

ボリュームデータの細線化とグラフマッチングを用いた 事例ベース人体姿勢推定

田中秀典[†] 中澤篤志[†] 町田貴史[†] 竹村治雄[†]

[†] 大阪大学大学院 情報科学研究科

あらまし 特殊なマーカや計測装置の装着を必要としないビジョンベースの人体姿勢推定（モーションキャプチャ）手法が数多く研究されている。しかし、これまで複数カメラを用いた研究において人体部位の接合等によるトポロジ変化に対応できる手法はほとんど提案されていなかった。我々はこの問題に対し、入力された人体形状データをグラフに変換し、あらかじめ用意したデータベースと比較することで、入力形状データの人体部位対応問題を解決し、安定的に姿勢推定を行う手法を提案する。まず、複数のカメラから人の動きを撮影し、視体積交差法によってボリュームデータを求め、トポロジや連結関係を維持しながら細線化処理を行う。次に、得られた線図形（スケルトン）の各部分と人体部位とを対応づけるために、スケルトンを属性付きグラフで表現し、あらかじめ用意したモデルグラフデータベース（MGDB）内のグラフと対応づける。MGDBには、人体の取りうるさまざまなトポロジの事例がグラフ化されて収められており、グラフの各ノードと人体の部位があらかじめ対応付けられているため、これらとグラフマッチングを行うことで入力スケルトンの人体部位を推定できる。得られた部位情報からスケルトンを適切に多関節に近似することで、人体の関節位置および角度を得ることができる。8台のカメラで人体を撮影し動作実験を行い、従来手法では困難だった体のトポロジが変化する姿勢に対しても、本手法が正しく推定を行うことが確認できた。

キーワード 姿勢推定, モーションキャプチャ, ボリュームデータ, 細線化, グラフマッチング, 事例ベース

Example Based Approach for Human Pose Estimation using Volume Data and Graph Matching

Hidenori TANAKA[†], Atsushi NAKAZAWA[†], Takashi MACHIDA[†], and Haruo TAKEMURA[†]

[†] Graduate School of Information Science, Osaka University

Abstract In this paper, we propose a novel marker free motion capture method by using volume data. A volume data is reconstructed from multiple camera views at each frame through visual hull based method. Then, the thinness process is performed to identify the structure of the volume. Here, we can get the model skeleton graph in which a body and limbs are expressed as the nodes, and links express the connectivity between them. We compare the acquired graph and the graphs in the Model Graph Database (MGDB) and find the most similar one. The MGDB contains the example graphs which express the human body postures. Because the nodes of the MGDB graph are labeled according to the body portions, we can know the portions of input graph (skeleton) from the graph matching result. Finally we fit the skeleton and human body portions' model by using the identification results of the body portions. The experiment result shows the validity of our approach.

Key words Pose Estimation, Motion Capture, Volume Data, Thinning, Graph Matching, Example Based Method

1. はじめに

人体の姿勢・動作を復元するモーションキャプチャ手法は、グラフィックスやVR、医療、ロボティクス、デジタルアーカイブ等の幅広い分野において利用されている。従来の手法が、光学マーカや磁気マーカなどを装着しなければならなかったのに対し、近年ではマーカの装着を必要とせず、画像のみから人

体姿勢復元を行う方法が提案されている [1], [2]。単一あるいは複数のカメラから得られた画像から、人物の三次元空間中の位置および関節角度を推定する。観測環境の様々な制約により、用いる画像特徴量や手法、推定する人体パラメータの自由度などは異なる。

単一の画像を用いた手法として、画像内の人物領域（シルエット）の形状をモデル化し、入力画像を比較する手法 [3] ~

[6] が代表的だが、本来的に二次元画像から三次元データを復元するという不良設定問題であるため、かなりよい初期推定値が必要であり、オクルージョンや追跡の失敗時に復帰できないという問題がある。すなわち、従来のモーションキャプチャシステムと同様の性能を得ることはできない。これを解決するために、複数視点からの画像を用いる方法も提案されている。高橋らは複数の視点から対象人物を撮影し、シルエット画像の輪郭の特徴点から人体の姿勢を推定する手法を提案している [7]。正面、側面、上面の 3 方向にカメラを設置した三眼視システムを用いて撮影した人物画像のシルエットの輪郭情報を解析し、自己回帰モデルを用いて特徴点を追跡する。しかし、長時間のオクルージョンによって撮影されない特徴点の追跡がうまくいかない、腕が胴体等に接触して分離できないときは肌色情報を用いなければならない等の制約がある。米元らはリアルタイムでの姿勢推定を行うため、肌色領域を特徴点として追跡し、与えられた多関節モデルと対象のシルエットを比較することで姿勢を得ている [8]。しかし、特徴点が限られているという制約上、体運動の自由度に制約があり、任意の姿勢やトポロジ変化に対応することができない。

特徴点を用いた上記手法に対し、複数視点の画像から三次元ポリウムデータを復元し、ここから姿勢を推定する試みも行われている。Mikić らの手法 [9] は、人体の各部位を楕円球や円筒等の CAD 図形で近似し、体全体を多関節モデルとして表す。関節のパラメータを変化させながら、モデルとポリウムデータをフィッティングする。また、胴体に腕が密着した姿勢や両足の密着した状態に対しても、階層的な複数のモデルを用いることで対応できる。しかしこの階層的モデルは木構造をしているため、頭に体が触れた場合や両手を接触させた場合などのループのある姿勢には対応できない。またポリウムデータとモデルのフィッティングには大きな計算コストがかかり、追跡に失敗した場合の復帰が難しい。Caillette らの手法 [10] は、ポリウムデータを色を基準に blob 群に近似し、それらを木構造を持つ人体モデルにフィッティングし姿勢を推定する。特に体と四肢が接近する姿勢において、色情報を用いることによって正しく姿勢推定を行えることが示されている。一方で前者と同様、モデルの運動自由度が大きい場合フィッティングにコストがかかり、初期値設定の問題や追跡の失敗に対する復帰が困難であるという問題がある。

これに対し、ポリウムデータから直接関節角を推定する試みも行われている。Chu らの手法 [11] は、ポリウムデータの各点を Isomap 空間 [12] に投影する。Isomap 空間は、データの隣接性を考慮した測地距離を用いてデータを別の空間に投影するため、屈曲した腕や脚も Isomap 空間中では直線状となる。これを利用して、Isomap 空間中でデータを分割して体の各部位を求め、それらを元の空間で再分割することで関節モデルを得ている。この手法は、各フレームの処理がほぼ独立しており、またポリウムから直接多関節構造を見つけるボトムアップアプローチをとっているため、与えられた多関節モデルとマッチングする手法に比べ、初期値の設定が必要なく、動作の安定性や追跡失敗からの復帰についても優れているといえる。一方で、

ポリウムデータの Isomap 空間への変換に時間がかかり、またデータの隣接性を利用しているため、手と手、手や体等、体部位同士がつながった場合（トポロジ変化）に対応できないという問題がある。

我々の提案手法は Chu らの手法と同様、得られたポリウムデータから直接多関節構造を導出するアプローチを取るが、計算コストのかかる Isomap 空間への変換を行わず、入力ポリウムデータを直接細線化し構造として把握する手法を用いる。また、彼らの手法では不可能であった、人体の様々なトポロジ変化に対応するために、事例ベースの手法を導入する。まず事前に、人のとりうる様々な姿勢をグラフ化しモデルグラフデータベース (MGDB) として用意しておく。入力データも同様にグラフ化し、グラフマッチングにより構造の最も近い MGDB 内の候補を選び出すことで、入力データのトポロジおよび体部位の判定を行う。この判定結果から、入力データの各部分を部位に応じて適切な数で分割することで、人体の関節位置を得ることができる。本手法により、従来手法では不可能だった以下の点を実現している。

初期値問題と処理の安定性 ポリウムデータをボトムアップ的に処理するため初期値が不要であり、安定的に推定が行える。**処理の高速化** ポリウムデータを直接細線化しグラフ化するため、多関節マッチング法や Isomap を使う手法に比べ、処理が高速である。

人体のトポロジ変化への対応 手と手の接合や手と胴体の接合など、人体のとりうる様々なトポロジをあらかじめ事例として保持し、入力データとグラフマッチング手法により比較することで、人体の様々な構造変化に対応可能である。

以降各章にて、提案手法の概要と詳細、実験結果を述べ、終章にてまとめを述べる。

2. 提案手法

提案手法の流れを図 1 に示す。本手法は、多視点画像からのポリウムデータ復元と細線化処理、入力データのグラフ化と MGDB を用いたトポロジおよび人体部位の判定処理、細線化データと部位判定結果から関節角度を推定する処理の 3 つに分けられる。以降、これらの手法の詳細を述べる。

2.1 多視点画像からのポリウムデータの復元

対象人物の画像を天井に固定された複数のカメラから撮影し、視体積交差法によって時系列ポリウムを求める。本実験環境では、8 台のカメラが人物を周囲より見下ろす環境を設定した。各カメラは、平面パターンを用いてレンズの非線形歪みを校正し、次に既知の大きさのキャリブレーションボックスを用いて、カメラパラメータを求めておく [13], [14]。入力画像から色相を考慮した背景差分を用いることで対象の領域を抜き出し、カメラパラメータを用いて対象の三次元ポリウムデータを復元する。得られた結果を図 2 に示す。

2.2 ポリウムデータの細線化

ポリウムデータは対象人物の概形を表現するが、そのままでは人体の方向や構造を明確に把握することは難しい。姿勢を解析する上で、この体積を持ったデータを線図形に変換するこ

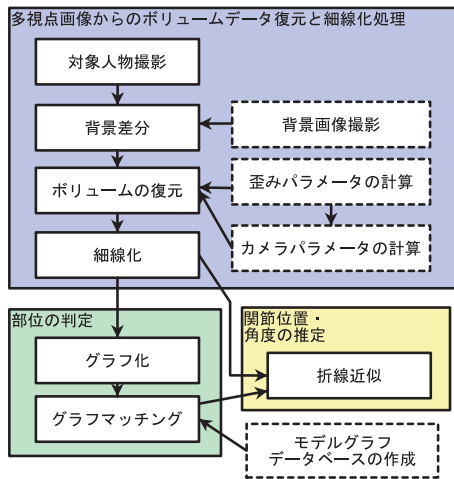


図1 提案手法の流れ



図2 復元された人体のボリュームデータ



図3 細線化によって得られた線図形(スケルトン)の例

とができれば、方向・長さ・接続関係などが明らかになり、構造の解析が容易になる。そこで前節で得られたボリュームデータを、ボクセルからなる3次元2値画像とみなし、細線化処理を行う。細線化の結果得られる図形(スケルトン)は、交差点、空洞等の特殊な場所を除いて太さが1であり、端点以外には消去可能な画素を含まない。また、原図形のトポロジを保存し、その中心を通るものとする。

我々は、これらを満たす手法として、斉藤らによって提案された一連の手法を用いている[15]。まず、入力となるボリュームデータにはノイズが含まれるため、ボリューム全体に対して膨張・収縮処理を行い、微細な突起や穴を取り除く。次に、ボリュームデータに2乗ユークリッド距離変換を行う。これは、存在する各ボクセルに対し、もっとも表面に近いボクセルまでの距離の二乗を与えるものである。すなわちユークリッド距離は、その画素の物体内での「深さ」を表している。次に、以下の手順によって余分なボクセルを除去し、細線化されたボクセル列を得る。

- (1) 最小の距離値を持つ境界画素の集合を近傍画素の配置により分類。
 - (2) 同一種類の境界画素から消去可能画素を逐次的に消去。
 - (3) 各々の種類の境界画素について処理2を行う。
 - (4) 処理1~3を、消去可能画素が存在する間繰り返す。
- ここでは、点の消去において原図形のトポロジを保存される

かを判定する処理が必要である[16]。この手法は、図形の境界画素の距離値の小さい画素から順に消去することで結果の線図形が図形の中心に位置することが期待でき、図形の回転による影響が少ない。また、境界画素のグループ分けはその画素の近傍の非0画素の個数に基づいて行われる。近傍の非0画素が少ない画素のグループから順に処理を行うことにより、小さな構造から偶発的に得られる不要な線図形(ヒゲ)の発生を抑制している。ボリュームデータに細線化を行った例を図3に示す。

2.3 人体部位の判定

細線化によって得られたボクセル列の接続性を評価するとスケルトンの構造が得られる。次にスケルトンのどの部分(枝)が体のどの部位に相当するかを判定する。本節では、まず細線化によって得られるスケルトンが様々なトポロジを持つことを述べ、グラフマッチングを用いて腕/足/頭/胴体等の人体部位とスケルトンの対応を把握する方法について述べる。

2.3.1 スケルトンのトポロジ

前節で述べたようにスケルトンは太さ1の線図形であり、2次元画像の線図形と同様の特徴点を考えることができる。すなわちスケルトン中の画素は図4のように交差点、端点、接続点の三つに分類される。端点は人体部位の末端部分を表し、交差点は胴体につながっている腕/足/頭の付け根を表す場合と、体が他の部位と接触して穴があるためにできるものが考えられる。また、交差点と端点に区切られた互いに隣り合う接続点の集合(枝)が各部位の位置と向きを表す。

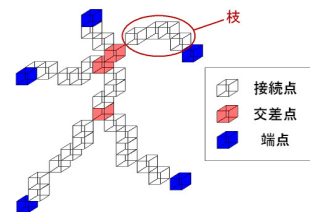


図4 スケルトン中の特徴点と枝

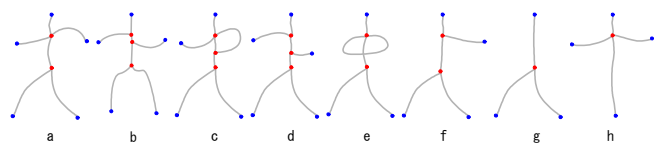


図5 様々なトポロジを持つスケルトン

図5aのように、スケルトンの一つ一つの枝が四肢が頭や胴と離れており、両腕の付け根が一つの交差点で表される場合は、各枝は腕/足/頭/胴体に1対1で対応する。またその接続関係から、どれが胴を表すかは明らかである。しかし一般的な姿勢を考えると、体の一部が触れ合うことで、他のトポロジを持つスケルトンが得られることがある。このような場合、それぞれの枝は人体部位の一部または全部、もしくは複数の部位に対応することになる(図5-c,d,e,f,g,h)。また、人物が同じ姿勢をとっても、細線化の際に交差点の位置関係が変化する場合がある(図5-a,b)。さらに、スケルトンには元のデータから起因するノイズやヒゲなど、本来の人体姿勢とは無関係な構造が得ら

れる場合もある．このように，人体のスケルトンデータは各種の要因により，様々なトポロジに変化する．

2.3.2 MGDB を用いた部位の判定

以上の問題に対処するため本手法では，細線化の結果生じると予想できる様々なトポロジを表すグラフを，あらかじめ多数用意しておき，モデルグラフデータベース (MGDB) として保存しておく．入力されたスケルトンはグラフ化され，MGDB 内のモデルグラフ群と比較される．最も類似すると判断されたモデルグラフから，入力スケルトンのトポロジを判別できる．また，モデルグラフの各枝には人体の部位情報も保存されているため，グラフマッチングによる枝同士の対応結果から，入力スケルトンの枝の人体部位判定を行う (図 6)．この手法により，人体のトポロジが変化した場合もボリュームデータと人体部位との対応が得られる．また，スケルトンをグラフに変換して扱うことで情報量や計算量が削減できる．グラフは人体の位置や姿勢に不変な表現であるため，あらかじめ準備するモデルグラフ個数は，トポロジの個数程度に削減できる．

a) スケルトンのグラフ表現

スケルトンのグラフ表現は，その各枝をグラフのノードとして表現し枝の連結関係をエッジと表すことで行われる．各ノードには，スケルトンの全長で正規化した枝の長さ，枝に属する (正規化された) ボリュームの体積が，属性値として与えられる (図 7)．

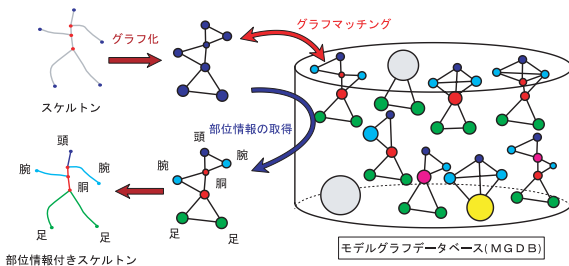


図 6 MGDB を用いた部位の判定

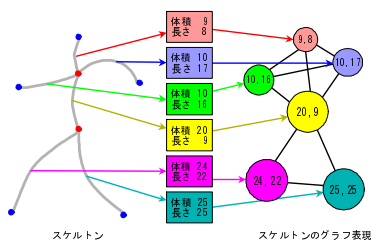


図 7 スケルトンのグラフ表現と属性値

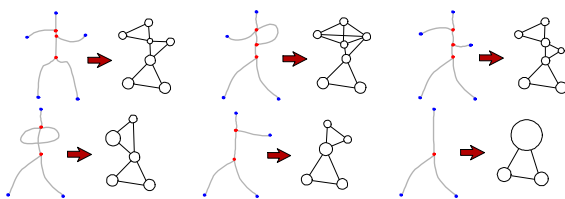


表 1 人体部位情報ラベルと関節数

FULL(4)	HEAD(2)
BODY(1)	LOWBODY(1)
ARM_A(3)	ARM_B(3)
LOWARMA(1)	LOWARMB(1)
LEG_A(3)	LEG_B(3)
HEAD_BODY_ARMS(2)	HEAD_UPARM(3)
HEAD_UPARMS_BODY(2)	HEAD_BODY_UPARMS(2)
BODY_ARM(1)	BODY_UPARM(1)
BODY_LEGS(3)	ARM_C(6)
UNLABELED(1)	

b) MGDB の作成

MGDB は，実際に撮影された人体のボリュームデータや CAD ソフトウェアによる人体の CG データから得たボリュームデータをグラフ化し，人体部位および属性値をラベル付けることで作成される．人体部位情報として用いるラベル，及びそれに対応する部位の途中に現れる関節数を表 1 に示す．

c) グラフマッチング

本手法におけるグラフは画像処理のノイズによるトポロジ変化や人体形状の個人差等による属性値の違いなどが考えられるため，エラー訂正を考慮したマッチングが行われる必要がある．そのため，Edit-Distance [17] に基づくグラフマッチング手法を用いる．これは，比較する 2 つのグラフを一致させるための編集操作列 (削除，追加，属性変更) のうち，もっともコストが最小であるものを，マッチングの評価値として使用する手法である．グラフマッチング問題は NP 完全であるが，グラフのノード数が小さいこととモデルグラフデータベースを部分グラフ群に分割し比較する Messmer らの手法 [18] を用いることで，比較的高速に処理が行える．各編集操作には，操作の種類，対象ノードの属性値の差等によりコストを定義する必要があり，本手法では以下の式で与える．

$$\text{cost}(\text{del_node}(n)) = n.\text{volume}$$

$$\text{cost}(\text{del_edge}(e)) = 0.1$$

$$\text{cost}(\text{sub_node}(n1, n2)) = k_1 |n1.\text{volume} - n2.\text{volume}| + k_2 |n1.\text{length} - n2.\text{length}|$$

$$\text{cost}(\text{sub_edge}(e1, e2)) = 0$$

なお，ノード n に対して $n.\text{volume}$ は体積の属性値を， $n.\text{length}$ は長さの属性値を表す．エッジは属性を持たせていないので，属性変更のコストを 0，削除のコストを 0.1 と定数にした．ノードの属性変更にかかるコストは，それぞれのノードに対応する部位同士の非類似性を表す．ここでは体積と長さ両方の差を考慮し，その線形和でコストを表現する． k_1, k_2 の係数である．また，ノードの体積属性が大きいものほどボリュームデータ内で大きな構造をもつものから作られたノードであると考えられるため，ノードの削除には体積に比例するコストを与える．

2.4 関節位置・角度の取得

得られたスケルトンの各枝と，人体部位の対応情報に基づき，

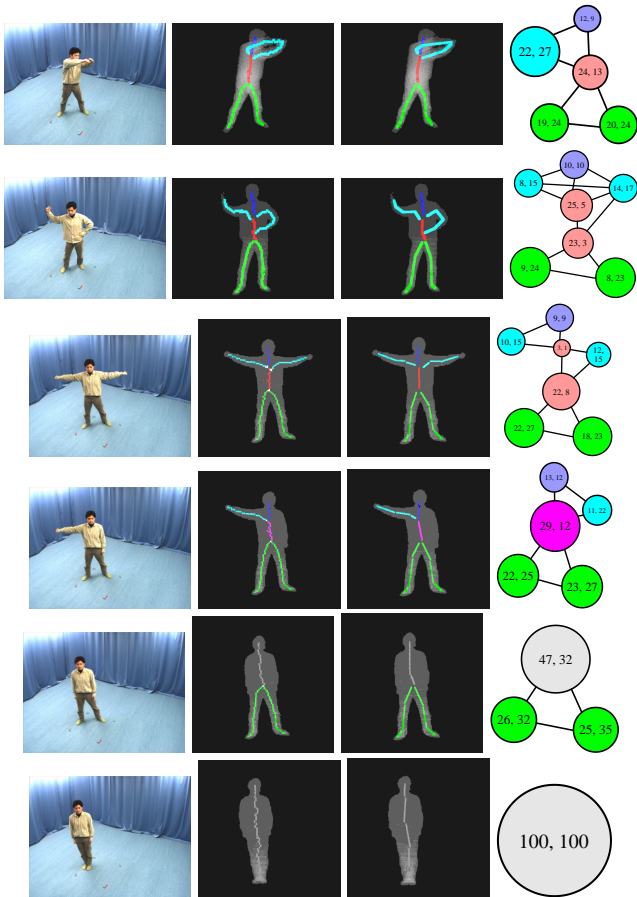


図 10 入力画像と姿勢推定結果 左より人物画像，部位判定されたスケルトン，関節推定結果，モデルグラフ

する状況が生じる．これは，本手法のみならず他の手法でも問題になる状況であるが，カメラ台数を増やしたり，配置を工夫する，ボリューム復元手法を改善することで解決できる．また，細線化の際にできる「ヒゲ」がスケルトンの構造を変えてしまう誤り例や，よく似た部位同士（腕が短く復元された場合の腕と頭等）の対応を誤る例が見られた．これに対しては，ヒゲの出やすいケースはあらかじめ MGDB に用意しておく方法や，現在定数としているマッチングコストの係数を変化させる等の工夫が必要であると考えられる．特に，マッチングコストの係数に関しては，誤りサンプルから学習によって最適値を得る手法を導入することを検討している．腕上部が胴体と接する場合や，脚の上部が互いに接する場合，関節が短く推定されるなどのケースもいくつか見られた．これに関しては [10] のように，関節の長さに対する知識，関節位置の時系列的な推定を用いる必要があると考えられる．

4. ま と め

本研究では，人物の姿勢を推定するために人体の形状データを線図形に変換し，その長さや角度を元に関節の角度を推定する手法を提案した．人体の構造は様々なトポロジに変化する可能性があるが，提案手法では，人のとりうる様々な構造事例をデータベースとして準備し，それらと入力データをマッチングすることで，人体の構造や部位の判定を行う手法を提案してい

る．また，実験により様々なトポロジを持つ人体形状に対して部位の対応が正しく判定されることが確かめられた．

今後の課題としては，部位の判定性能の向上や，関節位置推定手法の改善などがあげられる．具体的には，ボリュームデータのトポロジ判定に適したグラフ属性とマッチングアルゴリズムの検討，モデルグラフデータベースの最適化，時系列の情報を用いた推定精度の向上等が挙げられる．これにより，姿勢推定の安定性や精度をより向上させるだけでなく，人体の左右の判定などの，より高度な認識も行うことができると考えている．

文 献

- [1] T.B. Moeslund and E. Granum: "A Survey of Computer Vision-Based Human Motion Capture," CVIU, Vol. 81, No. 3, pp. 231-268 (2001)
- [2] D.M. Gavril: "The Visual Analysis of Human Movement: A Survey," CVIU, Vol. 73, No. 1, pp. 82-98 (1999)
- [3] C.R. Wren, A. Azarbayejani, T. Darrell and A.P. Pentland: "Pfinder: Real-Time Tracking of the Human Body," IEEE Trans. on PAMI, Vol. 19, No. 7, pp. 780-785 (1997)
- [4] 左藤明知, 川田聡, 大崎喜彦, 山本正信: "多視点動画画像からの人間動作の追跡と再構成", 信学論 D-II, Vol. J80-D-II, No. 6, pp. 1581-1589 (1997)
- [5] K. Rohr: "Towards Model-Based Recognition of Human Movements in Image Sequences," CVGIP: Image Understanding, Vol. 59, No. 1, pp. 94-115 (1994)
- [6] C. Sminchisescu and B. Triggs: "Kinematic Jump Processes For Monocular 3D Human Tracking," Proc. of CVPR 2003, pp. 69-76 (2003)
- [7] 高橋和彦, 坂口竜己, 大谷淳: "実時間非接触非装着型 3 次元人物姿勢推定法に関する一考察", 信学論 D-II, Vol. J83-D-II, No. 5, pp. 1305-1314 (2000)
- [8] 米元聡, 有田大作, 谷口倫一郎: "多視点動画画像処理による実時間全身モーションキャプチャシステム", 映像情報メディア学会誌, Vol.54, No.3 (2000)
- [9] I. Mikić, M. Trivedi, E. Hunter and P. Cosman: "Human Body Model Acquisition and Tracking using Voxel Data," IJCV, Vol. 53, No. 3, pp. 199-223 (2003)
- [10] F. Caillette and T. Howard: "Real-Time Markerless Human Body Tracking with Multi-View 3-D Voxel Reconstruction," Proc. of BMVC, Vol. 2, pp. 597-606 (2004)
- [11] Chi-Wei Chu, O. Chadwicke Jenkins and Maja J Mataric: "Markerless Kinematic Model and Motion Capture from Volume Sequences," Proc. of CVPR 2003, Vol. 2, pp. 475-482 (2003)
- [12] J. B. Tenenbaum, Vin de Silva, and J. C. Langford: "A Global Geometric Framework for Nonlinear Dimensionality Reduction", Science 22 December 2000: 2319-2323 (2000)
- [13] 加藤博一, M. Billingham, 浅野浩一, 橋啓八郎: "マーカー追跡に基づく拡張現実システムとそのキャリブレーション", 日本バーチャリアリティ学会論文誌, Vol. 4, No. 4, pp. 607-616 (1999)
- [14] 井口征士, 佐藤宏介: "三次元画像計測", 昭晃堂 (1990)
- [15] 齋藤豊文, 森 健策, 鳥脇純一郎: "ユークリッド距離変換を用いた 3 次元デジタル画像の薄面化及び細線化の逐次型アルゴリズムとその諸性質", 信学論 D-II, Vol. J79-D-II, No. 10, pp. 1675-1685 (1996)
- [16] 鳥脇純一郎: "3 次元デジタル画像処理", 昭晃堂 (2002)
- [17] R. Ambauen, S. Fischer and H. Bunke: "Graph edit distance with node splitting and merging, and its application to diatom identification," Proc. of International Workshop on GBRPR, pp. 95-106 (2003)
- [18] B.T. Messmer and H. Bunke: "A New Algorithm for Error-Tolerant Subgraph Isomorphism Detection," IEEE Trans. on PAMI, Vol. 20, pp. 493-504 (1998)