

しわや折れ重なりに着目した画像特徴量による布の状態表現

山崎 公俊^{†,††} 稲葉 雅幸[†]

[†] 東京大学大学院情報理工学系研究科 〒105-0123 東京都文京区本郷 7-3-1

^{††} 科学技術振興機構 さきがけ

E-mail: †{yamazaki,inaba}@jsk.t.u-tokyo.ac.jp

あらまし 本稿では、目の前に置かれた布製品を観測し、その結果から状態を把握したり把持点を決定できるような、ロボティクスでの応用が見込まれる認識課題を取り扱う。無造作に置かれた布について視覚認識を行う場合、それが様々な形状を取りうるため形状に依存しない特徴記述を行う必要がある。そこで、しわや布の折れ重なり、輪郭などに着目し、これらを画像処理によって強調した後に特徴量へ変換する方式を提案する。提案方式は、入力画像に多スケール・多方向のガボールフィルタを適用し、それらの結果を組み合わせることで布に見られる上述したような各現象を抽出しようとするものである。実験では、様々な置き方をした布のデータセットを用いて、提案する記述法によって選出される類似画像が把持点決定へ適用可能かどうかを検証する。

キーワード 布, 画像特徴, 布操作, マルチスケールフィルタ

1. はじめに

人間の生活する日常環境には、様々な種類の布製品が存在している。ロボットなどの自動機械による生活支援を考えたとき、それらに求められる作業の一つに洗濯物の取扱いのような柔軟物操作がある。本稿では、日常環境で支援行動を行うロボットのための視覚機能として衣類の状態認識に着目し、無造作に置かれた衣類を操作するために得るべき特徴表現について検討する。本研究の成果によって対応しようとする課題の一つは、目の前に無造作に置かれた布製品があったときに、それを畳むために都合の良い把持点を見つけるというものである。図1にその例を示す。畳むことを目的として掴み位置を決めるとき、人間であれば二点の把持位置を選ぶことが多い。左下の写真に示すように、布を持ち上げたときにそれを広げられた状態にできれば次の畳み動作へスムーズに移行できる。この動作を双腕ロボットに行わせることを考えたとき、布の見た目情報と把持点の組み合わせをデータベースに保存しておき、現在の見た目情報とそれらをマッチングさせることで把持点を選出することが一つの方法である。

ロボットによる物体操作においては、剛体を対象とした場合と比べ、布のような柔軟物ではその状態認識に様々な課題がある。剛体を対象とした従来研究において画像を用いる場合、形状モデルを与えられれば三次元幾何表現とのマッチングが適用できる。また、テクスチャが多い物体であれば、SIFT [6] の記述子を抽出して局所領域の類似性を1対1で評価するアプローチが取られる。しかしながら、布を対象とする場合はこれらのアプローチを適用することが難しい。前者については、たとえ衣類の形状を観測できたとしても、それによく当てはまる



図1 Folding a cloth

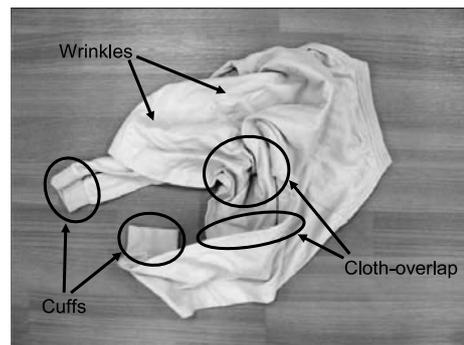


図2 A clothes

形状モデルを得るのが困難であることが一つの理由である。後者については、それらの記述子が平行移動や回転に不変な表現は可能である一方で、局所領域自体が変形することは想定されておらず、対応づけができないためである。

衣類の操作を主目的とした研究はロボティクスの分野で従来から存在し [3], [5], [9]. 認識機能についても検討が行われている。Kitaら [4] は、ステレオカメラを用いて

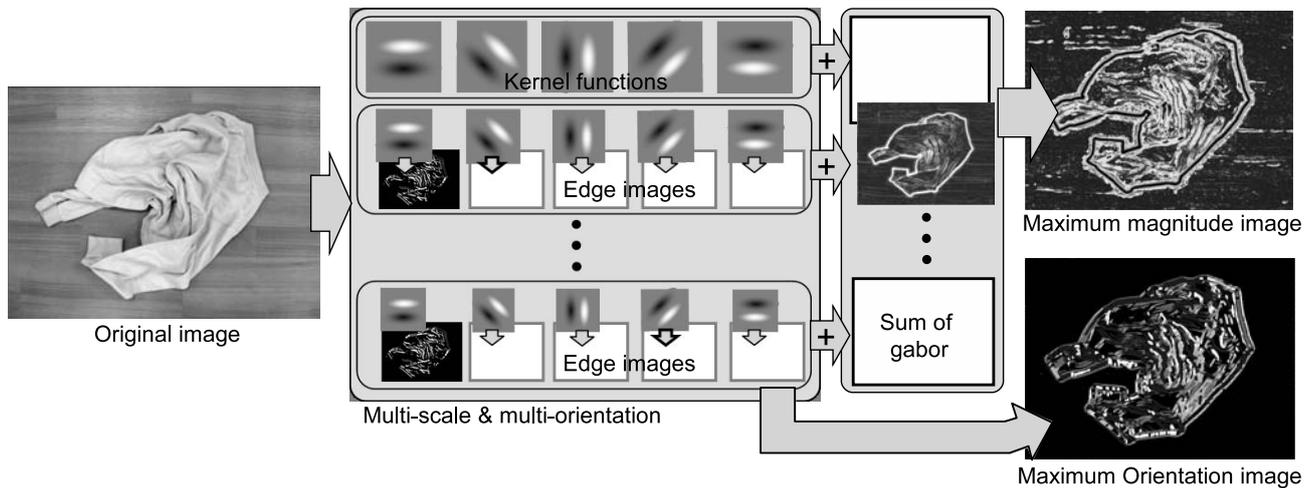


図3 Multi-scale & multi-orientation filtering. In the maximum magnitude image, the brighter pixel indicates a pixel which reacts in the larger scale parameters. In the maximum orientation image, the difference of pixel radiance indicates the direction of each wrinkle region.

得た対象衣類の三次元点群に対して一層の可変形状モデルを当てはめ、比較的複雑な形状の衣類の双腕操作を実現している。この研究では、衣類が把持されているところから認識行動がスタートしており、無造作に置かれている衣類のどこを把持点にするかという問題は扱っていない。それゆえに、衣類の種類はあらかじめ与えられている。Millerら[7]も衣類の状態モデル表現を提案しており、境界部分に着目したモデルの記述方式を提案している。また、双腕型のロボットを用いたタオル操作などについても実現しており、その様子は動画サイトで見る事ができる。同グループによる文献[8]では、ハンドタオルがテーブル上に置かれた状態からまず把持点が一つ決定され、ロボットが片手で対象物をつまみあげたところから新たな認識が始まる。冒頭に示したような人間の畳み行動に比べて、この手順は効率が低い。

筆者らによる先行研究として、衣類の発見がある[10]。この研究が上述した関連研究と異なるのは、これまでの研究が布製品の置き場所や背景を限定して進められてきたのに対し、日常環境下から布製品を「見つける」方法を示したことである。文献[10]では、抽出すべき視覚情報として布の「しわ」が着目され、無造作に置かれた状況での衣類発見と、実ロボットによる掴み上げについて報告している。本研究ではその成果の一部を引き継いでおり、布製品の種類に依らず観測できる現象に着目する。そして、それに基づいた特徴表現法を提案する。

本稿の構成は以下の通りである。次章では、布の状態記述に有効な情報を強調するための画像処理手法について述べる。3章では、特徴量の設計指針と算出方法について述べる。5章では実画像を用いた検証結果について述べ、6章でまとめる。

2. 衣類の状態を記述するための画像特徴

図2に、無造作に置かれた衣類の例を示す。綿製の長袖Yシャツであるが、ここから得られる見た目情報を以下の三種類に分けて考える。

- (1) ボタンや襟袖
- (2) しわ部分
- (3) 布の折れ重なり部分、輪郭

これらのうち(1)は布製品の種類に依る部分である。例えば襟はYシャツにはあるがタオルにはついていない。一方で(2)(3)はすべての布製品が持ちうるものであり、またそれらの状況を把握することでロボットハンドによる把持点の推測などが可能になると考えられる。そこで本研究では、しわや布の折れ重なり部分の状態を抽出できるような画像特徴の定義を試みる。

2.1 画像フィルタバンクを利用した情報抽出

布製品を撮影した画像領域には緩やかな帯状のエッジが検出されることに着目する。この領域は、比較的周期の長い周波数を密に含んでいるとも言える。

この性質を画像データから抽出するため、振幅や窓幅、方向成分などを様々に変えたガボールフィルタ(次節参照)を適用し、以下の二種類の画像を生成する。

(1) Maximum Orientation Image (MO 画像): この画像における各画素には、様々な方向成分を持つフィルタカーネルを用いて原画像に畳み込み積分を行った結果、最も強い値を記録したときの方向成分を納める。

(2) Maximum Magnitude Image (MM 画像): この画像における各画素には、様々な大きさのフィルタカーネルを用いて原画像に畳み込み積分を行った結果、最も強い値を記録したときのカーネルサイズを納める。

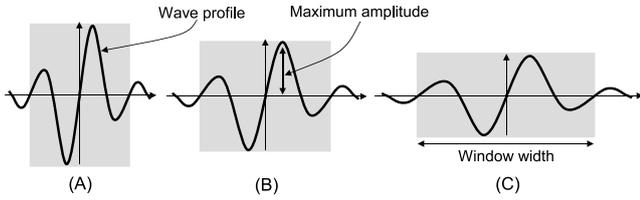


図4 Wave profiles

これらの画像を用いて、布製品のしわの分布や方向・輪郭等を分析する。また、特定のカーネルサイズのフィルタをかけた画像に着目することで、布の折れ重なり部分の候補領域を抽出する。

図3は、このフィルタリングの概要を示している。原画像に対して、様々なカーネル関数を持つガボールフィルタが適用され、二種類の画像が出力される。

2.2 ガボールフィルタ

二次元のガボールフィルタは、任意の方位、周波数を選択できるフィルタである。ウェーブレット変換ではスケールスペース構築などに利用されており、以下の式で表される。

$$g(\mathbf{x}, \theta, \sigma_x, \sigma_y) = \frac{1}{\sqrt{2\pi\sigma_x\sigma_y}} e^{-a} \cos(2\pi f x_\theta + p), \quad (1)$$

ただし、

$$a = -\frac{1}{2} \left(\frac{x_\theta^2}{\sigma_x^2} + \frac{y_\theta^2}{\sigma_y^2} \right) \quad (2)$$

$$x_\theta = (x - u_x) \cos \theta + (y - u_y) \sin \theta$$

$$y_\theta = -(x - u_x) \sin \theta + (y - u_y) \cos \theta$$

である。ここで、 f は周波数領域を示している。 σ_x^2, σ_y^2 はカーネルサイズを決める値であり、図4に示すように f は σ に連動させて決定する。なお、本研究においては x, y を同様に扱うため、以後まとめて σ と記す。すなわち、 $f(\mathbf{x}, \theta, \sigma_x, \sigma_y)$ も $f(\mathbf{x}, \theta, \sigma)$ と表記する。 x, y は現在ピクセルの座標であり、 u_x, u_y はガウス分布の中心座標である。 p は位相を示す変数であり、本研究ではここに $\pi/2$ を設定することでエッジ検出器を構成する。

ガボールフィルタには指向性があるため、式(1)の θ の設定によって特定方向のエッジを強調することが可能である。しわ等の方向を調べるためには、この θ を変化させながらフィルタリングを施せばよい。具体的には、 $-\pi \leq \theta < \pi$ の間で8分割した各方向でのエッジ検出を行う。ただし、各方向でのフィルタリングにおいて負の値を示した画素については、値を0に補正する。

2.3 多スケール・多方向フィルタリングの手順

MO画像とMM画像を生成するための前処理として、式(1)の σ と θ を様々に変えたガボールフィルタが用いられる。MO画像 $I_{ori}(\mathbf{x})$ における各画素の算出式を以下に示す。

$$I_{ori}(\mathbf{x}) = \operatorname{argmax} F_1(\mathbf{x}, \theta), \quad (3)$$

ここで、 \mathbf{x} は各画素の位置であり、 θ は式(2)におけるカーネル関数の傾き角度である。 $F_1(\cdot)$ は、 θ と近傍画素によって定まる連続関数である。

$$F_1(\mathbf{x}, \theta) = \int_w f(\mathbf{x}) g(\mathbf{x} + \mathbf{x}_0, \theta) d\mathbf{x}_0, \quad (4)$$

ここで、 $f(\mathbf{x})$ は入力画像を示している。 w は畳み込み積分のための窓領域であり、 σ の設定値によって決まる。

実際には、 $-\pi \leq \theta_n < \pi$ の範囲で複数個の θ が $(\theta_1, \theta_2, \dots, \theta_N)$ として設定されるので、 $I_{ori}(\cdot)$ はこれらの離散値での結果から算出される。フィルタリングの後、それぞれの画素で、最も大きな輝度値を示したときの θ_i を調べ、それをMO画像の画素値とする。この処理によれば、布のしわ部分を楕円領域に近似して抽出することができる。

一方で、MM画像の算出方法は以下のものである。まず、ある σ について、上述したように多方向でのフィルタリングを適用する。その後、全ての画像について同じ位置にある画素の画素値を足し合わせ、一枚の画像を生成する(図3における 'Sum of gabor')。本節では、この画像を中間画像と呼ぶ。

σ を様々に変えて同様の処理を行うことで、複数枚の中間画像が得られる。MM画像は、これらの画像で、最も大きな輝度値を示したときの σ の値を各画素の画素値としたものである。この画像を $I_{mag}(\mathbf{x})$ と記すと、計算式は以下のように表される。

$$I_{mag}(\mathbf{x}) = \operatorname{argmax} F_2(\mathbf{x}, \sigma), \quad (5)$$

ここで、

$$F_2(\mathbf{x}, \sigma) = \int_\theta \int_w f(\mathbf{x}) g(\mathbf{x} + \mathbf{x}_0, \theta, \sigma) d\mathbf{x}_0 d\theta. \quad (6)$$

窓幅 w と周波数 f は σ から自動的に決定される。本研究の実装では、窓の幅は $w = 6 \times \sigma$ として定め、周波数については $f = 1/(3 \times \sigma)$ とした。一方で、波形の最大振幅の値は以下の式によって求めた。

$$\lambda_{max} = \frac{cons}{x}. \quad (7)$$

上述の設定の元で $cons = 0.3171$ とした。これは、図4における横軸と波形の間部分の面積が、全ての σ で同じになるように調整した結果から求めた値である。

3. スケール・回転・平行移動に不変な特徴量記述

前述した多スケール・多方向のフィルタリングにより複数の画像処理結果が得られる。本研究では、ここから布の状態を表現するための特徴記述を三種類考案する。

まず、画像中から衣類領域が抽出でき、その領域を囲むような円形領域が定義されるとする。この領域の中心を C_c 、半径を R_c とする。

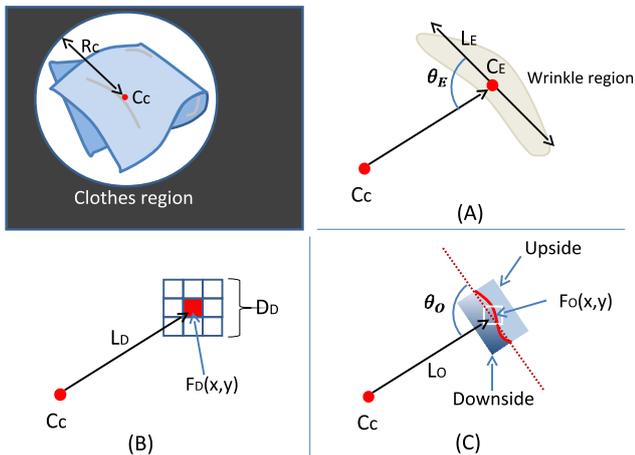


図 5 Feature description

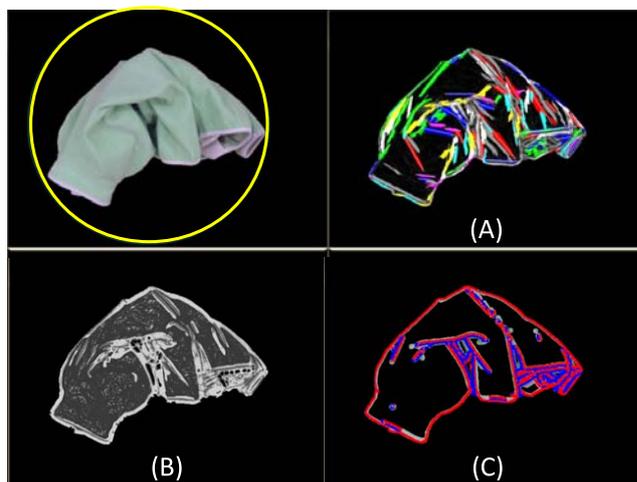


図 6 Examples of the results of image feature extraction

3.1 しわの分布と方向に着目した特徴量

MO 画像では、しわのような滑らかな輝度変化があるところで強い反応が得られ、それらの領域を方向成分によって分割することで、楕円の集合体とも言えるような処理結果が得られる。その分布を以下の三変数の特徴空間へ投影し、ヒストグラムを生成する。

ある楕円領域について、以下を算出する。

- (1) 楕円領域の長軸の長さ R_c の比
- (2) C_c と楕円領域の中心座標 C_E を結ぶ線分 L_E の長さ L_c と、 R_c の比
- (3) 線分 L_E と楕円長軸との相対角度 θ_E

これを衣類領域内部に存在する全ての楕円領域について行い、上述の各要素を軸とする三次元空間へ投影する。そして、その空間を三次元グリッドに区切り、各グリッドに納まっている値の数を取り出すことで、ヒストグラムを得る。こうして得られる表現は、スケール・回転・平行移動に不変である。

3.2 布のしわや折れ重なりへの密度に着目した特徴量

MM 画像では、しわの多くある部分とそうでない部分、境界部分などで異なる反応が得られる。この結果から、布に属する画像領域の全ての画素について、以下のような三種類の特徴表現を得る。

- (1) 指定した画素の位置 (x, y) から C_c の間の距離 L_D と、 R_c との比
- (2) 指定した画素の画素値 $F_D(x, y)$
- (3) 近傍の画素との差分の合計

$$D_D = \sum_{i,j \in W} (F_D(x, y) - F_D(x+i, y+j))$$

これらの結果を、前節と同様に三次元空間へ投影し、ヒストグラム化する。

3.3 布の折れ重なりへの位置と方向に着目した特徴量

比較的短い波長におけるフィルタリング結果に着目することで、布の境界や布の折れ重なり部分のある程度抽出することが可能である。さらに、そのようにして得た注目領域の各画素について、折れ重なりの上側の布に属するものと下側の布に属するものとに分ける。そこから以下のような三種類の特徴を求める。

- (1) 注目画素の位置 (x, y) から C_c の間の距離 L_O と R_c との比
- (2) 注目画素が折れ重なりの上どちらに属するかの連続値で評価した値 E_O
- (3) 注目画素が属している折れ重なり部分の方向成分と線分 L_E との相対角度 θ_O

これらの結果を、他の特徴表現と同様に三次元空間へ投影し、ヒストグラム化する。

3.4 特徴量の統合

上述した三種類の特徴量を統合して、布の状態を一つの特徴量として表現する。それぞれの特徴は三次元空間上の分布であるが、スケール・回転・平行移動に不変となるよう設計しているため、三次元ヒストグラム化した後に一次元のベクトルとして並び替えて利用する。具体的には、特徴空間の各軸を n 等分割し、 n^3 個の三次元グリッドを生成する。それぞれのグリッドに存在する特徴の数を計数し、ヒストグラムの各ピンの値とする。

本研究の実装では、 $n = 5$ とした。すなわち統合した特徴量の次元は、 $3 \times 5^3 = 625$ となる。

4. 実画像を用いた検証

4.1 適用例

図 6 に、提案する三種類の特徴を可視化したものの例を示す。対象衣類は厚手の生地のパーカーであり、左腕



図 7 Image dataset

部分が上になった状態で置かれている。左上の画像は原画像から衣類領域を切り出した結果である。画像中の黄色円は、後述する領域分割の結果によって衣類領域を覆うように定められるものである。右上の画像(A)は、しわの分布と方向に着目した処理(3.1節)の結果を可視化したものである。滑らかな濃度変化を持つ部分として抽出された領域を、角度によって色付けを変えて示している。これらの領域は、それぞれで楕円近似された後、特徴量に変換される。左下の画像(B)は、布のしわや折れ重なり密度に着目した処理(3.2節)の結果である。画素の濃度の違いは、フィルタが最大の反応をしめしたときの σ の違いを示している。しわや折れ重なり部分に明るい画素が集中しているが、これは σ を小さく設定したフィルタで強い反応が得られたことを意味する。また、より高解像度もしくは近接撮影した画像に対して同様の処理を適用することで、素材の違いによる反応を得られる可能性もあり、今後の課題として考えている。右下の画像(C)は、衣類の折れ重なり上下関係の評価(3.3節)した結果である。赤い画素は上側、青い画素は下側に属する布領域を示している。

4.2 データセット

特定の布製品について、それを無造作に置いて撮影することでデータセットを生成した。パイル地のハンドタオルを対象物とし、テーブルの天板や床面を背景にして約2200枚の画像を撮影した。図7に撮影画像の一部を示す。

これらのデータセットに対して、以下の処理を行う。まず、画像をVGA(640×480)にリサイズした後、Mean-shiftベースの領域分割[1]により布領域のみを抽出するためのセグメンテーションを行う。この手法では、中間結果として領域の近接性を表現したグラフと、境界エッジの情報が生成される。グラフでは、エッジに近いピクセルには小さな重み値が与えられ、Mean-shiftの結果を得るための補助情報として利用される。領域分割の結果はマスク画像として元画像と紐付けて保存しておく。さらに、布領域が抽出された画像に対して3章で述べた特徴量を求め、同様に保存しておく。以後、このように保存した特徴量を学習データと呼ぶ。

4.3 把持点探索を考慮した特徴量評価

本研究ではロボット等の自動機械による布製品の操作を背景としていることから、提案する特徴量が把持点を決定する助けとなりうるかどうかの視点から評価を行う。このため、タオルの把持点として二つの隣り合うコーナー部を選択できれば良いとし、画像に写り込んだタオルの状態を、その選びやすさの観点から5種類の難易度に分ける。

Level 5(最も難しい) :

つまみあげられるコーナー部が画像中に見つからない

Level 4(難しい) :

コーナー部が一ヶ所見つかる。もしくは離散的に複数箇所で見つかる

Level 3(やや難しい) :

コーナー部が二ヶ所以上見つかる。ただし、その二つの間が連続してつながっていることは判断しがたい

Level 2(比較的容易) :

コーナー部が二ヶ所以上見付き、それらの間の布の稜線がひねりを持ちながらも連続していることが視認できる。

Level 1(容易) :

コーナー部が二ヶ所以上見付き、それらの間の布の稜線が直線的になっていることが確認できる

例えば図7の左上の画像では、左側にコーナー部を二つ観測でき、その間がほぼ直線につながっているため、Level 1と判断される。

評価の手順は以下のとおりである。まず、データセットからランダムに画像を一枚選び出し、特徴量を算出する。それを、自分自身を除いた全ての学習データと比較し、もっとも似ていると判断される特徴量についてその元画像を選択する。その画像と入力画像について上述のレベル分けを手で行い、同じレベル値であれば対応付けが成功したもとする。なお、特徴量の類似度計算にはL1ノルムを用いた。

4.4 比較のための従来手法

比較のための方式として、提案する特徴表現の他に以下の二種類を実装した。

SIFT記述子を用いたBag of Keypoints マスク画像を用いて布領域のみを抽出した画像から、SIFT記述子を用いて多数の128次元ベクトルを得る。それを全てのデータセットについて行ったのち、Bag of Keypointsの流れに沿ってヒストグラムを用いた比較により評価できるようにする。

まず、クラスタリングにより特徴空間で任意の数のクラスタに分割し、各画像について一つのヒストグラムを生成する。本実験では、クラスタリングにはbayon[11]を用い、repeated bisectionにより100個のクラスタを生成した。

スケール調整後の布領域に関する Sum of Absolute Difference (SAD) 入力画像と学習画像からそれぞれ布領域を抽出した後、お互いの R_C が等しくなるようなスケールリングを行い、さらに画像には窓幅 $w = 3$ の平滑化を行う。その後、円形内部について Sum of Absolute Difference を取る。もしこの値が小さくなれば、二枚の画像は似ているものと判断する。

4.5 比較結果

ランダムに選び出した 50 枚の入力画像について、nearest neighbor により類似画像を選出し、それぞれで把持位置選択の容易さを調べた。入力画像の level の内訳については、Level 2 が最も 30 枚と最も多く、次いで Level 3 が 9 枚、Level 4 が 8 枚、Level 1 が 4 枚であった。すなわち、これらの入力データのうち 68% は、二点把持によって広げた状態で持ち上げられる置き方となっている可能性が高い。

入力画像から得た特徴量と学習データを比較し、最も対応のよかった学習データについてその原画像を読み出し、把持点の選びやすさ (Level) を調べた。提案特徴量を用いた場合の Level の一致率は $33/50 = 66\%$ となり、三種類の方式の中で最も高い値を示した。なお、BoK $20/50 = 40\%$ 、は SAD は $22/50 = 44\%$ の正解率であった。また、入力画像と類似画像の間の Level の違いの合計値を算出したところ (提案特徴量, BoK, SAD) = (23, 48, 41) となり、提案特徴量の選出結果では二つの画像の把持容易さがより近くなっていた。

傾向としては、図 7 における左上画像のように把持位置が容易に見つかる画像を対象にした場合に、提案方式は特に正解率が高かった。すなわち、布の稜線部分の滑らかなつながりをよく特徴化できたと考えられる。また、布が二つ折になっていた場合に、その折り目部分が高曲率のループ形状となることがあるが、そのような部位を持つ画像同士の特徴間距離は小さくなる傾向があった。

以上より、提案方式を用いた布領域全体の特徴量記述により、操作の難易度や操作方法の指針決定のための情報が得られる可能性が示された。さらに、データベース側の画像について把持点情報を人手で付加し、その部分の局所特徴量を出しておくことで、現在の入力画像から把持位置を効率的に探索することが可能になると思われる。

5. おわりに

本稿では、目の前に置かれた布製品の状態を把握したり把持点を決定することを念頭に置いた布の特徴量記述について述べた。無造作に置かれた布について視覚認識を行う場合、それが取りうる様々な形状に対応するため、スケール・回転・平行移動に不変な特徴を提案した。ここでは、しわや布の折れ重なり、輪郭などに着目し、これらを画像処理によって強調した後に特徴量へ変換する

方式となっている。実験では、様々な置き方をした布のデータセットを用いて、提案する記述法によって選ばれる類似画像が把持点決定や状態認識との親和性が高いことを確認した。

謝 辞

本研究は、JST 戦略的創造研究推進事業さきがけの一環として行われたものである。

文 献

- [1] C. Christoudias, B. Geogescu and Peter Meer: "Synergism on Low Level Vision," Int'l Conf. on Pattern and Recognition, pp. 150 - 155, 2002.
- [2] M. Galun, E. Sharon, R. Basri and A. Brandt: "Texture Segmentation by Multiscale Aggregation of Filter Responses and Shape Elements," Proc. of IEEE Int'l. Conf. on Computer Vision, pp. 716-723, 2003. <http://mplab.ucsd.edu/tutorials/tutorials.html>
- [3] K. Hamajima and M. Kakikura: "Planning Strategy for Unfolding Task of Clothes - Isolation of clothes from washed mass -," in Proc. of Int'l. Conf. on Robots and Systems, pp. 1237 - 1242, 2000
- [4] Y. Kita, F. Saito and N. Kita: "A deformable model driven method for handling clothes," Proc. of Int. Conf. on Pattern Recognition, 2004.
- [5] H. KOBORI, Y. KAKIUCHI, K. OKADA and M. INABA: "Recognition and Motion Primitives for Autonomous Clothes Unfolding for Humanoid Robot," in Intelligent Autonomous Systems 11, pp.57-66, 2010.
- [6] D. G. Lowe: "Distinctive image features from scale-invariant keypoints," Int'l Journal of Computer Vision, vol. 60, No. 2, pp. 91-110, 2004.
- [7] J. Maitin-Sp Shepard, M. Cusumano-Towner, J. Lei and P. Abbeel: "Cloth Grasp Point Detection based on Multiple-View Geometric Cues with Application to Robotic Towel Folding," Int'l. Conf. on Robotics and Automation, pp.2308 - 2315, 2010
- [8] S. Miller, M. Fritz, T. Darrell and P. Abbeel: "Parametrized Shape Models for Clothing," In the proceedings of the International Conference on Robotics and Automation (ICRA), 2011.
- [9] E. Ono, H. Okabe, H. Ichijo and N. Aisaka: "Robot Hand with Sensor for Cloth Handling," In Proc. 1990, Japan, U.S.A. Symp. on Flexible Automation, pp. 1363-1366, 1990.
- [10] K. Yamazaki and M. Inaba: "A Cloth Detection Method Based on Image Wrinkle Feature for a Daily Assistive Robots," IAPR Conf. on Machine Vision Applications, pp.366-369, 2009.
- [11] "Bayon: a simple and fast clustering tool," <http://code.google.com/p/bayon/>
- [12] "Tutorial on Gabor Filters," <http://mplab.ucsd.edu/tutorials/tutorials.html>