

ゴルフスイング診断のための画素値の時間ヒストグラムと グラフカットに基づく人体シルエット抽出

非会員 神山 泰典* 非会員 松本 俊昭* 非会員 白井 良明**
非会員 島田 伸敬** 非会員 植田 勝彦***

Silhouette Extraction based on Time Sequence Histograms and Graph Cut for Golf Swing Diagnosis

Yasunori Koyama*, Non-member, Toshiaki Matsumoto*, Non-member, Yoshiaki Shirai**, Non-member,
Nobutaka Shimada**, Non-member, Masahiko Ueda***, Non-member

(2012 年 5 月 9 日受付, 2012 年 7 月 7 日再受付)

We propose a novel method of extracting a human silhouette from a golf swing movie taken by a mobile phone camera for automatic golf swing diagnosis. The difficulties of the task are that we cannot use the background subtraction method and that the camera is not fixed. First we correct small camera movements during a swing sequence by tracking features in the obvious background. Based on the fact that the histogram of the pixel value (hue and brightness) at the background pixel distributes in a narrow interval, we classify each pixel into three categories: background pixel, human pixel, and mixture pixel. We propagate pixel values of the background from the background pixels to the mixture ones. Finally we apply a graph cut method to extract the human silhouette at each image. Experimental results showed that the proposed method is promising.

キーワード : 人体シルエット抽出, 手ブレ補正, ヒストグラム, 特徴追跡, グラフカット

Keywords : silhouette extraction, blurring correction, histogram, feature tracking, graph cut

1. はじめに

本研究の目的は、携帯カメラで撮影した動画から、ゴルフ診断に用いることのできる人体シルエットを抽出することである。

従来から人体姿勢推定の研究は盛んに行われてきた。多数のマーカを貼り付けるモーションキャプチャ法、光切断法で 3 次元データを得て形状を求める方法⁽¹⁾、一定の背景で多数のカメラでシルエットを抽出して 3 次元形状を抽出してから姿勢を推定する方法⁽²⁾⁽³⁾、多視点カメラからの動画から、3 次元データを直接求めることなく、それぞれの画像

処理を融合する方法⁽⁴⁾などがあるが、いずれも特別な装置を必要とする。

単眼画像からのシルエット抽出も研究されている。多くは背景差分を用いているが、あらかじめ背景画像を得ることができない場合は適用できない。

多数の背景サンプルを用いて背景の特徴を学習しておき、未知画像の背景を識別する方法もある⁽⁵⁾⁽⁶⁾が、あらかじめ背景を学習することができない場合には適用できない。

注目領域と背景を分割する手法として、グラフカット⁽⁷⁾がある。しかし、あらかじめ代表的な人体領域と背景領域が得られない場合は、適切な分割ができないことが多い(例を後に示す)。

動画から移動する人体を追跡する研究も多数あるが、追跡では人体の関節運動を求める必要がなく、人体全体としての移動のみを推定している⁽⁸⁾。

人体姿勢推定に関しては、手の上げ下げなどの単純な認識や隠れのない間接角度を誤差 20 度ほどで推定する試み⁽⁸⁾があるが、スイング診断には精度が十分でない。シルエットから、姿勢辞書を利用してより多様な姿勢を推定する研究⁽⁹⁾もあるが、辞書にある姿勢を認識するだけで、高精度の

* 立命館大学 理工学研究科
〒525-8577 滋賀県草津市野路東 1-1-1
Graduate School of Science and Engineering, Ritsumeikan University
1-1-1, Nojihigashi, Kusatsu-shi, Shiga 525-8577, Japan
** 立命館大学 情報理工学部
〒525-8577 滋賀県草津市野路東 1-1-1
College of Science and Engineering, Ritsumeikan University
1-1-1, Nojihigashi, Kusatsu-shi, Shiga 525-8577, Japan
*** 住友ゴム工業 (株)
〒651-0071 兵庫県神戸市中央区筒井町 2-1-1
Sumitomo Rubber Industries, Ltd.
2-1-1, Tsutsui-cho, Chou-ku, Kobe, Hyogo 651-0071, Japan

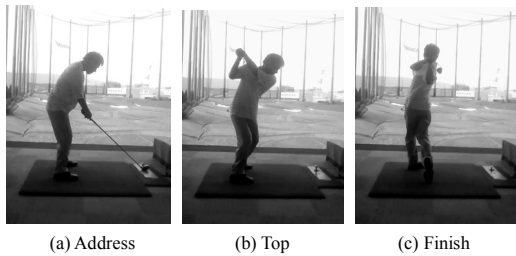


Fig. 1. Typical position images from a swing movie

関節角度を必要とするスイング診断には十分でない。

近年, 手軽にゴルフスイングの診断を受けられるシステムの開発の試みもある。ETGA スウィングレッスン⁽¹⁰⁾は, 腰に携帯電話をつけ, モーションセンサーを用いて診断しているが, 人体姿勢は直接把握していない。ダンロップスイング自動診断システム⁽¹¹⁾は, 室内で2台の固定TVカメラを用いて撮影した背景画像とゴルフスイングの動画画像から背景差分によって抽出した人体シルエットに基づいて診断している。屋外でも使えるシステムも研究されている⁽¹²⁾が, 固定TVカメラ2台を用いるので, 使用環境が限られている。

任意の場所でより手軽に利用できるようにするため, ユーザがカメラ付き携帯電話で撮影したゴルフスイングの動画画像 (以下, スイング動画画像) (Fig. 1 参照) からスイング診断できるシステムの開発が望まれている。そのためには以下の条件を考慮しなければならない。

- (1) 背景画像を事前に撮影しておけない。
- (2) 診断専用の衣服やマーカを使用できない。
- (3) 撮影中に手ブレが生じる。

本論文は, これらの条件下での人体シルエットの手法を述べる。なお, 診断に関しては室内用のシステム⁽¹⁰⁾と同様であるので, ここでは述べない。

提案手法は, スイング動画画像の背景に対応する画素の画素値はほぼ一定で, 人体に対応する画素値は変化することに基づいて人体シルエットを抽出する。そのため, まず, スイング動画画像の各画素がほぼ同じ背景に対応するように手ブレを補正し, 各画素の時間変化に対する画素値のヒストグラムから背景候補を得る。次にその結果を画像中に伝播させて人体シルエットを求める。その結果, 人体シルエット周辺で誤抽出する場合がある。そこで, さらにグラフカット処理によってノイズを軽減する。グラフカット性能は, 初期領域とコストの定義に依存する。ここでは, 初期領域は時間ヒストグラムによって得られている。また, コストも得られている人体と背景領域の画素値に基づいて設定している。

本論文の中心は時間ヒストグラムによって人体シルエットを抽出する手法の提案である。そして, それが有効に機能することを示すための手ブレ補正と, 後処理としてノイズを軽減するためのグラフカット処理も述べる。

2. 手ブレ補正

固定していないカメラ付き携帯電話でスイング動画画像を

撮影した場合, 手ブレにより, 各画像が変形する。画像の変形は, シルエット抽出や手足の移動推定の障害となるため補正する必要がある。手ブレは3次元の平行移動と回転の自由度があるが, 画像の変形は微小であるため, アフィン変換を用いて近似し, 背景がほぼ静止するようにする。

時刻 $t=0$ 画像の特徴点をその後 ($t=1, 2, \dots, n$) の画像で追跡し, 特徴点と追跡した点 (以下, 対応点という) の座標の差が最小になるよう時刻 $t=1, 2, \dots, n_t$ の画像を変換する。以下に具体的な処理を示す。

(1) 時刻 $t=0$ のアドレス姿勢の画像は人によって変化が少ないので, 大きめの人体マスクを事前に作成する。人体マスク以外 (仮背景領域) は, 画像系列の大部分で背景となる。仮背景領域から手ブレ補正に用いる特徴点を検出する。特徴点の候補の検出には, Harris オペレータを用いる。すなわち, 画素 (x, y) における輝度値を $I(x, y)$ とすれば, Harris 関数 $R(x, y)$ は次式となる。

$$R(x, y) = \det \mathbf{M}(x, y) - k (\text{tr} \mathbf{M}(x, y))^2$$

$$\mathbf{M}(x, y) = \begin{bmatrix} \left(\frac{\partial I(x, y)}{\partial x} \right)^2 & \left(\frac{\partial I(x, y)}{\partial x} \frac{\partial I(x, y)}{\partial y} \right) \\ \left(\frac{\partial I(x, y)}{\partial x} \frac{\partial I(x, y)}{\partial y} \right) & \left(\frac{\partial I(x, y)}{\partial y} \right)^2 \end{bmatrix}$$

..... (1)

ここで, $\det$ は行列式, tr は対角成分の和を表し, k は定数とする。 $R(x, y)$ が閾値 (ϕ) 以上の座標を特徴点の候補とする (ここでは $k=0.04$, $\psi=300$)。さらに非極大抑制によって候補を絞る。

(2) 時刻 $t-1$ の画像の特徴点 $p_1^{t-1}, \dots, p_n^{t-1}$ を基に, 時刻 t の画像の対応点を検出する。対応点の候補 $p_1^t, \dots, p_n^t$ の検出には Lucas-Kanade 法を用いる。これらの候補は, 仮背景領域に動物体が現れる場合には, 正確な対応点とならないことがある。そこで, 各対応点の候補 p_i^t の近傍領域と p_i^{t-1} の近傍領域をテンプレートマッチングによって対応をチェックし, 対応しない候補点を除く。

(3) 特徴点 p_i^t の座標を $\mathbf{x}_i^t$ とすると, アフィン変換後の座標 $\mathbf{x}_i'^t$ は次式である。

$$\mathbf{x}_i'^t = \mathbf{A}^t \mathbf{x}_i^t + \mathbf{B}^t \quad \text{..... (2)}$$

ここで, $\mathbf{x}=(x, y)^T$ であり, 変換パラメータ $\mathbf{A}^t, \mathbf{B}^t$ は, 時刻 t の画像を変化するための2行2列, 2行1列の行列である。時刻 t の画像のアフィン変換には, 次式に示す差の2乗和 d^t を最小とする $\mathbf{A}^t, \mathbf{B}^t$ を用いる。

$$d^t = \sum_{s=1}^n (\mathbf{x}_s'^t - \mathbf{x}_s^t)^2 \quad \text{..... (3)}$$

上記の処理を用いて時刻 $t=1, 2, \dots, n_t$ の画像をアフィン変換し, 手ブレを補正する。

Fig. 2 に, 補正前, 補正後の各姿勢とアドレス姿勢との差分画像を示す。手ブレ補正により, 背景領域の差分が減少したことが分かる。

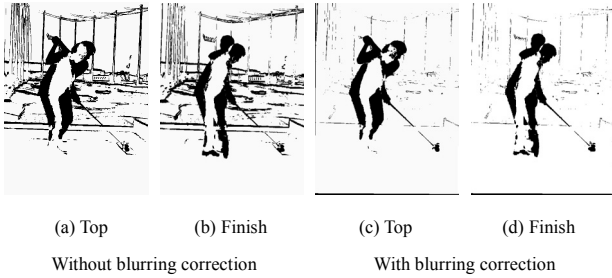


Fig. 2. Subtraction images between address and other typical position images in Fig. 1

3. 時間ヒストグラムに基づく人体シルエット抽出

各位置における画素値の時間変化に対するヒストグラムは、背景に長時間対応する位置では、背景に対応する画素値による頻度が一定の区間に集中するため、背景に対応する画素値の区間（以下、背景区間）を決める事ができる。一方、あるときは人体に、別なときには背景に対応する場合は、この方法で背景区間を決める事ができない。この場合、近傍画素で既に決まっている背景区間から背景区間を決める（伝播とよぶ）。足元は、逆光の影響で暗くなり色が似ている場合が多いので、別の手法を用いる。

〈3・1〉 ヒストグラムの作成 まず、カラー画像の画素値 R, G, B （0 から 1 に正規化）を色相 h 、彩度 s 輝度 b に変換する。変換は、

$$I_{\max} = \max\{R, G, B\}, \quad I_{\min} = \min\{R, G, B\}$$

$$D_r = R - (G + B)/2, \quad D_g = \sqrt{3}(G - B)/2$$

を用いて次式で表す。

$$h = \text{atan2}(D_g, D_r) \quad \dots\dots\dots (4)$$

$$s = \begin{cases} (I_{\max} - I_{\min}) / (I_{\max} + I_{\min}) & I_{\max} + I_{\min} \leq 1 \\ (I_{\max} - I_{\min}) / (2 - I_{\max} - I_{\min}) & I_{\max} + I_{\min} > 1 \end{cases} \quad \dots\dots\dots (5)$$

$$b = 0.299R + 0.587G + 0.114B \quad \dots\dots\dots (6)$$

このうち、領域分けに有効な、色相 h と輝度 b のヒストグラムを用いる。 h と b を 0～99 の整数値にスケール変換し、色相ヒストグラム N_h と輝度ヒストグラム N_b を作成する。

時刻 t の画像の x における画素を x' とし、 x' の色相、輝度を $h(x')$ 、 $b(x')$ ($0 \leq h(x'), b(x') \leq 99$) とし、すべての時刻にわたるヒストグラムを作成し、以下で表す。

$$N_b = (n_b(0), n_b(1), \dots, n_b(99))$$

$$N_h = (n_h(0), n_h(1), \dots, n_h(99))$$

ただし、色相は周期性があるとみなし、 $0 \leq h \leq 99$ 以外も繰り返しているとみなす。

〈3・2〉 ヒストグラムの性質 まず、典型的な例として、人体領域の画素値が一様でないとし、Fig. 3 の各画像の丸印について考える。右の丸の位置は常に背景に対応（以下、

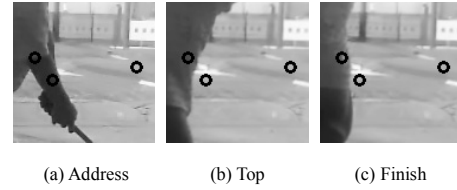
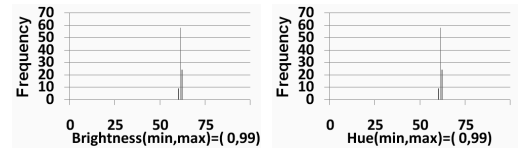
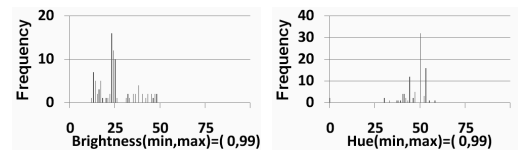


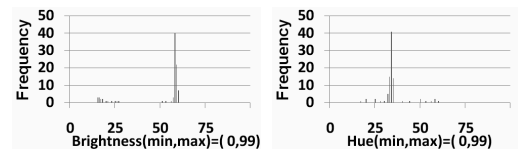
Fig. 3. Three classes of pixels



(a) Histograms at the right circle in Fig. 3



(b) Histograms at the left circle in Fig. 3



(c) Histograms at the middle circle in Fig. 3

Fig. 4. Histograms of three classes of pixels in Fig. 3

背景位置）するため、輝度ヒストグラム、色相ヒストグラムは、頻度が一定区間に集中する（Fig. 4(a)参照）。左の丸は常に人体に対応（以下、人体位置）し、頻度が分散するため、ヒストグラムのピークは比較的低く、ピーク以外にも小さくない頻度をもっている（Fig. 4(b)参照）。中央の丸の位置は背景と人体どちらにも対応（以下、混合位置）するため、一定区間に集中した頻度とそれ以外の分散した小さい頻度の両方を持つ（Fig. 4(c)参照）。

もし、人体領域の画素値が一様であれば、人体領域でもヒストグラムは集中する。しかし、背景と人体とどちらにも対応する場合は、頻度が一定区間に集中しないとみなすことができる。

〈3・3〉 背景区間の決定 ヒストグラムの頻度が集中した区間（以下、集中区間）があれば、それをてがかりにして背景に対応する画素値の区間（背景区間）を求める。ここで、集中区間とは、区間の幅が小さく、区間内の頻度が占める割合が大きい区間である。具体的には、輝度ヒストグラム N_b の集中区間 $[b_l, b_r]$ は、次式を満たすものとする。

$$\left. \begin{aligned} &1 \leq b_r - b_l \leq \alpha \\ &\sum_{b_l=b_i}^{b_r=b_i} n_b(b_i) / \sum_{b_i=0}^{99} n_b(b_i) \geq \theta \\ &n_b(b_i) \neq 0 \text{ for } b_i \in [b_l, b_r] \end{aligned} \right\} \quad \dots\dots\dots (7)$$

ここで, α , θ は閾値で, 手ブレの程度や動物体がどれだけの時間含まれることを許容するかに依存する (ここでは実験的に, $\alpha=12$, $\theta=0.6$ としている)。同様に, 次式を満たすように色相ヒストグラム N_h の集中区間 $[h_l, h_r]$ を決める。

$$\left. \begin{aligned} 1 \leq h_r - h_l \leq \alpha \\ \sum_{h_i=h_l}^{h_r} n_h(h_i) / \sum_{h_i=0}^{99} n_h(h_i) \geq \theta \\ n_g(h_i) \neq 0 \text{ for } h_i \in [h_l, h_r] \end{aligned} \right\} \dots\dots\dots (8)$$

輝度と色相の集中区間 $[b_l, b_r]$, $[h_l, h_r]$ がいずれも得られた場合は, それぞれの区間を背景区間候補とする。

人体と背景の輝度が類似している場合, 集中区間 $[b_l, b_r]$ に人体に対応する区間を含むことがある。そこで, 色度 (色相や彩度) の差によって背景区間候補を制限する。

色度の差は, 単純に h - s 空間のユークリッド距離とすると, 彩度の小さいところでの色相の変化が過剰に評価されるので, h を角度とし, s を半径とする極座標が望ましい。また, 小さい s での変化を重視するようにするために, s の対数を用いることにする。その結果, 色度空間の直交座標 x, y は次式で表すことができる⁽¹¹⁾。

$$\left. \begin{aligned} x &= (\log_{10}(99s+1)) \cos h \\ y &= (\log_{10}(99s+1)) \sin h \end{aligned} \right\} \dots\dots\dots (9)$$

色度の差はこの空間でのユークリッド距離とする。以後, 2 つの色相と彩度 h_1, s_1 と h_2, s_2 の色度差を $D_c((h_1, s_1), (h_2, s_2))$ で表す。

まず, 区間 $[b_l, b_r]$ の頻度の最大値を与える輝度 b_m は背景に対応するとし, それを構成する画素集合の色相, 彩度の平均 $h_{ave}(b_m)$, $s_{ave}(b_m)$ と, b_i ($b_l \leq b_i \leq b_r$) に対応する画素集合の色相, 彩度の平均 $h_{ave}(b_i)$, $s_{ave}(b_i)$ の色度差 $D_c((h_{ave}(b_m), s_{ave}(b_m)), (h_{ave}(b_i), s_{ave}(b_i)))$ が閾値 σ_c 以上の b_i を集中区間から除く。得られた区間の中で b_m を含み, (7) 式を満たす区間を N_b の背景区間候補 $[b_l, b_r]$ とする。

同様に, $[h_l, h_r]$ の頻度の最大値を与える h_m を構成する画素集合の輝度の平均 $b_{ave}(h_m)$ と h_i ($h_l \leq h_i \leq h_r$) に対応する画素集合の輝度の平均 $b_{ave}(h_i)$ の差の絶対値 $D_b(b_{ave}(h_m), b_{ave}(h_i))$ が閾値 σ_b 以上の h_i を集中区間から除く。得られた区間の中で h_m を含み, (8) 式を満たす区間を N_h の背景区間候補 $[h_l, h_r]$ とする。なお, 閾値 σ_c と σ_b は, いずれも可能な値の範囲の 7.5% としている。

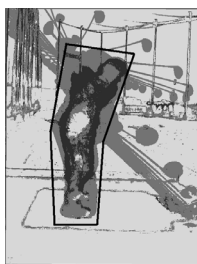


Fig. 5. Three classes of pixels

暫定的に, 全ての頻度が背景区間候補に入る位置を背景領域, 60%以上の頻度が背景区間候補に集中している位置を混合領域, それ以外の位置を人体領域とする。Fig. 5 に処理結果を示す。薄い灰色の領域が背景領域, 濃い灰色が混合領域, 黒が人体領域である。

〈3・4〉 背景区間の伝播

人体となる時間が長い場合は, 〈3・3〉節の手法では背景区間を決定できない。そこで, 既に背景区間が決められている画素の背景区間を 8 近傍画素で背景区間が決められていない画素へ伝播させて, 背景区間を決める。

背景区間が決められている画素 $\mathbf{x}$ から近傍画素 $\mathbf{x}_n$ への伝播を考える。まず, 画素 $\mathbf{x}$ の背景区間 $[b_l, b_r]$, $[h_l, h_r]$ に類似した $\mathbf{x}_n$ の区間 (以下, 背景区間候補) を決定し, つぎに, 背景区間候補から色が類似していない区間を除くことで位置 $\mathbf{x}_n$ の背景区間 $[b_l'', b_r'']$, $[h_l'', h_r'']$ を決める。以下に手順を示す。

(1) 背景区間候補を決定できるのは, 以下の 4 つの条件を満たす b_m'' と h_m'' が存在する場合である。

1) $\mathbf{x}$ の輝度ヒストグラム集中区間 $[b_l, b_r]$ の最大値を与える輝度 b_m に対応する画素集合の色相, 彩度の平均 $h_{ave}(b_m)$, $s_{ave}(b_m)$ と, $\mathbf{x}_n$ における同様の最大値を与える b_m'' に対応する画素集合の色相, 彩度の平均 $h_{ave}(b_m'')$, $s_{ave}(b_m'')$ との色度差 $D_c((h_{ave}(b_m), s_{ave}(b_m)), (h_{ave}(b_m''), s_{ave}(b_m'')))$ が閾値 ρ_c 以下。

2) $|b_m'' - b_m| \leq \beta$ (ここでは $\beta=2$)。

3) 色相ヒストグラムに関しても 1) と同様に, 色相の最大値を与える h_m と h_m'' に対する輝度の平均の差が閾値 ρ_b 以下。なお, 閾値 ρ_c と ρ_b は, ここではいずれも可能な値の範囲の 2.5% としている。

4) $|h_m'' - h_m| \leq \beta$ 。

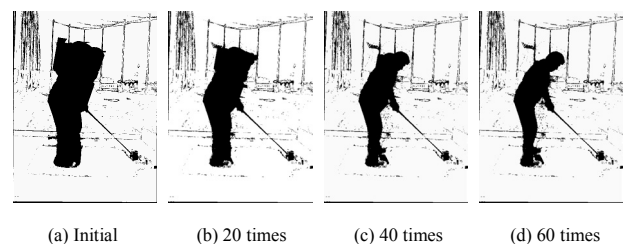
以上の条件を満たす場合は, 背景区間候補を $[b_l - \gamma, b_r + \gamma]$ $[h_l - \gamma, h_r + \gamma]$ とする (ここでは, $\gamma=4$)。

(2) 〈3・3〉節で背景候補を決めた場合と同様に, 輝度ヒストグラムの区間 $[b_l - \gamma, b_r + \gamma]$ から b_m'' に対応する色度との差が大きくなる輝度値を除き, その結果を背景区間 $[b_l'', b_r'']$ とする。また, 色相ヒストグラムの区間からも同様に輝度値の差が大きい色相値を除き, 背景区間 $[h_l'', h_r'']$ を決める。

新しく背景区間が決まった画素からも, その近傍に伝播を繰り返す。新たな背景区間が見つからなくなった時点で伝播を終える。アドレス時のシルエットが伝播により決まっていく例を Fig. 6 に示す。

〈3・5〉 足元のシルエット抽出手法

足元付近の背景



(a) Initial (b) 20 times (c) 40 times (d) 60 times

Fig. 6. Sequence of background propagation (shown in white)



Fig. 7. Feet search region foot

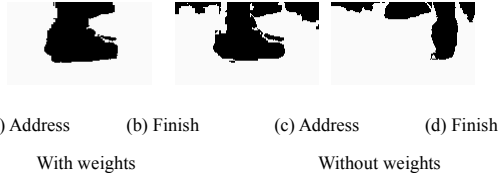


Fig. 8. Extracted feet silhouette

は、影や照明条件などによって一様でない場合が多いので、複数の背景サンプルを用いる。以下に処理手順を示す。

(1) アドレス姿勢の手元（シャフトの上端）とボールの位置）を基準に足元シルエットの探索領域を決める（Fig. 7 太枠内）。

(2) 背景サンプルは、探索領域内の下縁、右縁、左縁に沿って 5pixel 間隔で配置した n_q 個の小領域（10×10 単位は pixel）とする（Fig. 7 白枠内）。小領域 q の平均の色相、彩度、輝度を h_q, s_q, b_q とし、 $\mathbf{v}_q = (h_q, s_q, b_q)^T$ を背景サンプル q の画素値とする。同様に、探索領域内の位置 $\mathbf{x}$ の画素に対して、色相 h_x 、彩度 s_x 、輝度 b_x の画素値を $\mathbf{v}_x = (h_x, s_x, b_x)^T$ と表す。画素値 $\mathbf{v}_x$ と背景サンプル値 $\mathbf{v}_q$ の色の差 D_i を次式で表す。

$$D_i(\mathbf{v}_x, \mathbf{v}_q) = D_c((h_x, s_x), (h_q, s_q)) + D_b(b_x, b_q) \quad \text{.....(10)}$$

人体領域に対応する画素は、次式で示す背景サンプル値との差の重みつき和が閾値以上とする。

$$D_s(\mathbf{v}_x) = \sum_{q=1}^{n_q} w_q D_i(\mathbf{v}_x, \mathbf{v}_q) \quad \text{.....(11)}$$

ここで、重み w_q は、すべての背景サンプルが同程度に人体領域となりにくいように、下式を満たすように決める。

$$D_s(\mathbf{v}_q) = \text{constant. for } q=1, \dots, n_q \quad \text{.....(12)}$$

Fig. 8 に、処理結果の例を示す。

〈3・6〉 人体シルエットの決定 各時刻の画像の人体シルエットを抽出する。時刻 t の画素 $\mathbf{x}^t$ を、輝度 $b(\mathbf{x}^t)$ と色相 $h(\mathbf{x}^t)$ が背景区間に含まれている場合は背景、それ以外の場合は人体として抽出する。頻度の全てが背景区間に対応している領域を背景領域、背景区間が存在する領域を混合領域、それ以外を人体領域とする。結果を Fig. 9 に示す。

一般に、コントラストが強い背景領域では、2つの原因で幅の狭いノイズが残る：1) 突発的に現れる背景により、中間色が生成され、背景区間が伝播できないため、2) 補正しきれなかった手ブレによる微小な背景の位置ずれが、画素

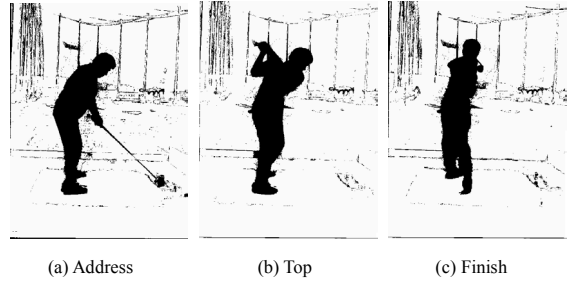


Fig. 9. Result of silhouette extraction for Fig. 1

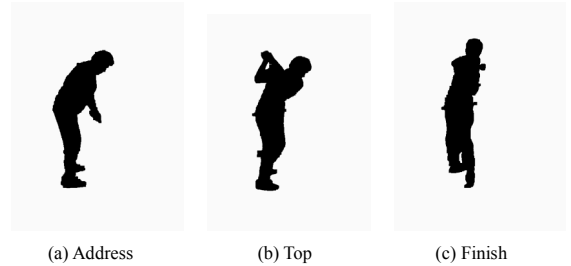


Fig. 10. Silhouettes by reducing noise for Fig. 9

値の異なる 2 つの背景区間を作るためである。これらは、収縮膨張を施すことにより除去できる（Fig. 10 参照）。

4. グラフカットを用いたシルエットの修正

背景と人体の色がほぼ等しい画素は、背景区間と人体区間の区別できないため、大きなノイズ（Fig. 10(b)の膝裏）となり、膨張収縮では削除できない。そこで、人体と背景の領域を求めるため、グラフカットアルゴリズムを適用する。

まず、その初期領域としては、ヒストグラムに基づいて求めたシルエット領域を用いる。

グラフカットでは、画像を、画素が頂点とし、隣接画素間に枝があるグラフとみなし、画素 $\mathbf{x}$ と人体との間の枝のコスト $P_h(\mathbf{x})$ 、背景との間の枝のコスト $P_b(\mathbf{x})$ 、および隣接画素間のコストを定義し、領域分割によって切られる枝のコストを最小とする。それぞれのコストは以下のように定める。

まず、 $P_h(\mathbf{x}), P_b(\mathbf{x})$ としては、画素値に基づくコストと、前節で述べた背景区間の伝播に基づくコストを考える。前者は、従来から用いられている人体領域および背景領域のカラーヒストグラムに基づく値 $P_h^c(\mathbf{x}), P_b^c(\mathbf{x})$ である。

後者は背景区間の伝播の信頼性の高い場合はコストを高くする。前節の伝播では、背景である隣接画素の背景区間が類似であれば、伝播をした。ここでは、差が大きいほど人体に近いので、差を人体との接続コストとする。画素 $\mathbf{x}$ の近傍を $\mathbf{x}_1 (= \mathbf{x} - (1, 0)), \mathbf{x}_2 (= \mathbf{x} - (0, 1))$ とする。 $\mathbf{x}$ での背景区間 $[b_l^x, b_r^x], [h_l^x, h_r^x]$ の最大値を与える輝度と色相を b_m^x, h_m^x とし、同様に $\mathbf{x}_1, \mathbf{x}_2$ での背景区間 $[b_l^{x_1}, b_r^{x_1}], [h_l^{x_1}, h_r^{x_1}], [b_l^{x_2}, b_r^{x_2}], [h_l^{x_2}, h_r^{x_2}]$ の最大値を与える輝度と色相を $b_m^{x_1}, h_m^{x_1}, b_m^{x_2}, h_m^{x_2}$ とする。

$$\left. \begin{aligned} P_h^p(x) &= \sum_{i=1}^2 \left(D_b(b_m^{x_i}, b_l^{x_i}) + D_h(h_m^{x_i}, h_l^{x_i}) + |b_m^{x_i} - b_m^{x_j}| + |h_m^{x_i} - h_m^{x_j}| \right) \\ P_b^p(x) &= 1.0 - P_h^p(x) \end{aligned} \right\} \dots\dots\dots (13)$$

$P_h^p(x)$, $P_b^p(x)$ は確率でないが、和が 1 になるように $P_b^p(x)$ を定めてもグラフカットには影響がない。画素 $\mathbf{x}$ の、HSV によるコストに伝播処理を考慮したコストを加えた人体背景接続コスト $P_h(x)$, $P_b(x)$ を以下に示す。以上より、

$$\left. \begin{aligned} P_h(x) &= (P_h^c(x) + P_h^p(x)) \\ P_b(x) &= (P_b^c(x) + P_b^p(x)) \end{aligned} \right\} \dots\dots\dots (14)$$

人体背景接続コストの計算結果を Fig. 11 に示す（白領域ほど人体領域である確率が高い）。

隣接画素間コストは、隣接画素 $\mathbf{x}_1, \mathbf{x}_2$ の画素値 $\mathbf{v}_{x_1}, \mathbf{v}_{x_2}$ との色の差を用いて次式で表す。

$$P_h(\mathbf{x}_1, \mathbf{x}_2) = \kappa / D_l(\mathbf{v}_{x_1}, \mathbf{v}_{x_2}) \dots\dots\dots (15)$$

ここで、 κ は重みである。

この 2 つのグラフ接続コストを使用し、2 値の大域最小化

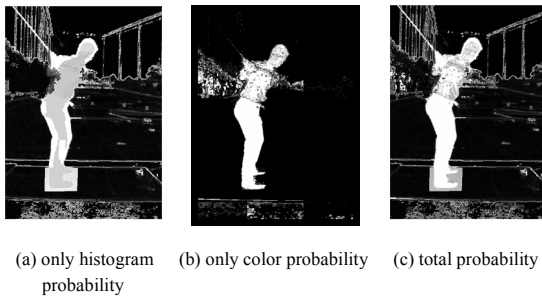


Fig. 11. Human probability

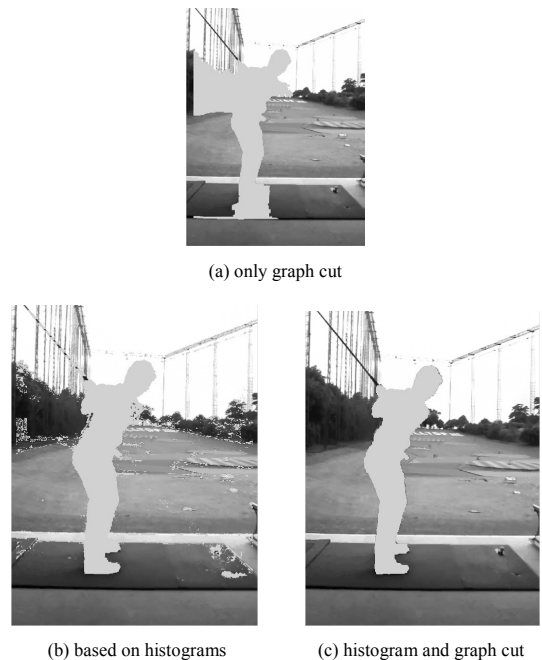


Fig. 12. Result

アルゴリズム⁽⁹⁾によるグラフカットをおこない、人体シルエットを抽出する。

Fig. 12(a)は、最初に人体背景領域と背景領域を小さめに推定し、各領域の色ヒストグラムと隣接画素間のコストを用いたグラフカットの結果、同図(b)は、3 章で述べた画素値の時間ヒストグラムに基づく結果、同図(c)はさらにここで述べたグラフカットの結果である。

グラフカットだけでは類似した色を判別することができない（例えば Fig. 12(a)）。時間ヒストグラムに基づく方法では、人体と背景の境界にノイズが残ることがある（Fig. 12(b)）。ここで提案する方法は、類似した色にも強く、人体周りのノイズの影響も軽減できることがわかる。場所人が異なる 10 種類の動画像に対して妥当な結果をえることができた。

5. おわりに

ゴルフスイングの診断に必要となる人体シルエットをスイング動画像から抽出する手法を述べた。まず、各位置の画素ヒストグラムから背景に長時間対応する位置の背景区間を決定し、それらを画像中に伝播させ背景区間を決め、シルエットを抽出する。得られたシルエットおよび伝播処理における画素値の差に基づいてグラフカット法を適用して人体シルエットを修正する。今後の課題としては、多数の例に対する検証、多様な携帯端末からの分解能の異なる画像への対処、夜間に服や髪など黒い部分の人体シルエットが欠けてしまう問題への対処などである。

文 献

- (1) T. Funatomi, M. Iiyama, K. Kakusho, and M. Minoh : “Accurate 3D scanning of trunk swaying human body”, Proc. Symposium of MIRU, Vol.1, pp.565-570 (2004) (in Japanese)
船富卓哉・飯山将晃・角所 考・美濃導彦：「身体動揺を考慮した人体形状の計測」，画像の認識・理解シンポジウム予稿集，Vol.1, pp.565-570 (2004)
- (2) N. Date, H. Yoshimatsu, D. Arita, and R. Taniguchi : “Vision-based Real-time Human Motion Sensing and Its Applications”, Proc. Symposium of MIRU, Vol.1, pp.678-683 (2004) (in Japanese)
伊達直人・吉松寿人・有田大作・谷口倫一郎：「多視点動画像解析による身体の実時間姿勢推定」，画像の認識・理解シンポジウム予稿集，Vol.1, pp.678-683 (2004)
- (3) H. Tanaka, A. Nakazawa, T. Machida, and H. Takerura : “Example Based Approach for Human Pose Estimation Using Volume Data and Graph Matching”, IEICE Transactions on Information and Systems, Vol.J91-D2, No.6, pp.1580-1591 (2008) (in Japanese)
田中秀典・中澤篤志・町田貴史・竹村治雄：「ボリウムデータの細線化とグラフマッチングを用いた事例ベース人体姿勢推定」，信学論，Vol.J91-D2, No.6, pp.1580-1591 (2008)
- (4) H. Kawashima and T. Matsuyama : “Multi-Viewpoint Gesture Recognition by an Integrated Continuous State Machine”, IEICE Transactions on Information and Systems, Vol.J85-D-II, No.12, pp.1801-1812 (2002) (in Japanese)
川嶋宏彰・松山隆司：「連続状態モデル間の相互作用に基づく多視点動作認識」，信学論，Vol.J85-D-II, No.12, pp.1801-1812 (2002)
- (5) A. Elgammal, R. Duraiswami, D. Harwood, and L. Davis : “Background and Foreground Modeling Using Nonparametric Kernel Density Estimation for Visual Surveillance”, Proc. of the IEEE, Vol.90, No.7, pp.1151-1163 (2002)

- (6) K. Kim, T. H. Chalidabhongse, D. Harwood, and L. Davis: "Real-time foreground-background segmentation using codebook model", Elsevier Journal of Real-Time Imaging, Vol.11, No.2, pp.172-185 (2005)
- (7) Y. Boykov and V. Kolmogorov: "An Experimental Comparison of Min-Cut/Max-Flow Algorithms for Energy Minimization in Vision", IEEE Trans. on PAMI, Vol.26, No.9, pp.1124-1137 (2004)
- (8) Y. Shirai, R. Okada, and T. Yamane: "Robust visual tracking by integrating various cues", M. Vince and G. D. Hager Eds., Robust vision for vision-based control of motion, IEEE Press, Chap.5, pp.53-65 (2000)
- (9) K. Onishi, T. Takiguchi, and Y. Arik: "3D Human Posture Estimation Using the HOG Features from Monocular Image", Meeting on Image Recognition and Understanding collection of paper, IS3-16, pp.960-965 (2008) (in Japanese)
大西克則・滝口哲也・有木康雄:「HOG 特徴に基づく単眼画像からの人体 3 次元姿勢推定」, 画像の認識・理解シンポジウム予稿集, IS3-16, pp.960-965 (2008)
- (10) R. Okada and B. Stenger: "Human Posture Estimation using Silhouette-Tree-Based Filtering", Proc. Symposium of MIRU, pp.64-79 (2006) (in Japanese)
岡田隆三・シュテンガビヨン:「シルエットを用いた Tree-Based Filtering による人体の姿勢推定」, 画像の認識・理解シンポジウム予稿集, pp.64-79 (2006)
- (11) www.fmworl.net/product/phone/f-01b/images/info46.pdf
- (12) M. Ueda, Y. Shirai, and M. Ohnuki: "Automatic Golf Swing diagnosis System Using Video from TV camera", Proc. DIA Workshop, pp.228-232 (2006)
植田勝彦・白井良明・島田伸敬・大貫正秀:「TV カメラからの映像を用いたゴルフスイング自動診断システム」, 動的画像処理実用化ワークショップ 2006 講演論文集, pp.228-232 (2006)
- (13) T. Mochizuki, M. Ueda, Y. Shirai, and Y. Shimada: "Image Feature Extraction for Automatic Outdoor Golf-Swing Diagnosis", Proc. View Workshop, pp.53-58 (2007) (in Japanese)
望月智則・植田勝彦・白井良明・島田伸敬:「屋外環境下におけるゴルフスイング自動診断のための画像特徴抽出」, ビジョン技術の実用ワークショップ論文集, pp.53-58 (2007)

神山泰典



(非会員) 1986 年 4 月 24 日生。2009 年 3 月立命館大学情報理工学部知能情報学科卒業。2011 年 3 月同大学大学院理工学研究科情報理工学専攻博士前期課程修了。同年 4 月 NTT コミュニケーション (株) 入社。現在, (株) NTT プララ出向中。

松本俊昭



(非会員) 1987 年 4 月 10 日生。2010 年 3 月立命館大学情報理工学部知能情報学科卒業。2012 年 3 月同大学大学院理工学研究科情報理工学専攻博士前期課程修了。同年 4 月 JR 東日本 (株) 入社。

白井良明



(非会員) 1941 年 8 月 3 日生。1964 年 3 月名古屋大学工学部卒業。1969 年 3 月東京大学大学院工学系博士後期課程修了。工学博士。同年 4 月電子技術総合研究所入所。1985 年制御部部长。この間 1971 年 8 月より 1 年間米国 MIT 人工知能研究所客員研究員。1988 年 4 月大阪大学工学部教授。この間 1997 年 4 月より 3 年間東京大学大学院工学研究科教授併任。2005 年 4 月立命館大学情報理工学部教授。2012 年 4 月同総合科学技術研究機構客員教授。人工知能学会, ロボット学会, 情報処理学会, 機械学会, 電子情報通信学会, IEEE Computer Society 各会員。

島田伸敬



(非会員) 1969 年 12 月 30 日生。1992 年 3 月大阪大学工学部卒業。1997 年 3 月同大学大学院工学研究科博士後期課程修了。博士 (工学)。1997 年 4 月より大阪大学工学部電子制御機械工学科助手。2004 年 4 月立命館大学情報理工学部知能情報学科准教授。2012 年 4 月同学科教授。その間 2007 年 4 月より 1 年間米カーネギーメロン大学ロボティクス研究所客員准教授。

コンピュータビジョン, ヒューマンインターフェース, 知能ロボットの研究に従事。電子情報通信学会, 情報処理学会, 人工知能学会, 日本ロボット学会, IEEE Computer Society 各会員。

植田勝彦



(非会員) 1974 年 2 月 16 日生。1999 年 3 月同志社大学院工学研究科機械工学卒業。同年 4 月住友ゴム工業 (株) 入社。2002 年からゴルフスイング自動診断システムの開発に従事。現在に至る。