

距離画像を入力とするハンドリングのための モデル駆動型衣類状態推定

○松川徹(筑波大学) 喜多泰代(産業技術総合研究所)

A deformable model driven method using a range image for handling clothes

*Tetsu Matsukawa, Graduate School of University Of Tsukuba,

Yasuyo Kita, National Institute of Advanced Industrial Science and Technology(AIST)

Abstract — For the purpose of handling clothes by a robot, we are studying a deformable model driven method which estimate cloth state by matching the predicted shape of the target clothes with its observed shape. Though our previous work have shown good potential of this strategy[3], it did not have enough generality and robustness for practical applications. To improve this point, we start considering 3D matching between 3D predicted shape and range data obtained by a trinocular stereo vision. In this paper, we report our system using a real-time trinocular stereo vision[4] and a cloth simulation software Maya[5] and preliminary experimental results using this system.

Key Words: Object Recognition, Deformable Object, Range Image, Cloth Simulation, Handling Clothes

1. はじめに

家庭内では、衣服や紙等の変形を生じる物体が溢れており、これらの物体をロボットが的確に扱うためには柔軟物体の形状を認識できる視覚機能が必要となる。

最近では、家庭用ロボットでの利用を想定した衣類のハンドリングに関する研究がいくつか報告されている。金子らは対象輪郭の幾何学的特徴量に基づいた展開動作計画を立てている[1]が、日常環境において物体輪郭を厳密に抽出することは困難であると思われる。大澤らは、洗濯・乾燥後の数種類の衣類を対象として、ロボットによる展開と分類を同時に行い、作業台上に乗せ、道具を利用して折りたたむシステムを開発している[2]が、展開開始状態として数回の持ち替え操作により少数の形状に収束させた状態を用いるため、余分な持ち替え動作が必要となっている。さらに、これらの手法では個別の衣類に対し人手により作業計画を立てる必要があり、ロボットが日常環境で様々な衣類を扱うことを考えると、より汎用性の高い状態推定方法が望まれる。

一方我々は、衣類のような自在に変形する物体を観測データのみからボトムアップに認識することは困難であると考え、対象に対する大まかな知識をもとにシーンに対する予測を行い、予測と観測画像の情報を照合することで状態を推定するモデル駆動型のアプローチ(Fig1)を提案している。これまでに予測モデルとの2次元画像照合を基本とする手法で状態推定を行いこのアプローチの有望性を示した[3]。

最近では、複数台のカメラによりシーンの密な3次元情報である距離画像を実時間で生成する技術が確立されて来ている[4]。そこで、距離画像を入力として用いることにより、背景からの衣類領域抽出を容易にし、詳細な3次元情報を利用し状態推定に頑健性を持たせることを目指す。

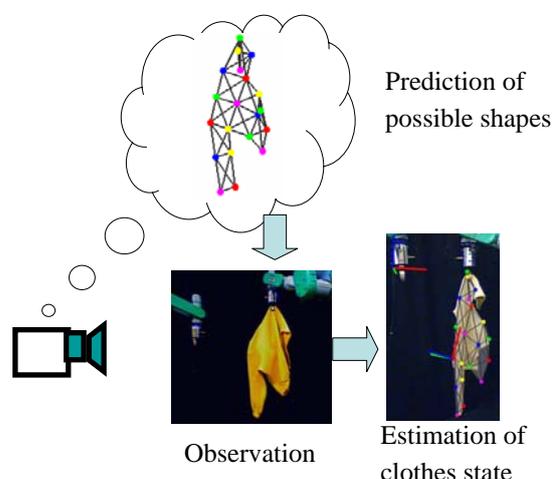


Fig.1 the concept of proposing deformable model-driven method of estimating the state of clothes

本研究報告では、マニピュレータを用いる前に吊るし上げた衣類の観測距離画像を用いて(Fig.2), 状態推定の予備実験を行ったので報告する。

2. 課題設定

本報告では、双腕マニピュレータとステレオカメラからなるハンドアイシステムにより衣類を折り畳むタスクにおけるサブゴールとして、衣類の両肩をハンドで保持し展開した状態にすることを課題とする。一方のハンドで保持した状態の衣類の特定の部位をもう一方のハンドで掴み持ち上げるという動作を基本として、この基本動作の繰り返しによりこの課題を実現する。本報告では基本動作の内、ある1点で保持された未知の衣類の状態を代表形状と照合することにより推定する手法を検討する。この状態を推定すれば、[3]で提案した方法と同様に、ハンドリングに必要な情報を算出することが可能である。

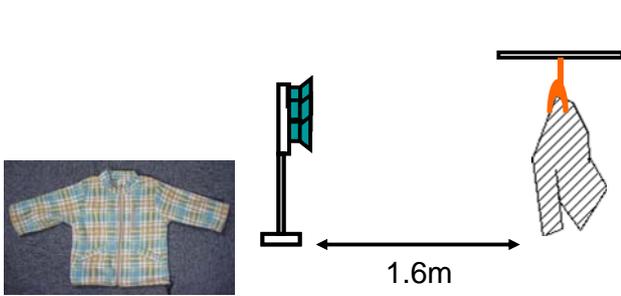


Fig.2 toddler pullover(left), environment of experiment(right)

3. システム構成と対象衣類

入力装置に用いるリアルタイム 3 眼ステレオビジョンシステム[4]は、IEEE1394 インターフェースを備え複数カメラの自動同期機能を持つ Point Grey Research Inc. の Dragonfly 3 台 (基準/右/上) のカメラを入力とし、マルチベースラインステレオと双方向視差探索を組み合わせたブロックマッチングに基づく 320×240 程度の距離画像を汎用 PC ワークステーション (Intel XEON 3.8GHz ×2) 上に 1 秒間に 30 枚生成可能である。現時点では静止した状況の衣類の状態推定を対象とし、その内の 1 フレームのみを用いる。今回の実験では、入力カメラの画角の関係上、子供用トレーナーを対象としている。(Fig.2)

衣類形状の予測には、衣服の変形シミュレーションが行える CG(ComputerGraphics)ソフトウェアの 1 つである Alias|Wavefront(現 Autodesk)の Maya4.5[5] の Cloth 機能を用いる。

4. 状態推定法

4.1 起こりえる形状の予測

対象としたトレーナーの形状の情報に基づき Fig3(a) に示す前身頃と後ろ身頃の離れることのないことを仮定した平面のモデルを Maya Cloth で生成する。このモデルは厳密な衣類の情報を与えなくとも使えるように同一種類の衣類全てに対応できる汎用モデルを目指している。本報告では、厳密ではないおおまかな形状や柔らかさ等、ある程度の衣類の情報はわかっている場合を想定し、モデルの形状、ノード点数、曲げに対する衣類の抵抗、伸縮のパラメータを手動で設定している。

このモデルを用いて、Fig3(a)の赤い点で示す縁からほぼ等間隔に選んだ全 18 個の代表点の内、ある 1 つの代表点が保持され吊るされた状況をシミュレーションする。我々が想定している保持は小面積を保有する平らなグリッブによるものであり、画像が常にグリッブ面から垂直に撮影される状況を想定している。ここで、Y 軸を重力方向に一致させた左手系 3 次元座標系を基準座標系とする。まず、対象モデルを XZ 平面に水平に Fig3(a)の状態に配置する。小面積を有するグリッブでの保持の効果を代表点に近い縁の 2 点のノードを固定することで模擬し、

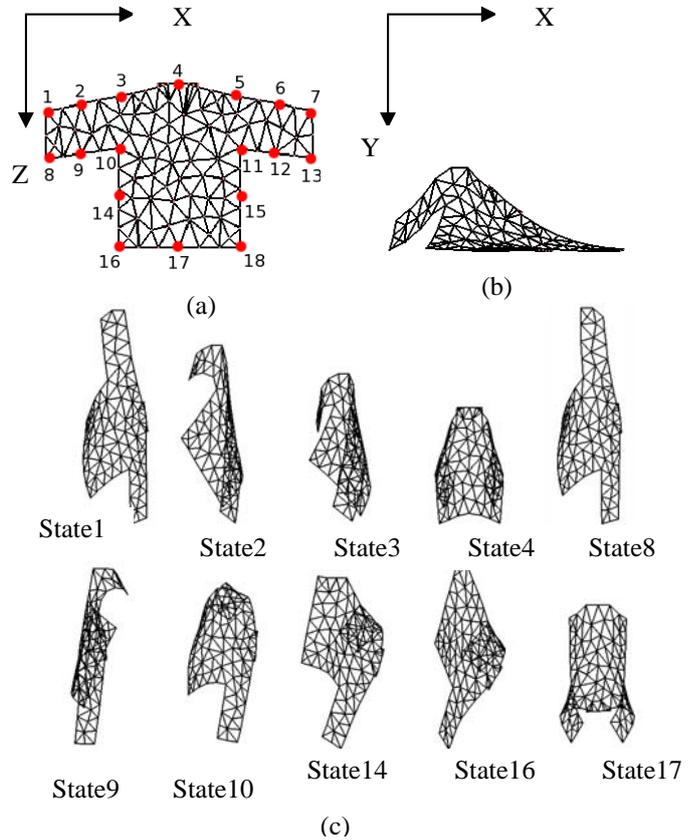


Fig.3 Model shapes: (a) common pullover model; (b) simulation process; (c) possible 3D shapes when the pullover is held at a point

他のノードに対し重力を平衡状態に達するまで働かせる。Fig3(b)はその過程を表している。

これを全 18 個の代表点に対し行い得られた 18 種類の予測形状モデル (以下モデル) に縁の近い代表点で掴み吊るされた全ての状態を代表させる (Fig3(c))。予測形状モデルの 3 次元座標とハンドの位置を maya の API を利用し作成した MEL (MayaEmbedded Language) コマンドにより、予測形状データベースに保存する。

4.2 観測画像と予測との照合

4.2.1 カメラ座標系

カメラ座標系は 3 台のカメラの内、基準カメラを原点に取り観測方向に z 軸の正、重力方向を y 軸の正、観測方向の右側を x 軸の正とした直交座標系とする。距離画像には欠損の生じていない画素(u,v)に対し、対応するカメラ座標系での 3 次元座標値(x(u,v), y(u,v), z(u,v))が格納されている。以下の座標の記述はカメラ座標系での記述である。

4.2.2 衣類領域抽出

ハンドリングを行う際、ハンドの近傍の空間に対象物体以外が存在しない状態であることは自然な設定であると考えられる。この仮定により、保持点以下一定範囲内の物体を抽出することで衣類の色やテクスチャによらず衣類領域を抽出できる。保持ハンドの 3 次元位置情報(hand.x,hand.y,hand.z)は既知であるので、 $hand.y \leq y(u,v)$ でありかつ $|hand.z - z(u,v)|$

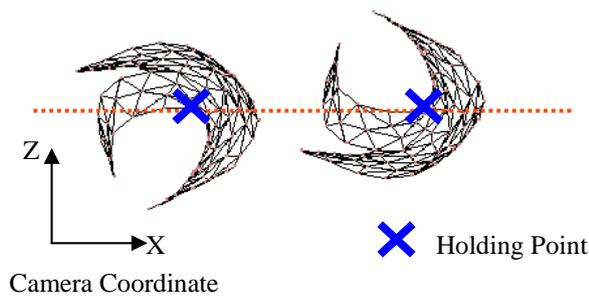


Fig.4 two possible symmetrical shapes

| ≤ 閾値 A を満たす画素(u,v)平面上の連続領域を衣類領域として入力距離画像より抽出する。衣類の広がり方を考慮して閾値 A を 300mm とした。

4.2.3 全長に基づく候補選定

距離画像より衣類領域を抽出したら、処理時間短縮のため全長情報に基づいて候補を選定する。衣類領域の最下点を (lowest.x, lowest.y, lowest.z) と表し、全長を | hand.y - lowest.y | と定義する。この全長の一定の誤差の範囲内に正しい状態が含まれている。そこで、| データの全長 - モデルの全長 | / データの全長 ≤ 閾値 B であるモデルを候補として残す。正しい状態のモデルが確実に残るように閾値 B を 0.2 とした。

4.2.3 保持位置を通る x-y 面に対象な候補

代表点を保持した場合に対し、現実には様々な状態が存在するものを 1 つの予測形状に代表させてモデルを作成したが、この予測に対しハンド保持部分を通る x-y 面に対象な形状 (以下、前後対象形と表記) があり得る (Fig.4)。ハンドリングを行う場合、第 2 ハンドが掴むべき部位の座標やハンドの進入方向はこの 2 種類で大きく異なってくる [3]。そこで、得られている 18 種類の代表形状にそれぞれの面対象な形状を算出し付加した 36 種類の形状で全ての見え方を代表することとする。

4.2.4 モデル距離画像の生成

観測距離画像とモデルを照合するためにモデルを観測距離画像と同一形式のものにして置く。ハンドの保持位置(hand.x, hand.y, hand.z)にモデル上の保持位置に対応する部分を合わせて配置し、グリップ面から垂直な方向から観測したモデルの見え方距離画像をカメラキャリブレーションにより既知である基準カメラの射影行列を用いて OpenGL のデプスバッファ等を利用し、生成する。

4.2.5 距離画像照合

各画素(u,v)において観測距離画像上での衣類領域の z 座標 (Dataz(u,v)) とモデル距離画像上の z 座標 (Modelz(u,v)) の双方の値が存在し、| Dataz(u,v) - Modelz(u,v) | ≤ 閾値 C を満たす画素(u,v)を重なり点とする。重なり点の点数を数え、次式で定義する

距離オーバーラップ比 ROR(Range Overlap Ratio)を評価値とする。

$$ROR = (\text{重なり点数}) / (\text{観測距離画像点数}) + (\text{重なり点数}) / (\text{モデル距離画像点数})$$

ROR (0~2.0) は 2 次元画像の重なり距離による制限を加えたものであり、数値が大きくなるに従い高い重なり度合いを示している。閾値 C の設定によって照合性能に違いが生じるが、実験による経験値により閾値 C を 100mm と設定した。

5 の実験では、18 種類のモデルからその前後対象形を考慮した 36 種類のモデルの内、4.2.2 の処理で残ったモデルに対し、モデル距離画像の生成、観測距離画像との照合を行い、ROR が最大のモデルを照合結果として選定する。

5. 実験

観測データの再現性を考察するために一度衣類を振った後に、4.1 で示した代表点の内、左右対称な場所を除く 10 個の代表点(1,2,3,4,8,9,10,14,16,17)に近い位置で対象衣類を保持した状態の距離画像を各々 5 枚撮影し、合計 50 枚の距離画像を入力とした。

今回の実験では、ハンドの 3 次元保持位置 (hand.x, hand.y, hand.z) は手で与えた。全ての画像で、複雑な背景下にあるテクスチャを有した衣類を保持したシーンであるが、衣類の領域抽出が行えた。入力装置の特性により若干の膨張領域や欠損領域が見られたが本研究目的としては十分な精度であると考えられる。全長に基づく候補選定により照合する 36 個のモデルを平均 19.5 個に減らすことが出来た。

Fig.5 に結果の一例(代表点 2 で保持)を示す。Fig.5(a) は基準画像の濃淡画像 (左) と距離画像 (右) を示し Fig.5(b)(c)(d) は距離画像より抽出した衣類領域とモデル距離画像との重なりを ROR が大きい順に前後対象形と共に示しており、赤で示す部分が重なり領域、青がデータのみ存在する領域、緑がモデルのみ存在する領域、紫がモデルとデータの双方が存在するが、双方の z の絶対差が閾値 C よりも大きいため重なりとしない領域を示す。実際の保持代表点 (State2) のモデルとの ROR が最も大きくなっており (Fig.5(b))、同一保持位置 (State5 内) においても前後対象形を正しく区別することが出来た。

Table.1 に全照合結果を示す。50 枚中 38 枚で正しい状態が選定された。誤選定は、代表点 2,3,9 を保持した場合に集中していた。代表点 2,3 を保持した場合には肩の折れの度合いが予測と大きく異なり、幅が変わっていることが原因であった。この例 (State3) を Fig.6 に示す。望ましいモデル Fig.6(b)(State3) では観測データより幅が狭くなり、異なるモデルである Fig.6(c) のモデル (State10) との重なりより小さくなっている。このような場合には、2

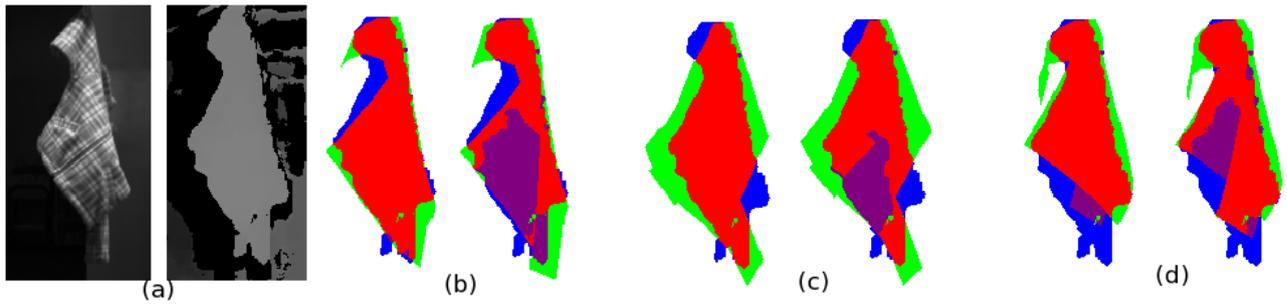


Fig.5 example of model matching (a)overbation image of State2: intensisty image(left), range image(right), (b)(c)(d)(e) matching with model range image :(b) State2(ROR=1.73,0.86), (c)State18(ROR=1.57,1.15), (d)State3(ROR=1.52,1.29)

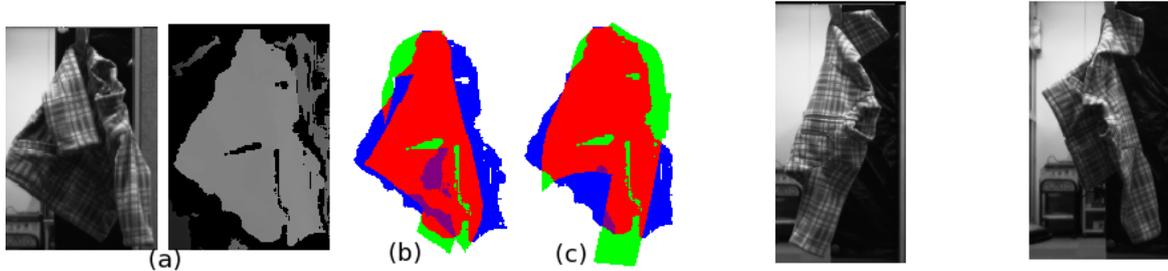


Fig.6 example of State3 (a)observation(intensity image(left), range image(right)), (b)(c)matching with model range image: (b)State3(ROR=1.40), (c)State10(ROR=1.53).

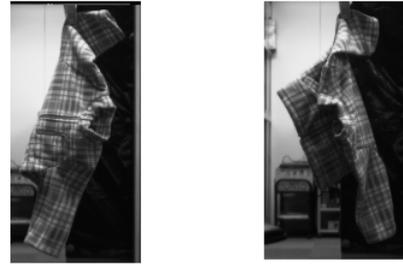


Fig.7 example of state 9

Table.1 Estimation results

Input Data No.	1	2	3	4	8	9	10	14	16	17	Sum
Success	5/5	3/5	0/5	5/5	4/5	2/5	5/5	5/5	5/5	5/5	39/50

次元画像に対してではあるが文献[3]で行っていたような幅正規化等の処理を考えるとよりある程度この違いを吸収できると考えられる。

代表点 9 を保持した場合では袖の下側の中心を保持していることにより、もう一方の袖の位置関係によって回転が引き起こされていることによりモデルが観測画像と大きく異なる形状をしていることが誤選定の原因であった (Fig.7)。今回の実験で代表点 9 に関しては 1 通りの状態ではなく、この回転により 2 通りの状態が現れる確率が高いことが確認された。そのため、代表点 9 に限ってはこの保持位置に対し 1 通りの予測では不十分であり、この回転を考慮した 2 通りの予測を行う必要があると考えられる。

6. まとめ

[3]で提案した衣類の状態をモデル駆動型に推定する手法に、実用のための頑健性を持たせるため、入力に詳細な 3 次元情報を用いることを検討中である。本報告では距離画像入力化に関する検討実験を行い次の結果を得た。

- (1)衣類近傍に他の物体が存在しないことを仮定して、背景の複雑さや衣類の色やテクスチャによらず 3 次元空間内の一定範囲を抽出することで衣類領域抽出が行えた。
- (2)距離画像照合により、多くの場合で望ましい保持

位置のモデル及び前後対象形の選定による状態推定が行えた。

今後は実際にロボットハンドを用いて実験を行い、本状態推定法の有効性検証を行う計画である。

謝辞

ステレオビジョンシステムを提供頂いた植芝俊夫主任研究員をはじめ、産総研インタラクティブビジョングループの人々に感謝する。また、本研究は科研費(19300066)の助成を受けた。

参考文献

- [1] M.Kaneko and M.Kakikura: Planning strategy for putting away laundry – Isolating and unfolding task -, Proc. Of the 4th IEEE International Symposium on Assembly and task Planning, pp.429-434, 2001.
- [2] F.Osawa, H.Seki, Y.Kamiya, “Clothes Folding Task by Tool-Using Robot”, Journal of Robotics and Mechatronics, Vol.18, No.5, pp618-625, 2006.
- [3] Y.Kita, F.Saito, and N.Kita :”A deformable model driven visual method for handling clothes”, Proceeding of the 2004 IEEE International Conference on Robotics& Automation, pp.3889-3895, 2004.
- [4] T.Ueshiba:”An Efficient Implementation Technique of Bidirectional Matching for Real-time Trinocular Stereo Vision”, Proc. 18th Int’l Conf. on Pattern Recognition, vol.1, pp.1076-1079, 2006.
- [5] David A.D.GOULD, 中村達也訳: COMPLETE MAYA PROGRAMMING(日本語版), ボーンデジタル, 2004.