

8.	シーンクラスと景観認識	エラー! ブックマークが定義されていません。
8.1	はじめに	2
8.2	実験	3
	8.2.1 交差学習による自動インデキシング	3
	8.2.2 CSC の自動生成	4
8.3	考察	5
8.4	風景解析への応用	5
8.5	実験	7
	8.5.1 学習データの生成	エラー! ブックマークが定義されていません。
8.6	教示データの作成	エラー! ブックマークが定義されていません。
8.7	認識実験	エラー! ブックマークが定義されていません。
8.8	考察	7
8.9	MPEG-Based Vision の実現に向けて	エラー! ブックマークが定義されていません。
	8.9.1 境界部分の検出	11
	8.9.2 周期的変化の検出	11
8.10	考察	11
8.11	おわりに	12
9.	予測と認識の相互作用	13
9.1	はじめに	13
9.2	めざすシステム	14
	9.2.1 予防安全	14
	9.2.2 認識器の自動適応	14
	9.2.3 渋滞予測や車両認識との比較	15
9.3	過去の統計に基づく予測	16
	9.3.1 予報に至るプロセス	16
	9.3.2 特徴ベクトルの定義	16
9.4	実験と考察	17
	9.4.1 実験条件	17
	9.4.2 予測性能の評価	17
9.5	認識率に対する変化	17
9.6	おわりに	17

8. シーンクラスと景観認識

計算機にシーンを解釈させる場合，認識対象となる有限個のシーンクラスを事前に定義しておくことが実際には好都合である．本章ではまず，特徴と意味の間の相互作用を有する統合的な景観認識手法を提案する．この概念に基づき，シーンクラスを考察する．次に，MPEG ベースの交叉学習タスクを4つの主観的シーンクラス（SSC）における5つの代表的シーンの間で行う．3番目に，計算的シーンクラス（CSC）の自動生成について実験する．今回，その機械生成した CSC は人間の定義による SSC に対して十分な一致を示さなかったが，いくつかの実際的な制約のもとではその CSC は SSC に関係づけることが可能であることが推察される．

Keyword シーンクラス，CSC，SSC，交叉学習，パン，チルト，白波の検出，木々の揺れの検出

8.1 はじめに

本章ではシーン間の交叉学習の性能に基づいてシーンクラスを生成する．まず，特徴と意味との間の相互作用の能力を有する統合的なシーン認識手法を提案する．この概念に基づき，シーンクラスを考察する．次に，4個の主観的シーンクラス（SSC）における5つの代表的なシーンの間で MPEG ベースの交叉学習タスクを実行する．3番目に計算的シーンクラス（CSC）の自動生成について実験する．

Table 8.1 Abstract of test sequences

No.	Name	Description	Subjective Scene Class
1	CHIRYU	General road in suburban residential area.	3
2	EXPO	Suburban toll road.	12
3	AOKI	Main street which has buildings on both sides.	4
4	APITA	General road in residential area. Buildings are on both sides of the road.	4
5	FOREST	Parking lot of a park along the road.	7

8.2 実験

既に6章で示した5種類の代表的なシーン(図6.1, 表8.1)について, 様々な交叉学習実験をおこなった. 表8.2の時間的特性に従って, 主要オブジェクトの認識率を計算し, 学習データ, 場所, 時間, 季節, 天候との関係を考察した.

8.2.1 交差学習による自動インデキシング

認識画像と学習画像の間で場所が異なる場合, もっとも影響力の大きい要因はシーン中のオブジェクトの配置の変化である. 学習画像に出現しないようなオブジェクトが認識対象画像に存在する場合には認識結果は決して真にならない. オブジェクトの配置に関する差異はSSCの差となって表れる.

しかし, さまざまな走行シーンを考察すると, より細かいピッチの時間スケールでオブジェクト配置や位置関係が変化する例が多く存在することがわかる. すなわち図8.1で示したSSCは静止画を2次元図形のパターンと領域ラベルによって分類しているにすぎない. ところが人間はそのSSCをもとにして3次元的な記憶をたどり, 最も適当なラベルを主観に従って割り当てる. そしてその主観とは主として複数の静的なオブジェクトの間の3次元的位置関係に由来する.

より正確なSSCを動画に対して定義するには, 形状や色などの視覚的特徴, ランドマークや交通関連の要因を含む地理的要因, に加えて, 車両, 歩行者など動的オブジェクトを航路すべきである.

これらの要因は上述した視点とターゲットアプリケーションに従って組み合わせられる. その結果, これらの要因に適したラベルを詳細化し再定義することが必要になる. まず, SSCにとって最も重要な要因である場所に焦点をあてる.

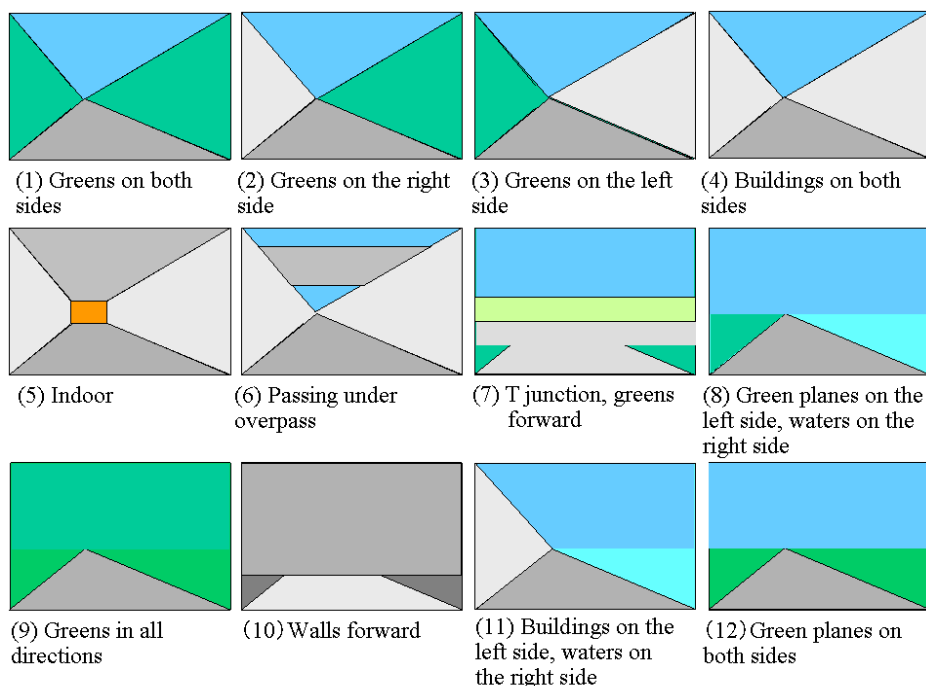


Fig. 8.1. Subjective scene classes.

Table 8.2 Experimental results of cross learning.

No.	Sequence to be learned	Remark	Season	Time	Weather	Amounts of learning	Sequence to be recognized and recognition rate (%)				
							1a	2a	3a	4a	5a
1a	070629-0826_CHIRYU	Representative sequence of 'CHIRYU'	Summer	Morning	Fine	3446	74.1	74.0	71.3	51.1	90.2
1b	070705-0819_CHIRYU	Date is close to 1a	Summer	Morning	Fine	670	70.8	70.7	67.4	61.8	92.6
1c	070927-0820_CHIRYU	Season and configuration ('road is under construction') are different from 1a	Autumn	Morning	Fine	699	55.6	-	-	-	-
1d	071023-0830_CHIRYU	Season is different from 1a	Autumn	Morning	Fine	1993	71.0	-	-	-	-
1e	060510-0750_CHIRYU_A	Weather and configuration of first frame ('no big trees') are different from 1a	Spring	Morning	Cloudy	1039	34.2	-	-	-	-
1f	060510-0750_CHIRYU_B	Season and weather are different from 1a	Spring	Morning	Cloudy	1592	75.7	-	-	-	-
1g	061124-0809_CHIRYU	Season is different from 1a	Autumn	Morning	Fine	821	77.9	-	-	-	-
1h	061213-0804_CHIRYU	Season is different from 1a	Winter	Morning	Fine	1723	74.2	-	-	-	-
1i	070305-0807_CHIRYU	Season and weather are different from 1a	Spring	Morning	Rain	1636	72.0	-	-	-	-
1j	070312-1120_CHIRYU	Time and weather are different from 1a	Spring	Before noon	Cloudy	1737	76.3	-	-	-	-
2a	060726-1713_EXPO	Representative sequence of 'EXPO'	Summer	Evening	Fine	1429	39.3	74.0	67.0	75.9	67.8
3a	070628-1734_AOKI	Representative sequence of 'AOKI'	Summer	Evening	Fine	1375	42.4	65.8	77.3	48.9	74.9
4a	070628-1744_APITA	Representative sequence of 'APITA'	Summer	Evening	Cloudy	789	41.9	70.8	76.8	76.2	60.5
4b	070628-1305_APITA	Time is different from 4a	Summer	Day	Cloudy	1134	-	-	-	67.3	-
5a	060514-1748_FOREST	Representative sequence of 'FOREST'	Spring	Evening	Fine	1453	60.7	70.6	62.4	41.8	94.0

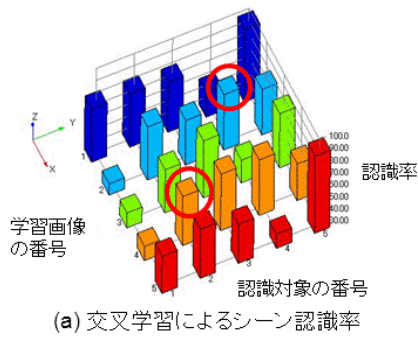
図 6.7 と図 6.8 は Seq.3a がターゲット画像で Seq.4a が学習画像である場合の実験結果である。このシーン認識率は viewpoint に従って定義され、主として画像中で大面積を占めるオブジェクトに影響を受ける。Seq.3a では空と道路がかなり高いオブジェクト認識率を維持しており、これが安定したシーン認識率につながっている。ここでは時間平均で 76.8% という良好な値である。

8.2.2 CSC の自動生成

次に、CSC の自動生成に関して実験する。簡単化のため、シーン認識率における view は *all* に設定する。今、ある学習画像 I_L とある認識対象画像 I_E の組み合わせについて、シーン認識率 Q_{all} が以下の条件：

$$Q_{all} \geq Q_{th_csc} \quad (1)$$

を満たすならば、 I_L と I_E は同じ CSC に属し、そうでなければ異なる CSC に属するとする。今回、閾値 Q_{th_csc} を 0.7 に設定し、上記実験で得た CSC の関係を図 8.5 に示す。この結果、シーン APITA (SSC=4) とシーン EXPO (SSC=12) は同じ CSC に属する。また、AOKI (SSC=4) の SSC は APITA と同じだが、AOKI の CSC は APITA と異なっている。これは、認識率が view に高い割合で依存するため、もっともらしい結果である。View が図 8.5 における SSC に従って良好に定義されていればこのミスマッチは改善されるであろう。この場合、空は APITA や EXPO の CSC において支配的な要因であった。AOKI は SSCno.4 の鍵となるビルディングに関して学習能力を提供できなかったと考えられる。これはカメラの角度とビルディングの高さに由来する。



No.	名前	説明	SSC
1	CHIRYU	General road in suburban residential area.	3
2	EXPO	Suburban toll road.	12
3	AOKI	Main street which has buildings on both sides.	4
4	APITA	General road in residential area. Buildings are on both sides of the road.	4
5	FOREST	Parking lot of a park along the road.	7

(b) 各シーンの主観的シーンクラス (SSC)

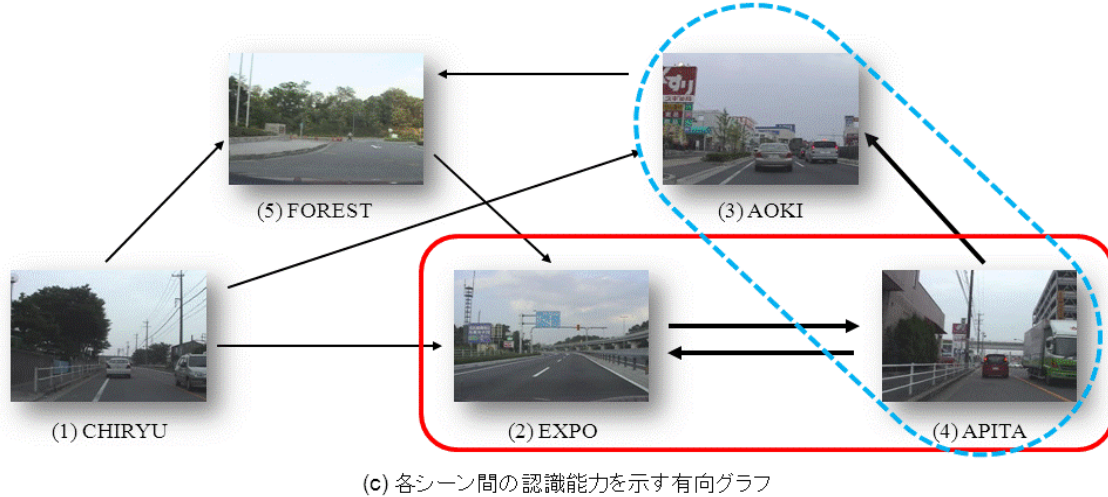


図 8.2 計算的シーンクラス (CSC) の生成

Fig. 8.2 Generation of Computational Scene Class (CSC).

8.3 考察

SSC は人間の直観的な主観的感覚に基づいてシーンを分類しようとする際に便利である。しかし、交叉学習で生成した CSC は常に SSC と一致するとは限らない。過去に獲得した回帰係数行列を活用してシステムが未知シーンを認識するためには、SSC と CSC の間の関係をより明確に定義しなければならない。

評価の view に関する規定が変われば、2つのシーン間の有向距離を計算する際のシーン認識率が変化するため、CSC グラフを修正することになる。語彙間の既存の知識ベースの関係を 3 章の (3.53)式と(3.54)式で導入すれば、複数の SC は良い方向に組織化され SSC に整合すると予想される。これは自動記述を目指す際の設計上の問題である。この段階で難しいことは、時間スパン、カメラからの物理的距離、カメラアングル、特定のオブジェクト（人、車両、ランドマーク他）であり、これはシーンの評価に重大な影響を与える。本章では、議論を単純明快にするため、多数の潜在的な値や、象徴的な意味（交通標識など）、あるいは動的特徴などを有する特定のオブジェクトは考察対象から外した。

8.4 風景解析への応用

以上、6 章で述べた動画画像の交叉学習の概念を発展させて 2 つのシーン間の有向距離を提案し、CSC の自動生成を試みた。MPEG ベースのシーン認識手法では個々のブロック画素を単位として意味との対応付けを計算できるため、シーンクラスの構成に適していると考えられる。

今回、自動生成した CSC は人間が定義した SSC と十分に一致していない。しかし、いくつかの実際的制約のもとでは CSC を SSC に連関づけることは十分可能である。3 章では信号的特徴と意味との間の相互作用を数値処理で行うことも提案した。これに関しては風景解析、予測と認識の相互作用（9 章）を踏まえた上で、再度 10 章で取り上げ、シーンクラスの認識に関する定量的評価を試みる。

今回、公園や海岸などの景観画像を例に取り、本手法の効果と今後の展望を考察した。人間が

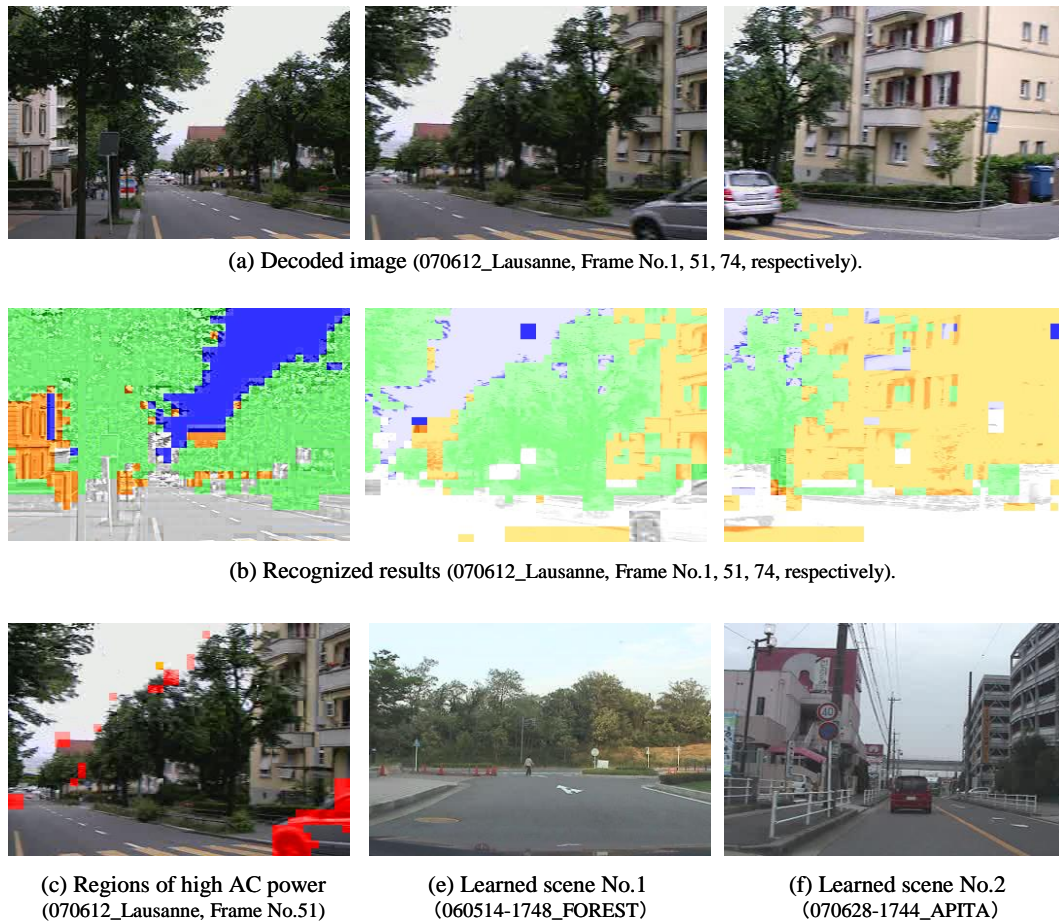


Fig.8.3 Recognition of street scene

図 8.3 道路景観の認識.

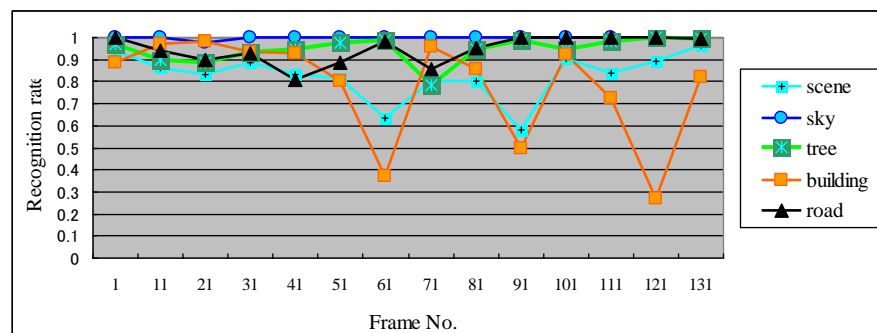


Fig.8.4 Temporal changes of recognition rates.

図 8.4 認識率の時間変化

手に持って撮影するカメラの映像について多変量回帰分析、主成分分析、時系列解析を適用する。また、前章までに取り上げた車載カメラ映像との違いについても考察する。

8.5 実験

一般的に、多変量回帰を含む統計的認識手法を用いる際、個々のオブジェクトの認識率は学習データの選び方にきわめて依存する。この傾向は学習量が固定であっても変わらない。3章および6章で述べたように、MPEG-1 フルフレームビデオを学習データとして処理する際、学習データの基本単位は1個のマクロブロックである。この学習単位をイベントと呼ぶ。また、学習量の基本単位は1フレームについて定義され、フレームあたりの最大イベント数はMPEG-1では330になる。この定数による規定は、学習操作において各オブジェクトが学習フレーム中で占める面積の割合を考慮する際に重要である。具体的な学習データの増やし方については6章で述べた。

図8.3は市街地の景観認識の実験結果である。図8.3(a)は認識対象とするMPEG復号画像を示す。図8.3(a)の左はI picture (intra frame), 図8.3(a)の中央と右はB picture (bi-directionally interpolated frame)である。ここで表示上、差分ブロックとイントラブロックを区別するために、同じインデックスのもので前者を濃い色に、後者を薄い色で表示した。

式(3.4)の学習タスクにおいて、図8.3(e)と図8.3(f)に示した2つのMPEGシーケンスを用いた。6章で示した手続きに従って、各MPEGシーケンスは独立に処理されてそれぞれに対応した学習データを生成する。最後に2つの学習データをマージして一つの学習データとする。これは2599イベントのデータを含み、式(3.17)の計測行列として用いた。図8.4は10フレームの区間における認識率の時間変化を示す。

8.6 考察

6章で述べた車載カメラによる走行映像の景観認識に関する知見を以下に要約しておく。

- (K1) 教示データ数1000イベント前後でシーン認識率は飽和傾向を示す。教示データが2000イベント以上になると逆に認識率が低下するケースもある。
- (K2) 空、街路樹、道路などシーン中で継続して広い面積を占有するオブジェクトの認識率は教示量の増大に対して飽和傾向を示す。
- (K3) 小面積のオブジェクトでも色や位置に特徴のある剛体では教示量の増大に対して認識率も増大する。
- (K4) 歩行者のように小面積かつ形状や位置が変化する非剛体の認識率は極端に低い。
- (K5) 認識率の時間変化は大面積を継続するオブジェクトで小さく、小面積あるいは生成消滅の激しいオブジェクトでは大きい。
- (K6) シーン認識率は画面内のオブジェクト数が少ないほど安定する。
- (K7) 状況要因が同じでも認識率が動的に変化する場合がある。人通りや交通量、移動物体の色、雲の配置などの動的変化がその要因である。
- (K8) 空、街路樹、道路などの大領域のオブジェクト認識率は季節の変化に対しても比較的安定しており、シーン認識率は低下していない。
- (K9) 急激な天候の変化や景観の変化(工事、建物の生成消滅、大面積の移動体、撮像範囲など)

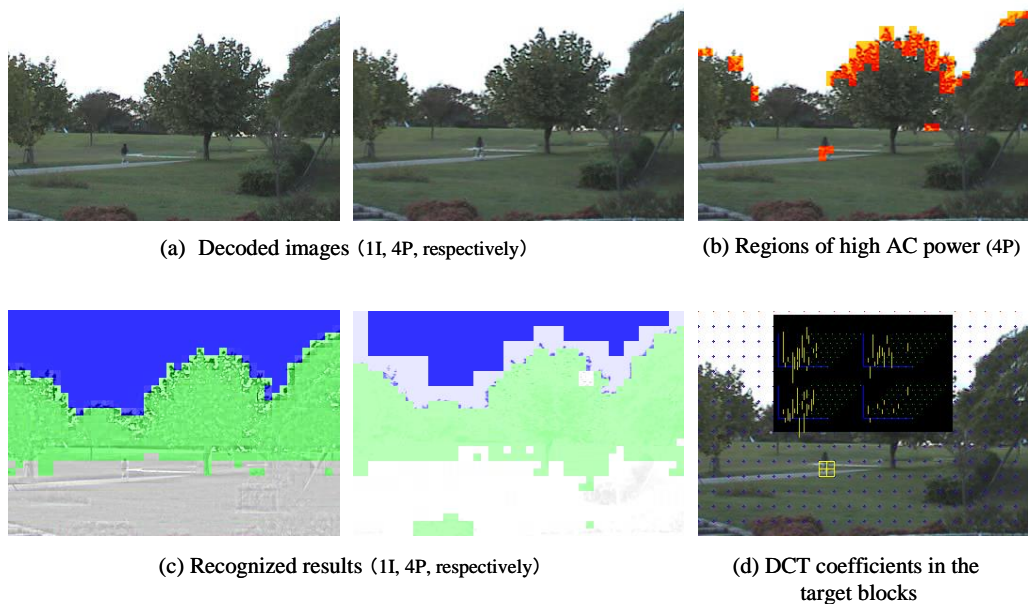


図 8.5 公園風景の解析

(a) 再生画像, (b) AC 成分の電力が高い領域, (c) 景観認識の結果, (d) 観測窓における DCT 係数

Fig.8.5 Analysis of a scene in a park.

は認識率を変動させる。

(K10) 学習画像とは異なる場所の画像を認識する場合でも、空や道路など大面積のオブジェクト認識率が高い値を保つならばシーン認識率も安定する。

(K11) 小面積のオブジェクトや出現、接近、消失の頻度が高いオブジェクト（標識、看板、対向車、歩行者など）が評価指標（視点）の対象に入っている場合、シーン認識率の変動は大きい。

(K12) 場所も時間も学習画像と異なる画像では平均シーン認識率は低い。

(K13) 2つの画像間で役割（学習と認識）を相互に交換した場合、認識率の対称性は必ずしも成立しない。

以上は走行時の車載カメラの場合であり、視野内に時々刻々とオブジェクトが生成消滅するという困難さがある。しかし、人間が手に持って撮影するカメラ映像の場合に比べてシーン構造をモデル化しやすいという特徴もある⁵⁾。今回図 8.3(a)で示した風景は人の手にあるカメラで撮影したもので、位置の変化は少ないがパン・チルトによりその後のカメラ姿勢が大きく変化する（水平方向は約 180 度）映像である。視点のインデックスを $O_{sightseeing} = \{\text{空, 樹木, 建物, 道路}\}$ とした場合、シーンの前半部分では比較的良好に追従できているが、シーンの後半ではやや劣化し、シーン全体の平均シーン認識率は 82.8%であった（第 8.4 図）。劣化の原因はパニングと見られ、差分ブロック処理の改善が今後の課題となる。

もう一つの問題として、建物の領域で空と誤判定される場合があるが、これは色やテクスチャのみでは人間でも判定困難な箇所である。この解決には次の方策が考えられる：

- 1) より多くの類似映像を学習させる
- 2) 周囲との位置関係や形状から意味を修正する[A9][A11]

1) は特定のシーンクラスに適合する学習データをユーザ（あるいはシステム開発者）が選択

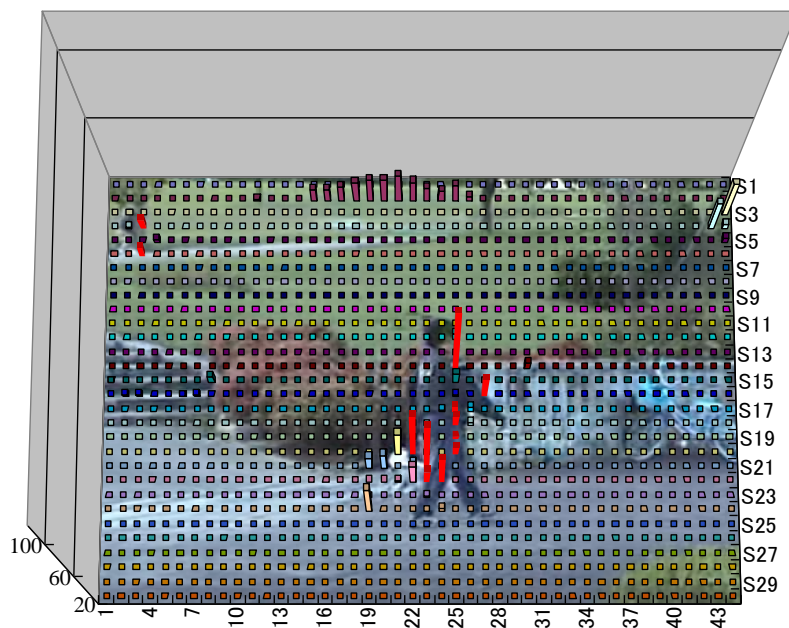


Fig.8.6 Detection of walkers.

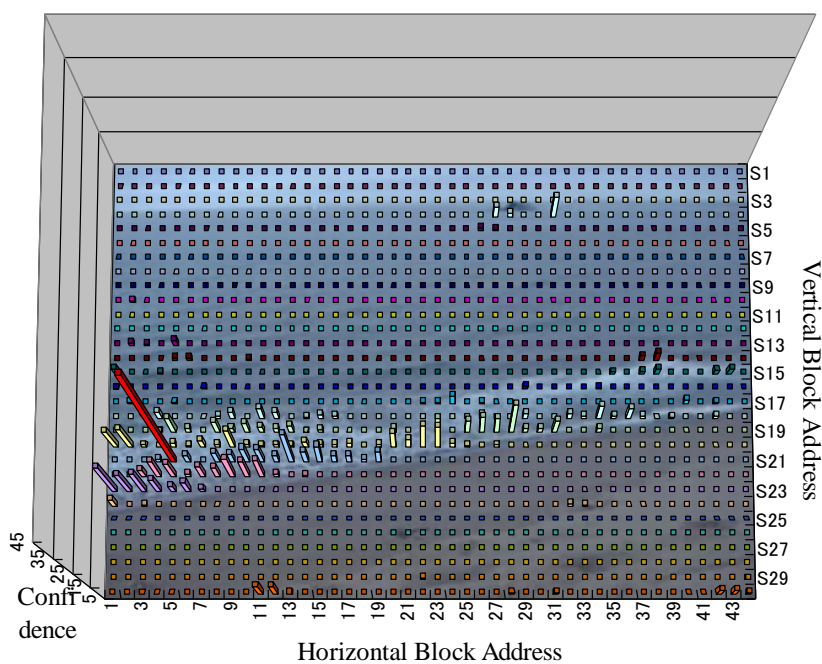


図 8.7 白波の検出

Fig.8.7 Detection of white waves.

的に増やすということの意味する。2) は判定規則を追加するかあるいはその規則そのものを獲得できるような知識処理を導入することが必要になる。

本手法ではマクロブロックあたりの特徴次元数が少ない (25 次元) にもかかわらず、フレーム平均のシーン認識率は SIFT (Scale-Invariant Feature Transform) 特徴量を用いた手法を 30% 以上も上回る場合がある [A12]。これは SIFT が特徴的なテクスチャに着目して局所特徴を抽出する反面、

空間変化が少ない大面積のオブジェクトの分類には適していないためと考えられる。さらに MPEG はフレーム間の特徴抽出が行えるため、動画全体を捉えるような景観認識では演算量と認識性能の両面で有利となる。

8.7 様々なシーンへの適用

MPEG-Based Vision の付加価値を高めるには景観認識に加えて遠方の人や動物，ランドマークなども認識できることが望ましい。車載カメラの視野に限定すれば，MPEG データで記述された動き補償差分の DCT 係数に時系列解析を施すことによって歩行者を検出する試みは 7 章で述べた。

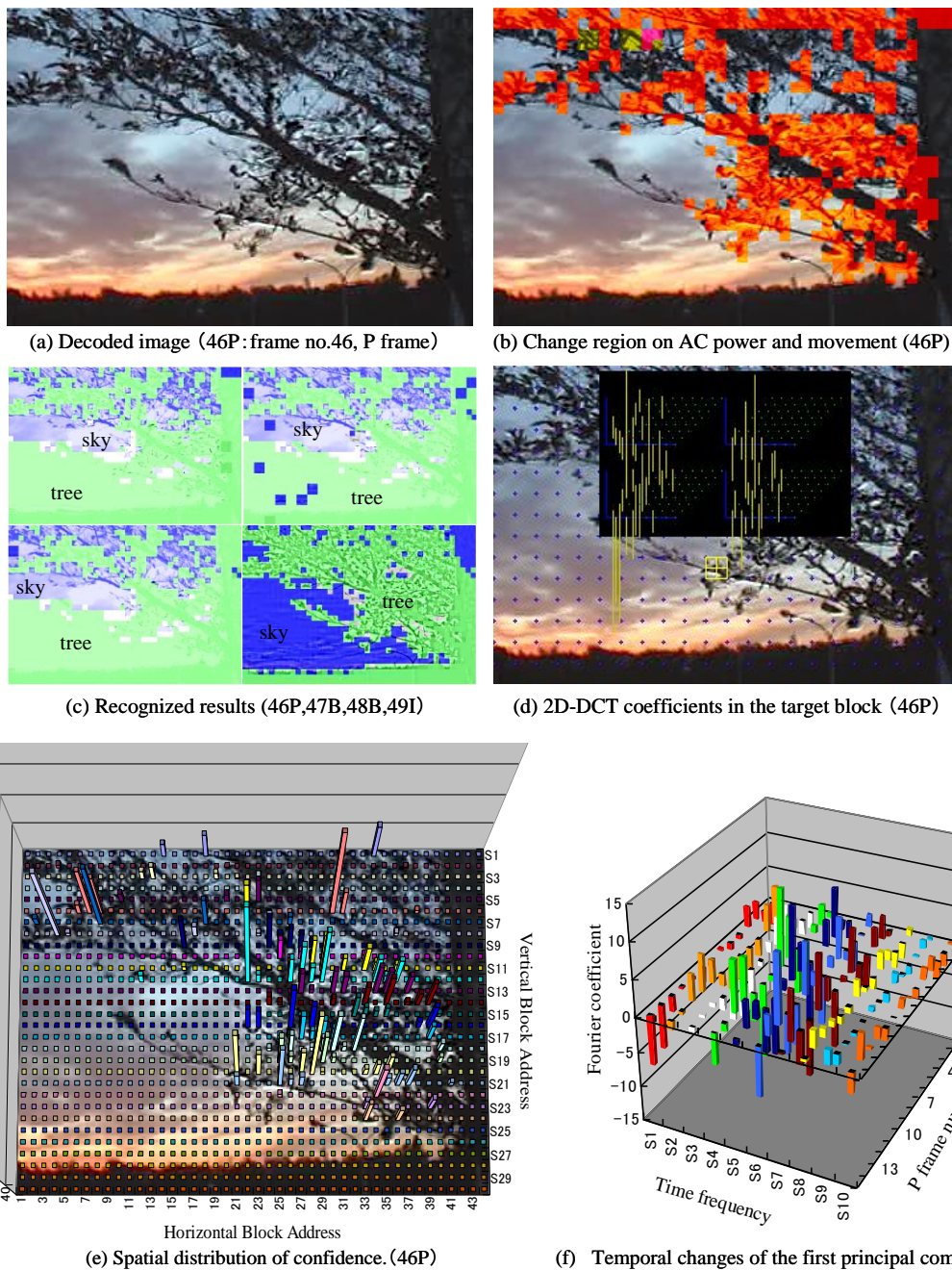


図 8.8 木々の揺れの検出

Fig.8.8 Detection of movements of trees.

また、MPEG の動きベクトル情報を利用した 3 次元運動推定や顔検出の試み、TRECVID[C10]における映像検索のコンソーシアムなども注目される。

このような流れを受け、ここでは DCT 係数を用いて風景に含まれる時間的変化を認識するための 2 つの手法を説明する。

8.7.1 境界部分の検出 8

画素ブロック内の AC 成分電力は絶対値電力の場合、基本的には式(7.1)に従って計算される。図 8.5 は公園風景のシーンの P フレームに適用した場合のシーン認識結果である。AC 成分の電力の閾値判定で木や歩行者の境界部分が検出されることがわかる。

8.7.2 周期的変化の検出

MPEG 特徴量を用いて風景に含まれる周期的な時系列変化を認識することを試みる。1 秒間に 10 フレーム存在する P ピクチャ（予測画像）を用いて DCT 係数の時系列を含む特徴ベクトルを構成し、線形主成分分析（LPCA：the linear principal component analysis）による歩行者検出（3 章）を適用した結果を図 8.6 に示す。図中の確信度は各ブロック画素の PCA について上位 3 主成分から算出した相対値である。AC 成分の電力のみで検出した場合よりも誤検出が低減され、遠方と近距離の両方で歩行者が検出できている。同様に、海岸における白波の部分（図 8.7）、公園の樹木において風で揺れている部分（図 8.8）の検出を行った。また、図 8.8(d)の観測ブロックについて第 1 主成分の時系列に FFT を施し、空間スペクトルの時間的周期性(最大周期 1 秒)を解析した(図 8.8(f))。自然風景に見られる、形状変化は複雑だが巨視的には繰り返しを感じさせる現象はこの手法で判別できる可能性がある。

8.8 考察

シーン中で空間的には形状が複雑であっても、時間的には何らかの繰り返し挙動を示しているとみられる現象は上記の手法で判定できると仮定する。この仮定に基づき、学習量を極力増やさないようにしてさらに性能を改善するいくつかの戦略を以下に要約する：

- S1) 適切なシーンクラスを選択する、あるいはシーンクラスを適応化させる
- S2) 過去の統計や時間的連続性、あるいは空間的連続性をもとに正解の推定値を自動生成する
- S3) 信号処理で得られたインデックスをセマンティクスで修正する
- S4) 状況要因の活用
- S5) KPCA や SVM などの高性能学習手法
- S6) 動きベクトルその他の次元を用いた P フレームと B フレームの認識性能の改善
- S7) PCA-FFT のような特徴ベクトルの時系列解析。

特に S1)に関して、シーンクラスは上記の実験に関して事前に指定されていないが、実験はシーンが地上にあるものと仮定して行われた。すなわち、事前定義の辞書では水面に関する語彙はまったくなかった。これがこの段階の出発点となる。したがって図 8.7 の入力シーンの認識結果は海や水の波に関するインデックスをまったく持たない。空と道路のみ存在する。そこで、図 8.7 の結果はシーンクラス関係グラフにおいて遷移のトリガーとなる。これは、「道路表面は動的に動かない」と「道路表面は流体のようなテクスチャを生成しない」という経験則に基づいている。それでもなお、水面とさざめく木々や、何か複雑な動的テクスチャを持つものとの違いを証明す

る問題が残されている。

図 8.7 において，最も高い値を示す赤のバーはこの位置のブロック画素がサンプルされ，時系列特徴ベクトルの LPCA において自己交叉学習に用いられたことを意味する．他のブロックで小さくない確信度を有するものは推定された白波の部分である．砂の表面や海の平面については，有効な確信度のブロックはほとんど存在しない．したがって，白波は上述の信号処理で識別され，その白波に隣接する表面は AC 電力を用いて動的部分と識別されたものであり，それは海面であるという判断に至るであろう．そしてこれは上部で空の平面とも隣接している．この空の面はまず第 1 段階で認識されることに注意せねばならない．他方，このかなり静的な面で道路と判断されているものは砂の平面（砂浜）と判定でき，これはほとんど言語レベルの常識を用いた結果である．

図 8.8 に示す実験において，木々の揺れ動く部分は PCA-FFT により検出された．これは黄昏，やや曇りがちの暗い空，そして木々の間の境界部分において誤検出を低減するための治療法となる．これらの境界は図 8.8(c)が示すように，第 1 段階のシーン認識においては部分的に劣化している．図 8.8(b)(e)(f)に基づく経験的規則は，もし適当なシーンクラスが選択されたならば認識結果を改善するであろう．

多変量回帰分析においても主成分分析においても非線形手法は鋭い判別特性を持つ反面，汎用性の獲得が難しい．一方，線形手法の場合は比較的容易に過去の学習特性や教示データを再利用できる．学習型の認識手法では人手で正解を教示する作業が最も高負荷であるため，その低減は切実に望まれる．そこで，学習データと語彙，シーンクラスをアプリケーションに応じて体系的に効率よく作成し，データベース化していくことが今後の課題となる．だが，データベース構築の初期段階では認識能力へのフィードバックは十分ではない．そこで，

- 1) 不完全な認識器を複数個連携させることで GT に近い ET を獲得する
- 2) 未整備だがすでに電子的に記述されたデータを教示データに変換する
- 3) システムが適宜ユーザに問い合わせをする機構（音声対話など）を具備する

などが考えられる．

8.9 おわりに

構成オブジェクトや状況要因（場所，時間帯，季節，天候）に大きな変化がなければ，平均シーン認識率は概ね良好である [A12]．本手法はフルフレームの映像をひとまとまりの景観として認識する際に有効であり，シーンクラスを自動判別できる可能性もある [A11]．しかし，事前に学習した定型的構造では記述困難なオブジェクト，学習範囲に納まらない特定の構造物，人，動物，さらには波や風に揺れる樹木といった微妙な時間的变化を示すオブジェクトについては本章で紹介したような新たな認識演算を追加する必要がある．そのシステムを構成する際には，どこまでが信号処理で、どこまでが画像を観測する人間側の知識処理に依存する部分なのかを明確にすることはアルゴリズムの構築上，大変重要である．そして，この両者をつなぐ最も実地的な手段の一つが 8 章で提案したシーンクラスであると考えられる．

以上を踏まえて次章以降では，シーンクラスによる適応化により最適な教示データの選択や特徴ベクトルの拡張を自動化する手法について考える．

9. 予測と認識の相互作用

本章では、車載カメラや定点カメラ、携帯電話などから得た歩行者の出現頻度に関する情報をセンターやカーナビの記憶装置に集積して得られる時空間分布データベースに基づいて、想定する時間帯に所定の対象区間内で出現する歩行者の頻度を予測するシステムを提案する。

キーワード 出現頻度, 予測, 論理予測率, 状況ベクトル, 主成分回帰分析

9.1 はじめに

前章までに示した確信度の時系列を推定する方法では、過去の学習データにのみ依存し、初期値の影響は考慮しなかった。システムがまったく過去の動作記憶をもたず、毎回、初期設定された状態で認識動作を行う場合は学習データのみが性能を決定する。ところが、その初期設定が対象シーンに適合しない場合は認識結果は劣悪となる。そこで、1章で説明した3Sモデルのように状況、形状、スペクトルといった異なるモードでの認識結果を初期設定に考慮することで適応的に性能を切り替えることが考えられる。また、仮に同じモード内であっても過去に試行した認識結果が、次の認識動作に反映されるならば、最終的には良好な認識性能を獲得することが期待される。さらには、通信などにより他の機械が獲得した経験を集合知として活用できれば人間が教示すべきデータ量は格段に低減する。

近年、プローブカーを用いた渋滞予測システムが実用化され、今後もこのような動的なセンサーネットワークを利用した様々な新しい情報サービスが台頭するであろう。一方で、車載器のみならず、携帯電話でも天気情報や映像情報をニーズに応じて各ユーザがアップロードするというヒューマンプローブも急速に普及し、2010年の段階で約160万台の携帯電話がその情報サービスを活用している。このような超分散的な情報センシングに基づく情報提供は集合知の第1段階とみなすことができる。

Sasaki ら[A9]はこのような動向を踏まえて歩行者の出現を天気予報のように予報するシステム

表 9.1 予報のうれしさ

Table 9.3 Benefits of forecasts.

大分類	クラス	一般的な時間変化	発生頻度のピーク	平均発生頻度	経験による予測可能性	予防安全における予報のうれしさ	
						クラス既知	クラス未知
よく通る場所	大通り	大	大	大	大	小	
	駅前	大	大	大	大	小	
	商店街	大	中~大	中~大	大	小	
	学校周辺	大	大	大	中	中	
	病院周辺	大	中~大	大	中	中	
	24hコンビニ周辺	中	中	中	中~大	中	
	スーパー周辺	大	中~大	大	中	中	
	住宅街	小	中	小	小	大	
	郊外交差点	不明~小	小~中	不明~小	小	大	
郊外単路	不明~小	小~中	不明~小	小	大		
よく知らない場所	大通り	大	大	大	中	中	大
	駅前	大	大	大	中	中	大
	商店街	大	中~大	中~大	中	中	大
	学校の近く	大	大	大	中	中	大
	病院の近く	大	中~大	大	中	中	大
	24hコンビニ周辺	中	中	中	中	大	大
	スーパー周辺	大	中~大	大	中	中	大
	住宅街	小	中	小	小	大	大
	郊外交差点	不明~小	小~中	不明~小	小	大	大
郊外単路	不明~小	小~中	不明~小	小	大	大	

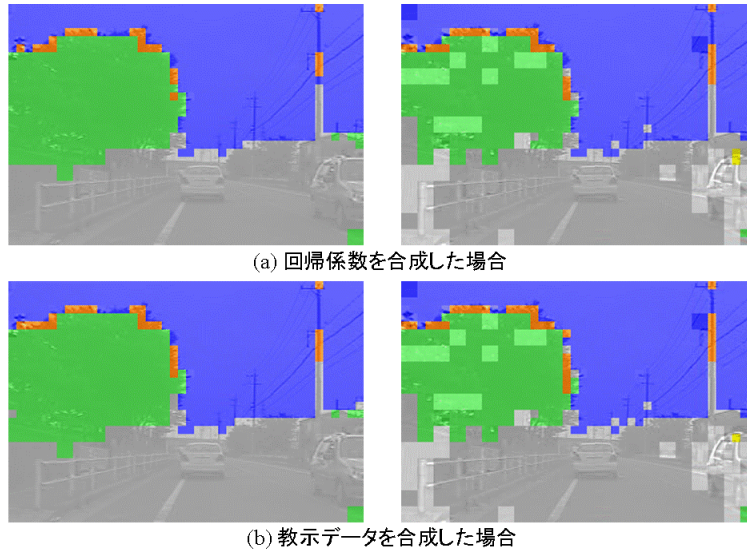


図 9.1 インデックス画像の比較

Fig.9.1 Comparison of indexed images.

のコンセプト「人予報」を提案した（図 5.1）[A14]。ここでは、小規模ではあるが3年間の走行映像をもとに予測に関する予備実験を行っている。本章ではこの内容に沿って、時空間的に分散する複数の画像認識器を連携させることで新たに発生する効果について考察する。

9.2 めざすシステム

9.2.1 予防安全

想定する日時の走行予定経路上において、歩行者や自転車の出現が予報できれば、ここで、システムは主として車載カメラの視野内にある移動体（車両、自転車、歩行者）を検出し、警報を発するとともに検出データを蓄積・通信できるものとする。

まず第1の恩恵は、人予報の結果、対象地域を通過する前に警報、危険回避、装置設定（速度制御も含む）、旅行計画変更、運転者変更（運転計画）を行うことが可能になり、予防安全を促進できるということである。例えば全く見知らぬ土地をレンタカーで運転する場合、歩行者や自転車と出くわす確率が高い場所や時間帯を事前に知ることができればドライバーの精神的負荷は軽減される。この効果の本質は、ドライバーにとって予測不可能な事象についてシステムがアドバイスする点にある。したがって、表 9.1 のように、走行場所のクラス分類または歩行者出現の有無が未知である場合に予報の恩恵は大きい。一方、駅や大通りなど常時人通りが多い場所では予報の恩恵は小さくなる。

近年、欧米やオーストラリア、日本などで実験が行われている ISA（Intelligent Speed Adaptation）[I4]との連携が考えられる。ISA は特定地域において自動的に速度が出なくする試みであり、現在はスクールゾーンなど固定エリアで実験が進められている。このエリアを動的に設定することが人予報によって可能になる。

9.2.2 認識器の自動適応

第2の恩恵は、予測結果と現地の認識結果を通信で連携させることにより、車載器の認識性能を自動的に改善できることである。これには以下の形態がある。

C1) 天候や環境などの外部データを利用する。

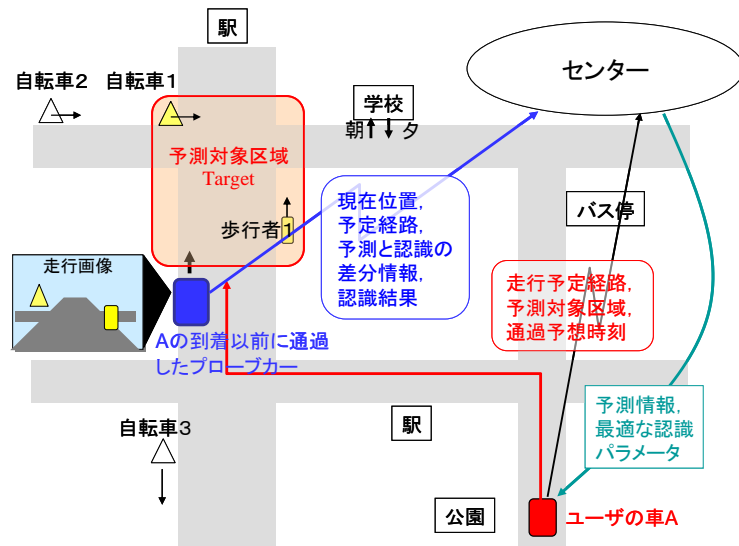


図 9.2 プローブカーの協調処理.
Fig. 9.2 Cooperative process between probe cars.

9.3 過去の統計に基づく予測

9.3.1 予報に至るプロセス

「ある車 A が予測対象地域 Target における人予報をユーザに提示する」というシナリオを図 9.2 の状況設定で考える。まず、Target に向かう A は予測条件をセンターに送信する。一方、A の通過以前に Target 近辺を撮像したプローブカーは認識結果をセンターに送信している。センターは算出した予測情報と認識パラメータを A に配信する。これをもとに A は Target 近辺の集客施設（学校、駅、バス停、公園など）の影響や走行経路上の時間経過に応じて予測と観測を繰り返し、

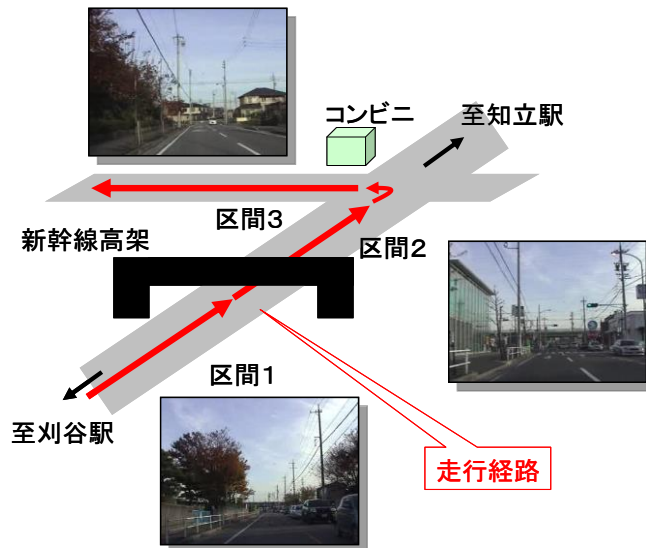


図 9.3 実験条件
Fig. 9.3 Experimental conditions.

ユーザにとって最適なタイミングで予報を提示する。

9.3.2 特徴ベクトルの定義

ある区間の歩行者出現頻度は、観測時も予測時も共通して特徴ベクトルの形式で表現できる。表 9.2 は図 9.3 に示した走行区域を対象として、オブジェクトの分類を {男性, 女性, 子供, 自転

車, 不明, 犬} × {右, 左} の合計 12 種類で考えた場合である. このとき, ある車載カメラで観測する歩行者出現状況の特徴ベクトルは 12 次元となる. しかし, 現状の画像認識技術ではこのような複雑な分類で高い認識率を達成することは困難である. そこで, 今回は走行映像を目視で認識した結果からイベントリストを作成し, 正解値および画像認識出力の模擬データとした.

9.4 実験と考察

9.4.1 実験条件

図 9.3 に示した区域で 3 区間の経路を対象とし, 05 年から 07 年にかけて撮影した走行映像から, 今回 156 イベントの GT 値を目視で作成した. これを状況ベクトルと特徴ベクトルの対で表現したものが表 9.2 の実験データである.

9.4.2 予測性能の評価

156 イベント全体を用いて主成分回帰係数を算出し, 自己回帰予測を行った結果とそれらに対応する GT 値を図 9.4 に示す. 歩行者出現に関する時空間分布のおおまかな傾向は類似しているが, 出現頻度の一致は困難であることが容易にわかる. また, GT の急激な変化に対して予測値は追従できていない. 一般に統計予測結果は理想的には正解データに近づくことが望ましいが, 人の出現に関しては予測値を人数で一致させることは非現実的と見られる. そこで予防安全の目的に立ち返り, 出現数ではなく, 出現の有無に関する論理値の一致度で判断する評価尺度を考え, これを論理予測率と呼ぶことにする. 156 個のイベント中 1 個を未知とし, 残り 155 個の GT を学習データとする主成分回帰予測を各イベントについて行い, 平均した結果を表 3 に示す. ただし, ここではオブジェクトを 4 種類 ({左, 右} × {歩行者, 自転車}) にグルーピングし, 犬は対象から外した. 以上により, 人予報の仮目標として想定した 70% を超える予測性能が達成された.

9.5 認識率に対する変化

式 (3.73) では認識率に対する ET の変化を表すモデル関数を導出した. これを用いれば, 認識率に対する予測率の変化を見積もることができる. 図 9.5 にその結果を示した. これより, 今回の地域データに関する限り, 画像認識率が 70% 以上あれば予測率の仮目標値 70% をクリアできる見通しがあることがわかった.

9.6 おわりに

車載カメラや定点カメラ, 携帯電話などから得た歩行者出現情報をセンターやカーナビに集積し, 想定する時間帯の走行予定区間における歩行者出現を予報するシステムを提案した. これは警報, 自動速度制御, 認識器の自動適応, マーケティング情報の提供, などの応用が期待できる. 今回, 05 年から 07 年にわたる 156 件の走行映像から目視で歩行者と自転車を抽出し, 状況行列を用いた主成分回帰分析を施すことで, 70% を超える平均論理予測率を確認した. 更に認識率の定義に基づいて ET のモデル関数を導出し, 画像認識率が最低 70% あれば予測率の仮目標値 70% を達成しうることを示した. 今後は様々な地域での検証, 画像認識器との結合, 複数の認識結果の統合, 予測から認識へのフィードバック, 予測結果の多値化などが課題となる. 一方で GT を極力少なくし, ET で予測効率を高めることが実用化の鍵となる.

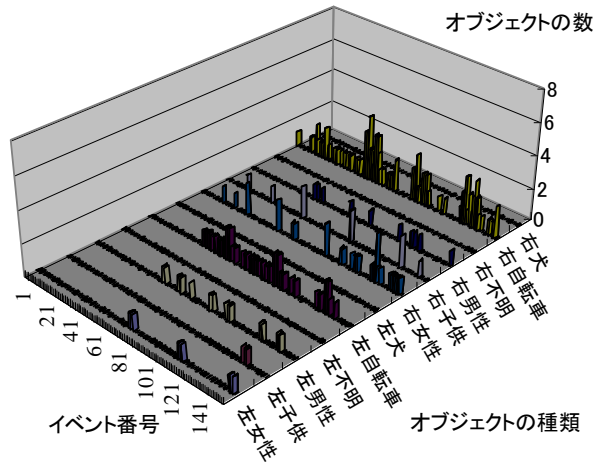


図 9.1(a) 自己回帰による予測結果

Fig. 9.2(a) Predicted results by autoregression.

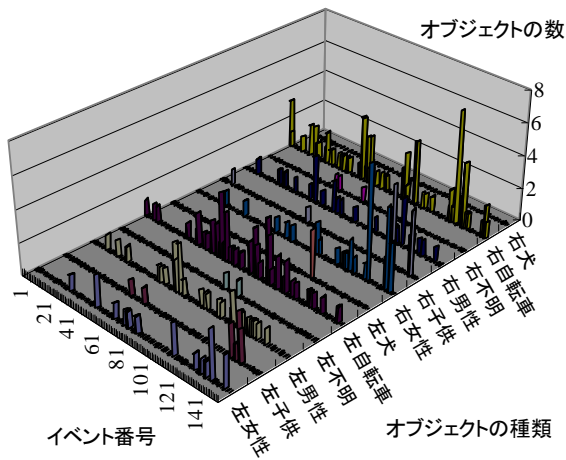


図 9.4(b) イベントの正解値

Fig. 9.4(b). Ground truth of events.

表 9.3 論理予測率

Table 9.5 Accuracy of logical prediction.

予測数	学習数	平均論理予測率 [%]				
		左歩行者	左自転車	右歩行者	右自転車	全体
156	155	72.4	73.7	70.5	68.6	71.3

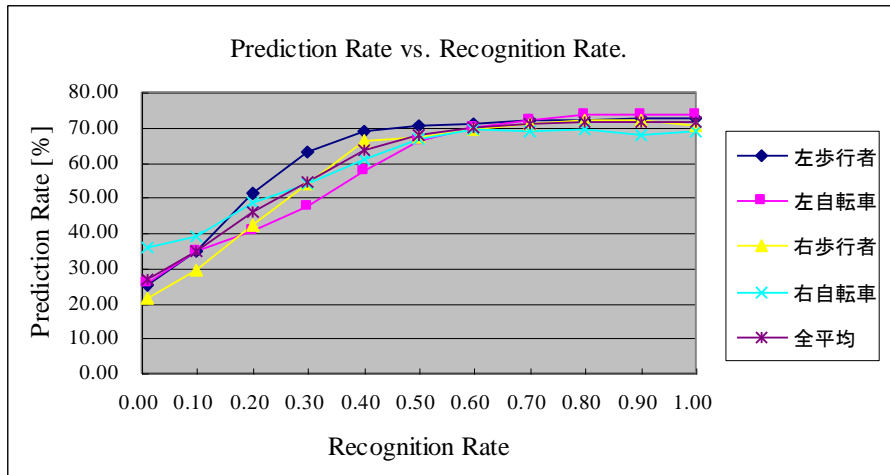


図 9.2 認識率に対する予測率の変化
 Fig. 9.3. Prediction rate vs. recognition rate.