領域分割に基づくSIFT特徴を用いた物体識別

永橋知行*，藤吉弘亘（中部大学），金出武雄（カーネギーメロン大学）

Object Type Classification Using Structure-Based Feature Representation

Tomoyuki Nagahashi*, Hironobu Fujiyoshi (Chubu University)
Takeo Kanade (Carnegie Mellon University)

Abstract

Current feature-based object type classification methods information of texture and shape based information derived from image patches. Generally, input features, such as the aspect ratio, are derived from rough characteristics of the entire object. However, we derive input features from a parts-based representation of the object. We propose a method to distinguish object types using structure-based features described by a Gaussian mixture model. This approach uses Gaussian fitting onto foreground pixels detected by background subtraction to segment an image patch into several sub-regions, each of which is related to a physical part of the object. The object is modeled as a graph, where the nodes contain SIFT(Scale Invariant Feature Transform) information obtained from the corresponding segmented regions, and the edges contain information on distance between two connected regions. By calculating the distance between the reference and input graphs, we can use a k-NN-based classifier to classify an object as one of the following: single human, human group, bike, or vehicle. We demonstrate that we can obtain higher classification performance when using both conventional and structure-based features together than when using either alone.

キーワード：物体識別, SIFT, 領域分割，混合正規分布
(Object type classification, SIFT, Region segmentation, Gaussian mixture model)

1. はじめに

一般的に移動体の識別を困難にしているのは，照明等の環境変化や物体の見え方の違い等さまざまな不確定な要素である。移動体識別に用いるには，このような不確定要素に不変である特徴量を選択する必要があり，これまでに，形状やテクスチャに基づく特徴量が用いられている[(1)~(5)]。

文献 (1) では，人と自動車の 2 クラスの識別に dispersedness と Area という 2 つの特徴量を抽出し，線形判別関数を用いて識別する手法が提案されている。また，文献 (2) では，前述のパラメータに加え，縦横比や，移動体のサイズ変化を考慮するためにカメラのズームパラメータを入力特徴とし，ANN による識別を実現している。しかし，これらの特徴量は形状のみに着目したものであるため，その識別能力には限界がある。文献 (3) では，物体の n 次モーメントなどのモーメント特徴を主体とした 11 次元の形状特徴を入力とし，判別空間中で重み付き K 最近傍法を適用した線形判別分析により，物体を人 2 種，車両 3 種，その他の計 6 クラスへの識別を 91% の精度で実現している。しかし，画像の見えをベースとした特徴量のみを用いるこの手法は，対象物の向きによる変動を吸収することができないため，画像を分割して，見えが同じになるようにしなくてはならないというデメリットがある。次に，テクスチャに着目した特徴量として，画像からエッジのヒストグラムを入力として，SVM により人を検出 (2 クラス識別）する手法が提案されている[(4)]。また，文献 (5) では空間と時間変化に着目し，Haar-like wavelets を用いた AdaBoost による人の検出法を提案している。文献 (4) (5) の手法は，人の検出を対象としているが，用いられている特徴量は自動車等の多クラス識別への応用が可能であると考えられる。これらの特徴量は，検出された領域全体から抽出される大域的な特徴量であるといえる，

一方，文献 (6) では，画像中の物体認識に bag-of-keypoints と呼ばれる局所領域の特徴量を用いた手法が提案されている。bag-of-keypoints では，位置情報を無視して画像を局所特徴の集合として捉えることで，画像のカテゴライズを実現している。通常，bag-of-keypoints では局所特徴間の位置関係は考慮しないが，文献 (7) は平行移動とスケールの変化に影響を受けないように位置情報を考慮する手法を提案しており，認識率を向上させている。しかし，これらの手法は画像のカテゴライズを目的としている為，大まかな種類への分類は可能であるが，人と二輪車のように物体の見え方が酷似しているものに対しての識別は困難であると考えられる．

そこで，本手法では，SIFT(Scale Invariant Feature Transform) 特徴量に基づくベクトル量子化ヒストグラムを領域分割した各領域ごとに抽出し，それらの関係をグラフを用いて表現する。領域分割には，混合正規分布モデルを用いて検出領域を各構造の主領域ごとに分割する手法を用いる。参照グラフとのグラフマッチングにより移動体を人，複数の人，自動車，二輪車のマルチクラスに識別する手法を提案する。

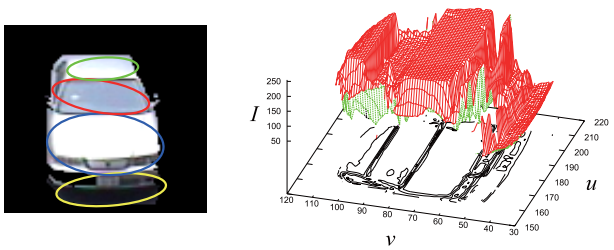

図 1　検出領域に対しての混合正規分布の当てはめ例
Fig. 1.　Example of GMM fitting for detected pixels.

2.　構造に基づく特徴量抽出

本章では，移動体識別における特徴量の抽出法について述べる。まず，背景差分により移動体の検出を行う。検出された物体領域に対して混合正規分布を当てはめ，検出領域を分割する。分割後の各領域に対して SIFT 特徴量を抽出し，ベクトル量子化ヒストグラムを用いて特徴を表現する。以下に，各処理の詳細を示す。

〈2･1〉　混合正規分布を用いた領域分割　物体の内部パターンを記述するために，検出領域に対して混合正規分布を当てはめる [8] [9]。物体領域の座標 (u, v) と輝度 I を $\boldsymbol{x}_i = \{u_i, v_i, I_i\}^T$，混合正規分布パラメータを $\boldsymbol{\Phi} = \{\alpha_j, \boldsymbol{\phi}_j = (\boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j)\}_{j=1}^{c}$ としたとき，$\boldsymbol{x}$ に対して，式 (1) の確定的アニーリング EM(DAEM:Deterministic Annealing EM) アルゴリズム [11] を用いて混合正規分布パラメータ $\boldsymbol{\Phi}_{ML}$ を推定する。

$$\boldsymbol{\Phi}_{ML} = \arg\max_{\boldsymbol{\Phi}} \sum_{j=1}^{c} (\alpha_j \cdot p_j(\boldsymbol{x}|\boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j))^{\beta}$$

$$p(\boldsymbol{x}|\boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j) = \frac{1}{\sqrt{(2\pi)^3 |\boldsymbol{\Sigma}_j|}} \cdot \exp\left\{-\frac{1}{2}(\boldsymbol{x} - \boldsymbol{\mu}_j)^T \boldsymbol{\Sigma}_j^{-1} (\boldsymbol{x} - \boldsymbol{\mu}_j)\right\} \cdots\cdots (1)$$

$p_j(\boldsymbol{x}|\boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j)$ は，平均 $\boldsymbol{\mu}_j$，共分散行列 $\boldsymbol{\Sigma}_j$ であり，$\boldsymbol{\phi}_j = \{\boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j\}$ の各正規分布である。β は DAEM アルゴリズムの温度パラメータである。この β を変化させることにより，EM アルゴリズムの問題点であった初期値への依存性を軽減することができる。また，α_j は混合比で，$\alpha_j > 0$，$\sum_{j=1}^{c} \alpha_j = 1$ を満たす。図 1 に，4 つの正規分布を当てはめ後，$\boldsymbol{\Phi}_{ML}$ で表される 3 次元の混合正規分布を 2 次元画像平面 (u, v) 上に投影した例を示す。各正規分布は，物体の内部パターンとなる主領域（クラスタ）を表しており，これを物体構造の記述に利用する。

推定されたパラメータ $\boldsymbol{\Phi}_{ML}$ から，各ピクセル $\boldsymbol{x}$ がどの正規分布 $\boldsymbol{\phi}_i$ に属しているかを次式により求め，領域分割を行う。

$$C_i = \arg\max_i p_i(\boldsymbol{x}|\boldsymbol{\phi}_i) \cdots\cdots (2)$$

図 2(c) は，異なるクラスに対するセグメンテーション結果である。一般に領域分割手法として用いられている Mean-Shift によるクラスタリング手法 [12] は，図 3 に示すように，同一色で繋がる自動車の側面と背面を同じクラスタに，側面のガラスとボディは別のクラスタに分けられる。また，提案手法では混合正規分布を用いるため，適した分布数を設定することで，微細なテクスチャの変化に影響されず，側面のガラスや車体を同一クラスタに分けられるため，物体構造に基づいた領域分割が可能である。これは，$\boldsymbol{x}_i = \{u_i, v_i, I_i\}^T$ の 3 次元空間でクラスタリングを行うため，側面と背面といった構造ごとへの分割が可能であると考えられる。

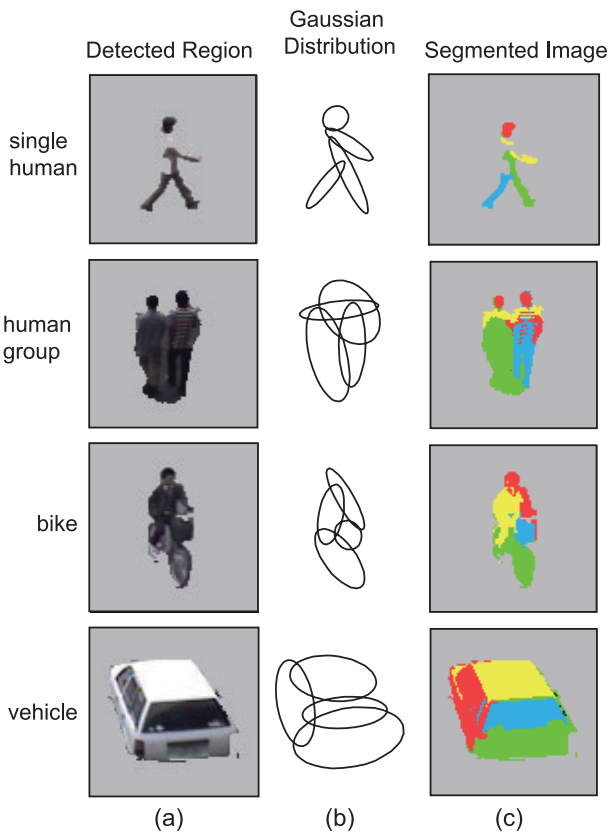

図 2　GMM によるセグメンテーション例
Fig. 2.　Examples of GMM-based segmentation.

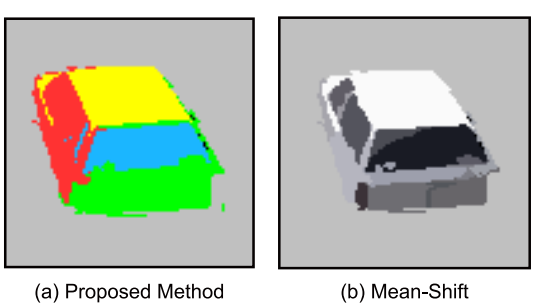

図 3　Mean-Shift セグメンテーションとの違い
Fig. 3.　Segmentation examples of proposed method and Mean-Shift

〈2･2〉　特徴量抽出　領域分割後の各ピクセルに対し

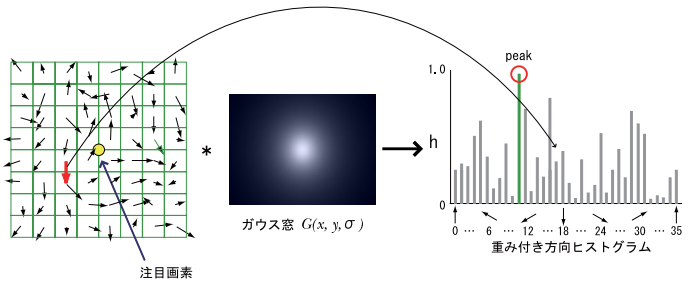

図 4　重み付き方向ヒストグラム
Fig. 4.　Weighted orientation histogram.

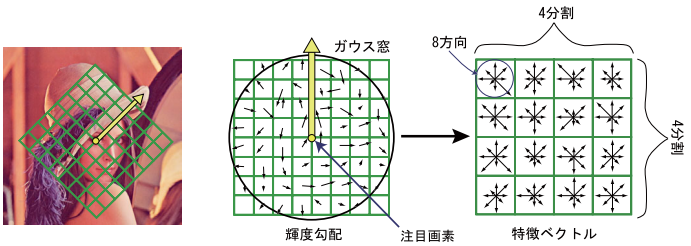

図 5　SIFT 特徴量の記述
Fig. 5.　SIFT feature extraction.

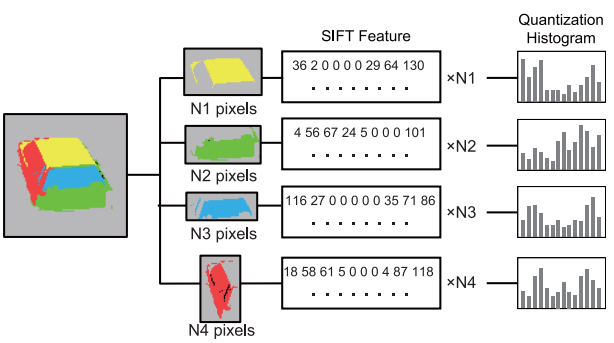

図 6　特徴量抽出
Fig. 6.　Feature extraction.

て，SIFT descriptor [13] に基づく特徴量を抽出する。以下に，各処理の流れを示す。

〈2•2•1〉 SIFT Descriptor　SIFT descriptor は，あるピクセルの代表輝度勾配方向を決定し，その方向を基準とした輝度勾配ヒストグラムを作成し，多次元ベクトルで特徴を記述する。はじめに，注目画素の代表輝度勾配方向を決定する。画像 $L(x, y)$ の輝度勾配方向 $\theta(x, y)$ と大きさ $m(x, y)$ は以下の式により求められる。

$$m(x,y) = \sqrt{f_x(x,y)^2 + f_y(x,y)^2} \quad \cdots\cdots\cdots\cdots (3)$$

$$\theta(x,y) = \tan^{-1}\left(\frac{f_y(x,y)}{f_x(x,y)}\right) \quad \cdots\cdots\cdots\cdots\cdots (4)$$

このとき，

$$f_x(x,y) = L(x+1,y) - L(x-1,y) \cdots\cdots\cdots\cdots (5)$$

$$f_y(x,y) = L(x,y+1) - L(x,y-1) \cdots\cdots\cdots\cdots (6)$$

である。輝度勾配の大きさ m を方向 θ を用いて，各方向のヒストグラムを次式より作成する。

$$w(x,y) = G(x,y,\sigma) \cdot m(x,y) \quad \cdots\cdots\cdots\cdots\cdots (7)$$

$$h_\theta = \sum_x \sum_y w(x,y) \cdot \delta[\theta, \theta(x,y)] \quad \cdots\cdots\cdots (8)$$

$G(x, y, \sigma)$ はガウス分布である。また，θ は全方向を 36 分割したものを使用する。このヒストグラムの最大値の方向をその位置での代表輝度勾配方向とする（図 4）。

この代表輝度勾配方向を基準とした周囲の輝度勾配ヒストグラムを作成する。正規分布から得られる領域を 4x4 の領域に分割し，それぞれの位置で 8 方向の輝度勾配ヒストグラムを作成する。4x4 の領域にそれぞれ 8 方向ヒストグラムを作成するため，128 次元ベクトルの特徴量を持つことになる（図 5）。この 128 次元の SIFT 特徴量を各ピクセルごと抽出する。

〈2•2•2〉 ベクトル量子化ヒストグラム　各領域ごとにベクトル量子化ヒストグラムを作成する。はじめに，ベクトル量子化に用いるコードブックを作成する。コードブックは，参照データから抽出したすべての SIFT 特徴慮を用いる。それらの SIFT 特徴量を LBG アルゴリズムを用いてクラスタリングをし，コードブックを作成する。次に，各領域を構成するピクセルの SIFT 特徴量をコードブックに従い符号化を行う。符号化した各符号を各領域ごとにヒストグラムを作成し，その面積で正規化を行う。この正規化したベクトル量子化ヒストグラムを図 6 に示すように分割した領域の特徴量として利用する。

3.　グラフマッチングによる物体識別

各領域ごとに得られるベクトル量子化ヒストグラムを特徴量をグラフで記述し，グラフマッチングにより物体識別を行う。このとき，物体領域全体から得られる大域的特徴量と，物体領域を混合正規分布によりセグメンテーションし各領域から得られる局所的特徴量のそれぞれのマッチングコストを計算する。この大域的特徴量と局所的特徴量を統合させ，k-NN 法を用いて識別を行う。提案する物体識別の流れを図 7 に示す。

〈3•1〉 グラフによる特徴量の記述　各領域毎に得られたベクトル量子化ヒストグラムをグラフにより表現する。グラフは頂点（ノード）と，頂点を結ぶ辺（エッジ）によって構成される。分割された各領域をノードとし，ベクトル量子化ヒストグラムをノード特徴量として記述する。エッジには，各領域を構成する正規分布の中心のユークリッド距離をエッジ特徴量として記述し，各ノードの位置関係を表現する。図 8 にグラフを記述した例を示す。

〈3•2〉 グラフマッチング　作成したグラフ間のコストをグラフマッチングにより計算する．ノードの集合を $\boldsymbol{N} = \{\boldsymbol{n}_1, \ldots, \boldsymbol{n}_4\}^T$，エッジの集合 $\boldsymbol{E} = \{\boldsymbol{e}_{12}, \ldots, \boldsymbol{e}_{34}\}^T$ とする。参照グラフ $\boldsymbol{T}$ と，入力グラフ $\boldsymbol{X}$ 間のマッチングコストを以下の式より求める。

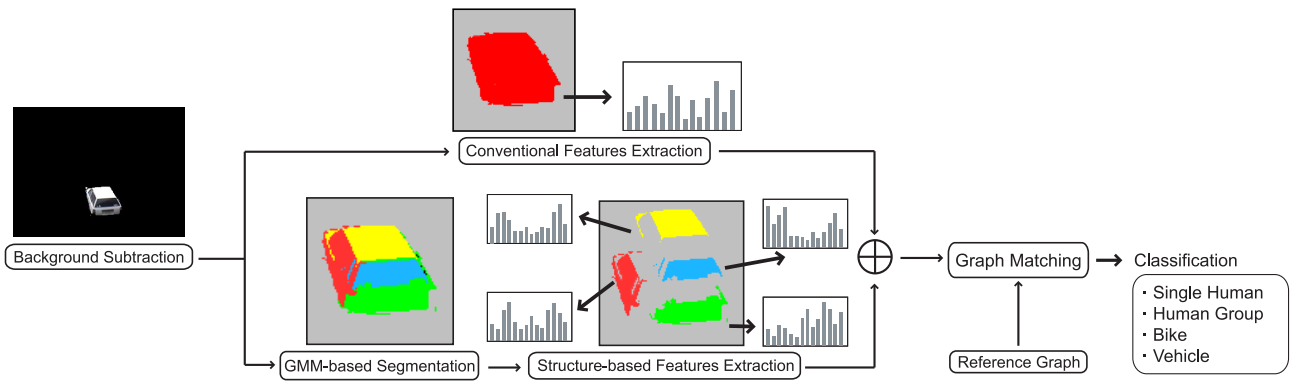

図 7　物体識別の流れ
Fig. 7.　Outline.

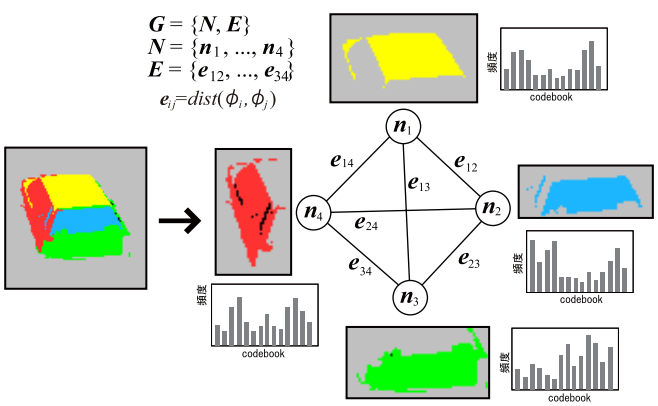

図 8　グラフによる記述
Fig. 8.　Graph description.

$$cost(\boldsymbol{T},\boldsymbol{X}) = \sum_{i=1}^{4} ||\boldsymbol{n}_i^t - \boldsymbol{n}_i^x|| + \sum_{j=1}^{6} ||\boldsymbol{e}_j^t - \boldsymbol{e}_j^x|| \cdots (9)$$

このとき，$\boldsymbol{T}$ と $\boldsymbol{X}$ のノードの対応は未知であるため，$\boldsymbol{T}$ と $\boldsymbol{X}$ のノードの全ての組み合わせについてコストを計算し，最小値を $\boldsymbol{T}$ と $\boldsymbol{X}$ のマッチングコストとする。

$$Cost(\boldsymbol{T},\boldsymbol{X}) = \min_i \{cost(\boldsymbol{T},\boldsymbol{X}_i)\} \cdots\cdots\cdots\cdots (10)$$

分割した各領域から得られる特徴量のコスト $Cost_l$ と領域全体から求める大域的特徴量のコスト $Cost_g$ を統合率 α を用いてマッチングコストを計算する。

$$Cost = \alpha \cdot Cost_l + (1-\alpha) \cdot Cost_g \cdots\cdots\cdots\cdots (11)$$
$$(0 \le \alpha \le 1)$$

この入力パターンと参照パターンとのコストから，kNN 法を用いて識別クラスを判定する。

4.　評価実験

〈4•1〉　実験データ　　構造情報に基づく特徴量の評価実験を行う。識別クラスとして，人 (SH)，人複数 (HG)，二輪車 (BK)，自動車 (VH) の 4 クラスを対象とする。評価用データとして，3 年間分の映像データから 30 分からなる映像シーンを選び，計 23 時間の映像データベースを作成し，その中から各識別クラス 200 パターン，計 800 パターンを用いた。作成するコードブックのサイズは 32，当てはめる正規分布数は 4 とする。図 9 に実験に使用した画像例を示す。

表 1　識別結果 [%]
Table 1.　Classification rate [%]

		α						
		0.0	0.1	0.3	0.5	0.7	0.9	1.0
Class	SH	75.6	80.8	77.5	74.7	71.4	70.9	71.8
	HG	80.4	87.1	85.7	85.7	85.7	85.2	85.2
	BK	86.3	87.7	86.3	87.2	86.3	85.3	85.8
	VH	97.3	96.8	95.9	95.9	96.4	96.4	96.4
	合計	85.0	88.2	86.4	85.9	85.0	84.5	84.9

〈4•2〉　実験結果　　統合率 α を変化させたときの識別結果を表 1 に示す。大域的特徴量と構造に基づく特徴量を統合することで，3.2% 識別率を向上させることができた。表 2 に，$\alpha = 0.1$ での識別結果をコンフュージョンマトリクスに示す．構造情報を加えることで大域情報のみ ($\alpha = 0$) を使用した際に人を二輪車と誤識別したパターンを減少さ

表 2　識別結果 ($\alpha = 0.1$)
Table 2.　Confusion matrix($\alpha = 0.1$)

		out					
		SH	HG	BK	VH	correct	rate[%]
in	SH	172	24	16	1	172	80.8
	HG	9	182	16	2	182	87.1
	BK	15	10	185	1	185	87.7
	VH	7	0	0	212	212	96.8
	合計					751	88.2

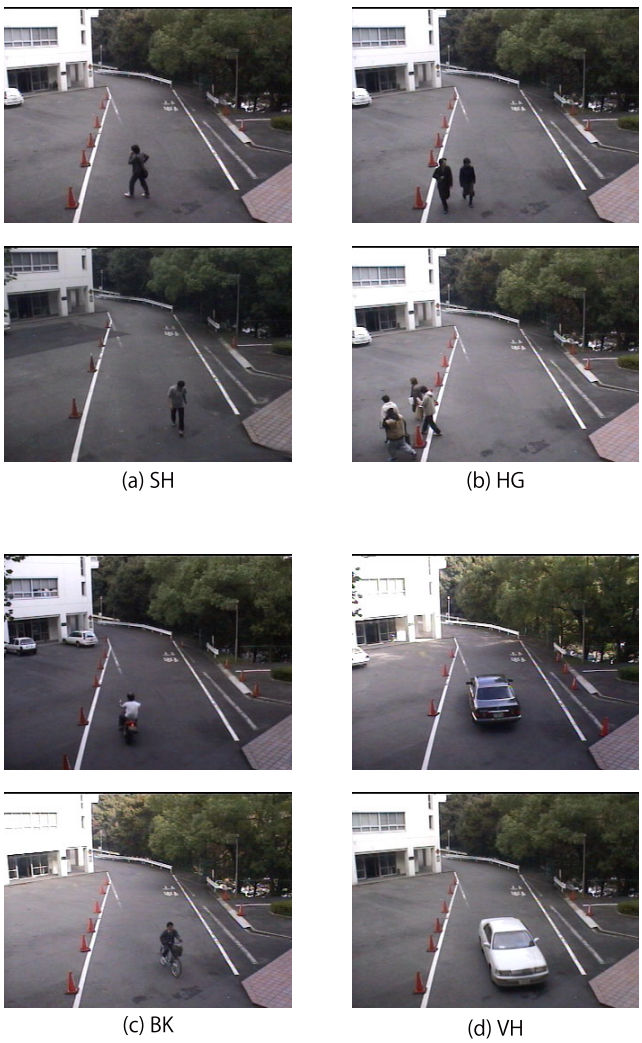

図 9　実験データ例
Fig. 9.　Example of video image.

せることができた。図 10 に人と二輪車の領域全体と各領域ごとに得られたベクトル量子化ヒストグラムを示す。人と二輪車の大域的特徴量は非常に似ていることがわかる．従って，大域的に捉えた特徴の記述では，誤識別が起こる可能性が高い．一方，構造に基づく特徴量は，図 10 の (a)(b) と (e)(f) のように上半身の特徴量は似ているが，(c)(d) と (g)(h) のように下半身部分の特徴量が異なっているのがわかる．そのため，部分的に物体の見え方が似ている人と二輪車の誤識別を抑制することができたと考えられる．図 11 に，構造に基づく特徴量を加えたことで正解した例を示す。

5.　おわりに

本稿では，SIFT を用いた物体構造に基づく特徴量の記述と物体識別法を提案した。評価実験により，従来の大域的特徴量に構造に基づく特徴量を追加し，統合することにより識別率を約 3% 向上させることができた。特に，特徴量を構造ごとに記述することで，人と二輪車の下半身の違

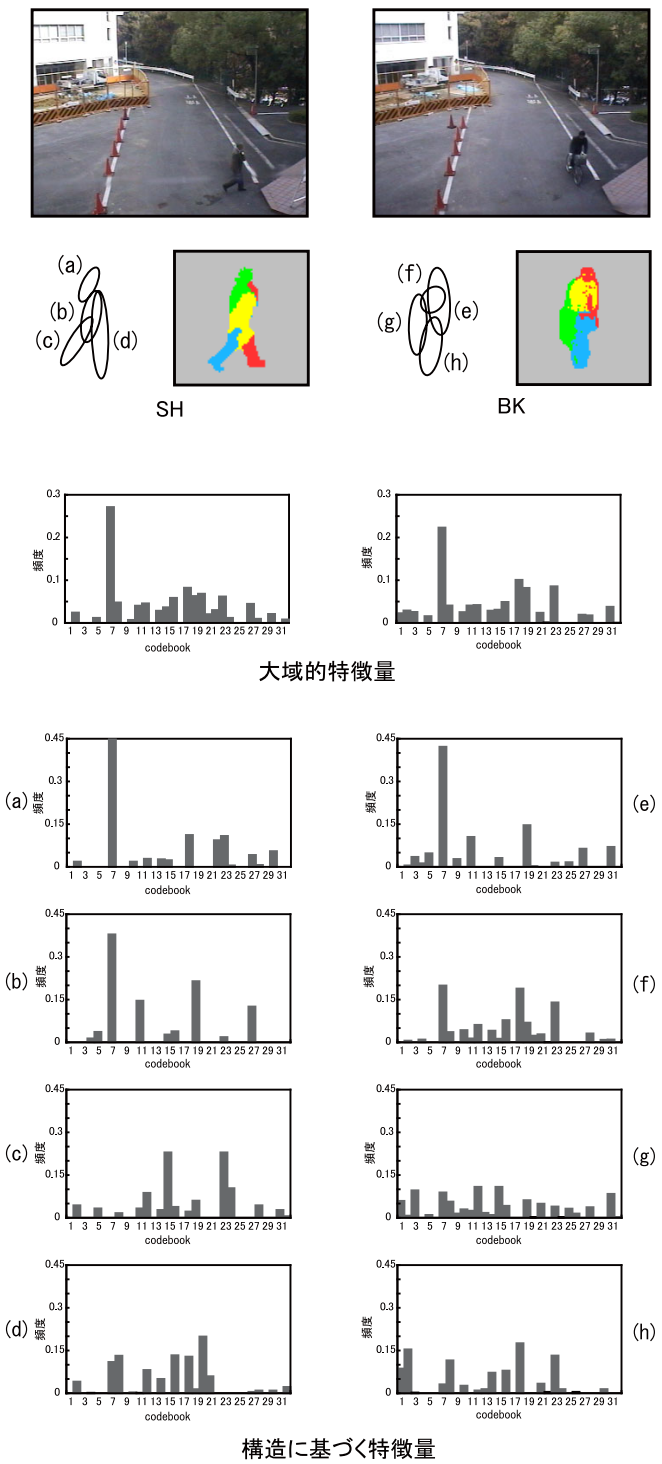

図 10　人と二輪車の特徴量の違い
Fig. 10.　Structure-based feature for mis-classification pattern (single human and bike).

いを明確にすることができ，人と二輪車の識別精度を向上させることができた。

図 11　正解データ例
Fig. 11.　Example of object classification.

参考文献

（1） A. Lipton, H. Fujiyoshi, and R. S. Patil, “Moving target detection and classification from real-time video.”, Proc. of the 1998 Workshop on Applications of Computer Vision (1998).

（2） R. Collins, A. Lipton, H. Fujiyoshi, and T. Kanade, “Algorithms for cooperative multisensor surveillance”, Proc. of the IEEE, Vol. 89, No. 10, pp1456 - 1477, (2001).

（3） O. Hasegawa and T. Kanade, “Type Classification, Color Estimation, and Specific Target Detection of Moving Targets on Public Streets”, Machine Vision & Applications, Springer, Vol.16, No.2, pp116-121 (2005).

（4） N. Dalal and B. Triggs, “Histograms of oriented gradients for human detection”, In Proc. of the IEEE Conference on Computer Vision and Pattern Recognition, pp886-893, San Diego, CA (2005)

（5） P. Viola, M. J. Jones, D. Snow. “Detecting Pedestrians Using Patterns of Motion and Appearance” Proc. of the Ninth IEEE International Conference on Computer Vision, Volume 2 (2003).

（6） G. Csurka, C. Bray, C. Dance, and L, Fan, “Visual Categorization with Bags of Keypoints”, Proc. of ECCV Workshop on Statistical Learning in Computer Vision, pp. 1-22 (2004).

（7） R. Fergus, L. Fei-Fei, P. Perona, and A. Zisserman, “ Learning Object Categories from Google’s Image Search”, Proc. of IEEE International Conference on Computer Vision, pp. 1816-1823 (2005).

（8） M. Seki, K. Sumi, H. Taniguchi, and M. Hashimoto, “Gaussian Mixture Model for Object Recognition”, MIRU2004, vol. 1, pp. 344-349 (2004) (in Japanese).

（9） N. Hirata, M. Seki, H. Okuda, and M. Hashimoto, “Vehicle Detection using Gaussian Mixture Model from IR Image”, IEICE Technical Report PRMU2005, pp. 37-42, (2005) (in Japanese).

（10） C. Stauffer and W. E. L Grimson, “Adaptive background mixture models for real-time tracking”, n Proc. of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 246-252 (1999).

（11） N. Ueda, and R. Nakano, “Deterministic Annealing EM Algorithm”, IEICE Journal(D-II), Vol. J80-D-II, No. 1, pp. 267-276 (1997).

（12） D. Comaniciu, and P. Meer, “Mean Shift Analysis and Applications”, Proc. IEEE Seventh Int’l Conf. Computer Vision, vol. 2, pp. 1197-1203 (1999).

（13） D. G. Lowe, “Distinctive image features from scale-invariant keypoints”, International Journal of Computer Vision, 60(2), pp. 91-110 (2004).