Fisher Vectorを用いた スマートフォン上でのリアルタイム高精度物体認識システム

河野 憲之[1,a)]　柳井 啓司[1,b)]

1. はじめに

近年、クアッドコアのスマートフォンが主流になり性能が大きく向上している。従来の画像認識を行うシステムは、サーバ上でコストの高い画像認識を行うためユーザ数の増加に対応できない。そこで、複数コアあるスマートフォンの計算資源のみを活用し、リアルタイムにかつ高精度に画像認識を行う手法が望まれている。さらに、スマートフォン上で画像認識を行うことにより、ネットワーク環境に依存しない、通信コストがかからないという利点がある。

また、近年、高速な線形識別器に適した画像表現が提案されている。従来の一般物体認識は識別に非線形識別器を用いることが一般的であり、コストが非常に大きかった。だが、画像表現を改良することにより、従来の画像表現に非線形識別器を用いるよりも大幅に性能が向上することが示されてる。特に、Fisher Vector が認識性能が高い結果になっている [1]。

しかし、スマートフォン上で Fisher Vector を用い、高速かつ高精度に物体認識をするシステムは存在しない。そこで、複数コアを活用しスマートフォンの計算資源のみを用い、リアルタイムかつ高精度に認識を行う手法を提案する。

従来のスマートフォン上で認識するシステムよりも大幅に性能が高く、さらに高速であることを実験により示す。また、コストの非常に大きい従来のシステムと比較し本手法の有用性を示す。

2. 関連研究

スマートフォンから利用できる画像認識アプリケーションを紹介する。Google Goggles[*1] は物体認識システムとして有名なアプリケーションである。しかし、認識対象は視覚的変化のない特定物体であり、本研究では、一般物体を認識対象にしている。Maruyama ら [4] は、特徴量は 144 個のみ局所色ヒストグラムの BoF 一つであり、直接線形識別器に適用している。そのため、視覚的変化が大きい一般物体認識では認識精度が極めて低い。我々は以前、SURF-BoF と色特徴を用い、χ^2 kernel feature map を適用し、非線形識別器と同等の精度で食事画像認識をするシステムを提案した [2,3]。SURF は強力であるがコストが高いためクアッドコアで並列化している。だが、認識時間のほとんどを SURF の特徴記述とコードワード割り当てに消費し、50 種類の認識で 0.26 秒かかっていた。そのため、同時に認識する対象が多くなるとリアルタイムに認識することができない。また、認識精度についても従来のサーバサイドの認識手法に劣っていたため改善の余地がある。

本研究では、スマートフォン上でリアルタイムかつ高精度に画像認識をする手法として、局所特徴量に計算コストの小さい HoG と色特徴を用い、画像表現に Fisher Vector を用いる。

3. 提案手法

画像特徴量には、色と勾配 2 種類の局所特徴量を用いる。色特徴は、局所パッチを 2×2 に分割したサブ領域から各ピクセルの RGB 値それぞれについて平均と分散とした。ゆえに、24 次元特徴ベクトルとなった。勾配特徴は、HoG を局所パッチから抽出することで 32 次元の特徴ベクトルを得た。HoG は、SIFT と類似しているが、SIFT、SURF よりも高速に抽出することが可能であり、その分特徴点を密にとることが可能である。画像が大きい場合には 3K ピクセルにリサイズし、局所特徴はともに 2 スケール、6 ピクセルごとの dense sampling で抽出した。そして、得られた局所特徴群は、Fisher Vector [6] で表現した。

Fisher Vector は、BoF と異なり局所特徴群の 1 次と 2 次の統計量も含めることにより量子化誤差を軽減している。確率密度関数に GMM を仮定し、局所特徴群をモデル化した。そして、認識性能を高めるために L2 正規化とパワー正規化を適用した [6]。GMM の混合数はともに 32、局所特徴はともに PCA を適用し 24 次元にした。ゆえに、ともに 1536 次元の特徴ベクトルとなった。また、学習画像を増やすために左右反転した画像も学習データに加えた。

分類は、one-vs-rest 線形 SVMs により多クラス分類を

[1] 電気通信大学大学院 情報理工学研究科 総合情報学専攻 〒182–8585 東京都調布市調布ヶ丘 1–5–1
[a)] kawano-y@mm.inf.uec.ac.jp
[b)] yanai@cs.uec.ac.jp
[*1] http://www.google.com/mobile/goggles/

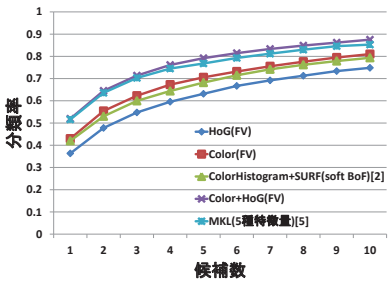

図 1 各手法による分類率

表 1 平均実行時間

	実行時間 [sec]
従来手法 [2]	0.26
提案手法	0.065

行った。

3.1 実装

実装は、クアッドコアデバイスを想定し並列処理を行った。色特徴、HoG それぞれ 2 コアずつ用い並列化した。HoG は最初に勾配方向と勾配強度を算出しておくことで高速に局所特徴を抽出した。GMM により負担率を求める際や、平均や分散に関する勾配を求める際に必要になる項で、オフラインに計算可能な部分はあらかじめ計算しておき、ルックアップテーブルを作成しておいた。また、SVM はオフラインで学習しておいた。そして、認識時に用いる値は全てメモリにロードしておいた。

4. 実験

データセットは我々で構築している各 100 枚以上ある 100 種類、合計 12,905 枚からなる食事画像を使用し評価実験を行った。特徴量を抽出する領域は、人手で与えられた正しい食事領域とした。検証、テストに各 20 枚、残りを学習に使用し評価することを、ランダムにデータを入れ換え 5 回繰り返しその平均値で評価した。次に、画像データが与えられ、全ての識別器の評価値を得るまでの時間 (認識時間) を計測した。今回実験に使用したデバイスは、Galaxy NoteⅡ(1.6GHz 4 コア Android4.1) であり、[2] と同様である。

実験結果を図 1 に示した。提案手法 (Color+HoG (FV)) と、比較として特徴量単体の場合 (Color(FV)、HoG(FV))、以前の手法 (ColorHistogram+SURF(soft BoF))、また、Matsuda ら [5] は複数品の食事の認識が主であるが、単品の食事においても分類率を評価している。そのため、[5] で示されている単品の食事において食事領域が与えられた際の食事 100 種類の分類率を示した (MKL)。[5] では、3 種類の局所特徴と 2 種類の大域特徴を使用し局所特徴は hard BoF で表現し、5 種の特徴量 (合計 46,452 次元) を非線形 χ^2RBF カーネル MKL-SVM により分類している。そのため、非常にコストが高い。実験結果は、候補を 5 つ提示した際、本手法では 79.2%を達成した。色特徴のみの場合では、70.6%、HoG のみの場合では 63.2%、以前の手法では 68.2%、[5] らの手法は 76.8%であった。実験結果より色特徴のみで以前の手法よりも高い分類率となり、食事画像認識において色特徴を Fisher Vector で表現することにより認識性能が大きく向上することがわかった。また組み合わせることで性能が向上し、実験データは異なるがコストが非常に高い [5] の結果と同等以上の性能であった。ゆえに、本手法の有効性が示された。

次に、平均認識時間について実験結果を表 1 に示した。実験結果より、以前の手法が 0.26 秒に対して本手法では 0.065 秒とより高速に認識可能であった。ゆえに、リアルタイム認識に適していることがことが示された。

認識精度と速度の実験により、本手法の有効性が示された。また、極めて高速かつサーバサイドの認識手法 [5] と同等以上の性能を示す結果となりスマートフォン上でのリアルタイム高精度物体認識が可能であることを示した。

5. まとめ

スマートフォン上でリアルタイムかつ高精度に画像認識を行う手法として、局所特徴量に HoG と色特徴を用い、Fisher Vector で表現し、線形識別器で分類する手法を提案した。

100 種類の食事に対して正しい食事領域が与えられた時、候補を 5 つ提示した際に 79.2%の認識精度であった。それは、以前の手法と比較して 11.0%の精度向上である。また、サーバサイドのコストが非常に高い認識手法 [5] と同等以上の性能が示された。そして、認識速度は、0.065 秒であった。それは、従来の 0.26 秒と比較して 75.0%高速化である。実験により本手法の有効性と、スマートフォン上でのリアルタイム高精度物体認識が可能であることを示した。

提案手法は、応用例として我々が開発しているモバイル食事画像認識アプリケーション (http://www.foodcam.jp)、食事画像認識 Twitter Bot(@foodimg_bot) に取り入れられている。

参考文献

[1] Chatfield, K., Lempitsky, V., Vedaldi, A. and Zisserman, A.: The devil is in the details: an evaluation of recent feature encoding methods, *BMVC* (2011).

[2] Kawano, Y. and Yanai, K.: Real-time Mobile Food Recognition System, *CVPR International Workshop on Mobile Vision (IWMV)* (2013).

[3] 河野憲之, 柳井啓司.: 料理画像認識を用いたモバイル食事記録システム. 情報処理学会コンピュータビジョン・イメージメディア研究会, (2013).

[4] Maruyama, T., Kawano, Y. and Yanai, K.: Real-time Mobile Recipe Recommendation System Using Food Ingredient Recognition, *ACM MM Workshop on IMMPD* (2012).

[5] Matsuda, Y., Hoashi, H. and Yanai, K.: Recognition of Multiple-Food Images by Detecting Candidate Regions, *ICME* (2012).

[6] Perronnin, F., Sánchez, J. and Mensink, T.: Improving the Fisher Kernel for Large-Scale Image Classification, *ECCV* (2010).