

CG モデル画像による教師画像を利用した車種認識の一手法

シャミラ モホッタラ 影澤 政隆 池内 克史

〒 153-8505 東京都目黒区駒場 4-6-1

東京大学生産技術研究所

{shirmi,kagesawa,ki}@cvl.iis.u-tokyo.ac.jp

本稿では、車両の3次元コンピュータグラフィックスモデル(以下CGモデルと呼ぶ)から作成した教師画像を用いたロバストな車種認識システムを紹介する。本システムの認識手法の基礎は、binary features法である。この手法は、局所的な特徴に注目した物体認識手法である固有窓法を元にしており、オクルージョンや照明条件の変動に対してロバストである。我々はこれまで、種々の実験を通してその有効性を確認してきた。特に、車両のCGモデルを利用して教師画像を作成することで、認識性能を損なうことなく教師画像の収集にかかる労力を大きく減少させることができた。しかしながら、binary feature法を基礎とする認識方法では、1車種を認識する性能は高いが、複数車種をきちんと認識させる場合は性能がやや落ちるといった問題点がある。本論文では、binary feature法に次の3つ手法を組み合わせることで、よりロバストな車種認識手法の提案を行う。3つの手法とは、(1)特徴量抽出手法の改良、(2)非存在領域の利用、(3)テンプレートマッチング、である。最後に、提案手法の有効性を、セダン、ワゴン、ハッチバック、ワンボックスの4車種を認識する場合を例にとり屋外実験にて確認した。

A Method for Vehicle Class Recognition Using CG Models

Shirmila MOHOTTALA Masataka KAGESAWA Katsushi IKEUCHI

Institute of Industrial Science, University of Tokyo

Komaba 4-6-1, Meguro-ku, Tokyo 153-8505

E-mail: {shirmi,kagesawa,ki}@cvl.iis.u-tokyo.ac.jp

This paper describes a robust method for recognizing vehicle classes. In our previous work, we have developed a vehicle recognition system based on local-feature configuration, which is a generalization of the eigen-window method. This system could recognize one vehicle class very accurately, but there have been limitations in recognizing several classes, when they are quite similar to each other. In this paper, we describe the improvements of our recognition system to distinguish four classes, namely sedan, wagon, mini-van and hatchback. The system requires training images of all target vehicle classes. These training images are easily created using a 3-dimensional computer graphic (3D-CG) tool. Using CG training images dispenses with much of the trouble of collecting real training images, and causes no effect on accuracy.

Keywords: *Vehicle Recognition, Computer Graphics, Eigen-window method*

1. はじめに

ITSにおける基礎的かつ重要な部分を担うテーマとしてセンシング技術がある。そのなかでも、路側カメラによる車両認識は次のような理由からもっとも期待されている技術である。路側カメラによる画像センシ

ングは、低コスト、広範囲のセンシング、速度や種類が判別可といった利点を有しており、さらに監視用モニタとしても使えることから盛んに研究されている。我々はこの画像センサに注目して、画像認識の手法を応用したロバストな車両認識システムを提案し、実験によりその実用性を示してきた [8]。

我々のシステムの特徴は、従来の車両認識システムが道路のモデルを持ち、道路との差分により車両を検出するのに対し、車両のモデルを持つことで認識を行うことにある。したがって、従来のシステムでは、車両の存在は認識できても、車両自体の詳細情報、すなわち、年式および型番といったレベルでの車種認識は不可能なものがほとんどであった。別アプローチとしてナンバープレート読み取り技術があるが、これにはプライバシーの問題から、車種認識への応用としては使いづらい。カメラを複数台利用して、横からの車両形状から車種を判定するシステムの提案もある [2] が、本研究では 1 台のカメラのみを用いるものとする。また、我々のシステムでは、特定車種 (n 年式 A 型車) といったレベルの認識から始めて、現在は、クラス認識、すなわち、セダン型乗用車を認識する、という研究を行っている。

我々のシステムはビューベースの認識システムであるので、あらかじめ多くの教師画像を利用してモデルを作成する。そのため、実用性の観点から有効な教師画像を多数収集する手間と労力が問題として挙げられていた。この問題を解決する手法として、車両の 3 次元コンピュータグラフィックスモデル (以下 CG モデルと呼ぶ) で教師画像を生成することを提案し、その有効性を示した。

以上の手法はセダン型乗用車の認識といった 1 つの車種やクラスを認識する場合には有効であったが、複数クラスの認識においては、あるクラス A の物体の一部が他のクラス B の物体の全部または一部と類似している場合に、クラス A の入力画像に対してクラス B の認識を応答する場合があります。この問題はビューベースである限り完全に解決することは不可能であるが、本論文では、この問題に対して、(1) 特徴量抽出手法の改良、(2) 非存在領域の利用、(3) テンプレートマッチング、の 3 手法を組み合わせることでクラス認識率を 85% 以上にあげることに成功したので、その手法を紹介する。

2. 車両認識手法

まず、本研究で用いた、屋外でもロバストな認識を行える binary feature 法について説明する。

本手法は次のようなビューベースの認識手法である：まず、あらかじめ認識対象車両について、いろいろな環境 (姿勢や昼夜、天候の違い等) の教師画像を多数用意し、これから認識対象車種のモデルを作成する。次に、入力画像各々について、このモデルとのマッ

グを投票を利用して行い、車両の車種及び入力画像中での位置を検出する。

Binary Feature 法 ([4]) は固有空間法 [1] から派生した固有窓法 [3] の応用と考えられる。これらの手法についての詳細は省略するが、固有窓法の著しい特徴として次のことに注意しておく。

- 認識対象物体の一部がオクルードされていても検出可能である。
- 認識対象物体の平行移動にロバストである。
- 認識対象物体領域を抽出する必要がない。

2.1 Binary Feature 法

Binary Feature 法は、上述の固有窓法をベクトル量子化を用いて二値画像に適用した手法であると考えられることができる。二値化エッジ画像を用いることにより、前述の特徴を損なうことなく、より安定した処理と計算の高速化を実現している。

原理は以下の通りである。まず、教師画像に対しガウシアンフィルタで太線化、ラプラシアンフィルタでエッジ検出を行う。次いで固有空間法の辞書ウィンドウと同様に特徴点 (binary features) 検出を行いウィンドウを設ける。二値画像における特徴点としては、近傍との相関性が低い点を選択する [5]。

選択された binary features は Lloyd のアルゴリズム [6] でコードに圧縮される。このようにして各 binary feature がどのコードに属するかを求めて、その対応するコードと教師画像中での位置を記憶しておく。

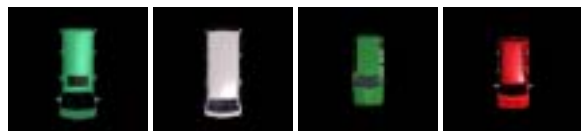
認識においては、入力画像を教師画像と同様に二値エッジ画像とし、全ての点にウィンドウを設け、コード化する (最も近いコードを求める)。コード化とは、モデル作成時に求めたコードに対して、入力画像中のウィンドウがどれに最も近いかを求めることである。固有窓法と同様に、入力画像中のウィンドウの位置と求めたコードに対応する binary feature の位置の相対座標に投票し、閾値でピークを検出して対象物体を認識する。

Binary Feature 法は固有窓法と同様に小さいウィンドウを用いているため、背景の変動やオクルージョン、画像中における物体位置の平行移動にも対応できる。また固有窓法に比べ計算が大幅に単純化されており、特に浮動小数点演算を行う必要がないため、処理は早くなる。また二値化エッジ画像を用いているため、グレースケールの画像を直接用いるよりもハイライト等の光源環境の変化にロバストである。屋外画像を扱う車両認識システムに適していると考えられるので我々の認識

システムにこの Binary Feature 法を用いることにした。



(a) セダン (b) ワゴン



(c) ワンボックス (d) ハッチバック

図.1 CG で作成した教師画像。

2.2 CG 画像の利用

前述のような認識手法では多数の教師画像を用意する必要がある。そこで、この教師画像収集の労力を低減するために、我々は CG モデルを利用して教師画像を製作する手法を提案している。近年の CG (コンピュータグラフィックス) 技術は非常に進んでおり、市販のモデリング、レンダリングツールを利用することでリアルリティの高い CG 画像を物体の形状や光源環境、カメラの視点などを自由に選択して簡便に製作することができ、また、教師画像中の物体と背景の区別も明らかである。よって、CG 画像を利用することにより、任意視点からの画像処理用テスト画像作成を非常に簡単に得ることができる。図 1 に実験で使用した CG モデルから作成した教師画像の例を示す。

このようにして得られた CG モデル画像でも、車両認識の教師画像として十分実用となることが示されている ([9])。

3. 提案認識システム

上述の手法で、特定の 1 車種クラスのモデルを作り、それを認識する場合は十分な性能が出ることが既に示されているが、複数車種のモデルを作成して同時に認識を行う場合は次のような問題がある。

- あるクラス A の物体の一部が他のクラス B の物体の全部または一部と類似している場合に、クラス A の入力画像に対してクラス B の認識を応答する

以下、この問題を「部分問題」と呼ぶことにする。



図.2 「部分問題」。

車種認識の場合は、図 2 のような場合が、典型的なケースである。図 2 左においては、セダン型車両の後部トランクを除いた部分がワゴン型車両に類似しているため、そのトランク部分を除いた部分にセダン型車両があると誤認識している。同図右においては、ワンボックス車両の後部を除いた部分がより小さい車両であるハッチバックと類似しているために後者と誤認識した例である。ニューベースシステムの場合、1 枚の入力画像だけが与えられている時には、この「部分問題」は本質的に解決不可能である。しかしながら、この「部分問題」に対する応答をよくすることは可能である。以下では、(1) 特徴量抽出手法の改善、(2) 非存在領域の利用、(3) テンプレートマッチング、の 3 手法を組み合わせることで応答を改善できたのでその報告を行う。

3.1 特徴量抽出方法の改善

一般に、画像から特徴量を抽出する場合、それが連続な抽出関数であれば、特徴量はいくつかの極値付近に集中する傾向がある。あるクラスの特徴量が一部に集中すると、その部分と似た部分において「部分問題」が発生する可能性がある。したがって、教師画像に含まれる認識対象物体からの特徴量抽出は、なるべく広い範囲に渡って行われる方が望ましい。

また、クラス認識を行う場合は、教師画像中の認識対象物体の大きさが異なることが考えられる。窓の大きさは一定であるので、特徴窓の数を一定とすれば、対象物体の大きさに応じて特徴量として抽出される部分が大きく変化する (図 3)。特徴量抽出領域が少ない場合、結果として特徴量が偏って抽出されたのと同じことになるため、「部分問題」が起こりやすくなる。

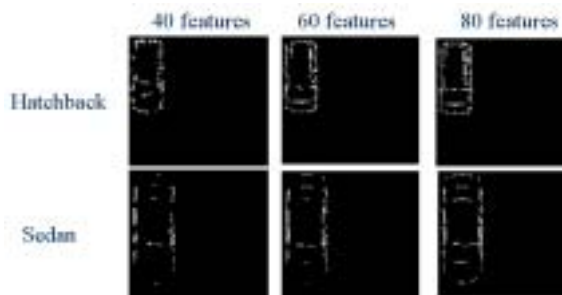


図.3 特徴量数と抽出部分。

以上の2点を解決するには、特徴量を広く浅く抽出するように変更する必要がある。具体的には、前者の問題に対しては、1つの特徴窓の回りには、他の特徴窓が近くに来ないようにする。後者の問題に対しては、物体の大きさに応じて抽出する特徴窓数を決定する。

以上をまとめて、次の2つの改善案を提案する。

1. $F(x, y)$ を点 (x, y) での特徴量、 F が (x_0, y_0) で最小値(最大値)をとる時、その近傍 $\{(x, y); |x - x_0| < d_x, |y - y_0| < d_y\}$ (d_x, d_y はしきい値)において、 F の値を次のように変更する: $F(x, y) = \max$ (または $F(x, y) = \min$), ただし、 \max, \min は、それぞれ F の取りうる最大値と最小値。
2. 教師画像 T 内に存在する認識物体の大きさが a 、取得する特徴窓数を n とする時、 $n = Ka + C$ (K, C は定数)として定める。

今回対象とした4つの車種クラス認識の場合、 $d_x = W_x = 7, d_y = W_y + 4 = 11$ (W_x, W_y は窓の大きさ)とし、 K, C の検討は行わず、セダン、ワゴンの場合に $n = 65$ 、ハッチバック、ワンボックスの場合に $n = 36$ と決定した(図4)。



図.4 特徴窓数の変更(左:ハッチバック, 右:セダン)。

3.2 非存在領域の利用

図5に、入力がセダン型車両、出力がワゴン型車両である典型的な「部分問題」が発生している例を示す。このような「部分問題」が発生する場合、小さい方のクラス(この場合はワゴン)の候補領域の外側は本来背景となるはずである。したがって、クラスとして認識された部分の外側が背景と大きく異なる場合、「部分問題」が発生している可能性があるとして判定する。ただし、車両や建物の影のように、当然外側に存在する、背景には含まれないエッジがでる可能性もあるので、物体の外側適当な範囲で、背景と異なるエッジが多数の場合に、仮説クラス B を棄却することにする。より正確には次のようになる。認識クラス候補がクラス B の場合、その教師画像中のマスク画像(物体がある部分で1ない部分で0)を BM とし、その BM の境界近傍マスク画像(エッジ画像) $M(b)$ を次のように定義する: $M(b)(x, y) = 1 \Leftrightarrow BM(x, y) = 0, d(x, BM) < b, d(y, BM) < b$

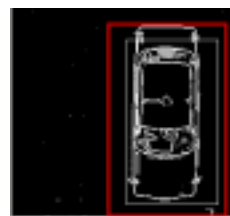


図.5 非存在領域の利用。

図5では、外側の長方形と内側の長方形に囲まれた部分が $M(b)$ となる。

このとき、 $M(b)$ エリアにおける入力画像と背景画像が大きく異なる場合に仮説「クラス B 」を棄却する。以上をまとめると次のようになる。

- 画像 BE を背景エッジ画像、入力画像のエッジ画像を IE 、認識クラスの候補を B 、クラス B の教師画像におけるマスク画像を BM 、 $M(b)$ を BM の境界近傍マスク画像とする。また、エッジ画像のハミング距離を D_H と書く。

この時、 $D_H(IE \text{ and } M(b), BE \text{ and } M(b)) > th_b$ (th_b はしきい値)ならば仮説「クラス B 」を棄却する。

3.3 テンプレートマッチング

今まで述べて来た認識アルゴリズムは、基本的に局所特徴量のみを利用したものである。局所特徴量だけに注目しているだけでは「部分問題」から逃れることは困難である。そこで、認識対象クラスの大域的性質であるテンプレート(エッジ画像)を用意し、これまでのアルゴリズムでクラス候補と判定された部分にそのクラスのテンプレートと比較することで大域的な性質を反映した認識結果とするものである。まとめると、次の処理となる。

- 入力画像のエッジ画像を IE 、認識クラスの候補を B 、クラス B の教師画像におけるマスク画像(物体がある部分で1ない部分で0)を BM 、クラス B のテンプレートエッジ画像を TB とする。この時、 $D_H(IE \text{ and } BE, TB) > tm$ (tm はしきい値)ならば、仮説「クラス B 」を棄却する。

図6に、テンプレートの例を示す。

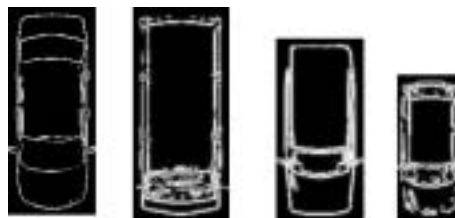


図.6 テンプレートの例。

3.4 提案認識システム

以上の処理を加えて、最終的に以下のようなシステムを構築した。

認識システムは、教師画像として CG レンダリングツールで製作した CG 画像を教師画像として利用し、Binary Feature 法を用いてモデル作成、認識処理を行う。また、高速化を図るため、並列画像処理ボード IMAP vision を利用し、その上に実装した。

我々の認識システムでは、入力画像に対して背景差分などの前処理がなくても機能するが、今回の実験では、エッジ画像の背景差分を行った。この理由は、認識対象とした画像が道路上から真下を見下ろすものであり、したがって、エッジをとるとセンターラインを含む道路標識が多数直線として登場し、その結果、車両の特徴の一部として認識されることがわかったからである。本認識システムでは、差分により余計な部分が削られても、それはオクルージョンと同じことになるため、既に述べた固有窓法の著しい性質により認識には支障がない。

ただし、投票方法について、教師画像が CG であることを考慮し投票空間にも平滑化フィルタをかけた。

認識対象クラスとしてセダン、ワゴン、ワンボックス、ハッチバックを考え、次のようにしてクラス認識を行った。

1. 各クラスに対して CG の車両を 6 台ずつ用意し、それぞれのモデル（したがって合計 24 台分のモデル）を用意する。
2. 各教師画像からエッジを抽出し、特徴量抽出を行う。
3. [特徴量抽出方法の改善] 特徴量抽出の際、モデルの大きさに応じて特徴窓の数を変え、かつ、特徴窓同士が x 方向で $W_x = 7$ ピクセル以上、 y 方向で $W_y + 4 = 11$ ピクセル以上離れるように抽出する。
4. 抽出された特徴量から binary feature 法により圧縮を行い、クラス認識用データベースを作成する。
5. 入力画像を 1 枚取得する。
6. 入力画像に対し、教師画像と同様の方法でエッジ抽出、特徴量抽出を行う。
7. ただし、特徴量の数および特徴窓間の距離については制限しない。
8. こうして得られた入力画像中の特徴窓と各クラス認識用データベースのモデルの特徴窓とでマッチングをとり投票を行う。
9. 投票数が閾値以上の場合に、類似度の高い車両とその属するクラスを候補として抽出する。

10. [非存在領域の利用] 候補として得られた類似度の高い車両について、その境界外側の適当な範囲 $M(b)$ において、エッジ量がしきい値以上の場合にその車両とクラスを候補から外す。
11. [テンプレートマッチング] 残った類似度の高い車両のクラスが 2 つ以上ある場合、その上位 2 つのクラスの車両のテンプレートと入力画像を比較し、類似度が近い方を最終的に採択するクラスとして決定する。

4. 屋外実験

屋外で撮影された実画像を入力画像とし、提案手法である CG モデルから作成した教師画像を利用した車両認識システムの性能を検証した。本実験では、カメラは道路上方から真下を見下ろす画角で設置されているものと仮定した。この画角を採用した理由は ETC (自動料金収受システム) やロードプライシングなどにおいて、地上側通信機器が意図している車両と通信しているかを確認するために詳細な車両情報が必要と思われるからである。

こうした画角の場合、本手法の利点は、影 (車両及び建物の影) によるノイズ発生にロバスト、車両の前後が入力画像から切れていても検出可能、走行位置によらず検出可能、として反映される。

実際の実験では、カメラを歩道橋上に真下に向けて設置し入力画像を撮影した。CG から製作した各クラスの教師画像の一部は図 1 に示した通りである。

表 1 に、クラス認識比較対照表を示す。上から順に、「部分問題」について何も対策を施さなかった場合、特徴抽出改善のみを行った場合、特徴抽出改善と非存在領域を利用の 2 つのみを施した場合、最終結果、を示す。

Classified as	Real Class			
	Sedan	Minivan	Hatchback	Wagon
Sedan	7	0	0	1
Minivan	3	3	1	2
Hatchback	3	0	10	10
Wagon	9	2	2	3

(a) 「部分問題」について何も対策を施さなかった場合：正解率 35%

Classified as	Real Class			
	Sedan	Minivan	Hatchback	Wagon
Sedan	16	2	0	2
Minivan	0	3	0	0
Hatchback	0	5	11	1
Wagon	6	3	2	13

(b) 特徴抽出改善のみを行った場合：正解率 65%

Classified as	Real Class			
	Sedan	Minivan	Hatchback	Wagon
Sedan	19	0	0	2
Minivan	0	7	0	0
Hatchback	0	3	11	0
Wagon	2	5	2	14

(c) 特徴抽出改善と非存在領域を利用の2つを施した場合：正解率 77%

Classified as	Real Class			
	Sedan	Minivan	Hatchback	Wagon
Sedan	19	0	0	1
Minivan	0	10	0	0
Hatchback	0	3	13	0
Wagon	2	2	0	13

(d) 最終結果：正解率 86%

表 1 認識実験結果 .

「部分問題」対策を行うにつれ、だんだんと誤認識が減少していく様子が見て取れる .

認識実験の様子を図 7 に示す (今回提案手法) . 位置も含めて車両を正確に検出した場合を成功としている . 結果画像には、検出した車両が長方形で示されている .

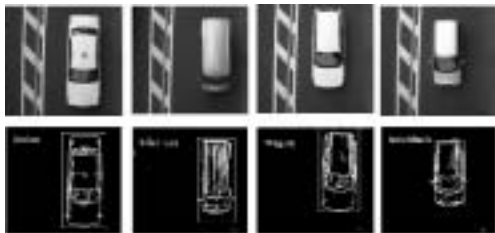


図.7 実験結果画像例 . 上:入力画像, 下:認識結果 .

なお、参考までに図 8 に認識に失敗した場合を示す . 左端の例は、セダンをワゴンと誤った例であるが、両クラスの主な違いは後部トランクがあるかどうかであるので、このような画像から両者を区別するのは原理的に不可能である . 真中の例は、エッジ画像からだけでは、人間でもハッチバックとワンボックスを区別するのは困難である . 右端の例は、逆に、あまりエッジがでなかったため、本システムの特徴である、隠れがあっても検出可能という性能が裏目に出て、ほぼ同じ大きさのハッチバックとして認識された例である .

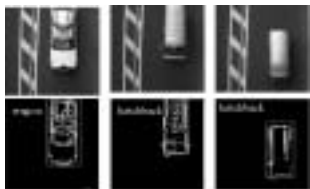


図.8 誤認識の例 . 上:入力画像, 下:認識結果 .

5. まとめ

CG モデルから教師画像を作成するシステムでも、十分な精度で4つの車種クラス認識が可能であることが示された . 特に、あるクラスの一部が別のクラスの全部または一部と似ている場合に誤認識をする問題 (ここでは「部分問題」と名づけた) について、3つの改善策を施し、その有効性を屋外実験を通じて確認した .

現在のシステムは入力として静止画が1枚与えられることを仮定しているが、交通監視画像など、動画として得られる場合は、同一車両について数枚の入力画像が得られることが期待でき、それらの処理結果に基づいてよりロバストなクラス認識ができると考えられる .

参考文献

- [1] H. Murese and S. K. Nayar, "Visual Learning and Recognition of 3-D Object from Appearance," International Journal of Computer Vision, vol.14, no.1, pp.5-24, Jan. 1995.
- [2] 大山宏, 村上伸一, "交通車輛の車種自動識別法に関する一検討", 画像電子学会誌, 第 25 巻, 第 5 号, pp.486-494, 1996.
- [3] K. Ohba and K. Ikeuchi, "Detectability, Uniqueness, and Reliability of Eigen-Windows for Stable Verifications of Partially Occluded Objects," IEEE Pattern Recognition and Machine Intelligence, vol.19, no.9, pp.1043-1048, Sept. 1997.
- [4] J. Krumm, "Object Recognition with Vector Quantized Binary Features," Proc. IEEE Computer Vision and Pattern Recognition '97, pp.179-185, June 1997.
- [5] J. Shi and C. Tomasi, "Good Features to Track," Proc. IEEE Computer Vision and Pattern Recognition '94, pp. 593-600, June 1994.
- [6] R. M. Gray, "Vector Quantization," IEEE ASSP Magazine, pp. 4-20, April 1984.
- [7] 佐藤いまり, 佐藤洋一, 池内克史 "物体の陰影に基づく光源環境の推定" 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「Physics-based VisionとCGの接点」特集号, Vol. 41, No. SIG 10 (CVIM 1), pp. 31-40, December 2000.
- [8] 吉田, 影澤, 塘中, 池内 "局所特徴認識アルゴリズムによる車両の認識" 電子情報通信学会 PRMU 研究会, 101, 302, pp. 9-14, Sep. 2001.
- [9] 吉田, モホッタラ, 影澤, 池内 "CG モデル画像による教師画像を利用した車両認識の試み", 第 1 回 ITS シンポジウム, December 2002.