

# 手話認識のための見えの学習による手領域検出と形状識別

山田 寛<sup>†</sup> 松尾 直志<sup>†</sup> 島田 伸敬<sup>†</sup> 白井 良明<sup>†</sup>

<sup>†</sup> 立命館大学理工学研究科  
滋賀県草津市野路東 1-1-1

E-mail: †{yamada,matsuo}@i.ci.ritsumeikan.ac.jp, ††{shimada,shirai}@ci.ritsumeikan.ac.jp

あらまし 手話では、背景に肌色を含む場合の手の形状認識が必要である。ここでは、様々な手の見えを学習することで手領域を検出し、手の3次元形状を推定する方法を提案する。特徴にはHSV表色系のHSヒストグラムと勾配方向ヒストグラムを用い、AdaBoostによる手の検出器を作成し、手領域の検出を行う。顔や手の一部は誤検出しやすいので、まず手・顔領域を検出し、次にその中から手領域を検出する。検出された手領域を識別器にかける。識別対象が手のため、色情報を使わない代わりにHistograms of Oriented Gradients(HOG)特徴を入力に用いる。検出実験と識別実験により提案手法の有効性を示す。

キーワード 手話認識, 手形状識別, 手領域検出, Histograms of Oriented Gradients, AdaBoost

## Hand Detection and Hand Shape Classification Based on Appearance Learning for Sign Language Recognition

YAMADA YUTAKA<sup>†</sup>, MATSUO TADASHI<sup>†</sup>, SHIMADA NOBUTAKA<sup>†</sup>, and SHIRAI YOSHIAKI<sup>†</sup>

<sup>†</sup> Graduate School of Science and Engineering, Ritsumeikan University  
1-1-1 Noji-higashi Kusatsu-shi Shiga

E-mail: †{yamada,matsuo}@i.ci.ritsumeikan.ac.jp, ††{shimada,shirai}@ci.ritsumeikan.ac.jp

**Abstract** This paper proposes a way of detecting a hand and classifying the shape by learning various appearances. The hue-Saturation histogram and the gradient histogram are used for hand detection. To detect a hand, we use two classifiers. The first classifier detects face and hand. The second classifier separates hand from face. The detected hand region is classified by shape classifiers using the histograms of oriented gradients. Results of detecting and classifying experiments show the validity of our method.

**Key words** sign language recognition, hand shape classification, hand detection, Histograms of Oriented Gradients, AdaBoost

### 1. はじめに

動画像からの手話の認識は、時間的に変化する手話の特徴を抽出する処理と、抽出された特徴系列から手話を認識する処理からなる。

手話単語は手の構え(形状), 向き, 提示位置, 大局的な運動からなる形態素の時系列での連続により語の意味が決まるといわれている[1]。画像処理を利用した特徴量の取得は話者への負担は少ないが, 特徴抽出が難しい。そこで, 特別な色をつけた普通の手袋を装着し手話を行い, その画像系列から簡単な手話を認識する研究が行われた[2]。また, 赤外線カメラを用いた温度画像を用いた手法[3]も行われたが, いずれも手話者への負担がある。

普通のカメラから画像処理によって特徴抽出を行う研

究では, オプティカルフローを利用した手法[4], [5]や動きベクトルを用いた手法[6], フレーム間差分を利用した手法[9]により, 手の提示位置, 大局的な運動は取得できるが手形状に関する特徴は十分考慮されていない。

川東ら[13], 今川ら[10], 山本ら[11], 柳ら[12]は色情報でセグメンテーションを行い手領域を取得し, 手形状特徴を求めている。しかし, これらの手法は手領域が他の肌色領域と重複する時には正確な手領域が得られず特徴抽出が難しい。

今井ら[14]は手領域の輪郭点からCGで作成した手の3次元形状の見え方と照合することで手の3次元形状を推定する手法を提案している。しかし, 膨大なパターンとの照合が必要なため, 計算コストが大きい。

その後, 手話認識は, 米国, フランス, 台湾, 日本な

どで研究されているが [15]，特徴抽出の困難さから実用には至っていない．

本論文では画像からの手話認識で困難な課題である手領域抽出および手形状特徴抽出を扱う．まず，色と勾配方向ヒストグラムを入力特徴として，検出を行う．重複時の画像も多数学習しておき，重複にも頑健な検出器を作成する．手と顔は特徴が似ているため識別が難しい．そこで，手・顔と背景の識別器と手と顔の識別器の2種類の識別器を用いて検出する．

次に検出された手領域に対し，ある手形状と別形状を識別する識別器を複数用いて，手形状の識別を行う．識別対象が全て手であるため，色情報による分類は難しいと考えられる．そこで，人物検出に用いられている Histograms of Oriented Gradients(HOG) 特徴 [7], [8] を識別器の入力に用いる．識別器には AdaBoost [16] を用いて学習を行う．

## 2. 色とエッジ特徴による手領域の検出

手領域の特徴として，以下の2点が考えられる．

- 指の輪郭に沿ってエッジが出ること
- 色が肌色であること

そこで，エッジと色情報を入力として手領域の検出を行う．

### 2.1 入力特徴

エッジや色情報を表現するために，本研究では勾配方向ヒストグラムと HSV 表色系による色ヒストグラムを入力特徴として手領域の検出を行う．

#### 2.1.1 勾配方向ヒストグラム

輝度画像から勾配画像を作成し，勾配方向ヒストグラムを作成する．各ピクセルの輝度  $L$  から勾配強度  $m$  と勾配方向を求める．

$$m(x, y) = \sqrt{f_x(x, y)^2 + f_y(x, y)^2} \quad (1)$$

$$\theta(x, y) = \tan^{-1} \frac{f_y(x, y)}{f_x(x, y)} \quad (2)$$

$$f_x(x, y) = L(x + 1, y) - L(x - 1, y) \quad (3)$$

$$f_y(x, y) = L(x, y + 1) - L(x, y - 1) \quad (4)$$

計算された勾配強度  $m$  と勾配方向  $\theta$  を用いて， $M \times M$  ピクセルの小領域を1セルとしてセル毎に，輝度の勾配方向ヒストグラムを作成する．ただし，計算された勾配方向は， $0^\circ \leq \theta \leq 360^\circ$  となるが，本手法では勾配方向の向きを考慮する必要がないため  $0^\circ \leq \theta \leq 180^\circ$  とする．方向を  $20^\circ$  ずつに分割するため，9方向の勾配方向ヒストグラムとなる．

スケールが異なるヒストグラム値を同等に扱うため，ヒストグラムを作成した領域の面積で正規化する．正規化されたヒストグラム値を下式のように並べて勾配方向ヒストグラム  $\mathbf{h}^e$  を得る．

$$\mathbf{h}^e = (b^1, b^2, \dots, b^9) \quad (5)$$

ここで  $b^i$  はヒストグラムの  $i$  番目の値を表す．

#### 2.1.2 色ヒストグラム

色情報は色味と明度が分離できる HSV 表色系で表現する．まず，RGB 表色系から円柱形の HSV 表色系へと変換を行う．その後，H 値，S 値を用いて2次元ヒストグラムを作成する．しかし，S 値が小さい場合は，H 値の値に信頼性がない．一方，S 値が小さくとも明度の差は生じるため V 値は異なる．そこで，S 値が小さい画素では，H 値でヒストグラムを作成するのではなく，V 値についてのヒストグラムを作成する．本研究では H 値を12分割，S 値を4分割してヒストグラムを作成する．その内，S 値がもっとも小さい値をとる場合は12分割した V 値のヒストグラムを作成する．得られたヒストグラム値を下式のように並べ，色ヒストグラム  $\mathbf{h}^c$  を得る．

$$\mathbf{h}^c = (v^1, v^2, \dots, v^{12}, b_2^1, \dots, b_2^{12}, \dots, b_4^1, \dots, b_4^{12}) \quad (6)$$

ここで  $v^i$  は明度の  $i$  番目のヒストグラム値， $b_j^i$  はヒストグラムの色相が  $i$  番目，彩度が  $j$  番目のヒストグラム値を表す．

#### 2.1.3 多重解像度のヒストグラム特徴

手領域は向き，手の形状にバリエーションがあるため，手領域の検出には局所的，大域的な特徴のどちらも用いる必要があると考えられる．

そこで，複数解像度でヒストグラムを求め，それぞれのヒストグラムを並べて入力特徴ベクトルとする．本研究では図1のように局所領域と大域的な領域でヒストグラムを作成し特徴とする．

注目している領域全体に対して作成したヒストグラムを Level 1 とする．Level  $N$  のヒストグラムは注目領域を縦横  $2^{N-1}$  等分した  $2^{2N-2}$  個の領域に対し作成されるそれぞれのヒストグラムである．

Level  $N$  までのヒストグラムを並べて下式のようなヒストグラム特徴  $\mathbf{H}_N$  を得る．

$$\mathbf{H}_N = [\mathbf{h}_1(1, 1), \mathbf{h}_2(1, 1), \mathbf{h}_2(1, 2), \dots, \mathbf{h}_n(2^{n-1}, 2^{n-1})] \quad (7)$$

ここで， $\mathbf{h}_n(i, j)$  は  $2^{2n-2}$  個に分割されたうち， $i$  行  $j$  列目の領域に対する Level  $n$  のヒストグラム値を並べたベクトルである．

最終的に得られる入力特徴ベクトル  $\mathbf{x}$  を下式に示す．

$$\mathbf{x} = (\mathbf{H}_N^c, \mathbf{H}_N^e) \quad (8)$$

ここで， $\mathbf{H}_N^e$  は勾配方向ヒストグラムを式7のように Level  $N$  まで並べたベクトル， $\mathbf{H}_N^c$  は色ヒストグラムを Level  $N$  まで並べたベクトルを表す．本手法では  $N = 3$  として1197次元の特徴ベクトルを作成する．

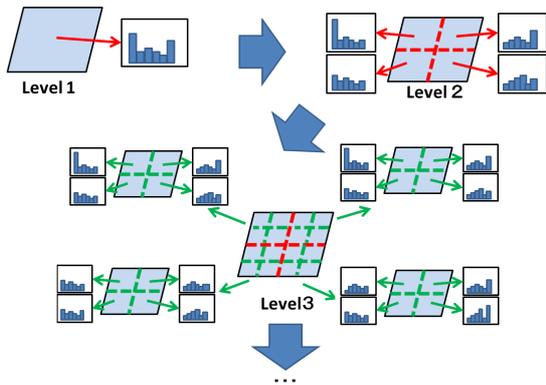


図 1 階層的なヒストグラムによる情報の表現

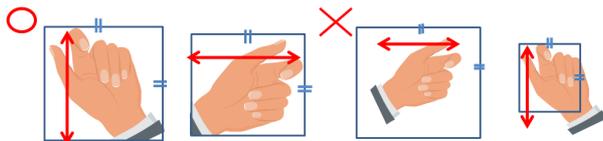


図 2 学習サンプルの取得例

## 2.2 2種類の識別器による手の検出

手領域を検出するため、AdaBoost [16] による識別器を作成し、手の検出器とする。手領域と顔領域は色特徴が類似しており、さらに重複時には手領域と顔領域の特徴が合わせて出力される。そこで、手・顔領域と背景領域を識別する識別器と手と顔領域を識別する識別器の2種類の識別器を作成する。Adaboost による識別器の出力  $F(x)$  は複数の弱識別器の出力  $f_t(x)$  の線形結合により下式のように表現される。

$$F(x) = \sum_t \alpha_t f_t(x) \quad (9)$$

ここで、 $\alpha_t$  は弱識別器  $f_t$  に対する重みを表す。弱識別器には OpenCV ライブラリ [17] による深さ 1 の決定木を用いて実装を行っている。

学習には手話画像から手動で取得した手の画像と手の一部、顔が映った画像、および背景画像を用いる。特徴量の次元を揃えるため、本手法では学習サンプルは縦、横の長さが等しい画像を用いる。さらに、手の画像を取得する際は画像内の手のスケールを揃えるため、図 2 に示すような、以下に示す条件を満たすようにする。

- 手領域の下端が矩形の下端に接する
- 手領域が矩形の左右の端に接する、もしくは上下の端に接する

学習に用いたサンプルの例を図 3 に示す。手・顔領域と背景を識別する識別器を構築する際は図 3(a)(b) をポジティブサンプル、図 3(c) をネガティブサンプルとしている。また、手領域と顔領域を識別する識別器を構築する際は図 3(a) をポジティブサンプル、図 3(b) をネガティブサンプルとして学習する。それぞれ強識別器には 300 個の弱識別器を持つものとした。重複時も検出できるよ



(a) 手領域の学習サンプル



(b) 顔領域の学習サンプル



(c) 背景の学習サンプル

図 3 学習サンプル

う手領域のサンプルには顔との重複時の画像も含めた。本研究では図 3(a) のような手領域画像を約 9000 枚、図 3(b) のような顔、もしくは顔や手の一部が写った画像を約 7000 枚、図 3(c) のような背景画像を約 13000 枚用いて学習を行う。

作成した 2 種類の識別器を用いて手領域の検出を行う。まず、入力画像に検出ウインドウを準備し、ウインドウに対して手・顔領域の検出を行う。検出ウインドウは正方形の領域とし、ウインドウのサイズは実験的に 64x64 から 160x160 の間で 16 ずつ大きさを変え、画像全体を探索する。それぞれの検出ウインドウに対し、特徴を抽出し、顔・手領域と背景領域の識別器  $F_s$  にかける。 $F_s$  の値を用いて、手の候補領域ウインドウの集合  $W$  を下式で定義する。

$$W = \{w | F_s(x(w)) > t_s\} \quad (10)$$

ここで、 $w$  は検出ウインドウ、 $x(w)$  は  $w$  における特徴ベクトル、 $t_s$  は閾値を表す。得られた  $W$  に対し、手と顔を識別するための識別器  $F_h$  にかける。 $F_h$  を可視化した結果例を図 4 に示す。画像中の輝度値は検出ウインドウの中心  $(x, y)$  における  $F_h$  の値を表す。黒い領域は、 $F_s(x(w)) < 0$  となった領域である。図 4 より、手の付近で  $F_h$  の出力値が高いことが分かる。

そこで、下式により手領域  $\hat{w}$  を推定する。

$$\hat{w} = \arg \max_{w \in W} F_h(x(w)) \quad (11)$$

図 5 のように手領域を検出することができる。

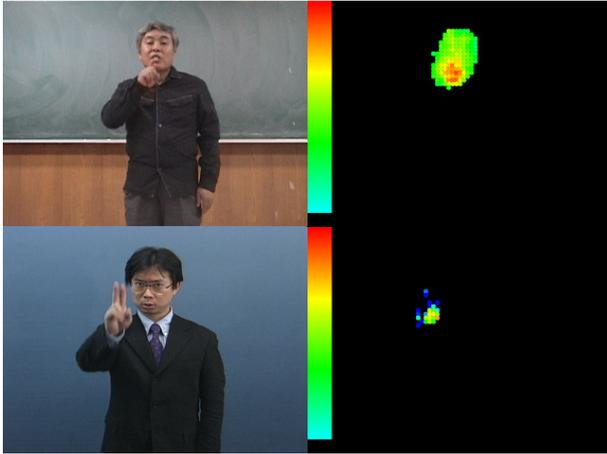


図 4 手の検出器の出力値



図 5 手領域の検出

### 3. HOG 特徴による手形状の識別

得られた手領域に対して手形状の識別を行う。同じ肌色の手領域同士の識別を行うので、色情報による識別は難しいと考えられる。そこで、物体検出に用いられる Histograms of Orientated Gradients(HOG) 特徴を入力特徴に用いる。AdaBoost を用いてある手形状か否かを識別する識別器を手形状毎に作成し手形状の識別を行う。

#### 3.1 HOG 特徴

HOG 特徴は勾配方向ヒストグラムを近接するヒストグラムのノルムで正規化して得られる特徴 [8] で、物体の形状を表すことが可能である。

本研究では学習画像を 48x48 ピクセルとし、まず、2.1.1 節で述べた手法で 6x6 のセル毎に  $h^e$  を作成する。勾配方向ヒストグラムでは境界付近のエッジは複数のセルにまたがるため、エッジ強度が分散してしまう。そこで、もともとの小領域のグリッドをグリッド幅の半分だけずらした位置でもヒストグラムを作成する。境界付近で 2 つ以上の領域に分散してしまったエッジ強度もずらしたセルでは強い特徴とすることができる。

作成した輝度の勾配方向ヒストグラムを 3x3 セルを 1 ブロックとして正規化を行う。i 行 j 列のセル  $(i, j)$  の  $h^e$  を  $h_{ij}^e$  とすると、ブロックあたりの特徴量  $b$  は

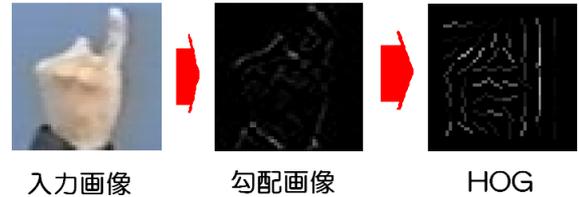


図 6 HOG 特徴量

$b = (h_{ij}^e, h_{i+1, j}^e, h_{i+2, j}^e, h_{i, j+1}^e, \dots, h_{i+2, j+2}^e)$  と表すことができる。正規化後の特徴ベクトルを  $B$  として、次式により正規化する。

$$B = \frac{b}{\sqrt{\|b\|^2 + 1}} \quad (12)$$

ブロックを 1 セルずつ移動させ、正規化を行う。最終的に得られる特徴量  $x^H$  は  $k$  番目のブロックの特徴量を  $B_k$  として、下式で表す。

$$x^H = (B_1, B_2, \dots, B_N) \quad (13)$$

ここで、 $N$  はブロックの数を表す。図 6 に HOG 特徴を可視化した例を示す。学習画像より、4941 次元の入力特徴を得る。

#### 3.2 複数の 1 対他クラスの識別器による識別

$x^H$  を入力特徴として手形状  $s$  に合致するか否かを識別する識別器  $C_s$  を形状ごとに作成し、手形状の識別を行う。

##### 3.2.1 識別対象とするクラス

本研究では手話単語の認識に必要となる、図 7 に示す 28 クラス + 1 クラス (手ではない) の識別を行う。

識別する手形状はシルエットが似ている、もしくは隠蔽が生じることが多い手形状を選んだ。HOG 特徴は見えが異なると大きく異なるため、手形状は同じでも向きが異なる場合は別クラスとして識別を行う。学習サンプルには、各クラス手動で取得した 200 ~ 300 枚の手領域の画像を用いた。

##### 3.2.2 識別処理

AdaBoost により、 $C_s$  を作成する。学習サンプルは手領域の検出器の学習に用いた図 3(a) のような手領域画像の内、形状  $s$  の画像をポジティブサンプル、 $s$  以外の画像をネガティブサンプルとして学習を行う。

図 8 に各識別器の出力結果例を示す。グラフの横軸はクラス、縦軸が出力を表している。正解クラスの識別器の出力値が最も大きいことが分かる。

このことより、下式により手形状  $\hat{s}$  を推定する。

$$\hat{s} = \arg \max_{s \in S} C_s(x^H) \quad (14)$$

ここで、 $S$  は分類対象の手形状の集合を表す。ただし、誤検出された領域を識別に用いる場合もあるため、 $C_s(x^H)$  の値が負の場合は対象が手でないとする。



図 7 分類するクラスの画像例

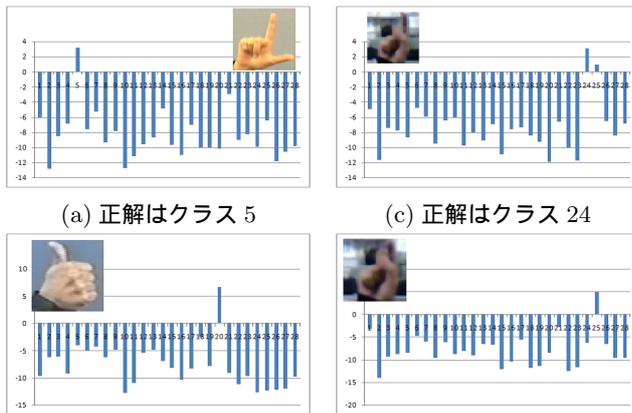


図 8  $C_s$  の出力結果例

#### 4. 手領域検出実験と手形状識別実験

提案手法の有効性を確かめるため、右手について検出器、識別器を作成し手領域の検出実験と、手形状の識別実験を行った。

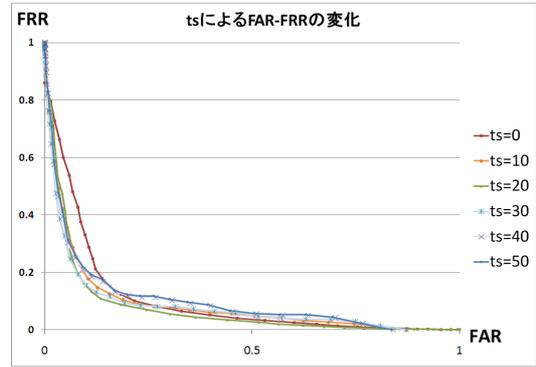


図 9 FAR-FRR 曲線

#### 4.1 手領域検出実験

作成した検出器そのものの性能を確かめるため検出器の閾値を変えながら False Accept Rate(他人受入率), False Reject Rate(本人拒絶率)を調べた。さらに、入力特徴量の有効性を確かめるため、色情報のみ、エッジ情報のみを入力特徴とした場合と検出結果を比較した。また、2段階の検出による効果を確認するため、検出器を1種類のみにしたものと検出結果の比較を行った。

##### 4.1.1 検出器の識別性能確認実験

検出器の手領域の識別性能を確かめるため、 $t_s$  を 0, 10, 20, 30, 40, 50,  $F_h$  の閾値を-100 から 100 の間で5刻みで変更しながら FAR, FRR の変化を調べた。FAR, FRR は下式により求める。

$$FAR = \frac{FP}{TN + FP}, FRR = \frac{FN}{TP + FN} \quad (15)$$

ここで、FP は非手領域で  $F_s, F_h$  とともに閾値以上の場合、TP は手領域で  $F_s, F_h$  とともに閾値以上の場合、FN は手領域で  $F_s, F_h$  いずれかが閾値以下となったウインドウの数を表している。入力画像には学習画像とは別の 100 枚の手話画像を用いた。画像の解像度は 640x480 もしくは 720x480 ピクセルである。なお、手動で取得した手領域の中心座標  $C_h$  と検出ウインドウの中心座標  $C_w$  の距離が 20 ピクセル以下のウインドウで、サイズが手動で取得した領域のサイズ  $\pm 20$  の場合を正解領域とする。

図 9 に FAR-FRR 曲線を示す。図中の閾値は  $t_s$  を表している。また、図 10, 11 に閾値による手領域の検出結果の違いを示す。左から右へは  $t_s$  を 0 から順に 10 ずつ増やしたもので、上から下へは  $F_h$  の閾値を-20 から順に 5 ずつ増やした時の検出結果を表している。

図中の 6 つのグラフの内、 $F_s$  の閾値が 20 のグラフの (FAR, FRR) が (0.11, 0.14) の時原点からの距離が最小となった。その時の  $F_h$  の閾値は-15 であった。そのため、 $F_s$  の閾値を 20,  $F_h$  の閾値を-15 とすると最も効果的な検出器を作成できると考えられる。

しかし、図 11 の場合そのような閾値を設定すると手が未検出となる。誤検出の場合は識別処理の際に手では

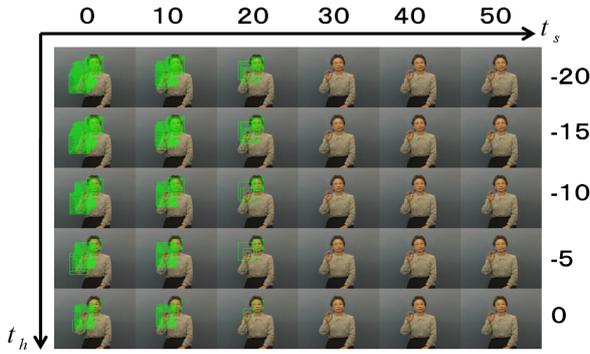


図 10 検出結果の閾値による変化 (1)

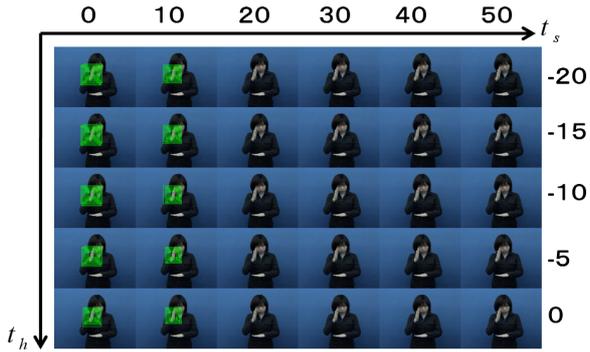


図 11 検出結果の閾値による変化 (2)

表 1 入力特徴による検出結果の比較

失敗 (1):他の場所の  $F_h(x) > 0$  が最大となり誤検出  
 失敗 (2):  $F_h(x) < 0$  又は  $F_s(x) < 0$  のため誤検出

入力特徴	成功	失敗 (1)	失敗 (2)	処理時間
色のみ	68.2%	30.4%	1.4%	165(s)
勾配方向のみ	38.1%	29.1%	36.8%	117(s)
色 + 勾配方向	78.3%	21.5%	0.2%	187(s)

ないと識別できるが、未検出の場合は識別不可能である。そのため本研究では  $t_s$  の閾値を 0 とすることとした。

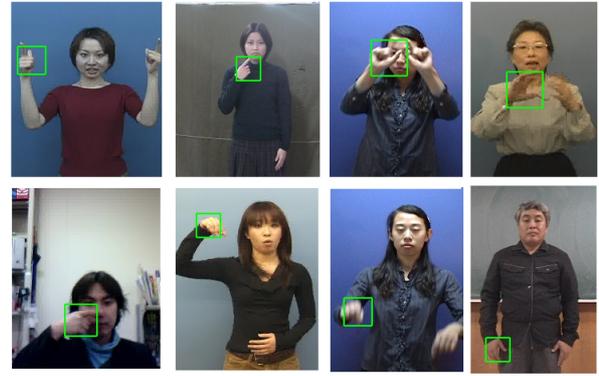
#### 4.1.2 特徴量の比較実験

入力特徴の有効性を確かめるため  $H_N^c$  のみ、 $H_N^s$  のみを入力特徴とした場合で検出性能を比較した。実際の手話動画を入力として、計 1000 フレームの画像に対して提案手法を用いて手領域の検出を行った。表 1 に検出実験の結果を示す。

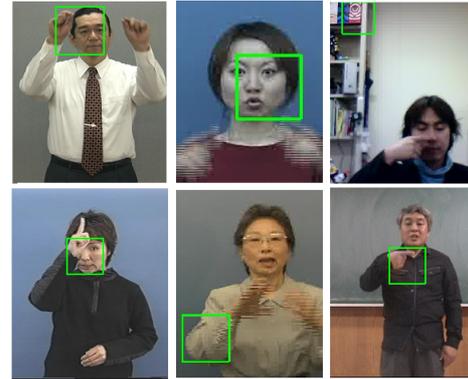
表 1 より、提案手法が色のみ、勾配方向のみを特徴とする場合と比べて効果的であるといえる。しかし、1 フレームの処理に平均 187 秒かかっており今後処理時間を短縮する必要がある。

図 12 に提案手法による検出結果例を示す。

図 12(a) のように隠蔽時や複雑背景でも手領域の検出ができた。また、従来手法では色によるセグメンテーションを行うだけであったため、半袖の人物で手領域を取得する場合、肘なども含めて手領域を取得する問題があったが、図 12(a) 左上のように半袖の人物に対し手領



(a) 検出実験成功例



(b) 検出実験失敗例

図 12 手領域検出実験結果例

域のみを取得することができている。

一方、図 12(b) に示すような約 20%の画像では検出に失敗した。検出失敗の原因は大きく分けて以下の 2 つが考えられる。

- 入力画像中の手領域がぼけている場合
- 入力画像中の手形状を学習していない場合

これらの場合、正解領域に対する  $F_h(x)$  が小さくなり、未学習の背景や顔を誤検出した。

毎フレーム画像全体から手領域の検出を行っていたが、前フレームの検出位置を利用して探索範囲を限定することで誤検出を減らし、処理時間を短縮できると考えられる。

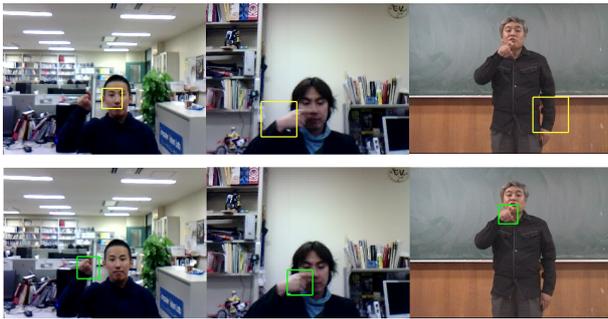
#### 4.1.3 検出器の比較実験

2 種類の検出器による検出の有効性を確かめるため、1 種類の検出器だけを用いた場合と  $F_s, F_h$  を併用した場合の検出結果の比較を行った。入力には 4.1.2 節で用いたものと同じ画像を用いた。比較用の検出器  $F_c$  の学習には、 $F_s, F_h$  の学習に用いた手領域画像 9000 枚をポジティブサンプル、顔領域や手の一部が映った画像および背景画像およそ 20000 枚をネガティブサンプルとして用いた。また、特徴の数を揃えるため  $F_c$  は 600 個の弱識別器をもつものとする。

表 2 に実験結果を示す。誤検出、処理時間ともに改善した。図 13 に改善例を示す。 $F_c$  のみの場合、顔や背景の肌色を誤検出するケースが多いが、 $F_s, F_h$  を併用した場

表 2 検出器による検出結果の比較

検出器	成功	失敗	処理時間
$F_c$ のみ	44.1%	55.9%	220(s)
$F_s, F_h$ を併用	78.3%	21.7%	187(s)



上段: $F_c$  のみ 下段: $F_s, F_h$  併用

図 13  $F_s, F_h$  による検出結果の改善

表 3 手動で取得した手領域に対する識別率の比較

識別器	識別成功	識別失敗
C-SVM	69.3%	30.7%
-SVR	74.2%	24.7%
Boost	80.9%	19.2%

合では正しく検出出来ている。また、 $F_c$  のみの手領域の判定では毎回 600 個の弱識別器にかける必要があるが、 $F_s, F_h$  を併用した場合は大半の検出ウィンドウが  $F_s$  (弱識別器は 300 個) で除外されるため、処理時間が改善したと考えられる。

## 4.2 手形状識別実験

作成した識別器の性能を確かめるため、手動で取得した手領域に対する識別実験を行った。さらに、提案手法全体の評価を行うため自動検出した手領域に対して提案手法による識別を行った。

### 4.2.1 手動で取得した領域の識別実験

識別器の性能を確認するため、手動で手領域を取得した約 900 枚の画像に対して第 3 章で作成した手形状識別器によって手領域の識別を行った。また識別性能を比較するため OpenCV ライブラリを用いて以下の 2 種類のサポートベクターマシン (SVM/SVR) を作成し、識別能力を比較した。

- 1 つの C-SVM (マルチクラスの識別が可能)
- 28 個の -SVR (各クラス毎に実数値を出力)

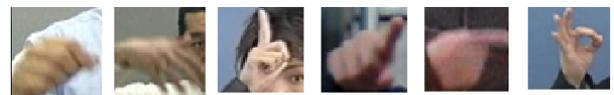
SVM, SVR を作成する際のカーネルは RBF 関数、ペナルティに関するパラメータ  $C=10.0$  とした。作成した C-SVM にはおよそ 7500 個、-SVR には平均およそ 7300 個のサポートベクターが含まれている。-SVR については提案手法同様に出力値が最大のクラスを識別結果として識別を行う。結果を表 3 に示す。C-SVM に比べて約 10%、-SVR に比べて約 5% 提案手法の方が識別率が良い結果となった。図 14(a) に識別に成功した例を示す。左 2 つのように輪郭が似ているが別の形状 (右側



(a) 識別成功例 (上段: 入力画像, 下段: 正解クラスの画像例)



(b) 誤識別例 (上段: 入力画像 中央: 識別結果の例 下段: 正解クラスの画像例)



(c) 非手領域と識別した例

図 14 予備実験結果例

は人差し指を曲げている) の手形状を識別している。また、中央のように隠蔽を起こしている場合も識別に成功している。また奥行き方向に指を伸ばしているような手形状も識別に成功しており、輪郭が似ている、肌色領域同士で隠蔽を起こしている場合にも提案手法が有効であると考えられる。

図 14(b)(c) に識別に失敗した例を示す。識別器が誤識別をした入力動作途中のものが多い。学習サンプルは実際の手話動画から取得したものがほとんどであるが、はっきりとした形状の時だけを学習画像として取得している。その為、少しぼけた場合や手形状が変化する途中のサンプルがなく、誤識別を起こしたと考えられる。しかし、識別が難しい場合は動きがあまりない時の識別結果から、手形状を推定することができると考えられる。また、動作途中の場合、手形状は手話認識時には重要な要素でないため、少しの誤識別は許容される。

### 4.2.2 自動検出した領域の識別実験

手領域の検出実験で検出された領域に対して、手領域の種類を識別する実験を行った。予備実験同様、比較対象として SVM を用意した。取得した手領域の内、明らかに手領域を検出していない領域は、実験対象としない。

実験結果を表 4 に示す。また、図 15 に実験の結果例を示す。

SVM と比較すると約 2 倍の識別率となった。図 15(a) は一連の画像系列で、重複したまま手形状が変わっているが提案手法により形状の識別が可能となった。しかし、予備実験と比較すると 30% 識別率が下がった。図 15(b) に識別失敗例を示す。識別失敗例では、手領域の検出の

表 4 自動検出した手領域に対する識別率の比較

識別器	識別成功	識別失敗
C-SVM	25.9%	74.1%
Boost	51.8%	48.2%



(a) 識別成功例 ( 上段 : 手領域検出結果 , 中段 : 手領域検出例 , 下段 : 識別結果の例 )



(b) 識別失敗例

図 15 手領域識別実験結果例

時点では成功したといえるが、いずれの画像も識別時には非手領域であるという出力を得た。学習用の画像は図 2 に示すような手領域に外接する長方形を左右もしくは上下に広げて正方形としたものである。そのため、学習画像中の手の大きさは手形状毎に正規化される。一方、検出器のウィンドウサイズは 16 刻みで変化する。そのため、検出ウィンドウ内に写る手のサイズと学習画像中の手のサイズに誤差が生じる場合がある。そのため、識別率が下がったと考えられる。今回の実験では  $\hat{w}$  が手領域から少しずれる場合や、顔や服を大きく含む場合があったが、 $\hat{w}$  以外では手領域の識別をしていない。 $C_s(x^H)$  が最大な点が必ず  $F_h(x)$  も最大になるとは限らないため、 $\hat{w}$  に対してのみ手領域の識別を行うのではなく、周辺のウィンドウでも識別を行うことで識別率が上がるのではないかと考えられる。

## 5. おわりに

見えを学習することで、手話画像中から手領域を検出し、さらに検出した手領域から手形状を識別する手法を提案した。手領域検出実験より、複雑背景中や隠蔽時、被験者が半袖の服を着ている場合においても提案手法により手領域の検出ができることを確認した。また、手形状識別実験より提案手法によって輪郭が類似している形状や隠蔽時でも識別が可能であることを確認した。今後は、手形状の識別率の改善を図り、手話認識システムへの応用を目指す。

## 文 献

- [1] 長嶋祐二, 神田和幸, "手話のコンピュータ処理," "電子情報通信学会誌, Vol.84, No.5, pp.320-324, May. 2001.
- [2] L. Yoshino et al, "Recognition of Japanese sign language from image sequence using color combination," Proc. 3rd Int. Conf. Image Processing, pp.16-19, (1996)
- [3] 吉富康成, 永山しづえ, 杉山雅祥, "温度画像処理による手軌跡の抽出と手話認識," "信学技報, SP2002-112, Vol.102, No.418, pp.21-26, Oct.2002.
- [4] 岡澤裕二, 堀内靖雄, 市川薫, "オプティカルフローによる手話の大局的動作の認識について," "信学技報, PRMU2002-77, Vol.102, No.317, pp.39-44, Sep.2002.
- [5] 星野聖, 小渡信, 神里志穂子, 新垣武士, "手話認識のための 3 次元手指運動推定 "信学技報, NLP2000-170, Vol.100, No.681, pp.43-49, Mar.2001.
- [6] 李昌宏, 中園薫, 長嶋祐二, 張鴻徳, "動きベクトルを用いた手話単語分類," "信学技報, WIT2003-62, Vol.103, No.746, pp.65-70, Mar.2004.
- [7] Yuji Yamauchi and Hironobu Fujiyoshi, "People Detection Based on Co-occurrence of Appearance and Spatiotemporal Features," International Conference on Pattern Recognition 2008(ICPR2008), Dec.2008.
- [8] N.Dalal and B.Triggs, "Histograms of Oriented Gradients for Human Detection," IEEE Computer Vision and Pattern Recognition, 886-893,(2005)
- [9] 木場洋介, 石川真澄, "動画像からの移動物体抽出と手話動作認識への応用," "信学技報, NC2002-170, Vol.102, No.729, pp.221-226, Mar.2003.
- [10] 今川和幸, 谷口倫一郎, 有田大作, 松尾英明, 呂山, 猪木誠二, "カメラを用いた手話認識における見えの違いを考慮した手話の局所特徴認識," "映像情報学会誌, Vol.54, No.6, pp.848-857, Jun.2000.
- [11] 山本貢嗣, 野村健, 南角吉彦, 後藤富朗, 北村正, "HMM に基づく日本手話認識のための特徴の統合に関する検討 "電子情報通信学会総合大会講演論文集, D-12-121, p.271, Mar.2005.
- [12] 柳哲, 柳生雄午, 徳田恵一, 北村正, "手の動作と形状を用いた HMM 手話認識 "電子情報通信学会総合大会講演論文集, D-12-119, pp.285, Mar.2004.
- [13] 川東香菜, 白井良明, 島田伸敬, 三浦純 "手話の HMM 作成のための状態分割 "信学技報, WIT2005-21, Vol.105, No.67, pp.55-60, May.2005.
- [14] 今井章博, 島田伸敬, 白井良明, "複雑背景下におけるモデルの照合誤りを考慮した手指形状推定 "信学論 (D) Vol.J91-D, No.3, pp.784-792, Mar.2008.
- [15] S.C.W. Ong et al, "Automatic Sign language Analysis: A Survey and the Future beyond Lexical Meaning" IEEE Trans. PAMI, Vol.27, No.6, pp.873-891, Jun.2005.
- [16] Friedman, J. H., Hastie, T. and Tibshirani, R., "Additive Logistic Regression: a Statistical View of Boosting," The Annals of Statistics, Vol.28, No.2, pp. 337-374, Apr. 2000
- [17] The Open Computer Vision Library; <http://sourceforge.net/projects/opencvlibrary/>