不均衡データからのランダムフォレストを利用した高速高次特徴抽出

Fast Extraction Method of High-Level Feature Using Random Forests from Imbalanced Training Data

正会員 河合 吉彦[†], 正会員 住吉 英樹[†], 正会員 藤井 真人[†],
正会員 柴田 正啓[†], 正会員 馬場 口登^{††}

Yoshihiko Kawai[†], Hideki Sumiyoshi[†], Mahito Fujii[†], Masahiro Shibata[†] and Noboru Babaguchi^{††}

Abstract We propose a method of using a random forest algorithm to quickly detect semantically high-level features such as specific objects. The random forest has a lower computation cost than that of the common algorithm such as a support vector machine (SVM). However, it cannot cope with training data that have a large bias in the number of negative and positive examples. We improve the conventional training algorithm to ensure sampling the data with equal probability from each class when creating bootstrap samples, which increases the classification accuracy. Experiments on the Caltech-101 dataset resulted in a recall of 64.3% and precision of 71.1%, which were comparable to those of conventional methods. The average time needed for training and for detection were reduced to one sixteenth and one twenty-seventh that of SVM, respectively.

キーワード：高次特徴抽出，不均衡学習データ，ランダムフォレスト

1. ま え が き

大量の映像データの中から，所望の映像を効率的に探索するためには，色や動きといった表層的な特徴だけでなく，映像の意味内容を考慮した解析が重要である．このような解析技術のひとつに高次特徴抽出がある．高次特徴抽出は，映像から具体的なオブジェクトやイベント（高次特徴と呼ぶ）を検出する技術であり，学習データの変更によって様々な種類の高次特徴に対応できる汎用的な手法を目的とした技術である．高次特徴抽出では，映像から取得した代表フレームから特徴ベクトルを算出し，機械学習によって分類するというアプローチが一般的である．これまでに，SIFT¹⁾などの局所領域における特徴^{2)~4)}や，色やテキストチャ⁵⁾などの大域的な特徴に基づく特徴ベクトルを，サポートベクターマシン（SVM）で判別する手法が提案されている．

従来手法における問題のひとつとして，学習データ数や

特徴ベクトルの次元数の増大に起因する計算コストの増加がある．膨大な映像データを処理する場合，計算時間は非常に重要な要素となる．そこで本稿では，従来のSVMに比べ，より高速な処理が期待できるランダムフォレスト⁶⁾を利用した高次特徴抽出手法を提案する．しかし，ランダムフォレストは，学習データに大きな偏りがあると分類精度が低下するという問題があり，本手法にそのまま適用することができない．ほとんどの場合において，ある高次特徴が出現する映像区間（正例）よりも，出現しない映像区間（負例）が圧倒的に多いためである．そこで本手法では，このようなデータの偏りに対応できる新たな学習アルゴリズムを提案する．具体的には，学習データから部分集合を生成する処理過程において，従来の重み付けによる補正ではなく，サンプリング方法の修正によってデータ偏りの影響を軽減する．高次特徴を判別するための画像特徴ベクトルについては，局所特徴と大域特徴の両方を組合せて算出し，本提案の学習アルゴリズムで分類することによって，高次特徴抽出の高速化と検出精度の向上を図る．

2. 高次特徴抽出手法

提案手法の概要を図1に示す．まず，bag-of-visual-wordsと呼ばれる局所特徴に基づくアプローチ³⁾と，画像の大域的な特徴を組合せることによって，入力画像から特徴ベク

2010年8月2日受付，2010年10月7日再受付，2010年10月15日採録

[†] NHK 放送技術研究所

(〒157-8510 世田谷区砦 1-10-11, TEL 03-5494-3387)

^{††} 大阪大学 大学院 工学研究科

(〒565-0871 吹田市山田丘 2-1, TEL 06-6879-7744)

[†] Science and Technical Research Laboratories, NHK

(1-10-11 Kinuta, Setagaya-ku, Tokyo, 157-8510, Japan)

^{††} Graduate School of Engineering, Osaka University

(2-1 Yamadaoka, Suita-shi, Osaka, 565-0871, Japan)

図 1 提案手法の概要
Overview of the proposed method.

トルを算出する。次に、それらをランダムフォレストで分類し、画像中に特定の高次特徴が出現しているか否かを判定する。以降では、各処理の詳細を説明する。

2.1 局所特徴の算出

事前に、学習データから算出した SIFT 特徴を k 平均法で M クラスにクラスタリングし、各クラスの代表ベクトルを求めておく。次に、入力画像から SIFT 特徴を算出し、代表ベクトルとの距離を考慮して M 次元の特徴ベクトルを算出する。ベクトルの i 番目の要素 v_i は式 (1) で算出する⁷⁾。

$$v_i = \sum_{j=1}^B \sum_{k=1}^{C_j} \frac{1}{2^{j-1}} \text{sim}(p_k, w_i), \quad 1 \leq i \leq M \quad (1)$$

C_j は SIFT 特徴点から j 番目に距離が近い代表ベクトルが w_i であるような特徴点の総数を表し、 $\text{sim}(p_k, w_i)$ は、SIFT 特徴 p_k と代表ベクトル w_i との類似度を表す。本手法では $B = 4$ とした⁷⁾。

さらに、画像内における SIFT 特徴点の位置を反映できるように、入力画像を縦横 2×2 および 1×3 のグリッド領域に分割して、各領域ごとに特徴ベクトルを求め、それらをつなぎ合わせることでフレーム全体の特徴ベクトルとする。分割数は予備実験の結果を考慮して決定した。

2.2 大域特徴の算出

大域特徴としては、以下に示す 3 種類を利用する。

(1) 色モーメント

色モーメントは、色分布を反映した特徴量である。入力画像を HSV および $L^*a^*b^*$ 色空間に変換し、各コンポーネント c ($c \in \{h, s, v, l, a, b\}$) に対する画素値の平均 μ_c 、標準偏差 σ_c 、歪度の立方根 s_c を特徴量とする。画像を 5×5 の合計 25 個のグリッドに分割し、各グリッド領域に対して特徴量を算出する。分割数は予備実験に基づいて決定した。

(2) ハールウェーブレットテクスチャ

ハールウェーブレットは、画像のテクスチャを反映した特徴量である。画像を 3×3 のグリッドに分割し、各領域に対して 2 次元ハールウェーブレット変換を 3 段階適用する。次に、それぞれのサブバンド領域に対して画素値の分散を算出し特徴量とする。画像の分割数などは予備実験に

基づいて決定した。

(3) ローカルバイナリパターン

ローカルバイナリパターン⁸⁾は、注目画素に対する周辺画素の濃度の大小パターンを表す特徴量である。半径 R の円周上の P 個の画素から算出されるローカルバイナリパターン $L_{P,R}$ は式 (2) で表される。

$$L_{P,R}(x, y) = \begin{cases} \sum_{p=0}^{P-1} \delta_{P,R}(x_p, y_p), & \text{if } U_{P,R}(x, y) \leq 2 \\ P + 1, & \text{otherwise} \end{cases} \quad (2)$$

$\delta_{P,R}$ は、注目画素 (x, y) より周辺画素 $(x + x_p, y + y_p)$ の画素値が大きい場合には 1 を、小さい場合には 0 を返す関数である。ここで、 $x_p = R \cos \frac{2\pi p}{P}$ 、 $y_p = R \sin \frac{2\pi p}{P}$ である。また、 $U_{P,R}$ は $\delta_{P,R}$ の系列において 0 と 1 が変化する箇所の総数を表し、式 (3) で算出される。

$$U_{P,R}(x, y) = |\delta_{P,R}(x_{P-1}, y_{P-1})| + \sum_{p=1}^{P-1} |\delta_{P,R}(x_p, y_p) - \delta_{P,R}(x_{p-1}, y_{p-1})| \quad (3)$$

提案手法では、解像度変化に耐性を持たせるため、 $(P, R) = (8, 1), (16, 2), (24, 3)$ について⁸⁾、 $L_{P,R}$ の頻度ヒストグラムを求める。

3. 不均衡データに対応したランダムフォレスト学習

ランダムフォレスト⁶⁾はアンサンブル学習の一種であり、多数の決定木を組合せることによって、精度の高い識別器を実現する。アルゴリズムを以下に示す。 N は生成する決定木の総数、 M は特徴ベクトルの次元数を表す。

- (1) 学習データから N 個の部分集合（ブートストラップサンプル） $\Theta_1, \Theta_2, \dots, \Theta_N$ を作成する。ブートストラップサンプルは、学習データから復元抽出でランダムサンプルすることによって作成する。
- (2) Θ_i を用いて決定木 T_i を学習する。決定木の各ノードの分岐においては、 M 個の全特徴量の中から m 個 ($m < M$) をランダムに選択し、その中から最良のものを選択する。良否の判定には、ジニ係数を用いる。木は最大まで成長させ、枝刈りはしない。
- (3) 分類結果は、学習された決定木 $T_1, T_2, \dots, T_N$ の投票によって決定する。

このアルゴリズムは、学習データにおける各クラス（正例と負例）のデータ数に大きな差がない場合は良好に動作するが、データ数の差が大きくなると、数の多いクラスに偏った識別器が学習されるという問題がある。これは、ジニ係数に基づく分岐では、多数派のクラスを重視した特徴量が選択されるためである。これに対して従来手法⁹⁾では、各クラスのデータ数に応じて、データに重み付けする手法を提案している。つまり、少数クラスのデータに大きな重みを与え、それらが誤って分類された際のペナルティを大きくする。この手法は、偏りの度合いが小さい場合には有効

表 1 実験結果
Experimental results.

高次特徴	SVM			提案手法		
	再現率	適合率	F 値	再現率	適合率	F 値
airplane	46% (358/770)	99% (358/363)	0.63	82% (632/770)	95% (632/664)	0.88
bonsai	11% (11/ 98)	69% (11/ 16)	0.19	38% (37/ 98)	49% (37/ 76)	0.43
car_side	74% (69/ 93)	97% (69/ 71)	0.84	99% (92/ 93)	97% (92/ 95)	0.98
chandelier	18% (14/ 77)	67% (14/ 21)	0.28	25% (19/ 77)	40% (19/ 47)	0.31
faces	17% (70/405)	90% (70/ 78)	0.29	62% (251/405)	67% (251/377)	0.64
faces_easy	42% (84/405)	97% (84/ 87)	0.59	79% (320/405)	79% (320/406)	0.79
ketch	25% (21/ 84)	78% (21/ 27)	0.38	61% (51/ 84)	29% (51/178)	0.39
leopards	92% (157/170)	99% (157/159)	0.95	88% (150/170)	97% (150/155)	0.92
motorbikes	66% (508/768)	100% (508/510)	0.80	74% (570/768)	99% (570/574)	0.85
watch	20% (42/209)	84% (42/ 50)	0.32	35% (73/209)	60% (73/122)	0.44
平均	41.3%	86.4%	0.56	64.3%	71.1%	0.68

である。しかし、偏りが非常に大きい場合、ブートストラップサンプルには非常に大きな重みを持つ少数のデータが含まれるため、これらに過度に適応した識別器が学習される恐れがある。また、少数クラスのデータがブートストラップサンプルに全く含まれないという場合もありうる。

そこで提案手法では、ブートストラップサンプルの作成において、各クラスから等確率でデータを復元抽出するようアルゴリズムを改良する。学習データの総数に比べて、ブートストラップサンプルのデータ数が十分に小さい場合には、少数クラスからも多様性のあるデータが選択され、より汎化能力の高い識別器の生成が期待できる。提案手法におけるブートストラップサンプル Θ_i の作成手順を示す。説明では、 Θ_i に含まれる学習データの総数を D とおく。

- Θ_i を初期化, $\Theta_i \leftarrow \phi$
- Θ_i のデータ数が D になるまで、以下を繰り返す。
 - (1) 無作為にクラス c を選択する。
 - (2) クラス c に属するデータを復元抽出でランダムサンプルし、 Θ_i に追加する。

上述の処理を N 回実行し、 $\Theta_1, \Theta_2, \dots, \Theta_N$ を作成する。

4. 実験

Caltech-101⁹⁾ データセットを利用した評価実験により、提案手法の有効性を検証した。Caltech-101 は 101 種類の高次特徴からなる合計 9,144 枚の画像セットである。画像の数は高次特徴によってばらつきがあり、少ないもので 31 枚、多いもので 800 枚となっている。

4.1 実験の設定

101 種類の高次特徴から無作為に 30 枚ずつ、合計で 3,030 枚の画像を選択し、学習データとした。残りはテストデータとした。本手法の目的は、大量のデータから特定の高次特徴を抽出することである。そのため実験では、ある高次特徴とその他の高次特徴を分類する識別器を高次特徴ごとに学習した。学習データにおける正例は 30 枚、負例は 3,000 枚となる。なお本実験では、101 種類の高次特徴のうち、特にデータ数の多い 10 種類を抽出対象とした。学習時のパラメータについては、決定木の本数 N を 500、ブートストラップサンプルのデータ数 D を学習データの総数の 5%、

表 2 処理時間の比較結果
Comparison of processing time.

手法	学習	抽出	合計
SVM	210.6 秒	377.7 秒	588.3 秒
RF	9.0 秒	13.6 秒	22.6 秒
提案手法	13.3 秒	13.8 秒	27.1 秒

決定木の分岐における特徴量候補数 m を特徴次元数 M の 2% とした。また比較手法としては、様々な手法で広く利用されている RBF-SVM (SVM) と、重み付けによって補正する従来のランダムフォレスト法 (RF) を用いた。抽出精度の評価には、下式で定義される再現率 (R) と適合率 (P)、F 値 (F) を用いた。

$$R = \frac{N_b}{N_g}, \quad P = \frac{N_b}{N_o}, \quad F = \frac{2 \cdot R \cdot P}{R + P} \quad (4)$$

ここで、 N_b は本手法で正しく検出された高次特徴の総数を表し、 N_g は正解の高次特徴の総数を表す。 N_o は、本手法によって検出された高次特徴の総数を表す。再現率は未検出の少なさを、適合率は誤検出の少なさを反映した評価値である。また F 値は、トレードオフの関係にある再現率と適合率を統合した評価値を表す。

4.2 実験結果

表 1 に 10 種類の高次特徴に対する実験結果を示す。SVM は、平均再現率が 41.3%、平均適合率が 86.4% という結果となった。平均 F 値は 0.56 となり、比較的、良好な結果が得られた。また提案手法は、平均再現率が 64.3%、平均適合率が 71.1% となり、平均 F 値は 0.68 となった。SVM と比較しても遜色のない良好な結果が得られた。それに対して RF は、いずれの高次特徴も全く抽出することができなかった。そのため、表中に結果を示していない。本実験で使用した学習データは、正例と負例の数に 100 倍の偏りがあったため、少数の正例に過度に適応した決定木が学習されたものと思われる。また、ブートストラップサンプルに正例が全く含まれないケースもあったものと考えられる。このようにデータの偏りが大きい場合、RF の重み付けによる補正では充分に対応できなかったといえる。

次に、処理時間の比較結果を表 2 に示す。実験に使用した計算機のスペックは、CPU が Intel Xeon 3.5GHz、RAM

図 2 再現率-適合率グラフ
Recall-precision graph.

が 4GB である。提案手法と RF は、SVM と比較して計算時間が非常に短くなった。提案手法と SVM を比較した場合、学習時間は約 16 分の 1、抽出時間は約 27 分の 1 となった。なお、SVM における処理時間の多くは、特徴ベクトルをスケーリングするための処理で占められていたが、それらを除いた場合でも、提案手法は SVM の 3 分の 1 程度の処理時間であった。また、SVM は、表 2 以外に、学習パラメータを推定するためのグリッドサーチに約 12 時間を要した。提案手法は非常に計算コストが低いため、大規模なデータセットを検索対象とする場合や、より多くの種類の高次特徴を扱う場合においても、実用的な時間で処理を完了することができる。

学習データに含まれる負例の数を調整することによって、データ数の偏りを軽減した場合、抽出精度がどう変化するかを調べた。RF と提案手法に対する実験結果を図 2 に示す。図中の各点に対するラベルは学習データに含まれる正例と負例の比率を示している。提案手法については、負例の割合が増加するに従い、適合率が段階的に向上した。種々の負例が学習データに加えられることによって、誤検出が軽減されたものと考えられる。RF については、ほとんど精度が得られなかった。学習データに含まれる正例の絶対数が少なかったためと考えられる。そこで、ブートストラップサンプルのデータ数を増やした場合（図中、RF, D=25%）についても調査したが、やはり提案手法には及ばなかった。抽出精度の向上のためには、学習データに種々の負例を含める必要があるといえるが、従来の RF では、このような偏りの大きな学習データに対応できないといえる。

最後に、他の代表的な従来手法との比較について述べる。Caltech-101 を利用した研究の多くは、入力画像を 101 種類のいずれかのクラスに分類することを目的としている。手法の目的が異なるため、単純な比較は困難であるが、これらの従来手法⁴⁾¹⁰⁾の認識率は 65%前後となっている。それに対して、本手法は平均再現率が 64.3%、平均適合率が 71.1%であり、十分な精度を実現できていると考える。

5. む す び

本稿では、ランダムフォレストを利用した高速な高次特

徴抽出手法を提案した。本手法では、学習データにおける正例と負例の数に大きな偏りがある場合においても、識別精度を低下させない新たな学習アルゴリズムを提案した。具体的には、ブートストラップサンプルの生成処理において、各クラスから等確率でデータをサンプリングすることにより、正例と負例の両方から多様性のあるデータを選択するようにした。Caltech-101 データセットを対象とした評価実験では、平均再現率 64.3%、平均適合率 71.1%となり、サポートベクターマシンや従来の重み付けによる補正方法と比較して良好な結果が得られた。また、学習処理と抽出処理に要した計算時間は、それぞれ、サポートベクターマシンの約 16 分の 1、約 27 分の 1 にまで削減できた。

今後は、他のデータセットや他の高次特徴に対して実験を実施し、手法の有効性を確認したい。また、大量の映像データを対象とした手法についても検討をすすめたい。

〔文 献〕

- 1) D.G. Lowe, "Object recognition from local scale invariant features," in Proc. IEEE Int. Conf. Computer Vision, pp.1150-1157 (1999)
- 2) C. Schmid and R. Mohr, "Local grayvalue invariants for image retrieval," IEEE Trans. Pattern Analysis and Machine Intelligence, **19**, 5, pp.530-535 (1997)
- 3) G. Csurka, C. Bray, C. Dance and L. Fan, "Visual categorization with bags of keypoints," in Proc. ECCV workshop on Statistical Learning in Computer Vision, pp.59-74 (2004)
- 4) S. Lazebnik, C. Schmid and J. Ponce, "Beyond bags of features: spatial pyramid matching for recognizing natural scene categories," in Proc. IEEE Computer Vision and Pattern Recognition, pp.2169-2178 (2006)
- 5) D.-D. Le, S. Satoh and T. Matsui, "NII-ISM, Japan at TRECVID 2007: High level feature extraction," in Proc. TRECVID Workshop (2007)
- 6) L. Breiman, "Random forests," Machine Learning, **45**, pp.5-32 (2001)
- 7) Y.-G. Jiang, C.-W. Hgo and J. Yang, "Towards optimal bag-of-features for object categorization and semantic video retrieval," in Proc. ACM Int. Conf. Image and Video Retrieval, pp.494-501 (2007)
- 8) T. Ojala M. Pietikainen and T. Maenpaa, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," IEEE Trans. Pattern Analysis and Machine Intelligence, **24**, 7, pp.971-987 (2002)
- 9) <http://www.vision.caltech.edu/archive.html>
- 10) H. Zhang, A.C. Maire and J. Malik, "SVM-KNN: discriminative nearest neighbor classification for visual category recognition," in Proc. IEEE Int. Conf. Computer Vision and Pattern Recognition, pp.2169-2178 (2006)

かわい よしひこ
河合 吉彦 1999 年，大阪大学基礎工学部情報工
学科卒業。2001 年，同大学大学院修士課程修了。同年，
NHK 入局。放送技術局を経て，2005 年から放送技術研
究所勤務。博士（工学）。メディア処理の研究に従事。正
会員。

すみよし ひでき
住吉 英樹 1980 年，広島県立広島工業高校電気
科卒業。同年，NHK 入局。広島放送局を経て，1984 年
から放送技術研究所勤務。現在，放送技術研究所 専任研
究員。工学博士。コンピュータを応用した番組制作シス
テム，メタデータ制作システムの研究に従事。正会員。

ふじい まひと
藤井 真人 1983 年，名古屋大学大学院工学研究
科電気工学専攻修士課程修了。同年，NHK 入局。1987
年から放送技術研究所勤務。1991 年，カーネギーメロン
大学客員研究員。1998～2001 年，ATR 人間情報通信研
究所室長。現在，放送技術研究所 主任研究員。博士（情
報科学）。画像認識，視覚のバイオサイバネティクス，立
体画像情報処理に関する研究に従事。正会員。

しばた まさひろ
柴田 正啓 1979 年，京都大学工学部電子工学科卒
業。1981 年，同大学大学院修士課程修了。同年，NHK
入局。新潟放送局，放送技術研究所，技術局，放送技術局
を経て，現在，放送技術研究所 人間・情報科学研究部 部
長。博士（情報学）。情報検索，映像インデキシング，映
像制作システム，エージェント技術の放送への応用の研
究，FIPA，MPEG-7，ARIB の標準化活動に従事。正
会員。

ばば ごろ のぼる
馬場口 登 1979 年，大阪大学工学部通信工学科
卒業。1981 年，同大学大学院修士課程修了。1982 年，愛
媛大学工学部助手。大阪大学工学部助手，講師，産業科
学研究助教授を経て，現在，大阪大学大学院工学研究
科教授。1996～1997 年，カリフォルニア大学サンディエ
ゴ校・文部省在外研究員。工学博士。メディア処理，人工
知能の研究に従事。PCM2006，IAS2009 Best Paper
Award，FIT2009 論文賞などを受賞。正会員。