H-010

複雑背景下における単眼カメラによるハンドジェスチャの認識
Hand Gesture Recognition with a Monocular Camera in Complex Background

根本 祐也† 中島 克人†
Yuya Nemoto Katsuto Nakajima

1. はじめに

単眼カメラによるハンドジェスチャ認識は，安価にシステムを構成できるが，手の認識速度と精度に課題がある．特に複雑背景下ではノイズ除去等のための前処理に時間を要する．また，手形状の変化全てに対応する事は困難であり検出精度に影響が出る．

本稿では，複雑背景下での実時間ハンドジェスチャ認識システムを目的とした高速な手領域検出手法を提案すると共にその評価結果を報告する．

2. 提案手法の概要

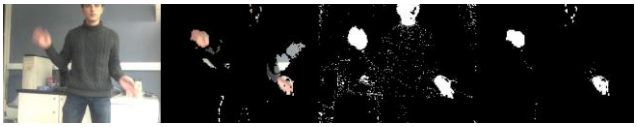
本研究ではハンドジェスチャ認識を，(1) 手の位置と腕の方向の検出，(2) 検出位置での手形状識別，(3) 手形状の遷移によるジェスチャ認識，の 3 段階で行う．これにより，実時間処理のための高速化と認識の精度向上を図る．本稿では，(1)に関して，高速な手領域検出手法提案する．

2.1 手の位置と腕の方向の検出

我々は，ジェスチャが動きを伴う事に着目し，まず，前処理として動きのある肌色領域を前景として抽出し，次に，平滑化やモロフォロジ演算等を施すことなく，積分画像を用いて直接前景画素の多く含まれる領域を見つけ，更に以下に示すように，その領域内の大まかな解析により，手の形状に関わらず，手の位置と腕の方向を高速に検出する．

2.1.1 前景抽出

肌色画素を単純に抽出すると，人の肌色部分以外に背景にあるダンボール等の肌色に似た物体までが抽出されてしまう．そこで，ジェスチャ中の手の認識が出来れば良いとの考え方の基に，動きのある肌色画素を手の候補領域(前景)とする．動きのある部分の抽出には，照明変動に強い動的背景推定・背景差分法[1]を用いる．また，肌色は，[2]より色相（Hue）の6≦Hue≦38の範囲とする．図1 (d)で前景抽出結果の例を示す．なお，抽出した「動きのある肌色」を本節以降では単に「肌色画素」と呼ぶ．



(a) 元画像 (b) 背景差分 (c) 肌色抽出 (d) 提案手法

図 1 前景抽出の結果

2.1.2 手の位置の検出

手は人体の末端器官であり，必ず手首の先に位置する事から，図 2 に示す 2 重矩形の探査窓で手の候補領域を検出する．即ち，この 2 重矩形を用いて，以下の条件を満たす領域を手の候補領域として検出する．

・内側領域：ある一定の多数の肌色画素が占める．
・縁領域：ある一定の範囲でまとまった肌色画素が占め，その他は肌色画素が殆どない

なお，この条件を満たす肌色画素の占有比率の閾値を，$\alpha l \leqq$ 内側領域$< \alpha h$ (αl, αh：閾値)，$\beta l \leqq$ 縁領域$< \beta h$ (βl, βh：閾値)と定義する．

図 2 2 重の矩形

2.1.3 縁領域における腕の方向の検出

縁領域における手首位置を推定は，手の形状認識やジェスチャの認識の精度向上に寄与する．

ここでも，積分画像による高速化を図るために，縁領域を図 3 右のように 16 のブロックに分割し，それぞれの肌色画素比率の計測を行う．縁ブロックにおいて，肌色画素の占有比率が閾値 γ 以上 のものが 1 つ，または，2 つだけであり，かつ，後者の場合， それが連続する場合にのみ，そこが手首であると判断する．連続しない場合は，肌色画素数が多いものを手首の候補として扱う．この時，指はブロック内の肌色画素の占有比率が低いため，結果として無視される．

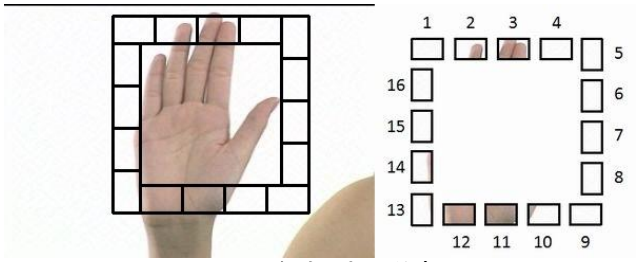


図 3 縁領域の分割

2.1.4 閾値の設定

内側領域と縁領域，縁領域における各ブロックにおける肌色画素の占有比率の適切な閾値を設定するための調査を行った．調査対象には手話の指文字 106 種類（五十音，アルファベット，数字）の画像[3]（図 2，3，4）を使用した．肌色画素比率の閾値 αl と αh,の組 βl と βh の組，γ について，それぞれ 5 パターン設定し，合計 125 パターンの組み合わせについて F 値（適合率，再現率の調和平均）を求めた．

なお，手領域検出の成功条件は(1) 探査矩形の幅が手の甲の幅より大きい，(2) 親指が立っている場合は親指の付け根を含む，(3) 推定された手首位置と検出矩形の中心を結んだ線から手の傾きが分かる，の 3 つを条件とした．この条件を満たさない検出結果の例を図 4 に示す．

† 東京電機大学大学院未来科学研究科
Graduate School of Science and Technology for Future Life, Tokyo Denki University

調査の結果，肌色画素比率は，内側領域が $\alpha l = 70\%$，$\alpha h = 80\%$，縁領域が $\beta l = 30\%$，$\beta h = 40\%$，縁領域における各ブロックが，$\gamma = 90\%$の時，F 値が 0.957 と最も高くなった．しかし，実環境においてはノイズ等の影響を受ける事を考慮し，システムとしては，内側領域と縁領域の閾値を共に 10%下に広げ，$\alpha l = 60\%$，$\alpha l = 20\%$とすることとした．また，手の位置検出においても，少々の位置ずれは許容し，検出結果が手の位置であるものは全て検出成功とする．

図 4　条件を満たさない結果

2.1.5 多段階判定

フレーム内走査中の手領域判定は，内側領域の判定，縁領域の判定，縁領域におけるブロックでの判定の 3 つである．しかし，非手領域に対して全ての判定を同時に行うのは効率が悪い．そこで，各判定段階的に行い，非手領域を早期に破棄することで探査を効率化する．

3. 実験

3.1 実験とその環境

実環境での処理速度の計測と，公開データセットを対象とした検出精度評価を行った．走査矩形の幅は対象領域の短辺長から，20×20 画素までサイズを 0.9 倍ずつ縮小する．なお，フレーム内で一度検出に成功した場合，そのフレーム内での更なる検出は行なわない．また，今回の実験に用いる検出器は片手検出器である．

3.2 処理速度評価

検出対象が存在する場合（図 5 左）と存在しない場合（図 5 右）における，任意の連続した 1000 フレームを対象に処理速度の計測を 3 回施行した．結果の平均処理速度を表 1 に示す．なお，この実験における計測環境は PC(CPU：Intel Core(TM) 2 Duo CPU(1.07GHz)，メモリ：2GB，OS：Windows7)，カメラ(BUFFALO BSW20K04H)である．実験対象の解像度は QVGA(320×240 画素)，フレームレートは 30fps である．

表1　処理速度の評価結果

検出対象	前景抽出	積分画像生成	手領域検出
有り	18.7(msec)	1.87(msec)	1.61(msec)
	22.18(msec)		
無し	17.59(msec)	1.83(msec)	2.08(msec)
	21.5(msec)		

(a) 検出対象が存在する　(b) 検出対象が存在しない

図 5　検出速度評価に用いた画像例

3.3 検出精度評価

データセットには再現性確保のため Spruyt らによって作成された，手の追跡精度評価の為のビデオシーケンス[4]を用いた．これは解像度が 320×240 画素，フレームレートが 25fps である．評価対象は公開されている計 8 種類のデータセットの中から#2,3,6,7,8 を対象とした．これらのデータセットは両手を含むため，今回の実験では探査領域を左 3/5 および右 3/5（128×240 画素）とし，さらに，探査領域内に片側の手のみであるものを選択している．表 2 に結果を示す．表 2 において，検出対象は片側の手のみのであるフレームの総数，正誤はそれぞれ，検出対象に対しての検出成功数と誤検出数の比率である．また，F 値は上記の結果から求めた適合率，再現率の調和平均である．

表 2　認識精度の評価結果

		#2	#3	#6	#7	#8
左	検出対象	1750	1988	2025	2024	1738
	正（%）	0.702	0.705	0.626	0.684	0696
	誤（%）	0.047	0.081	0.268	0.028	0.114
	F 値	0.802	0.789	0.661	0.799	0.769
右	検出対象	1631	1698	1954	1507	1695
	正（%）	0.742	0.596	0.560	0.659	0.771
	誤（%）	0.102	0.269	0.092	0.066	0.087
	F 値	0.805	0.639	0.678	0.764	0.830

4. まとめ

本稿では，複雑背景下での実時間ハンドジェスチャ認識システムのための手領域検出手法を提案した．実験の結果から，本手法は，実時間処理のために十分高速な手法であると言える．また，肌の色に似た背景であっても動きの情報を組み合わせることにより手領域の検出が可能である事から，複雑背景下で利用出来ると言える．

なお，顔や肘等も手の候補領域として誤検出し易い事も判明している．そのため，これらの誤検出は許容するが，後続処理で手であるかどうかの確実な判定を行う必要がある．複数の手領域候補の検出に対応する必要があるのは当然である．また実験により，環境光やカメラの動きによって，背景差分法による動きの抽出の精度が低下し，肌色画素が上手く得られない状況もある事が分かったため，これらの改善を今後の課題とする．

参考文献

[1] 篠崎,他, “実時間物体追跡に適した動的背景推定法と背景差分法”, 知能と情報（日本知能情報ファジィ学会誌）, Vol.24, No.2 (2012).
[2] J.Sherrah,et.al,"Skin Colour Analysis", (2001). http://homepages.inf.ed.ac.uk/rbf/CVonline/LOCAL_COPIES/GONG1/cvOnline-skinColourAnalysis.html
[3] IPA「教育用画像素材集サイト」 http://www2.edu.ipa.go.jp/gz/
[4] V. Spruyt, et.al, "Real-Time Hand Tracking By Invariant Hough Forest Detection", ICIP (2012).