

# 領域分割結果の投票処理を用いた 背景に頑健な衣服領域抽出

金 正文<sup>†</sup> 山名 早人<sup>‡ §</sup>

<sup>†</sup> 早稲田大学大学院基幹理工学研究科 〒169-8555 東京都新宿区大久保 3-4-1

<sup>‡</sup> 早稲田大学理工学術院 〒169-8555 東京都新宿区大久保 3-4-1

<sup>§</sup> 国立情報学研究所 〒101-8430 東京都千代田区一ツ橋 2-1-2

E-mail: {kim, yamana}@yama.info.waseda.ac.jp

**あらまし** 画像に存在する衣服は人物特徴の一部として様々な研究において用いられており、画像から衣服領域を特定する処理が重要となっている。しかし、従来の衣服領域抽出手法は、対象となる衣服や背景に制限が存在し、不得意な背景を含む画像に対しては精度を大きく低下させるといった問題がある。そこで本稿では、GrabCut と投票処理を用いることで、背景に頑健な衣服領域抽出手法を提案する。本手法は衣服が持つ主要色の分布を抽出することで衣服領域を表すマスク画像を自動的に生成する。次に、画像上を走査する注目領域を設置し、注目領域の位置とマスク画像の情報に合わせて領域分割のシードを変更しながら衣服領域抽出を行う。そして得られた各領域抽出結果から、より衣服である可能性の高い画素を投票処理によって求めることで、画像内の衣服領域を推定する。実験では、多様な背景を含む画像から人力で衣服領域をトリミングした画像を正解画像とし、この正解画像と各手法より得た領域抽出画像を比較することで F 値による精度算出を行った。実験の結果、提案手法と同様に GrabCut を利用している既存手法と比べ、簡単な背景を含む画像において 3%、複雑な背景を含む画像において 16%の精度向上を得た。

**キーワード** clothing segmentation, GrabCut, voting, color detection, MeanShift

## 1. はじめに

現代社会において衣服は、単に風雨や寒冷などから人を保護する用途にのみ用いられるばかりでなく、人物の持つ個性、地位、主義や思想を象徴するものとして、人を表す上で欠かせないものとなっている。コンピュータビジョンにおいても、衣服は人物特徴の一部として人物の検出や認識[1][2][3]、姿勢推定[4][5][6]、写真のスケッチ化[7]など、幅広い分野の研究において利用されている。しかし衣服は画像において普遍的な形状を持たないため、画像から衣服特徴を直接得るためには背景変化、姿勢変化、他の物体によるオクルージョンといった様々な問題を解決する必要がある。特に画像から背景の影響を受けずに衣服特徴を得るためには、画像から衣服領域のみを取り出す領域抽出と呼ばれる処理が必要となる。衣服の領域抽出は衣服特徴を用いる様々な研究が大量の画像を処理しなければならないことから、全て自動で行われることが望ましい。

衣服領域を自動で抽出する既存手法として、予め衣服の形状をモデルとして定義することで、モデル画像との形状比較に基づく領域抽出を行う手法[8][9]がある。しかし、これらの手法には「定義されていない衣服形状に対応することができない」「極端な形状変化に対応することができない」などといった問題がある。また、

Interactive GraphCuts[10]や GraphCuts を拡張した GrabCut[11]による領域分割を利用することで衣服の種類や形状を問わない領域抽出を実現する手法も存在する[12][13]。GraphCuts は、画像にシードと呼ばれるラベルを付与することで領域分割を行い、画素毎に前景、背景のラベルを割り当てる手法である。シードには前景と背景を表す二種類のラベルが存在し、GraphCuts ではこの二種類のシードをユーザが指定することで、シード情報に基づいた領域分割を行う。GraphCuts を用いた衣服領域抽出手法は、衣服や背景であると思われる領域に対して付与するシードを画像特徴から自動的に決定することにより、事前知識を必要としない衣服領域抽出を実現している。しかし、これらの手法ではシードを衣服の形状に沿って与えることができない。つまり、シードを持たない領域に一樣な背景が存在していないような複雑な背景の場合、領域抽出が正しく行われず、精度が下がる。

そこで本稿では、GrabCut と投票処理を用いることで衣服の形状や背景に頑健な衣服領域抽出を提案する。本手法では衣服を構成する主要な色分布を入力画像から抽出することで、衣服領域を表すマスク画像を、事前知識を与えずに生成する。次に、入力画像上を走査する注目領域を設置し、注目領域の位置とマスク画像

の情報に合わせて領域分割のシードを変更しながら衣服領域抽出を行う。そして、投票処理を行うことで、得られた各領域抽出結果から「より衣服である可能性の高い」画素を衣服領域として抽出する。

本稿では以下の構成をとる。まず2節では関連研究について述べ、3節では提案手法について述べる。次に4節で実験について述べ、最後に5節でまとめを述べる。

## 2. 関連研究

本節では画像から衣服領域を抽出する既存の研究について述べる。なお本稿で紹介する領域抽出手法は、衣服領域の抽出処理において人手による操作を必要としない手法に限定する。

### 2.1. 衣服の形状情報を用いる手法

衣服の形状情報を用いる衣服領域抽出手法は、衣服の種類や形を予め定義することで衣服のモデルを作成し、衣服モデルとの形状マッチングによって衣服領域を抽出する手法である。

2010年、Hasanら[8]はスーツ形状に特化した空間形状モデルを作成し、スーツの衣服領域抽出を行った。Hasanらの手法は、ワイシャツ、ジャケット、ネクタイの3つの構造からなるモデルをMarkov確率場(Markov random field)[14]によって自在に変形させ、80%の精度でスーツ領域を抽出している。しかしこの手法は写真に登場する人物が全てスーツを着ていることを前提としているため、スーツ以外の衣服には手法を適用できないといった問題がある。

2011年、Weberら[9]は上下の衣服を区別した人物画像をマスクとして大量に学習させることで、上下の各衣類領域を分割可能な衣服領域抽出手法を提案した。Weberらの手法では、まず、人手で上下の衣服領域を区別した人物画像を学習画像として大量に与え、学習画像が持つ衣服の平均形状マスク画像を生成する。このマスク画像を、肩や袖口などの部分形状を比較する学習器を用いてテスト画像と形状比較し、一致した領域を切り出していくことで衣服領域を抽出する。しかし、マスクの形状が学習セットに依存しており、また、各学習器が対応できる形状変化の範囲が狭いことから、衣服の形状が学習画像の平均値と極端に異なる画像に対しては精度が大幅に低下するといった問題がある。

### 2.2. 衣服の形状情報を用いない手法

衣服の形状情報を用いない衣服領域抽出手法は、事前知識に基づいて作成される厳密に形状を定義したモデルを必要としない手法である。このような条件を満たす領域抽出手法として、GraphCuts[10]やGrabCut[11]による領域分割を用いた研究がある。

2008年、Y. Huら[12]は、Trimapと呼ばれる前景か

背景か解らない部分領域を内包する人体包含矩形を作成し、これをGrabCut[11]による領域分割結果にあわせて拡張、収縮しながら更新させていくことで衣服領域を抽出する手法を提案した。Y. Huらの手法は顔検出結果と画像の色特徴以外の特徴を必要としないため、様々な画像において手法を適用することができる。しかしtrimapの更新はGrabCutの結果に依存するため、trimap内に衣服や背景以外の物体が入りこむと、処理を何度繰り返してもその物体が除外されないといった問題がある。

2008年、Z. Huら[13]は、画像が持つ色やエッジの特徴から領域分割に用いるシードを決定することで、いかなる衣服モデルも必要としない衣服領域抽出手法を提案した。Z. Huらの手法ではまず、顔の位置から大まかな胴体包含矩形を設定することで衣服領域に最も多く含まれる色の領域を抽出し、抽出した主要色領域の形状から胴体矩形領域を推定する。次に、画像から抽出したエッジを直線に近似し、両端が胴体包含矩形の内側に位置する近似直線の頂点をドロネー三角形分割によって連結することで輪郭線を取得する。最後に、取得した胴体矩形領域の内側を前景シード、輪郭線の外側を背景シードとしたGraphCutsによる領域分割を行うことで、衣服領域を抽出する。しかし、Z. Huらの手法はエッジ特徴からシードを決定するため、テクスチャに弱いといった問題がある。また、前景シードと背景シードの間にシードが付与されない領域が生まれてしまうため、シードが存在しない領域に一律な背景が存在しない場合、領域抽出が正しく行われず、といった問題が存在する。

### 2.3. 既存手法の問題点

衣服の形状情報を用いる衣服領域抽出手法はモデルを予め定義するため、モデルで想定されていない衣服の種類や形状変化には手法を適用することができない問題があった。また衣服の形状情報を用いない手法は、領域分割を行う上で必要なシードの情報が衣服の形状に沿って与えられないため、シードが指定されていない領域の背景が一律でなかった場合、領域抽出が正しく行われず問題があった(図1)。

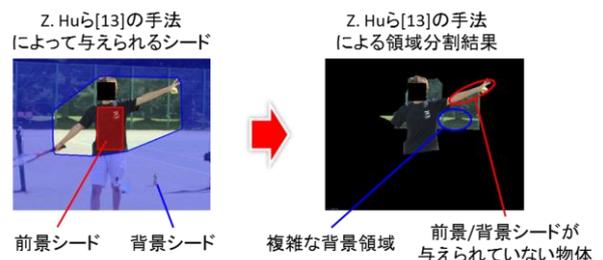


図1 シードが存在しない領域を含む領域分割

一方で、シードを与えすぎることによって生じる問題も存在する。画像に前景シードを与えすぎた場合、本来、背景領域として取り除かれるべきある領域が、衣服領域として誤って検出されてしまう問題が生まれる(誤検出領域の増加)。この問題は、ユーザが与えるシードラベルには間違っただけの情報に含まれていないことが原因となっている。そのため GraphCuts のシードを画像特徴から自動的に抽出する手法において画像のノイズ等が原因でシード情報に誤りが生じると、ユーザがそのシードを修正しない限り GraphCuts は誤ったシード情報に基づいた形で誤った領域分割を行ってしまう。以上のことから、GraphCuts による領域分割を用いた手法では、シード量が足りないと複雑な背景やテクスチャを持つ画像に弱くなる一方で、シード量を増やすと抽出結果において誤検出領域が現れる危険性が増すことがわかる。

### 3. 提案手法

本節では提案手法である領域分割結果の投票処理を用いた衣服領域の自動抽出手法について述べる。

2 節にて示した既存手法の問題点に対し、提案手法では、衣服を構成する色の分布を入力画像から抽出することで事前知識を必要としない衣服形状マスクを生成する。そして領域分割のシードを、マスクを基に衣服の形状に沿って与えることで、複雑な背景やテクスチャに頑健な衣服領域抽出を実現する。また提案手法では、投票処理によって「衣服である可能性がより高い領域」を選択することによって、十分なシードを与えつつも領域抽出結果における誤検出領域の出現を抑えることを可能としている。

提案手法は以下の 3 つの処理からなる。

1. 顔検出による人物位置取得(3.1)
2. 主要色抽出に基づくマスク画像の生成(3.2)
3. 投票処理を用いた衣服領域の自動抽出(3.3)

#### 3.1. 顔検出による人物位置取得

本提案手法ではまず、入力画像に存在する人物の位置を検出するために Haar-like 特徴による弱識別器の Boosting[15]を用いた顔検出を行う。次に、顔と胴体の位置と大きさには相互関係が成り立つ[13]ことに着目し、画像から検出した顔の位置とスケールから人体を包含する矩形を設定する。本稿では実験的に、横幅が顔の 7 倍、縦幅は顔の縦幅の 6 倍、位置は胴体領域近傍を外接するように配置する(図 2 左)。

#### 3.2. 衣服の主要色抽出に基づくマスク画像の生成

提案手法では、3.1 で取得した人物の位置情報を用いて、衣服領域を表すマスク画像を自動的に生成する。ここで、マスク画像は衣服と思われる領域に 1 の値を、

背景と思われる領域に 0 の値を持つ二値画像である。提案手法は、画像から衣服を構成する主要色を抽出することで、衣服の形状を持ったマスク画像を生成する。

##### 3.2.1. 衣服領域内接矩形の設定

提案手法ではまず、顔の位置情報を用いることで衣服に内接する矩形を設定する。本手法では、この内接矩形に含まれる全ての色を画像全体から抽出することで、衣服を構成する主要色に基づく衣服形状マスクを生成する。本稿では実験的に、顔領域と同サイズの内接矩形が、顔の位置から顔の縦幅分開けた位置に配置されるよう設定している(図 2 右)。

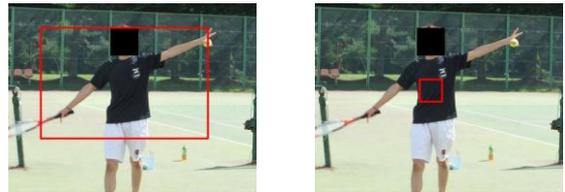


図 2 各矩形の設定

左:人体領域包含矩形

右:衣服領域に内接する矩形

##### 3.2.2. 衣服を構成する色の抽出

色の抽出は、RGB からなる入力画像に減色処理を施したものを HSV 色空間に変換し、色相を表す H チャネルのみを取り出すことで行われる(図 3 中央)。このとき、減色後の色数は  $16 \times 16 \times 16$  色とした。なお提案手法では、入力画像に予め MeanShift による画像セグメンテーション処理[16]を施すことによって、テクスチャや影による領域の細分化やノイズの影響を低減している(図 3 左)。次に提案手法は、3.2.1 で設定した内接矩形において、内接矩形に含まれる全ての色を H チャネル画像全体から抽出する(図 3 右)。内接矩形の内側に存在する画素の色を  $C = \{C_i | 1 \leq i \leq N\}$ 、入力画像 I における  $(x, y)$  座標の色を  $I(x, y)$  としたとき、内接矩形内に存在する画素の色を全て抽出したマスク画像 M は式(3.1)によって定義される。

$$M(x, y) = \begin{cases} 1 & \text{if } I(x, y) \in C \\ 0 & \text{else if } I(x, y) \notin C \end{cases} \quad (3.1)$$



図 3 HSV 色空間からの色相抽出

左:減色+MeanShift によるテクスチャ結合画像

中央:減色画像からの H チャネル抽出画像

右:内接矩形内に存在する画素の色を抽出した画像

### 3.2.3. マスク画像の生成

最後に、マスク画像  $M$  からノイズ領域を除去する。提案手法では抽出された色分布領域の中から、輪郭の持つ面積が最大となる領域が最も衣服領域を表す領域であると仮定し、それ以外の領域を除去することで衣服の形状を持ったマスク画像を生成する。マスク画像の生成結果を図 4 に示す。

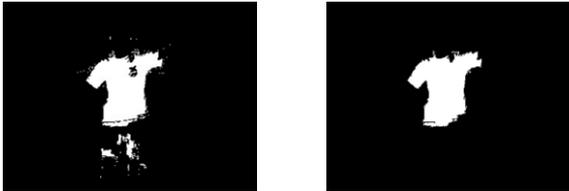


図 4 マスク画像の生成

左:ノイズ除去前

右:ノイズ除去後(マスク画像)

### 3.3. 投票処理を用いた衣服領域の抽出

3.2 で生成したマスク画像を基に衣服領域抽出を行う。提案手法では、衣服領域の抽出に GrabCut[11]を使用する。GrabCut は画像に矩形を指定することで前景と背景の色分布を GMM でモデル化し、前景らしさ、背景らしさの尤度を算出することで高精度な領域分割を実現する GraphCuts の発展手法である。また、GraphCuts と同様に、各画素にシードラベルを与えることで領域分割の精度を向上させることもできる。

GrabCut は各シードのラベルをユーザがインタラクティブに指定することでシードに基づく領域分割を行う手法であるため、一度でも誤ったシードを指定してしまうとユーザがそのシードを修正しない限り誤った領域分割を行ってしまう。ここで、3.2 で生成したマスクを GrabCut のシードとして与えることを考える。マスクは画像から主要色を抽出することで衣服の形状に沿った領域を持つ画像である。しかしマスクには細部において、画像に含まれるノイズの影響で衣服以外の領域を衣服領域であると判定してしまった誤ったシード情報が含まれる。このため、マスクをそのままシードとした領域分割を行うと、領域抽出結果に衣服以外の領域も含めてしまう(図 5 の左)。そこで本手法では、3.1 で設定した人体領域包含矩形をグリッド状に分割し、分割したグリッド領域の一つに注目したシードを入力画像全体に設定する。そして、注目領域を変更させながら領域分割を行うことで、結果の異なる複数の領域抽出画像を取得する。提案手法は、この取得した各抽出画像の中から「より衣服である可能性の高い」画素を投票処理で選択することで、入力画像から衣服である可能性の高い領域のみを最終的に出力する(図 5 の右)。

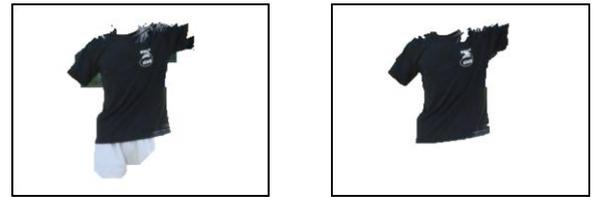


図 5 投票処理による抽出結果の比較

左:シード指定のみ

右:提案手法

#### 3.3.1. 包含矩形の分割と注目領域の設定

3.1 で決定した人体領域包含矩形を  $8 \times 8$  のグリッドに分割し、分割したグリッド上を走査する注目領域を包含矩形内にセットする。この時、注目領域の縦横幅は分割したグリッドと同幅、初期配置は最も左上に位置するグリッドと同位置、1 回の処理で移動する距離は、グリッドの縦横幅の半分とする。このため、 $8 \times 8$  のグリッドの場合、注目領域は  $15 \times 15$  回セットされる。提案手法では、注目領域がセットされる度に注目領域の位置に基づいたシードを画像全体に設定する(図 6)。

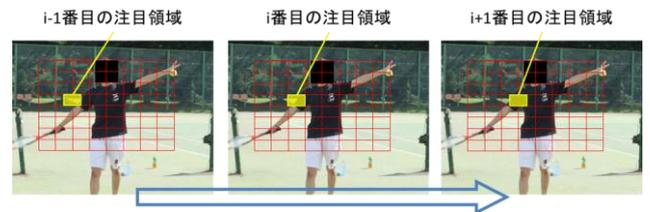


図 6 人体領域包含矩形のグリッド分割と注目領域の走査

#### 3.3.2. シードの設定

本手法では、3.2.3 で生成したマスク画像と 3.3.1 で決定した注目領域の情報を基に、入力画像  $I$  全ての画素にシードを設定することで GrabCut による領域分割を行う。なお、提案手法では GraphCuts で与えられるような前景シード(FGD)と背景シード(BGD)の他に、前景であると思われるピクセル(PR\_FGD)と背景であると思われるピクセル(PR\_BGD)にもシードを指定する。

人体領域を包含する矩形領域を  $R$ 、 $R$  内にセットされた  $i$  番目の注目領域を  $G_i$  とし、 $G_i$  に注目したとき画像全体に付与されるラベルのうち、 $(x,y)$  座標におけるシードラベルを  $S_i(x,y)$ 、各座標におけるマスクの有無を  $M(x,y)$  (マスク画像において  $(x,y)$  座標にピクセルが存在する時 1、それ以外は 0) としたとき、 $S_i(x,y)$  は式(3.2)によって以下のように定義される。

$$S_i(x,y) = \begin{cases} BGD & \text{if } (x,y) \notin R \\ PR\_BGD & \text{else if } (x,y) \notin G_i \\ FGD & \text{else if } M(x,y) = 1 \\ PR\_FGD & \text{otherwise} \end{cases} \quad (3.2)$$

式(3.2)は図 7 に示すように、マスク画像  $M$  によって衣服領域であると判断された画素に FGD シードを付

与する。また、人体領域包含矩形  $R$  の外側にある画素には BGD シードを，人体領域包含矩形  $R$  の内側且つ注目領域  $G_i$  の外側にある画素には PR\_BGD シードを，注目領域  $G_i$  の内側且つマスク画像  $M$  によって衣服領域と判断されなかった画素には PR\_FGD シードを付与する。提案手法では，その時注目している  $G_i$  によって画像に与えるシードの設定を変更させることで，一枚の入力画像から，異なる出力結果を持つ領域分割画像を複数個取得する。

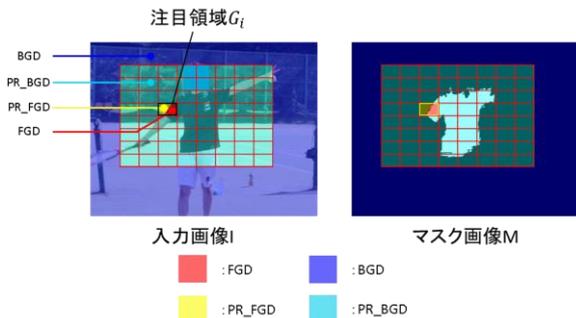


図7 シードピクセルの設定

### 3.3.3. GrabCut によるグリッド単位の領域抽出

3.3.1 において設定されたシードの情報を基に GrabCut を行うことで，注目領域ごとに衣服領域抽出結果を取得する。3.3 において，マスクによって与えられるシード情報には一部において誤ったシード情報が含まれると述べた。ここで得られた各抽出結果を見ると，誤ったシード情報を含む注目領域における領域抽出処理では，衣服領域と背景領域のどちらも含んだ抽出結果が得られる。一方で，誤ったシード情報を持たない注目領域における領域抽出処理では，衣服領域のみが抽出された結果が得られる。マスクにおいて誤ったシード情報は正しいシード情報よりも少ないことを考えると，複数回行われる各領域抽出結果において衣服領域は背景領域に比べて出現頻度が高い。このことから，各領域抽出結果をピクセル値による投票処理を用いて統合し出現頻度の高い画素のみを選択することで，衣服領域である可能性が高い領域のみを残した領域抽出結果を得ることができる。

実際の出力において，GrabCut の領域抽出結果はシードラベルで出力される。GrabCut を行った後のシードラベルを  $S_i(x, y)$ ， $i$  番目にセットされた注目領域を  $G_i (1 \leq i \leq I)$  としたとき，投票空間  $V(x, y)$  への投票は式 (3.3), (3.4) に従って行われる。

$$V_i(x, y) = \begin{cases} 1 & \text{if } S_i(x, y) = \text{FGD or PR\_FGD} \\ 0 & \text{otherwise} \end{cases} \quad (3.3)$$

$$V(x, y) = \sum_{i=1}^I V_i(x, y) \quad (3.4)$$

ここで， $V_i(x, y)$  は注目領域  $G_i$  がセットされたときの  $V(x, y)$  への投票を示している。投票空間  $V$  は入力画像と同じ大きさを持ち，「衣服領域である可能性が高い」画素を投票数によって表している。

式 (3.4) で求めた投票空間を基に，衣服領域である可能性が高い画素のみを結果として出力する。提案手法では，投票空間  $V(x, y)$  の値が閾値  $TH$  以上である画素は衣服領域であるとして出力する。投票空間  $V$  において，画素が存在する  $V(x, y)$  の取りうる最小値を  $MIN(V(x, y))$ ， $V(x, y)$  の取りうる最大値を  $MAX(V(x, y))$  としたとき， $TH$  は式 (3.5) によって定義される。

$$TH = \frac{MAX(V(x, y)) - MIN(V(x, y))}{2} \quad (3.5)$$

投票空間  $V$  において，衣服領域に属する画素における  $V(x, y)$  の値は  $MAX(V(x, y))$  に近くなり，衣服領域に属さない画素における  $V(x, y)$  の値は  $MIN(V(x, y))$  に近くなる。このため，提案手法では閾値  $TH$  を  $MAX(V(x, y))$  と  $MIN(V(x, y))$  との中間値に設定した。投票空間  $V$  と，閾値  $TH$  によって決定された，提案手法による最終的な衣服領域抽出結果を図 8 に示す。

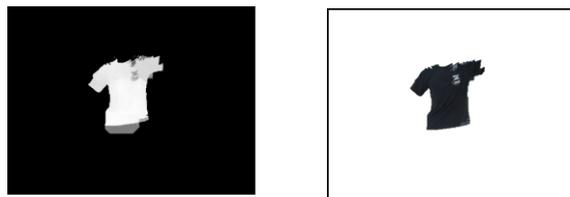


図8 衣服領域の抽出結果

左:投票空間  $V$

右:提案手法による衣服領域抽出結果

## 4. 衣服領域の抽出実験

本節では既存の衣服領域抽出手法と提案手法の領域抽出結果を比較することで提案手法の有効性を評価する。本稿では衣服領域の抽出条件を以下のように定義する。

1. 抽出処理は全て自動で行われる
2. 一枚の静止画像から抽出される
3. 衣服領域を領域の形状に沿って抽出する
4. 画像に複数の人物が存在する時は最も手前に位置する人物の衣服領域を抽出する
5. 上半身に着る衣服の領域を抽出する
6. 衣服形状の種類は問わない
7. 対象となる人物が重ね着をしていた場合，一番外側にある衣服領域を抽出する

精度比較対象としては，2 節で述べた既存研究のうち，定義された条件を全て満たす手法である Z.Hu ら [13] の手法を実装しベースラインとした。

#### 4.1. 実験に使用したデータ

実験では、二種類の入力画像を用いた。二種類用意したのは、既存手法が得意とする背景を持つ画像と、不得意な背景を持つ画像、それぞれにおける各手法の性能変化を確認するためである。

##### 4.1.1. 実験に使用する衣服画像の条件

実験に用いる衣服画像の条件を以下に示す。

1. 入力画像には衣服領域抽出において対象となる人物が写しだされている
2. 対象となる人物の顔が検出可能である
3. 対象となる人物の顔の下には衣服が存在し、遮蔽物などによって衣服領域が欠如していない
4. 衣服を構成する主要色が衣服領域内接矩形内に存在する

実験では、上記の条件に基づき独自で撮影した画像 20 枚と、Flickr[17]にて収集した 30 枚の画像を用いた。

##### 4.1.2. 背景に基づく入力画像の分類

4.1.1 の画像について、衣服領域と隣接する背景領域に複雑なエッジやテクスチャが存在しない画像を「一様な背景を多く含む画像」、存在する画像を「複雑な背景を多く含む画像」と定義することで 25 枚ずつ二種類に分類し、それぞれの画像を入力画像とした実験を行う。一様な背景を多く含む画像の一例を図 9 に示し、複雑な背景を多く含む画像の一例を図 10 に示す。



図 9 一様な背景を多く含む画像の例



図 10 複雑な背景を多く含む画像の例

なお本稿では、入力画像から衣服領域を人手により切り出したトリミング画像を予め作成し、正解画像として用いた。

#### 4.2. 結果と考察

本稿において行った実験内容は、大きく分けて以下の二つである。

**実験1. 「一様な背景を多く含む画像」による抽出実験**

**実験2. 「複雑な背景を多く含む画像」による抽出実験**

本実験では、提案手法と既存手法それぞれにおいて衣服領域の自動抽出を行い、各手法によって得られた衣服領域の抽出結果画像(以下出力画像)と正解画像を、ピクセル一致度で比較することで精度を算出する。

正解画像において画素が存在する全領域(衣服領域)を「正解領域」、出力画像において画素が存在する全領域を「抽出領域」、正解画像の座標(x,y)に画素が存在し、出力画像の座標(x,y)に画素が存在しない領域を「未検出領域」、出力画像の座標(x,y)に画素が存在し、正解画像の座標(x,y)に画素が存在しない領域を「誤検出領域」とする。このとき、

**適合率:**抽出領域における衣服領域の割合

$$= \frac{\text{抽出領域} - \text{誤検出領域}}{\text{抽出領域}}$$

**再現率:**手法において抽出された正解領域の割合

$$= \frac{\text{正解領域} - \text{未検出領域}}{\text{正解領域}}$$

$$F\text{値} = \frac{2 \times \text{適合率} \times \text{再現率}}{\text{適合率} + \text{再現率}}$$

と定義することで F 値を計算し、F 値を基に既存手法と提案手法を比較する。

実際の実験では 2 値化した出力画像と正解画像を用いることで、画像から定義した各領域の条件と合致するピクセルの総数をカウントし領域の値とした。そして各領域の値を基に適合率、再現率、F 値を算出した。

##### 4.2.1. 実験結果

実験 1 における各手法の領域抽出結果例を図 11、精度平均を表 1 に、実験 2 における各手法の領域抽出結果例を図 12、精度平均を表 2 に示す。なお、図 11、図 12 の比較画像は、各手法による抽出結果と正解画像をピクセル一致度で比較した結果を画像で表現したものである。比較画像において、正解領域は白、誤検出領域は赤、未検出領域は青のピクセルで表現される。

##### 4.2.2. 各実験結果の考察

はじめに実験 1 の結果を考察する。図 11 より実験 1 において既存手法と提案手法はどちらも高い精度で衣服領域を抽出していることが確認された。また各比較画像より、提案手法の抽出結果は既存手法と比べて誤検出領域の出現を抑えることに成功している様子が確認された。一方で、投票処理を行うことで未検出領域が逆に増加してしまった画像も確認された。これは「より衣服領域である可能性が高い」画素を投票処理で求める際に、衣服領域の一部も背景領域とともに除去してしまったことが原因として考えられる。この問題は投票時に用いる閾値を現在の値より低く設定することによって解決される。以上の内容を踏まえて精度を比較すると、提案手法は既存手法と同程度の精度で衣服領域を抽出できていることが表 1 より確認できた。提案手法の精度において、投票処理を行う前と後で精度を比較すると投票処理後の再現率は前述の未検出領域の増加によって投票処理前より低下していることが解

入力画像	Z.Huらの手法 (既存手法)	比較画像 (既存手法)	提案手法 (投票無し)	提案手法 (投票あり)	比較画像 (提案手法)

図 11 実験 1 における各手法の領域抽出結果例

入力画像	Z.Huらの手法 (既存手法)	比較画像 (既存手法)	提案手法 (投票無し)	提案手法 (投票あり)	比較画像 (提案手法)

図 12 実験 2 における各手法の領域抽出結果例

表 1 実験 1 における各手法の精度平均

手法	適合率	再現率	F 値
既存手法	<b>0.967</b>	<b>0.880</b>	<b>0.917</b>
提案手法 (投票無し)	<b>0.952</b>	<b>0.954</b>	<b>0.950</b>
提案手法 (投票あり)	<b>0.976</b>	<b>0.925</b>	<b>0.945</b>

表 2 実験 2 における各手法の精度平均

手法	適合率	再現率	F 値
既存手法	<b>0.736</b>	<b>0.890</b>	<b>0.786</b>
提案手法 (投票無し)	<b>0.926</b>	<b>0.910</b>	<b>0.913</b>
提案手法 (投票あり)	<b>0.964</b>	<b>0.897</b>	<b>0.915</b>

る。一方で投票処理後の適合率は投票処理前より増加しているため最終的な F 値はほぼ同じ値となっている。

次に実験 2 の結果を考察する。実験 2 で用いた画像における既存手法の領域抽出結果には、実験 1 と比べてより多くの誤検出領域が含まれていることが図 12 において確認できる。これは、領域分割を行う上で必要なシードの情報が衣服の形状に沿って十分に与えられないため、シードが指定されていない領域の背景が一様でなかった場合、領域抽出結果において誤検出領域を多く残してしまうからである。これに対し提案手法の領域抽出結果からは、投票処理を行うことで複雑な背景を多く含む画像においても誤検出領域の増加を抑えることに成功している様子が確認された。なお実験 2 においても、実験 1 と同様に投票処理を行うことで未検出領域が逆に増加してしまった画像が存在した。こちらの問題も投票時に用いる閾値を現在の値より低

く設定することで解決することができるが、一方で閾値を下げることによって誤検出領域が増加してしまうため、実験 2 において誤検出領域の増加と未検出領域の増加はそれぞれトレードオフの関係にあるといえる。

以上の内容を踏まえて精度を比較すると、提案手法は既存手法と比べ、再現率を保ったまま適合率を大きく増加させたことによって最終的な F 値で既存手法を大きく上回ったことが表 2 から確認された。これは投票処理によって「より衣服領域である可能性が高い」画素を決定してあげることで、複雑な背景を多く含む画像においても誤検出領域の出現を抑えることに成功したことに起因される。提案手法の精度において、投票処理を行う前と後とで精度を比較すると、投票処理後の再現率は前述の未検出領域の増加によって投票処理前より低下していることがわかる。一方で、投票処理後の適合率は投票処理前より増加しているため、最

最終的な F 値は同値となっている。

本稿において行った各実験をまとめる。既存手法と提案手法の精度比較実験の結果、提案手法は既存手法と比べて簡単な背景を含む画像において 3%、複雑な背景を含む画像において 16%、F 値が向上したことが表 1 表 2 から解った。特に実験 2 における適合率を比較すると、提案手法は既存手法と比べて 31%もの精度向上を得たことが確認された。以上のことから、投票処理を用いることで複雑な背景を多く含む画像においても、誤検出領域の出現を抑えることで精度を向上させることを目的とする提案手法の有用性が確認された。

また本稿では、統計的な有意性を検証するために検定を行った。本実験は標本サイズが小さく一部に異常値が含まれているのでパラメトリックな検定は適さない。このためノンパラメトリック検定であるマン・ホイットニーの U 検定を行った。検定の結果、実験結果において、実験 1 の適合率の P 値が 0.6834、再現率の P 値が 0.0247、F 値の P 値が 0.0059 となった。また、実験 2 の適合率の P 値が 0.0009、再現率の P 値が 0.0316、F 値の P 値が 0.0454 となった。よって有意水準 5%において、既存手法に比べて提案手法の適合率、再現率、F 値が向上することは、実験 1 の適合率を除きすべて統計的に有意であることがわかった。実験 1 の適合率において統計的な有意性が見られなかった原因としては、一部の画像において提案手法の適合率が既存手法の適合率と比べ低かったことが原因として考えられる。これらの画像において提案手法の精度が低下した原因としては、光や影の影響による輝度変化や衣服の模様(ストライプ模様)の存在によって画素の伝搬が失敗してしまったこと等が考えられる。

## 5. まとめ

本稿では、GrabCut と投票処理を用いることで衣服の形状や背景に影響されない衣服領域抽出手法を提案した。本手法では、衣服を構成する主要な色分布を入力画像から抽出することで、衣服領域を表すマスク画像を事前知識無しに生成する。次に、入力画像上を走査する注目領域を設置し、注目領域の位置とマスク画像の情報に合わせて領域分割のシードを変更しながら衣服領域抽出を行う。そして、投票処理を行うことで、得られた各領域抽出結果から「より衣服である可能性の高い」画素を衣服領域として抽出する。

実験では、提案手法と同様に GrabCut を利用している既存手法と比べ、簡単な背景を含む画像において 3%、複雑な背景を含む画像において 16%の精度向上を得た。今後は明暗変化に強い衣服領域抽出と、重ね着や衣服領域内節矩形の位置制限によって発生する提案手法の適用制限の緩和が課題となる。

## 謝辞

本研究の一部は、科研費(基盤(B)21300038)によるものである。

## 参考文献

- [1] N. Sprague and J. Luo: "Clothed people detection in still images," Proc. of the IEEE Int'l. Conf. on Pattern Recognition, pp.585-689, 2002.
- [2] A. C. Gallagher and T. Chen: "Using context to recognize people in consumer images," IPSJ Trans. on Computer Vision and Applications, pp.115-126, 2009.
- [3] W. Gao, H. Ai and S. Lao: "Adaptive contour features in oriented granular space for human detection and segmentation," Proc. of the IEEE Computer Vision and Pattern Recognition, pp.1786-1793, 2009.
- [4] R. Ronfard, C. Schmid and B. Triggs: "Learning to parse pictures of people," Proc. of European Conf. on Computer Vision, pp.700-714, 2002.
- [5] Z. Hu, G. Wang, X. Lin and H. Yan: "Recovery of upper body poses in static images based on joints detection," Pattern Recognition Letters, pp.503-512, 2008.
- [6] V. Ferrari, M. Marin-Jimenez and A. Zisserman: "Pose search: retrieving people using their pose," Proceedings of the IEEE Computer Vision and Pattern Recognition, pp.1-8, 2009.
- [7] H. Chen, Z. Xu, Z. Q. Liu and S. C. Zhu: "Composite templates for cloth modeling and sketching," Proc. of the IEEE Computer Vision and Pattern Recognition, pp.943-950, 2006.
- [8] B. S. Hasan and D. C. Hogg: "Segmentation using deformable Spatial Priors with Application to Clothing," British Machine Vision Conference, pp.1-11, 2010.
- [9] M. Weber, M. Baumli, R. Stiefelhofen: "Part-based clothing segmentation for person retrieval," IEEE Int'l. Conf. on Advanced Video and Signal-Based Surveillance, pp.361-366, 2011.
- [10] Y. Boykov and M. P. Jolly: "Interactive Graph Cuts for Optimal Boundary & Region Segmentation of Objects in N-D Images," IEEE Int'l. Conf. on Computer Vision and Pattern Recognition, pp.731-738, 2004.
- [11] C. Rother, A. Blake, and V. Kolmogorov: "GrabCut - interactive foreground extraction using iterated graph cuts," ACM Trans. Graphics, Vol.23, No.3, pp.309-314, 2004.
- [12] Y. Hu: "Human Body Region Extraction from Photos," IAPR Conf. on Machine Vision Applications, 2007.
- [13] Z. Hu, H. Yan and X. Lin: "Clothing Segmentation Using Foreground and Background Estimation Based on the Constrained Delaunay Triangulation," Pattern Recognition Letters, pp.1581-1592, 2008.
- [14] S. Geman and D. Geman: "Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images," IEEE Trans. on Pattern Analysis and Machine Intelligence, pp.721-741, 1984.
- [15] P. Viola and M. Jones: "Rapid Object Detection using a Boosted Cascade of Simple Features," Proc. of IEEE Conf. on Computer Vision and Pattern Recognition, pp.511-518, 2001.
- [16] D. Comaniciu and P. Meer: "Mean shift: A robust approach toward feature space analysis," IEEE Trans. Pattern Analysis and Machine Intelligence, Vol.24, No.5, pp.603-619, 2002.
- [17] Flickr, <http://www.flickr.com/>, (2012.1.31.accessed).