

動画像を対象とした一般物体認識のための時空間 CoHOG 特徴量の検討

中村 彰吾[†] 出口 大輔[†] 高橋 友和^{††} 井手 一郎[†] 村瀬 洋[†][†] 名古屋大学 大学院情報科学研究科 〒464-8601 愛知県名古屋市千種区不老町^{††} 岐阜聖徳学園大学 経済情報学部 〒500-8288 岐阜県岐阜市中鶯1丁目38番地

E-mail: †snakamura@murase.m.is.nagoya-u.ac.jp, †{ddeguchi,ide,murase}@is.nagoya-u.ac.jp,

††ttakahashi@gifu.shotoku.ac.jp

あらまし Web上の大量の動画像を効率よく検索するための重要な要素技術の一つとして、動画像に映っている物体を認識する技術が挙げられる。従来、静止画像を対象とした一般物体認識は盛んに研究が行われてきたが、動画像を対象とした一般物体認識の研究はほとんど行われていない。動画像を対象とした一般物体認識では、動画像中の様々なフレームから得られる形状特徴と動き特徴の双方を効果的に利用することが重要となる。本報告では、歩行者検出などで高い性能を持つ CoHOG (Co-occurrence Histograms of Oriented Gradients) を時間方向に拡張した時空間 CoHOG 特徴を提案する。時空間 CoHOG 特徴は、動画像中の局所領域での時空間勾配方向の共起ヒストグラムである。認識には、BoF (Bag of Features) 表現とカーネル SVM を用いる。実験では、Web上から収集した10カテゴリ、計1,000本の動画像を用い、オプティカルフロー特徴および SIFT 特徴と比較することにより、時空間 CoHOG 特徴の有効性を確認した。

キーワード 一般物体認識, 動画像, 時空間 CoHOG 特徴, BoF 表現

A Study on Spatio-Temporal CoHOG Features for Recognition of Generic Objects in Video

Shogo NAKAMURA[†], Daisuke DEGUCHI[†], Tomokazu TAKAHASHI^{††},Ichiro IDE[†], and Hiroshi MURASE[†][†] Graduate School of Information Science, Nagoya University

Furo-cho, Chikusa-ku, Nagoya-shi, Aichi, 464-8601 Japan

^{††} Faculty of Economics and Information, Gifu Shotoku Gakuen University

Nakauzura 1-38, Gifu-shi, Gifu, 500-8288 Japan

E-mail: †snakamura@murase.m.is.nagoya-u.ac.jp, †{ddeguchi,ide,murase}@is.nagoya-u.ac.jp,

††ttakahashi@gifu.shotoku.ac.jp

Abstract Recognizing objects in videos is one of the important technologies to search a large amount of videos efficiently on the Web. Recently, generic object recognition has been actively studied for still images, but almost not for videos. As for the generic object recognition in a video, it is important to use both the shape features and the motion features obtained from multiple frames in the video efficiently. In this paper, we propose spatio-temporal CoHOG (Co-occurrence Histograms of Oriented Gradients) features. This is an extension of CoHOG features that provide a high performance for pedestrian detection and others. The spatio-temporal CoHOG features are co-occurrence histograms of oriented spatio-temporal gradients in local regions in a video. In the recognition, a BoF (Bag of Features) representation and a kernel SVM are employed. We conducted an experiment on 1,000 videos including 10 categories collected from the Web. The experimental results showed the effectiveness of the spatio-temporal CoHOG features compared with conventional optical flow features and SIFT features.

Key words Generic object recognition, Video, Spatio-temporal CoHOG features, BoF representation

1. はじめに

近年、Web 上には大量の動画が存在し、それらを効率よく検索する技術が求められている。検索するためのクエリの一つとして、動画に現れる物体が挙げられる。動画中の物体を認識することができれば、ユーザは自分の見たい動画を容易に探し出すことができる。例えば、自動車を検索クエリとすることで、図 1 のような動画を検索できる。一般的に、これらの検索は動画に付随したタグ（テキスト）を用いて行われる。しかしながら、このようなタグはユーザが主観的に付けるものであるため、表記ゆれなどが原因で正しく検索できない場合があり、そもそもタグが付随しないものも多い。そのため、動画中の物体を計算機で認識する技術が必要となる。

Web 上の動画に含まれる物体は様々であるため、特定の物体に依存しない認識手法が必要である。実世界に含まれる物体を計算機が一般的な名称で認識する手法は一般物体認識と呼ばれ、近年盛んに研究が行われている [1]。一般物体認識はカテゴリ内で見た目のバリエーションが大きい物体を扱うため、高い認識率を得ることが難しく、解決すべき課題が多い。

従来、静止画像を対象とした一般物体認識手法に関する研究が盛んに行われている。代表的な手法として、SIFT (Scale-Invariant Feature Transform) [2] を用いたものが挙げられる。SIFT は、画像の回転や照明変化に頑健な局所特徴であることが知られており、抽出した特徴点の周辺領域のエッジやテクスチャなどの形状特徴を 128 次元の特徴ベクトルで記述するものである。これに対して、動画を対象とした一般物体認識においては形状特徴に加えて、動き特徴を用いることが有効である。

Kläser らは、HOG (Histograms of Oriented Gradients) [3] の勾配を時間方向に拡張した特徴を提案している [4]。この特徴は、動画の形状情報だけでなく隣接フレーム間の動き情報も得ることができ、動作認識に有効とされている。一方、複数の特徴量の共起をとることで対象の記述能力を向上させる研究が行われている。それらの研究の一つとして、Watanabe らは勾配の共起を利用した CoHOG (Co-occurrence Histograms of Oriented Gradients) を提案している [5]。画像空間中の様々な位置関係で勾配の共起をとることで複雑な形状が表現可能となり、歩行者検出に有効とされている。

本報告では、これらの特徴の性質を組み合わせた時空間 CoHOG 特徴を提案する。具体的には、従来の CoHOG における勾配と共起を時間方向に拡張する。そうすることで、形状と動きに対して高い記述力を得る。

以降、2 節で時空間 CoHOG 特徴とそれを用いた認識手法について述べる。そして、3 節で実験の方法と結果を述べる。4 節では、3 節で述べた実験の考察に加えて、時空間 CoHOG 特徴と他の特徴を統合した場合の効果について調査した結果を述べる。最後に、5 節で本報告をむすぶ。

2. 時空間 CoHOG 特徴を用いた認識手法

2.1 時空間 CoHOG 特徴を用いた認識手法の概要

図 2 に、手法の流れを示す。まず、入力動画から単位区



図 1 Web 上の動画

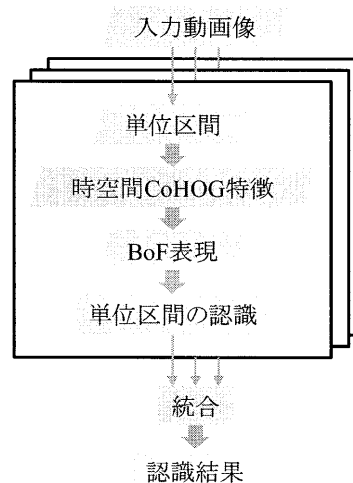


図 2 時空間 CoHOG 特徴を用いた認識手法の流れ

間を抽出し、そこから時空間 CoHOG 特徴を抽出する。単位区間は、連続した複数フレームからなり、認識を行う単位である。そして、単位区間から抽出された特徴を BoF (Bag of Features) [6] 表現し、カテゴリ毎に用意した 2 クラス識別器を用いて対象カテゴリに属する信頼度を算出する。最後に、全単位区間の信頼度から、入力動画の対象カテゴリに属する信頼度を決定する。以降で、各段階について詳しく述べる。

2.2 時空間 CoHOG 特徴の抽出

単位区間から一定間隔で時空間局所領域（ブロック）をサンプリングし、ブロック毎に時空間 CoHOG 特徴を抽出する。図 3 に、ブロックのサンプリングの様子を示す。以下で時空間 CoHOG 特徴量の計算方法について述べる。

まず、ブロック内の全画素で時空間勾配 $g(x, y, t)$ を計算する。 $I(x, y, t)$ を画素の輝度として、 $g(x, y, t)$ の各要素を次のように求める。

$$\begin{cases} g_x(x, y, t) = I(x+1, y, t) - I(x-1, y, t) \\ g_y(x, y, t) = I(x, y+1, t) - I(x, y-1, t) \\ g_t(x, y, t) = I(x, y, t+1) - I(x, y, t-1) \end{cases} \quad (1)$$

次に、ブロックを複数の領域（セル）に分割し、各セルで平均勾配 $\bar{g}(x, y, t)$ を計算する。そして、各平均勾配 $\bar{g}(x, y, t)$ を正 20

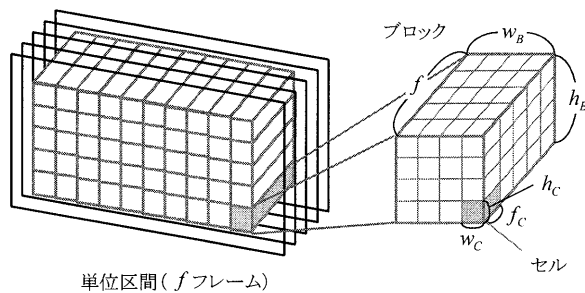


図 3 単位区間からのブロックのサンプリング

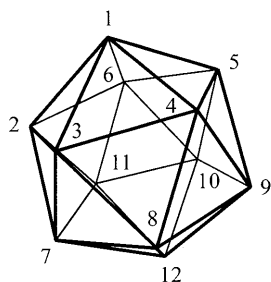


図 4 正 20 面体 (図の数字は量子化方向を表す)

面体の頂点である 12 方向に量子化する。正 20 面体の頂点の座標は、 $\varphi = \frac{1+\sqrt{5}}{2}$ として $(\pm 1, \pm\varphi, 0)$, $(0, \pm 1, \pm\varphi)$, $(\pm\varphi, 0, \pm 1)$ である。図 4 に、正 20 面体を示す。

ブロック内の各セルで計算した勾配方向を利用して、勾配方向の共起ヒストグラムを作成する。これが時空間 CoHOG 特徴である。以下で、図 5 に沿って作成の流れを説明する。

(a) まず、基準セルとの位置関係により、63 種類のペアを形成する (ただし、基準セル自身も含む)。なお、手前側にペアを形成しないのは、対称性による冗長性を排除するためである。

(b) 各ペアの位置関係にあるセルの組み合わせをブロック内から列挙し、その勾配方向のペアをヒストグラムに投票していく。

(c) 各ペアのヒストグラムを連結することで最終的なヒストグラムとする。ただし、基準セル同士のペアはそのセル自身の勾配方向のみでヒストグラムを作成する。また、勾配強度が 0 に等しいセルは利用しない。

時空間 CoHOG 特徴量は $(12 \times 12 \times 62) + 12 = 8,940$ 次元となり、非常に次元が大きくなる。そのため、PCA (Principal Component Analysis) により d 次元に圧縮する。

2.3 時空間 CoHOG 特徴の BoF 表現

認識を行う前処理として、単位区間の全ブロックで作成した時空間 CoHOG 特徴を BoF 表現する。BoF 表現は画像を局所特徴量のヒストグラムで表現する手法であり、一般物体認識の分野で広く用いられている。この手法は学習段階と認識段階に分かれており、一般に以下の手順で行う。

学習段階では、まず各学習用画像から複数の局所特徴量を抽出する。そして、全学習用画像の全局特徴量を k -means クラスタリングすることにより、visual word を生成する。visual word は、局所特徴量を表す特徴ベクトルをベクトル量子化したものである。各局所特徴量を最も類似する visual word とし

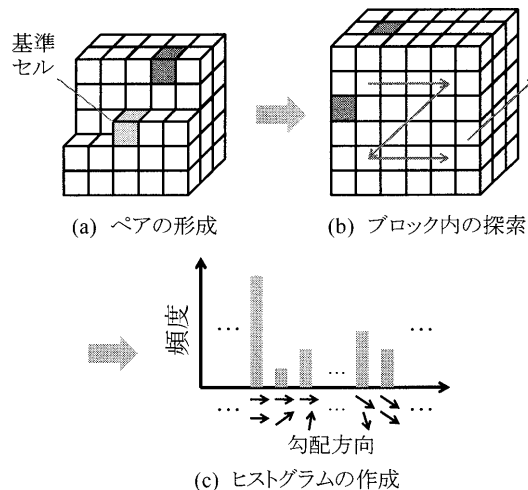


図 5 時空間 CoHOG 特徴の抽出

て表現することにより、各学習用画像は visual word の出現頻度のヒストグラムで記述される。認識段階では、まず学習段階と同様に、各入力用画像から局所特徴量を抽出する。そして、学習段階で生成された visual word を用いて、各入力用画像を visual word の出現頻度のヒストグラムで表現する。

2.4 識別器の学習

学習動画の単位区間で作成した時空間 CoHOG 特徴の BoF 表現を用いて、識別器を学習する。なお、識別器はカテゴリ毎に用意しておく。識別器には、カーネル SVM (Support Vector Machine) [7] を用いる。カーネル SVM は高い認識性能を持っていることが知られ、様々な画像認識問題に応用されている。本研究では、カーネル関数として χ^2 カーネルを用いる。 χ^2 カーネルは以下の式で表される。

$$k(\mathbf{x}, \mathbf{y}) = \exp\left(-\gamma \sum_i \frac{(x_i - y_i)^2}{x_i + y_i}\right) \quad (2)$$

χ^2 カーネルは画像分類において最も性能が良いカーネルであると報告されている [8]。 γ には、全学習ベクトルの組み合わせにおける平均 χ^2 距離の逆数を設定する。

2.5 単位区間の認識

事前に学習しておいた識別器を用いて、単位区間の認識を行う。識別器の出力結果としては、分離超平面からの符号付き距離が得られる。この値を対象カテゴリに属する信頼度とする。

2.6 統合

単位区間 t の信頼度 c_t を用いて、入力動画の認識を行う。入力動画の単位区間数を T とすると、最終的な信頼度 \hat{c} は次式によって決定する。

$$\hat{c} = \frac{1}{T} \sum_{t=1}^T c_t \quad (3)$$

動画に映っている物体を認識したい場合、 $\hat{c} > 0$ となるカテゴリを認識結果とする。ただし、映っている物体が 1 種類という条件ならば、 \hat{c} が最大のカテゴリを認識結果とする。また、クエリを与えて動画を検索したい場合は、 \hat{c} をそのまま利用する。

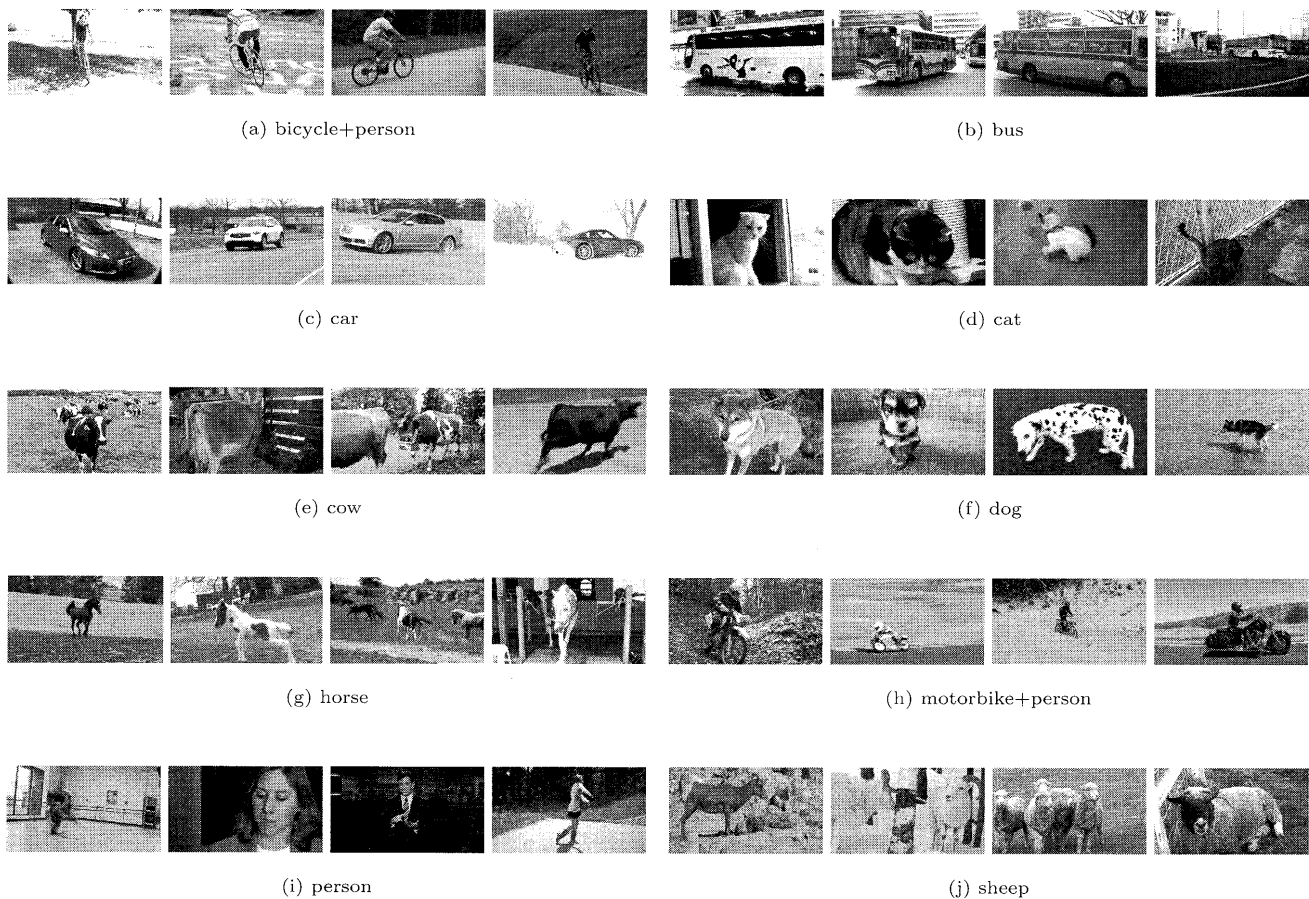


図 6 実験で使った動画の例

3. 実 験

3.1 実験データセット

データセットは、YouTube [9] から動画を収集して構築した。動画中の一般物体のカテゴリは、PASCAL Visual Object Classes Challenge 2006 [10] で使用されたデータセットと同一の 10 カテゴリとした。具体的には、bicycle, bus, car, cat, cow, dog, horse, motorbike, person, sheep である。図 6 にデータセットの一部を示す。データセット中の動画は対象物体が映るように切り出した。しかし、オクルージョンや画面外へのはみ出しで物体の一部が映っていないフレームが存在した。また、異なるカテゴリの物体が同時に映っている動画も収集した。例えば、bicycle, motorbike が含まれる動画については、すべて person も同時に映っているものであった。そのため、これらの動画を person の識別器に学習させる際は、すべてポジティブデータとして扱った。

データセットの動画像数は 1,000 本であり、各カテゴリ 100 本 (person は bicycle, motorbike と同時に映っている動画像も含めると 300 本) とした。また、動画像のフレームサイズは $1,280 \times 720 \sim 1,920 \times 1,080$ [pixels], フレーム数は 45~270 であった。

3.2 実験方法

提案する時空間 CoHOG 特徴を用いた手法と、オプティカル

フロー特徴、SIFT 特徴を用いる以下の 2 つの手法を比較した。

比較手法 1: 認識手法は、2.2 節をオプティカルフロー特徴に置き換え、2.3 節の BoF 表現を行わないものとした。オプティカルフロー特徴としては、隣接フレーム間のオプティカルフローを用いた。フローの抽出にはブロックマッチング法を利用した。各フレームの各ブロックに対してフローを計算し、単位区間でフローの方向ヒストグラムを作成した。

比較手法 2: 認識手法は、2.2 節を SIFT 特徴に置き換えたものとした。SIFT 特徴点の検出には、DoG (Difference of Gaussian) を用いた sparse な手法とグリッドサンプリングやランダムサンプリングを用いた dense な手法があるが、本研究では一般物体認識でより有効とされる後者の dense な手法を用いた。スケール変化に頑健な特徴とするため、特徴点検出は複数スケールのグリッドサンプリングにより行った。

各手法を用いる際のパラメータは、次のように設定した。時空間 CoHOG 特徴については、単位区間のフレーム数を $f = 5$ 、ブロックサイズを $w_B = h_B = 80$ 、セルサイズを $w_C = h_C = 5$ 、 $f_C = 1$ 、PCA 適用後の次元数を $d = 500$ とした。オプティカルフロー特徴については、ブロックサイズを 20×20 、方向ヒストグラムのビン数を 36 とした。SIFT 特徴については、サンプリング間隔を 80、スケールを 8, 16, 24, 32 の 4 種類とした。時空間 CoHOG 特徴と SIFT 特徴の BoF における visual word の数は 500 とした。

表 1 各特徴を用いた手法の Mean AP

特徴	Mean AP
時空間 CoHOG 特徴	0.454
オプティカルフロー特徴	0.394
SIFT 特徴	0.414

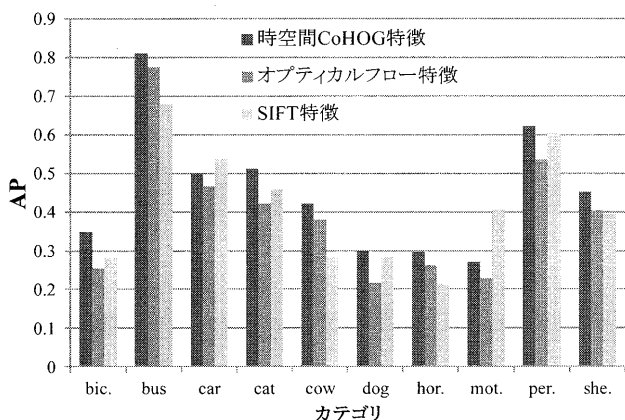


図 7 各特徴を用いた手法のカテゴリ毎の AP

3.3 評価方法

評価には、2-fold cross validation による Mean AP (Average Precision) を用いた。Mean AP はカテゴリ毎に算出した AP の平均値である。AP は検索結果を評価する際に用いられる評価基準であり、以下のように求める。まず、入力動画像群を対象カテゴリに属する信頼度が高い順に並び替える。そして、ランキング上位から順に調べていき、正解であればその時点での適合率を計算する。すべての正解データが現れたら、次式で AP を計算する。

$$AP = \frac{1}{V} \sum_{v=1}^V Pr(v) \quad (4)$$

$Pr(v)$ は対象カテゴリに属する動画像 v が現れたときの適合率、 V は対象カテゴリに属する動画像数である。本研究の場合、person では $V = 150$ 、それ以外のカテゴリでは $V = 50$ となる。AP は Precision-Recall 曲線と座標軸に囲まれた領域の面積に相当する。

3.4 実験結果

表 1 に、各手法の Mean AP を示す。この表から、オプティカルフロー特徴、SIFT 特徴を用いるよりも時空間 CoHOG 特徴を用いるほうが精度が高いことがわかる。このことから、時空間 CoHOG 特徴の有効性を確認した。

4. 考 察

4.1 特徴の比較に関する考察

図 7 に、各特徴を用いた手法のカテゴリ毎の AP を示す。この図から、ほとんどのカテゴリで時空間 CoHOG 特徴の精度が最も高いことがわかる。表 2 に、cat のランキング上位 10 位までの結果を示す。この表は、cat の識別器に全カテゴリの動画像を入力した際、信頼度が上位 10 位以内であった動画像の実際のカテゴリを表している。

表 2 cat のランキング

(a) 時空間 CoHOG 特徴
(AP = 0.553)

順位	カテゴリ
1	cat
2	cat
3	cat
4	person
5	cat
6	cat
7	cat
8	cat
9	cat
10	cat

(b) オプティカルフロー特徴
(AP = 0.400)

順位	カテゴリ
1	car
2	dog
3	cat
4	cat
5	person
6	cat
7	cat
8	dog
9	cat
10	dog

表 3 motorbike のランキング

(a) 時空間 CoHOG 特徴
(AP = 0.275)

順位	カテゴリ
1	motorbike
2	person
3	motorbike
4	car
5	horse
6	car
7	horse
8	person
9	horse
10	car

(b) SIFT 特徴
(AP = 0.439)

順位	カテゴリ
1	motorbike
2	motorbike
3	motorbike
4	motorbike
5	motorbike
6	motorbike
7	horse
8	bicycle
9	sheep
10	dog

car と motorbike については SIFT 特徴に及ばなかった。これらのカテゴリについては、オプティカルフロー特徴の精度も低いことから、動きが形状に比べて有効ではなかったと考えられる。表 3 に、motorbike のランキング上位 10 位までの結果を示す。時空間 CoHOG 特徴のランキングでは person, car, horse の動画像が上位に入っていた。これらの動画像を実際に見てみると、直線的に走行していない motorbike や、踊っている person、浅瀬で水しぶきを上げながら走る horse などが映っていた。したがって、特殊な動きを伴う動画像が高い信頼度を得た可能性が高く、直線的に走行する motorbike の動画像を検索できなかったと考えられる。現段階では、時間方向の共起を 2 フレーム先までしか用いていないため、さらに離れた時間間隔でペアをとることでより良い動き特徴が得られる可能性もある。

4.2 特徴の統合に関する考察

本節では参考として、2 節で提案した時空間 CoHOG 特徴、および 3 節で比較手法として取り上げた SIFT 特徴・オプティカルフロー特徴を統合した場合の効果について調査した結果を述べる。

4.2.1 特徴の統合手法

特徴の統合には、MKL (Multiple Kernel Learning) [11] を用いる。これは、複数の識別器のカーネルを線形結合すること

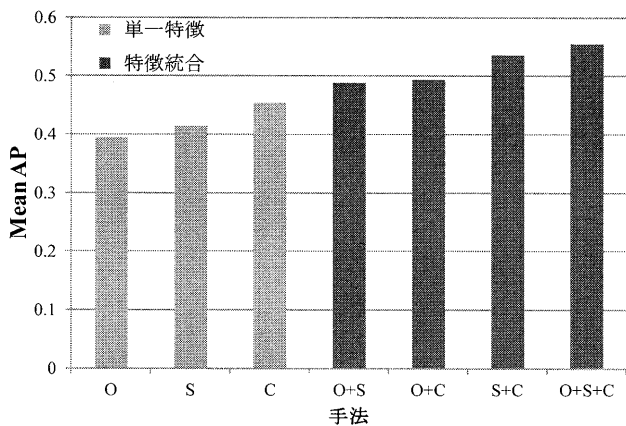


図 8 各特徴を用いた手法および特徴統合による手法の Mean AP (O はオプティカルフロー特徴, S は SIFT 特徴, C は時空間 CoHOG 特徴を表す)

で統合カーネルを作成する手法である。本来 MKL はカーネル選択のために作られたが, Varma らは MKL における各カーネルを特徴に対応付けて統合することで画像認識を行う手法を提案した [12]。そこで, 本節でもカーネル選択ではなく特徴統合のために MKL を用いる。統合カーネル $K(\mathbf{x}, \mathbf{y})$ は特徴 f のカーネル $k_f(\mathbf{x}_f, \mathbf{y}_f)$ から次式のように作成できる。

$$K(\mathbf{x}, \mathbf{y}) = \sum_{f=1}^F \beta_f k_f(\mathbf{x}_f, \mathbf{y}_f) \quad (5)$$

F は特徴数, β_f は特徴 f の重みを表す ($\beta_f \geq 0, \sum_{f=1}^F \beta_f = 1$)。また, 各特徴のカーネル $k_f(\mathbf{x}_f, \mathbf{y}_f)$ には, 2 節同様 χ^2 カーネルを用いる。

4.2.2 特徴の統合実験

特徴統合の有効性を調査するため, 各特徴のみを用いる手法と特徴統合を行う手法を比較した。実験条件は基本的に前節と同じであり, 各特徴 f のカーネル重み β_f は均一重み ($\beta_f = \frac{1}{F}$) とした。

各特徴を用いた手法および特徴統合による手法の Mean AP を図 8 に示す。この図から, 単一特徴を用いるよりも特徴統合を行う方が認識精度が高くなっていることがわかる。このことから, 特徴統合の効果を確認した。また, 特徴統合の中でも 3 種類すべての特徴を統合する手法が最も認識精度が良かった。すなわち, 時空間 CoHOG 特徴は形状と動きの双方を記述するものの, SIFT 特徴とオプティカルフロー特徴には認識に有効な情報が他にあることがわかる。特に, オプティカルフロー特徴よりも SIFT 特徴と統合した方が認識精度が向上することから, 時空間 CoHOG 特徴に形状情報が不足していると考えられる。そのため, 空間方向をより重視した勾配の量子化が必要である。

5. むすび

本報告では, 動画像を対象とした一般物体認識のための時空間 CoHOG 特徴量を提案した。時空間 CoHOG 特徴は CoHOG における勾配と共起を時間方向に拡張したものである。単位区

間から一定間隔で時空間 CoHOG 特徴を抽出し, PCA により次元を圧縮し, BoF 表現することで単位区間を記述した。そして, カーネル SVM による認識を単位区間毎に行い, それらの結果を平均することにより, 入力動画像の認識を行った。実験では, YouTube において収集した 1,000 本の動画像を使用し, 時空間 CoHOG 特徴とオプティカルフロー特徴, SIFT 特徴を比較した。オプティカルフロー特徴は, 各隣接フレーム間のオプティカルフローのヒストグラムを作成することによって抽出した。SIFT 特徴は, 各フレームからグリッドサンプリングにより特徴点を抽出し, BoF 表現することによって抽出した。実験の結果, 時空間 CoHOG 特徴を用いた手法が最も高い認識精度となることを確認した。

今後の課題としては, より離れた時間間隔でのペアの形成を始めとした適切なパラメータ設定や, 空間方向をより重視した勾配の量子化が挙げられる。

謝辞 日頃より熱心に御討論頂く名古屋大学村瀬研究室諸氏に深く感謝する。本研究の一部は, 科学研究費補助金による。また, 本研究では画像処理に MIST ライブラリ (<http://mist.murase.m.is.nagoya-u.ac.jp/>) を使用した。

文 献

- [1] 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, November 2007.
- [2] D.G. Lowe, “Object recognition from local scale-invariant features,” Proc. 7th IEEE Int. Conf. on Computer Vision, pp. 1150-1157, September 1999.
- [3] N. Dalal, B. Triggs, “Histograms of oriented gradients for human detection,” Proc. 2005 IEEE Int. Computer Society Conf. on Computer Vision and Pattern Recognition, vol. 1, pp. 886-893, June 2005.
- [4] A. Kläser, M. Marszałek, and C. Schmid, “A spatio-temporal descriptor based on 3D-gradients,” 19th British Machine Vision Conference, pp. 995-1004, September 2008.
- [5] T. Watanabe, S. Ito, and K. Yokoi, “Co-occurrence histograms of oriented gradients for pedestrian detection,” Proc. 3rd Pacific-Rim Symposium on Image and Video Technology, Lecture Notes in Computer Science, vol. 5414, pp. 37-47, January 2009.
- [6] G. Csurka, C.R. Dance, L. Fan, J. Willamowski, and C. Bray, “Visual categorization with bags of keypoints,” Proc. ECCV 2004 Workshop on Statistical Learning in Computer Vision, pp. 1-22, May 2004.
- [7] V. Vapnik, “Statistical learning theory,” Wiley-Interscience Publication, 1998.
- [8] J. Zhang, M. Marszałek, S. Lazebnik, C. Schmid, “Local features and kernels for classification of texture and object categories: A comprehensive study,” International Journal of Computer Vision, vol. 73, no. 2, pp. 213-238, June 2007.
- [9] YouTube – Broadcast Yourself –, <http://www.youtube.com/>
- [10] The PASCAL Visual Object Classes Challenge 2006, <http://pascallin.ecs.soton.ac.uk/challenges/VOC/voc2006/>
- [11] G.R.G. Lanckriet, N. Cristianini, P. Bartlett, L.E. Ghaoui, and M.I. Jordan, “Learning the kernel matrix with semidefinite programming,” Journal of Machine Learning Research, vol. 5, pp. 27-72, 2004.
- [12] M. Varma, D. Ray, “Learning the discriminative power-invariance trade-off,” Proc. 11th IEEE Int. Conf. on Computer Vision, pp. 1-8, October 2007.