

テクニカルノート

多くの画像が共有する 「一般クラス」に着目した訓練画像の選択

加藤 ジェーン^{1,a)} 白須 遼^{1,†1} 王 彧^{1,b)} 間瀬 健二^{1,c)}

受付日 2013年8月13日, 採録日 2013年9月13日

概要: 写真データ (フォトアルバム) を効率的に整理・管理するために, 画像の内容に基づく高度な分類技術が求められている. 本論文では, 分類器の性能を高く保ちつつ, ユーザのラベリング作業量を削減することにより, 写真データの分類を効率化する手法を提案する. 具体的には, 分類対象となる写真データから, 分類器の学習において効果が高いと見込まれる少数の訓練サンプルを自動選択する. “skin”, “water” といった, 多くのフォトアルバムで共有される普遍的な概念である “一般クラス” に着目し, 一般クラスに関する知識を普遍的な画像から獲得したうえで, 分類対象となるローカル画像に適用することにより, 多彩かつ特徴的な訓練サンプルを探索してユーザに提示することを可能としている. 21,424 枚の写真からなる保育園のフォトアルバムを用いた検証実験では, 一般クラスを考慮せずに単純なクラスタリングにより分類器を学習した場合や, 訓練サンプルをランダムに選択して分類器を学習した場合と比較して, 分類精度が同等であれば, 学習に必要なサンプル数 (ラベリング量) を 2 分の 1 ないし 3 分の 1 に削減することに成功した.

キーワード: 一般クラス, 訓練画像の選択, マルチクラスタリング, 個人写真の管理

Selecting Training Images Using Underlying Common Classes

JIEN KATO^{1,a)} RYO SHIRASU^{1,†1} YU WANG^{1,b)} KENJI MASE^{1,c)}

Received: August 13, 2013, Accepted: September 13, 2013

Abstract: In this paper, we proposed a novel method to effectively select discriminative training samples for local photo classification or management. We introduced a concept widely shared by most of photo albums, underlying common classes, and based on this concept we catch more colorful and more characteristic/discriminative training samples. We first learn valuable feature dimensions and adaptive sampling methods from images of each single common class, and then apply acquired information/knowledge to target local photos so that discriminative training samples can be found out. We have evaluated our proposed method on 21,424 photos taken in daily life of a nursery school to solve two classification problems. Experimental results show that our method is superior to not only randomly sample selecting method but also traditional sample selecting method without introducing common classes. Since underlying common class we depend on is a general concept, a wide range of application of proposed method can be expected.

Keywords: underlying common class, training image selection, multi-clustering, personal photo management

¹ 名古屋大学大学院情報科学研究科
Graduate School of Information Science, Nagoya University,
Nagoya, Aichi 464-8603, Japan

^{†1} 現在, オークマ株式会社
Presently with OKUMA Corporation

a) jien@is.nagoya-u.ac.jp

b) ywang@nagoya-u.jp

c) mase@nagoya-u.jp

1. はじめに

近年, 家庭用デジタルカメラの普及およびハードディスクをはじめとする記録媒体の大容量化・低価格化により, 個人などが所有する写真データの量は日々増加している. そのため, 写真データを効率的に整理・管理する技術が求

められている。

比較的簡易な写真整理のアプローチは、写真データ自体に含まれる撮影日時や場所などの情報に基づいて写真データをソート・分別するというものである。多くのデジタルカメラ製造メーカからソフトウェアが提供されていることもあり、現在も頻繁に用いられている。しかし、このアプローチでは、写真の視覚的特徴が示す意味内容は、まったく考慮されない。写真整理の場面においては、写真の意味情報も多くユーザの関心事であるため、写真の意味情報を反映した写真整理手法が必要不可欠である。

そこで近年は、画像内容の類似度比較に基づいて、画像に対して自動ラベル付けをするアプローチが開発されてきた。多くの研究開発事例では、ラベル付けの対象を特定の物体に限定したうえで、物体の検出および認識アプローチ(黄瀬 [1] など)をとる。なかでも、顔認識を利用した写真整理手法は現在、実用化が進んでいる (Google Picasa [2], NTT Docomo 社のフォトコレクション [3] など)。これらの事例では、画像中の顔領域検出結果と、ユーザによる個人特定用タグの入力を用いて、写真データの管理を効率化することが可能である。ラベリング済みかつ最も視覚的に類似した顔領域のタグを推薦することにより、新たな顔領域へのラベリングを支援する。また、顔以外の物体の検出に基づく写真整理手法の例としては、写真整理に画像認識を導入するためのプラットフォームである kooba [4] があげられる。kooba は新聞や雑誌などの印刷物の記載、ワインのビンや CD に貼付された文字ラベルなどを認識することで写真整理を促進させることができる。

しかし、これらの事例では、認識の対象範囲が特定の物体に限定されている。対象範囲を広げるためには、実世界における物体のアピアランスの膨大な変化に対応する必要がある、これはきわめて困難な課題である [5]。上記の kooba においても、アピアランスの変化が大きい物体については低い認識率にとどまるか、まったく認識できないケースが存在することが報告されている。また、実際的な場面において、ユーザが写真に付加しようとするラベルは、内容が曖昧かつ抽象的であり、また、特定の物体への関心にとどまらず、写真全体のシーンを反映するなど複雑な面も持ち合わせることが多い。したがって、画像内容の単純な類似度比較に基づく自動タグ付けでは、ユーザの要望を反映したタグを提供することが困難である。

そこで、ユーザが自ら事前に定義したクラス、およびそれらに対応するラベルを付された少量の訓練データセットを用いて分類器の学習を行い、写真のクラス分類を行うことが有効な解決策として考えられる。しかし、このアプローチを採用する場面においても、以下2つの相反する要望に応えることが求められている。

(1) 精度良い分類のためには、できるだけ多くのラベリングを行い、豊富な訓練サンプルを準備して分類器の学

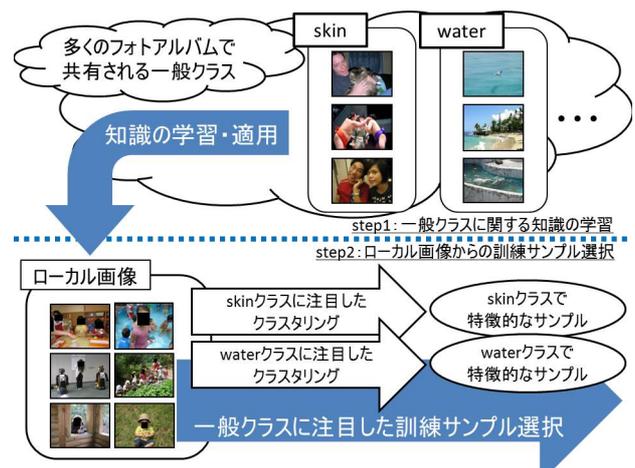


図 1 提案手法概略

Fig. 1 Outline of our proposed method.

習をすることが望ましい。

(2) ラベリングは手作業で行われるため、多くの人的および時間的コストを要する。ユーザの労力削減のためには、できるだけラベリング量を少なくしたい。

このジレンマを解決するために、本研究では、個人のフォトアルバムといったローカル画像 (ユーザの手元にある) を対象として、ラベルなしサンプルの中から、学習に効果的なサンプルを自動選択することで、少数の訓練サンプルを自動抽出する手法を提案する。

一般に、分類器の性能は、学習の際の訓練サンプルの数だけでなく、その良し悪しにより大きく左右される。訓練サンプルを単純に少数にした場合、過学習により分類器の汎化性能が著しく損なわれる可能性がある。したがって、高精度の分類結果を得るためには、多様で豊富な特徴パターンを含む訓練サンプルを選び出し、これらによって訓練データセットを構築する必要がある。また、類似の特徴を持つサンプルを重複して訓練データセットへ加えることを避けなければならない。本提案手法を用いれば、分類器の学習効果が少ないサンプルへのラベリングを回避でき、良質な訓練サンプルへのラベリングをユーザに提案することが可能となる。低労力で比較的良質な分類器の学習がなされることで、結果として効率的な写真整理を実現できる。

本研究がとるアプローチの概略を図 1 に示す。本研究において提案する訓練サンプル選択手法では、多くのフォトアルバムが共有していると期待される、画像中の潜在的な概念に注目する。本論文ではこの概念を“一般クラス”と呼ぶことにする。たとえば、あるフォトアルバムに含まれる人間が写っている写真は、“human skin”という一般クラスを持つと考えることができる。また、別のフォトアルバムに含まれるプールや湖が写っている写真は、“water”という一般クラスを持つと考えることができる。

また、各ローカル画像から見ると、それぞれの画像には一般クラスが潜在的に含まれており、その含まれる一般ク

ラスにより特徴づけられると考えられる。つまり、視覚的に特異な特徴パターンを持つローカル画像は、潜在的に含まれる一般クラスによって、特徴づけられる面が大きいと考えられる。したがって、一般クラスに注目してサンプルを選ぶことができれば、その選ばれたサンプルは、多様な特徴パターンを含むと期待できる。

これらの一般クラスに注目した訓練サンプル選択を行うために、本提案手法では、2段階の学習を行う。まず、一般クラスに関する知識を、あらかじめラベル付けされた(汎用的な)画像群から抽出する。その後、分類対象のローカル画像に対して上記の知識を適用することで、学習に有用な訓練サンプルを選択する。

以下、2章では関連研究および本研究の位置づけ、3章では一般クラス概念について、その性質と有用性を述べる。4章では提案手法の詳細を述べ、5章では提案手法検証実験の結果および考察を報告する。最後に、6章では本研究のまとめと今後の課題について論じる。

2. 関連研究および本研究の位置づけ

本研究のように、少量のラベルあり訓練サンプルのみを用いて比較的高性能な分類器の学習を行う手法については、多くの研究成果が報告されている。

半教師あり学習 [6] は、分類器の学習において、ラベルありサンプルだけでなく、ラベルなしサンプルおよびそれらに対するラベル推定結果を訓練サンプルとして使用することにより、ユーザのラベリング量を削減できる分類器学習アプローチである。代表例は、Self-Training アルゴリズム [7] である。同アルゴリズムでは、まず、少量のラベルありサンプルを用いて分類器を学習し、この分類器を用いてラベルなしサンプルのラベルを推定する。次に、最も信頼度の高い一部のラベル推定結果および対応するサンプルを訓練サンプルとして追加し、分類器の更新を行う。以上の処理を、推定結果に対する閾値などによる終了条件に達するまで繰り返し行う。

一方、能動学習 [8] は少量のラベルありサンプルを用いて学習した分類器による評価結果に基づき、ラベルなしサンプルの中から、分類器の性能向上に寄与すると期待されるサンプルを選出し、ユーザにラベリングを提案する学習アプローチである。ここでラベリングされたサンプルは訓練サンプルに追加され、分類器が逐次的に更新される。学習効果が低いサンプルへのラベリングは回避されるので、ユーザのラベリング回数を抑えながら分類器の性能を向上させることができる。サンプルの選択基準として、「現在の分類境界の近くに位置するサンプルを選ぶ」や「Training Error を最小化するようなサンプルを選ぶ」など様々な種類が存在する [9], [10]。

上記2つのアプローチはいずれも分類器ベースの学習手法であり、少量のラベルありサンプルで初期化された分類

器を大量のラベルなしサンプルを用いることで改良し、比較的良好な分類器を得ることが可能である。本研究の提案手法は、これらの分類器ベースのアプローチと同じ目的を持つが、分類対象のラベルの情報をいっさい用いないという点で大きく異なる。分類器ベースのアプローチでは、分類対象のラベルの情報から学習した分類器の評価に基づき有用な訓練サンプルを選択するが、提案手法では各サンプルの視覚的な特徴に基づいて訓練サンプルを選択する。したがって提案手法は、分類器ベースのアプローチで用いる分類器を初期化するためのサンプルを提供することが可能であり、相補的な関係であると考えられる。

ラベリングコストを削減するためのもう1つのアプローチとして、ドメイン適応という手法がある。ドメイン適応 [11] は、解くために必要なデータや知識(目標ドメインのもの)が乏しい問題を解決するために、別の関連した問題のデータや学習結果(元ドメインのもの)を再利用する手法である。本研究では、ラベル付けされたデータセットから一般クラスに関する知識を学習した後、その結果をラベルなしローカル画像における訓練サンプル選択に利用する。この点で本研究は、ラベルあり元ドメインでの学習結果をラベルなし目標ドメインに転移させるドメイン適応の一種であると考えられる。しかし、これまでに報告された画像分類にドメイン適応を利用する研究 [12], [13], [14] では、目標ドメインにおけるサンプルのラベル推定を目的としたものが多く、本研究のように訓練サンプルの選択方法の最適化を目的としたものは報告されていない。

3. フォトアルバムにおける一般クラス概念

3.1 一般クラスが持つ性質

個人またはグループが撮影および所有しているフォトアルバムにおいては、撮影趣旨や撮影者の嗜好により、収録されている写真の内容は様々である。それにともない、写真整理における分類方法もフォトアルバムごとに様々だと予想される。

しかし、各フォトアルバムの撮影趣旨などとは別に、多くのフォトアルバムにおいて広く共有される概念は存在すると考えられる。本論文ではこの概念を“一般クラス”と呼ぶ。たとえば、図2に示すように、写真の中に海、池またはプールなど「水が存在する」ことを表現する一般クラス“water”や、「人間の肌が存在する」ことを表現する一般クラス“human skin”は、多くのフォトアルバムが潜在的に共有している。言い換えれば、一般クラスは、「河口湖が写っている」や「イギリス人の Mr. Bean が写っている」といった具体的かつ限定的な情報を表現するものではなく、やや抽象的かつ広汎な情報を表現するものである。一般クラスは、写真における一定の画像特徴を持つ物体領域の有無に関する情報であることが多い。本論文では、物体領域の有無は、既存の分類技術または視覚的特徴を用い

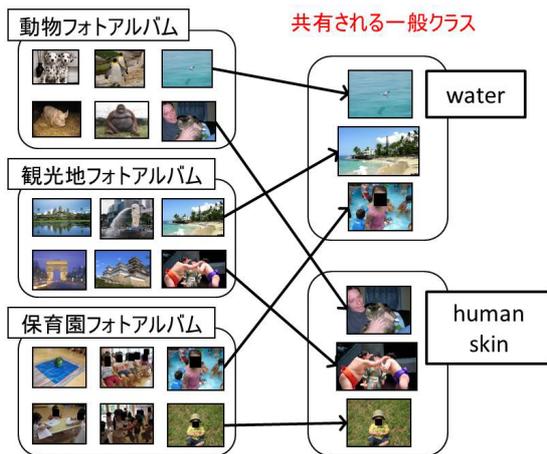


図 2 複数のフォトアルバムにわたって共有される一般クラス
 Fig. 2 Common classes shared by multiple photo albums.

ることにより、比較的安定した精度で判別することが可能であると想定する。

以上をまとめると、本論文でいう「一般クラス」は、写真の内容についての抽象的な情報の記述である。多くのフォトアルバムにおいて共有される普遍的な概念であり、低次元視覚的特徴を用いて抽出することが可能であるという性質を持つ。

一般クラスは、多数のフォトアルバムで共有されるという好ましい性質を持つほか、一般クラス自身の学習のために必要な訓練データが比較的容易に準備できるという利点もある。たとえば、ある一般クラスの名称を Web 上の検索エンジンに入力することで対応する画像を収集することや、Flickr [15] をはじめとする写真共有サイトにアップロードされている画像および付加されたタグを訓練データとして利用することもできる。また、物体認識に関する他の研究で利用されている既存のデータセットを利用することも可能である。

本研究では、一般クラスを設計することにより、分類対象のフォトアルバムにおいて特異なパターンを持つサンプルを、一般クラスに注目して探索できる点に着目している。提案する訓練サンプル選択手法では、一般クラスに関する知識を既存のデータセット（物体認識に関する研究で利用されているもの）から抽出し、分類対象であるローカル画像の訓練サンプル選択に導入することにより、ローカル画像における写真分類の低労力化を図る。

3.2 一般クラスに関する知識

提案手法は、個人やグループのフォトアルバムなど、まったくラベル付けされていないローカル画像を入力として、ユーザが事前に定義したクラスへの写真分類に寄与すると期待される訓練サンプルを自動選択する。ユーザに対して、このような訓練サンプルに対するラベリングを推薦することによって、比較的高精度な分類器の学習を、少ない

ラベリング量で行うことができる。

写真分類において視覚的に特異な特徴パターンを持つサンプルは、有用な訓練サンプルとなりうる。したがって提案手法では、まず、対象のローカル画像にクラスタリングを適用することによって特徴的なサンプルを選別する。また一方で、明示的にラベル付けは行われていないが、対象のローカル画像の中にも一般クラスは潜在すると考えられることから、ローカル画像に潜在する複数の一般クラスに注目して個別にクラスタリングを行い、特徴的なサンプルをそれぞれ探索する。そのため、提案手法では、事前に一般クラスに関する以下の 2 つの知識を学習する。

- 各一般クラスを表現する適切な特徴次元群
 各サンプルから計算される特徴ベクトルと各一般クラスとの対応を考えると、特徴ベクトルの中には、ある一般クラスの有無を決定づけるような特異な特徴次元と、そうではない特徴次元が混在していることが予想される。たとえば、画像の色を表現する特徴を用いる場合、画像に肌色が含まれているかどうかを示す特徴次元は一般クラス“human skin”の有無を、青色に反応する特徴次元は一般クラス“water”の有無をそれぞれ知るために重要な手がかりとなりうる。ローカル画像においてこのような特徴次元のみを用いてクラスタリングを行うことにより、各一般クラスに注目したクラスタリング結果を得ることが可能である。
- クラスタリング結果からの適切なサンプル選択方法
 ローカル画像においてクラスタリングを行うことにより、各サンプルはクラスタの重心近くに位置する、または遠く（クラスの境界付近）に位置するという性質で大きく二分される。ここで、各一般クラスにおいて特徴的なサンプルを得るためにはどちらの性質のサンプルを主に選択すればよいかを学習しておき、その選択方法に従うことにより、ローカル画像におけるクラスタリング結果においても、各一般クラスの中で特徴的なサンプルを得ることが可能である。

複数の一般クラスに注目して特徴的なサンプルを選択することによって、ローカル画像に単純なクラスタリングを適用する場合よりも、より多彩な訓練サンプルを細かく探し出すことが可能となる。また、一般クラスは数多くのフォトアルバムに共有される普遍的な概念であることから、提案手法は強い汎化能力を持つことが期待できる。

4. 提案手法

すでに述べているとおり、提案手法は、一般クラスに関する知識学習フェーズ（以下、学習フェーズと呼ぶ）と、ローカル画像からの訓練サンプル選択フェーズ（以下、選択フェーズと呼ぶ）からなる。以下、それぞれの詳細について述べる。

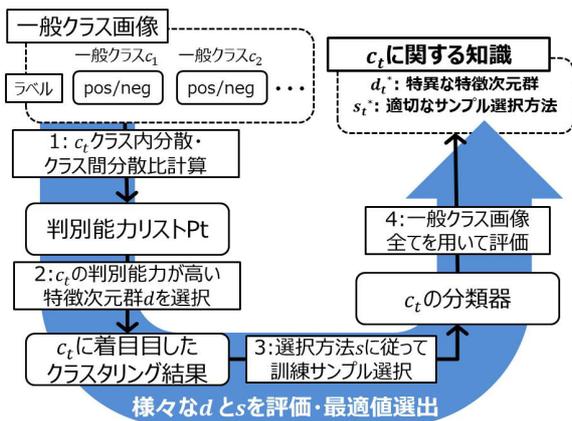


図 3 学習フェーズの処理の流れ

Fig. 3 Processing flow of learning phase.

4.1 一般クラスに関する知識学習フェーズ

第1段階である学習フェーズの目的は、各一般クラス $c_t \in (c_1 \dots c_T)$ の中で、特徴的なサンプルを選択するために必要な知識を学習することである。ここで d_t^* は一般クラス c_t を判別することのできる適切な特徴次元群とする。また、 s_t^* は対応する一般クラス c_t におけるクラスタリング結果からの最も適切なサンプル選択方法とする。 d_t^* および s_t^* を学習するために、あらかじめラベル付けされた一般クラスのデータセットを用いる。本論文ではこのデータセットを一般クラス画像と呼ぶ。一般クラス画像に対して、訓練サンプルの選択および分類器の評価を行うことで、各一般クラスに関する知識として d_t^* および s_t^* の適切な値を決定する。学習フェーズの概略を図3に示す。なお、学習フェーズは、一般クラスのデータセット内の処理であり、ローカル画像には無関係である。

4.1.1 知識学習アルゴリズム

本項では、一般クラスに関する知識学習の詳細について述べる。 $(d_1^* \dots d_T^*)$ と $(s_1^* \dots s_T^*)$ の学習は、一般クラス画像を用いて以下のように行われる。なおここで、一般クラス画像には、対応する一般クラス $(c_1 \dots c_T)$ の有無がラベル付けされているものとする。

まず、一般クラス画像から算出された l 次元の特徴ベクトルにおいて、全特徴次元の各 c_t に対するクラス間分散・クラス内分散比を計算する。次に、これらのクラス間分散・クラス内分散比に基づいて、各特徴次元が各 c_t の判別にどの程度有効かを表す判別能力リスト $P_t \in (P_{t1} \dots P_{tl})$ を得る。 $P_t = (p_{t1} \dots p_{tl})$ は各特徴次元がどの程度 c_t を判別できるかを表し、高い値を示すほど判別能力が高いことを意味する。 P_t を計算するためのアルゴリズムを Algorithm 1 に示す。

次に、 P_t を用いて d_t^* と s_t^* を学習する。そのための知識学習アルゴリズムを Algorithm 2 に示す。各 c_t について、 d_t^* と s_t^* の候補となる様々な特徴次元群 d とサンプル選択方法 s を用いて、一般クラス画像から訓練サンプルを選択

Algorithm 1 Compute the discriminability list P

Require: training samples: $D = \{(f_i, g_i)\}_{i=1}^n, g_i \in (c_1 \dots c_T)$

Require: number of data points in c_j : n_j

Require: element-wise multiplication: \otimes

Require: element-wise division: \oslash

- 1: $m := \frac{1}{n} \sum_{i=1}^n f_i$ {total mean of all samples}
- 2: $S_T := \sum_{i=1}^n (f_i - m) \otimes (f_i - m)$ {total variance of all samples}
- 3: for $t := 1$ to $num-class$ do
- 4: $m_t := \frac{1}{n_t} \sum_{i \in c_t} f_i$ {mean for class c_t }
- 5: $\sigma_t^2 := \sum_{i \in c_t} (f_i - m_t) \otimes (f_i - m_t)$ {within-class variance for class c_t }
- 6: end for
- 7: $S_W := \sum_{j=1}^T \sigma_j^2$ {total within-class variance}
- 8: $S_B := S_T - S_W$ {total between-class variance}
- 9: $P := S_B \oslash S_W$
- 10: return P

する。そして、選択した訓練サンプルを用いて分類器を学習し、その性能評価結果に基づき、各 d と s の中から d_t^* と s_t^* となるものを選出する。以下、同アルゴリズムの処理手順についてさらに詳細に述べる。

まず、 P_t に基づいて c_t を表現する適切な特徴次元を h 次元選択する。選択される特徴次元群 d は、各 c_t に対してクラス間分散・クラス内分散比が最大であるような特徴次元の上位 h 個である。次に、 d のみを用いて一般クラス画像に対してクラスタリングを実行することにより、 c_t に注目したクラスタリング結果を得る。最後に、 s に従ってクラスタリング結果から訓練サンプルを n 個選択する。 s はクラスタリング結果において、クラスタ重心に近いサンプル (centroid) およびクラスタ重心から遠いサンプル (border) のどちらを主に選択するかを表す値である。このように選択した訓練サンプルで学習した分類器の分類性能は、一般クラス画像全体をテストサンプルとして用いることで評価する。

上記の手順により、 d と s の様々な組合せを評価し、最も良好な分類性能を示した際の値を現在のサンプル選択数 n における最適値 d_t と s_t として記録する。これらの処理を n を変化させながら実行し、すべての n を通して選ばれた d_t および s_t を最終的な学習結果である d_t^* と s_t^* として採用する。

以上により学習された $(d_1^* \dots d_T^*)$ および $(s_1^* \dots s_T^*)$ は、次の選択フェーズで、ローカル画像において一般クラスに注目した訓練サンプル選択を行うために利用される。

4.2 ローカル画像からの訓練サンプル選択フェーズ

第2段階である選択フェーズの目的は、分類対象となるラベルなしローカル画像を入力として、複数の一般クラスに注目して、それぞれ特徴的なサンプルを選択を行うことにより、分類器の学習に有用な訓練サンプルを得ることである。訓練サンプル選択は、学習フェーズによって学習された知識 $((d_1^* \dots d_T^*)$ および $(s_1^* \dots s_T^*)$) に基づいて行わ

Algorithm 2 Learn dimensions and selection method

(d_t^*, s_t^*)

Require: training samples: $D = \{(f_i, g_i)\}_{i=1}^N$

Require: discriminability list: P

Require: selecting methods M

Require: evaluation function $L(d, s, n)$

```

1: for  $t := 1$  to  $num-class$  do
2:   for  $n := num-patt1$  to  $num-patt2$  do
3:      $L^* := 0$ 
4:     for  $h := num-dime1$  to  $num-dime2$  do
5:       Add the first  $h$  elements of  $P_t \in P$  to  $d$ 
6:       Clustering on  $D$  with dimension set  $d$ 
7:       for all  $s$  in  $M$  do
8:         Select  $n$  samples from resulting clusters by  $s$ 
9:         if  $L(d, s, n) > L^*$  then
10:           $d_t^n := d$ 
11:           $s_t^n := s$ 
12:           $L^* := L$ 
13:        end if
14:      end for
15:    end for
16:  end for
17:  Choose the most frequently appeared  $d_t^n$  and  $s_t^n$  as  $(d_t^*, s_t^*)$ 
18: end for
19: return  $(d_t^*, s_t^*)$ 
    
```

Algorithm 3 Select valuable training samples

Require: target local images: $I = \{f_i\}_{i=1}^M$

Require: number of training samples to be selected: n_{tr}

Require: discriminative dimensions for class c_t : d_t^*

Require: sample selection method for class c_t : s_t^*

```

1:  $I_{tr} := \Phi$ 
2: for  $t = 1$  to  $num-class$  do
3:   select discriminative features  $I' = \{f'_i\}_{i=1}^M$  based on  $d_t^*$ 
4:   clustering on  $I'$ 
5:   obtain clustering result  $R_t$ 
6: end for
7: for all  $R_t \in (R_1 \dots R_T)$  do
8:    $n := n_{tr}/num-class$ 
9:    $n := [n]$  {the maximal integer that does not exceed  $n$ }
10:  select  $n$  samples from  $R_t$  by  $s_t^*$  and put them into  $I'_{tr}$ 
11:   $I_{tr} := I_{tr} \cup I'_{tr}$ 
12: end for
13: return  $I_{tr}$ 
    
```

について述べる. $(d_1^* \dots d_T^*)$ および $(s_1^* \dots s_T^*)$ を用いた訓練サンプル選択アルゴリズムを Algorithm 3 に示す.

まず, 対象のローカル画像に対し, すべての一般クラス $(c_1 \dots c_T)$ ごとに, 学習された d_t^* と同一の特徴次元選択を行う. 次に, ローカル画像に対して, 上で選択された特徴次元のみを用いたクラスタリングを, すべての一般クラス $(c_1 \dots c_T)$ ごとに, それぞれ独立に実行し, 各 c_t に注目した T 個のクラスタリング結果を得る. さらに, 各クラスタリング結果から, それぞれ同じ量の訓練サンプルを $(s_1^* \dots s_T^*)$ に基づいて選択する. 最後に, 選択された訓練サンプルをユーザに提示し, 写真整理に利用したいと望むクラス分類でのラベリングを求める.

5. 検証実験

提案手法の有効性を検証するため, 日常生活を記録したフォトアルバムを用いて, 検証実験を実施した. 本章ではこの実験について述べる.

本実験の目的は, フォトアルバム中の写真分類において, 提案手法がどの程度分類器の学習に有用な訓練サンプルを選択できるか, また, 提案手法を用いることによりユーザのラベリング量をどの程度減らせるかを検証することである. そこで, 計 21,424 枚の写真データを含むフォトアルバムにおける写真分類を対象として実験を行い, 提案手法の性能を 2 つのベースライン手法と比較する. そのベースライン手法とは, 一般クラスを考慮せずに単純なクラスタリングによる訓練サンプルを選択して分類器を学習した場合, および訓練サンプルをランダムに選択して分類器を学習した場合である.

5.1 実験手順

本実験の全体像は, 次のとおりである (図 5).
 (1) ローカル画像および一般クラス画像から 1,440 次元の

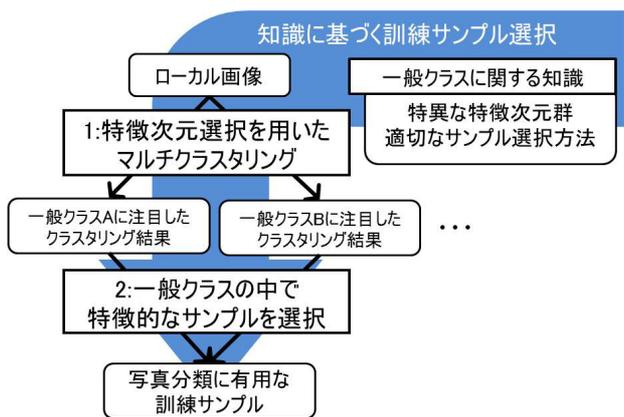


図 4 選択フェーズの処理の流れ
 Fig. 4 Processing flow of selecting phase.

れる.

まず, 各一般クラス c_t の判別に有効な特徴次元 d_t^* のみを用いたクラスタリングを, 対象のローカル画像に対してそれぞれ独立に行う. 同一画像が複数のクラスのクラスタに含まれることが許される. 本論文では, 上記の処理をマルチクラスタリングと呼ぶ. これにより, $(c_1 \dots c_T)$ に注目した T 個のクラスタリング結果を得る. 次に, s_t^* に従って, 各クラスタリング結果からそれぞれ適切なサンプル選択を行う. 選択フェーズでの処理により, ローカル画像における分類器の学習に有用なサンプルをユーザに提供することが可能となる.

選択フェーズの概略を図 4 に示す.

4.2.1 訓練サンプル選択アルゴリズム

本項では, ローカル画像からの訓練サンプル選択の詳細

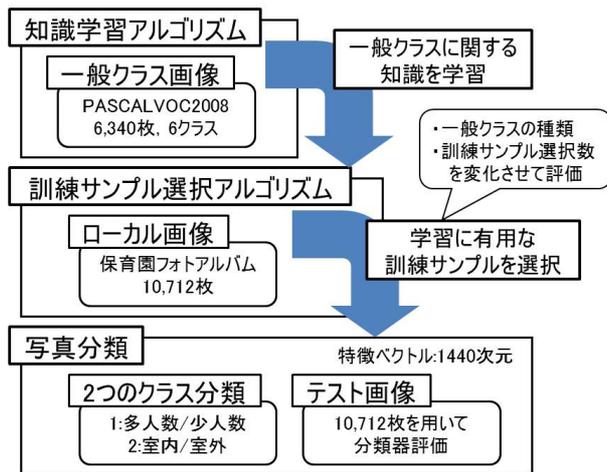


図 5 提案手法検証実験の概要

Fig. 5 Summary of the experiment for verifying our proposed method.

特徴ベクトル (後述) を抽出。

- (2) 一般クラス画像 6,340 枚および 6 種類の一般クラスのラベルから、一般クラスに関する知識学習アルゴリズムを用いて知識を学習。
- (3) フォトアルバムのうち半数のローカル画像 10,712 枚を入力として、訓練サンプル選択アルゴリズムを用いて、複数の一般クラスの中で特徴的なサンプルを選択。
- (4) 上記で得られた訓練サンプルに対し、写真整理のためにあらかじめ定義したクラスに従って手動でラベリングを行い、分類器を学習。
- (5) フォトアルバムの残り 10,712 枚をテスト用ローカル画像として分類し、選択した訓練サンプルで学習した分類器とベースライン手法との性能比較を行うことにより、提案手法の有効性を検証。

なお、本実験では 2 つのパラメータを変化させながら提案手法を評価した。

1 つめは、提案手法において注目する一般クラスの数と種類である。本実験では、注目する一般クラスと提案手法の性能との関係を調べるために、用意した 6 種類の一般クラスの中の任意の組合せ (計 63 パターン) すべてを用いて、それぞれ個別に訓練サンプル選択および分類器の学習・評価を行った。

2 つめは、提案手法を用いてローカル画像から選択する訓練サンプルの数である。本実験では、訓練サンプル選択枚数の変化にともなう提案手法の性能変化の傾向を調査するために、訓練サンプル選択枚数を変化させつつ、提案手法による訓練サンプル選択と分類器の学習および評価を行った。

5.2 実験設定

5.2.1 ローカル画像

本実験では、対象のローカル画像として、保育園の日常



図 6 保育園フォトアルバムに含まれる写真の一部

Fig. 6 Some photos in a nursery school album.

生活を記録した計 21,424 枚からなるフォトアルバムを用意した。保育園フォトアルバムに含まれる写真の例を図 6 に示す。

このフォトアルバムには主に、園児と保育士が遊んでいる場面や、お遊戯をしている場面、昼食をとっている場面、そして運動会や卒園式、ハロウィンといった特別な行事に参加している場面をそれぞれ撮影した写真が含まれている。写真はすべて 800 × 600 の解像度で撮影されている。

本実験ではこのアルバムを均等に 2 分割し、一方を提案手法への入力画像として使用し、他方を分類器の評価におけるテスト画像として使用した。

5.2.2 学習において使用するクラス定義および分類器

本実験では、一般クラスに関する知識の学習に使用する一般クラス画像として、PASCALVOC2008 データセット [17] および Farhadi らの研究 [18] において提供されたアノテーションデータを用いた。

PASCALVOC2008 データセットは、一般物体認識における代表的なベンチマークであり、人間や数々の動物、車や建造物といった多種多様な物体を画像中に含んでいる。Farhadi らは画像中の物体が持つ「属性」に注目し、各物体の外接矩形それぞれに物体が持つ属性の情報をラベルとして PASCALVOC2008 の画像群に付加した。計 6,340 個の物体に対して「形状」、「部品」、「材質」の 3 種類の属性が付加されている。

本実験では、この 3 種類の属性のうち物体の「材質」を特定する属性を、多くのフォトアルバムで共有される性質であると考え、対応するアノテーションデータを一般クラスのラベルとして採用した。使用する一般クラスは“Glass”, “Metal”, “Skin”, “Wool”, “Vegetation”, “Wood” の計 6 種類とした。各一般クラスが表現する写真の性質を表 1 に示す。また、PASCALVOC2008 から物体領域の画像をクリッピングし、アノテーションに基づいてラベル付けした計 6,340 枚の画像を一般クラス画像とした。一般クラス

表 1 各一般クラスが表現する写真の性質

Table 1 Correspondences between photos and common classes.

一般クラス名	表現する写真の性質
Glass	コップやグラスなど, ガラス製の物を含む.
Metal	車や鉄骨など, 金属製の物を含む.
Skin	人間の肌を含む.
Wool	布や繊維を含む.
Vegetation	鉢植えや草むらなどの植物を含む.
Wood	フローリングの床や椅子など, 木製の物を含む.

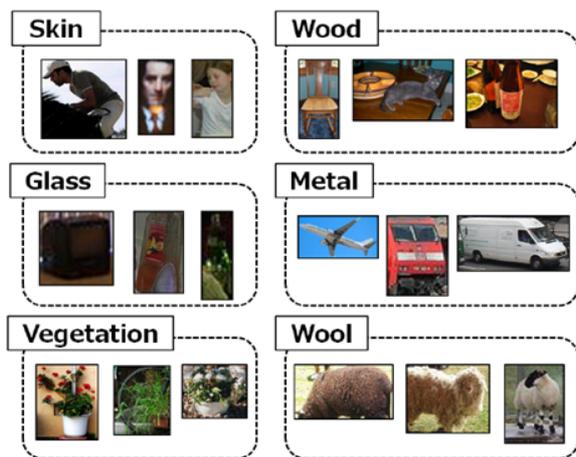


図 7 一般クラス画像の例

Fig. 7 Some examples of common class images.

画像の例を図 7 に示す.

一般クラスの d_t^* と s_t^* を学習するにあたっては, 各一般クラス c_t について選択次元数 h の値を 100 から 500 まで 100 単位で評価するとともに, サンプル選択方法 s を以下の 3 つの内容で評価した.

- [centroid] クラスタ重心に最も近いサンプルを選択.
- [border] クラスタ重心から最も遠いサンプルを選択.
- [one-half-each] 上記規則で得たサンプルの両方を半数ずつ選択.

知識学習アルゴリズムにおける分類器の評価は, サンプル選択数 n を 100 から 1,000 まで 100 単位で変化させてそれぞれ独立に行われ, 各 h で選ばれた d_t および s_t に基づき, d_t^* と s_t^* を多数決により決定する. 分類器は動径基底関数カーネル SVM [19], [20] を用いた. また, 分類器の学習では特徴ベクトル中のすべての次元を用いた.

5.2.3 特徴

本実験では, 各画像につき以下の 4 種類の特徴を計算し, 結合することで得られる 1,440 次元の特徴ベクトルを用いた. 学習, 選択の両フェーズいずれにも利用した.

- 色ヒストグラム
色情報を表現する特徴として, 画像を RGB 空間で 512 色に減色し, 各色に対応するピクセル数のヒストグラムを用いる (512 次元).
- エッジ
物体がどの程度多く存在しているかを示す特徴として

物体の輪郭を表現する 16 次元のエッジ特徴を用いる. 画像を均等に 4×4 の領域に分割し, CannyEdge 検出器を適用して得られたエッジピクセルの数から得る.

- SIFT を用いた bag-of-visual words 表現
局所的かつ特徴的なパターンの有無を表現する特徴として, SIFT [21] を記述子として利用した 400 ワードの bag-of-visual words 表現 [22] を用いる. コードブックは PASCALVOC2011 からの 5,717 枚の画像を用いて計算した 400 個の visual words を用いる. 画像中の各キーポイントから 128 次元の SIFT 特徴量を計算し, コードブックに従ってベクトル量子化を行うことで 400 本のピンを持つヒストグラムを得る (400 次元).
- GIST
画像全体のシーンを表現する特徴として, 512 次元の GIST 特徴量 [23] を用いた. 数種類の方向および周波数の Gabor フィルタを適用することにより計算する.

5.2.4 分類において使用する分類器およびクラス定義

分類器として, 知識学習アルゴリズムと同様に, 動径基底関数カーネル SVM [19], [20] を用いた. また, 分類器の学習において, 提案手法により選択された訓練サンプルの特徴ベクトル中のすべての次元を用いた.

また, 保育園フォトアルバムにおける写真分類は, 以下の 2 つのクラス定義に基づいて行った. これらは保育士へのインタビューにより, 彼らが写真整理で用いたいという要望があったクラスである. 選択された訓練サンプルへのラベリングは, これら 2 つのクラス定義に基づいて行われた.

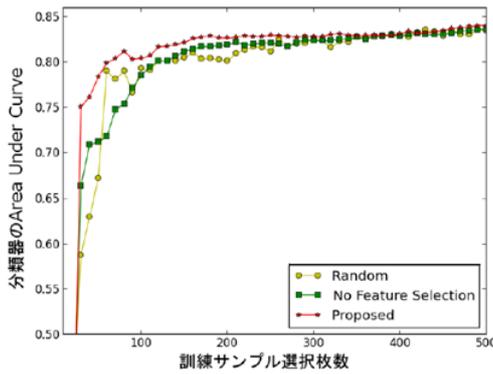
- クラス定義 A
写真中に多数の人が存在するか否かにより分類. 4 人以上の人間が含まれている場合は多人数クラスと判定. 4 人未満の場合は少人数クラスと判定.
- クラス定義 B
写真が室内でまたは室外で撮られたものであるかにより分類.

5.2.5 分類器の評価指標

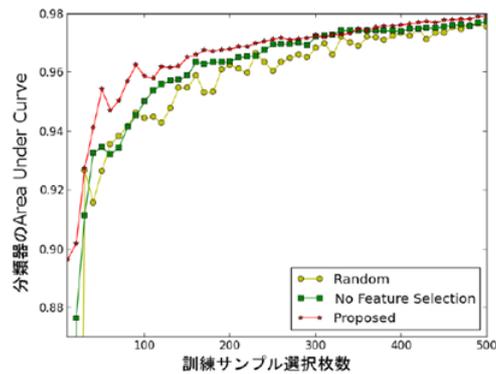
本実験では, 分類器の評価指標として AUC (Area Under Curve) を用いた. AUC はテストサンプルに対する分類結果の ROC (Receiver Operating Characteristic Curve) の下部の面積である. ここで ROC 曲線とは, 2 クラス分類器を用いて分類器のパラメータを変化させながらテストサンプルを分類した結果における真陽性率を縦軸および偽陽性率を横軸として描画される, 分類器の性能を表す曲線である. AUC の値が高いほど良質な分類器であることを示す. 本実験では, テストサンプルに対する分類器のラベル推定結果における閾値 (手動で決定) により, テストサンプルの正例と負例を判定した.

5.2.6 ベースライン手法

本実験では, 提案手法と比較するために以下 2 つの手法をベースライン手法として設定した.



(a) クラス定義 A における提案手法の評価結果



(b) クラス定義 B における提案手法の評価結果

図 8 提案手法の識別精度およびベースライン手法との比較

Fig. 8 Recognition accuracy of our proposed method and comparison with baseline methods.

表 2 各一般クラスからの知識学習結果

Table 2 The results of common class learning.

一般クラス	特徴次元数 h^*	サンプル選択方法 s^*
Vegetation	400	centroid
Skin	300	one-half-each
Wood	200	centroid

(1) 一般クラス考慮なし (単純なクラスタリング)

一般クラスをまったく考慮せずに訓練サンプル選択を行う。すなわち、特徴次元選択を行わずに特徴ベクトル中のすべての次元を用いてクラスタリングを1度だけ行い、各クラスタの重心に近いサンプルのみを訓練サンプルとして選択する。提案手法において一般クラスに注目することが訓練サンプルの選択に与える影響を評価する。

(2) ランダム選択

ローカル画像の中からランダムに選択したサンプルを訓練サンプルとする。ただし、ランダム選択はその能力に大きな分散が存在することが予想されるため、本実験ではランダム選択を計5回独立に実行し、各分類器のAUCの平均を評価する。提案手法が実際に分類器の学習に効果的な訓練サンプルを選択できているかどうかを評価する。

5.2.7 クラスタリング・アルゴリズム

提案手法におけるすべての処理を通じて、クラスタリングには Repeated Bisection アルゴリズム [16] を採用した。同アルゴリズムは分割最適化クラスタリングの1つであり、各サンプル間のコサイン類似度を評価尺度として用いる。

5.3 実験結果および考察

一般クラスに関する知識学習アルゴリズム (学習フェーズ) による、6種類の一般クラスからの学習結果を表2に示す。計63パターンの組合せすべてのうち、一般クラス “Skin”, “Wood”, “Vegetation” に注目した場合に、提案手法は最も良い性能を示した。そこで、ローカル画像からの訓練サンプル選択の場面 (選択フェーズ) では、上記3

つの一般クラスに注目した場合の提案手法の性能について述べることにする。

提案手法およびベースライン手法により選択された訓練サンプルを用いて学習したクラス定義 A における分類器の評価結果を図8(a)に、クラス定義 B における分類器の評価結果を図8(b)に示す。ここでは訓練サンプル選択枚数を横軸として、選択枚数を10サンプルから500サンプルまで10サンプル単位で変化させつつ、各選択枚数に対する分類器のAUCをプロットした。いずれもAUCの値が高いほど良い。

クラス定義 A および B の両方の場合において、提案手法を用いて学習した分類器は、ベースライン手法を用いた場合と比較すると、100枚以下という少量の訓練サンプルを選択した場合について、顕著に良い性能を示した。50枚の場合を例にあげると、クラス定義 A において、提案手法では、約0.80のAUCを示しているのに対して、2つのベースライン手法では、同等の性能に達するために120枚のラベリング量を必要としている。また、クラス定義 B においては、提案手法では約0.95を示しているのに対し、ベースライン手法のうち、一般クラス考慮なし (単純なクラスタリング) の場合は、同等のAUC値に達するまでに120枚のラベリング量を、またランダム選択の場合は、150枚のラベリング量を必要とした。

以上から提案手法は、分類器の性能をベースライン手法と同等の値に保ったまま、ラベリング量をおよそ1/2から1/3に削減することに成功したと考えられる。この結果は、ユーザによるラベリング作業を少量に抑えつつ、良い性能を持つ分類器を提供するという本研究の目的を十分に満たすものであり、写真整理の低労力化において提案手法が有効であることが実証されている。

また、計63パターンの一般クラスの組合せに対する実験結果から、保育園フォトアルバムの中でよく現れる一般クラスに注目した場合に、提案手法の性能が向上するという結果が得られている。すなわち、良い性能を示した一般

クラスは、実際に画像内容とも相関があった。まず、保育園フォトアルバムには、一般クラス“Vegetation”に対応する植物の葉や樹木を含む写真が多く存在していた。また、一般クラス“Skin”に対応する人間を含む写真、および“Wood”に対応する木製の床やテーブルを含む写真がそれぞれ多く存在していた。以上から、提案手法の性能と、注目する一般クラスの種類には画像内容に基づく関係があり、提案手法の性能を高めるためには、対象のローカル画像の内容に関係する一般クラスを用いることが必要であることがうかがえる。

6. おわりに

本研究では、フォトアルバムにおける写真分類を効率化することを目的として、フォトアルバム（ローカル画像）から分類器の学習に有用な訓練サンプルを選択する新しい手法を提案した。提案手法では、一般クラスに関する2つの知識を一般クラス画像から獲得し、ローカル画像に適用することで一般クラスを考慮した訓練サンプルの選択を可能にした。一般クラスに関する知識学習フェーズにおいては、一般クラス画像に対して訓練サンプルの選択および分類器の評価を行うことで各一般クラスに関する知識として適切な値を決定する。また、ローカル画像からの訓練サンプル選択フェーズにおいては、上記で得られた知識を用いて対象のローカル画像にマルチクラスタリングを適用し、クラスタリング結果から学習に有用な訓練サンプルを選択する。これらの処理により、多彩かつ特徴的な訓練サンプルを一般クラスごとに細かく探し出すことを可能にした。また、一般クラスは数多くのフォトアルバムに共有される普遍的な概念であることから、提案手法は強い汎化能力を持つことが期待できる。

計 21,424 枚からなる保育園フォトアルバムを用いた提案手法検証実験では、提案手法は、2種類のベースライン手法と比較して、比較的少量の訓練サンプルを選択したときに、分類器の性能を同等の値に保ちつつ、ラベリング量をおよそ 1/2 から 1/3 に削減することができることを示している。

今後の課題として、第 1 に、対象のローカル画像に適した一般クラスを自動設定する方法を開発する必要がある。実用的な場面において提案手法を適用するためには、ローカル画像の内容に基づき、潜在する一般クラスを自動的に抽出する必要がある。この課題は、Flickr など、画像共有サイトから画像およびその画像に付与されたタグを収集することで解決できると考えられる。たとえば、対象のローカル画像と視覚的特徴が類似している画像を探索し、それらの画像を一般クラス画像、多く付与されているタグを一般クラスとして設定する。この場合、一般クラスに関するユーザのラベリングは不要である。

第 2 に、多数のフォトアルバムにおいて提案手法の評価

を行うことによって、提案手法の汎化能力および一般クラスの有用性を調査することがあげられる。3章で述べたように、数多くのフォトアルバムにおいて普遍性が高い一般クラスを導入することにより、提案手法は高い汎化能力を有することが期待できる。

第 3 に、本研究で提案した一般クラス概念を写真分類以外の課題に適用して、有用性を評価することがあげられる。一般クラスは写真以外にも、文書や音声の分類においても有用であると予想している。文書分類の例でいえば、各文書における特定の単語の出現頻度や、文章のパターンから算出される種々の素性は、一般クラスの性質論で述べた「低次元特徴」にあたる。これらの低次元特徴を用いることで、文書内に「政治についての記述が含まれる」「機械についての記述が含まれる」といった情報を表現する一般クラスを抽出することができると考えられる。

謝辞 本研究の一部は、情報通信研究機構 (NICT)、総務省戦略的情報通信研究開発推進制度 (SCOPE) (No.131306004)、JSPS 特別研究員奨励費 (No.2410663) の支援によるものである。

参考文献

- [1] 黄瀬浩一：特定物体認識，電子情報通信学会技術研究報告，PRMU，パターン認識・メディア理解，Vol.109，No.306，pp.79-87 (2009).
- [2] Google Inc.: Google Picasa, available from (<http://picasa.google.com>).
- [3] NTT docomo: フォトコレクション，入手先 (<http://www.nttdocomo.co.jp/service/entertainment/photo-collection/>).
- [4] Kooaba Japan: kooaba, available from (<http://kooaba.jp/technology>).
- [5] 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：コンピュータビジョンとイメージメディア，Vol.48，No.SIG-16(CVIM-19)，pp.1-24 (2007).
- [6] Zhu, X.: Semi-Supervised Learning Tutorial, *Proc. International Conference on Machine Learning* (2007).
- [7] Yarowsky, D.: Unsupervised Word Sense Disambiguation Rivaling Supervised Methods, *Proc. 33rd Annual Meeting on Association for Computational Linguistics*, pp.189-196 (1995).
- [8] Dasgupta, S. et al.: A Tutorial on Active Learning, *Proc. International Conference on Machine Learning* (2009).
- [9] Tong, S. and Koller, D.: Support Vector Machine Active Learning with Applications to Text Classification, *Journal of Machine Learning Research*, pp.45-66 (2001).
- [10] Seeger, M.: Learning with Labeled and Unlabeled Data, Technical Report, Institute for ANC, Edinburgh, UK (2000).
- [11] 神鳥敏弘：転移学習，人工知能学会誌，Vol.25，No.4，pp.572-580 (2010).
- [12] Thrun, S.: Is Learning The n-th Thing Any Easier Than Learning The First?, *Advances in Neural Information Processing Systems-8*, Mozer, M.C., Touretzky, D.S. and Hasselmo, M.E. (Eds.), pp.640-646, MIT Press, Cambridge, MA (1996).
- [13] Quattoni, A. et al.: Transfer Learning for Image Classification with Sparse Prototype Representations, *CVPR*,

- pp.1-8 (2008).
- [14] Wang, H. et al.: Dyadic Transfer Learning for Cross-domain Image Classification, *ICCV*, pp.551-556 (2011).
 - [15] Flickr, Yahoo! Inc.: available from <http://www.flickr.com/>.
 - [16] Ziao, Y. et al.: Comparison of Agglomerative and Partitional Document Clustering Algorithms, Technical report, Department of Computer Science, University of Minnesota (2002).
 - [17] Visual Object Classes Challenge: available from <http://pascallin.ecs.soton.ac.uk/>.
 - [18] Farhadi, A. et al.: Describing Objects by Their Attributes, *CVPR*, pp.1778-1785 (2009).
 - [19] Vapnik, V.: *The Nature of Statistical Learning Theory*, Springer (1995).
 - [20] Bishop, C.M.: *Pattern Recognition and Machine Learning*, Springer (2006).
 - [21] Lowe, D.G.: Distinctive Image Features from Scale-Invariant Keypoints, *International Journal of Computer Vision*, Vol.60, No.2, pp.91-110 (2004).
 - [22] Nowak, E. et al.: Sampling Strategies for Bag-of-Features Image Classification, *ECCV* (2006).
 - [23] Oliva, A. et al.: Modeling the Shape of the Scene: A Holistic Representation of the Spatial Envelope, *International Journal of Computer Vision*, Vol.42, No.3, pp.145-175 (2001).



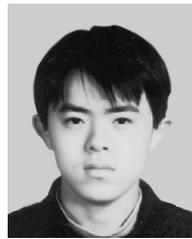
加藤 ジェーン (正会員)

1993年名古屋大学大学院工学研究科情報工学専攻博士後期課程修了。同年富山大学工学部助手。1999~2000年Oxford大学客員研究員。現在、名古屋大学大学院情報科学研究科准教授。移動体追跡，行動認識，人物同定，映像要約，機械学習等に関する研究に従事。博士(工学)。IEEE Computer Society シニア会員。



白須 遼

2011年名古屋大学工学部電気電子・情報工学科卒業。2013年名古屋大学大学院情報科学研究科博士前期課程修了。現在、オークマ株式会社に勤務。



王 彧

Computer Society 会員。

2013年名古屋大学大学院情報科学研究科博士後期課程修了。2012年より日本学術振興会特別研究員。現在、名古屋大学大学院情報科学研究科研究員。博士(工学)。コンピュータビジョン，機械学習の研究に従事。IEEE Com-



間瀬 健二 (正会員)

1979年名古屋大学工学部電気工学科卒業。1981年同大学大学院工学研究科情報工学専攻前期課程修了。同年日本電信電話公社(現在NTT)入社。1988~1989年米国MITメディア研究所客員研究員。1995~2002年(株)国際電気通信基礎技術研究所(ATR)研究室長。2002年より名古屋大学教授。現在、同大学大学院情報科学研究科所属。人工知能学会1999年度論文賞，同2013年功労賞受賞。電子情報通信学会フェロー。博士(工学)。