想起背景を用いた移動物体抽出に関する研究

神田 嵩臣* 藤本 健司**

A Study of Object Extraction Using Recalling Background

Takaomi KANDA* Kenji FUJIMOTO**

ABSTRACT

Recently, demand for object extraction technology and object recognition technology has been increasing in many disciplines. To improve accuracy of these technologies, researchers have studied background subtraction method which extract objects. However, many conventional methods have some restrictions, because it's difficult to deal with background transition caused by alteration of ambient light or background objects. Consequently, the camera must be fixed to obtain moving images in most cases. In this paper, we propose a new object extraction method that is able to use a moving camera. This method takes advantage of properties of Neural Networks, ability of learning and recalling images and robustness. After a Neural Network learns background images, if an image including objects is inputted, the Neural Network outputs the recalling background image that corresponds to the input. Then, the recalling image is postprocessed to detect objects. The experimental results showed that the proposing method is able to extract objects under various situations.

Keywords : neural network, background subtraction method, object extraction

1. はじめに

近年，デジタルビデオカメラなどの電子光学機器の発達により，様々な分野で物体抽出技術や物体認識技術の需要が増加している．しかし，物体抽出技術においては，広範囲に渡って動作させようとすると全方位カメラといった特殊なカメラや複数台のカメラが必要となり，高価であったり設置が複雑となったりする問題がある．また，一般物体認識技術に関しては，長年に渡って研究が行われてきたにもかかわらず，人間の顔の認識を除いては，ほとんど実用的な精度に至っていない(1)．したがって，これらの問題を解決しようと現在も多くの研究が行われている．

ここで，物体の領域を精度よく抽出することは，物体抽出技術だけではなく，物体認識技術の精度を向上させるためにも必要である．この物体領域の抽出手法は様々に存在するが，動画像において物体の抽出を行う手法としては，背景差分法が広く知られている．しかし，従来の背景差分法は，木々の揺れといった背景物体の変動や照明の変化などへの対応が難しく，様々な雑音が前景として抽出されてしまう．さらに，ほとんどの場合でカメラは静止していなければならず，使用には様々な制約が伴う．この制約の内，輝度ヒストグラム法を用いた手法(2)や，混合ガウス分布を用いた手法(3)などで，背景の変化に対応できることが知られている．しかし，いずれもカメラは静止していなければならず，加えて，急激な背景の変化には対応が難しい．

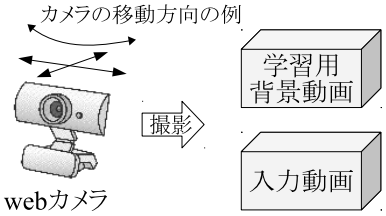
また，近年ではコンピュータの性能の急速な向上により，膨大な計算量が必要となる処理も容易に行えるようになってきているため，複雑系を用いた研究が注目を集めている．その中の1つに，生体の神経回路網を模倣したニューラルネットワーク(Neural Networks: 以下NN)(4)を用いたものがある．これは，人間の脳のシナプスの可塑性を模した学習機能と，外部からの入力に応じて自分自身の構造を変えていく自己組織化の機能を有しており，周囲の状況に合わせた柔軟な認識を行うことができる．

そこで本研究では，NNが有する学習・想起能力とロバスト性を利用することにより，従来の背景差分法のカメラが静止していなければならないという制約を解消できると考えた．したがって，移動している1台のwebカメラより取得した画像に対して物体抽出が行える背景差分法の実現を目的とする．これが実現できれば，広範囲に渡る物体抽出が安価となることに加えて，移動するロボットに搭載したカメラを用いた物体抽出も可能となるため，物体抽出技術や物体認識技術のさらなる発展が期待できる．

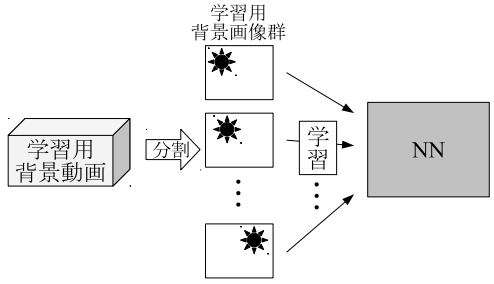
*専攻科 電気電子工学専攻
**電子工学科 准教授

2. 提案する手法

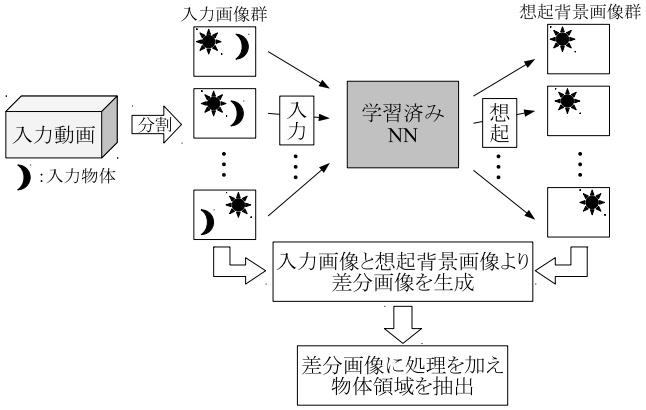
2.1 提案する手法の概要 今回提案する手法の概略図を図1に示す．まず，図1(a)に示すように，webカメラを移動させながら背景として学習させる動画(学習用背景動画)とその背景に物体が入った動画(入力動画)を撮影する．そして，図1(b)に示すように学習用背景動画から任意枚数の画像を切り出し，NNにその画像を1枚ずつ学習させていく．学習が終了した後，図1(c)に示すように入力動画から任意枚数の画像を切り出し，学習済みのNNに入力画像を1枚入力すると，それに対応して想起された背景画像(想起背景画像)が出力される．この入力画像と想起背景画像の差分をとり，処理を加えることにより物体抽出を行う．



(a) 動画の取得



(b) 背景の学習



(c) 物体の抽出

図1 提案する手法の概略図

2.2 NNの学習 図1(b)で示したように，提案する手法はNNの学習が必要となる．今回，NNとしては3層フィードフォーワード型NNを使用し，誤差逆伝搬(Back Propagation)法[4]で学習を行う．

NNへの入力としては，以前の研究[5]ではRGB表色系で表される画像の全画素の3要素(R,G,B)を用いていた．しかし，RGB-YC_bC_r変換式[6]により画像をYC_bC_r表色系に変換し，事前調査を行った結果，多少精度が悪くなるものの入力要素として輝度Yのみを用いても問題ないことがわかった．今回はNNの出力や学習を行う速度を高速化するために，入力として全画素の輝度Yのみを用いることにした．したがって，入力層および出力層のユニット数をn個，画像の高さをL_{height}[画素]，幅をL_{width}[画素]とすると，$n = L_{\mathrm{height}} \cdot L_{\mathrm{width}}$である．また，隠れ層のユニット数は任意の個数とする．

2.3 物体抽出のための画像処理 図1(c)で示したように，物体抽出を行うために差分画像に画像処理を加える必要がある．その手順を図2に示す．まず，差分画像には入力画像や想起背景画像に含まれていたノイズがそのまま引き継がれるため，メディアンフィルタを適用することでノイズの軽減を行う．次に，ある閾値により2値化を行い，物体領域として抽出された反応画素(1:白)とそれ以外(0:黒)に分割する．最後に，必要に応じて物体抽出を抑制する処理を行う．この処理を「面積制限処理」と呼ぶこととする．次の項より，これらの具体的な処理方法を説明していく．

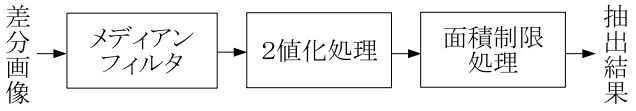


図2 差分画像から物体抽出を行う手順

2.3.1 メディアンフィルタ ここでは，3×3のメディアンフィルタを用いる．これは，注目画素とそれに8隣接している画素の輝度値の中央値を求め，注目画素に中央値を与えるフィルタである．

2.3.2 2値化処理 今回，2値化処理を2種類用意した．それらの処理手順を以下に示す．

(1) 1つの閾値を用いた2値化
  (i) 閾値tで2値化
(2) 2つの閾値を用いた2値化
  (i) 閾値tで2値化
  (ii) 反応画素(1:白)を起点として8隣接している画素のみ拡大閾値t_{ex}で2値化
  (iii) 反応画素が増加しなくなるまで(ii)を繰り返す

(1)の閾値の決定には，以前の研究[5]と同様に柔軟な閾値決定を行える判別分析法を用いる．ここで，0〜255の間の閾値決定では物体がない場合でも背景領域を物体として抽出してしまうため，閾値決定範囲の下限を差分画像の平均輝度値$D_{\mathrm{ave}}+30$とする．(2)の2つの閾値の決定には，平均輝度値D_{ave}を変数として閾値tを求める関数，およびその閾値tを変数として拡大閾値t_{ex}を求める関数を統計的に作成し，使用する．

2.3.3 面積制限処理 8隣接している反応画素数を面積としたとき，それが10画素以下ならば，その部分の画素の値を0にする処理を行う．ノイズが抽出された場合は小さな面積となることが多いと考えられるため，これによりノイズの抽出の軽減を図る．

3. 評価方法

3.1 評価用画像の作成 物体抽出の評価を行うにあたって，入力画像のどの部分が物体領域であるかをあらかじめ知っておく必要がある．今回はその領域を目視で確認し，手動で設定する．図3に評価用画像の作成例を示す．まず，図3(a)の学習用背景画像と図3(b)の入力画像を目視で確認し，物体の輪郭に沿って点を打つ．この際に，連続して打った2点間を直線で補間する．そして，輪郭線で領域を分割した後に物体領域を選択することで，図3(c)に示すような物体領域(白)と背景領域(黒)に分けられる．ここで，輪郭線付近の物体の有無については間違いが多くなってしまうため，輪郭線とその4隣接画素を無効領域(灰色)とした．

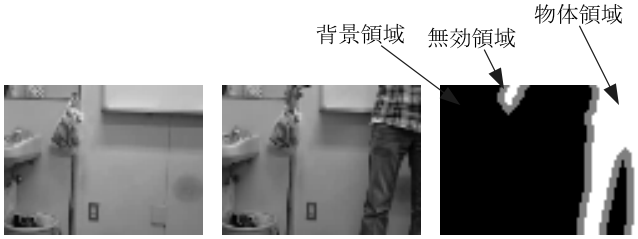


図3 評価用画像の作成例

3.2 物体抽出の評価 図4は，差分画像を画像処理することにより生成された抽出結果と，その入力画像に対応する評価用画像を用いて評価を行った例である．この例における評価用画像は概念的なものであり，実際に6×6画素の画像では，このような無効領域の評価用画像は作成できない．抽出結果の反応画素は，評価用画像のどの領域に属するかによって「正反応」,「無効反応」,「誤反応」に分けられる．

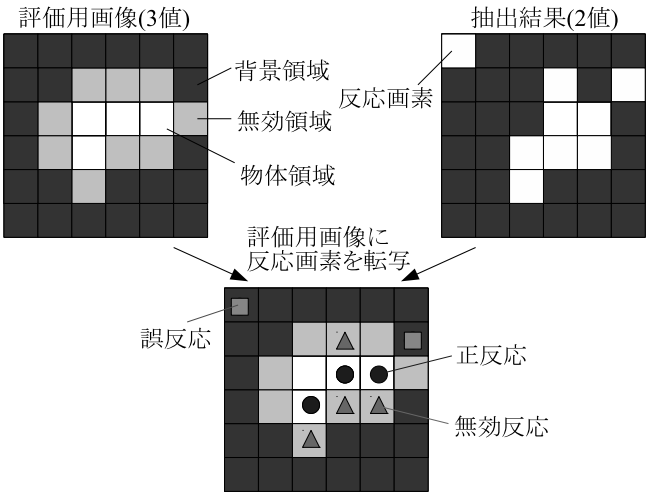


図4 抽出結果と評価用画像を用いた評価例

目的とする処理によっては，誤反応がある程度多くとも正反応が多ければよい場合と，その逆の場合が存在する．しかし，今回は正反応1つと誤反応1つを同等の価値とみなし，次のような評価式で評価値を求めることにする．

$$\text{評価値[画素]} = \text{正反応数[画素]} - \text{誤反応数[画素]} \quad (1)$$

また，正反応率と誤反応率は以下のとおりとする．

$$\text{正反応率[\%]} = \frac{\text{正反応数[画素]}}{\text{物体領域面積[画素]}} \times 100 \quad (2)$$

$$\text{誤反応率[\%]} = \frac{\text{誤反応数[画素]}}{\text{背景領域面積[画素]}} \times 100 \quad (3)$$

ここで，評価値は物体抽出の精度を相対的に評価する指標となる．一方，正反応率はどの程度物体を抽出できたか，誤反応率はどの程度背景を抽出してしまったかを表し，絶対的な評価の指標となる．

4. 実験方法

4.1 事前準備 まず初めに，図1(a)にも示したように，学習用背景動画と入力動画を用意する必要がある．これらは，簡易なスイング機構に取り付けたwebカメラ(Logicool HD Webcam C270)により，320×240画素で180秒間撮影し，取得した．今回は，カメラの動きはスイングのみとした．その後，それらの動画を80×60画素に圧縮し，それぞれ5000枚の学習用背景画像と300枚の入力画像に分割した．さらに，その入力画像に対応する評価用画像も300枚作成した．

以上の手順を異なる場所で5セット繰り返し，それぞれを“S1”，“S2”，“S3”，“S4”，“S5”と名付けた．最後に，各セットの学習用背景画像をそれぞれNNに同程度学習させた．このとき，中間層のユニット数は1000個とした．

ここで，学習用背景画像の例としてS1の画像の一部を図5に示す．また，各セットにおいて300枚を総計した評価用画像の各領域面積を表1に示す．

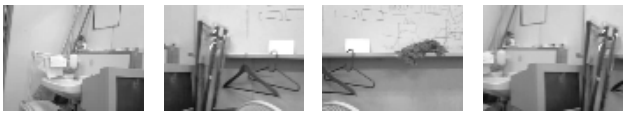


図5 S1の学習用背景画像の一部

表1 評価用画像の各領域面積

	物体領域面積[画素]	背景領域面積[画素]
S1	61935	1355531
S2	76412	1326989
S3	82141	1316062
S4	87431	1327157
S5	58417	1355923

4.2　1つの閾値を用いた場合

4.2.1　1つの閾値で得られる最大の評価値の確認

最初に，画像1枚ごとに評価値が最大となる閾値を決定したとき，その最大の評価値がどの程度の値になるかを確認した．なぜなら，この値が1つの閾値を用いたときの目標値であり，以降に示す結果の比較対象となるからである．評価値が最大となる閾値の決定は，評価用画像を利用することにより実現した．ここで，面積制限処理は適用しないものとした．

4.2.2　判別分析法の適用　ここでは，判別分析法により閾値を決定することで，4.2.1項で求めた最大の評価値と，どの程度の差が見られるかを確認することが目的である．したがって，メディアンフィルタを適用した差分画像に対して，判別分析法を用いて閾値を決定し，2値化を行ったとき，各セットでどのような評価が得られるかを確認した．ここでもまた，面積制限処理は適用しないものとした．

4.2.3　面積制限処理の適用　次に，4.2.2項と同様に判別分析法による2値化を行った後，面積制限処理を適用することによって，評価がどのように変化するかを確認した．

4.3　2つの閾値を用いた場合

4.3.1　2つの閾値で得られる最大の評価値の確認

4.2.1項と同様に，評価用画像を利用することで，画像1枚ごとに評価値が最大となる閾値と拡大閾値の組み合わせを決定し，最大の評価値を求めた．

4.3.2　関数による閾値決定の適用　まず，閾値tと拡大閾値t_{ex}を求める関数を作成する必要がある．これには，S1，S2，S3で作成された差分画像の平均輝度値D_{ave}と，4.3.1項で求めた閾値tと拡大閾値t_{ex}を使用した．そして，D_{ave}-t平面とt-t_{ex}平面にこれら3セット900点をプロットし，最小二乗法により3次多項式のフィッティング関数を求めた．これらのフィッティング関数を使用して，各セットにおいて2つの閾値を決定し，2値化を行い，どのような評価になるかを確認した．

4.3.3　面積制限処理の適用　最後に，4.3.2項と同様にフィッティング関数による2値化を行った後，面積制限処理を適用することによって，評価がどのように変化するかを確認した．

5.　実験結果

5.1　1つの閾値を用いた場合

5.1.1　1つの閾値で得られる最大の評価値の確認結果

評価用画像を用いて評価値が最大となる閾値を1枚ごとに設定した場合の1セットあたり300枚を総計した評価結果は表2のとおりである．これより，理想的な閾値を選択した場合，評価値は19000～42000画素程度であり，それに伴う正反応率は約39～62%，誤反応率は約0.5～1%となった．したがって，単純な処理のみでも良好な精度が得られる可能性があることがわかった．さらに，学習条件の改善や得られた反応画素を用いた補間などを行うことによって，この精度を向上させることができると考えられる．

表2　1つの閾値による最大の評価値とそれに伴う正反応率・誤反応率

	評価値[画素]	正反応率[%]	誤反応率[%]
S1	27363	54.24	0.46
S2	24432	46.09	0.81
S3	18840	39.15	1.01
S4	41565	61.58	0.93
S5	23017	62.12	0.98

5.1.2　判別分析法の適用　次に，判別分析法により閾値を決定した場合の評価結果を表3に示す．これより，評価値は8800～19000画素程度であり，正反応率は約17～31%，誤反応率は約0.3～0.4%となった．この結果を表2と比較すると，判別分析法による2値化では最大の評価値に対して約39～63%の評価値を得られていることがわかった．また，正反応率と誤反応率はともに減少している．

表3　判別分析法による2値化の評価

	評価値[画素]	正反応率[%]	誤反応率[%]
S1	13492	31.01	0.42
S2	15353	27.63	0.43
S3	8765	17.42	0.42
S4	18852	26.60	0.33
S5	9069	25.17	0.42

5.1.3　面積制限処理の適用結果　判別分析法によって2値化を行った後，面積制限処理を適用した結果を表4に示す．面積制限処理を適用していない表3に比べて，評価値はS4を除いて各セットで2000画素程度増加している．S4においても約400画素の減少のみであり，悪化はわずかである．このことから，面積制限処理は有効であることがわかった．しかし，面積制限処理によって，10画素以下の面積の物体が抽出不可能になることに注意する必要がある．また，正反応率と誤反応率は，当然のことながらどちらも減少した．

表4　判別分析法による2値化後の面積制限処理の評価

	評価値[画素]	正反応率[%]	誤反応率[%]
S1	15601	29.73	0.21
S2	17188	25.91	0.20
S3	10887	15.44	0.14
S4	18421	24.44	0.22
S5	11339	23.04	0.16

5.2 2つの閾値を用いた場合

5.2.1 2つの閾値で得られる最大の評価値の確認結果

ここでは，最大の評価値が得られる閾値と拡大閾値の組み合わせを1枚ごとに設定し，1セットあたり300枚を総計した評価を表5に示す．この結果より，理想的な2つの閾値を選択した場合，評価値は32000〜57000画素程度であり，それに伴う正反応率は約54〜74%，誤反応率は約0.2〜0.9%となることがわかった．これより，1つの閾値を用いた場合よりも2つの閾値を用いたほうが，より高い可能性を有することが確認できた．

表5 2つの閾値による最大の評価値とそれに伴う正反応率・誤反応率

	評価値[画素]	正反応率[%]	誤反応率[%]
S1	40451	69.71	0.20
S2	36782	58.14	0.58
S3	32033	53.71	0.92
S4	56759	74.27	0.62
S5	32714	67.41	0.49

5.2.2 関数による閾値決定の適用結果 まず始めに，S1，S2，S3に対して，評価用画像を用いて評価値が最大となる2つの閾値を1枚ごとに設定した．そして，このときの差分画像の平均輝度値D_{ave}と閾値tと拡大閾値t_{ex}の関係を確認した．この平均輝度値D_{ave}と閾値tの関係を図6に，閾値tと拡大閾値t_{ex}の関係を図7に示す．また，それぞれ図6，図7中に示したフィッティング関数は3次多項式であり，最小二乗法によってフィッティングを行った．図6を見ると，フィッティング関数近傍に大半の点がプロットされているが，D_{ave}が35〜40付近でtが0付近といった特異点が数点見られる．これは入力画像のほとんどが物体領域であるときに起こりうる．また，図7を見ると$t = t_{\mathrm{ex}}$の位置に多くプロットされている．さらに，その数はtが小さいほど多い傾向にある．これは入力画像のすべてが背景領域の場合，tが小さくなり，かつ物体領域を増加させるt_{ex}が不要になるためである．

次に，フィッティング関数を用いて2つの閾値を決定した．これはD_{ave}が評価用画像を用いることなく得られるため，D_{ave}からtを求め，そのtよりt_{ex}を求めることで決定できる．この評価結果を表6に示す．表6の結果より，フィッティング関数により2つの閾値を決定した場合は，評価値は13000〜22000画素程度であり，正反応率は約23〜42%，誤反応率は約0.2〜0.6%となることがわかった．この評価値を表5に示した最大の評価値と比較すると，約39〜55%の値が得られており，その効率は判別分析法と比べても遜色がないことがわかった．また，判別分析法で得られた評価値(表3)と比較すると，いずれも判別分析法より高い評価値が得られていることがわかる．これより，S1，S2，S3を基に

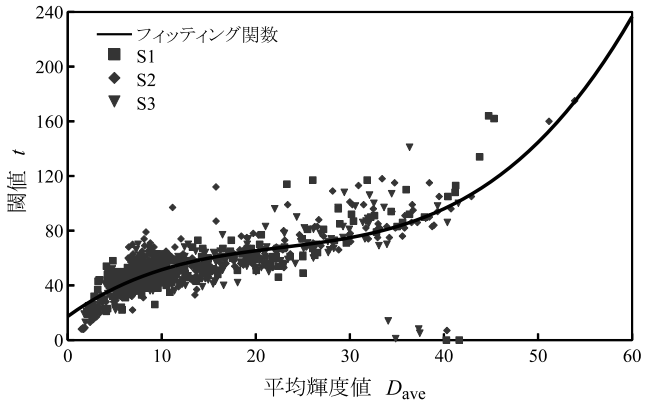


図6 平均輝度値D_{ave}と閾値tの関係

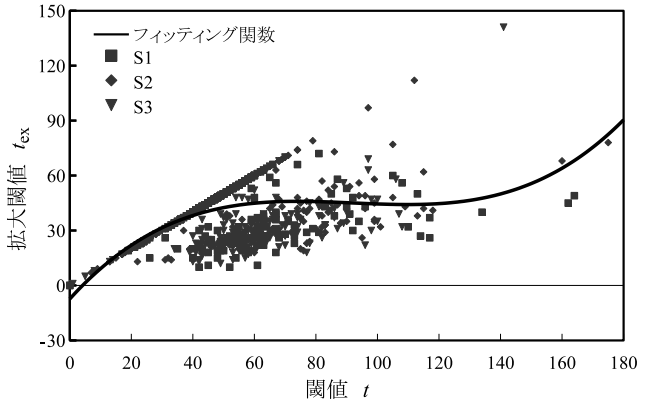


図7 閾値tと拡大閾値t_{ex}の関係

表6 フィッティング関数による2値化の評価

	評価値[画素]	正反応率[%]	誤反応率[%]
S1	20004	42.35	0.46
S2	20320	37.13	0.61
S3	13090	22.80	0.43
S4	22299	29.28	0.25
S5	16376	33.34	0.23

作成した関数であるが，S4やS5にも適用できることが確認できた．ただし，今回は用意できたセット数が少なかったため，さらにセット数を増やして検証していかなければならない．また，今回の手法ではtの理想値と計算値のずれによって，t_{ex}の基となる反応画素がなくなってしまった可能性がある．この問題を解決することによって精度が向上すると考えている．

5.2.3 面積制限処理の適用結果 5.2.2項と同様にフィッティング関数によって2値化を行った後，面積制限処理を適用した結果を表7に示す．この結果を表6に示した面積制限処理なしの結果と比較すると，すべてのセットの評価値が平均500画素ほど増加していることがわかった．これより，5.1.3項と同様に面積制限処理が有効であることが確認できた．また，正反応率や誤反応率を比較すると，その変化は判別分析法の後に面

表7 関数による2値化後の面積制限処理の評価

	評価値[画素]	正反応率[%]	誤反応率[%]
S1	20684	42.21	0.40
S2	20896	37.01	0.56
S3	13519	22.53	0.38
S4	22551	29.10	0.22
S5	16890	33.12	0.18

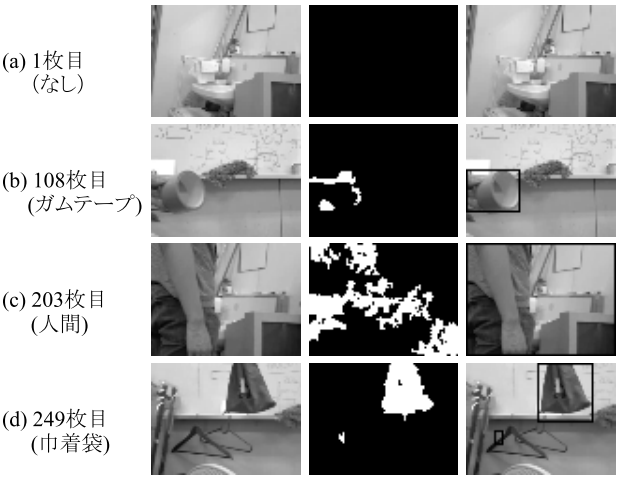


図8 S1の入力画像の一部とそれに対する抽出結果と可視化画像(枚数の下の括弧内は入力物体を表す)

積制限処理を適用したときよりも幾分小さい．これは，拡大閾値が面積を増加させるためだと考えられる．

最後に，この処理を行った各セットのうち，S1の入力画像の一部とそれに対する抽出結果，可視化画像を図8に示す．可視化画像は，抽出結果において8隣接している反応画素を長方形状に線で囲ったものを入力画像上に表示している．ここで，長方形が重なる場合にはそれらを統合した．図8を見ると，図8(a)は物体がなく誤反応もない，よい結果が得られた．また，図8(b)は物体を抽出できているものの，欠損が多い結果となった．図8(c)では物体の一部しか抽出できておらず，誤反応も多くなってしまった．この可視化画像は，抽出結果の反応画像が画像全体に広がっているため，画像の一番外側に線が引かれている．誤反応が多くなった原因としては，物体が大きいために背景の想起が不安定になったことが考えられる．図8(d)は物体をよく抽出できているが，少し誤反応もみられた．

6. まとめ

今回，スイングカメラを用いた物体抽出法を構築し，実空間を撮影した動画を用いて実験，評価を行った．2値化の手法としては2種類用意したが，1つの閾値を用いるよりも2つの閾値を用いたほうが，よりよい精度の物体抽出を実現する可能性があることがわかった．ここで，2つの閾値の決定方法として提案したフィッティング関数を用いた方法は，最大の評価値に対して39〜55%程度の評価値が得られることが確認できた．また，正反応率は約23〜42%，誤反応率は約0.2〜0.6%となり，簡単な処理で精度がよい物体抽出を実現できた．さらに，面積制限処理も評価値の向上には有効であることがわかった．

しかし，よりよい精度を実現するためには，現状で問題点が3つ存在する．1つ目は，物体抽出に欠損がでてしまうことである．これは，背景想起によるノイズも原因の1つとしては考えられるが，主な原因は背景と物体の輝度値の差が小さいときに抽出ができなくなることである．この解決のためには，輝度値以外のほかの要素を用意する必要がある．2つ目は，物体が大きくなると背景想起が不安定になることである．また，3つ目はNNの学習誤差が出力層のユニットによってばらつくことで，誤反応が発生しやすくなることである．これらに対しては，NNの構造や学習方法を変更していかなければならない．

今後は，これらの問題点の解消に加えて，より柔軟に環境に適応する手法の構築を目指す．このためには，並進しているカメラに対する実験も行う必要がある．また，環境光が変化する場合への対応も行わなければならない．これらの環境への適応は，従来の背景差分では困難とされている．しかし，我々が提案した手法では，以前の研究[5]においてこれらに対応できる可能性が示唆されており，柔軟な対応が期待できる．

参考文献

(1) 柳井啓司：「一般物体認識の現状と今後」, 情報処理学会論文誌, Vol. 48, No. SIG 16(CVIM 19), pp. 1-24, 2007.

(2) 早坂光晴, 富永英義：「動画像からの背景画像生成を用いた移動物体抽出法に関する一検討」, 情報処理学会研究報告 オーディオビジュアル複合情報処理, AVM29-1, pp. 1-6, 2000.

(3) 島田敬士, 有田大作, 谷口倫一郎：「混合ガウス分布による動的背景モデルの分布数増減法」, 画像の認識・理解シンポジウム(MIRU2006), pp. 746-751, 2006.

(4) 吉冨康成：「ニューラルネットワーク」, 朝倉書店, 2002.

(5) 藤本健司, 神田嵩臣：「ニューラルネットワークによる背景想起を利用した物体抽出システムの開発」, 神戸高専研究紀要, 第50号, pp. 69-74, 2012.

(6) Radiocommunication Sector of ITU：“Studio encoding parameters of digital television for standard 4：3 and wide-screen 16：9 aspect ratios”, Recommendation ITU-R BT. 601-7, pp. 2-3, 2011.