C-03

複数特徴量の重み付け統合による一般物体認識

GENERIC OBJECT RECOGNITION BASED ON WEIGHTED INTEGRATION OF MULTIPLE FEATURES

須賀 晃†　　滝口 哲也‡　　有木 康雄‡
Akira Suga　　Tetsuya Takiguchi　　Yasuo Ariki

１．はじめに

本稿では，複数の特徴量の重み付け統合による一般物体認識手法を提案する．近年，一般物体認識では SIFT 特徴を量子化した Bag-of-Features を用いた手法が注目を浴びている．しかし，物体内部の輝度変化が少ないものや，他カテゴリでも共通する特徴を多く含むものについては，SIFT 特徴だけでは認識率が低いという問題も見られた．本手法では，SIFT 以外にも SIFT では用いていなかった色情報や，大域的形状情報の HOG 特徴を用いて認識を行う．加えて，Saliency Map を用いて物体領域を大まかに抽出し，特徴重要度抽出理論を用いて，各物体の識別的な特徴量を自動的に学習する手法を提案する．実験の結果，従来の Bag-of-Features だけでは認識が難しかったカテゴリに対応し，特徴量の重み付けにより識別に有効な特徴を学習することでより精度を向上させることができた．

２．特徴量

本章では，本手法で用いる特徴量について述べる．

2.1　局所特徴量 (Bag-of-Features)

Bag-of-Features[1]では，図１に示すように，まず各学習画像から SIFT 特徴[2]を抽出する．SIFT は特徴点の検出と記述を同時に行うアルゴリズムで，スケールの異なるガウシアンフィルタにより得られた平滑化画像の差分画像から極値を検出し，その周辺領域を４×４ブロックに分割し，ブロックごとに８方向の勾配方向ヒストグラムを作成する．これにより回転・スケール変化にロバストな１２８次元の特徴量が抽出できる．次に，得られた全 SIFT 特徴に対して SIFT 特徴空間上で k-means クラスタリングを行う．このクラスタリングによって得られた各クラスタを visual word とみなし，visual vocabulary を構築する．１枚の画像に対し，visual word のヒストグラムを 1000 次元の特徴ベクトルで表現する．

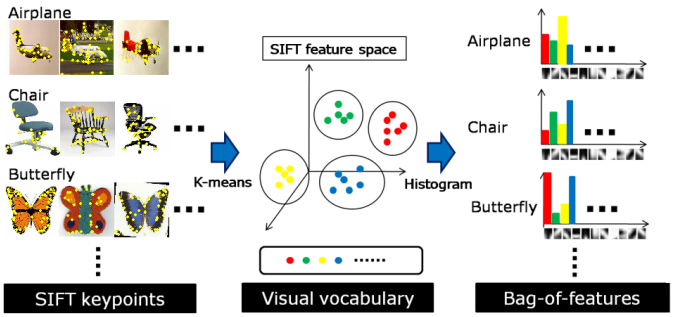

図１．局所特徴量

†神戸大学大学院工学研究科

‡神戸大学自然科学系先端融合研究環

2.2　色特徴量

SIFT 特徴では，グレースケール画像からの輝度変化を用いているが，色情報は用いていない．しかしカテゴリによっては色情報が重要な場合が考えられる．そこで，画像から得られる RGB のヒストグラムを色特徴として用いる．色特徴はまず図２のように画像をブロックに分割する（図では２×２だが実際には４×４に分割している）．各ブロックで色空間を４×４×４＝６４次元のヒストグラムで表現する．それらを統合した１０２４次元の特徴ベクトルで画像を表現する．

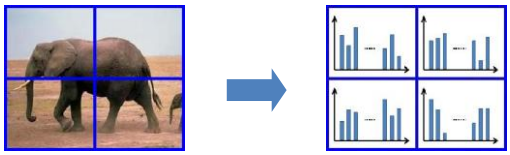

図２．色特徴量

2.3　形状特徴量(Histograms of Oriented Gradients)

SIFT 特徴は局所的な領域での特徴量であり，大域的な情報が用いられていない．そこで大域的な形状を表現できる HOG 特徴を用いる．HOG 特徴[3]は，SIFT のように局所領域の輝度の勾配方向をヒストグラム化した特徴量であるが，SIFT は特徴点に対して特徴量を記述するのに対し，HOG は大域的領域に対して特徴量を記述する．そのため，図３に示すように大まかな物体形状を表現することができる．各ピクセルの勾配強度 $m(x,y)$ と勾配方向 $\theta(x,y)$ は以下の式で求められる．

$$m(x,y)=\sqrt{f_x(x,y)^2+f_y(x,y)^2} \tag{1}$$

$$\theta(x,y)=\tan^{-1}\frac{f_y(x,y)}{f_x(x,y)} \tag{2}$$

$$\begin{cases} f_x(x,y)=L(x+1,y)-L(x-1,y) \\ f_y(x,y)=L(x,y+1)-L(x,y-1) \end{cases} \tag{3}$$

輝度勾配方向は次のように定義される．

$$\tilde{\theta}(x,y)=\begin{cases} \theta(x,y)+\pi, & if \quad \theta(x,y)<0 \\ \theta(x,y) & otherwise \end{cases} \tag{4}$$

算出された輝度勾配画像を，図３に示すようにセルと呼ばれる $c_w \times c_h$ 画素からなる小領域に分割し，それぞれの領域において輝度勾配方向ヒストグラムを作成する．

最後に，$b_w \times b_h$ セルで構成されるブロックと呼ばれる領域を設定する．1 セルずつオーバラップさせながら正規化する．あるブロックの特徴ベクトルを v，ブロック内での位置 $(i, j), \{1 \le i \le b_w, 1 \le j \le b_h\}$ にあるセルのヒストグラムを h_{ij} としたとき，次式により正規化を行う．

$$h'_{ij} = \frac{h_{ij}}{\sqrt{\|v\|_2^2 + \varepsilon}} \quad (\varepsilon = 1) \tag{5}$$

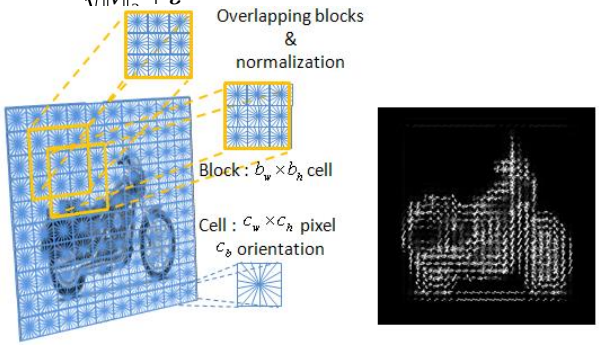

図３．形状特徴量

3． Saliency Map による物体領域の抽出

各物体の識別的な特徴量を抽出する上で，背景から得られる特徴はその精度に影響を与えてしまう．そこで本手法では Saliency Map を用いて物体領域を大まかに特定し，Saliency 値に基づきその領域の特徴に重みをおいて特徴を抽出することで精度を高める．Saliency Map とは，画像中の視覚的注意を引く領域を抽出する手法であり，一般的に画像中の物体は背景に比べて顕著性が高いと考えられる．本手法では Map を作る際に解像度を落とさない Achanta らの手法[4]を用いた．図４にその流れを示す．

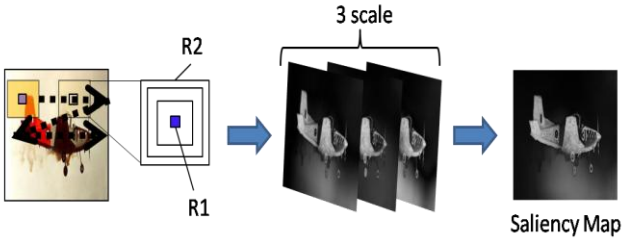

図４．Saliency Map

まず，ピクセルサイズの内側領域 R1 と，異なるスケールの外側領域 R2 からなるフィルタを用意する．座標 (i, j) における Saliency の値は R1 と R2 領域内の平均特徴ベクトルの差から次式のように算出する．

$$c_{i,j} = D\left[\left(\frac{1}{N_1}\sum_{p=1}^{N_1} v_p\right), \left(\frac{1}{N_2}\sum_{q=1}^{N_2} v_q\right)\right] \tag{6}$$

ここで，N_1，N_2 はそれぞれ R1，R2 領域内のピクセル数であり，v_p，v_q はそれぞれ R1，R2 領域内の各ピクセルの L*a*b 色空間における特徴ベクトルである．異なるスケールサイズのフィルタによって得られた各 Map を足し合わせたものを最終的な Saliency Map とする．

4. 特徴重要度による重み付け

本章では，各カテゴリにおいて，認識を行う上でより識別的な特徴量を学習するための，特徴重要度の算出法について述べる．

4.1 TF-IDF

tf-idf の理論[5]を用いて，tf(特徴量の出現頻度) と idf(逆出現頻度) の 2 つの指標から，そのカテゴリを認識する上でより識別的な特徴量に対し大きな重みを与える．カテゴリ c における特徴量 f の重要度 $tfidf_f^c$ は，次の式で求められる．

$$tfidf_f^c = tf_f^c \cdot idf_f^c \tag{7}$$

$$tf_f^c = \frac{\sum_{n=1}^{N_c} HI_f^c(n)}{N_c} = \frac{\sum_{n=1}^{N_c}\sum_{i=1}^{DIM_f} \min\left(h_i(n), h_i(\bar{n})\right)}{N_c \cdot DIM_f} \tag{8}$$

$$idf_f^c = \log\frac{D}{d_f^c} \tag{9}$$

ここで，N_c はカテゴリ c 中の学習画像枚数，DIM_f は特徴量 f の次元数である．$h_i(n)$ はカテゴリ c の画像 n，特徴量 f のヒストグラムにおける i 番目の次元の値，$h_i(\bar{n})$ は画像 n 以外に対する特徴量 f のヒストグラムの値であり，また，D は画像データ総数，d_f^c は特徴量 f のヒストグラムインターセクションがカテゴリ c 内の平均ヒストグラムインターセクション値 θ 以上の画像枚数である．そのカテゴリ内でよく類似しているほど tf_f^c は大きな値となる．また，そのカテゴリと他のカテゴリとの類似性が低いほど idf_f^c は大きな値となる．この重み付けにより，識別性の高い特徴量ほど大きな重みとなる．

4.2 相互情報量

相互情報量[6]は，2 つの要素が共起して現れやすい度合いを統計的に数値化したものである．特徴量 f とカテゴリ c の共起度の尺度として，この相互情報量 $I(f,c)$ を用いる．本稿では，$I(f,c)$ を以下のように定義する．

$$I(f,c) = H(f) + H(c) - H(f,c) \tag{10}$$

$$= -\log P(f) - \log P(c) + \log P(f,c) \tag{11}$$

$$= \log\frac{P(f,c)}{P(f)P(c)} \tag{12}$$

$$P(f) = \frac{\sum_{c=1}^{K}\left[\frac{\sum_{n=1}^{N_c}\sum_{i=1}^{DIM_f} \min\left(h_i(n), h_i(\bar{n})\right)}{N_c \cdot DIM_f}\right]}{\sum_{f=1}^{L}\sum_{c=1}^{K}\left[\frac{\sum_{n=1}^{N_c}\sum_{i=1}^{DIM_f} \min\left(h_i(n), h_i(\bar{n})\right)}{N_c \cdot DIM_f}\right]} \tag{13}$$

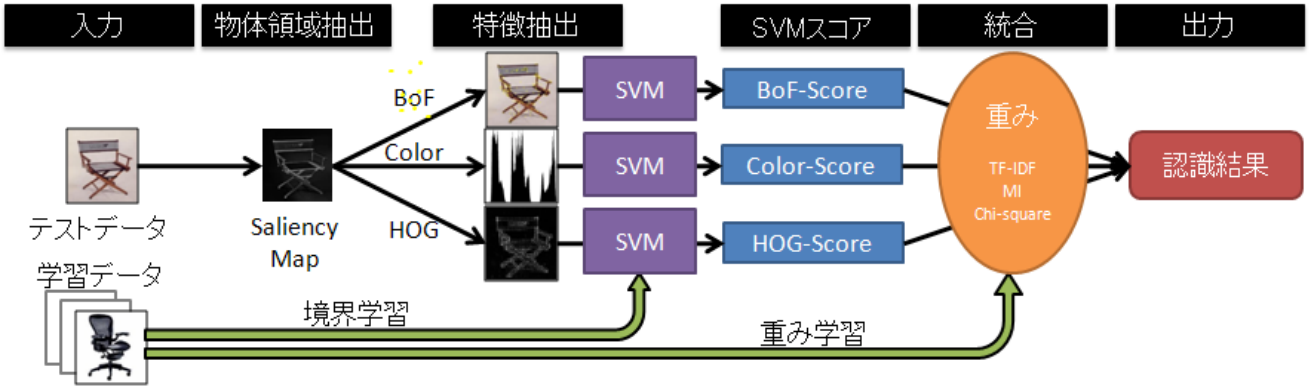

図５．認識の流れ

$$P(c)=\frac{\sum_{f=1}^{L}\left[\frac{\sum_{n=1}^{N_c}\sum_{i=1}^{DIM_f}\min\left(h_i(n),h_i(\bar{n})\right)}{N_c\cdot DIM_f}\right]}{\sum_{f=1}^{L}\sum_{c=1}^{K}\left[\frac{\sum_{n=1}^{N_c}\sum_{i=1}^{DIM_f}\min\left(h_i(n),h_i(\bar{n})\right)}{N_c\cdot DIM_f}\right]} \quad (14)$$

$$P(f,c)=\frac{\frac{\sum_{n=1}^{N_c}\sum_{i=1}^{DIM_f}\min\left(h_i(n),h_i(\bar{n})\right)}{N_c\cdot DIM_f}}{\sum_{f=1}^{L}\sum_{c=1}^{K}\left[\frac{\sum_{n=1}^{N_c}\sum_{i=1}^{DIM_f}\min\left(h_i(n),h_i(\bar{n})\right)}{N_c\cdot DIM_f}\right]} \quad (15)$$

ここで，Kはカテゴリの数，Lは特徴量の数である．以上のように，特徴量 f とカテゴリ c の相互情報量を用いて重み付けを行う．

4.3 χ二乗値

２組の定性的変数間の関連の強さの有無を統計的に検討するための手法として χ 二乗値による手法[7]がある．本稿では，表１のような２×２分割表を各特徴量，カテゴリ毎に作成し，特徴量とカテゴリ間の関連性を求める．

表１．分割表

	カテゴリ	！カテゴリ
特徴	O_{fc}	$O_{f\bar{c}}$
！特徴	$O_{\bar{f}c}$	$O_{\bar{f}\bar{c}}$

２×２分割表の場合の χ^2 値は，以下の式により求められる．

$$\chi^2=\frac{N\left(O_{fc}O_{\bar{f}\bar{c}}-O_{f\bar{c}}O_{\bar{f}c}\right)^2}{\left(O_{fc}+O_{f\bar{c}}\right)\left(O_{fc}+O_{\bar{f}c}\right)\left(O_{f\bar{c}}+O_{\bar{f}\bar{c}}\right)\left(O_{\bar{f}c}+O_{\bar{f}\bar{c}}\right)} \quad (16)$$

ここで，O_{fc} はカテゴリ c における特徴量 f の平均ヒストグラムインターセクションを表し，$\bar{f}$ は f 以外の特徴量，$\bar{c}$ は c 以外のカテゴリを表している．N はその総数である．特徴量 f とカテゴリ c の関連の強さを表す χ^2 値を重みとして用いることで，特徴量とカテゴリの関連度が強いほど大きな重みとなる．

5. 認識

本章では，前章で述べた特徴重要度抽出法により得られた重みを用いて認識結果を統合する手法について述べる．認識の一連の流れを図５に示す．まず，学習データから Bag-of-Features, Color, HOG の各特徴を抽出し，特徴ごとに SVM(Support Vector Machine)を用意し境界面を学習する．同時に，tf-idf，相互情報量，χ 二乗値により各特徴の重みを算出しておく．入力画像が与えられると，同様にまず各特徴を抽出する．そして特徴量 f の SVM によって算出されるカテゴリ c の SVM スコア S_f^c を算出する．算出されたスコア S_f^c に対し，特徴重要度抽出により得られた重みをそれぞれ以下の式のように統合し，最終的に最も高い値が得られたカテゴリを認識結果 c' として出力する．

$$c'=\arg\max_{c}\sum_{f=1}^{L}W_f^c\cdot S_f^c \quad (17)$$

ここで，W は各手法による重みを表している．

6. 実験

6.1 実験条件

実験は Caltech101 データベース[8]を用いた．学習データは１カテゴリあたり 20 枚で，それぞれ Bag-of-features, Color, HOG の各特徴量を抽出し，各カテゴリの画像を正データ，他カテゴリのデータを負データとして SVM により境界を学習する．テストデータは 1 カテゴリあたり 15 枚用意し，101 個のカテゴリに対して認識実験を行った．実験では，まず各特徴量単独での認識実験を行い，次に３特徴を線形的に統合したものと，提案手法である３特徴の重み付け統合 による結果の比較を行った．また，

Saliency Map により物体領域抽出を行った場合と行わなかった場合の精度について比較を行った．

6.2 実験結果

各特徴量単独での認識結果を表２に示す．

表２．各特徴量による認識率 (%)

特徴量	BoF	Color	HOG
認識率	45.2	37.2	53.0

単独特徴量での認識実験では，HOG 特徴による結果が最も高い結果となった．HOG は大まかな形状を表現できるため，カテゴリ内のアピアランス変化の大きい Caltech の画像データに対して有効であったと思われる．

次に，TF-IDF，相互情報量，χ 二乗値を用いてカテゴリごとに識別に有効な特徴量を学習し，重み付けにより認識実験を行った結果を表３に示す．

表３．重み付け統合による認識率 (%)

特徴量	BoF+Color+HOG			
重み	なし	TF-IDF	MI	Chi
認識率	51.3	58.1	57.8	56.6

結果をみると，TF-IDF によって重み付けを行ったものが最も良い結果となった．また，重み付けを行うことにより，重みを付けずに線形的に統合したものよりも全体的に精度が向上していることが確認できた．

また，図６に Saliency Map を用いた効果を示す．Saliency Map を用いた場合の方が全体的に精度が向上していることが分かる．線形統合の場合よりも重み付け統合の場合の方が上がり幅が大きいのは，物体のより正確な重みが得られたためと思われる．

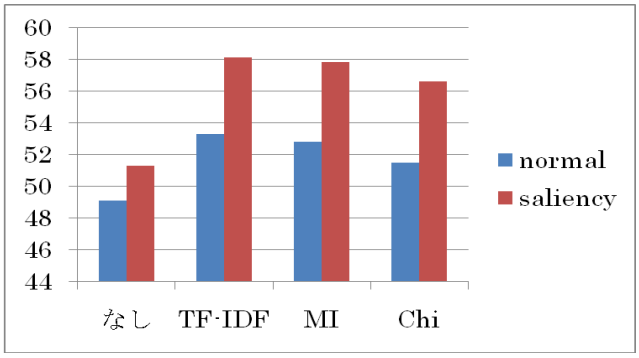

図６．Saliency の効果

最後に，各特徴量において重みの割合が最も大きかったカテゴリを，表４に重み付け手法ごとに示す．またそのカテゴリの画像データのサンプルを図７に示す．

表４．重みが大きかったカテゴリ

	TF-IDF	MI	Chi
BoF	Joshua_tree	Yin_yang	Windsor_chair
Color	Strawberry	Sunflower	Sunflower
HOG	Motorbikes	Motorbikes	Dragonfly

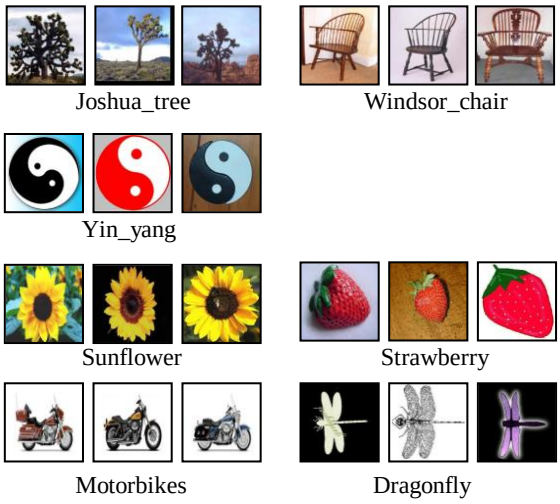

図７．画像データの例

7． おわりに

本稿では，複数特徴量を用いて，特徴重要度抽出によりカテゴリごとに重要特徴に重みを付けて統合する認識手法を提案した．実験では Caltech101 データベースを用いて認識精度の比較を行い，複数特徴量を線形的に統合したものよりも高い結果が得られた．また，Saliency Map を導入し，大まかに物体領域を重視して特徴を抽出することで，より正確な特徴量の重みが得られ精度が向上した．今後は，使用する特徴を増やし，他の重み学習法との比較を行っていく予定である．

参考文献

[1] G. Csurka, “Visual categorization with bags of keypoints,” Proc. of ECCV Workshop on Statistical Learning in Computer Vision, 1-22, 2004.

[2] D.G. Lowe, “Distinctive image features from scaleinvariant keypoints,” Journal of Computer Vision, vol.60, 2, 91-110, 2004.

[3] N. Dalal, B. Triggs, “Histograms of oriented gradients for human detection,” Proc. of IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 886-893, 2005.

[4] R.Achanta, “Salient Region Detection and Segmentation” *6th International Conference on Computer Vision Systems*, 66-75, 2008.

[5] G. Salton, C. Buckley, “Term-weighting approaches in automatic text retrieval,”, Information Processing \& Management, 24, 5, 513-523, 1988.

[6] H. C. Peng, F. Long, C. Ding, “Feature selection based on mutual information: criteria of max-dependency, max-relevance, and min-redundancy,” IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 27, No. 8, pp.1226-1238, 2005.

[7] C. D. Manning, H. Schutze, “Foundations of Statistical Natural Language Processing,” MIT Press, 1999.

[8] Caltech 101, http://www.vision.caltech.edu/Image_Datasets/Caltech101/