

構造に基づく特徴量を用いたグラフマッチングによる物体識別

永橋 知行† 藤吉 弘亘† 金出 武雄††

† 中部大学 工学部 情報工学科

†† カーネギーメロン大学ロボティクス工学研究所

E-mail:kida@vision.cs.chubu.ac.jp, hf@cs.chubu.ac.jp, tk@cs.cmu.edu

あらまし 従来、物体識別には形状やテクスチャ情報に基づく特徴が識別器への入力として用いられている。これらの入力特徴は、物体領域全体を大局的に捉えたものであり、その物体の構造情報は用いられていない。そこで、本稿では混合正規分布モデルを用いて記述した物体の構造パターンに基づく識別法を提案する。本手法は、検出された物体領域に混合正規分布モデルを当てはめ領域分割を行う。分割後の各領域から求めたテクスチャと形状に基づく特徴量をノード、2つの正規分布間の距離をエッジとするグラフを作成する。各クラスの参照グラフと入力グラフとの距離を算出し、kNN法を用いて移動体を自動車/人/複数の人/二輪車に識別する。評価実験の結果、全体から得られる特徴量に構造情報を加えることで、識別率を向上させることができた。

Structure-based Object Type Classification using Graph Matching

Tomoyuki Nagahashi† Hironobu Fujiyoshi† Takeo Kanade††

†Dept. of Computer Science, Chubu University

††The Robotics Institute, Carnegie Mellon University

E-mail:kida@vision.cs.chubu.ac.jp, hf@cs.chubu.ac.jp, tk@cs.cmu.edu

Abstract Current feature-based object type classification methods use texture and shape based information derived from image patches. Generally, input features such as aspect ratio are derived from some rough characteristics of the entire object. However, we derive input features from parts-based representation of the object. In this paper, we propose a method to distinguish object types using structure-based features described by a mixture of Gaussian distribution. Our approach uses the Gaussian fitting onto object image to segment into several sub-regions, each of which is related to a physical part of the object. We model the object as a graph, where the nodes contain texture and shape information obtained from the corresponding segmented regions, and the edges contain the distance information between two connected regions. Calculating the distance of the reference graph and the input graph, we can classify an object into single-human/human-group/vehicle/bike using k-NN based classifier. We demonstrate that we can obtain higher classification performance when we use both conventional features and structure-based features compared to using only conventional features or only structure-based features.

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など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る。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, 2, 3]。文献 [1] では人、自動車の 2 クラスの識別に、dispersedness と Area という 2 つの特徴量を線形判別関数に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、ノイズその他の計 6 種への識別を 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では、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。

これらの移動体識別に使用される特徴量は、検出された移動体領域全体から得られる特徴量である。一方、関心は物体の内部パターンの構造に着目し、検出された物体に混合正規分布を当てはめることにより物体を複数の主領域の大まかな形状と位置関係を表現し、物体認識に利用する手法を提案した [5], [6]。この手法では、物体の色の違いや個体差等に影響されにくいという特徴がある。しかし、形状が似ているクラス間（例えば人と二輪車等）では、混合正規分布当てはめ結果はほぼ同等となり、識別が困難となる場合がある。そこで、従来用いられてきた形状やテクスチャから得られる特徴量を物体構造に基づいて記述することで、高精度な物体識別の実現が期待できると考えられる。

本稿では、関心によって提案された混合正規分布の当てはめによる物体構造の記述法を拡張する。混合正規分布モデルを用いて検出領域を主領域に分割し、各領域から得られる特徴量をグラフを用いて表現する。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。

2 物体構造に基づく特徴量の抽出

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。図 1 に提案する移動体識別の流れを示す。移動体の検出には、ピクセル毎のしきい値を適応的に設定するフレーム間差分 [4] を用いる。フレーム間差分により得られた移動体領域には、図 1(b) に示すように背景が含まれているため、正確に物体構造を捉えることが困難である。そこでグラフカット [8] を用いて誤検出した背景の除去を行う。次に、グラフカットにより得ら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特徴量を抽出し、グラフを用いてその関係を記述する。以下に、各処理の詳細を示す。

2.1 グラフカットによる背景除去

混合正規分布モデルによる当てはめの際、未検出領域の発生や、背景領域の誤検出が含まれると物体構造を正確に表現できない。そこで、グラフカットを用いてより正確な物体領域のセグメンテーションを行う（図 3）。

グラフカットは、ピクセルをノードとしたグラフを作成し、そのグラフの最小カットを求めることで、物体と背景のセグメンテーションを行う。グラフ G は、頂点 V (ノード) とそれらをつなぐ線 E (エッジ) の関係を表したものであり、 $G = \langle V, E \rangle$ と表す。画像からグラフを作成する場合、ピクセルの集合を P 、近傍ピクセルの集合を N 、またターミナルを表現するノードである “source(object)” を s 、 “sink(background)” を t とした際、ノード V とエッジ E 以下の式で表される。

$$V = P \cup \{s, t\} \quad (1)$$

$$E = N \cup_{p \in P} \{\{p, s\}, \{p, t\}\} \quad (2)$$

このようなグラフ G をフローネットワークと呼び、その構成を図 2(a) に示す。近傍ピクセル P 間のエッジを n -link、各ピクセル p から s や t に接続しているエッジを t -link と呼ぶ。注目ピクセルを $p \in P$ 、その近傍ピクセルを $q \in N$ 、ノード p と q をつなぐエッジを $\{p, q\}$ とするとき、 $\{p, q\}$ のエッジコスト

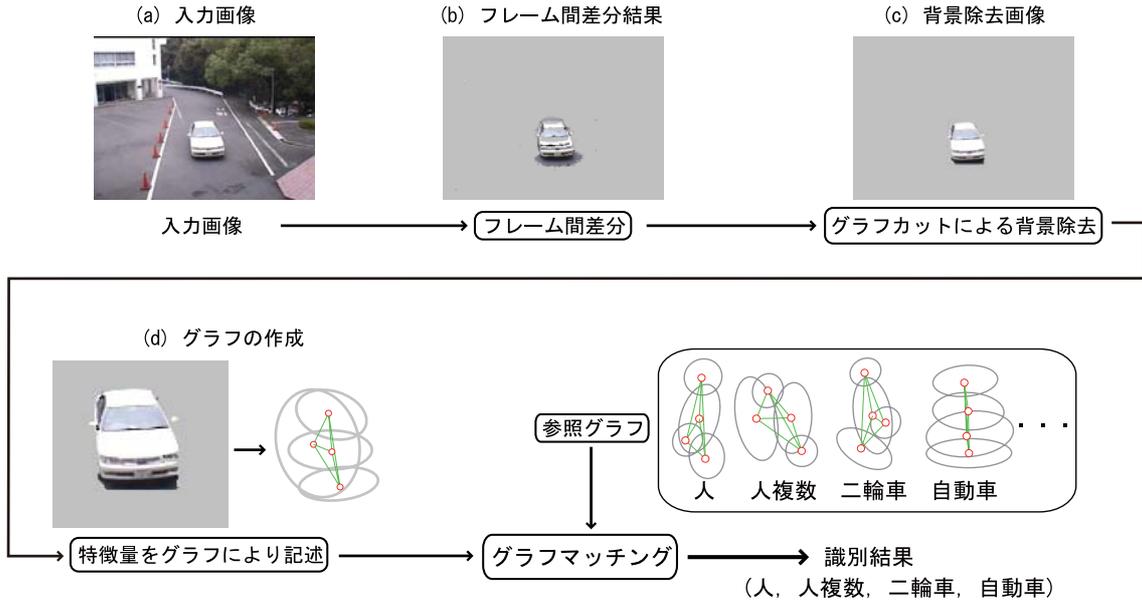


図 1: 物体識別の流れ

$B_{\{p,q\}}$ は次式により求められる .

$$B_{\{p,q\}} = \exp\left(-\frac{(I_p - I_q)^2}{2\sigma^2}\right) \cdot \frac{1}{\text{dist}(p,q)} \quad (3)$$

ここで, I_p はピクセル p の輝度値, $\text{dist}(p,q)$ はピクセル間の距離, σ は近傍との連結度に関連するパラメータであり経験的に決定する. t-link は, あらかじめ物体領域と背景領域のヒント (seed) となるピクセル \mathcal{O} (“object” seed) と \mathcal{B} (“background” seed) を与えることにより, 次式から決定される .

$$\{p, s\} = \begin{cases} \lambda \cdot R_p(\text{“bkg”}) & , p \notin \mathcal{O} \cup \mathcal{B} \\ K & , p \in \mathcal{O} \\ 0 & , p \in \mathcal{B} \end{cases} \quad (4)$$

$$\{p, t\} = \begin{cases} \lambda \cdot R_p(\text{“obj”}) & , p \notin \mathcal{O} \cup \mathcal{B} \\ 0 & , p \in \mathcal{O} \\ K & , p \in \mathcal{B} \end{cases} \quad (5)$$

$$R_p(\text{“obj”}) = -\ln \Pr(I_p | \mathcal{O})$$

$$R_p(\text{“bkg”}) = -\ln \Pr(I_p | \mathcal{B})$$

$$K = 1 + \max_{p \in \mathcal{P}} \sum_{q: \{p,q\} \in \mathcal{N}} B_{\{p,q\}}$$

λ は n-link との関係を表す比例係数である .

グラフカットによるセグメンテーションを行う場合, t-link のエッジコストを求めるために seed を決定しなければならない. 文献 [8] では, seed をインタラクティブにユーザが与えるためセグメンテーションのプロセスは自動化されていない. そこで, 本稿ではフレーム間差分の結果から物体と背景の seed

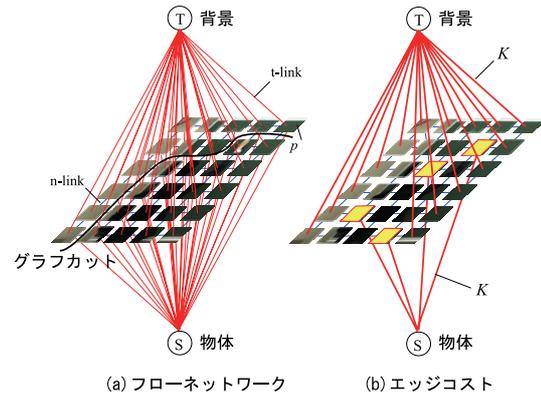


図 2: グラフカット

を決定する. まず, フレーム間差分により検出された領域を包含する大きさの矩形領域を抽出する. 切り出された矩形領域には物体が矩形領域内部の中央に存在する. そのため, 図 3(a) に示すように矩形の外周は背景である可能性が高い. そこで, 矩形の外周を “background” seed として与える. 次に, 矩形領域内の各ピクセルを座標値 (u, v) と輝度値 I の 3次元空間にてクラスタリングを行う. クラスタリングにより得られた各クラスの中心座標の (u, v) を図 3(c) のように “object” seed として与える .

各エッジコストを式 (4), (5) から求めることによりグラフ \hat{G} が完成する (図 2(b)). 作成した \hat{G} において, s と t を分割するエッジコストの和が最小となるようにエッジを切断し, 2つのグラフへ分割する. \hat{G} を分割した2つのグラフのうち, s を含むグラフのノードが物体, t を含むグラフのノードが背景

となる．グラフカットによる背景除去結果を図3(d)に示す．

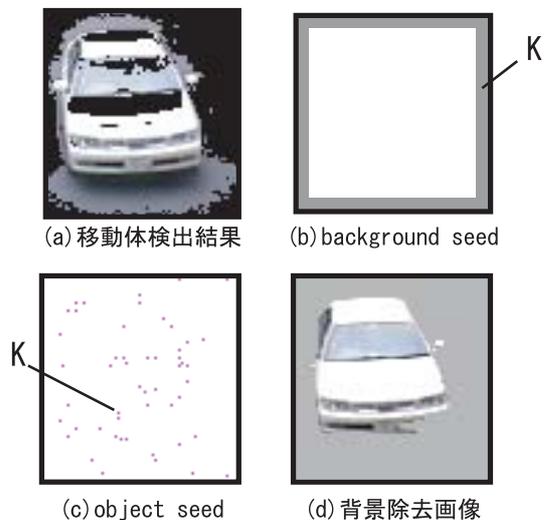


図3: グラフカットによる背景除去

2.2 混合正規分布の当てはめ

物体の内部パターンを記述するために，背景除去後のパターンに混合正規分布を当てはめる [5]．物体領域の座標 (u, v) と輝度 I を $\mathbf{x}_i = \{u_i, v_i, I_i\}^T$ ，混合正規分布パラメータを $\Phi = \{\alpha_j, \phi_j = (\mu_j, \Sigma_j)\}_{j=1}^c$ とする． \mathbf{x} に対して，式 (6) の確定的アニーリング EM (DAEM: Deterministic Annealing EM) アルゴリズム [9] を用いて混合正規分布パラメータ Φ_{ML} を推定する．

$$\Phi_{ML} = \arg \max_{\Phi} \sum_{j=1}^c (\alpha_j \cdot p_j(\mathbf{x} | \mu_j, \Sigma_j))^\beta \quad (6)$$

$$p(\mathbf{x} | \mu_j, \Sigma_j) = \frac{1}{\sqrt{(2\pi)^3 |\Sigma_j|}} \exp \left\{ -\frac{1}{2} (\mathbf{x} - \mu_j)^T \Sigma_j^{-1} (\mathbf{x} - \mu_j) \right\} \quad (7)$$

ただし， $p_j(\mathbf{x} | \mu_j, \Sigma_j)$ は，平均 μ_j ，共分散行列 Σ_j であり， $\phi_j = \{\mu_j, \Sigma_j\}$ 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4に，正規分布を当てはめ後， Φ_{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 (クラスター) を表しており，これを物体構造の記述化に利用する．

2.3 混合正規分布による領域分割

推定されたパラメータ Φ_{ML} から，各ピクセル \mathbf{x} がどの正規分布 ϕ_i に属しているかを次式により求め，領域分割を行う (図4(c))．

$$C_i = \arg \max_i p_i(\mathbf{x} | \phi_i) \quad (8)$$

一般に領域分割手法として用いられている Mean-



図4: 物体構造のモデル化

Shift [10] は，図5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ーに，側面のガラスとボディは別のクラスターに分けられる．一方，提案手法は混合正規分布を用いるため，側面や背面といった物体構造に基づいた領域分割が可能である．

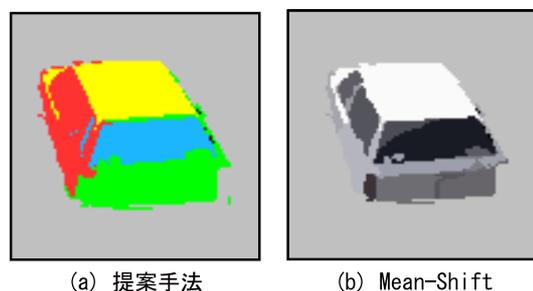


図5: Mean-Shift との違い

2.4 特徴量

領域分割後の各領域に対し，形状に基づく特徴量として正規分布の縦横比と傾き，テクスチャ情報に基づく特徴量として4方向のエッジ強度を用いる．以下に，各特徴量の算出方法を述べる．

2.4.1 形状情報に基づく特徴量

各正規分布の平均 μ_j , 分散共分散行列 Σ_j から正規分布の傾き θ と長径と短径を求める. 各領域に対して当てはめた正規分布の縦横比と傾きを次式により計算し, おおまかな形状を表す特徴量とする.

$$\text{正規分布の縦横比} = \frac{\text{正規分布の短径}}{\text{正規分布の長径}} \quad (9)$$

$$\text{正規分布の傾き} = \sin \theta \quad (10)$$

2.4.2 テクスチャ情報に基づく特徴量

分割された各領域におけるエッジ強度を4方向(垂直: V , 水平: H , 右上がり: R , 左上がり: L)について求める. まず, 次式に示す各方向 $v \in \{V, H, R, L\}$ のエッジ検出オペレータ h_v を用いて, 対象画素 (i, j) を含む近傍画素 Ω に対するエッジ強度 $g_v(i, j)$ とエッジ方向 $v'(i, j)$ を求める.

$$h_V = \begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix} \quad h_H = \begin{bmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ 1 & 2 & 1 \end{bmatrix}$$

$$h_R = \begin{bmatrix} 0 & 1 & 2 \\ -1 & 0 & 1 \\ -2 & -1 & 0 \end{bmatrix} \quad h_L = \begin{bmatrix} 2 & 1 & 0 \\ 1 & 0 & -1 \\ 0 & -1 & -2 \end{bmatrix}$$

エッジ強度

$$g_v(i, j) = \sum_{(m, n) \in \Omega} f(m, n) h_v(i - m, j - n) \quad (11)$$

エッジ方向

$$v'(i, j) = \arg \max_{v \in \{V, H, R, L\}} g_v(i, j) \quad (12)$$

求めたエッジ強度を各領域 O_i ごとの最大強度の総和で正規化したものを v 方向のエッジ強度特徴とする.

$$G_i(v) = \frac{\sum_{(i, j) \in O_i} g_v(i, j) \cdot \delta_K[v'(i, j) = v]}{\sum_{(i, j) \in O_i} \max_v (g_v(i, j))} \quad (13)$$

3 グラフマッチングによる識別

3.1 グラフによる記述

領域毎に得られた特徴量とその関係をグラフを用いて表現する. グラフ G は頂点 N (ノード) と,

頂点を結ぶ辺 E (エッジ) によって構成される. 各正規分布に対応するノード n_i には2.4で求めた特徴量を記述する. また, エッジ e_{ij} にはノード n_i と n_j に対応する正規分布間の距離を記述し, 各ノードの位置関係を表す(図6).

$$n_i = \{t_1, t_2, t_3, t_4, t_5, t_6\} \quad (14)$$

$$e_{ij} = \{\text{dist}(\phi_i, \phi_j)\} \quad (15)$$

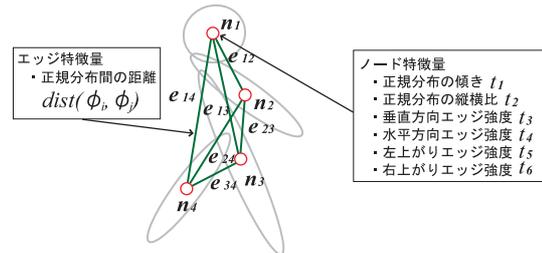


図6: 物体構造からグラフの作成

3.2 グラフマッチング

ノードの集合を $N = \{n_1, \dots, n_4\}^T$, エッジの集合を $E = \{e_{11}, \dots, e_{34}\}^T$ とする. 参照グラフ $T = \{N_t, E_t\}$ と, 入力グラフ $X = \{N_x, E_x\}$ 間のマッチングコストを以下の式より求める.

$$\text{cost}(T, X) = |N_t - N_x| + |E_t - E_x| \quad (16)$$

T と X のノードの対応は未知である. そこで, T と 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 T と X のマッチングコストとする.

$$\text{Cost}(T, X) = \min_{i \in \text{cl}=24} \{\text{cost}(T, X_i)\} \quad (17)$$

クラス毎に用意された複数の参照グラフと, 検出された領域から得られる入力グラフとのマッチングコストを計算し, kNN法を用いて入力グラフの識別クラスを判定する.

4 評価実験

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. 評価には, 各識別クラス200パターン. 計800パターンを用いる. 当てはめる混合正規分布の数は4とし, kNN法による識別では $k=5$ として実験を行った. 比較対象には, 検出領域全体から上記と同様に求めた特徴量を用いる. 図7に各クラスに対する識別率を示す.

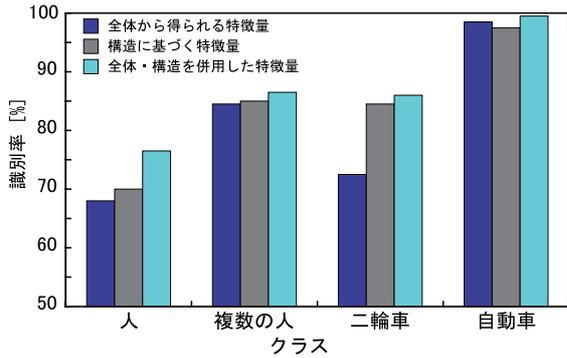


図 7: 識別結果

物体領域全体から得られる特徴量より構造情報に基づく特徴量のほうが高い識別率を得た。特に、二輪車の識別率は約 12.0% の向上が見られた。通常、人と二輪車は図 8 のように、正面からの見え方が似ているため誤識別を招き易い。提案手法では、構造情報を用いることにより物体の下部の特徴の違いを記述することができるため、識別率を向上させることができた。また、全体から得られる特徴量と構造情報に基づく特徴量を併用してグラフマッチングした結果、全てのクラスにおいて識別率を向上させることができた。これは、物体の特徴を大まかに捉え、かつ部分的な特徴を表現することで、クラス間の特徴の違いをより正確に捉えることができたためといえる。表 1 に 2 つの特徴量を併用したときのコンフュージョンマトリクスを示す。構造情報を用いることで、識別率性能を上げることができたが、人と二輪車間の誤識別が多いことがわかる。この解決には、新たな特徴量、例えばオブティカルフローを加えることが考えられ今後の課題としたい。

表 1: 特徴量を併用した識別結果

		出力					
		SH	HG	BK	VH	正解数	正解率
入力	SH	153	14	33	0	153	76.5
	HG	21	173	6	0	173	86.5
	BK	20	7	172	1	172	86.0
	VH	0	1	0	199	199	99.5
	合計					697	87.1

5 まとめ

本稿では、物体構造に基づく特徴量を用いたグラフマッチングによる物体識別法を提案した。評価実験により、物体構造に基づく特徴量の有効性を確認した。さらに、領域全体から得られる特徴量と構造

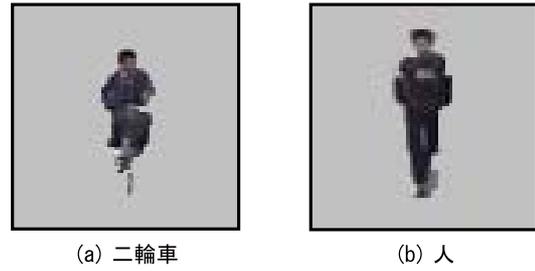


図 8: 人と二輪車の見え

情報に基づく特徴量を併用することにより、識別率を向上させることができた。今後は、さらなる識別能力の向上を目指し、有効な特徴量の検討と識別器の高性能化について検討を行う予定である。

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, October, pp. 1456 - 1477, 2001.
- [3] 長谷川修, 金出武雄, "一般道路映像中の移動物体の識別・色の推定と特定対象の検出", 情報処理学会論文誌, Vol.44, No.7, pp.1795-1807, 2003.
- [4] 小村剛史, 藤吉弘巨, 矢入郁子, 香山健太郎, 吉水宏, "歩行者移動支援のための道路状況配信システム", 情報処理学会 コンピュータビジョンとイメージメディア研究会 2004.
- [5] 関真規人, 鷲見和彦, 谷口博康, 橋本学, "物体認識のための混合正規分布モデル", 画像の認識理解シンポジウム MIRU2004, vol. 1, pp. 344-349, July 2004
- [6] 平田奈美, 関真規人, 奥田晴久, 橋本学, "混合正規分布モデルを用いた赤外線画像からの車両検出", 電子情報通信学会技術研究報告, pp37-42, 2005.
- [7] C. Stauffer and W.E.L. Grimson, "Adaptive background mixture models for real-time tracking", CVPR99, pp.246-252, 1999.
- [8] Yuri Boykov, Marie-Pierre Jolly "Interactive Graph Cuts for Optimal Boundary & Region Segmentation of Objects in N-D Images", ICCV, vol. I, pp. 105-112, 2001
- [9] 上田修功, 中野良平, "確定的アニーリング EM アルゴリズム", 電子情報通信学会論文誌 (D-II), Vol. J80-D-II, No. 1, pp. 267-276, 1997.
- [10] Dorin Comaniciu, Peter Meer, "Mean Shift Analysis and Applications", ICCV '99: Proceedings of the International Conference on Computer Vision-Volume 2, 1999.