ドットパターングローブを用いた画像からの手の三次元位置同定

3D Hand Position Tracking from Images with Dot Pattern Glove

豊浦 正広 † Matthew TURK‡
Masahiro TOYOURA† and Matthew TURK‡

† 山梨大学大学院医学工学総合研究部 ‡ カリフォルニア大学サンタバーバラ校コンピュータ科学部
† Interdisciplinary Graduate School of Medical and Engineering, University of Yamanashi
‡Department of Computer Science, University of California, Santa Barbara

E-mail : †mtoyoura@yamanashi.ac.jp, ‡mturk@cs.ucsb.edu

1. はじめに

手の動きに合わせて仮想物体を重畳表示できれば，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インタフェースは提案されてこなかった．

我々は，手の三次元位置および姿勢に合わせた仮想物体提示を実現するために，画像からこれらを抽出することを目指している．仮想物体の重畳表示は現実世界を観測して得られる画像に対して行われる．磁気センサで得られる三次元位置やデータグローブで観測した姿勢を仮想物体重畳の目的で使う場合には，得られる位置・姿勢と観測画像との位置合わせが必要となるが，観測画像から手の三次元位置および姿勢を獲得するときにはこれが不要となる．このことは単に処理時間を減らすだけでなく，仮想物体提示位置のずれを減らすことにも貢献する．

本論文では手の三次元位置・姿勢獲得の前段階として，手表面の三次元位置を画像情報のみから同定することを目的とする．用いるカメラは 2 台とし，移動環境でも利用可能となるようにする．本研究で獲得される手の三次元位置は，三次元空間上で手が存在する領域として表現される．これは，従来のジェスチャ認識 [4] や画像上の手領域抽出 [5] によって得られる関節角の集合や平面位置とは異なる．

手を少数台のカメラで位置および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．カメラ画像から自己隠蔽の起こりやすい物体の位置や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 [6]．多数のカメラを持つシステムを移動環境下で用いることは難しい．

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 [7] [8] [9]．衣服には，格子 [7]，三角形 [8]，ドット [9] などのパターンが与えられ，それぞれの構造は既知としている．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，特徴点を画像から抽出するために，規則正しく並んだパターンを要求する．このために，既知パターンなしでの手法の拡張は難しい．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，手に既知のドットパターンを持つグローブを装着させ，これをステレオカメラで観測することで，手表面位置の獲得を実現する．手表面位置の獲得のためには，特徴点自体の特徴量と，特徴点同士の三次元的な配置を用いるので，将来的にカメラの解像度が高くなり，手表面に識別可能で密な特徴点を抽出できるようになれば，ドットパターンを与えることなくシステムを実現することが可能である．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．

2. ドットパターンの設計

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．グローブ領域は白とした．

ドットのサイズは，ステレオカメラから観測でき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したい．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した．画角もこれに合わせて調整した．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. ドットパターンの抽出

あらかじめ獲得するモデルグラフと，各フレームで獲得されるデータグラフとのマッチングを求めることで，データグラフ上の各ドットが，モデルグラフ上のどのドットに対応するのかを識別する (図 1).

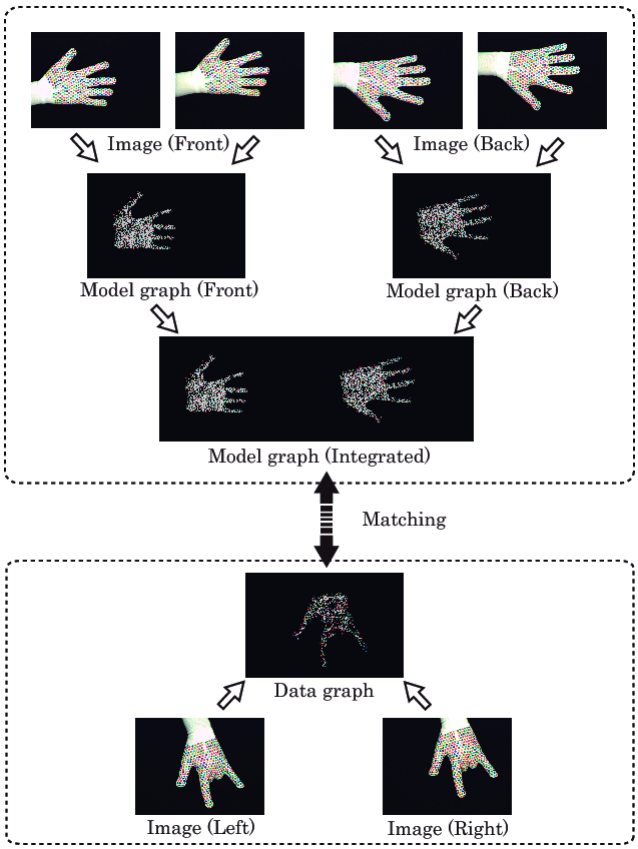


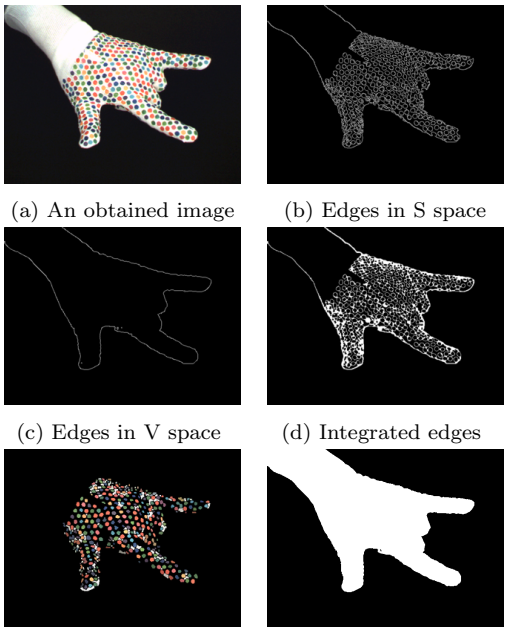
図 1　ドットパターンマッチング

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与え，各ドットの三次元位置を求める．色が近く，十分にエピポーラ拘束を満たすドット間に対応を与える．各ドット領域の三次元重心位置に，そのドット領域の平均色を持つノードを配置する．

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のそれぞれを観測して得られる 2 組のステレオ画像から，2 つのグラフを作成する．データグラフでは，各フレームの観測から片側からのみのグラフが得られる．観測画像には，ドット領域 R_D，グローブの下地領域 R_G，背景領域 R_B が含まれることになる．

HSV 色空間において，画素 p が (h_p, s_p, v_p) の値を持つとするとき，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条件で抽出できる．

$$R_D = \{p \mid s_p > s_{th}\}$$



(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　観測画像からのドット抽出

$$R_G = \{p \mid s_p \leq s_{th}, v_p > v_{th}\}$$
$$R_B = \{p \mid v_p \leq v_{th}\}$$

S 値と V 値を持つような画像を観測画像 (図 2(a)) から作成し，それぞれエッジ画像を求め (図 2(b)，(c))，得られるエッジを統合して膨張・縮退を行うことで (図 2(d))，面積の小さく S 値が大きな領域がドット領域として抽出できる (図 2(e)).

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 (図 2(f)) の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．

モデルグラフでは，裏表のそれぞれについて，色と隣接関係を持つ 2 つのグラフができる．これらのグラフは，それぞれのグラフ間にパスは再生成せずに，1 つのグラフとみなし，マッチングが行われる．

4. ドットパターンの識別

ドットの識別は，(1) ドットの色，(2) 近傍のドットの色，(3) 近傍のドットとの三次元距離に基づいて行われる．ここでは，(1)~(3) が手の姿勢の変化によってもほぼ変わらないことを仮定している．ドット間の距離が十分に近いときには，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ことが期待できる．このとき，(1)~(3) 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．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} を計算できる．

従来手法 [8] 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 [8]．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．

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

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

従来研究で提案されている p_{ij} の算出方法 [8] 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る．これにより，2 台のみのカメラからでも効率的にノード間の隣接関係を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 を用いて以下のように表すことができる．

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

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

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．

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 [9] 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. $(i, j) = \underset{i,j}{\text{argmax}}\ p_{ij}$ 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．

図 3 に，4 つのフレームのデータグラフ上で，識別されたドットが正しいドットとして識別された割合を示す．我々の提案する特徴量の有効性を示すために，式 (2) において特徴量に三次元的な距離の成分 φ を含めた場合と，φ を含めない場合の割合をそれぞれ示した．

いずれのフレームにおいても，φ を含めた場合に識別率が高くなっており，三次元的な距離の成分が識別のために有効に働くことが確かめられた．また，自己隠蔽の多い姿勢に対しては，ドットの識別率が低く，特に指先において，ドットの識別率が低かった．これは，指先では各ドットの近傍となるドットの数が十分に得られないためである．これを解決するためには，各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．

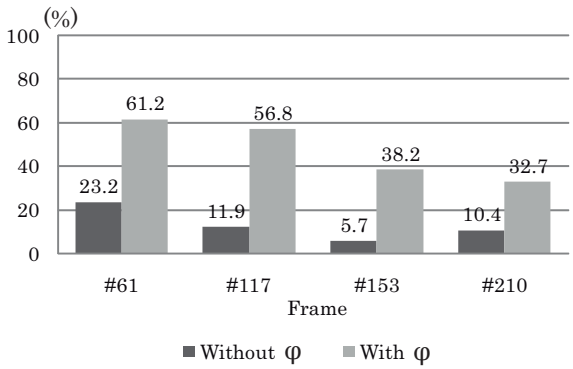
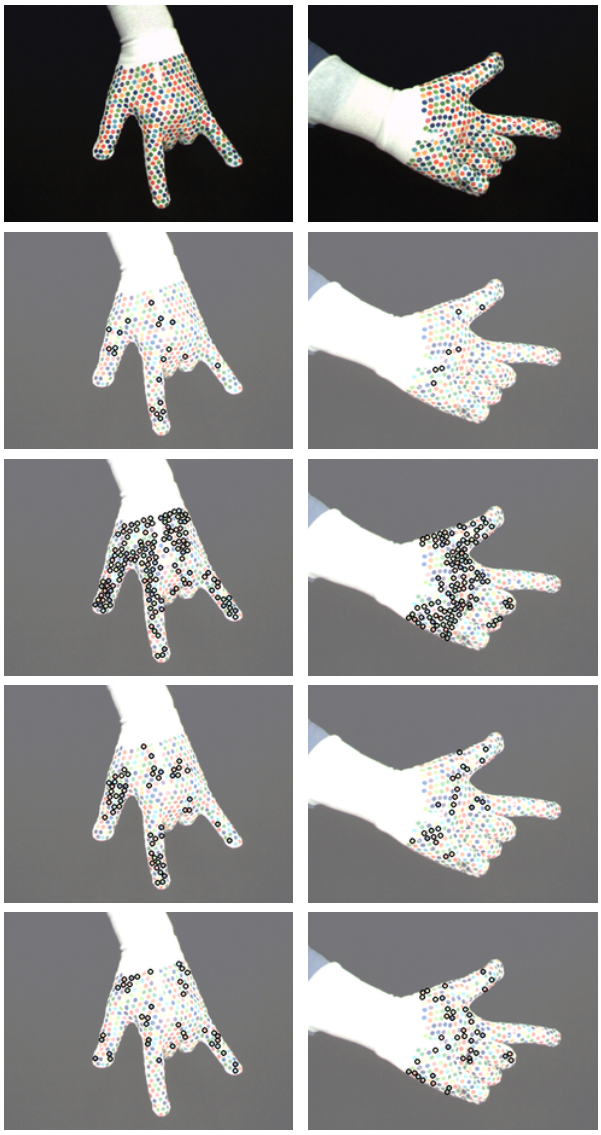


図 3　認識結果

6. まとめ

本研究では，ドットパターングローブを作成し，三次元的な近傍情報を含む特徴量によって，グローブ上のドットを識別する手法を提案した．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．これらの特徴量は，手の姿勢に不変な特徴量であるといえる．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区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



(a) 117th frame. (b) 153rd frame.

図 4　データグラフとモデルグラフのドット識別結果．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．

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．

また現在，ドットパターンのマッチングに 1 フレームで 2 分ほどかかっている．処理時間のほとんどは，画像上でのドット位置を正しく抽出するための領域分割に費やされている．候補となるドットの三次元位置のリストが得られた後の処理は，ラップトップ PC の処理によっても 30fps の更新速度を確保できている．領域分割やステレオ処理の処理を簡単化したり，ハードウェアによる処理に持ち込むことで，実時間処理が可能になると見込んでいる．

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

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

[9] 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.