ステレオカメラを用いた手表面追跡

豊浦　正広† 　 Matthew TURK††

† 山梨大学大学院 医学工学総合研究部　〒 400–8511 山梨県甲府市武田 4–3–11
†† Department of Computer Science, University of California, Santa Barbara
Santa Barbara, CA 93106-5110, U.S.A.
E-mail: †mtoyoura@yamanashi.ac.jp, ††mturk@cs.ucsb.edu

あらまし　本研究では，ステレオカメラによる手の三次元位置および姿勢の獲得のための手法を提案する．本研究の特徴は，多数のカメラではなくステレオカメラで観測することによって，移動環境下での手の位置・姿勢の獲得を実現することにある．手は関節が多く，自己隠蔽が起こりやすい．カメラ画像上での手の位置・姿勢を獲得するためには，多数のカメラを用いる必要があった．これに対し本研究では，ステレオカメラを用いて手表面の三次元的な特徴点の並びを抽出することで，手の三次元位置および姿勢の獲得を行う．手の姿勢に依らず表面パターンの局所的な三次元構造が変わらないことを利用して，手の姿勢に不変な特徴量を計算し，モデルとのマッチングを行う．これにより，部分ごとの位置および姿勢の獲得を可能にする．姿勢不変特徴量としては，色や SIFT などの特徴点自体が持つ特徴量と，近傍となる特徴点との相対位置関係を組み合わせたものを用いる．本研究では，既知のパターンを持つグローブをユーザに装着させることで画像特徴点を与えた．実験結果から，グローブ上の特徴点が手の姿勢に依らずに識別できることを示す．

キーワード　姿勢不変特徴量，ステレオカメラ，ドットパターングローブ，拡張現実感，モーションキャプチャ．

3D Hand Tracking with Stereo Cameras

Masahiro TOYOURA† and Matthew TURK††

† Interdisciplinary Graduate School of Medical and Engineering, University of Yamanashi
Takeda 4–3–11, Kofu, Yamanashi, 400–8511 Japan
†† Department of Computer Science, University of California, Santa Barbara
Santa Barbara, CA 93106-5110, U.S.A.
E-mail: †mtoyoura@yamanashi.ac.jp, ††mturk@cs.ucsb.edu

Abstract　We propose a method for extracting 3D position and posture of hands with stereo cameras. The main contribution of our research is that our method enables to track the position and posture in mobile environments by using stereo cameras. The hand is often observed with many occluded regions, since the hand has many joints. In previous methods, many cameras are required to extract the position and posture. In this research, the position and posture of the hand are estimated from 3D alignment of feature points on the surface of the hand. Stereo cameras enable to extract 3D alignment of the feature points. The position of each feature point is identified by matching with the model. Even if the surface is deformed, the local alignment of the feature points is not drastically changed. Each feature point is identified with pose-invariant feature that is a combination of the feature of the points and the relative position to the neighboring feature points. In this research, the feature points are given by wearing gloves with a known pattern. Experimental results show that the position of the feature points on gloves can be tracked in stereo images, which is not dependent on the posture of the hand.

Key words　Pose-inverient feature, stereo cameras, dot pattern glove, augmented reality, motion capture.

1. はじめに

カメラ画像上で人間の手を追跡できれば，人間の手を拡張現実感に利用することができる．手の動きに合わせて仮想物体を提示すれば，手は仮想世界とのインタフェースとなる．二次元的なマーカ [1] ，顔 [2] ，手 [3] などから得られる平面に合わせて仮想物体を提示する研究はこれまでになされてきたが，三次元的な情報を持つ仮想物体に対し，つかんだりはさんだりといった直接操作ができるような三次元的なインタフェースは提案され

てこなかった．

そこで本研究では，画像から人間の手の位置および姿勢を抽出することを目的とする．用いるカメラは 2 台とし，拡張現実感の実現のために移動環境でも利用可能となるようにした．得られる結果は，従来のジェスチャ認識 [4] や画像上の手領域抽出 [5] によって得られる離散的な関節角の集合やパターンの識別，平面的な領域とは異なる．

自由度の高い手に対して，少数台のカメラで位置・姿勢を獲得することは容易ではない．手は関節が多く，自己隠蔽が起こりやすいためである．カメラ画像上で多関節物体を獲得するためには，多数のカメラを用いる必要があった [6]．または，磁気センサやデータグローブなどを組み合わせて用いることで，カメラ画像上での手の位置・姿勢を求める手法も提案されてきた [7]．しかし，指の三次元姿勢を求めるような複雑な問題に対しては，これらの手法は適用できない．加えて，これらの装置を移動環境下で用いることは難しかった．

一方で，自由度の高い変形表面を追跡する研究に，既知パターンを持つ衣服の追跡を行うものがある．多数のカメラで得られる画像からの衣服の追跡をする手法が提案されてきた [8] [9] [10] ．衣服には，格子 [8]，三角形 [9]，ドット [10] などのパターンが与えられ，それぞれの構造は既知である．パターンを多数のカメラで観測することでモデルグラフを得る．多数のカメラで衣服を観測し，それぞれの画像上でデータグラフを抽出し，モデルグラフとのマッチングを行うことで変形表面の追跡を実現する．マッチングは，色および特徴点の並びに基づいて行われる．これらの手法では，規則並んだパターンがほぼ完全に観測できることが必要となる．パターンの抽出にパターン配置の規則性を用いるためである．よって，手のように自己隠蔽が観測されやすい対象にはそのまま適用できない．また，このような観測環境を持ち運ぶことは通常できない．

我々は衣類の追跡と同様，手に既知のドットパターンを持つグローブを装着し，これをステレオカメラで観測することで手の追跡を実現する．今回のパターンには，簡単な画像処理で抽出が可能なドットパターンを採用した．ステレオカメラを用いることで，パターンの三次元位置を抽出することができる．手の姿勢変化が起こってもドットパターンの局所的な三次元構造が変わらないことを利用して，手の姿勢に不変な特徴を抽出し，マッチングを行う．

将来的にカメラの解像度が十分に得られ，手表面で特徴点を密に抽出できるようになれば，既知のパターンを与えることなくシステムの実現が可能である．特徴点の特徴量と局所構造を手掛かりに，特徴点の同定を行う本研究の枠組みは，色特徴点以外の SIFT などにもそのまま応用できるためである．

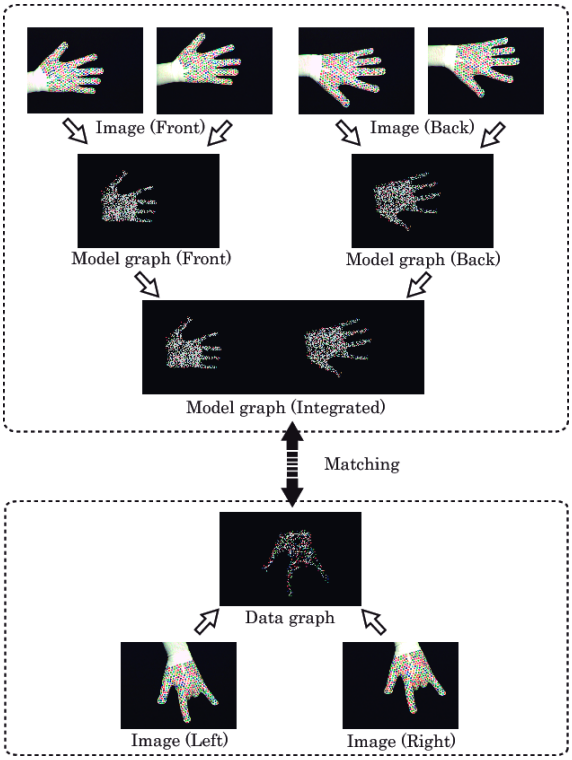


図 1　処理の流れ

2. 手 法 概 要

グローブに付与するパターンには，ドットパターンを採用した．ドットパターンは，格子パターン [8] や三角形パターン [9] に必要となるような高度な画像の領域分割を必要としないためである．グローブを白，背景を黒に設定する場合には，それぞれのドットは，鮮色を持つ小さな領域として，容易に抽出することができる．それぞれのドットの色およびサイズの設定方法については，3.1 で述べる．図 1 に提案手法の処理の流れを示す．

ステレオ画像からのグラフ作成

ステレオカメラで観測されるドット領域から，ドット間の対応を求める．ステレオカメラは校正済みであり，画像間で対応する領域が与えられると，その対応する領域の三次元位置を求めることができる．色が近く，エピポーラ拘束を十分に満たす 2 つのドット領域を対応する領域とみなす．それぞれのドット領域の重心位置に，その領域の平均色を持つノードを配置する．

モデルグラフの作成

モデルグラフ作成のためには，ステレオカメラでグローブを裏表の両面から観測して画像を得る．画像から，ドット領域とグローブ領域を抽出する．抽出手法については，3.2 で述べる．

次に，一定距離内にあるノード間にパスを作成する．パスを持つ 2 つのノードは，隣接関係を持つものと定義する．ただし，グローブ領域外に渡るようなパスは削除

する．これによって，異なる指の間に渡るようなノードの間に隣接関係が構築されないようになる．以上により，裏表のそれぞれについて，色と隣接関係を持つモデルグラフができる．裏表のグラフは，それぞれのグラフ間にパスは作成せずに，1 つのグラフとして扱う．

データグラフ上のドットの識別

データグラフの構築は，モデルグラフと同様に，互いに近いノードとパスを作成することで行う．パスを持つ 2 つのノードは，隣接関係を持つ．

データグラフとモデルグラフのマッチングを求めることで，ステレオ画像上の各ドット領域が，モデル画像上のどのドット領域に対応するのかを識別することができる．ドットの識別の手法は，4. で説明する．

3. ドットパターングローブ

3.1 ドットパターンの構成

ドットの色は，HSV 表色系における H の値によって表現する．H の値は照明環境の影響を受けにくいためである．予備実験を行い，設定環境で判別可能な色の数を調べた．実験環境では，同一ドットの中で H の値に 15 までのぶれが見られたので，$H = 30n(n = 0, \cdots, 11)$ の色を用いることとした．ただし，黄色 ($H = 120$) は輝度が高く，グローブ領域との識別がつかなかったため，黄色は使用しないこととした．また，グローブは白とした．

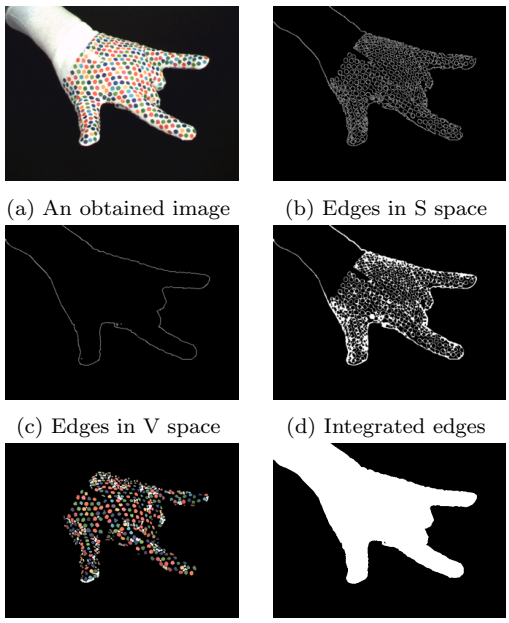
ドットのサイズは，ステレオカメラから観測される最小のサイズであることが望ましい．ドットサイズは小さく，かつ，ドット間の距離も小さいほうが，手の表面に多くのドットを配置することができるためである．ドットの数は，表面の特徴点密度を意味する．多くのドットが一度に観測される方が，マッチングによるドット識別が容易になるので，ドットは可能な限り多く配置する．ステレオカメラを目元に設置することを想定して，ユーザの手が観測されるであろう 40cm あたりに焦点距離を設定する．画角もこれに合わせて調整する．このときに識別可能なドットのサイズとして，半径 2.5mm を採用した．ドットはアイロンプリント可能な用紙を用いて作成した．

また，本研究では，背景は黒とし，グローブ領域が容易に抽出できる環境を設定した．将来的には，鮮やかさを示す S の値などを用いることによって，任意の背景下でもグローブ領域とドット領域が抽出可能であると予想される．肌色領域の抽出が参考となる [3].

3.2 ドットの抽出

グラフの作成手順は，モデルグラフとデータグラフに共通である．HSV 表色系でグラフは抽出される．画素 p の HSV 成分を (h_p, s_p, v_p) と書き表す．h_p にドット識別のための信号を含み，s_p および v_p には領域分割のための信号が含まれる．画像には，ドット領域 R_D，グローブの下地領域 R_G，背景領域 R_B が含まれる．S 値および V 値の閾値 s_{th}，v_{th} を用いると，それぞれの領域は以下のように書き表すことができる．

$$
\begin{aligned}
R_D &= \{p \mid s_p > s_{th}\} \\
R_G &= \{p \mid s_p \leqq s_{th}, v_p > v_{th}\} \\
R_B &= \{p \mid v_p \leqq v_{th}\}
\end{aligned}
$$



(a) An obtained image (b) Edges in S space
(c) Edges in V space (d) Integrated edges
(e) Extracted dot regions (f) Whole glove region

図 2　ドットグローブパターンの領域抽出

グラフの作成のためには，まず，ノードとしてのドット領域を抽出する．図 2(a) に示すような観測画像から，画素値に S の値を持つグレースケール画像を作成する．この画像上で Canny エッジを抽出すると，図 2(b) に示す画像が得られる．このエッジは，R_D と R_G の境界を示すことになる．ここで注意したいのは，このエッジにはドット領域 R_D と背景領域 R_B の境界は含まれないことである．

ドット領域 R_D と背景領域 R_B の境界を求めるために，画素値に V の値を持つグレースケール画像も作成する．この画像上で Canny エッジを抽出すると，図 2(c) に示す画像が得られる．このエッジは，R_D と R_G とを合わせた領域と，R_B の境界を示すことになる．2 つの種類の境界を足し合わせて，ドット領域 R_D とそれ以外の領域との境界を得ることができる．エッジが途切れないように，画像に膨張と収縮を数回施して得られる画像が図 2(d) である．領域ラベリングによって小さく閉じられた領域を得る．得られる領域のうち，$s_p > s_{th}$ を満たす領域をドット領域として抽出する．図 2(e) が得られるドット領域である．ステレオ画像のそれぞれで得られるドット領域 R_D から，色が近く，エピポーラ線上にある点をグラフのノードとして求める．ノードは，三次元位置と画像上での色を情報として持つ．

次に，隣接関係を表すパスを作成する．三次元空間上で一定距離内にあるものを隣接関係のあるノードとして定義し，パスを作成する．このとき，画像上のグローブ領域を参照し，グローブ領域外に投影される部分があるようなパスは削除する．グローブ領域 $R_D \cup R_G$ は，S の値に閾値を設け，膨張と縮退を繰り返すことで得る (図 2(f))．手に関しては，グローブ領域外にまたがるパスを持つ 2 つのノードは別々の指にあり，その間の三次元距離は変わると想定されるためである．

4. ドットの識別

4.1 姿勢不変特徴量

ドットの識別は，**(1) ドットの色**，**(2) 隣接するドットの色**，**(3) 隣接するドットとの距離**に基づいて行う．ドットの識別は，データグラフに含まれるひとかたまりの部分モデルとモデルグラフとのマッチングを行う処理として実現される．ここでは，ドット間の三次元距離がグローブの伸縮によってもほぼ変わらないと仮定している．手の姿勢変化が起こってもドットの色とドット間の相対的な位置関係は変わらない．色と距離を手の姿勢に不変な特徴として抽出し，マッチングを行う．この仮定の下で，モデルグラフのノード i がデータグラフのノード j に対応する確率 p_{ij} を計算する (図 3)．

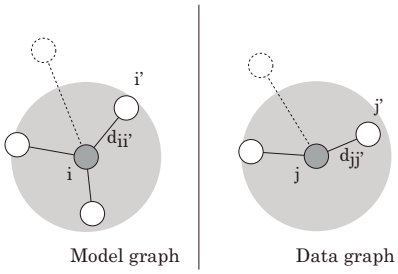


図 3　姿勢不変特徴量によるドット識別

文献 [9] には，ドットの持つ情報量の計算方法が与えられている．本研究ではドットの色の数は 11 色としたので，1 つのドットは $log_2 11 = 3.5[bits]$ の情報量を持つことになる．グローブ上にはおよそ 1000 個のドットが配置されているので，それぞれの点を一意に特定するには $log_2 1000 = 10.0[bits]$ の情報量が必要となる．それぞれのドットが N 個の近傍を持っているとし，それぞれの近傍に区別がないのであれば，あるドットとその近傍が与える情報量は $(N+1) \times 3.5 - log_2 N[bits]$ となる．$N \geqq 3$ であれば，1000 個のドットを一意に特定できることになる．

従来手法のように，画像上のドットの並びで近傍を識別する場合には，それぞれの近傍に区別をつけることは難しかった．アフィン変形されたパターンを観測する画像からは，近傍である以上の情報を抽出することができなかったためである．それに対して，我々はあるドットから三次元距離 d_{th} にあるドットを近傍とみなし，ドット間の距離によってどの近傍であるかを判定する．これにより，特徴点に規則的な並びを要求する必要がなくなり，将来的に高解像度画像が利用できるようになれば，手表面の SIFT 特徴量などに対して，本手法を適用することが可能である．ただし，近傍との距離は不変であると仮定できるように，十分に近いドットだけを近傍に指定する必要がある．

従来研究において p_{ij} は以下のように算出された [9]．モデルグラフ上のドット i の色を c_i とし，i の近傍であるノードを $i' \in \mathcal{N}^M(i)$ とする．同様に，データグラフ上のドット j の色を c_j とし，j に隣接するノードを $j' \in \mathcal{N}^D(j)$ とする．また，色 c_i と c_j の距離を $dst(c_i, c_j)$ とすると，p_{ij} は定数 σ_c を用いて以下のように表すことができる．

$$p_{ij} = c_{ij} \prod_{j' \in \mathcal{N}^D(j)} \max_{i' \in \mathcal{N}^M(i)} c_{i'j'} \tag{1}$$

$$c_{ij} = exp\left(-\frac{dst(c_i, c_j)^2}{2\sigma_c^2}\right)$$

従来研究で提案されている p_{ij} の算出方法 [9] では，画像上での隣接関係を利用しており，近傍関係が確定的に抽出できないときには，この近傍ノードはドットの識別に利用しないという方針を採用している．また，近傍との距離は考慮されない．

これに対して，本研究では，三次元空間中の距離を用いている．これにより，ノード間の隣接関係を 2 台のカメラからでも効率的に求めることができる．

ii' 間の距離を $d_{ii'}$ とする．同様に，jj' 間の距離を $d_{jj'}$ とする．我々の提案する三次元近傍情報を含んだ p_{ij} は，定数 σ_d, d_{th} を用いて以下のように表すことができる．

$$p_{ij} = c_{ij} \prod_{j' \in \mathcal{N}^D(j)} \max_{i' \in \mathcal{N}^M(i)} \varphi(d_{ii'}, d_{jj'}) c_{i'j'} \tag{2}$$

$$\varphi(d_{ii'}, d_{jj'}) = exp\left(-\frac{||d_{ii'} - d_{jj'}||^2}{2\sigma_d^2}\right)$$

p_{ij} は，i と j の色が一致し，$i' \in \mathcal{N}^M(i)$ と $j' \in \mathcal{N}^D(j)$ が等距離に同じ色として観測されるときに，最大値を取る．適当な閾値 d_{th} を設定することで，画像上で隣接する領域以外の領域も参照して，確率を計算する．

4.2 モデルマッチングによるドット識別

得られる p_{ij} から行列 P を作成する．行列 P から対応付けは winner-takes-all のアルゴリズム [10] によって得られる．データグラフ上のあるノードは，モデルグラフ上で高々1 つのノードと対応が与えられる．すでに対応点を持つノードには，他に対応するノードは与えれられない．このアルゴリズムは，モーションキャプチャシステムにおいて，各画像で得られるマーカ位置を統合するのによく用いられる．手順は以下のとおりである．

（1） $(i,j) = \underset{i,j}{\operatorname{argmax}}\ p_{ij}$ を満たす (i,j) の組を得る．
（2） $p_{ij} > 0$ であれば，(i,j) を対応する組として登録する．そうでなければ，処理を終了する．
（3） $\forall k\ p_{kj} \leftarrow 0$ ，$\forall l\ p_{il} \leftarrow 0$.
（4） (1) から (3) を繰り返す．

5. 実験結果

ステレオカメラの解像度は 640×480 であった．2 つのカメラは 1 枚のプレートの上に取り付けられ，あらかじめ校正した．$s_{th} = 30$，$v_{th} = 200$，$d_{th} = 20[mm]$，$\sigma_c = 7.5$，$\sigma_d = 200[mm]$ とした．

図 5 にドットの識別結果を示す．図 5(a) がモデルグラフ作成のためのステレオ画像のセットである．この画像から得られたモデルグラフを基に，図 5(b) から (d) の各フレームで得られるデータモデルの各ドットに対する識別を行った．図 4 に，データグラフ上で識別されたドットのうち，正しいドットとして識別されたドットの割合を示す．我々の提案する φ の有効性を示すために，φ を用いた場合 (式 (2)) と，φ を 1 として近傍であることのみを使った場合 (式 (1)) の割合をそれぞれ示した．

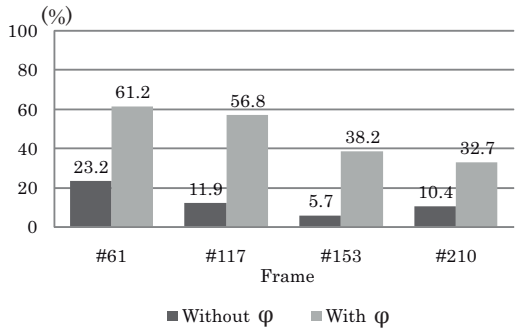


図 4　ドット識別結果

いずれのフレームにおいても，φ を用いた場合に識別率が高くなっており，φ の導入が有効であることが確かめられた．しかし，自己隠蔽の多いフレームにおいては，ドットの識別率が低い．特に指先においては，ドットの識別率が低いことが図 5 から見て取れる．これは，指先では近傍となるドットの数が十分でないためである．これを解決するためには，ドットのサイズを小さくしてドットの密度を上げ，これを抽出できるようにカメラの解像度も上げる必要がある．

6. まとめ

本研究では，手表面の特徴点を，**(1) 特徴点自体の色**，**(2) 近傍の特徴点の色**および **(3) 近傍特徴点との距離**によって識別した．ステレオカメラによって，それぞれの特徴点の三次元位置を抽出した．グローブ上のドットは，我々が提案する三次元的な近傍情報を含む特徴量によって，従来の特徴量を用いたときよりも精度よく抽出することができた．近傍情報を用いた特徴量は，手の姿勢に不変な特徴量であるといえる．

実験では，ドットの識別が完全ではなかった．しかし，手の三次元位置および姿勢を特定するためには，必ずしもすべてのドットが正しく識別されている必要はない．手の多関節モデルがあれば，これをデータにフィッティングさせることで位置・姿勢推定ができると考えられる．この識別率で手の三次元位置・姿勢が推定できるかどうかは，今後の研究で明らかにしなければならない．

識別率の向上のためには，サイズを小さくしてドットの密度を上げて，十分な数の近傍が観測される必要がある．また，この小さなドットを抽出できるようにカメラの解像度も上げる必要がある．将来的には，手のテクスチャを使ってグローブなしでの手の位置・姿勢を獲得したい．

文　献

[1] H. Kato, M. Billinghurst, I. Poupyrev, K. Imamoto, and K. Tachibana, “Virtual object manipulation on a table-top ar environment,” Proceedings of the International Symposium on Augmented Reality (ISMR2000), pp.111–119, 2000.

[2] J. Pilet, V. Lepetit, and P. Fua, “Fast non-rigid surface detection, registration and realistic augmentation,” International Journal of Computer Vision, vol.76, no.2, pp.109–122, 2008.

[3] T. Lee, and T. Höllerer, “Initializing markerless tracking using a simple hand gesture,” Proceedings of the IEEE/ACM International Symposium on Mixed and Augmented Reality (ISMAR), pp.259–260, November 2007.

[4] H. Guan, J.S. Chang, L. Chen, R.S. Feris, and M. Turk, “Multi-view appearance-based 3d hand pose estimation,” Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshop, pp.154-159, 2006.

[5] B. Stenger, A. Thayananthan, P.H. Torr, and R. Cipolla, “Model-based hand tracking using a hierarchical bayesian filter,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol.28, no.9, pp.1372–1384, September 2006.

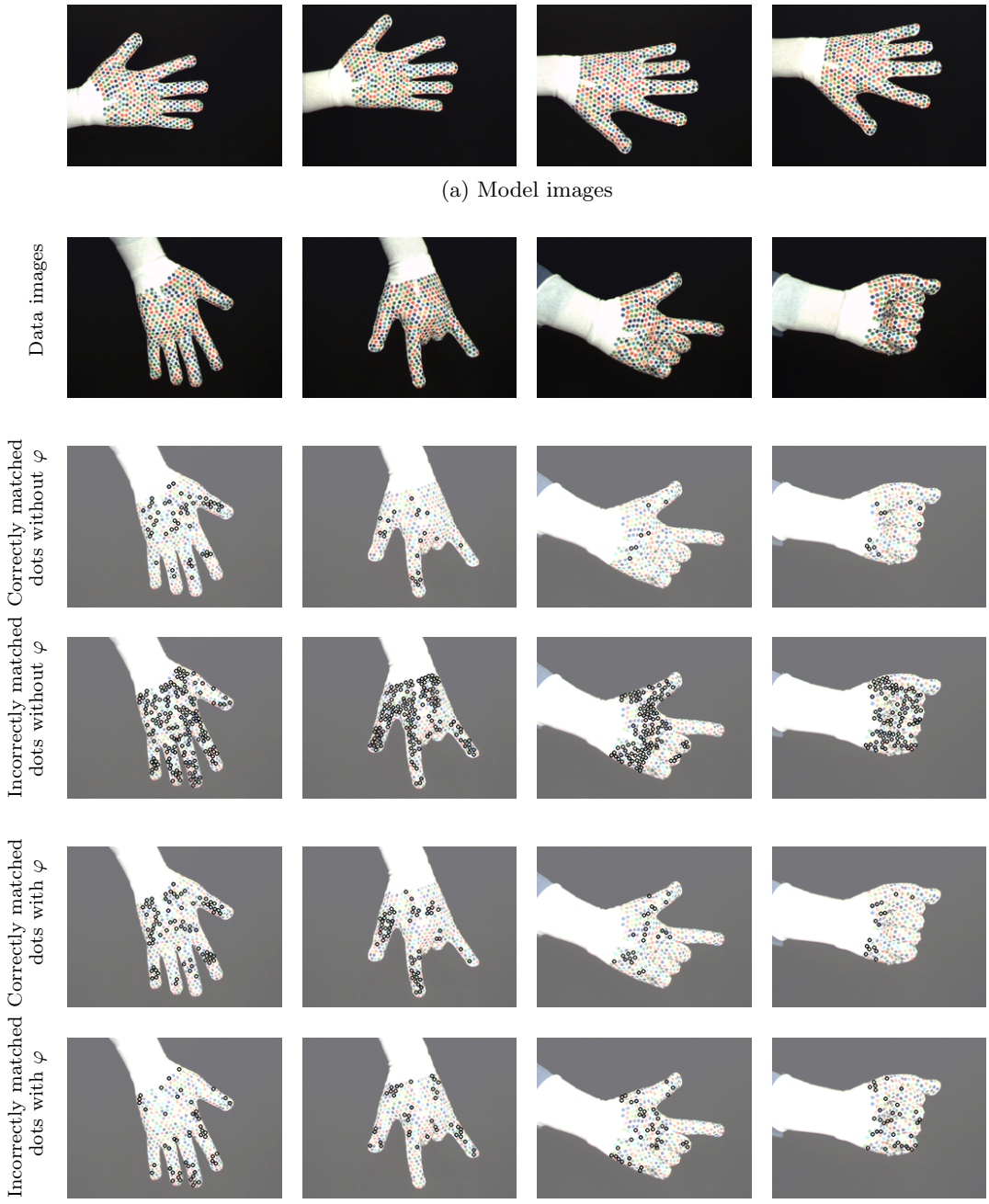
[6] J. Starck, and A. Hilton, “Surface capture for performance-based animation,” IEEE Computer Graphics and Applications, vol.27, pp.21–31, 2007.

[7] M. Minoh, H. Obara, T. Funatomi, M. Toyoura, and K. Kakusho, “Direct manipulation of 3d virtual objects by actors for recording live video content,” Second International Conference on Informatics Research for Development of Knowledge Society Infrastructure (ICKS’07), pp.11–18, January 2007.

[8] I. Guskov, S. Klibanov, and B. Bryant, “Trackable surfaces,” Proceedings of the ACM SIGGRAPH / Eurographics symposium on Computer animation, pp.251–257, 2003.

[9] R. White, K. Crane, and D.A. Forsyth, “Capturing and animating occluded cloth,” Transaction on Graphics, vol.26, no.3, 2007, Article 34.

[10] V. Scholz, T. Stich, M. Keckeisen, M. Wacker, and M. Magnor, “Garment motion capture using color-coded patterns,” Computer Graphics Forum, vol.24, no.3, pp.439–448, August 2005.



(a) Model images

(b) 61st frame. (c) 117th frame. (d) 153rd frame. (e) 210th frame.

図 5　データグラフとモデルグラフのドット識別結果．(a) モデルグラフを生成した 2 組のステレオ画像．(b)∼(e) ドット識別結果．1 段目はデータグラフを生成するステレオ画像のうちの 1 枚．2 段目，3 段目は φ を用いずにドットを識別した結果．2 段目は正しく識別されたドット，3 段目は正しく識別されなかったドット．4 段目，5 段目は φ を用いてドットを識別した結果．4 段目は正しく識別されたドット，5 段目は正しく識別されなかったドット．