

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

長谷川 修^{†1,†2,†3} 金出 武雄^{†4,†5}

人と車両が入り混じる、対面通行の一般道路映像中の移動物体を対象に、(1) {車両, 人物, その他} の識別および車両ならばその種別の識別, (2) 対象が含む主要な色の推定, を同時に行うビジョンシステムを試作した。識別する対象の種別は、実験システムの入力映像中に多く見られた、{小型作業車, セダン, バン, トラック, 人物(単数, 複数とも), その他(ノイズなど)} とし、また色は {red-orange-yellow, green, blue-lightblue, white-silver-gray, darkblue-darkgray-black, darkred-darkorange} の色グループから選択させた。入力映像として晴天もしくは曇天の午前9時から午後5時の間の映像を用いた実験の結果、種別と色の識別結果が双方ともオペレータの目視による識別結果と合致した場合を正答として、91.1%の正答率を得た。また本研究では、上記の機能を利用し、あらかじめ設定した特定対象(種別と色の組合せから特定される)を映像中から検出させる実験も行った。特定対象を {FedEx 社の集配車, 郵便車, パトカー} に設定して行った実験の結果、92.9%の正答率を得た。こうした識別や推定の機能は、統計的線形判別法(線形判別分析)と非線形識別則(重みつき K-最近傍法)を組み合わせる手法を導入して実現した。

Type Classification, Color Estimation, and Specific Target Detection of Moving Objects on General Streets

OSAMU HASEGAWA^{†1,†2,†3} and TAKEO KANADE^{†4,†5}

This paper describes a vision system that recognizes moving objects (cars and humans) on general streets. The system classifies objects based on shape appearance and estimates their colors from images of color video cameras set up toward a street. The input images were obtained between 9:00 a.m. and 5:00 p.m. in fine or cloudy weather conditions. The types set up in the system are {human, sedan, van, truck, mule (golf cart for workers), and others}, and the colors are {red-orange-yellow, green, blue-lightblue, white-silver-gray, darkblue-darkgray-black, and darkred-darkorange}. Moreover, the system can selectively extract specific targets, such as FedEx vans, police cars, mail vans and others from objects based on classified and estimated results. For classification and estimation, we cooperatively used a stochastic linear discrimination method (Linear Discriminant Analysis: LDA) and a nonlinear decision rule (weighted K-Nearest Neighbor rule: K-NN).

1. はじめに

近年セキュリティに対する意識が高まり、監視を目的としたカメラを街頭に多く見かけるようになったが、

屋外映像は太陽の位置や天候の影響を強く受けるため、そこに映し出される状況の自動把握は難しく、現状ではオペレータが目視により確認するのが一般的となっている。しかし、人間は同時に複数の映像をモニタリングすることは困難であり、また疲労や慣れによる注意力の低下なども避けられないため、そうした処理の自動化が望まれている。

そこで本研究では、人と車両が入り混じる、対面通行の道路映像中の移動物体を対象に、

- 移動物体を6種の種別に識別、
- 移動物体が含む主要な色を5色の色グループから推定、

†1 東京工業大学大学院理工学研究科付属情報工学研究施設
Imaging Science and Engineering Laboratory, Tokyo
Institute of Technology

†2 産業技術総合研究所脳神経情報研究部門
Neuroscience Research Institute, Advanced Industrial
Science and Technology (AIST)

†3 科学技術振興事業団さきかけ研究 21
PRESTO, Japan Science and Technology Corp. (JST)

†4 カーネギーメロン大学ロボティクス研究所
The Robotics Institute, Carnegie Mellon University

†5 産業技術総合研究所デジタルヒューマン研究センター
Digital Human Research Center, Advanced Industrial
Science and Technology (AIST)

今回の実験では、双方向に人や車両が通行する、片側1車線の中央分離帯のない道路の映像を用いた。

を行うことを主たる目的として、その実現のためのアルゴリズムの探求と試作システムの構築を行った。今回処理対象としたのは、おおむね対象の側面がとらえられる位置に設置した固定カラーカメラからの映像であり、映像入力時の天候と時刻は、晴天もしくは曇天の午前9時から午後5時の間とした。

ここで、本研究の処理対象は屋外環境を移動するため、様々な不確定要因に影響を受け、(1)他の物体の陰に隠れる、(2)日向から日陰に入る、(3)車体に他の物体が映り込む、(4)複数の対象物が重なる、といった様々な状況が生じることが考えられる。

これは、対象の種別の識別のために有用な情報(欠けやブレなどが少ない形状情報)と、色の推定のために有用な情報(強い太陽光の反射や影などの影響が少ない画像)がいつ得られるかは不定であることを意味している。

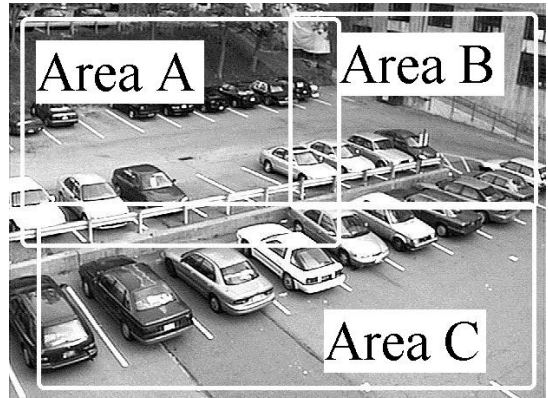
そこで本研究では、この問題に対処するため、対象をトラッキングしつつ各時刻で種別の識別と色の推定を処理結果に対する確信度(重み)つきで独立に行い、トラッキングシーケンスを通じてそれぞれ最も確信度が高かった結果を選択する手法を提案する。これにより、識別・推定結果の出力に時間遅れが生じることとなるが、このように種別の識別結果と色の推定結果から最も良好なものを独立に選択することによって、全体の処理パフォーマンスの向上が図れるものとする。本研究では、こうした工夫を通じ、不確定要因の大きい屋外環境における色の推定や、6種の対象の姿勢の多様性に対する適応性を探求することを主眼とする。

図1(a)に本システムへの入力画像の例を示し、図1(b)にここを通過した乗用車を切り出して、トラッキングした例を示す。

識別する対象の種別は、実験システムの入力映像中に多く見られた、{小型作業車、セダン、バン、トラック、人物(単数、複数とも)、その他(ノイズなど)}とし、また色は{red-orange-yellow, green, blue-lightblue, white-silver-gray, darkblue-darkgray-black, darkred-darkorange}の色グループから選択させた。

実験の結果、種別と色の識別結果が双方ともオペレータの目視による識別結果と合致した場合を正答として、91.1%の正答率を得た。

また本研究では、上記の機能を活用し、あらかじめ設定した特定対象(種別と色の組合せから特定される)が映像中に現れると、それらを検出させる実験も行う。



(a) 入力映像とその領域分割

(a) Input image and its segmentation.



(b) 入力映像からの対象の切り出しとトラッキング例

(b) A sample sequence of segmented and tracked target.

図1 システムへの入力映像とその領域分割

Fig.1 Input image to the system and its segmentation.

た。特定対象を{FedEx社の集配車、郵便車、パトカー}に設定して行った実験の結果、92.9%の正答率を得た。

従来研究としては、自動車専用道路の走行レーン上を一方に走る車や、料金所で停止した車の種別を識別をするシステムの構築事例は多く報告されている^{1)~4)}。しかし、CCDカメラから入力した対面通行の一般道路映像を対象に、映像中の移動物体の種別の識別と色の推定や、特定車両の検出を試みた事例は見当たらない。

Jollyらは、画像のほぼ中央に車体の側面(真横)が一定の大きさで完全に見えるようにした画像にDeformable Templateを適用して車体の輪郭を切り出し、輪郭の形状から{Sedan, Pickup Truck, Hatchback, Station Wagon, Van}のいずれであるかを識別する事例を報告している⁵⁾。ここで、処理対象とする車両の画像は、影や他車両の映りこみ、太陽光の反射の影響などがないように調整されている。これに対し本稿の提案手法では、一般道に向けたカメラからの入力映像をそのまま処理対象とし、図2~4に示す程度の向きや画像中の大きさの変動、また上述の日照条件・天候の変化は処理範疇に含めた。

処理対象が入力画像中に現れてから出ていくまでの過程。

Kanadeらは、市街地の映像から背景差分により運動物体を切り出し、輪郭線の複雑さと面積の比、撮影時のカメラのズーム値などのパラメータを3層のニューラルネットに学習させて、様々なズーム値で対象を識別するシステムを開発している⁶⁾。このシステムでは、背景差分処理時に基準とする背景画像を適応的に更新するといった工夫が加えられているが、識別種別は3種 {単数の人物、複数の人物、車両} となっている。

影沢らは、赤外線カメラの映像を利用し、車両の映像の重なりを許容しつつ、それらを独立に車両をオンラインで認識するシステムを報告している。しかし、乗用車、トラックといった車種の識別は今後の課題とされている⁷⁾。本稿に述べるシステムでは、映像中の重なりのある処理対象に対しては、後述する方法論により対処した。

2. 本システムの概要

2.1 ハードウェア構成

本研究で構築した実験システムは、2階建ての建物の屋上に設置した固定カラーカメラと、カメラからの入力映像を処理する市販のパソコン1台からなる。画像処理専用のハードウェアは用いていない。カメラは市販のCCDカメラであり、パソコンにはNTSC規格のビデオ信号を入力した。入力映像のサイズは、640 × 480ピクセルである。

2.2 ソフトウェア構成

パソコンのOSにはLinux、プログラムの記述にはC言語を用いた。実験システムは、(a) 画像から背景差分をベースに対象(車両や人物)を切り出してトラッキングするモジュール、(b) 切り出された対象の種別を識別するモジュール、(c) 対象の色を推定するモジュール、(d) 種別の識別・色の推定結果を総合して最終判定をするモジュール、の4つのモジュールからなる。

このうち(a)のモジュールには、本研究とは別に研究・開発されたLiptonらのモジュール⁸⁾を、同氏らからプログラムの供与を受けて利用した。このモジュールは日照条件や天候の変化に応じて適応的に背景画像を更新して対象を切り出すほか、切り出した対象に固有のIDを付けてトラッキングすることができる。

3. 対象の種別の識別

3.1 方針

本研究では車両と人物の双方を識別の対象とした。図2に示すように車両の形状は多彩であり、また人



Trucks



Vans



Sedans

(a) 領域 A から切り出した処理対象の画像例
(a) Sample images extracted from Area A.



Sedans



Vans

(b) 領域 B から切り出した処理対象の画像例
(b) Sample images extracted from Area B.

図2 処理対象の画像例

Fig. 2 Sample images of targets.

物の2人以上のグループでは人が様々な配置で観察される。

一般に画像対象の識別法には、対象の形状モデルを作成して利用するモデルベースの手法と、対象の「見え」をそのまま利用する「見え」ベースの手法がある。

このうちモデルベースの手法では、車種ごと(セダン、バンなど)の3次元モデルを回転・拡大・縮小させて生成した2D画像を、入力対象の画像とマッチングさせて識別するといった手法が考えられる。しかしこの方法では、識別対象のクラスごとにモデルを作る必要があり、本研究では図2のように各対象の形状が多彩であることから適用が困難であった。

これに対し「見え」ベースの手法は、対象の形状モデルを事前に作る必要がなく、またすべての対象を2次元の画像テクスチャとして統一的に処理できるなど、システム実装の観点からも好ましいと思われたため、本研究では「見え」ベースの識別法を採用した。

具体的には、ベクトル化した「見え」画像を線形判別分析(Linear Discriminant Analysis: LDA)した後、判別空間中で重みつきK-最近傍法を適用することによって、全体として非線形の識別の枠組みとした。

一般にK-最近傍法は、入力ベクトルの次元数が少ない場合は漸近誤差が“Bayes Error”(理論限界誤差)の2倍以内に収まるという際立った特長を持つことで知られるが、その識別精度はバイズ境界面付近のサンプル数によっては変動(低下)する。そこで本研究では、K-最近傍法に近傍ベクトルとの距離に応じた「重み」を導入することにより、その安定化を図った。

3.2 学習用サンプル画像の収集

図1(a)に実験システムへの入力映像の例を示す。本研究の処理対象の「見え」(向き、大きさなど)は、その切り出し位置によって大きく変化する。そこで、まず入力映像を図1(a)に示すArea A, B, Cの3つの領域に分け、それぞれの領域で「見え」が近くなるようにした。なお今回の実験では、駐車スペースに駐車する車や出庫する車は処理対象外とした。

映像中、道路は画面左端から駐車場の間(Area A)を通り、右にカーブして画面右端(Area B)を下方向に抜けており、識別対象(人や車両)は主にここを往来する。先の図1(b)はそうした往来のトラッキング例の1つである。

先に述べたように、本研究では図2~4に示す程度の向きや画像中の大きさの変動は、種別の識別の処理範囲に含めた。図2~4はいずれもArea AとBからトラッキングモジュール(2.2節参照)によって切り出された画像である。これらの図より、同一種別(あ

るいは特定対象)に識別すべき車両でも、その「見え」(車体の形状)は多彩であることが分かる。学習用サンプル画像には、こうした「見え」の異なるものを可能な限り広く集めた。

さらには、車体に描かれたデザインが異なったり、観察された画像中の位置や走行の方向により「見え」が異なった場合もすべてサンプルに含めた。また、晴天時には車両や人物の影も差分画像の一部として切り出されたが、太陽の位置により影のつき方が異なった場合もサンプルに含めた。

学習用サンプル画像の収集は、晴天または曇天の朝9時から日没までの間のオンライン入力映像、もしくはその時間帯に収録した計16時間分のビデオテープの映像から行った。最終的に、種別の識別のための学習用サンプル画像の総数は約2,000枚となった。

なお、図2~4には対象を分かりやすく示すため対象を囲むボックス内のテクスチャ画像を示したが、学習にはこれらの画像から図1(b)のような背景差分の濃淡画像を求めて用いた。

3.3 学習用サンプルへの種別ラベルの貼付け

収集したサンプル画像には、オペレータが目視のうえ、種別{セダン、バン、トラック、小型搬送車、人、その他}のラベルをつけた。サンプル画像があらかじめ定めた特定対象(実験ではFedEx社の集配車、郵便車、パトカーとした)である場合には、その旨のラベルもつけた(よってサンプル画像によっては、2つのラベルがつくこととなる)。

図3に、特定対象の学習用サンプル画像の例を示す。

3.4 種別(形状)の学習過程

学習用サンプル画像(背景差分の濃淡画像)を、幅と高さ方向にそれぞれ2倍に伸長して平滑化する。この背景差分の濃淡画像から、

- 濃淡画像を囲むボックスの幅と高さ(2個)
- ボックスにローカルに設定した座標系における、各軸(2軸)方向の濃淡画像部分の1, 2, 3次モーメント($2 \times 3 = 6$ 個)
- ボックスにローカルに設定した座標系における、濃淡画像部分の重心の座標値(2個)
- 濃淡画像部分の面積(1個)

の計11個のデータを算出して11次元のベクトルを構成し、これを各画像の特徴ベクトルとする。

像がある。

背景差分から切り出された画像には高さや幅が10ピクセル程度のももある。そこで画像を2倍に伸長してから平滑化することにより、伸長しない場合に比べ、よりスムーズな平滑化ができることを期待した。



図 3 特定対象の学習用サンプル画像例
Fig. 3 Sample images of specific targets.

上記の特徴ベクトルを Area A, B, C の各領域ごとにまとめ、さらに種別のラベルを参照して種別ごとのクラスを構成する。これを領域ごとに線形判別分析し、各領域に対応した種別識別用の判別空間を構成する。定式化すると、全クラスのクラス間変動行列を B 、クラス内変動行列を W とすると、

$$B = \sum_{k=1}^J n_k (\bar{x}_k - \bar{x})(\bar{x}_k - \bar{x})^T \quad (1)$$

$$W = \sum_{k=1}^J \sum_{i=1}^{n_k} (x_{k,i} - \bar{x}_k)(x_{k,i} - \bar{x}_k)^T \quad (2)$$

- J : クラス数
- n : 全サンプル数
- n_k : クラス k 内のサンプル数
- \bar{x} : 全ベクトルの平均ベクトル
- \bar{x}_k : クラス k の平均ベクトル
- $x_{k,i}$: クラス k に属する対象 i のベクトル
- T : 行列の転置記号

と表される。このとき、

$$W^{-1}B \quad (3)$$

の固有値問題を解き、固有値 λ_i と固有ベクトル b_i を求める。ここで、入力特徴ベクトルの次元数を N とすると、

$$\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_N \quad (4)$$

であり、これらに対応する固有ベクトル b_i を基底として張られた判別空間の軸の寄与率になっている。本研究では、この固有値の累積寄与率が 99.9% に達した時点で判別空間の次元の上限 (M 次元とする) とした。

$$0.999 \times \sum_{i=1}^N \lambda_i \geq \sum_{i=1}^M \lambda_i \quad (5)$$

よって、判別空間の最終的な次元数を L とすると、 M とクラス数 J との関係は以下となり、

$$L = \min(J - 1, M) \quad (6)$$

また入力される N 次元の画像特徴ベクトルの L 次元判別空間への変換は、

$$\begin{matrix} L \times 1 & L \times N & N \times 1 \\ y & = [b_1 b_2 \dots b_L]^T & x \end{matrix}$$

となる。

3.5 種別の自動識別過程

以下の識別過程は、すべてシステムにより自動処理される。

- (1) 時刻 t に背景差分により切り出された対象 (識別番号 id とする) を、高さ、幅方向に 2 倍に伸長して平滑化する。
- (2) 平滑化後の画像から、3.4 節に述べた 11 個のデータを算出し、これに基づき 11 次元の画像特徴ベクトル ($x_s^{(id,t)}$: s は “shape”) を構成する。
- (3) $x_s^{(id,t)}$ を、画像が切り出された領域に対応する種別識別用の判別空間に射影する。
- (4) 射影後のベクトルとユークリッド距離の近い K 個の学習用ベクトル $y_{s(k,c)}$ を約 2,000 個の学習用サンプルベクトルから抽出する。ここで、“ c ” は $y_{s(k,c)}$ が属する種別 (class) を、また “ k ” は近いものから k 番目であることを意味する。
- (5) $x_s^{(id,t)}$ と $y_{s(k,c)}$ の間の距離の逆数を以下のようにな種別ごとにまとめ、それを $x_s^{(id,t)}$ に対す

本研究では $N = 11$ である。また判別空間の次元数は、固有値の累積寄与率を 99.9% と大きくとったため、Area A, B, C のいずれの判別空間においても $M = 10$ となった。この id は対象の切り出しとトラッキングのモジュールによって割り振られる⁸⁾ (2.2 節参照)。実験では $K=10$ とした。以下の K も同様。

る各種別の重み “weight” と定義する .

$$weight(\mathbf{x}_s^{(id,t)}, S_i) \quad (7)$$

$$\stackrel{def}{\leftarrow} \sum_{if\ c=i} \frac{1}{\|\mathbf{x}_s^{(id,t)} - \mathbf{y}_s^{(k,c)}\|} \quad (8)$$

すなわち, $weight(\mathbf{x}_s^{(id,t)}, S_i)$ は, $\mathbf{x}_s^{(id,t)}$ に対する種別 S_i の重み値となる .

- (6) こうして求めた重みのうち最大の値となった種別

$$S^{(id,t)} \leftarrow \max_i (weight(\mathbf{x}_s^{(id,t)}, S_i)) \quad (9)$$

を $\mathbf{x}_s^{(id,t)}$ の種別の識別結果とする . この識別結果を, その重み値とともに保存する .

$$class(id, t) \quad (10)$$

$$= \{S^{(id,t)}, weight(\mathbf{x}_s^{(id,t)}, S^{(id,t)})\} \quad (11)$$

4. 対象の色の推定

4.1 方針

人間の視覚には色に対する恒常性があり, 物体の色を, 強い直射日光下や日陰といった様々な環境下で安定に認識 (知覚) する . たとえば「赤い車」を日向でも日陰でもつねに「赤い車」と認識する . しかし, 実際にそれらの画像の色成分を比較すると, かなり異なることが多い (同一色に分類すべきデータの分散が大きい) .

そこで本研究では, 様々な条件下で撮影した色サンプルにオペレータが自らの印象に基づいて色ラベルをつけた後, ラベル付きの色サンプルから色特徴ベクトルを算出して色の推定用の線形判別空間を構成した . 判定過程では, 判別空間内で重みつき K 最近傍法による判定を行った .

図 4 に, 晴天時の日向と日陰における同一対象の色の「見え」の変化例を示す . 図に示されるように, 同一対象であっても, 環境によってその色の「見え」は大きく変化する .

4.2 色の学習過程

- (1) 晴天時 (太陽光下, 日陰) もしくは曇天時の様々な時間の車両や人物のカラーテキストチャ画像のサンプルを, 晴天時/曇天時それぞれ別個に幅広く集める . 一般にサンプル画像には様々な色が含まれるため, それらからさらに単色の色サンプルを切り出す .



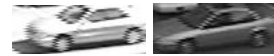
(a) 赤色 (小型搬送車)

(a) Red color (cart for workers)



(b) 白色

(b) White color



(c) 銀色

(c) Silver color

図 4 日向と日陰における色の「見え」の変化例

Fig. 4 Appearance differences between under sunlight and shadow.

切り出した色サンプルに対し, オペレータが自らの印象に基づいて, $\{\text{red-orange-yellow, green, blue-lightblue, white-silver-gray, darkblue-dark-green-black, darkred-darkorange}\}$ のいずれかのラベルをつける .

色サンプルの数は, 最終的に, 晴天時の映像から約 1,500, 曇天時の映像から約 1,000 となった . これらは晴天時の色サンプル, および曇天時の色サンプルとして別個に保存する .

- (2) 上記の色サンプル画像上の左上より, 幅, 高さ方向に $25 (= 5 \times 5)$ ピクセルを 2 ピクセルおきにサブサンプリングし, それらの色成分を (R, G, B) 表色系から以下に示す $(I1, I2, I3)$ 表色系⁹⁾に変換して平均する . これを, 各色サンプルの特徴ベクトル (3 次元ベクトル) とする .

$$I1 = \frac{(R + G + B)}{3.0} \quad (12)$$

$$I2 = \frac{(R - B)}{2.0} \quad (13)$$

$$I3 = \frac{(2.0 \times G - R - B)}{4.0} \quad (14)$$

- (3) 上記の各色サンプルの特徴ベクトルを, 晴天時のサンプルのベクトル, 曇天時のサンプルのベ

たとえば, FedEx 社の集配車の画像にはボディー色の白のほか, “FedEx” のロゴのペイント色である赤や青が含まれる .

$(I1, I2, I3)$ 表色系を用いたのは, 予備実験の結果, (R, G, B) 表色系もしくは (Y, I, Q) 表色系を用いた場合に比べ, 良好な識別結果が得られたことによる .

クトル, それぞれ別個に線形判別分析し, 晴天用と曇天用の2つの色の推定用判別空間を構成する.

- (4) 特定対象の画像の全ピクセルのデータを, 画像を収録したときの天候に応じた色の推定用判別空間に投影し, 全ピクセルに対してそれぞれ対応する(最も近傍の)色ラベルを求める. この結果を各ラベルごとに集計し, それらの値の比をデータベース化して保存する.

4.3 色の自動推定過程

以下の過程は, すべてシステムにより自動的に処理される.

- (1) 時刻 t に背景差分により切り出された対象の濃淡画像(識別番号 id とする)をマスクとし, 入力映像から対象 id のカラーテキストチャ画像を切り出す.
- (2) 上記の画像の全ピクセルの色成分を (R, G, B) 表色系から上述の $(I1, I2, I3)$ 表色系⁹⁾に変換し, 3次元の特徴ベクトル群 $x_c^{(id,t,n)}$ を構成する. ここで, $x_c^{(id,t,n)}$ の “ c ” は “color” を, また “ n ” は変換後のデータの n 番目であることを意味する.
- (3) $x_c^{(id,t,n)}$ を, データ収集時の天候に対応する判別空間に投影し, これを基準に K 個の近傍ベクトル $y_{c(n,k,c)}$ を学習用ベクトル(晴天用 1,500 個, 曇天用 1,000 個)から抽出する. ここで, $y_{c(n,k,c)}$ の “ k ” は近いものから k 番目であることを, また “ c ” は $y_{c(n,k,c)}$ が属する色グループ(class)を意味する.
- (4) $x_c^{(id,t,n)}$ と $y_{c(n,k,c)}$ の間の距離の逆数を以下のように色グループごとにまとめ, それを $x_c^{(id,t,n)}$ に対する各色グループの重み “weight” と定義する.

$$weight(x_c^{(id,t,n)}, C_i) \quad (15)$$

$$\stackrel{def}{\leftarrow} \sum_{if \ c=i} \frac{1}{\|x_c^{(id,t,n)} - y_{c(n,k,c)}\|} \quad (16)$$

ここで, $weight(x_c^{(id,t,n)}, C_i)$ は $x_c^{(id,t,n)}$ に対する色グループ C_i の重み値を示す.

- (5) 以上を全ピクセルに対して行い, 各ピクセルごとに算出した重み値を色グループごとに集計する. こうして求めた重みのうち最大となった色グループ,

$$C^{(id,t)} \leftarrow \max_i (weight(x_c^{(id,t)}, C_i)) \quad (17)$$

を $x_c^{(id,t)}$ の色の推定結果とし, その結果を重み値とともに保存する.

$$color(id, t) \quad (18)$$

$$= \{C^{(id,t)}, weight(x_c^{(id,t)}, C^{(id,t)})\} \quad (19)$$

5. 最終判定

5.1 判定過程

4章までに述べた処理は, 対象をトラッキング中に画像が切り出されるたびに施される処理である. 最終判定モジュールは, それらの処理結果を総合し, 最終的な判定を行う.

2.2節に述べたように, 切り出しとトラッキングのモジュールは, 入力映像中に新たな対象が現れると固有のIDをつけ, その動きをトラッキングする. 最終判定モジュールは, トラッキングシーケンスを通じて重みが最大となったクラスを種別・色グループそれぞれ独立に求め, 最終結果 x^{id} として出力する.

$$weight(x_s^{(id,t')}, S^{(id,t')}) \quad (20)$$

$$= \max_t (weight(x_s^{(id,t)}, S^{(id,t)})) \quad (21)$$

$$weight(x_c^{(id,t')}, C^{(id,t')}) \quad (22)$$

$$= \max_t (weight(x_c^{(id,t)}, C^{(id,t)})) \quad (23)$$

$$S^{id} = S^{(id,t')} \leftarrow class(id, t') \quad (24)$$

$$C^{id} = C^{(id,t')} \leftarrow color(id, t') \quad (25)$$

$$x^{id} = \{S^{id}, C^{id}\} \quad (26)$$

5.2 特定対象の判定

3.5節(4)に述べた, 種別の識別時に算出される上位 K 個の学習用ベクトルの中で, 特定対象のラベルを持つものの重みを特定対象ごとに集計する.

4.3節(4)で求めた各色グループの重み $weight(x_c^{(id,t)}, C_i)$ の比を算出する.

ここで集計後の重み値が閾値を超えた特定対象のうち, 最上位となったものの色のデータベース “ $db^{(id,t)}$ ” と, 上記の比の平均二乗誤差 “error” を求め, 誤差が閾値以下であれば, これを $x^{(id,t)}$ に対する特定対

この id は対象の切り出しとトラッキングのモジュールによって割り振られる⁸⁾ (2.2節参照).

閾値は実験に基づいて定めた.

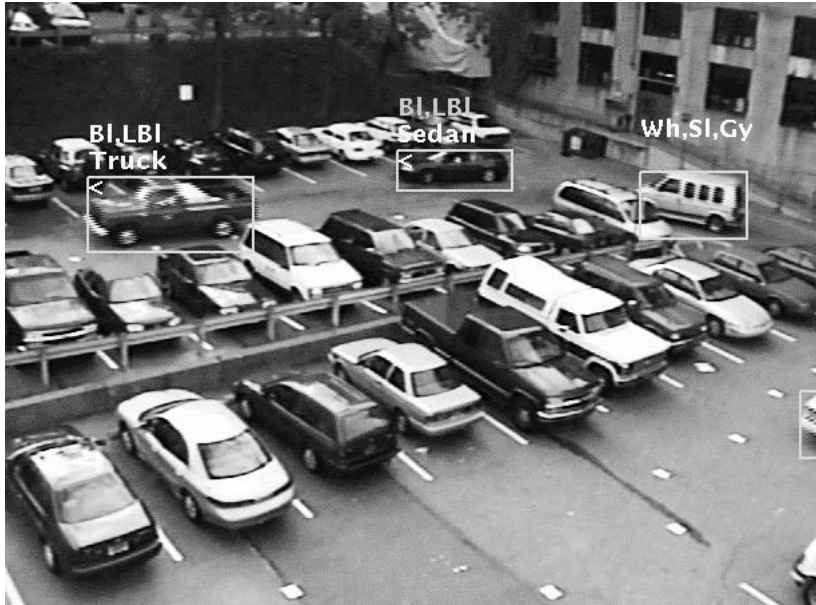


図5 処理画面の出力例

Fig. 5 Output example of processing screen.

象の判定結果 $Sp^{(id,t)}$ と判定する.

誤差が閾値を超えた場合, 上記の集計で第2位となった特定対象に対して同様の処理を行い, 誤差が閾値以下となれば, 第2位を特定対象と判定する. 第2位も閾値を超える, もしくは第2位の候補が存在しない場合は, 処理対象は特定対象ではないと判定する.

$Sp^{(id,t)}$ が求まった場合, その結果を上記の誤差値とともに保存する.

$$specific(id,t) \quad (27)$$

$$= \{Sp^{(id,t)}, error(\mathbf{x}^{(id,t)}, db^{(id,t)})\} \quad (28)$$

このうち, 誤差がトラッキングシーケンスを通じて最小となった特定対象を最終結果 Sp^{id} として出力する.

$$error(\mathbf{x}^{(id,t')}, db^{(id,t')}) \quad (29)$$

$$= \min_t (error(\mathbf{x}^{(id,t)}, db^{(id,t)})) \quad (30)$$

$$Sp^{id} = Sp^{(id,t')} \leftarrow specific(id,t') \quad (31)$$

6. 実験結果と考察

6.1 処理画面の表示例

ある時刻 t における, 本システム稼働中の処理画面例を図5に示す. 各処理対象を囲むボックスの左上部に, 算出した結果, (1) 特定対象の可能性 $Sp^{(id,t)}$,

(2) 色グループ $C^{(id,t)}$, (3) 種別 $S^{(id,t)}$ がこの順で示されている.

ここで, 色の推定結果は画面上では以下の略号で示した.

Rd,Or,Yl: red-orange-yellow,

Gr: green,

Bl,Lbl: blue-lightblue,

Wh,Sl,Gy: white-silver-gray,

DBl,DGy,Bl: darkblue-darkgray-black,

DRd,DOr: darkred-darkorange

画面上の文字色は, 表示中の各処理結果の確信度(重み)の値に応じて「(大)赤 → 緑 → 白 → 表示せず(小)」と変え, オペレータの了解性を高めた(これらの文字色は, 図5, 7, 8においては赤は白, 緑は灰色で示される). 表示色変更の閾値は, 実験に基づいて定めた.

本システムの処理速度(対象のトラッキングと切り出し + 種別の識別 + 色の推定 + 特定対象の検出)は, 入力映像中に処理対象が1つの場合, 毎秒2~3サイクルであった(CPU Pentium III, 400 MHzの場合. 切り出された画像の大きさに依存). この処理速度は純粋にCPUの処理能力に依存しており, より高速のCPUを用いることにより, 高速化を図ることができる.

6.2 処理画面の解釈

図5に示す例の種別の識別過程においては, 画面左のトラックと中央のセダンには図1(a)のArea A

表 1 実験結果
Table 1 Experimental results.

	Human	Sedan	Van	Truck	Mule	Others	Total	Correct	%
Human	67	0	0	0	0	7	74	67	90.5%
Sedan	0	33	2	0	0	0	35	33	94.2%
Van	0	1	24	0	0	0	25	24	96.0%
Truck	0	2	1	12	0	0	15	12	80.0%
Mule	0	0	0	0	15	1	16	15	93.8%
Others	0	2	0	0	0	13	15	13	86.7%
Total							180	164	91.1%

用の、画面右のミニバンには Area B 用の種別識別用判別空間が自動的に適用される。画面の右下に後部のみ見える車両があるが、画面の端にかかっているため処理の対象となっていない。画面のいずれかの端にかかった対象は、全体像がとらえられていないことが多いため、処理から外すアルゴリズムとした。

識別結果は、図中のトラックと中央のセダンについては明瞭に全体像が得られているため良好で、正しい結果が文字色赤（図 5 では白）で示されている。しかし、画面右のミニバンは最も特徴的な車体前部が駐車車両の陰になったため、処理結果の重みが低くなり表示に至らなかったと推測される。

色の推定では、この映像を収録した日は曇天であったため、オペレータが曇天用の色の判別空間を選択して用いた。推定結果としては、トラックとミニバンでは色情報のデータが安定して十分に得られたものと考えられ、適切な結果が赤色（図 5 では白）で表示されている。これに対し、中央のセダンは緑色（図 5 では灰色）の表示がなされているが、これは車体に他車の色が映り込むなどして色情報が不安定になったためと考えられる。

なお、図 5 の左のトラックと右のミニバンはある程度の速度で走行しており、映像にインタレース方式特有のブレが見られるが、こうしたブレは学習用サンプル画像にも見られ、そのまま「見え」として学習させた。

6.3 実験結果と考察

6.3.1 実験結果

実験は、天候と太陽の位置の双方を考慮し、晴天/曇天の午前 9 時～10 時、1 時～2 時、午後 4 時～5 時にカメラからの映像をオンラインでシステムに入力して行った。よって入力映像にはいずれも学習サンプルはない。映像中に現れた処理対象の数は 180 となった。

実験では、式 (20) の双方がともに目視による結果と一致した場合を正答としてカウントした。

まず、対象をトラッキングシーケンスの各時刻において切り出した、個々の濃淡画像に対する識別率の平

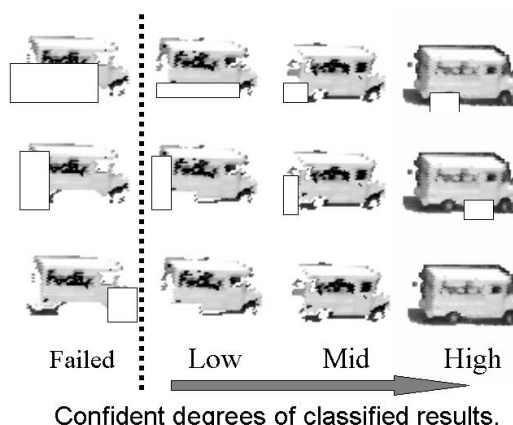


図 6 切り出しエラーと識別信頼度の関係の例（四角は人工的に加えた切り出しエラー）
Fig. 6 Examples between segmentation errors and classified confidences (Squares are artificial errors).

均は、81.2%であった。

次に本稿で提案する枠組みである、各時刻での種別の識別と色の推定の結果から、トラッキングシーケンスを通じてそれぞれ最も確信度が高かった結果を選択した場合、識別率の平均は 91.1%に向上した（表 1 参照）。

晴天と曇天では、曇天の映像の方が安定するため、若干結果が良かった。領域別では、図 1 (a) の Area A からの切り出し画像が Area B, C からの画像より若干良好に識別された。対象の向きは、識別結果にほとんど影響しなかった。

トラックの処理結果が他よりもやや低いのは、図 2 に示すように、トラックには様々な「見え」(形状)のものがある反面、通行量が少なく、学習サンプル数が十分に得られなかったためと考えられる。用いた手法の特性から、学習用サンプルを十分に補えば正答率は改善されると考えられる。

図 6 に、右向きに走るバンを切り出したデータと、その一部を人工的に削除したデータの例を示す。ここでは、了解性を高めるため背景差分の濃淡画像に対応するテクスチャデータを重ねて示している。最下段は

これらをオフラインで識別させた場合の識別結果の信頼度である．図 6 より，どの部位がどの程度削られると，識別の信頼度がどの程度低下するかが理解される．

図 6 最左列は正しく識別されなかったデータである．たとえば最左列最下段のデータは，図 5 の画面右のミニバンと同様，最も特徴的な個所であるフロント部分が削られたため，識別に至らなかったと考えられる．左から数えて第 2 列からは正しく識別され，右列のデータほど識別時の確信度が上昇した．これらの識別結果は，画面上の表示色としては (High → 赤)，(Mid → 緑)(Low → 白) で表示される．

6.3.2 考察

種別の識別や色の推定に失敗したケースとしては，以下のようなものが見られた．

<種別の識別> 対象どうしが重なった，学習サンプルにない対象が現れた(工用クレーン車など)，など．

<色の推定> 車体が太陽光を強く反射し，その部分が白色系の色として識別された，など．

本システムでは，対象が重なった画像に対しては以下のように対処した．

- (1) あらかじめ重なりのある画像を多く集めて学習用サンプルに含め，「その他」の種別とした．
- (2) 5.1 節の式 (20) の計算過程において， max 値の小さい「その他」の種別は処理対象から外すアルゴリズムとした．

図 7 に重なりのある画像の例を示す．

図 7(a) の画像は，画面右端から進入したバンと左から進入したセダンが重なった画像である．この例では，上記 (2) の理由により種別の識別結果の表示には至っていない(画面中の白い楕円の内部)．

図 7(b) の画像は同じくバンとセダンであるが，異なるシーケンスのもので，双方が重なっていない段階から統合的に切り出され(切り出しに失敗し)，信頼度低(白色)で Truck と判定されている．こうした誤認識は一般にその前後の対象ごとの識別結果が良好な場合には取り除くことができたが，そうでない場合には双方の対象が全体として誤認識されることもあった．

対象の影の影響については系統立てた実験は行っており，今後の課題であるが，図 2，図 4 の一部に見られるような影は処理対象として想定しており，学習データに含めている．

6.3.3 学習用/評価用データについて

本実験では，ある程度公共性のある場所に向けて据えたカメラからの入力映像を用いたので，入力映像中



(a) 車両が重なって観察された例 1
(a) A sample image of overlapped cars #1.



(b) 車両が重なって観察された例 2
(b) A sample image of overlapped cars #2.

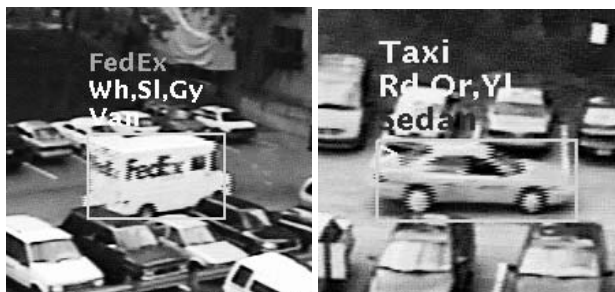
図 7 対象が重なった場合の処理例
Fig. 7 Examples of overlapped targets.

には一般道路と同程度の割合で量販車(頻出サンプル) やその他の車両(標準もしくは稀少サンプル) が見られた．

このうち，学習用データと評価用データに「まったく同一の車両」が含まれる割合は，20～30 台に 1 台程度であった．これに対し，双方のデータに含まれる「同一車種」の割合は，色違いの同じ車種を「同一」とする場合，幅広くデータを集めたため 6 割程度(10 台中 6 台程度)であった．一般に量販車はあらゆる時間帯で高い頻度で見られたため，それらの種別は比較的良好に(安定して)識別が可能であった．

しかしながら，全体として良好な識別率を得るためには，異なる環境の入力映像を用いることに，再学習を行うのが好ましいと考える．ただしその「手間」は，新環境で誤識別されたデータに対してのみ，それらを正しく識別するよう再学習するといった，微調整程度で済むものと考えられる．

なお今回用いた枠組みでは，対象は 11 次元の特徴ベクトルで表現したため，各データ(ベクトル)のサイズは数十バイト程度であり，数千のオーダのサンプルがあっても，データのサイズはたかだか数十キロバイトとなった．さらには学習法として線形判別分析を用いたため，学習にかかる時間は判別空間の構成に要する時間のみとなることから，全データを再学習したとしても，学習に要する時間はたかだか数秒であった．



(a) FedEx 社集配車

(a) FedEx Van

(b) タクシー

(b) Taxi

図 8 種別と色による特定対象の検出例：(a) あらかじめ設定した対象，(b) 種別と色から推定させた事前には設定していない対象（黄色のセダン：タクシー）

Fig. 8 Detection of specific target by type and color: (a) preset target, (b) non-preset target (yellow-sedan: taxi).

表 2 特定対象の検出結果

Table 2 Results of specific target detection.

	Detection Rate	False Alarm
FedEx Van	11/12 (91.0%)	3/88 (3.4%)
Mail Van	12/13 (92.3%)	7/87 (8.0%)
Police Car	16/17 (94.1%)	12/83 (14.5%)
Average	92.9%	8.5%

このように、扱うデータ量や、再学習や微調整の負荷が比較的少なく済む点も、本稿の提案手法の特長の 1 つと考える。

6.4 特定対象の検出

表 2 に特定対象の検出実験の結果をまとめ、図 8 に検出結果の表示例を示す。実験は、FedEx 社と郵便の集配車が午前と午後の集配時のみ現れるという制約から、数日にわたってこの時間の入力映像をビデオテープに記録して行った。入力した処理対象の数は、検出すべき特定対象を含め、100 個とした。

表 2 中、Detection Rate は特定対象画像を正しく検出した比率（正答数/特定対象の数）であり、False Alarm は非特定対象を特定対象と誤検出した数である。テスト画像は学習用サンプル画像には含まれていない。結果は False Alarm がやや大きい値となったが（特に Police Car）、Detection Rate はいずれも良好であった。これはセキュリティや監視などの観点から Detection Rate を優先したチューニングとしたためである。

郵便車の False Alarm は、実験場所付近の郵便車には図 3 に示すように 2 タイプあり、濃い赤系色のバンを、このうちの小さいバンとして誤検出したことなどによる。Police Car（パトカー）の False Alarm は、パトカーは図 3 に示すように白の大型セダンであっ

たため、他の白や銀色の大型セダンを誤検出したことによる。図 8(a) に、特定対象の 1 つとして設定した FedEx 社の集配車（Van）を検出した例を示す。

本実験では、さらに、あらかじめ学習させていない特定対象の検出も試みた。実験場所近辺のタクシーは一樣に黄色のセダンであったが、種別と色の識別結果を用いて red-orange-yellow 色のセダンをタクシーとして検出させたところ、Detection Rate で 95.2% の正答率を得た。図 8(b) に、このようにして検出されたタクシーの例を示す。

なお、「多様な特定対象を精度良く検出すること」を主たる目的とするならば、その目的により合致した方法論/アルゴリズムが別途考えうられる。

7. おわりに

本稿では、対面通行の一般道路映像中の、移動対象の識別とそれらの色の推定を行う手法を提案し、またそれらの機能を活用することによる、特定対象の検出の可能性を示した。

提案手法は、対象をトラッキングしつつ、各時刻で種別の識別と色の推定を処理結果に対する確信度つきで独立に行い、トラッキングシーケンスを通じてそれぞれ最も確信度が高かった結果を選択することを基本とする。対象の表現には「見え」ベースのアプローチを用い、また強い非線形性をともなって分布するデータの識別には、線形判別分析と重みつき K-最近傍法を組み合わせた手法を導入して対処した。

実験システムを構築して評価した結果、提案手法は不確定要因の影響を強く受ける屋外環境における、自動監視システムの実現に向けたアプローチの 1 つとして、可能性を有するものと考えられる。また提案手法は、

ビデオアーカイブの検索、タグづけといった分野などにも応用が可能と考える。

今回の実験では対象画像へのラベルづけはすべて人手によって行ったが、この過程は効率化が図れると思われる。たとえば、最初に各クラスの代表的な画像を学習サンプルとしてシステムに学習させ、システムには新たに学習サンプルにない画像を与えて識別させる。このとき誤識別されたものについてのみ、学習サンプルに加えて再学習させるといったことを繰り返せばよいと考える。

早朝と夕暮れの映像は今回は処理対象としなかったが、定性的な傾向として、太陽の高度が低いために影の影響が強く出ることや、対象が赤みがかかるといったことなどから、種別の識別・色の推定ともに不安定になった。

夜間については、高感度カメラと照明を併用することにより、本稿の提案手法はある程度機能するものと考えられる。

晴天/曇天以外の天候、特に降雨や降雪条件下での対象の識別や色推定については、降雨/降雪時の映像はホワイトノイズをかけたような映像になるといわれており、提案手法はある程度機能するものと考えられる。ただし、雨滴がレンズにつくと画像が歪むといった問題もあり、それらへの対策も別途必要である。

これらはいずれも今後の課題である。

謝辞 本研究の遂行にあたり多くの支援とコメントをいただいた、CMU VSAM Project メンバ各位、特に Robert Collins, Alan Lipton, Hironobu Fujiyoshi, Nobuyoshi Enomoto の各氏に感謝する。また多くの有益なコメントをいただいた査読者の方々に感謝する。

参 考 文 献

- 1) Koller, D., Weber, J., Huang, T., Malik, J., Ogasawara, G., Rao, B. and Russel, S.: Towards Robust Automatic Traffic Scene Analysis in Real-time, *Proc. International Conference on Pattern Recognition (ICPR94)*, pp.126-131 (1994).
- 2) Beymer, D., McLauchlan, P., Coifman, B. and Malik, J.: A Real-time Computer Vision System for Measuring Traffic Parameters, *Proc. Computer Vision and Pattern Recognition (CVPR97)*, pp.495-501 (1997).
- 3) Mantri, S., Bullock, D. and Garret, J. Jr.: Vehicle Detection Using a Hardware-Implemented Neural Net, *IEEE Expert*, pp.15-21 (Jan.-Feb. 1997).
- 4) Ikeda, T., Ohnaka, S. and Mizoguchi, M.:

Traffic Measurement with a Roadside Vision System — Individual Tracking of Overlapped Vehicles, *Proc. International Conference on Pattern Recognition (ICPR96)*, pp.859-864 (1996).

- 5) Jolly, M-P. D., Lakshmanan, S. and Jain, A.K.: Vehicle Segmentation and Classification Using Deformable Templates, *IEEE Trans. PAMI*, Vol.18, No.3, pp.293-308 (1996).
- 6) Kanade, T., Collins, R., Lipton, A., Burt, P., and Wixson, L.: System for Video Surveillance and Monitoring, *Proc. Darpa Image Understanding Workshop*, pp.3-24 (1998).
- 7) 影沢正隆, 上野信一, 池内克史, 柏木 寛: 並列処理ボード IMAP-Vision を利用した赤外画像による車両認識システム, 画像の認識・理解シボジウム講演論文集 II, pp.301-306 (2000).
- 8) Lipton, A., Fujiyoshi, H. and Patil, R.: Moving target classification and tracking from real-time video, *Proc. Workshop on Application of Computer Vision*, pp.8-14 (1998).
- 9) Ohta, Y. and Kanade, T. and Sakai, T.: Color Information for Region Segmentation, *Computer Graphics and Image Processing*, Vol.13, No.3, pp.222-241 (1980).

(平成 14 年 9 月 2 日受付)

(平成 15 年 5 月 6 日採録)



長谷川 修 (正会員)

1993 年東京大学大学院博士課程修了。博士(工学)。同年電子技術総合研究所入所。1999 年 6 月より 1 年間カーネギーメロン大学ロボティクス研究所滞在研究員。2001 年産業技術総合研究所主任研究員。2002 年 5 月東京工業大学大学院理工学研究科付属情報工学研究施設助教授。産業技術総合研究所脳神経情報研究部門併任。2002 年 11 月科技団さきがけ研究 21 兼任。パターン認識, 画像理解, マルチモーダルシステム等の研究に従事。電子情報通信学会, 人工知能学会, 日本認知科学会, 日本顔学会, IEEE-CS 等各会員。

**金出 武雄**

1974 年京都大学電子工学科博士課程修了(工学博士). 1976 年同助教授. 1980 年カーネギーメロン大学計算機科学科・ロボット研究所高等研究員. 1985 年同教授. 1992 年～2001 年カーネギーメロン大学ロボティクス研究所所長. 1993 年カーネギーメロン大学 U.A. and Helen Whitaker 記念教授. 1998 年カーネギーメロン大学 U.A. and Helen Whitaker 記念全学教授. 2001 年産業技術総合研究所デジタルヒューマン研究センターセンター長(非常勤). 計算機視覚, マルチメディア, センサー, 医用ロボット, 自律ヘリコプター等の研究に従事. National Academy of Engineering のメンバー, IEEE フェロー, ACM フェロー, American Association of Artificial Intelligence フェロー, C&C 賞, Joseph Engelberger 賞, JARA 賞等受賞.
