社団法人　電子情報通信学会
THE INSTITUTE OF ELECTRONICS,
INFORMATION AND COMMUNICATION ENGINEERS

信学技報
TECHNICAL REPORT OF IEICE.

SVMと逐次学習を併用したHOG特徴による手形状認識手法

武藤　亮介†　　嶋田　和孝†　　遠藤　　勉†

† 九州工業大学　情報工学部
福岡県飯塚市川津 680-4
E-mail: {r_mutou,shimada,endo}@pluto.ai.kyutech.ac.jp

あらまし　本稿では，少数のサンプルで学習したSVMと逐次学習により作成される識別器を併用したHOG特徴による手形状認識手法を提案する．HOG特徴を用いた形状認識ではSVMを用いて認識を行う研究が多数報告されているが，一般的に機械学習には十分な学習が必要であり，学習に存在しない人物の認識精度が低下することも起こり得る．本手法では，入力画像からHOG特徴を算出し，それをSVM，逐次学習により作成される識別器それぞれに入力する．それぞれの結果を一定条件で選択することにより，高精度で汎用性の高い手形状認識を行う．提案手法とSVM単体の結果と比較したところ、提案手法の有効性を確認することができた．
キーワード　手形状認識，HOG特徴，逐次学習

Hand Shape Recognition based on SVM and online learning with HOG

Ryousuke MUTOU†, Kazutaka SIMADA†, and Tutomu ENDOU†

† Kyushu Institute of Technology Computer Science and Systems Engineering
kawazu 680-4, Iizuka, Fukuoka
E-mail: {r_mutou,shimada,endo}@pluto.ai.kyutech.ac.jp

Abstract　In this paper, we proposed a combined method for hand shape recognition. It consists of support vector machines (SVMs) and an online learning algorithm based on the percepron. We apply HOG features to each method. First, our method estimates a hand shape of an input image by using SVMs. Also the online learning method with the perceptron uses the input image as training data if the data possesses a high confidence score in the recognition process. Next, we select the final hand shape from the outputs from the SVMs and perceptron by using the score from SVMs. We compared the combined method with SVMs. The experimental results show the effectiveness of the proposed method.
Key words　hand shape recognition, online learning, HOG

1. はじめに

近年，直感的に扱えるユーザインタフェース (UI) が重要視されており [1]，コンピュータを直感的に扱う手段の一つとして，iPhoneやニンテンドーDSなどに見られるようなタッチスクリーン技術を用いたUIが，私たちの身近な物となってきた．しかし，これらタッチスクリーン技術を用いたUIには，画面が大きいとすべての領域をタッチすることが困難といった大画面での課題や，直接触ると画面が汚れるといった問題があるため，画面を触らずに操作ができる非接触型のUIが必要とされている．非接触型UIの一つとして，手指情報を用いて操作を行うハンドジェスチャUIが挙げられる．ハンドジェスチャUIの構築には，手形状の認識が非常に重要となってくる．手形状の認識手法としては，データグローブや距離センサのような特殊な装置を使用する方法が多く取られているが，機材費用や設置場所などユーザへの負担が大きい．

また，特殊な装置を使わずにカメラのみで手形状認識を行う手法もある．こちらの手法ではユーザの負担は軽減されるが，使用できるジェスチャが限られる，ロバストな検出が行えないといった問題がある．また，肌色情報のみを用いて手指情報の取得を行うので，顔と手が重なると認識できない．

本研究では，誰でも直感的に作業が行えるUIの構築を目指し，HOG特徴による手形状認識手法を提案する．エッジベースの特徴であるHOG特徴を用いることにより，手と顔が重なった状態でも認識することが可能となる．HOG特徴を用いた形状認識ではSVMを用いて認識を行う研究が多数報告されてい

るが，一般的に機械学習には十分な学習が必要であり，学習に用いていない人物の認識精度が低下することも起こり得る．そこで，SVM と逐次学習器を併用することで，この問題に対応する．

2. 先行研究

まず提案手法の要となる HOG 特徴及び HOG 特徴を用いた従来手法の問題点について述べる．また，その問題点の対応法も合わせて本節で述べる．

2.1 HOG 特徴とそれを用いた従来手法

HOG 特徴とは，2005 年に Navneet Dalal と Bill Triggs によって提唱された特徴であり [2]，入力画像の勾配 (微分画像) を求め、それを局所領域ごとに勾配方向で区間分割してヒストグラムを取ったものを特徴とする．この特徴の性質は，局所的な幾何学的変化，明度変化に対して不変であるが，回転，スケール変化に対しては不変ではないとされている．

以下で，HOG 特徴の算出法について述べる．

（1） 入力画像を一定サイズに正規化する．

（2） 各ピクセルにおける勾配強度 m と勾配方向 θ を以下の式を用いて算出する．

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

$$\theta(u,v) = tan^{-1}\frac{(f_x(x,y)}{f_y(x,y)]} \tag{2}$$

$$f_x(x,y) = I(u+1,v) - I(u-1,v) \tag{3}$$

$$f_y(x,y) = I(u,v+1) - I(u,v+1) \tag{4}$$

（3） $5*5$ ピクセルを 1 セル、$0° - 180°$ を $20°$ ずつ 9 方向に分割し、1 セルごとに輝度勾配ヒストグラムを作成する．

（4） 各セルにおいて作成したヒストグラムを $3*3$ セルを 1 ブロックとして正規化を行い特徴量を算出する．正規化はブロックを 1 セルずつずらしながら全領域に対し行う．

上記の手法で算出した HOG 特徴と SVM を組み合わせて，人検出や車検出など一般物体の認識を行う研究が多数報告されている [2,4,7,8]．山内ら [7] は，大量のサンプルで学習した SVM で，画像全体を予め決められた大きさで切り取りながら，ラスタスキャンすることにより高精度な人物検出を行っている．また，山田ら [8] は HOG 特徴を用いることにより，手と顔が重なった状態でも高精度に手形状の認識することが可能であると報告している．

2.2 従来手法の問題点とその対応

事前実験として 2.1 節で述べた山内らの手法を参考に，図 1 のように手の領域が映った画像の HOG 特徴を学習した SVM を用いて，ユーザの上半身が映った画像から手の位置推定及び手形状認識を行った．その結果，以下のような問題点があることがわかった．

（1） 画像全体をラスタスキャンし手形状認識を行うのでは，1 フレームあたりの手の位置及び手形状の認識に多大な計算時間がかかる．

図 1 手の領域が映った画像

（2） 高精度な認識には，事前に十分な学習が必要である．

（3） 手の形は人によって指の長さなどが異なるため，学習した人物の手は高精度に認識できても，学習していない人物の手の認識精度は必ずしも高くない．

そこで提案手法では，問題点 1 に関しては，予め肌色領域を取得し，その周辺のみ認識することにより，認識時間を短縮する．問題点 2，3 に関しては，理論的にいえば，使用するユーザ毎の学習サンプルを大量に用意すれば，解決できる．しかし，ユーザは固定されているわけではないため，ユーザ毎に大量の学習サンプルを用意するのは，現実的ではない．そこで，逐次学習器を SVM と併用することで，入力された画像を認識すると同時に学習も行う．そして，SVM から出力される信頼度に応じて SVM と逐次学習器の出力結果を選択することにより，高精度な認識精度は維持しつつも上記の問題点 2，3 に対応する．逐次学習器にはパーセプトロンを用いた．

3. 提案手法

2 節で，従来手法の問題点とその対応法について簡単に述べた．本節では，提案手法の詳細について述べる．

3.1 手法の概要

まず，提案手法の概要について述べる．本手法は USB カメラを入力デバイスとし，取得した画像に HOG 特徴による認識処理を適用することにより，ユーザの手形状認識を行う．認識処理の流れを図 2 に示す．認識処理は大きく分けて，キャリブレーションプロセス，HOG 特徴抽出プロセス，手形状認識プロセスの 3 段階ある．以下で，各々のプロセスの詳しい手法について述べる．

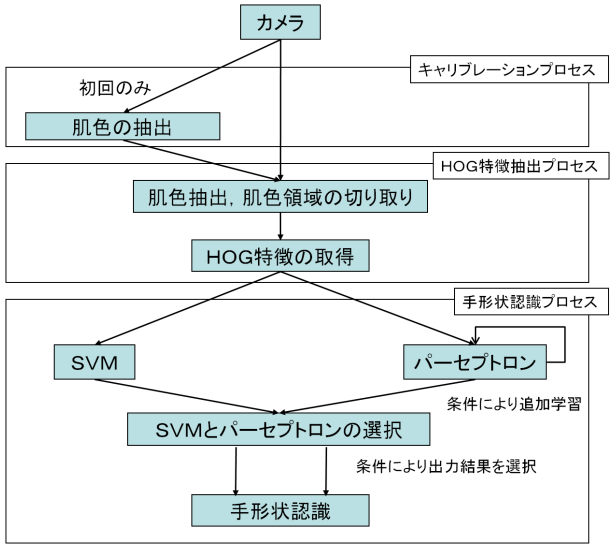


図 2 処理の流れ

3.2 キャリブレーションプロセス

本手法では，予め肌色領域を取得しその周辺のみ認識することにより，認識時間を短縮する．そのため，ユーザの肌色を抽出する必要がある．そこで，認識開始時にキャリブレーション (ユーザに指定の場所に手を合わせてもらう (図 3)) を行い，ユーザの肌色の抽出を行う．ユーザの肌色抽出法を以下に示す

（1） 画像を HSV に変換する．

（2） 図 3 の肌色領域内にある画素をランダムに 20 個取得する，取得した画素の中から H, S, V それぞれの最大値，最小値を算出する．

（3） 以下の範囲内の画素をユーザの肌色とする．

$$minH - 5 <= H <= maxH - 5 \quad (5)$$

$$minS - 5 <= S <= maxS - 5 \quad (6)$$

$$minV - 5 <= V <= maxV - 5 \quad (7)$$

3.3 HOG 特徴抽出プロセス

3.2 節で取得したユーザの肌色を画像全体から検索することにより，肌色領域を抽出する．続いて，取得した肌色領域から認識する領域の切り取りを行う．切り取り処理は以下の手順で行う．

（1） 肌色抽出画像をラベリングし、同じラベルの面積を算出する．

（2） ラベルの面積が閾値以下なら，その領域を含む矩形で切り取りを行う (図 4)．

（3） ラベルの面積が閾値以上なら，大きさの違う検索ウインドウで肌色周辺を 10 ピクセル飛ばしでラスタスキャンを行う．スキャンした領域内で肌色が一定面積以上ある場合のみ，切り取りを行う (図 5)．検索ウインドウの大きさは $90 * 100$, $60 * 80$，$50 * 60$ の 3 種類を用意した．

切り取った画像に対し，2.1 節で述べた HOG 特徴の算出法を適用することにより，切り取った画像各々の HOG 特徴を算出する．算出した特徴量を SVM とパーセプトロンを用いた逐次学習器それぞれに入力する．

3.4 手形状認識プロセス

本手法では，3.3 節により算出した HOG 特徴を SVM とパーセプトロンに入力し，SVM から出力される信頼度に応じて SVM とパーセプトロンの出力結果を選択することにより，手形状の認識を行う．以下では，まず SVM 及びパーセプトロンについて述べ，最後に SVM とパーセプトロンの選択法について述べる．

3.4.1 SVM

SVM は，Vapnik [6] が考案した Optimal Separating Hyperplane を起源とする超平面による特徴空間の分割法であり，現在，2 値分類問題を解決するための最も優秀な学習モデルの一つとして知られている．

しかし，SVM を単純に使用するのでは，2 値分類しか行えない．多くの手形状認識を行うには，SVM で多クラス分類を行えるようにする必要がある．本手法では，全てのクラスの組み合わせで 2 値分類の問題を作って多数決を取る手法で多クラス分類を行う．以上により，分類する画像の予測されるクラスと各クラス毎の信頼度を得る．

図 3　キャリブレーション

図 4　ラベルの面積が閾値以下の場合

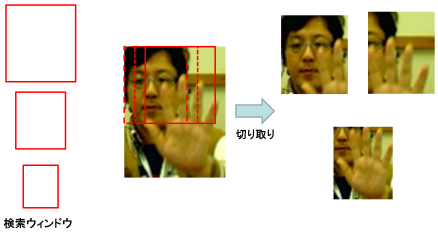


図 5　ラベルの面積が閾値以上の場合

3.4.2 パーセプトロン

本手法の逐次学習器にはパーセプトロンを用いる．パーセプトロンは，米国の Rosenblatt [3] が 1958 年に発表した，ニューラルネットの一種である．視覚と脳の機能をモデル化したものであり．学習を行いながらパターン識別をする．入力層，中間層，出力層の 3 つの部分からなる．入力層と中間層の間はランダムに接続されている．入力層には外部から信号が与えられる．中間層は入力層からの情報を元に反応する．出力層は入力層の答えに重みづけをして，多数決を行い，答えを出力する．

パーセプトロンの主な特徴は，次のとおりである．

- 入力層、中間層、出力層の三つの層から出来ている．
- 信号は一方通行で、入力層から出力層へと伝送される．
- 各層のユニット数は任意で決められる．
- 各層の入力値、出力値はともに 0 か 1 のみである．

パーセプトロンも SVM と同様に単純に使用するのでは，2 値分類しか行えないため，パーセプトロンでも多クラス分類を行えるようにする必要がある．本手法では以下に示す手順でパーセプトロンの多クラス分類を実現する．

（1） それぞれの手形状毎に，入力された画像が，その手形状なのか，又は手以外 (顔や腕の一部) なのか分類する識別器を作成しておく．

（2） 1 で作成した識別器に，分類したい画像を入力し，各手形状毎に識別器のスコアを算出する．

（3） 2 で出力されたスコアが一番高い識別器の手形状を，分類したい画像の手形状とする．

パーセプトロンを用いた逐次学習に限らず，逐次学習には学

習の早さと，正しい解への収束性にトレードオフの関係があるため，どのような条件下で追加学習するかが非常に重要とされている．そこで本手法は，まず予め全ての手形状の組み合わせで，入力された画像がどちらの手形状に近いかを判別する識別器を作成しておく．そして，

`First` ： 分類する際に値が最も高い識別器のスコア
`Second` ： 分類する際に値が 2 番目に高い識別器のスコア
`Fshape` ： `First` の識別器の手形状
`Sshape` ： `Second` の識別器の手形状
`FvsS` ： 入力された画像が `Fshape` と `Sshape` のどちらの手形状に近いか分類する識別器のスコア

とすると，以下に示す条件の時に分類された画像を用いて追加学習を行う．条件の値は実験的に決定した．

- `First` ＞ 1.0 かつ `FvsS` ＞ 0.7 の時，`Fshape` として追加学習
- 0 ＞ `First` ＞− 0.015 かつ `FvsS` ＞ 0.7 の時，`Fshape` として追加学習
- 0.015 ＞ `First` ＞ 0 かつ 0.015 ＞ `FvsS` ＞ 0 の時，手以外として追加学習

3.5 SVM とパーセプトロンの選択

本手法では，SVM から出力される信頼度に応じて SVM とパーセプトロンの出力結果を選択することにより，手形状の認識を行う．選択する条件としては，SVM より出力される認識結果の信頼度と手以外の信頼度に 0.3 以上の差がない時とした．0.3 という条件の値は実験的に決定した．処理の流れを図 6 に示す．

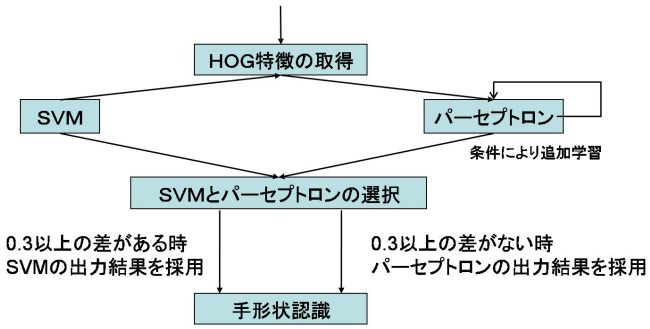


図 6　SVM とパーセプトロンの選択の流れ

4. 実　　験

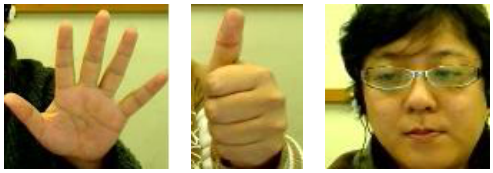
4.1 実験環境

まず，実験環境を以下に示す．

- PC
  - CPU: Intel(R) Xeon(R) X5450 3.00GHz
  - メモリ: 4.0GB
  - OS: Windows XP
- 使用したカメラ
  - Logitech 社 QuickCam Fusion
  - 画像サイズ: 320 × 240
  - カメラから PC への画像の転送速度：30fps
- 照明環境：白色蛍光灯下

4.2 評価実験

評価実験として，図 7 に示す 2 つの手形状と手以外(注1)の認識を提案手法を用いて行い，提案手法の認識精度を算出する．2 つの手形状それぞれには，図 8 のように手と顔が重なっている状態の画像も含まれる．SVM の実装には libSVM(注2)を使用した．また，パーセプトロンの実装には Oll(注3)を使用した．



Open　　Up　　Other

図 7　認識する手形状と手以外の図

Open　　Up

図 8　手と顔が重なっている状態

実験の詳細は以下のようにした．

- 学習データ
  - 男性 2 名，女性 2 名，計 4 名
  - 120 枚 (各形状一人 10 枚ずつと手以外 10 枚)
- テストデータ
  - 男性 2 名，女性 2 名，計 4 名 (男 1 名、女 1 名は学習に用いた人とは違う人)
  - 150 枚 (各形状 50 枚ずつと手以外 50 枚)
- 追加学習データ
  - 一人当たり時系列で取得した画像 480 枚
- HOG 特徴を算出する際の正規化サイズ
  - $80 * 90$

その結果を表 1 に示す．また，同じデータで学習した SVM 単体を用いて認識を行った実験結果も合わせて表 1 に示す．なお，表中の学習の有無は学習データとして使用した人物で有るか無いかを示しており，学習に使用した人物の学習データとテストデータは同一ではない．

その結果，SVM を単体で用いるより，提案手法の方が高精度であった．また，提案手法では学習した人物の手形状と学習をしていない人物の手形状の認識精度に偏りが見られないため，汎用性は高いといえる．

手形状認識プロセスの処理時間は，一枚あたり平均 0.03 秒

(注1)：Other には顔だけではなく，背景や服なども含まれる．
(注2)：libSVM: http://www.csie.ntu.edu.tw/ cjlin/libsvm/
(注3)：Oll: http://code.google.com/p/oll/wiki/OllMainJa

表 1　提案手法のハンドジェスチャ認識手法の認識精度

被験者		被験者 A	被験者 B	被験者 C	被験者 D
学習の有無		学習有り	学習無し	学習有り	学習無し
Open	提案手法	88%	92%	98%	**96%**
	SVM	88%	92%	98%	88%
Up	提案手法	**94%**	**98%**	100%	**96%**
	SVM	90%	96%	100%	90%
Other	提案手法	98%	**98%**	100%	98%
	SVM	98%	96%	100%	98%
個人平均	提案手法	**93%**	**96%**	99%	**96%**
	SVM	92%	92%	99%	92%
全体平均	提案手法	**96%**			
	SVM	94%			

であった．この処理時間では，手と顔が離れている状態 (図 9) なら，顔を切り取った領域と手を切り取った領域の 2 か所をのみ判別すればいいので，1 フレームあたり 0.06 秒で高速な手の検出及び認識ができる．

しかし，手と顔が重なっている状態 (図 10) ならば，3.3 節で説明したように，大きさの違う検索ウインドウで肌色周辺を 10 ピクセル毎にラスタスキャンを行うため，図 9 の時と比べ，多くの領域を判別しなければならない．検証したところ，手と顔が重なっている状態で正確な手の検出及び認識を行うには，30 か所程度の領域判別をする必要があり，1 フレームあたりに 1 秒程度必要であった．つまり，手と顔が離れている時には高速な認識が行えるものの，手と顔が重なった瞬間に急激に認識速度が低下する．これは UI としては致命的である．よって，現状の提案手法を UI として用いるには，カメラから見て手と顔が重ならないように作業を行うという制約を付けなければならない．

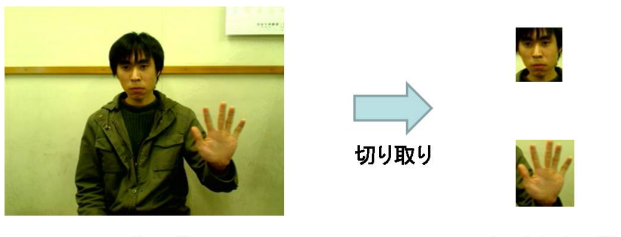


図 9　手と顔が離れている時

図 10　手と顔が重なっている状態

4.3　追加実験

また，追加実験として認識する手形状を，先ほど示した 2 種類の手形状に以下の 4 種類の手形状 (図 11) を加え，認識する手形状が多数ある場合においても提案手法が有効かどうか検証を行った．

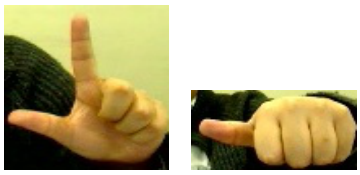


図 11　追加した手形状

実験の詳細は以下のようにした．

- 学習データ
  - 男性 2 名，女性 2 名，計 4 名
  - 560 枚 (各形状一人 20 枚ずつと手以外 20 枚)
- テストデータ
  - 男性 2 名，女性 2 名，計 4 名 (男 1 名、女 1 名は学習に用いた人とは違う人)
  - 140 枚 (各形状 20 枚ずつと手以外 20 枚)
- 追加学習データ
  - 一人当たり時系列で取得した画像 480 枚
- HOG 特徴量を算出する際の正規化サイズ
  - $80 * 90$

その結果を表 2 に示す．また，同じデータで学習した SVM 単体を用いて認識を行った実験結果も合わせて表 2 に示す．なお，こちらも先ほどと同様に，表中の学習の有無は学習データとして使用した人物で有るか無いかを示しており，学習に使用した人物の学習データとテストデータは同一ではない．

その結果，手の形状が多くなっても，提案手法は有効であるということがわかった．また，Side と Peace の認識精度が悪いことから，認識しやすい形とそうでない形があるということがわかった．しかし，これは今回テストデータが各形状 20 枚しか用意していないため，テストデータに偏りがあった可能性もある．そのため，今後は，この問題について十分に検討する必要がある．また，2 種類の形状を認識する時に比べれば精度は落ちた．このことから，高精度な認識を行う為には，認識する手形状は使用する目的に合わせて少なく決める必要があることがわかった．処理速度に関しても 1 枚平均 0.05 秒と増加してしまった．

表 2　提案手法のハンドジェスチャ認識手法の認識精度

被験者		被験者 A	被験者 B	被験者 C	被験者 D
学習の有無		学習有り	学習無し	学習有り	学習無し
Open	提案手法	100%	**95%**	95%	**95%**
	SVM	100%	85%	**100%**	75%
Up	提案手法	**95%**	100%	100%	**95%**
	SVM	90%	95%	100%	85%
L	提案手法	90%	**90%**	90%	**90%**
	SVM	**95%**	80%	**95%**	75%
Side	提案手法	**90%**	**75%**	**70%**	**70%**
	SVM	55%	40%	55%	40%
Down	提案手法	100%	**95%**	100%	**95%**
	SVM	100%	90%	100%	90%
Peace	提案手法	**95%**	**80%**	**85%**	**75%**
	SVM	60%	50%	70%	50%
Other	提案手法	100%	95%	**100%**	95%
	SVM	100%	**100%**	95%	95%
個人平均	提案手法	**95%**	**90%**	**91%**	**87%**
	SVM	85%	77%	88%	73%
全体平均	提案手法	**91%**			
	SVM	81%			

5.　提案手法の応用

5.1　iTunes 操作アプリケーション

提案手法の応用として，音声とジェスチャで操作を行うマルチモーダル UI を構築した．また，そのマルチモーダル UI の適用例として，iTunes の主要機能を，手形状と音声に割り当てることにより，気軽に離れた場所からでも操作できる iTunes 操作アプリケーションを作成した．音声理解には，我々の研究室で開発されている複数の音声認識器に基づく音声理解手法 [5] を用いた．この手法では，出力される音素が似ていれば，それはアプリケーションに対する命令発話とする．一方で，音素が似ていなければ，それは雑談だとすることにより，命令発話と雑談の聞き分けを行う．対応させた手形状は以下の図 12 ようになる．また，音声で認識できる命令としては“ 再生 ”,“ 停止 ”,“ シャッフル ”を定義した．第三者に実際に使用してもらったところ,「簡単に操作が行えて楽しい」といった感想を得た．単純な操作であれば本手法でも快適に操作できることを確認した．

6.　ま　と　め

本研究では，誰でも直感的に作業が行える UI の構築を目指し，SVM と逐次学習を併用した HOG 特徴による手形状認識手法を提案した．HOG 特徴を用いた形状認識には SVM を用いて認識を行う研究が多数報告されているが，SVM に限らず機械学習では，十分な学習が必要であり，学習に用いていない人物の手は高精度に認識できないという問題もあった．そこで，本研究では SVM とパーセプトロンを用いた逐次学習器を併用することにより，この問題に対応した．その結果，本手法は高精度な手形状認識ができ，汎用性も高かったが処理速度に問題があり，ユーザにカメラからみて顔と手が重ならないように作業を行うという制約を付けなければならなかった．しかし，手と顔が重なっても認識はできるため，今後領域の切り取り処理や HOG 特徴の算出処理を改善し，高速化を行うことにより，この制約を解消していきたい．

図 12　操作と手形状の対応

図 13　音声とジェスチャで操作を行う iTunes 操作アプリケーション

文　　献

[1] A. Dam, ”Post-WIMP User Interfaces” Communication of The ACM, February 1997, Vol.40, No.2

[2] N. Dalal, B. Triggs,“ Histograms of oriented gradients for human detection ”, CVPR05, 2005.

[3] R. Frank,“ The Perceptron: A Probabilistic Model for Information Storage and Organization in the Brain ”, Psychological Review 65 (6): 386-408, 1958.

[4] F. Han, Y. Shan, R. Cekander:,“ A Two-Stage Approach to People and Vehicle Detection with HOG-Based SVM ”, PerMIS, pp. 133-140, 2006.

[5] K. Shimada, Satomi Horiguchi and Tsutomu Endo,“ An Effective Speech Understanding Method with a Multiple Speech Recognizer based on Output Selection using Edit Distance ”, Proceedings of the 22nd Pacific Asia Conference on Language, Information and Computation (PACLIC22), 2008.

[6] V. N. Vapnik,“ Statistical Learning Theory ”, John Wiley & Sons, 1999.

[7] 山内　悠嗣, 藤吉弘亘, Bon-Woo Hwang, 金出武雄,“ アピアランスと時空間特徴の共起に基づく人検出 ”, 画像の認識・理解シンポジウム (MIRU08), pp.207-214, 2008

[8] 山田 寛, 島田 伸敬, 白井 良明,“ 画像処理による手話認識のための手形状識別 ”, 電子情報通信学会 2009 年総合大会, A-19-001, pp.331, 2009