

統計的推論に基づく景観認識の検討

笹木 美樹男

株式会社デンソー 基礎研究所

〒470-0111 愛知県日進市米野木町南山 500-1

あらまし 我々は車載カメラ映像について、ブロック画素単位の色判定と3次元モデルベース認識及びルールベース推論を用いて主要オブジェクトの位置関係やシーンの特徴を把握し、XML形式の映像記述を生成する方法を提案してきた。本論文では更に、オブジェクトの判定規範を確信度ベクトルという概念で統一し、過去の映像をもとに作成した教示データから統計的に推論する手法を提案する。

キーワード インデキシング、確信度、統計的推論、多変量回帰分析

Investigation on Scene Recognition on the basis of Statistical Reasoning

Mikio SASAKI

Research Laboratories, DENSO CORPORATION

500-1 Minamiyama, Komenoki-cho, Nisshin-shi, Aichi, 470-0111 Japan

Abstract We have been proposing a method to acquire three dimensional relative locations between major objects and scene features from on-board captured images, so as to generate scene description in XML format automatically. This method is based on color indexing of block pixels, three dimensional model based localization, and rule based reasoning. This paper further presents a statistical reasoning method to estimate confidence vectors that universally express probabilistic interpretation of each block pixels on the basis of training data extracted from past images.

Keyword Indexing, confidence, statistical reasoning, multi-variate regression analysis

1. まえがき

我々は、車両周辺の景観画像中に出現しうる複数のオブジェクトをあらかじめ認識辞書中の有限個の語彙として登録し、画像をブロック画素単位で色解析することで辞書中の語彙に対応付け、景観を概略記述する方法を提案した[1][2]。これは従来の画像理解手法が専ら3次元形状の獲得や領域分割に注力し、意味へのマッピングに関しては余り報告されていないことに起因する。我々は人間が視野を直感的に理解するてがかりは、コンテキストと色特徴、そしてその色が存在する2次元位置であると考えた。一方で、色と画面上の2次元位置をキーとするオブジェクトの認識は標識や信号機、先行車両など特定の対象については行われているが、それらはルールベースによる判断が大半であった。ところが、ルールベースによる認識ではシステム設計者が経験的に設定するパラメータが多くなるという弱点がある。しかも、ルールで記述できる判断条件は直感的なおおまかな特性には向いているが、細かい特性を記述する際には、煩雑で柔軟性のないルール群を増やしてしまうことも多い。

そこで本稿では、車載カメラ映像をもとに周辺環境を統計的に把握する方法を提案する。特に、サンプル画像をもとに画像特徴を学習し、確信度ベクトルを統計的に推論することで景観認識を試みる。

2. 確信度ベクトルによる環境認識

2.1. 確信度ベクトル

いま、画像中のある画素領域に認識結果として付与するインデックスの確からしさを確信度と呼び、インデックス A_i ($i=1, \dots, N$) に対する確信度を C_i で表す。このとき、

$$C = (C_1, C_2, \dots, C_N)^T \quad (1)$$

を確信度ベクトルと定義する。これにより、 N 個のオブジェクトのどれに相当するかという認識問題を確信度ベクトル C の推定に帰着させることができる。これは最終的に推定された C の最大成分 C_k に相当するオブジェクトが画素領域に対応するオブジェクトであると判定することで可能になる。

いま、ある認識処理が M 個の部分的な認識処理で構成されるとすると、

$$C = c_0 + \sum_{m=1}^M a_m c_m \quad (2)$$

といった部分ベクトル c_m の線形荷重で総合的な確信度ベクトルを算出することが考えられる。これはあるプロセスに含まれる部分的な認識処理のみならず、過去の認識結果や他の観測者の認識結果、あるいは他の手法による認識結果さえも初期値や部分処理として併用し、統合的に推定できることを意味している。

2.2. 確信度の線形回帰モデル

確信度ベクトルはさまざまな状況要因に影響される。そこで本稿では状況要因と確信度ベクトルの因果関係を次のような多変量線形回帰モデルで表現する。すなわち、

$$C = c_0 + \sum_{m=1}^M W_m s_m \quad (3)$$

とする。ただし、

c_0 : 初期の確信度ベクトル (N 次元列ベクトル)

W_m : 要因 s_m に対する回帰係数の組を表す N 行 L_m 列の行列

s_m : m 番目の要因を表す L_m 次元列ベクトル ($m=1, \dots, M$)

であり、更に、 $i \neq j$ ならば s_i と s_j は無相関であると仮定した。

上記の状況要因は大きく分けて二つのクラスで構成される。一つはシーンの状況であり、カメラが撮影した時間や場所、天候などがそれに相当する。そこで数量化理論 I 類に基づき、後述のようにダミー変数を導入してこれらの状況要因を数値表現する。

もう一つの状況要因クラスは画像特徴であり、撮影された画像中のあるブロック画素の色、動き、2 次元位置、推定される 3 次元位置などがそれに相当する。

3. 統計的推論手法

3.1. 車両周辺画像の特異性

一般に走行中の車両における車載カメラは道路路上にあり、安全運転支援の場合はまず車両進行方向を視野とすることが多い。一方で、道路周辺に存在する主要オブジェクトはいくつかのシーンクラスに分類して考えれば限定できることが多い。各シーンクラスにおいてもオブジェクトには静的なものと動的なものが存在する。しかもオブジェクトは完全に視界から消えたり、全く新たに出現するというケースが多々存在する。特に車両や歩行者などがその事例となる。また、天候や照明の変化、撮影方向の違いもある。更にはカメラの設置位置や車両による違いもある。

したがって景観画像の認識では、逐次的学習によって認識率が向上することに加えて、認識可能なオブジェクトの種類が増えていくことが期待される。また、入力画像から得られた特徴空間をある境界面で切っていくというクラスターリング手法では、判定結果が 1 か 0 になる上に、複数のカテゴリへの帰属度を表現できない。

3.2. 統計的推論規則の獲得

そこで我々は、特徴空間を分割して判別規範を求め

るのではなく、複数の解釈結果に対する確信度で構成される空間 C と入力画像や付随情報から得られる特徴空間 S との間の回帰関係 W を求めることを考えた。すなわち、サンプル映像中の個々のブロック画素に対する解釈結果を確信度として人手で教示し、多変量回帰分析にかけることで(3)式回帰係数行列 W_m を推定する。例えば、最小二乗法を適用すれば、 W_m^T の推定値 B_m は次式で計算できる。

$$B_m = (S_m^T S_m)^{-1} S_m^T Y \quad (m=1, \dots, M) \quad (4)$$

ただし、

$$S_m = [s_m(1) \dots s_m(k) \dots s_m(K)]^T \quad (5)$$

であり、 $s_m(k)$ は時刻 T_k ($k=1, \dots, K$) における s_m の値とする。また、 Y は確信度の履歴を表す行列であり、

$$Y = [Y_1 \dots Y_n \dots Y_N] \quad (6)$$

である。ここで、 Y_n は n 番目の語彙に対する確信度の履歴をあらわす K 次元列ベクトルであり、

$$Y_n = (y_n(1), \dots, y_n(k), \dots, y_n(K))^T \quad (7)$$

である。更に、 $y(k)$ は時刻 T_k における C の観測値を表す N 次元列ベクトルであり、

$$y(k) = (y_1(k), \dots, y_n(k), \dots, y_N(k))^T \quad (8)$$

である。

したがって、さまざまな状況要因の組み合わせに対して確信度ベクトル C を算出する推定式が決定できる。この C の最大成分に相当するインデックスを一次の認識結果とする。

3.3. 状況ベクトル

(3)式をより具体化すると、時刻 T_k における画像中の m_r 行 m_c 列に位置するマクロブロック $MBK(m_r, m_c, k)$ ($m_r=1, \dots, M_r, m_c=1, \dots, M_c$) が何のオブジェクトであるかを示す確信度ベクトル $C(m_r, m_c, k)$ は状況と画像特徴を説明変数として推定でき、多変量線形回帰モデルでは、

$$C(m_r, m_c, k) = W_S S_S(m_r, m_c, k) + W_P S_P(m_r, m_c, k) \quad (9)$$

と表現できる。ただし、

$S_S(m_r, m_c, k)$: $MBK(m_r, m_c, k)$ に関するシーンの状況を表す L_S 次元列ベクトル

W_S : 要因 S_S に対する回帰係数の組を表す N 行 L_S 列の行列

$S_P(m_r, m_c, k)$: $MBK(m_r, m_c, k)$ に関する画像特徴を表す L_P 次元列ベクトル

W_P : 要因 S_P に対する回帰係数の組を表す N 行 L_P 列の行列

である。この中で、状況ベクトル S_S は更に分解でき、例えば下記で構成される。

$$S_S(m_r, m_c, k) = (S_{nenv}^T S_{location}^T S_{time}^T)^T \quad (10)$$

ただし、

$S_{nenv}(m_r, m_c, k)$: $MBK(m_r, m_c, k)$ に関するシーンの自然環境情報を表す L_{nenv} 次元列ベクトル

$S_{location}(m_r, m_c, k)$: $MBK(m_r, m_c, k)$ に関するシーン

の位置情報を表す $L_{location}$ 次元列ベクトル

$S_{time}(m_r, m_c, k)$: $MBK(m_r, m_c, k)$ に関するシーンの時間情報を表す L_{time} 次元列ベクトル

である。

これらの状況ベクトルは色情報のようにそのまま数値で表現できるものもあるが、それ以外の代数演算が意味を持たないような情報では数量化理論 I 類を用いて次のように表す。

$$S_{time} = (\text{早朝, 朝, \dots, 夕, 夜})^T \quad (11)$$

$$S_{env} = (\text{晴, 曇, 雨, \dots, 暴風雨})^T \quad (12)$$

ただし、各状況ベクトル中の {早朝, 朝, ..., 晴, 雨, ...} などは各事象が真のとき 1, そうでないとき 0 を設定するものとする。これは各状況要因の確信度を 0 と 1 の間で正規化して表現したことに相当し、以後、代数的演算が可能になる。

4. 実験

4.1. 実験システム

市販デジタルビデオカメラをカーナビディスプレイ位置の背後に設置し、フロントガラス越しに前方の道路状況を撮影した。それぞれで撮影した DV 形式の映像音声で PC 上で MPEG-1 形式に変換し、実験システムの入力とした。今回は MPEG-1 のイントラフレームのみ認識対象としたため、資料映像に対して 1 秒間に 2 フレームのフレームレートで処理を行った。

4.2. 計測行列の作成

いま、状況行列 S と確信度行列 Y を用いて、

$$Q = [SY] \quad (13)$$

を計測行列と呼ぶことにする。ただし、

$$S = [s(I) \dots s(k) \dots s(K)]^T \quad (14)$$

であり、 $s(k)$ は時刻 T_k ($k=1, \dots, K$) における s の値とする。また、 Y は (6) 式と同様に確信度の履歴を表す行列である。ここで、 Y に含まれる確信度はサンプル映像中の各ブロック画素が何であるかを目視で判断し、教示することで獲得される。すなわち、その判定結果であるオブジェクトインデックスに対応する成分の値を 1 とし、それ以外は 0 と設定する。この計測行列から (4) 式の回帰係数行列を算出すれば、新たな入力映像中にブロック画素に対する確信度ベクトルを推定することができる。

今回 (3) 式における M を

$$M = I \quad (15)$$

とし、 s_j を s として、

$$s = (S_p^T S_s^T)^T \quad (16)$$

と定義する。ただし、

$$S_p = (S_{color}^T S_{block_position}^T)^T \quad (17)$$

$$S_{color} = (I_Y, I_U, I_V)^T \quad (18)$$

であり、 S_{color} は $MBK(m_r, m_c, k)$ に関するシーンの色

情報を表す L_{color} 次元列ベクトルである。(18) 式はブロック内平均色であり、DCT 係数ブロックの DC 成分に対応するが、テクスチャ情報に対応する AC 成分も含めるように次元を拡張できる。なお、今回は次元数が比較的少ないため、(9) 式のような分解は行わず、

$$C = Ws \quad (19)$$

とした。ただし、 W は N 行 L 列の行列であり、

$$L = L_p + L_s \quad (20)$$

である。

4.3. 認識率

確信度ベクトルの最大成分である最大確信度を求めると認識結果と判定すべきオブジェクトのインデックス $X(m_r, m_c, k)$ を決定できる。これと予め人手で求めておいた正解 $G(m_r, m_c, k)$ とを比較して一致するならば $MBK(m_r, m_c, k)$ の認識結果 $R(m_r, m_c, k)$ について、

$$R(m_r, m_c, k) = 1 \quad (21)$$

と表記する。一致しないならば、

$$R(m_r, m_c, k) = 0 \quad (22)$$

とする。これに基づき、

$$A = (100/(M_r M_c)) \sum_{i=1}^{M_r} \sum_{j=1}^{M_c} R(i, j, k) \quad (23)$$

を画面全体の認識率 (%) と定義する。

4.4. 認識実験の概要

認識実験に際しては 6 種類の資料映像 (各 5 分程度) から学習映像 I_{learn} と評価映像 $I_{estimate}$ を選び、学習映像による教示をもとにして評価映像の認識を行った。

ある映像 I を撮影した時間帯 $T(I)$ と場所 $P(I)$ の組み合わせという観点から、次の 4 種類の実験を考えた。

(Exp.1) $T(I_{learn}) = T(I_{estimate})$ かつ $P(I_{learn}) = P(I_{estimate})$

(Exp.2) $T(I_{learn}) = T(I_{estimate})$ かつ $P(I_{learn}) \neq P(I_{estimate})$

(Exp.3) $T(I_{learn}) \neq T(I_{estimate})$ かつ $P(I_{learn}) = P(I_{estimate})$

(Exp.4) $T(I_{learn}) \neq T(I_{estimate})$ かつ $P(I_{learn}) \neq P(I_{estimate})$

4.5. 教示データの作成

認識率算出のためには映像の各イントラフレームについて正解のインデックスを 330 個付与する必要がある。5 分の映像ならば 198,000 個となる。これは、教示のための計測行列を作成する際も同様であり、事実上困難となる。そこで、今回、重要なオブジェクト (車両, 道路, 空, 緑地, 信号, ビル, 高架, など) に絞って教示データを作成し、徐々に追加するという段階的な教示を行った。

5. 考察

(Exp.1) の結果の一例を図 1 に示す。ここでは、高架が前方に見えるシーンについて、教示データを増やすと徐々に道路や先行車両, 高架, 街路樹が識別されていく様子がわかる。

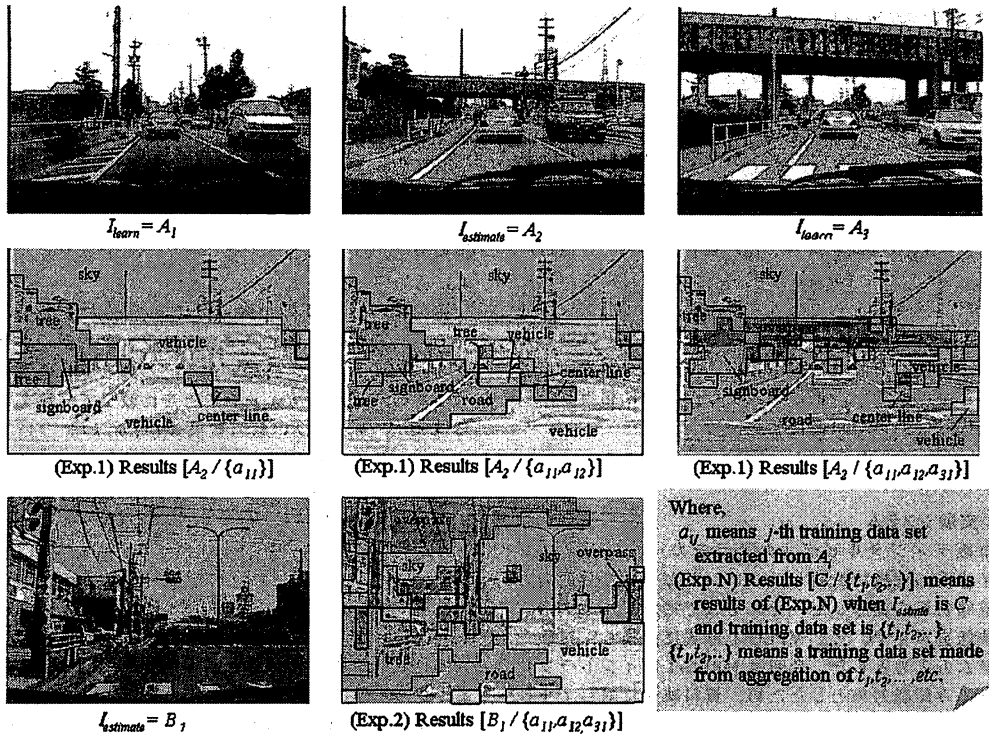


Fig.1 Experimental Results.

(Exp.2)の結果の一例を図1に示す。ここでは、特にクラスが異なるシーンを選択したために、学習映像の中に存在しないオブジェクト(例えばビル)が評価映像中に存在する場合は、認識結果は正解になりえない。

(Exp.3)では時間帯は照明や動的オブジェクトの発生状態に影響する。例えば、昼と夕方の違いがあるため誤認識が増える。

(Exp.4)では、オブジェクトの構成が類似しかつ照明が著しく異ならなければ認識率はさほど劣化しない。全体を概観すると次のような傾向が推察された。

- ・シーンクラスが同じであれば学習効果は高い。
- ・天候や時間帯の違いにより照明やオブジェクトの構成が変化する場合、認識率は著しく低下する。
- ・場所、天候、時間帯のすべてが同じである場合が最も認識率が高い。ただし、季節による景観変化(影の長さや樹木の色など)の影響は認められた。

6. むすび

サンプル映像の学習に基づく統計的推論によりブロック画素単位の確信度ベクトルを推定する認識手法を提案した。今回は目視により定性的な認識性能を確認する程度に留めたが、今後はさまざまなシーンにつ

いて定量的に認識率を評価していきたい。また、多変量回帰分析以外の認識器を用いた場合との比較も行っていく必要がある。

システム全体として考えれば、ルールベースの手法と今回提案した統計的推論の手法を組み合わせ、更にブロック間の連結性の判断や時系列分析を併用すれば認識率は向上すると思われる。一方で、従来の応用特定型の認識モジュール(歩行者、車両、標識など)との統合やプローブシステムによる分散協調の効果についても考察していきたい。

文 献

- [1] 笹木, 難波, "ITSにおける知的情報支援とそれを支える視覚通信複合技術 - 知識共有型の環境認識をめざして -", JSAI 2006 1E2-2 (2006).
- [2] 笹木, "色情報と知識処理に基づく車載カメラ映像のインデキシング", 情報処理学会研究報告, 2006-AVM-54 (2006).