

平成23年6月10日現在

機関番号：32657

研究種目：研究活動スタート支援

研究期間：2009～2010

課題番号：21800057

研究課題名（和文）視覚物体認識におけるクラスの創発，分割統合，階層化

研究課題名（英文）Finding of Object Class in Visual Object Recognition

研究代表者

日高 章理（HIDAKA AKINORI）

東京電機大学・理工学部・助教

研究者番号：70553519

研究成果の概要（和文）：

パターン認識に基づく視覚的物体認識では，認識器が取り扱うことのできるクラスがあらかじめユーザーが定めた特定のものに限られる，認識器を構築するために教師情報を付与した大量の学習用データが必要になる，などの問題がある．本研究ではこれらの問題への対処を目的とし，教師情報が付与されていない映像データから半自動的に教師情報付きデータを生成するためのスキームや，それらのデータを非教師なし学習によって分類する手法に関する研究を行った．

研究成果の概要（英文）：

In visual object recognition based on pattern recognition, unsupervised learning is a still important problem. In this research, we try semi-automatic generating of training samples from non-labeled video sequence by using unsupervised approach. We use grow cut and template matching based object tracking method for our sample extraction. In our experiments, we show that our approach can make the clusters of targets without teacher signals.

交付決定額

（金額単位：円）

	直接経費	間接経費	合計
2009年度	1,080,000	324,000	1,404,000
2010年度	980,000	294,000	1,274,000
年度			
年度			
年度			
総計	2,060,000	618,000	2,678,000

研究分野：工学

科研費の分科・細目：情報学・知能情報学

キーワード：パターン認識，機械学習，画像認識，クラスタリング

## 1. 研究開始当初の背景

パターン認識とは適切に設計された特徴抽出画像と識別関数を用いて入力パターンを出力クラスに正しく対応させることを目的とする計算手法のことである．特徴画像と

識別関数には通常多くのパラメータが含まれるため，学習用データのもとでそれらをあらかじめ最適化する必要がある．これは機械学習の手法を用いて行われる．

近年，識別関数については SVM など、また

画像認識の特徴抽出法については SIFT 特徴や bag of visual words 法などの優れた手法が提案されてきた。その結果、一昔前と比べれば著しく高速・高精度な画像認識器を構築することが可能になっており、最近ではカメラの顔検出機能や大規模画像データベース検索システムなどの様に、日常生活の一端でパターン認識技術を見かける機会も増えている。

一方、既存のパターン認識技術には次のような問題がある。

(1) 認識器はあらかじめ定められた特定のクラスしか取り扱えない。画像を顔と顔以外の2クラスに分類する認識器では顔表情の分類は行えないし、もちろん文字や動物の分類なども行えない。さらに、例えば学習用データが正面向きの顔のみだった場合、横向きや下向きの顔画像を「顔」と認識することも困難である。

(2) そのような対象の変動に対応するには、あらかじめ各向きごとに大量の学習用データを用意し、それを元に学習を行う必要がある。大量の学習用データの取得とラベル付けはパターン認識研究において最も労力がかかる部分であり、これは実際上大きな問題点である。

これらの制約は、画像認識の実用化を限られた分野に留めている大きな原因の一つであると考えられる。

## 2. 研究の目的

先述の通り、従来のパターン認識に基づく視覚的物体認識に関する研究では、認識器が取り扱うことのできるクラスがあらかじめユーザーが定めた特定のものに限られるという問題や、あらかじめ教師情報が付与された大量のサンプルデータを学習のために用意する必要があるという問題があった。

本研究ではこれらの問題への対処を試み、以下の目的で研究を行った。

(1) 教師情報が付与されていない動画データから、半自動的に教師情報付きデータを生成するためのスキームを開発する。

(2) そのようにして得たデータを主成分分析や階層的クラスタリング手法によって分析し、それらが実際に学習用データとして利用可能かどうかを検討する。

## 3. 研究の方法

(1) 本研究では、動画画像中から任意の時空間部分構造を自動的に取り出すため、Grow

Cut に基づく前景・背景のセグメンテーションと、テンプレートマッチングによる移動追跡を組み合わせた対象追跡を用いた。これらはいずれも事前の学習を必要とせず、与えられた初期情報のみに基づいて対象の切り出しや追跡を行うことが可能である。

これにより、無作為に選んだ動画中で任意の対象に追跡処理を行い、対象が存在する領域を連続的に切り出した。セグメンテーションの際には何らかの方法で大まかに前景・背景の初期情報を与える必要があるが、今回は切り出す対象として人間や動物を指定するため、それらが存在する領域を大まかに手動で与えた。

対象の追跡時間は1トラックあたり30フレームとし、これにより、実世界中で発生する様々な変動を豊富に含んだ1秒間分の対象を捉えた学習用サンプルが切り出される。このようなサンプルを学習に用いることで、様々な変動にロバストな認識器を自然に構築できるものと期待される。

例として、今回は主な対象を人（歩行者）と猫として動画から学習用サンプルの切り出しを行った。最初に先述の方法で1秒間分の学習サンプルを切り出した後は、それらを用いて最近傍決定法やテンプレートマッチングなどを行うことで、取得済みのサンプルと類似の領域を動画中から大まかに探すことが可能となる。また、取得済みのサンプルは、同様のサンプルを切り出す際のセグメンテーションにおける初期情報を自動的に与えるためにも利用可能である。

サンプルの切り出しを行った後は、それらの画像を特徴変換する必要がある。本研究で切り出される学習サンプルは、回転、剛体変形、非剛体変形、照明変動、部分遮蔽など、あらゆる変動を含みうる。そのため対象の局所構造に注目する画像特徴などは避け、対象の向きや姿勢などの変動を吸収しやすい特徴を用いることが望ましい。本研究では512次元の Bag of Visual Words 特徴と、輝度勾配ヒストグラム (HOG) 特徴を結合した特徴を用いた。

(2) (1)で得られた特徴ベクトルの集合に対して主成分分析と階層的クラスタリング手法を適用し、特徴空間におけるクラスタ構造を可視化した。

## 4. 研究成果

図1~8はGrow Cut法とテンプレートマッチングを用いた追跡により、動画中から自動的に切り出されたサンプルで、それぞれ30フレーム(1秒)の途中を等間隔で間引いた8フレーム分である。図1, 3, 5, 7は追跡結果の検出窓内の画像であり、図2, 4, 6, 8は追跡時のセグメンテーション結果である。

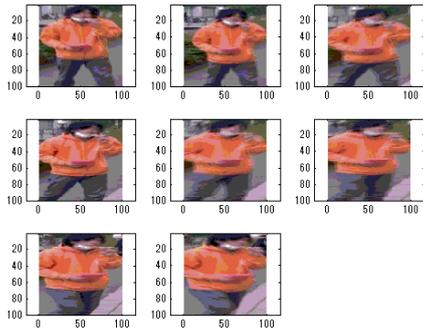


図 1 追跡結果 (歩行者 1)

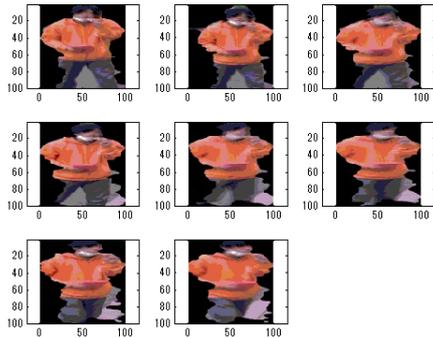


図 2 セグメンテーション結果

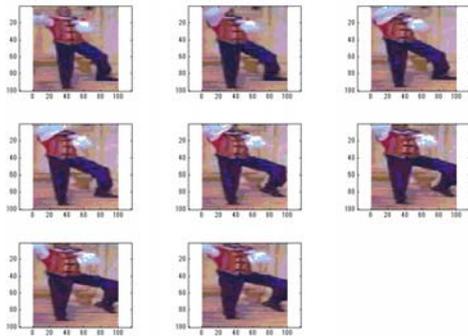


図 3 追跡結果 (歩行者)

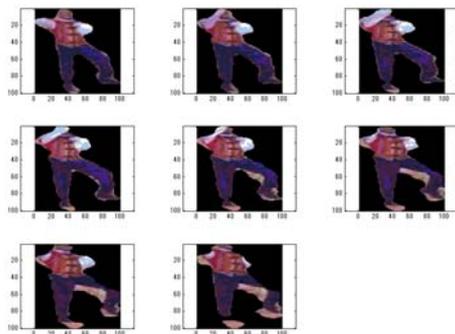


図 4 セグメンテーション結果

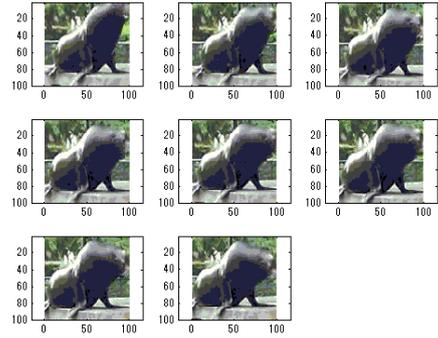


図 5 追跡結果 (オットセイ)

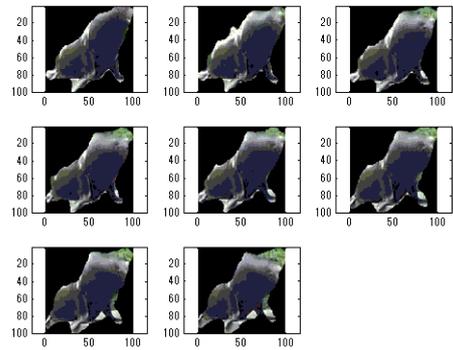


図 6 セグメンテーション結果

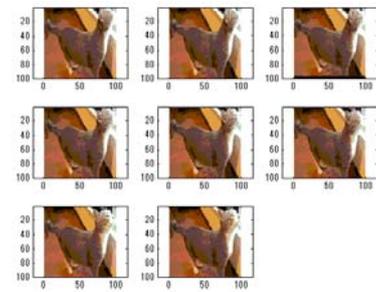


図 7 追跡結果 (猫)

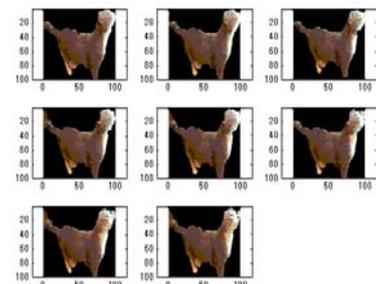


図 8 セグメンテーション結果

図 1~8 のサンプルを含め、オートセイ 3 サンプル (90 フレーム), 猫 6 サンプル (180 フレーム), 歩行者 10 サンプル (300 フレーム) の合計 19 サンプル (570 フレーム) の切り出し画像を前節で述べた特徴量に変換し, 主成分分析して可視化したものが, 図 9, 10 である.

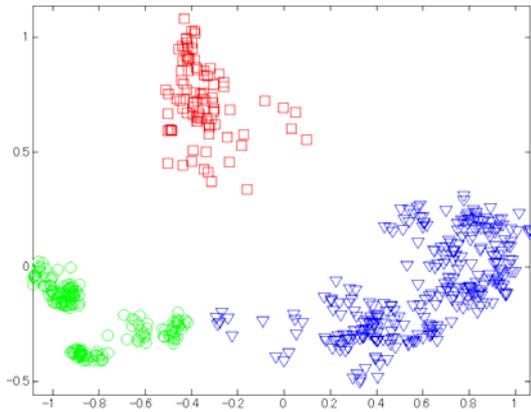


図 9 切り出し画像の主成分空間

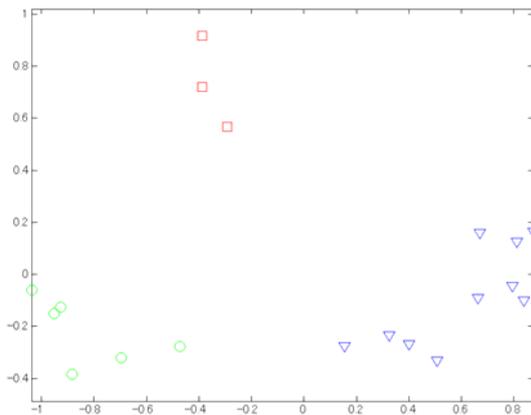


図 10 切り出し画像の主成分空間  
(各サンプルの重心のみ表示)

図 9, 10 では□記号がオートセイ, ○記号が猫, ▽記号が歩行者のサンプルを表す. 図 9 は全 570 フレームを表示し, 図 10 は各サンプルごとに 30 フレーム分の重心ベクトルのみを表示したものである. 図 9, 10 いずれの空間においても各クラスのサンプルが線形分離可能な分布になっており, 特に図 10 では各クラスのサンプルが能く離れて分布していることがわかる.

図 11 は 570 フレーム分の特徴ベクトルに対して ward 法で階層的クラスタリングを行った結果である. この結果では 2 段目の分岐の時点で全サンプルが 4 つの大きなクラスタに分割されるが, このうち左から 1 番目のクラスタにオートセイ, 2 番目のクラスタに猫,

3 番目と 4 番目のクラスタに歩行者の全サンプルがそれぞれ所属しており, 各クラスタではサンプルの混合が 0 となるという, よいクラスタリング結果を得た.

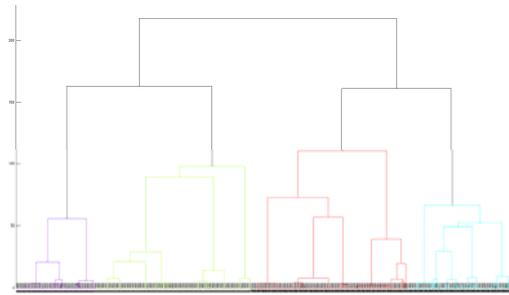


図 11 ward 法による分類結果

これより, 各クラスタに所属するサンプル同士を同一の「クラス」に属するものと考え, すなわちクラスタの番号をクラスラベルと見なすことで, 教師情報の付与されていない動画から切り出した画像セットに対して半自動的に教師情報を付与することが可能であることが示された.

以上から, セグメンテーションと対象追跡を行って半自動的に切り出した画像セットを学習に用いることの妥当性が示されたといえる. 本手法ではセグメンテーションの初期情報をユーザが指定する必要があり, 完全な自動化には至っていない. しかし, 乳幼児が親の視線や指さしによって外界への“興味”を促されることで効率的に知的能力を獲得していくように, 本研究のような問題でもユーザの利便に応じてコンピュータに簡単な指示を出し, その内容に応じて学習の優先度を変えていくことは, 十分合理的であると考えられる.

## 5. 主な発表論文等

(研究代表者、研究分担者及び連携研究者には下線)

〔雑誌論文〕 なし

〔学会発表〕 なし

## 6. 研究組織

### (1) 研究代表者

日高 章理 (HIDAKA AKINORI)  
東京電機大学・理工学部・助教  
研究者番号: 70553519