

2種類の異なる特徴を用いた特定物体検出

Object Detection Method Using Two Heterogeneous Features

小坂谷 達夫 伊藤 聡 窪田 進

■ KOZAKAYA Tatsuo

■ ITO Satoshi

■ KUBOTA Susumu

画像から特定の物体を検出する技術は、デジタルカメラでの顔の検出など様々な分野でその応用が見込まれており、近年活発に研究開発が行われている。しかし、多種多様なパターンを持つ物体を精度良く検出することは困難な課題である。

東芝は、人の顔の検出に利用される高速処理が可能な画像特徴と、人物検出に利用される物体の詳細な情報を抽出可能な高次元の画像特徴という2種類の異なる特徴量を用いることで、複雑なパターンを持つ特定の物体を高速かつ高精度に検出する手法を開発した。この手法を猫の顔検出に適用して評価実験を行った結果、性能面及び計算コスト面で有効性が高いことを確認できた。

Technologies to detect a specific object in images are expected to further expand to a wide range of applications, such as face detection functions for digital cameras and so on. However, high-accuracy detection of an object with a number of different patterns remains a difficult issue.

As a solution to this issue, Toshiba has developed an object detection framework that is generic and efficient enough to deal with complicated shapes and textures at high speeds based on two heterogeneous features: joint Haar-like features, and co-occurrence histograms of oriented gradients (CoHOGs). We evaluated the combination of these two distinct classifiers through a cat face detection problem, and confirmed the effectiveness of this method in terms of both performance and computational cost.

1 まえがき

近年、画像から特定の物体を検出する技術が実用化に向けて急速に進んでおり、例えばデジタルカメラでは人の顔を自動的に検出して合焦する機能は一般的なものとなっている。東芝も、AVノートPC (パソコン) Qosmio™に搭載している顔ナビ™などに人の顔の検出技術を利用している。人の顔以外にもペットの顔など特定の物体を検出するニーズも数多く存在するが、人の顔よりもはるかに複雑で多様なパターンを持つ物体に対し、人の顔の検出で使われた技術をそのまま適用すれば十分な性能が得られるとは限らない。

当社は、複雑なパターンを持つ特定の物体に対しても高速かつ高精度に検出する手法を開発した。この手法は、人の顔の検出に用いているJoint Haar-like特徴⁽¹⁾と、人物検出に用いているCoHOG (Co-occurrence Histograms of Oriented Gradients)特徴⁽²⁾それぞれによる識別器を直列に接続して利用する。Joint Haar-like特徴は、矩形(くけい)領域の平均輝度差に基づくHaar-like特徴と呼ばれる特徴量を複数組み合わせたもので、高速に計算できるという利点があるが、複雑なパターンの識別には適していない。一方、CoHOG特徴は、二つの輝度勾配(こうばい)の組合せで表現される共起関係から求めたヒストグラムに基づく数万次元に及ぶ非常に高次元の特徴量で、強力な識別能力を持つがHaar-like特徴と比較して計算に時間がかかる。

そこで、まずJoint Haar-like特徴を用いた識別器で画像内

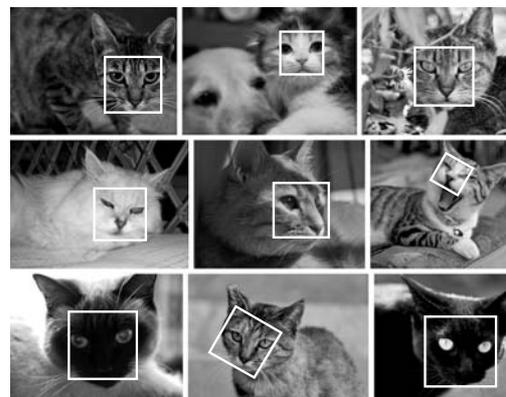


図1. 猫の画像例 — 猫画像データベースの例で、図中の矩形は開発した手法による検出結果である。

Examples of results of applying newly developed method

から候補を高速に抽出し、次に得られた少数の候補領域に対してCoHOG特徴による検証処理を行う。このように2種類の傾向のまったく異なる特徴を直列に接続して相補的に利用することで、高速な検出処理と高い検出精度の両方を実現した。

ここでは、猫の顔への適用を例に、開発した物体検出手法の仕組みと評価結果について述べる。

2 猫の顔の検出

猫はペットとして一般的であり、被写体としての人気も非常

に高い。しかし、猫の顔は個体ごとに毛並みや模様が様々に変化するため、人の顔と比較して検出することが難しい対象である。猫の画像の例を図1に示す。図中の四角は、開発した手法で検出した結果である。このように画像中から特定の物体を検出する技術は、大量に保存されている写真に対して、任意の被写体の表示や整理をするようなアプリケーションに適用することができる。

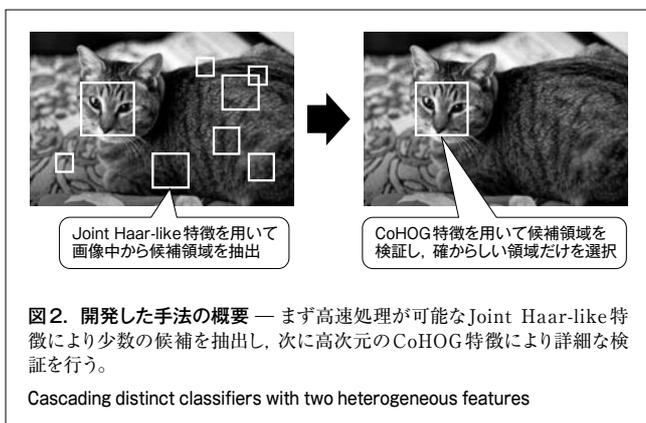
3 2種類の異なる特徴を用いた特定物体検出手法

開発した手法は大きく分けて二つのステップで構成される(図2)。まず、画像から様々な大きさや場所についてJoint Haar-like特徴を用いた識別器によって顔と非顔を判定し、顔らしい候補領域を抽出する。次のステップでは、抽出された各候補領域に対してCoHOG特徴を計算し、線形識別器により検出すべき顔かどうかを最終的に判定する。

前段のステップは、画像中の様々な位置や大きさを考慮した膨大な数の探索領域から、ごく少数の顔らしい候補領域に絞り込むことを目的としている。Haar-like特徴は、画像の解像度や候補領域のサイズに依存せずに高速に計算することができる⁽³⁾。ここでは、背景などを含む顔以外の領域を許す(過検出)が、真の猫の顔領域はなるべく見逃さないように識別器のしきい値を設定する。

後段の処理で用いるCoHOG特徴は、パターン内の輝度勾配の共起関係をヒストグラム化した共起行列に基づいて計算される。様々な輝度勾配ペアの組合せで表現される共起行列は、局所的な形状情報を詳細に抽出することができる。抽出されたCoHOG特徴は非常に高次元となるが、類似度の計算には線形識別器を用いるため高速に処理することができる。また、CoHOG特徴の抽出はJoint Haar-like特徴により候補領域を絞り込んだ後の処理になるため、CoHOG特徴での輝度勾配やヒストグラムの計算が全体の処理速度に与える影響は少ないと考えられる。

それぞれの特徴の詳細について、以下に説明する。



3.1 Joint Haar-like特徴

Haar-like特徴は、図3に示すような複数の矩形領域内の平均輝度差に基づいて物体か非物体かを判定する。個々のHaar-like特徴を用いた識別器は弱識別器とも呼ばれ、それぞれの弱識別器の識別性能は高くないが、ブースティング⁽³⁾と呼ばれる学習方法により、複数の特徴を統合した強識別器を構築することで高い識別率を実現している。

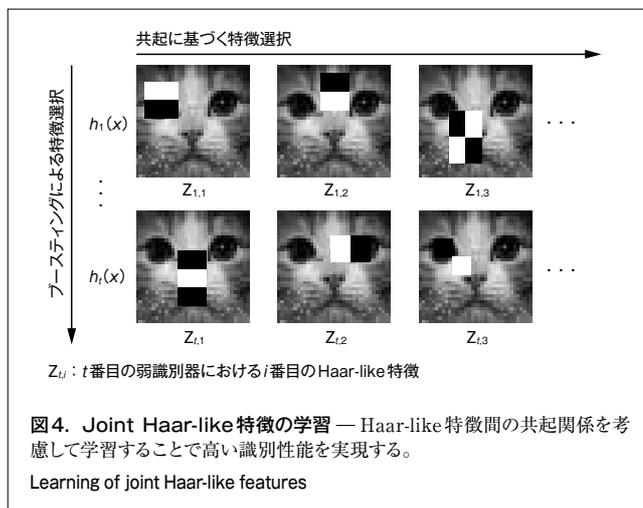
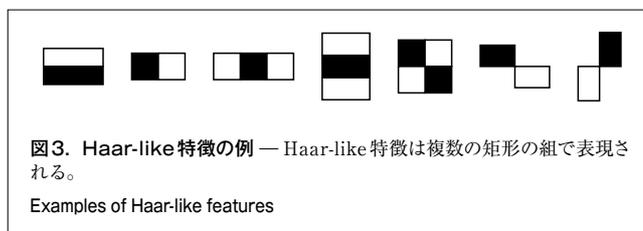
当社が開発したJoint Haar-like特徴は、Haar-like特徴の間の共起関係を考慮することで、更に識別能力を高めている。

Haar-like特徴は、2値化された値、つまり検出すべきサンプル(正例)であれば1、背景などの検出すべきでないサンプル(負例)であれば0を返すように定義されている。これに対しJoint Haar-like特徴は、F個のHaar-like特徴の出力を束ねたFビットの値として表現される(図4)。

Joint Haar-like特徴を用いた弱識別器($h_t(x)$)は、式(1)で定義される。ここで、 y はクラスラベル、 P_t はあるサンプル x が与えられた時のJoint Haar-like特徴の出力 j に対する各クラスの条件付き確率である。

$$h_t(x) = \begin{cases} +1 : P_t(y=+1 | j) > P_t(y=-1 | j) \\ \text{のとき} \\ -1 : \text{そのほかのとき} \end{cases} \quad (1)$$

Joint Haar-like特徴を学習するためには、Haar-like特徴のすべての組合せから最適な共起関係を見つけ出す必要がある。Haar-like特徴がM種類存在したときのF個の共起関係



は膨大な組合せになるため、全探索を行うことは現実的ではない。そこで、識別率が改善するように一つずつ特徴を追加していくことで、準最適な共起関係を高速に求められるように工夫している。また、条件付き確率 P_i に関しては、正例と負例のそれぞれの学習サンプルから計算することができる。

強識別器 $H(x)$ は、Joint Haar-like特徴を用いた識別器 $h_i(x)$ を T 個組み合わせる形で表現することができる。ここで、 a_i はブースティングによる学習で得られた弱識別器に関する重みである。

$$H(x) = \text{sign} \left(\sum_{i=1}^T a_i h_i(x) \right) \quad (2)$$

猫の顔の学習では、最終的な識別器は、10個の強識別器を直列に接続し、4,388個のHaar-like特徴で構成されており、正検出率が95.3%、過検出率は $2.92 \times 10^{-3}\%$ であった。一方、人の顔を学習させた場合は11個の強識別器でHaar-like特徴が3,169個、正検出率が95.6%、過検出率が $3.48 \times 10^{-7}\%$ であった。人の顔と比較して、猫の顔の学習には1.3倍の特徴数を用いてもはるかに低い検出性能しか達成できず、猫の顔のような複雑なパターンを従来の顔検出の枠組みだけで学習することがいかに困難であるかがわかる。

3.2 CoHOG特徴

CoHOG特徴は、輝度勾配の組合せによる2次元ヒストグラムから得られる共起行列に基づいて計算する。この共起行列は、画像中のある特定のオフセット（注目点からの相対位置）を持つ二つの輝度勾配のペアが、どのような出現頻度を持つかを表現している。

輝度勾配は画素ごとに式(3)で計算することができる。

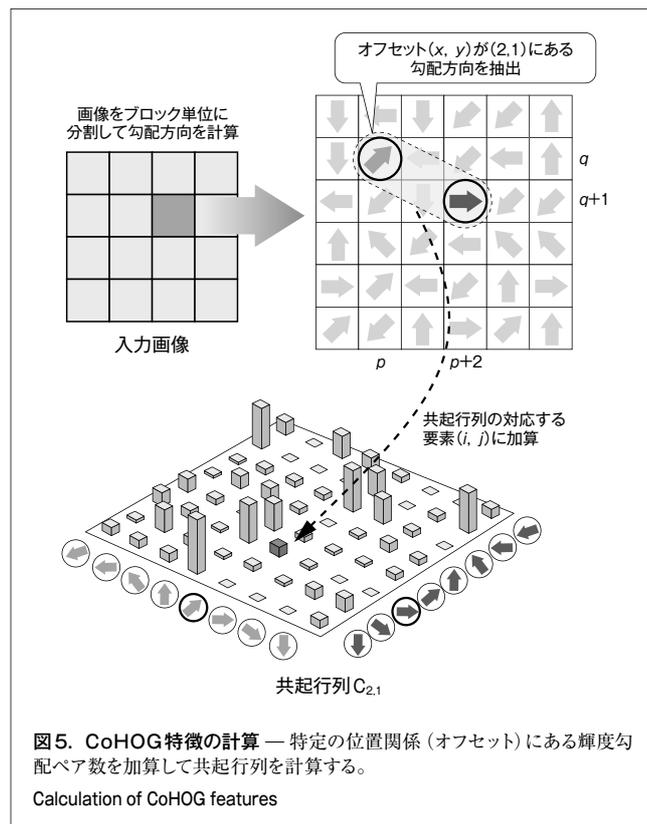
$$\theta = \arctan \frac{v}{h} \quad (3)$$

ここで、 v と h はエッジ検出フィルタで求めた縦方向と横方向の勾配を表しており、得られた角度 θ は、45度単位で分割して8種類のラベルとして用いる。ただし、勾配強度 $\sqrt{v^2+h^2}$ が所定のしきい値よりも小さいものに関しては共起行列の計算には使用しない。

入力されたパターンは、 $N \times M$ の小領域（ブロック）に分割され、パターン内の画素ごとに輝度勾配を計算しておく。共起行列 $C_{x,y}$ は小領域ごとに特定のオフセット (x, y) 単位で計算され、式(4)で示すように、輝度勾配ペア (i, j) に対応する共起行列の要素に値を加算する。

$$C_{x,y}(i, j) = \sum_{p=1}^n \sum_{q=1}^m \begin{cases} 1 : I(p, q) = i \text{ かつ} \\ \quad I(p+x, q+y) = j \\ \quad \text{のとき} \\ 0 : \text{そのほかのとき} \end{cases} \quad (4)$$

ここで、 $I(p, q)$ はある位置 (p, q) における輝度勾配方向のラ



ベル、 (n, m) は分割した領域のサイズである。

8方向の輝度勾配を二つ持ち、オフセット $(2, 1)$ の場合の共起行列の例を図5に示す。オフセットは注目する画素に対して4画素の範囲内で定義され、重複を取り除くと30個の組合せが存在する。最終的にゼロオフセットを含む31種類の共起行列が小領域ごとに計算され、それらすべての共起行列をつなぎ合わせてベクトル化したものがCoHOG特徴として用いられる。評価実験ではブロックを 7×7 に分割しており、この場合のCoHOG特徴の次元数は $(30 \times 64 + 8) \times 7 \times 7 = 94,472$ である。CoHOG特徴は極端に次元数の高い特徴量であるため、識別器には高速に計算できる線形識別器を利用して計算量を削減している。

4 評価実験

4.1 評価実験の準備

評価実験は、図1に示すような約10,000枚の公開猫画像データベース⁽⁴⁾を用いて行った。このデータベースをランダムに分割し5,000枚を学習用、2,000枚を学習の検証用、3,000枚を評価用として利用した。検出結果の判定は、検出された領域とあらかじめ人が与えた真の正解との重なり率が50%を超えた場合に正解とした。

データベースには、ほぼ正面を向いた猫の様々な大きさの顔が収録されている。学習パターンは、顔の大きさと傾きを補正



図6. 学習サンプルの例 — 様々な猫のパターンを用いて学習を行う。
Examples of training samples

するため、両目を基準にしてJoint Haar-like特徴用に 25×25 画素を、CoHOG特徴用に 35×35 画素を切り出し、それぞれの識別器を学習させた(図6)。

4.2 評価実験の結果

評価実験の結果として、正検出率と過検出数の関係を図7に示す。グラフの横軸は過検出したパターン数、縦軸は評価用の画像にあるすべての猫の顔に対して正しく検出された割合である。

従来手法として、猫の顔に独特な特徴に着目して高い性能を実現したZhangらの猫顔検出手法⁽⁴⁾との比較を行った。開発した手法は検出対象に特化した特徴を用いていないにもかかわらず、従来手法よりも高い正検出率を達成した。正検出率を76%に固定したときの過検出数は、Zhangらの手法の20に対して開発した手法は7であり、誤りを半分以下に削減することができた。この結果から、開発したJoint Haar-like特徴とCoHOG特徴というまったく異なる2種類の特徴を用いた検出手法が有効に働いていることが確認できた。

次に計算コストの面から開発した手法の有効性について評価するため、Intel®Core™(注1) 2 Quadプロセッサ(動作周波数: 2.66 GHz)を搭載したPCで開発した手法の処理時間

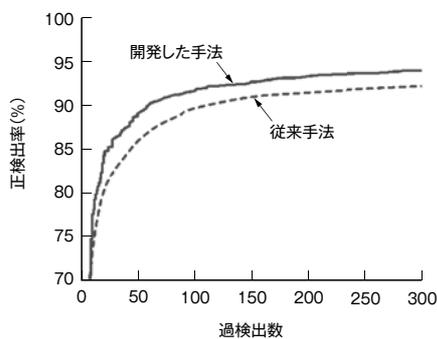


図7. 評価実験の結果 — 従来手法と比較して、開発した手法は同じ過検出数でより高い正検出率を達成した。

Results of experiments using conventional and newly developed methods

(注1) Intel, Intel Coreは、米国及びその他の国における米国Intel Corporation又は子会社の登録商標又は商標。

を計測した。380×326画素の入力画像に対して顔の傾きを補正するため30度ごとに12枚の回転画像を作成し、すべて検出させるのに約3秒かかった。処理時間の内訳は前段のJoint Haar-like特徴による識別器は95%、後段のCoHOG特徴による識別器ではわずか5%であった。つまり、CoHOG特徴による検証処理として5%の計算量を追加するだけで、Joint Haar-like特徴だけによる識別器と比較してはるかに高い検出性能が得られることを意味しており、計算コストの面でも開発した手法の有効性が高いことが確認できた。

5 あとがき

ここでは、Joint Haar-like特徴とCoHOG特徴というまったく性質の異なる2種類の特徴量を相補的に用いることで、複雑なパターンを持つ物体を精度良く高速に検出する手法を開発した。この手法を猫の顔の検出に適用して評価した結果、検出精度と処理速度の両方で開発した手法の有効性を確認できた。

今後、様々な物体へ適用範囲を広げ、画像をキーとした検索技術の研究・開発を進めていく。

文 献

- (1) Mita, T., et al. Discriminative feature co-occurrence selection for object detection. IEEE Trans. on Pattern Analysis and Machine Intelligence. 30, 7, 2008, p. 1257 - 1269.
- (2) Watanabe, T., et al. "Co-occurrence histograms of oriented gradients for pedestrian detection". The 3rd Pacific-Rim Symposium on Image and Video Technology (PSIVT2009), Lecture Notes in Computer Science (LNCS). 5414, Tokyo, Japan, 2009-01, IEEE. p.37 - 47.
- (3) Viola, P.; Jones, M. Robust real-time face detection. Int. Journal of Computer Vision. 57, 2, 2004, p.137 - 154.
- (4) Zhang, W., et al. "Cat head detection - how to effectively exploit shape and texture features". The 10th European Conference on Computer Vision, LNCS. 5305, Marseille, France, 2008-10 IEEE. p.802 - 816.



小坂谷 達夫 KOZAKAYA Tatsuo

研究開発センター マルチメディアラボラトリー 研究主務。
画像処理, パターン認識の研究・開発に従事。情報処理学会
会員。
Multimedia Lab.



伊藤 聡 ITO Satoshi

研究開発センター マルチメディアラボラトリー。
画像認識技術の研究・開発に従事。電子情報通信学会
会員。
Multimedia Lab.



窪田 進 KUBOTA Susumu

研究開発センター マルチメディアラボラトリー 主任研究員。
画像処理, パターン認識の研究・開発に従事。電子情報通信
学会
会員。
Multimedia Lab.