自然特徴点の追跡手法

カメラの位置姿勢を推定するために、入力画像から既知の特徴点を探索・追跡して画面上での 2 次元座標を求める必要がある。本手法では自然特徴点の情報を要素とするリス

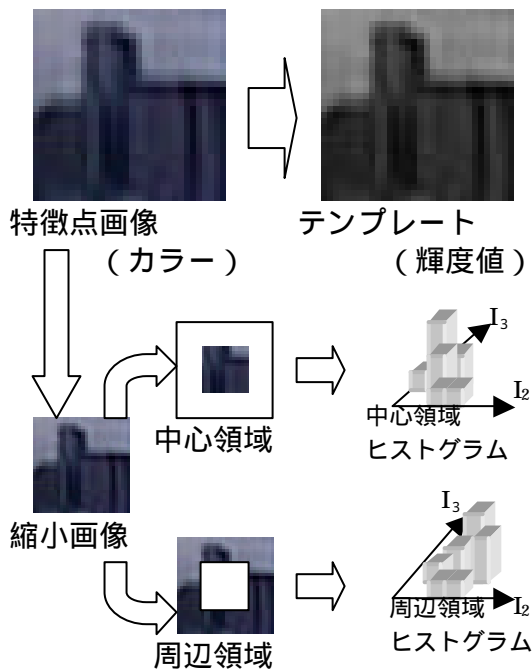


図 2 : 探索に用いる画像特徴量の計算

ト (特徴点リスト) を事前に設定し、登録されている特徴点を既知の特徴点として入力画像中から探索・追跡して画面座標を求める。特徴点リストには事前に計測した 3 次元座標情報と探索・追跡処理に用いられる画像特徴量が記録されている。

初期フレームでは入力画像全体を探索対象とするが、第 2 フレーム以降では、連続するフレーム間の特徴点の画面上の移動距離は小さいと仮定をおき、前のフレームでの画面座標を中心とする探索窓を設定する。

従来から対応点探索手法として多くの手法が提案されている。しかし、輝度値による類似尺度のみを用い、色の情報を利用しない手法も多い。本研究では色の情報も自然環境中で利用できる重要な特徴であると考え、色ヒストグラム[14][15]に着目した。色ヒストグラムを用いたマッチングは比較的高速で安定した探索が可能であるが、通常のテンプレートマッチングなどで利用できる画素の位置関係を利用できないため精度に問題が残る。

そこで、本手法では、精度を保ちつつ計算量を削減するために対応点探索を 2 段階に分けて処理する。第一段階では入力画像の縮小画像を用い、色情報によるマッチングにより対応点候補の数を絞る。次に第二段階では

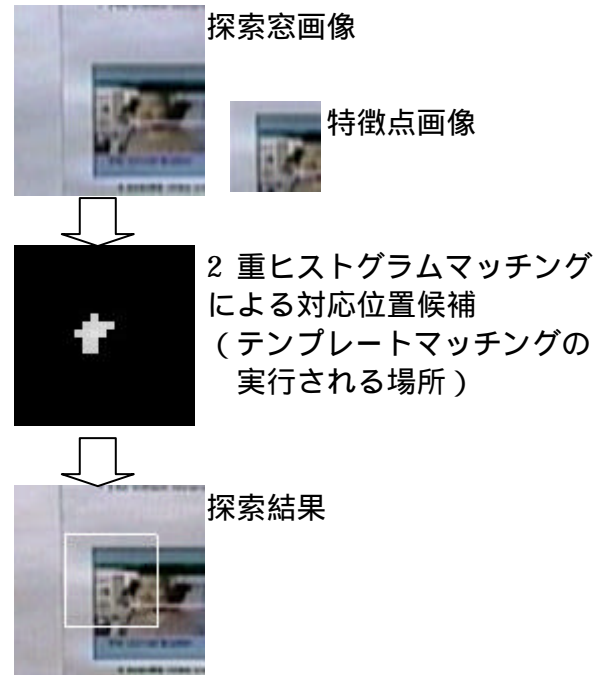


図 3 : 探索処理の様子

元のサイズの画像を用い、テンプレートマッチングによって輝度値の位置情報を用いた正確な対応点探索を行う。

さらに、計算量の削減と安定した追跡を行うために第一段階でのマッチングに利用する色情報として

- (1) セグメンテーションに有効とされる $I_1 I_2 I_3$ 表色系 [16] を用いて輝度情報を分離し、色の情報 (I_2, I_3 値) による 2 次元ヒストグラムを、
- (2) 特徴点中心領域と周辺領域の 2 領域に分けてそれぞれに対して生成

したヒストグラム (2 重色ヒストグラム) を利用する (図 2 参照)。これにより、通常の色ヒストグラムでは利用できない、色の位置情報のある程度保存したマッチング処理を行うことができる。

第一段階のマッチング処理の類似度にはこの 2 重ヒストグラムにおけるヒストグラムインタセクションを用い、第一段階での類似度が閾値以上の対応位置候補についてのみ第二段階のテンプレートマッチング処理を行う。第二段階では正規化相互相関を用いたテンプレートマッチングを行う。マッチングの結果、類似度の最大値が閾値以上であった場合にその位置を特徴点の画面座標値と

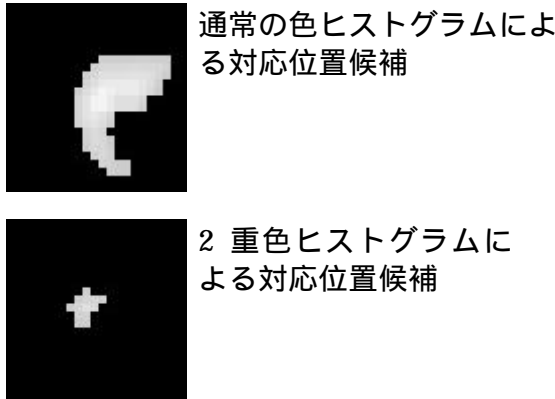


図 4：2重色ヒストグラムマッチングと通常の色ヒストグラムマッチングの比較

する。図 3 に実際のマッチング処理経過の例を示す。この例では2重ヒストグラムマッチングによってテンプレートマッチングを行う回数が約 2%にまで減少している。また、図 4 には図 3 で示した特徴点画像と探索窓画像に対して、通常の色ヒストグラムを用いてマッチングした場合と2重色ヒストグラムを用いてマッチングした場合の対応位置候補を示す。この図に示した例では2重色ヒストグラムによる絞込みにより、通常の色ヒストグラムを用いる場合に比べ、対応位置候補を13%程度に減少させている。

ここで探索窓内に探索対象となる特徴点が見つからなかったと判定された場合(マッチングの結果、類似度の最大値が閾値以下の場合)には、現実物体に隠蔽された、もしくは撮像範囲内から外れたと判断する。見つからなかった特徴点の画面座標は、特徴点の3次元座標が既知であるという仮定をおいているので、以降で求めるカメラの位置姿勢情報を用いて推定できる。この情報は次フレーム以降の追跡に用いられる。

3.3 カメラの位置姿勢の計算

カメラの位置姿勢は実世界に設定されている基準座標系からカメラ座標系への変換行列(モデルビュー行列) M で表すことができる。すなわち、ある点の基準座標系での座標値を w 、カメラ座標系での座標値を c とするとき $c = Mw$ と変換する行列である。図 5 に実世界基準座標系とカメラ座標系の関係を示す。

本稿では前節で画面座標を獲得した特徴点の個数に応じて任意の 6 点、平面上の 4

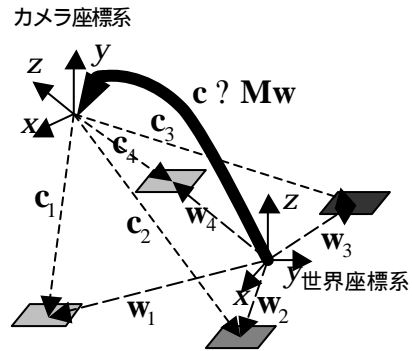


図 5：座標系とモデルビュー行列

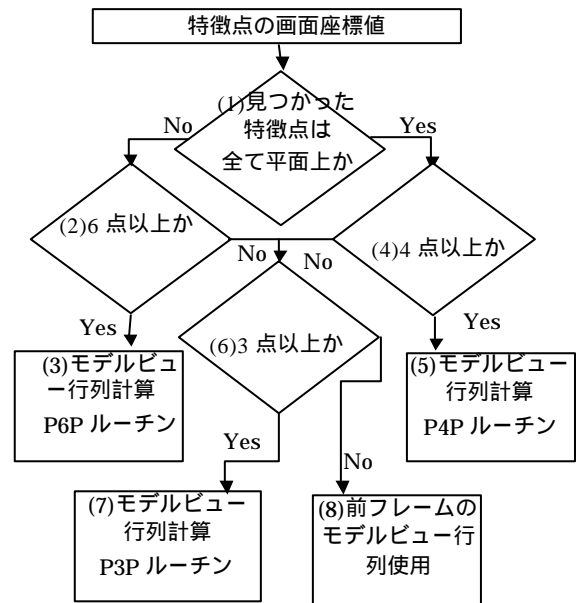


図 6：モデルビュー行列計算処理の概要

点、任意の 3 点の画面座標値からカメラの位置姿勢を推定する手法[17][18][19]を選択して使用する手法[11]を適用した。図 6 に各フレームでの処理の概要を示す。

- (1) 見つかった特徴点が全て平面上に存在するかを判定。全て平面上であれば(2)へ、されていないければ(4)へ。
- (2) 6 点以上の点が見つまっているか。見つかっていれば(3)へ、見つかっていなければ(6)へ。
- (3) 6 点を用いてカメラの位置姿勢を求める。終了。
- (4) 平面上の 4 点の特徴点が見つかっていれば(5)へ、されていないければ(6)へ。
- (5) 4 点を用いてカメラの位置姿勢を求め終了。
- (6) 3 点以上の特徴点抽出されていれば(7)

へ、されていないければ(8)へ。

- (7) 3点を用いてカメラの位置姿勢候補を求める。このとき、4点以上の点が追跡されている場合には、すべての点をすべての位置姿勢候補で画面座標系に再投影し、画面座標の距離の和が最も小さい位置姿勢候補を出力して終了。3点のみ抽出されている場合には、カメラの位置が前のフレームでのカメラの位置に最も近いものを選んで終了。
- (8) 前のフレームでの位置姿勢をそのまま用いる。

このとき、ステップ(3)、(5)、(7)で必要以上の特徴点数が抽出されている場合、カメラの位置姿勢を求めるために用いられる特徴点として、抽出されたものの中から特徴点リストの登録順に必要な数だけを選択して計算に用いる。

3.4 観測されない特徴点の位置の推定

前節までの処理により、モデルビュー行列 M が求まる。しかし、視点・視線の自由な変更を可能にし、カメラの位置と姿勢を安定に推定するためには、広範囲に存在する既知の特徴点を順次切替えて利用する必要がある。このため、各フレームで抽出・追跡に失敗した特徴点についてもモデルビュー行列 M を用いてカメラ座標系への変換を行ない、画面座標値を推定する。この座標値は画像枠外になることもあり得る。この処理により、実物体によって隠蔽されたり、撮像範囲外に存在するため追跡できない特徴点の位置を推定しておき、再び見えるようになったときや、撮像範囲内に新しく入って来た場合に追跡を再開もしくは開始することができる。

4. 動作確認実験

前節で提案した実時間カメラ位置姿勢推定手法の動作を確認するために、ウェアラブルビジョンシステムのヘッドセットと動画像の入出力および動画像のハードウェアテクスチャマッピング機能を持つ PC(SGI VWS540 PentiumIII Xeon 500MHz 4CPU) を用いて動作確認実験を行った。

提案手法を用いて3次元CGを合成した画像提示の様子を図7に示す。この例ではウェアラブルビジョンシステムの3次元情報取得機能[5][6]により生成したコアラのぬいぐるみのCGモデルを、その研究紹介用のパネ

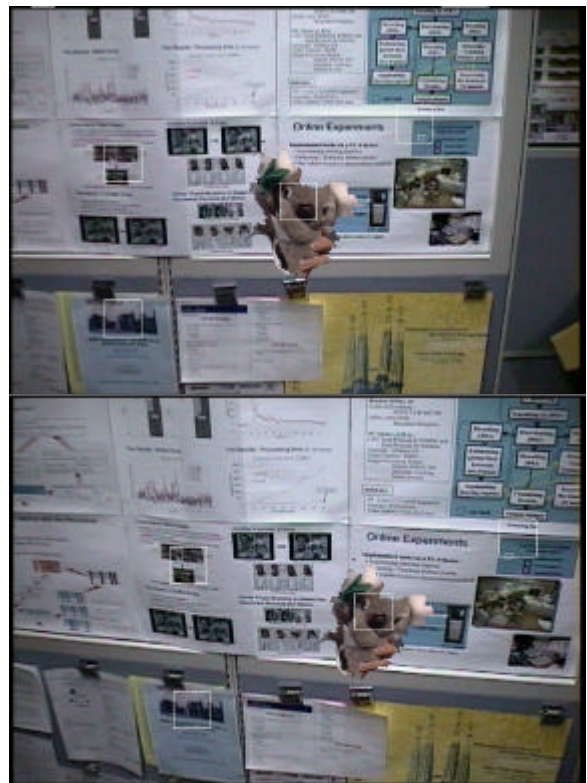


図7：合成実験結果画像

ルに対する付加情報として観察者に提示している。合成環境を提示する手法にはビデオシースルー方式による合成画像提示を採用した。試作システムの画像更新レートは、実行例に示す程度の仮想物体(約4000ポリゴン)を合成する場合で約5~10フレーム/秒である。入力画像の解像度は 360×243 画素、このとき追跡していた特徴点数は4点で、特徴点領域、探索窓領域はそれぞれ 24×24 画素、 48×48 画素であった。

5. 考察

本手法では現在、画像中の特徴点としてどの点を選ぶかは経験的に選択し、手動で設定している。また、設定した特徴点の3次元座標も事前に手動で測定し、データとして入力している。この作業量は今後、3次元情報取得機能[5][6]との連携により軽減されることが考えられる。また、パノラマ画像に基づく情報提示手法から得られる大局的な位置情報を用いることで、画像中を全探索する初期フレームにおける特徴点探索の処理を軽減することができる。

また、現在、画像のみから視点の位置姿勢を推定しているが、今後、ジャイロセンサや加速度センサ等、システムに組み込み可能な

センサと画像からの情報を併用することで
更なる高速化, ロバスト化が期待される[20].
6. むすび

本稿では, 人工的なマーカを用いずに既知
の自然特徴点を用いて入力画像からカメラ
の位置と姿勢を求める手法を提案し, 提案手
法によるウェアラブルビジョンシステムへ
の3次元情報提示機能の追加を検討した. 特
に自然特徴点の探索・追跡手法として2重色
ヒストグラムのヒストグラムインタセク
ションによるマッチングと, 正規化相互相関
によるテンプレートマッチングを2段階に用
いる特徴点の探索・追跡手法を提案した. ま
た, 提案手法の動作を確認するために実環境
画像に本手法を適用し, その動作例を示すこ
とで提案手法の有効性を確認した.

謝辞 本研究はリアルワールドコンピ
ューティング(RWC)プログラムの一環とし
て行われた.

参考文献

- [1] T. Starner, S. Mann, B. Rhodes, J. Levine, J. Healey, D. Kirsch, W. R. Picard, A. Pentland, "Augmented Reality Through Wearable Computing," Technical Report 397, MIT Media Lab. Perceptual Computing Section, 1997.
- [2] 蔵田, 興梠, 栗田, 村岡, 坂上, "気の効いたウェアラブルビジョンシステムの実現に向けて: 顔検出 識別とハンドマウスの実装", 信学技報 Vol.99, No.710, PRMU99-270, pp.65-72, 2000.
- [3] M. Kourogi, T. Kurata, K. Sakaue, and Y. Muraoka, "A Panorama-based Technique for Annotation Overlay and Its Real-time Implementation," In Proc. ICME 2000, 2000. (to appear)
- [4] 興梠, 蔵田, 坂上, 村岡, "パノラマベースドアニメーション - 全方位センサーによる大量コンテンツ生成の簡便化とジャイロセンサーを用いた注釈提示の安定化?", 第6回画像センシングシンポジウム (SSII2000), 2000. (to appear)
- [5] T. Kurata, J. Fujiki, M. Kourogi, K. Sakaue, "A Fast and Robust Approach to Recovering Structure and Motion from Live Video Frames," Proc. CVPR 2000, 2000. (to appear)
- [6] 蔵田, 興梠, 藤木, 坂上, "ウェアラブルカメラを用いた高速でロバストな対象物体の3次元復元", 第13回人工知能学会全国大会, pp. 103-104, 1999.
- [7] P. Milgram and F. Kishino, "A taxonomy of mixed reality visual display," IEICE Trans. on Information and Systems, vol.E77-D, No.12, pp.1321-1329, 1994.
- [8] Y. Ohta and H. Tamura Eds., Mixed Reality? Merging Real and Virtual Worlds, Ohmsha & Springer-Verlag, 1999.
- [9] T. Okuma, K. Kiyokawa, H. Takemura and N. Yokoya, "An augmented reality system using a real-time vision based registration," Proc. ICPR 98, pp.1226-1229, 1998.
- [10] 大隈, 竹村, 横矢, "拡張現実感システムのための画像からの実時間カメラ位置姿勢推定", 信学論 D-II, Vol.J82-DII, No.10, pp.1784-1792, 1999.
- [11] A. State, M. A. Livingston, W. F. Garrett, G. Hirotsu, M. C. Whitton, E. D. Pisano and H. Fuchs, "Technologies for augmented reality systems: Realizing ultrasound-guided needle biopsies," Proc. SIGGRAPH 96, pp. 439-446, 1996.
- [12] U. Neumann and Y. Cho, "A self-tracking augmented reality system," Proc. VRST 96, pp. 109-115, 1996.
- [13] K. N. Kutulakos and J. R. Vallino, "Calibration-free augmented reality," IEEE Trans. Visualization and Computer Graphics, Vol.4, No.1, pp.1-20, 1998.
- [14] M. J. Swain, and D. H. Ballard, "Color Indexing," International Journal of Computer Vision, Vol.7, No.1, pp.11-32, 1991
- [15] 村瀬, and V. V. Vinod, "局所色情報を用いた高速物体探索? アクティブ探索法-", 信学論 D-II Vol.J81-D-II No.9 pp.2035-2042, 1998.
- [16] Y. Ohta, T. Kanade, T. Sakai, "Color Information for Region Segmentation," Computer Graphics and Image Processing, Vol.13, pp.224-241, 1980.
- [17] 出口, "射影幾何学による PnP カメラ補正問題の統一的解法", コンピュータビジョン '90 シンポジウム, pp.41-50, 1990.
- [18] 中沢, 中野, 小松, 齋藤, "画像中の特徴点に基づく実写画像とCG画像との動画像合成システム", 映像情報メディア学会誌, Vol.51, No.7, pp.1086-1095, 1997.
- [19] R. M. Haralick, C.-N. Lee and K. Ottenberg, "Analysis and solutions of the three point perspective pose estimation problem," Proc. CVPR '91, pp.592-598, 1991.
- [20] 横小路, 菅原, 吉川, "画像と加速度計を用いたHMD上での映像の正確な重ね合わせ", 日本バーチャルリアリティ学会論文誌, Vol4, No.4, pp.589-598, 1999.