4. ルールベースによる景観認識

本章では，色情報と知識処理を併用することで比較的簡単に映像記述を生成する手法を提案する．まず，MPEG 形式で記録された車載カメラ映像について，符号中の DCT 係数から得た色情報をトリガーとしてブロック画素単位のオブジェクトインデックスを生成する．次に，3 次元モデルベース手法とルールベース推論を併用することで主要オブジェクトの 3 次元的位置関係やシーンの特徴を把握し，XML 言語の映像記述を自動生成した．さらに，この記述データをユーザ側で作成した問い合わせプロファイルと照合することで，自然言語に近いクエリで画像検索が行えることを実験で確認した．

4.1 はじめに

近年，放送・編集及びマルチメディアの分野を中心として ISO の MPEG-7 など映像音声コンテンツの意味内容を記述する規格が定められた**エラー！参照元が見つかりません。エラー！参照元が見つかりません。エラー！参照元が見つかりません。**．このような動きは，氾濫する膨大なコンテンツ群からユーザニーズにマッチしたものだけを即座に自動抽出し要約するという技術が，業務用機器のみならず民生機器においても切望されていることに呼応する**エラー！参照元が見つかりません。**．一般にそのような記述の自動生成は信号特徴レベルでは可能だが，パターンや意味のレベルでは画像理解の問題と同様，用途を限定しない限り困難である．

しかしながらホーム，パーソナル，エンタテインメント系では，近年のカメラ付き携帯電話や車載カメラの普及に伴い，記録や検索，情報提供の観点から映像記述への要求は拡大している．たとえば，有限個の定型的なインデックスを辞書で用意し，シーンの分類や視覚的特徴（支配色，動き，

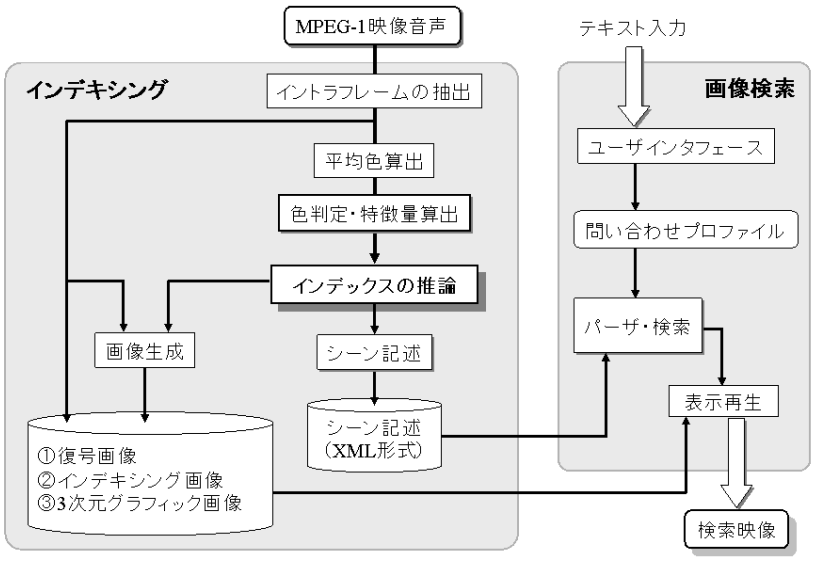


図 4.1 システムの構成

Fig. 4.1System architecture.

形状，明暗パターンなど）の記述を行うことは可能であり，実用性も高い．

本稿では車載カメラ映像を対象として，MPEG-1 形式の映像データから抽出した DCT 係数を色情報とし，画像処理と知識処理を併用したブロック単位のインデキシング手法を提案する．

4.2 車載カメラ映像のインデキシング

4.2.1 めざす応用

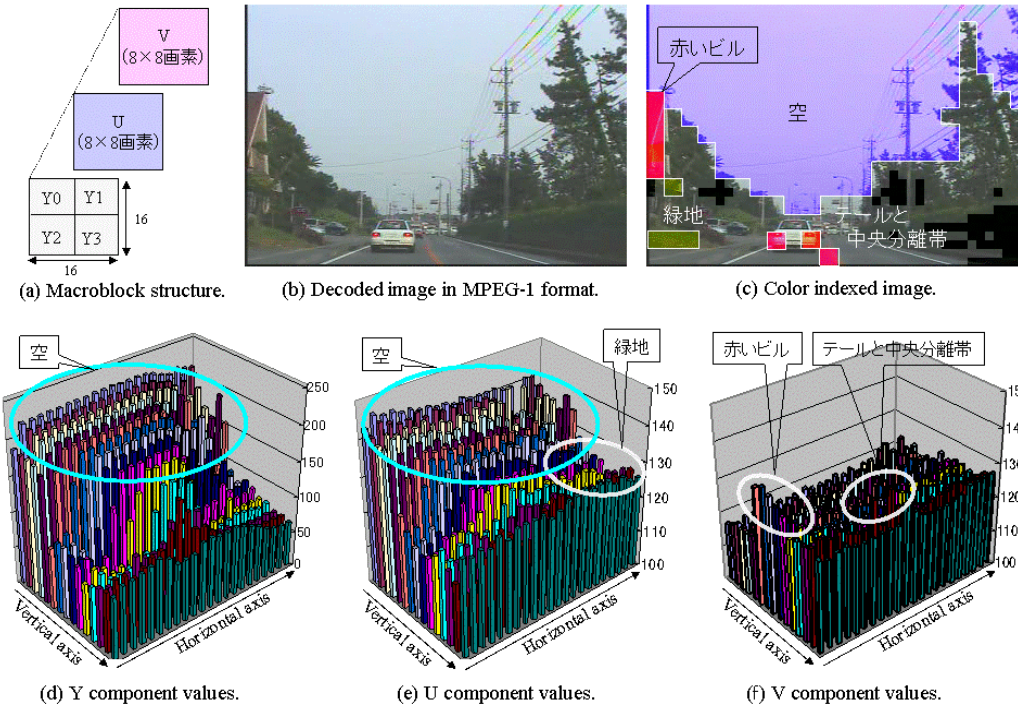
自動車にかかわる映像処理技術は，自動運転，安全運転支援，ドライバーモニター，駐車場監視など多岐にわたる．特に自動運転の分野では，ロボットの環境理解との共通部分も多く，今も精力的に研究が続けられている．一方，近年では交通事故発生時に不正確な陳述を回避するという観点から，事故を記録するドライブレコーダも普及しつつあるエラー! 参照元が見つかりません。．これらの動きを踏まえて我々は，走行中の道路状況を認識して適切な状況記述を付与し，車が獲得した映像記憶を検索するといった応用をめざしている．

4.2.2 インデックスの定義

車載カメラ映像に用いるインデックスとは，前方車両，対抗車両，道路標識，交差点，道路状況（渋滞，非渋滞），緑地，ビル，空，海岸，などである．それらは個々のオブジェクトに対応するものとシーン全体の特徴を表現するものに大別される．また，ある一つのシーンにはユーザの視点に基づく複数の解釈が可能であり，複数のインデックスを付与することもある．たとえば，“渋滞”，“事故発生”，“トンネルの入り口”，“桜が満開”などはひとつのシーンに付与できる．

4.2.3 システムの構成

映像のインデキシングは音声認識と同様に考えることができる．ある発話音声に単語または単



(a) Macroblock structure. (b) Decoded image in MPEG-1 format. (c) Color indexed image.

(d) Y component values. (e) U component values. (f) V component values.

図 4.2　YUV 成分の分析

Fig. 4.2　Analysis of YUV component.

表 4.1 色インデックスのリスト

Table 4.1 List of color index.

番号	表現語	色判定の対象となりうるオブジェクト								
		主として道路上にあるもの			場所が特定困難			主として道路外にあるもの		
1	青	青の標識	青の車	青信号	空	-	-	-	-	青のﾋﾞﾙ
2	緑	緑の標識	緑の車	青信号	-	-	-	緑地	街路樹	緑のﾋﾞﾙ
3	赤	赤の標識	赤の車	赤信号	-	-	-		街路樹	赤のﾋﾞﾙ
		停止ﾗﾝﾌﾟ	街灯	ﾍｯﾄﾞﾗｲﾄ	-	-	赤のﾗﾝﾌﾟ	赤いﾎﾟｽﾀｰ	赤い花	-
4	橙	橙の標識	橙の車	ｾﾝﾀｰﾗｲﾝ	空	工事看板	橙のﾗﾝﾌﾟ	-	-	橙のﾋﾞﾙ
5	黄	黄の標識	黄の車	黄信号	-	黄の看板	-	黄の花	街路樹	黄のﾋﾞﾙ
		-	-	ﾍｯﾄﾞﾗｲﾄ	-	-	-	黄のﾎﾟｽﾀｰ	-	-
6	白	ｶﾞｰﾄﾞﾚｰﾙ	白の車	白線	空	-	-	白い花	-	白のﾋﾞﾙ
		横断歩道	-	ﾍｯﾄﾞﾗｲﾄ	-	-	-	-	-	-
7	灰	道路	灰の車	-	空	-	-	-	-	灰のﾋﾞﾙ
8	黒	道路上の影	黒の車	車の影	夜空	-	ﾜｲﾊﾟｰ	-	-	黒のﾋﾞﾙ
-	(不定)	標識	対向車	先行車	歩行者	-	駐停車	-	-	ビル

語列を対応させる機構としては，単語のデータベースである辞書を用意し，その中の最も妥当性の高い単語を選出する手法が一般的である．もちろん，これでは辞書にない単語（未知語）を解として出力することはできないが，カーナビなど既に実用化された多くの分野ではこの方式が用いられている．

図 4.1 に本章のシステム構成を示す．インデキシング部では MPEG-1 形式[B1][B2]で符号化された映像を解析し，イントラフレームの色や信号変化に関する特徴を抽出する．そして，あらかじめ運転状況辞書やシーン分類辞書で定義された語彙のどれにあてはまるかを決定する認識計算を行う．その結果マクロブロック単位でインデキシングを行い，MPEG-1 の復号画像，インデキシングされたブロックを復号画像にオーバーレイして強調表示した画像，認識したオブジェクトの概略位置を 3 次元グラフィックス表示した画像，の 3 種類の画像を生成する．一方で，シーン中の主要オブジェクトの存在をフレーム毎に XML 形式で記述する．

画像検索部では上記の XML 形式の車載映像記述データとユーザ入力による問い合わせプロファイルの内容を照合し，検索画像を表示する．

なお，本システムでは次の 3 種類のカテゴリで色の言語表現を行った．

１） オブジェクトの色を表現する際の言語（表 4.1）

２） インデックス画像における各ブロックの色強調表示を行う際の色を表現する言語

３） オブジェクト位置の 3 次元グラフィックス表示を行う際の各ブロックの色を表現する言語

いずれも同一のカラーパレット上で各表現語と YUV 値および対応する RGB 値を管理した．

4.3 色インデキシング

4.3.1 DCT の DC 成分

MPEG 形式など DCT（離散コサイン変換）を用いた画像符号化形式では一般に，符号中に YUV 成分のおのおのについてブロック画素の DCT 係数が記述されている．中でもその DCT 係数の DC 成分はブロック画素の平均値に相当することは容易に導かれる．したがって MPEG-1 データ中の

DCT 係数の DC 成分をしきい値判定すれば，新たにブロック内平均処理を行わずともブロック単位の色判定が即座に行える．たとえば，ある道路画像について YUV 各成分の分布をブロック単位で示すと順に図 4.2 のようになる．

4.3.2 色によるオブジェクトの判定

インデキシングの対象としては，表 4.1 に示す 8 つの色インデックスに対応する代表的オブジェクトを考えた．そして図 4.1 にしたがって，MPEG-1 形式の圧縮画像データからイントラフレーム（予測や補間を用いない独立再生できるフレーム）を抽出し，マクロブロック単位（16×16 画素）で色特徴を解析した．具体的には，復号画像中で上記オブジェクトに対応する代表的なブロックを目視で選定し，次の 2 つの判定方法に必要な判定条件を経験的に設定した．

（方法 1）各成分の閾値判定

YUV の各成分値に対する閾値判定を論理的に組み合わせることで各ブロックの色判定を行う．各成分 A $(A=Y,U,V)$ を判定して論理値 $L(A)$を出力する判定規則は A_{upper} を上限，A_{lower} を下限として次式のようになる．

$$If\ (A_{upper} \geq A \geq A_{lower})\ then\ L(A)=1\ \ else\ \ L(A)=0 \tag{4.1}$$

（方法 2）ベクトル量子化

各ブロックの YUV 成分の代表値の組(Y,U,V)を色ベクトル $\boldsymbol{A}$ とみなす．一方，k 番目の色に対応する代表ベクトル $\boldsymbol{A}_R(k)$ $(k=1,\ldots,K)$で構成されるコードブックを予め設定する．このコードブックをもとに各ブロックの $\boldsymbol{A}$ についてベクトル量子化を行い，どの色であるかを決定する．ここで，

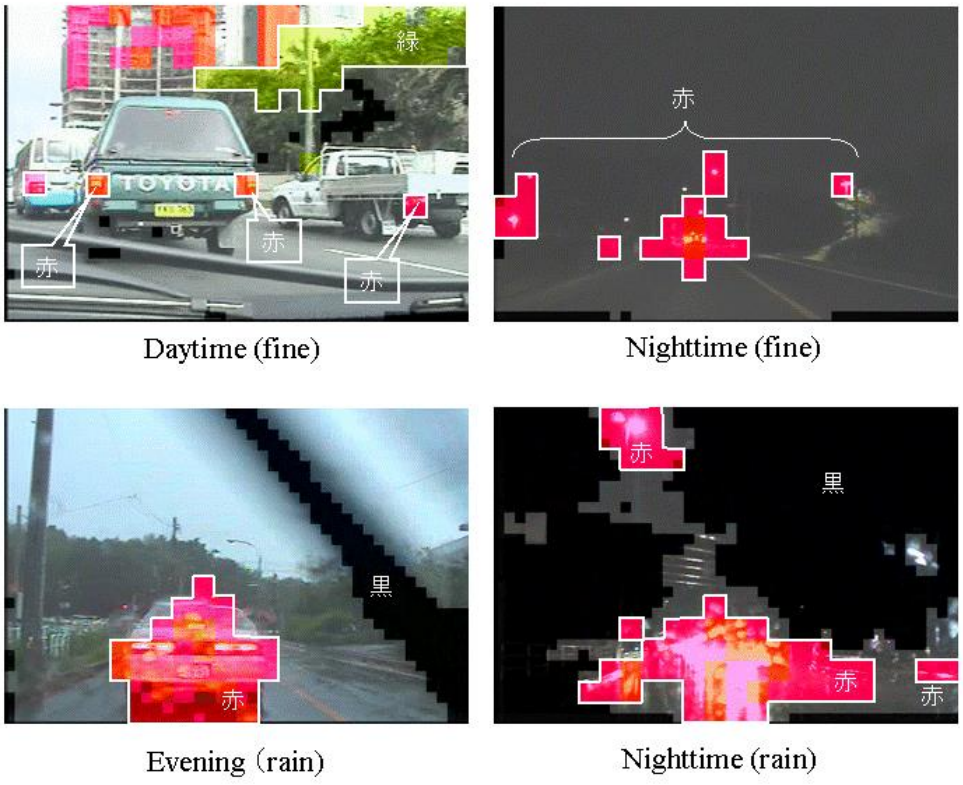


図 4.3 色インデックス画像の例

Fig. 3 Examples of color indexed images.

判定規範は次式で定義される重み付け絶対値距離の最小化を用いた．

$$d_A(A, A_R(k), w(k)) = \frac{1}{W(k)} \sum_{i=1}^{3} w_i(k) |A_i - A_{Ri}(k)| \tag{4.2}$$

ただし，

$\boldsymbol{A} = (A_1, A_2, A_3) = (Y, U, V)$

$\boldsymbol{A}_R(k) = (A_{R1}(k), A_{R2}(k), A_{R3}(k)) = (Y_R(k), U_R(k), V_R(k))$

$\boldsymbol{w}(k) = (w_1(k), w_2(k), w_3(k)),\ W(k) = w_1(k) + w_2(k) + w_3(k)$

であり，$\boldsymbol{w}(k)$は k 番目の色に対応する荷重係数ベクトルである．

4.3.3 車載カメラ映像の色インデキシング例

まず，図 4.2 の型式の YUV 成分の分布を観察した結果から（方法１）の判定条件を定め，これをもとに，（昼間，晴），（夕刻，雨），（夜間，晴），（夜間，雨）における車載カメラ映像のインデキシングを行った．

図 4.3 にその結果を抜粋して示す．目視の評価では，前方車両，対抗車線の車両，信号機，看板，などの赤色領域が良好に抽出されていることがわかった．また，夜間は予想以上にインデキシングが良好に行われていた．従来，画像処理は夜間に弱いという定説があったが，人間は決して真っ暗闇の中を運転するわけではない．ヘッドライトのみでは困難だが，街灯，他の車両，ビルの照明などが多く存在する道路環境では，夜間でも予想以上に主要なオブジェクトを捉えることができた．特に，特徴点に相当する街灯や信号機，ストップランプ，対抗車両などは昼間の画像よりも抽出されやすいことがわかった．

また，雨天ではワイパーが作動することとフロントガラスに雨だれが流れるために画像処理は

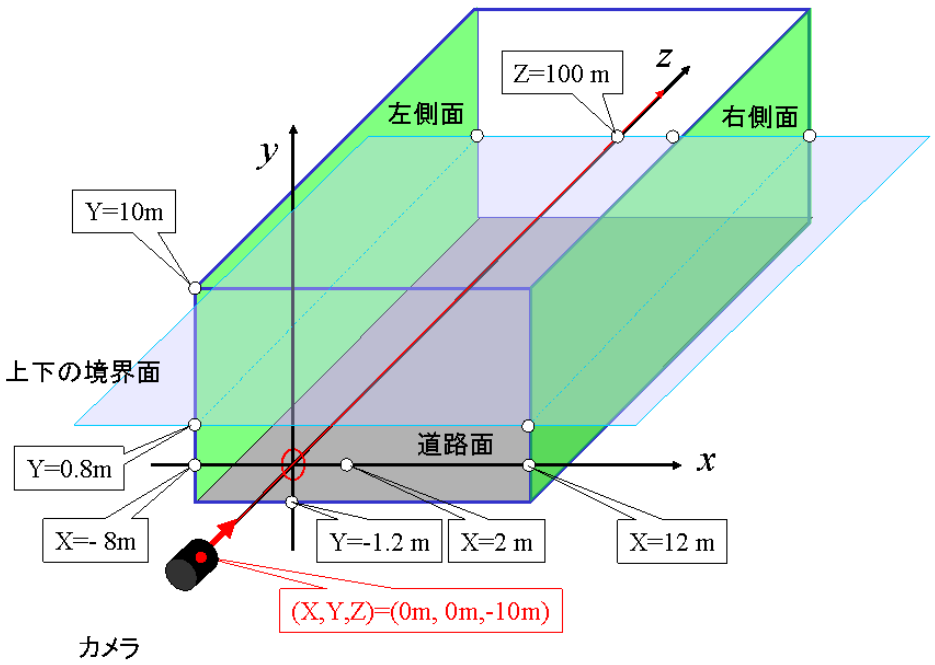


図 4.4 道路シーンの立体構造モデル

幾分乱れる．しかし，乱反射や拡散が一様な色領域を形成するため，大域的には赤色領域（前方車両のストップランプなど）を安定して抽出できそうだということがわかった．なお，画像中のワイパー部分は黒領域となるため，周期パターンを獲得することで比較的容易に除去できる．

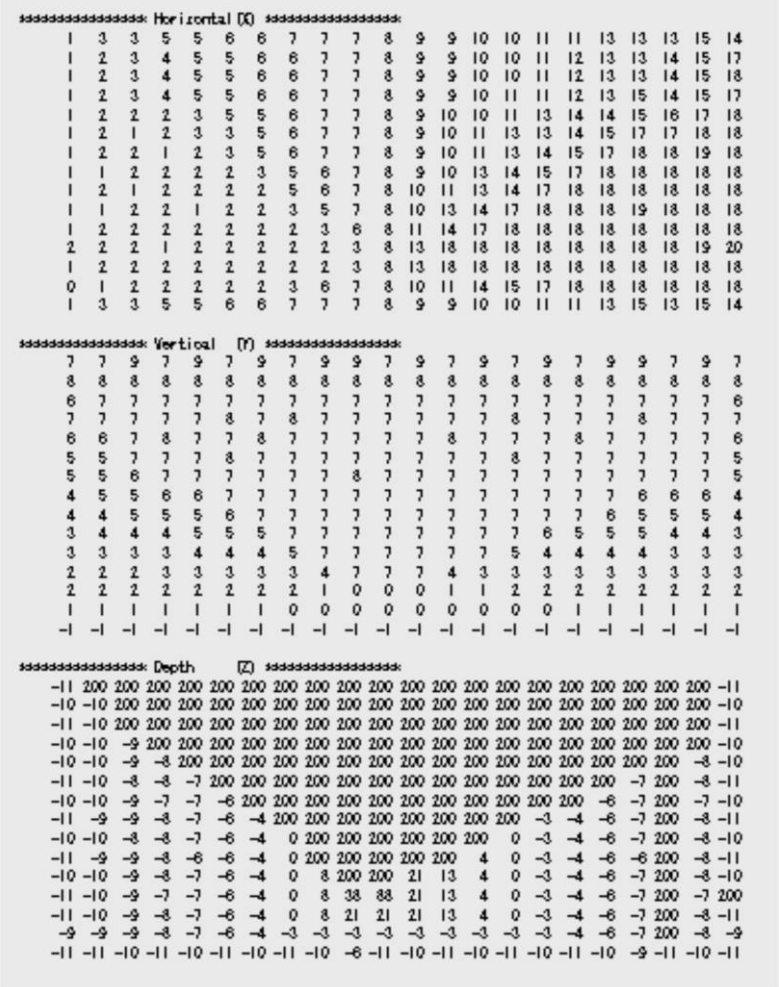


図 4.7 道路構造モデルから推定した距離情報

Fig. 4.7Distance information estimated from road structure model.

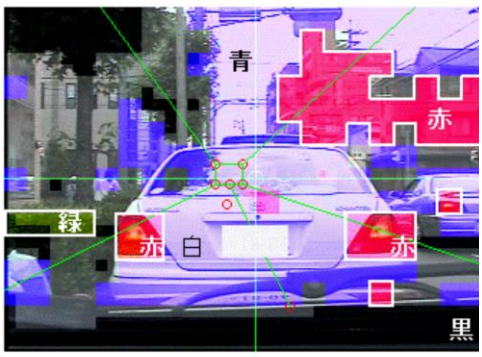


図 4.7 色インデックス画像

Fig. 4.7Color indexed image.

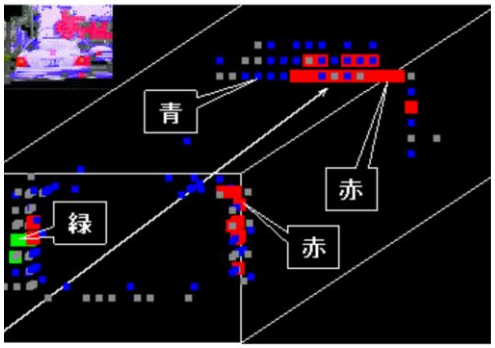
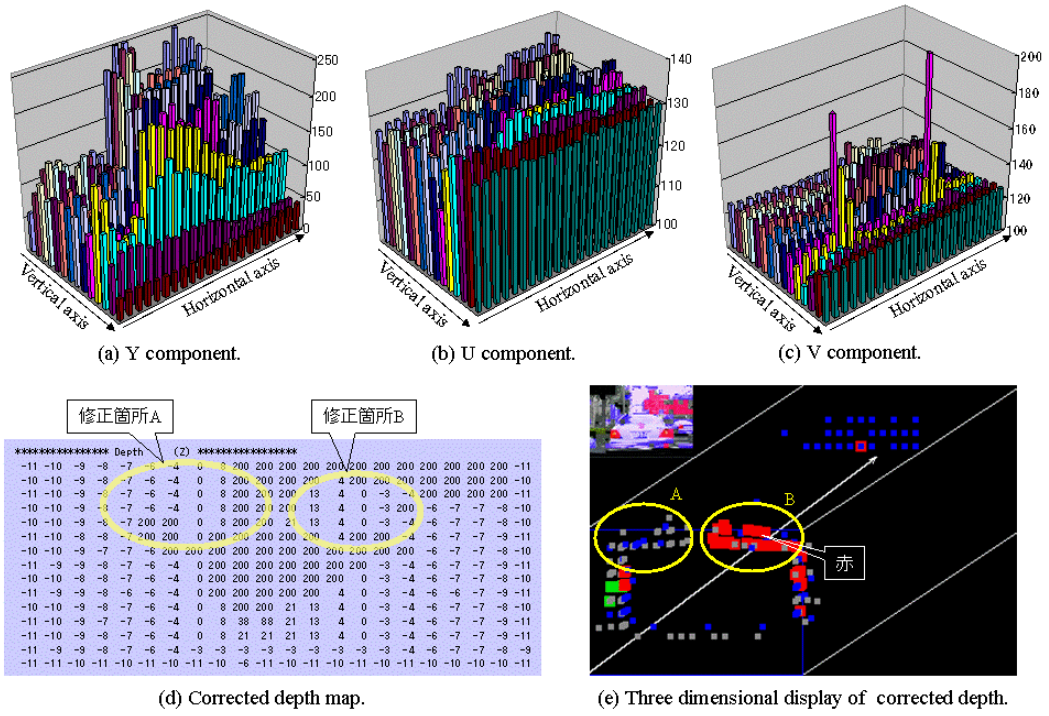


図 4.7 奥行きの３次元表示

Fig. 4.7Three dimensional display of



(a) Y component. (b) U component. (c) V component.

(d) Corrected depth map. (e) Three dimensional display of corrected depth.

図 4.8 空の情報に基づく奥行きの修正

Fig. 4.8Correction of depth on the basis of sky information.

4.4 シーンの構造モデル

単眼画像からシーンの 3 次元構造やオブジェクトの 3 次元位置姿勢を取得するには移動カメラを用いたステレオや因子分解法[D1][D2][D3]，対象物の 3 次元モデル知識を利用するモデルベース手法などが知られている[A2][A3][C11]．本研究では，静止画 1 枚からでもシーン記述を生成できるという観点からモデルベース手法を採用した．これにより，シーンの静的構造モデルを 2 次元画像に割り付け，色インデキシングされたマクロブロックに対応するオブジェクトの概略の 3 次元位置を割り出す．

4.4.1 道路の立体構造モデルの仮定

一般に，走行車両は局所的には前方に直進することが多いため，車載カメラによる道路画像では図 4.4 のような箱型の立体構造モデルを仮定することが考えられる[B3][B7][C12]．これに基づき，投影されたマクロブロックの各々について 3 次元座標を表すマップを作成した．図 4.5 は図 4.4 の箱型のシーン構造モデルのみから得られた距離マップである．これにより，図 4.6 における赤インデックスが付与された「赤いビル」の部分の奥行き情報は 3 次元グラフィックス表示で図 7 のようになる．ここでは空と同じ奥行き位置（便宜上 200m に設定）に「赤いビル」が表示されている．これは明らかに実シーンと相違しており，シーン構造モデルを修正する手段が必要となる．

4.4.2 空仮説によるモデルの修正

そこで，図 4.8(a) に示すマクロブロック単位の Y 成分分布に着目し，空の領域は次式を満たさねばならないという仮説（これを空仮説と呼ぶ）を設定した．

$$Y_{MBK\_mean}(m_r, m_c) \geq Y_{th} \tag{4.3}$$

ただし，

$Y_{MBK\_mean}(m_r, m_c)$：画像中，垂直方向で m_r 番目，水平方向で m_c 番目にあるマクロブロック MBK(m_r,m_c)の Y 成分の平均値．

Y_{th}：設定しきい値

である．なお，MPEG-1 形式のマクロブロック（16×16 画素）中には Y_0 から Y_3 の 4 つの Y ブロック（8×8 画素）が存在するため，次式で計算する．

$$Y_{MBK\_mean}(m_r, m_c) = \frac{1}{4}\sum_{i=0}^{3} Y_{BLK\_mean}(m_r, m_c, i) \tag{4.4}$$

ここで，$Y_{BLK_mean}(m_r, m_c, i)$は $MBK(m_r, m_c)$中の Y_i ブロック（$i=0,..,3$）の平均画素値である．

この空仮説を図 4.5 の距離情報マップに適用すると，図 4.8(d)のようなフィルタリング結果が得られる．これにより，上述の「赤いビル」の部分の奥行き情報は 3 次元表示で図 4.8(e)のように修正された．

一般に経験的事実として，夜間でない限り空は最も明るく広い領域の一つであり，屋外において面光源の役割を果たす．したがって空仮説は上記以外のシーンにも幅広く適用できると考えられる．

4.5　画像領域の分割

確信度を用いてオブジェクトの判別を行う際に，走行映像における経験的な特徴と道路構造モデルの情報を利用することが得策であろう．そこで，図 4.5 の距離マップに基づき，それが投影された画像面を｛左下，正面下，右下，左上，正面上，右上｝の 6 個の領域に分割した（図 4.9）．

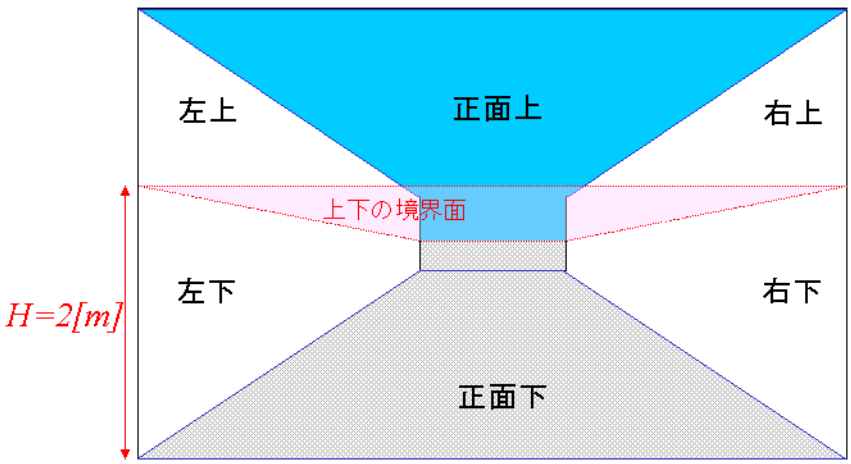


図 4.9 画像面上の領域設定

Fig. 4.9 Area definition on image plane.

以後，この領域情報を付して各ブロックの確信度ベクトルを管理する．

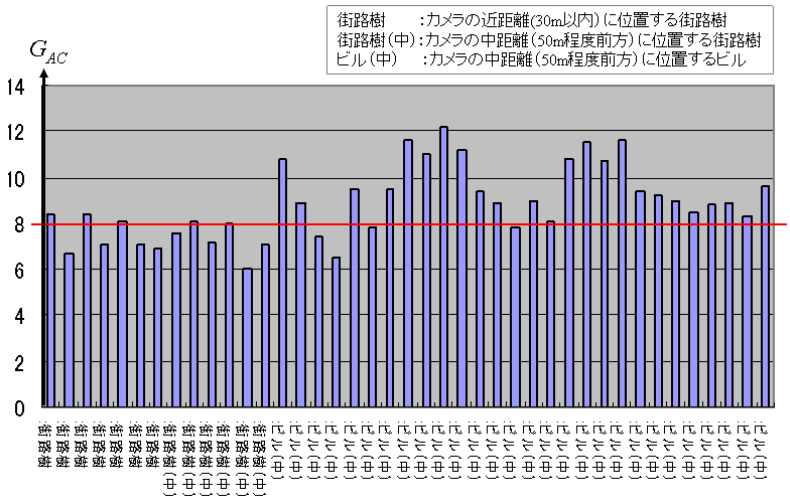


図 4.10　AC 係数における垂直成分の相対強度

Fig. 4.10 Relative intensity of vertical components in AC coefficients.

4.6　オブジェクトの判別

ここでは今回，画像検索条件として用いた個々のオブジェクトの判別手法を説明する．

4.6.1 青の標識

交差点案内標識や予告案内標識に代表される青の標識を認識させるには上述のベクトル量子化で対応した．具体的にはサンプルした複数ブロックについて平均色ベクトルを算出し，代表ベクトルとした．なお，天候や時間帯の違いで代表ベクトルに変化が見られたが，今回は個々の状況について個別に作成した．

4.6.2 先行車両

走行時の安全上重要なオブジェクトの一つとして，絶えず前方に先行車両を仮定する．これを

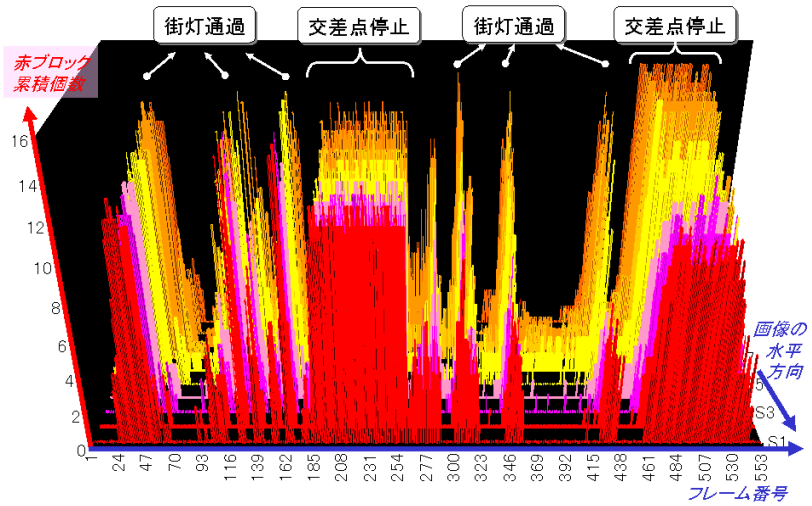


図 4.11赤領域（夜間）に関する時系列ヒストグラム

Fig. 4.11 Time series histogram regarding red regions (nighttime).

車仮説と呼ぶことにする．この仮説に基づき，ブロック画素が先行車両あるいはその一部に該当するという判定の確信度を計算した．具体的には以下の手順による．

まず，水平方向に並ぶ左右２つのテールランプ位置のマクロブロック（赤）を検出し，２つのランプ間の距離を画素値で表現した値を L_R とする．

次に，あらかじめ 5 段階で設定した車両後部の奥行き位置 $Z(m)\ (m=1,..,5)$ に対応するランプ間距離を画素値で表現した $L_V(m)$ のうち最も L_R に近い値を取るときの m に対応する Z(m)を先行車両の奥行き位置と判定する．この結果を距離マップの修正に反映した．

さらに，テールランプの輝度変化を判定することで点灯の有無を判別した．ここで，点灯時の判定は閾値判定とベクトル量子化に基づいて算出した確信度の最大値判定で行った．また，消灯時のランプ検出は主としてベクトル量子化に基づいて算出した確信度の最大値判定で行った．

なお，各フレームで得られた先行車位置の時間的連続性については，フレーム間の相対的な位置変化 D_Z がある閾値 D_{Zth} 以下のときのみ推定結果を採用し，そうでない場合は前フレームの値を保持するという弱い制約を加えた．

4.6.3 ビルと街路樹

ビルの判別に際しては，上述のようなブロック単位の色判定に加え，図 4.9 の領域情報とオブジェクトの大きさに関する先見的知識を利用することで確信度の算出を行った．即ち，図 4.9 の左側面あるいは右側面の空間位置に存在し，かつ同じ色インデックスを付与された隣接マクロブロック群がある個数以上のマクロブロック数で構成されるときに，それが“ビル”（家屋，マンション，コンビニなどを含む）であると判定する確信度を高める操作を行った．

今回，ビルに付与した色インデックスは｛赤，緑｝の２色としたが，ブロック画素の平均色と距離情報のみを用いた予備実験では“緑色のビル”と“街路樹”を誤判定するケースが目立った．そこで，一般に樹木ではビルなどの人工構造物に比べて AC 成分が多いと仮定し，次式で定義する AC 成分発生量 Q_{AC} を調べた．

$$Q_{AC} = \sum_{l=0}^{5} S_{AC}[l] \tag{4.5}$$

ただし，$S_{AC}[l]$ はブロック別の AC 成分発生量であり，$l=0,1,\ldots,5$ は順に Y_0,Y_1,Y_2,Y_3,U,V ブロックに対応する添え字を表す．すなわち，

$$S_{AC}[l] = \sum_{h=0}^{7}\sum_{v=0}^{7} |B_l(h,v)| - |B_l(0,0)| \tag{4.6}$$

であり，$B_l(h,v)$ は逆量子化後の MPEG-1 の DCT 係数ブロックの(v+1)行(h+1)列成分である．

サンプル画像で比較したところ，カメラに対して近距離から中距離（約 50m 以内）にある空間領域では，樹木の AC 成分発生量はビルの場合よりも明らかに高かった．一方，遠景の場合や符号化プロセスに起因してぼけが多い場合では AC 成分発生量にさほど差異はなかった．したがってあるしきい値 Q_{AC_tree} について，

$$Q_{AC} \geq Q_{AC\_tree} \tag{4.7}$$

を満たす場合を樹木，それ以外をビルと判定することは困難である．そこで，樹木の AC 成分発生量は水平方向と垂直方向で比較して余り偏りがなく，一方，ビルでは垂直方向に偏る傾向がある（水平方向の変化が少ない）ことに着目し，次式の特徴量を算出した．

$$G_{AC} = \sum_{l=0}^{5} \frac{G_V[l]}{G_H[l]} \tag{4.8}$$

$$G_H[l] = \frac{1}{64}\sum_{h=0}^{7}\sum_{v=0}^{7}(h+1)\,|B_l(h,v)| \tag{4.9}$$

$$G_V[l] = \frac{1}{64}\sum_{h=0}^{7}\sum_{v=0}^{7}(v+1)\,|B_l(h,v)| \tag{4.10}$$

この G_{AC} は 2 次元 DCT 係数の発生量に関する重心位置の傾きに相当する．映像データからサンプルした街路樹とビルのブロックについて，G_{AC} を算出した結果を図 4.10 に示す．この結果から経験的に，

$$G_{AC} \geq 8.0 \tag{4.11}$$

のとき樹木の確信度を高め，それ以外のときビルの確信度を高める操作を行った．

4.7 周囲状況の判別

ここでは，特に色情報の時系列変化に着目して周囲の状況を判別する方法を考察する．

4.7.1 時系列ヒストグラムの判定

図 4.11 は渋滞のない夜間の直線道路を 5 分間走行した映像をもとに，各イントラフレームで赤インデックスが付与されたマクロブロックの数を水平マクロブロック座標毎に垂直方向に累算した結果得られる時系列ヒストグラムである．

ここで，色インデキシングの予備実験結果（図 4.3）を踏まえ，夜間の光源がもたらす色領域は同一のオブジェクトに対しても YUV 空間内で広い変化範囲を取る傾向にあるという経験的な性質に留意した．即ち，この性質はベクトル量子化に要する代表色の選定を困難とすることから，夜

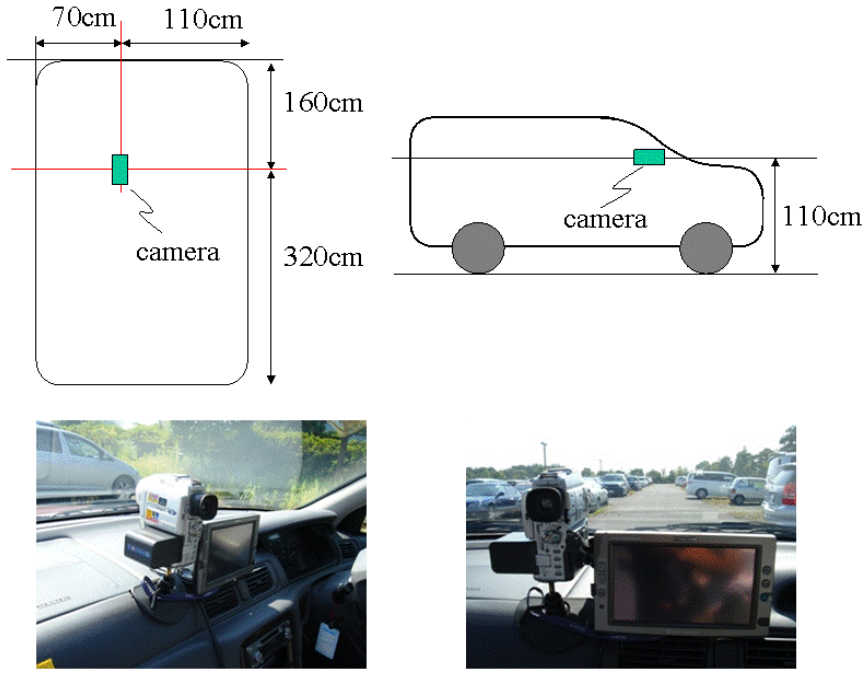


図 4.12　車載カメラ

Fig. 4.12.　On-board camera.

間の赤の検出は YUV の閾値判定で行った．

このようにして得られた時系列ヒストグラムをソース映像と対比することにより，次のような判定規則を経験的に作成した．

RN１）画面右側に累積される赤インデックスのピークは中央分離帯に設置された『街灯下を通過』している状況に対応する可能性が強い．

RN２）画面左側から中央部にかけて 10 秒以上にわたり観測される赤のピークは『交差点で停止』している状況に対応する可能性が強い．

RN3）赤インデックスの個数の急激な増大は『先行車両の停止ランプが点灯』した状況に対応する可能性が強い．

同様にまた，雨の日の夕刻の走行映像についても赤インデックスのヒストグラムを取り，次のような規則を経験的に作成した．

RE1）赤インデックスのヒストグラム値が特に画面左半分で増加する場合は『先行車に接近』している状況に対応する可能性が強い．

RE2）赤インデックスのヒストグラム値が特に画面右半分で増加する場合は『対向車が接近』している状況に対応する可能性が強い．

RE3）赤インデックスのヒストグラム値が画面中央部から連続的に増加する場合は『交差点に接近』している状況に対応する可能性が強い．

以上の各規則における“可能性が強い”は対応する状況への確信度を高める操作に反映される．

4.7.2 危険箇所の判定

フレーム間の差分電力が大きいほど運転中の視覚への負担は大きいと仮定し，次式でフレーム番号 n のシーンにおける危険度を算出する．

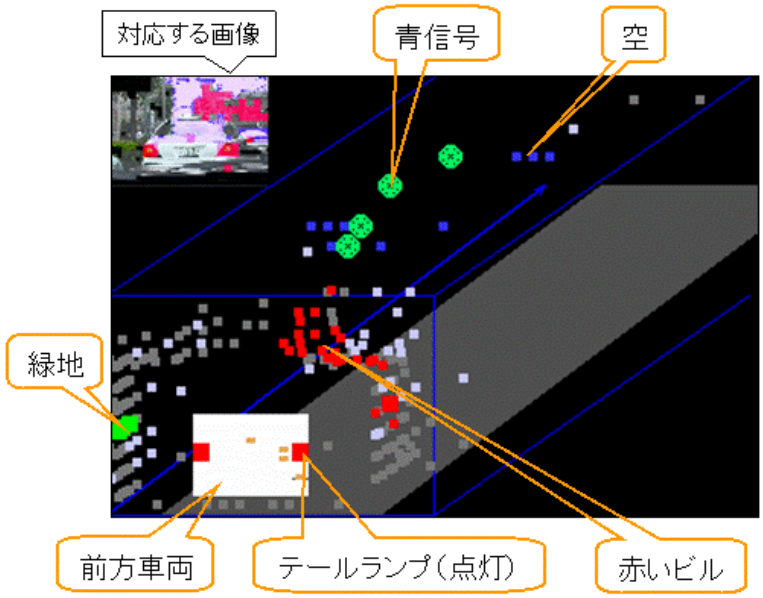


図 4.13 推定結果の 3 次元表示

Fig. 4.13 Three dimensional expression of estimated results.

表 4.2 映像検索実験の結果

Table 4.2 Experimental results of image retrieval.

映像						検索語	検索結果						
番号	時間帯	天候	道路の分類	撮影日	N_A		N_R	N_P	N_Q	S	a_r	a_p	F
1	朝	晴	住宅街 一般道	2004/6/10	300	緑色のビル＆注意箇所	12	10	0	0.04	1.00	0.83	0.91
2	朝	晴	住宅街 一般道	2005/11/21	202	青の標識	22	11	6	0.11	0.65	0.50	0.56
3	昼	晴	郊外 有料道路	2005/10/20	277	青の標識	34	26	6	0.12	0.81	0.76	0.79
4	夜	晴	中央分離帯のあるバイパス	2005/10/20	277	先行車の停止ランプ点灯	7	5	0	0.03	1.00	0.71	0.83
5	夜	晴	中央分離帯のあるバイパス	2005/10/21	60	街灯下を通過	9	9	1	0.15	0.90	1.00	0.95
6	夕	晴	住宅街に面した大通り	2005/10/20	277	先行車の停止ランプ点灯	3	3	3	0.01	0.50	1.00	0.67
7	夕	雨	郊外の一般道	2004/6/25	315	対向車が接近	48	48	31	0.15	0.61	1.00	0.76
平均	—	—	—	—	244.0	—	19.3	16.0	6.7	0.09	0.78	0.83	0.78

$$D_n = \sum_{k=0}^{K-1} \left\{ W_k \sum_{i=0}^{5} \left| Z_0[i, N_D(k), n] - Z_0[i, N_D(k), n\text{-}1] \right| \right\} \tag{4.12}$$

ただし，

$Z_0(i,j,n)$：フレーム番号 n の画像中の j 番目のマクロブロックにおける i 番目のブロックの平均画素値（DCT 係数の DC 成分）

K：危険領域の判定対象となる領域に含まれるマクロブロックの個数

$N_D(k)$：危険領域の判定対象となる k 番目のマクロブロックの番号

である．なお，画像周辺部では危険度とは無関係に，通常の直進走行においてもフレーム間変化は大きくなると考えられる．そこで，危険度の判定対象領域でも特に消失点周辺に近いほど重み付け W_k を大きくした．

4.8 実験

以上に基づき，映像中の各マクロブロックにオブジェクトインデックスを付与し，認識結果を XML で自動記述する．また，人間が直感的にその認識結果を把握できるように 3 次元グラフィックス表示を行う．さらに，XML の記述をもとに画像検索を行い，その検索効率を評価する．

4.8.1 実験システム

市販ディジタルビデオカメラ（DV 形式）をカーナビディスプレイ位置の背後に設置し，フロントガラスごしに前方の道路状況を撮影した（図 4.12）．撮影した映像音声を PC 上で MPEG-1 形式に変換し，映像データのみを抽出して図 4.1 のシステムの入力とした．

4.8.2 認識結果の視覚表現

本システムは，色インデックス画像（図 4.6）をトリガーとしてシーン構造モデル（図 4.4），空仮説（図 4.8），車仮説，領域情報（図 4.9），オブジェクト固有の特徴量，時系列ヒストグラムなどから得た一連の情報をルールベースで判定し，最終的なシーンの認識結果を出力する．ここで，画面中の各ブロック画素に対応する 3 次元座標を算出し，車両前方 200m の範囲内で本システムが認識したオブジェクト（あるいはその断片）のカテゴリと概略位置を視覚表現した（図 4.13）

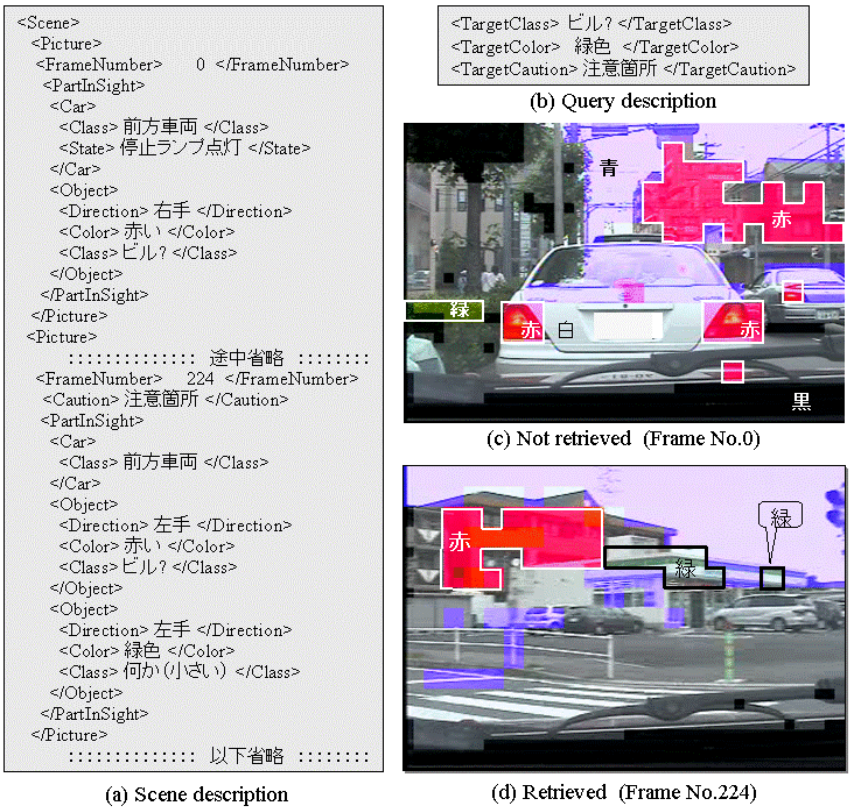


図 4.14 シーンの記述と検索

Fig. 4.14 Description and retrieval of scene.

4.8.3 メタデータの生成と画像検索

上記の色，オブジェクトなどのインデキシング結果を XML で表現したメタデータを図 4.14(a)に示す．これに対し，図 4.14(b)の形式で問い合わせプロファイルを作成した．これは「注意箇所」,「緑色のビル」といった自然言語表現に対応する．この問い合わせプロファイルでメタデータを検索した結果の一例を図 4.14(d)に示す．

検索の効果は検索結果をブラウズする際の画像枚数と画像内容の妥当性で判断する．今，全体の枚数 N_A に対して検索された画像の枚数を N_R とすると，ブラウズに要する時間の短縮率 S は

$$S = N_R / N_A \tag{4.13}$$

で表せる．また，検索された画像中で内容が妥当な画像の枚数を N_P，内容は妥当だが検索されなかった画像の枚数を N_Q とすれば，検索の再現率 α_r と適合率 α_p は次式で表せる[J1].

$$\alpha_r = \frac{N_P}{N_P + N_Q} \tag{4.14}$$

$$\alpha_p = N_P / N_R \tag{4.15}$$

したがって，検索効率として再現率と適合率の調和平均で定義される F 尺度[J1]を用いれば，

$$F = \frac{2}{\alpha_r^{-1} + \alpha_p^{-1}} \tag{4.16}$$

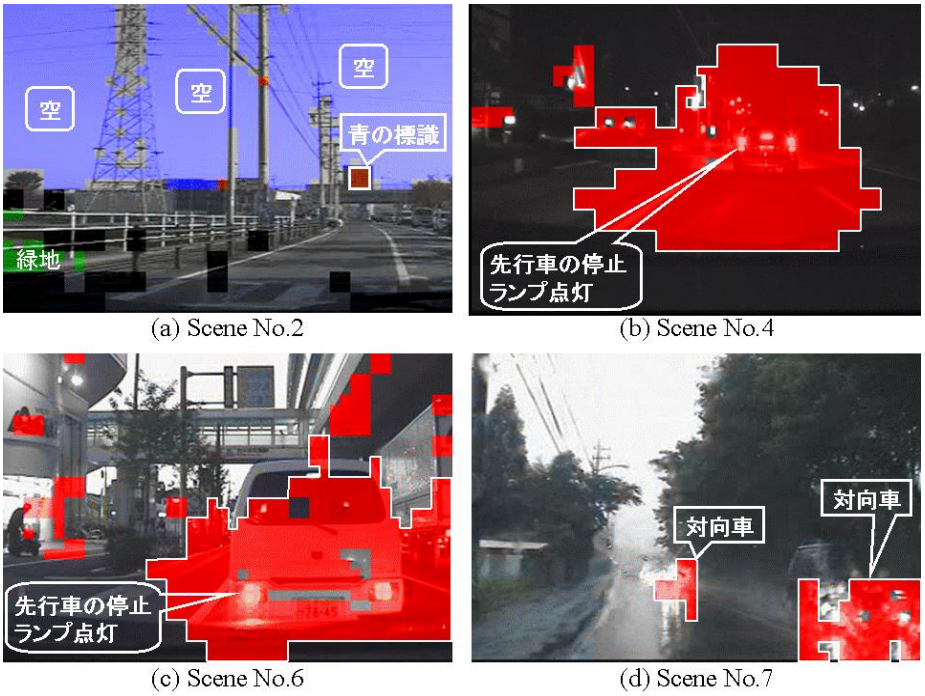


図 4.15 検索結果の例

Fig. 4.15 Examples of retrieved results.

となる．ヒューマンインタフェースの観点からは F（$0 \leqq F \leqq 1$）を最大化し，S（$0 \leqq S \leqq 1$）を小さく抑えるような検索特性が要求される．

2 フレーム／秒の MPEG-1 イントラフレームの時系列で構成される 2〜3 分程度のシーンを複数個用意し，検索実験を行った結果を表 4.2 と図 4.15 に示す．

4.9 考察

本稿の手法で車載カメラ映像から自動生成した XML 記述を用い，画像検索を行った結果，ユーザがブラウズに要する時間は平均で 1/10 以下に短縮でき，最短では 1/100 となった（表 4.2）．また，検索効率 F は平均で 0.7 以上，最大で 0.95 を示した．一方で，今回生成した映像記述のもとになるインデックス画像を目視で考察したところ，誤判定を含むが，特定色のビルや注意箇所，前方車両，信号などが指定した表示色でインデキシングされていることを確認した．今後は，より誤判定を低減させるために推論規則を充実させていくことが必要である．また，色判定の適応化など IF-THEN ルールのみでは記述困難な処理については 5 章以降に示すような統計的推論手法の併用が考えられる[A9][A10][A11][A12]

4.10 むすび

MPEG-1 形式の車載映像から 2 次元 DCT 係数を輝度・色情報として抽出し，ルール判定を施すことで，検索用途のインデキシングを実現できる見通しが立った．さらに，確信度を用いた知識処理や時系列処理を併用して比較的良好に映像記述が行える可能性を見出した．今後は天候や道路状況の変化に対応して推論特性を適応化できる認識手法に発展させたい．

5. 景観認識を用いた歩行者認識の改善

本章ではプローブカーベースの安全情報システムを想定し，その中で各車の歩行者認識性能を景観認識で改善する方法を述べる．想定するシステムは大きく分けて５つのパートからなる．すなわち，シーン解析，知識獲得，知識共有，予測推定，情報提示の５つである．中でも特に魅力的な応用のひとつとして，シーン記述データベースに統計的手法を適用することで歩行者の出現を予測する機能があげられる．この詳細は 10 章で述べるが，それに迫るためのキー技術をここでは紹介し，いくつかの実験結果を示す．

5.1 はじめに

従来，ドライバは交通渋滞や規制に関する情報を主として車載端末を通じて VICS (Vehicle Information and Communication System)その他の無線手段で知るのみであった．しかしながら，より確かな安全をすべてのドライバに約束するために，それぞれの車載端末は他の車両や歩行者に関してより詳細な挙動を知る必要がある．そこで，近年ではプローブカーベースの統合的な画像センシングと知識共有の方式が提案されている[A9][A14][A15]．これは運転における安全度（または逆に危険度）の予報を行おうとするものである．

5.2 システム

運転の安全に関するいくつかの数値的指標は心理的要因，交通要因，地理的要因などからなると考えられている．本章では，主として交通要因に焦点をあて，それらの基本的なデータがプローブカーシステムによって収集されていることを仮定する．

プローブカーシステムは ITS の分野では最もよく知られたセンサーネットワークの一つである．多くの ITS 関連の団体は試験的にこのプローブカーシステムを交通情報のリアルタイムデータマイニングツールとして用いてきた．いくつかの会社ではすでにそれを交通渋滞予測に用い，それがユーザの要求するどこのエリアでも行えるようにしている[I6][I7]．プロファイルベースの情報システムや画像データベースについても利便性への応用の観点から提案されている[I1][A5]．

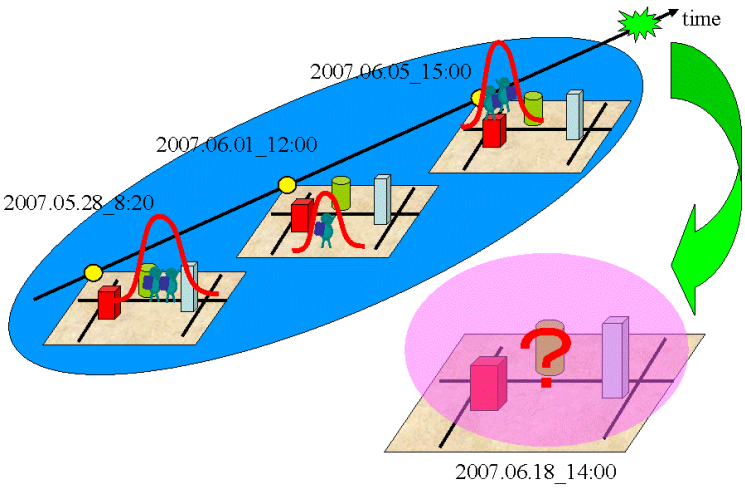


図 5.1 時空間分布に基づく歩行者の出現予測の概念

Fig. 5.16. Concept of prediction of pedestrian appearane from spatio-temporal distribution.

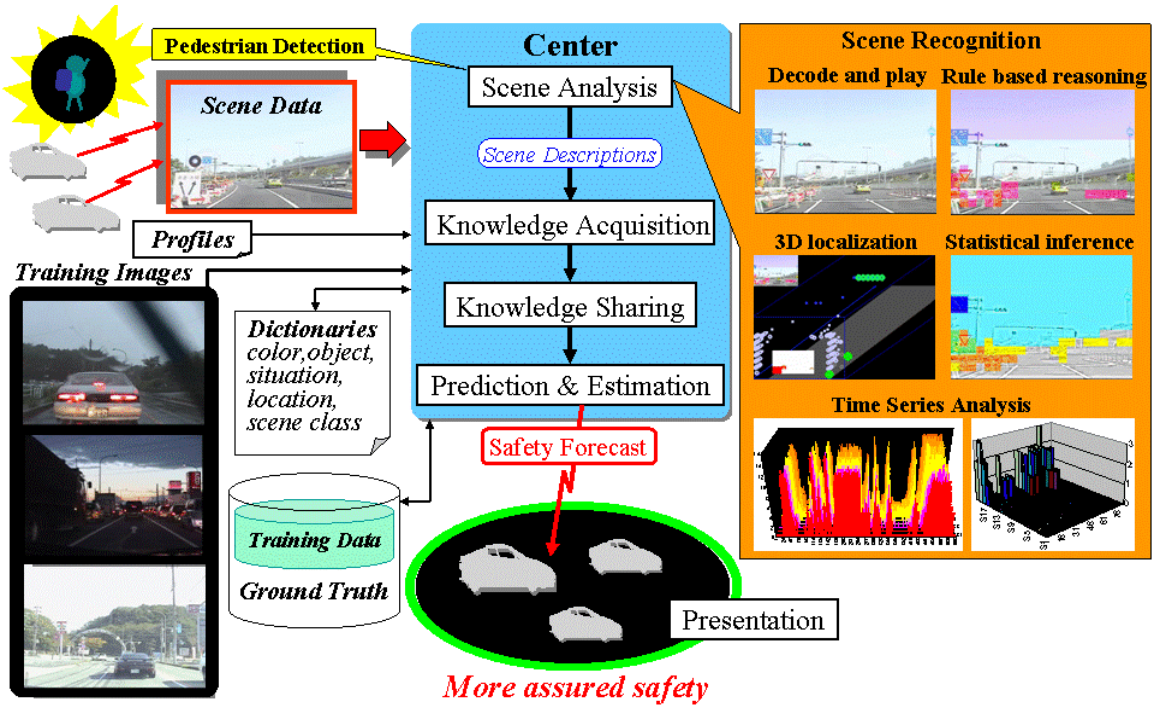


Fig. 5.19. Total scheme for safety forecast.

ここでは安全予測，とりわけ，歩行者の出現の予測に関して図 5.1 のようなコンセプトを提案する．この機能を技術的に実現するために，図 5.2 に示すような全体像を考える[A6][A9].このコンセプトは以下の 5 つのフェーズからなる：

(1) シーン解析

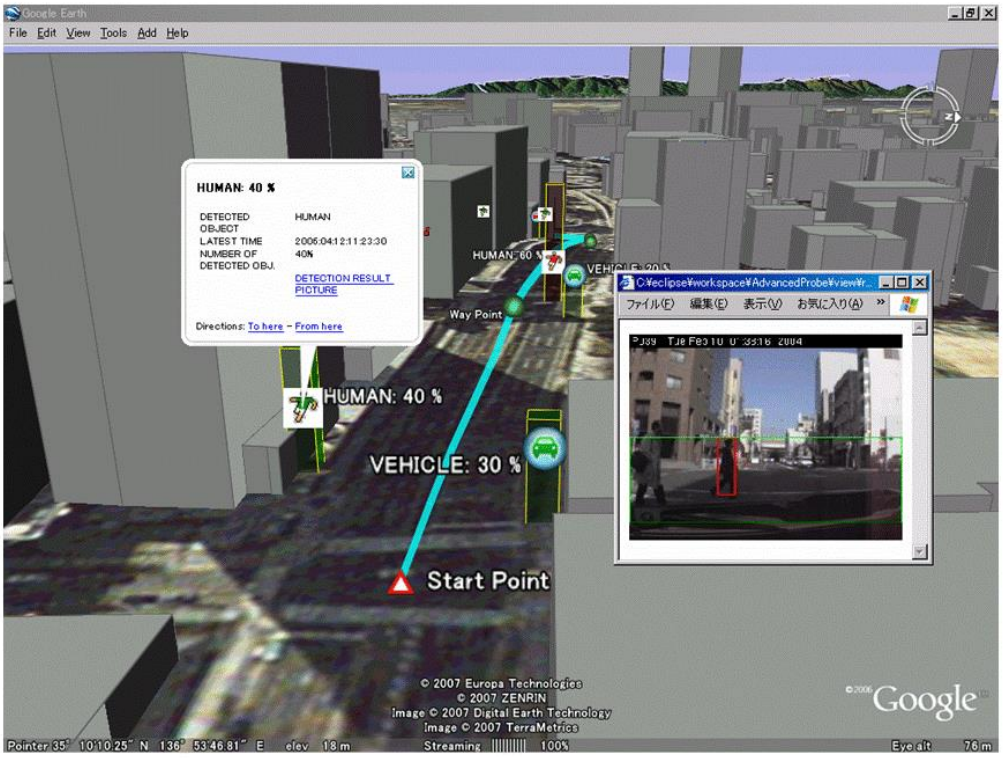


図 5.17. 出現確率の鳥瞰表示．

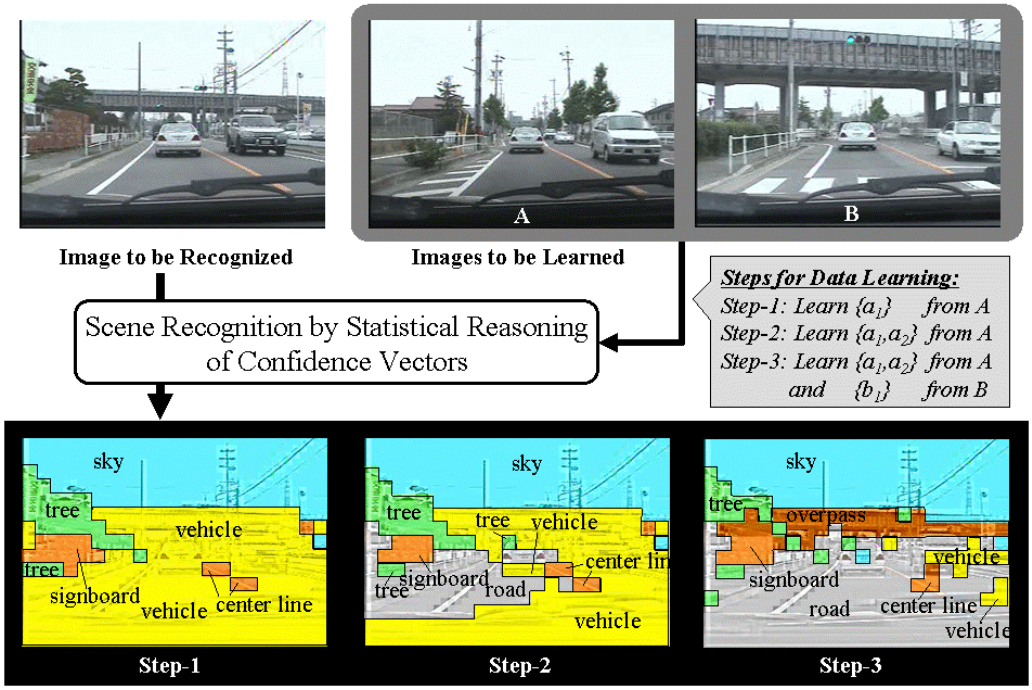


Fig. 5.20. Experimental results.

車載カメラで撮影されたそれぞれのシーンを解析して歩行者およびそれに関係する主要なオブジェクトや特徴量を検出する．それらは XML 形式のシーン記述として出力される．

(2) 知識獲得

上記のシーン記述を無線ネットワークを通じてセンターデータベースに蓄積する．そのデータベースに統計的解析と時系列解析を施し，知識を獲得する．

(3) 知識共有

上記の知識を共有することにより，各車におけるセンサや認識器はより訓練された特性を獲得する．さらには，人間が経験的にインデックスを付与したいくつかの画像はセンシングの Ground Truth として共有される．

(4) 予測と推定

歩行者出現に関する予測と推定は検出された歩行者の過去のデータベースについて多次元次空間解析を行えば可能である[A14][A15][G4]．Particle filtering [E3], [E6] [G4]はその中でももっとも有望な手法である．時間の観点ではトレンド，季節，週，曜日，および一日，時刻の成分が発生するであろう．同様に，それらはランドマーク要因や住居エリアに密接に関係する空間的成分も有するであろう．

(5) 情報提示

ある種の警告操作が最終的な目標であるとしても，システム設計者にとっては一瞥してその数値的背景をチェックできることが強く要求される．そこで，視覚的提示が熱望されている．我々は Google Earth ベースのバードビューモジュールを設計し，仮想都市画像上で安全度を重畳表示してみた（図 5.3）．

5.3 シーン解析

5.3.1 景観認識

第 3 章で述べた統計的手法により，ある画像領域が景観を構成する有限個のオブジェクトの投影であると判定する際の確信度をそれぞれのブロック画素について算出する．本章では歩行者認識の補助的手段としてシーン解析を行うため，2 フレーム／秒の INTRA 画像について MBK 単位で DCT 係数の多変量回帰分析を行った．これにより，各 MBK に付与されるべき未知の確信度ベクトル $\boldsymbol{C}$ を様々な状況要因の組み合わせによって推定することができる．この $\boldsymbol{C}$ の最大成分に相当するインデックスを第 1 次の認識結果とする．

図 5.4 に実験結果の一例を示す．過去の画像の特徴を事前にシステムに学習させておくことにより，システムは新たに入力された画像データ（事前に MPEG-1 規格で符号化）中の各画素ブロック(16×16)について確信度ベクトルを計算する．そして事前に用意した運転環境に関する 64 個の語彙を含む辞書において，最も適合するインデックスを一意に決定する．図 5.4 は適切な学習データの増加により，景観認識性能が段階的に向上していく様子を示している．

5.3.2 歩行者検出

確信度による景観認識(以下，CVSCR)とパターン認識による歩行者検出(以下，PPD) の統合化をはかる．PPD にはニューラルネットや SVD などがある．これは PPD が単独では疑似的な人間を識別できず，また，隠れや形状変化，および常識的判断に対しても様々な欠点を有するからである(図 5.5).そこで，本稿ではニューラルネットベースの PPD でかなりよく学習はしているが完全に適応はなされてない程度のものを用いる．そしてこれを一定のレベルに保ち，追加学習は行わなわなかった．これによって，PPD 単独の場合と PPD と CVSCR を統合した場合の比較を，歩行者検出性能に関して数量的に行うことができる．

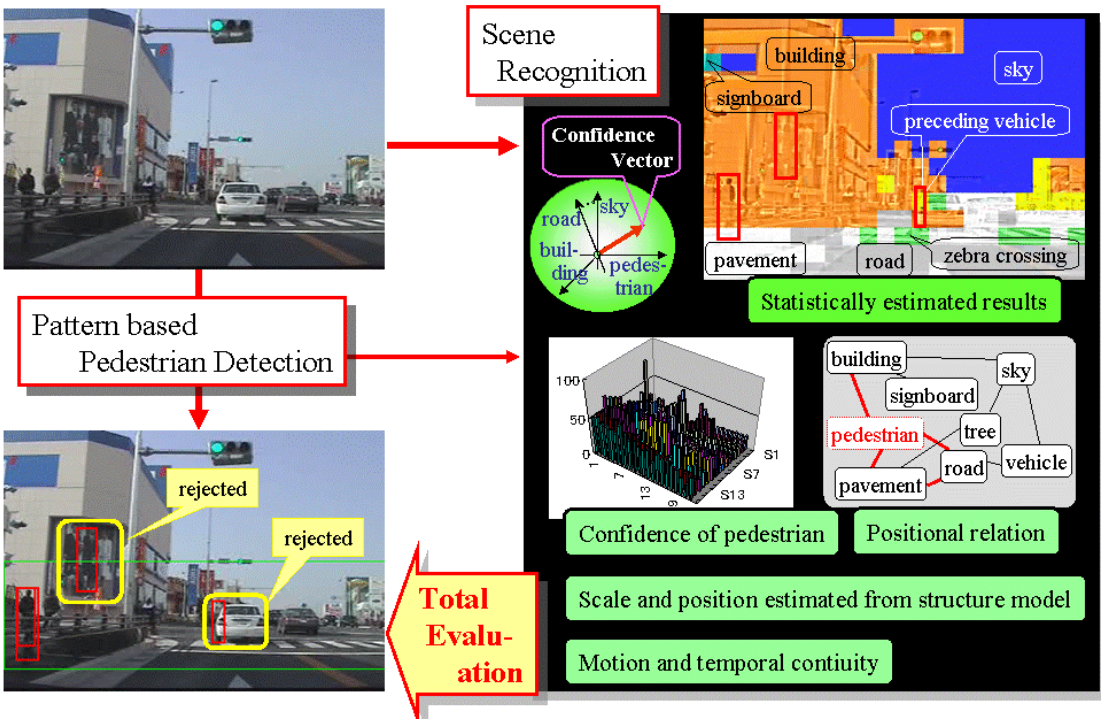


図 5.21. インテリジェントな歩行者検出のフロー．

Fig. 5.22. Flow of intelligent pedestrian detection.

5.4 実験結果

2004 年から 2005 年にかけて，JARI（日本自動車研究所）が行った名古屋プローブカー実験において，約 50 万枚の静止画像が専用車の車載カメラで撮影され，位置情報と時間データが付与された．撮影されたエリアは栄地区であり，名古屋でも最も人気のある中心街のひとつである．この画像セットを試験的に用いた．各画像のサイズは 360×240 画素である．一つの画像と次の画像の間の時間間隔は最小で 10 秒とされている．これらの画像を用いた提示フェーズの一例を図 5.3 に示す．これは人と車両の出現確率を GoogleMap 上で可視化したものである．

A. PPD の数値評価

しかし，本章で用いた PPD ではすべてのシーンにおいて認識結果が実際の人の出現状況に一致するとは限らない．樹木，ビル，壁の上のポスター，車両のエッジなどが PPD における形状パターンの認識プロセスにおいて，しばしば疑似人物を作り出し，誤検出を発生させる．これは現状の PPD が形状ベースの静止画パターン認識であり，しかも，衝突安全の観点から道路領域に特化した手法であるためである．また，屋外での車載カメラ撮影であるがゆえに様々な画像の不鮮明や隠れ状況を作り出し、未検出を誘発する。

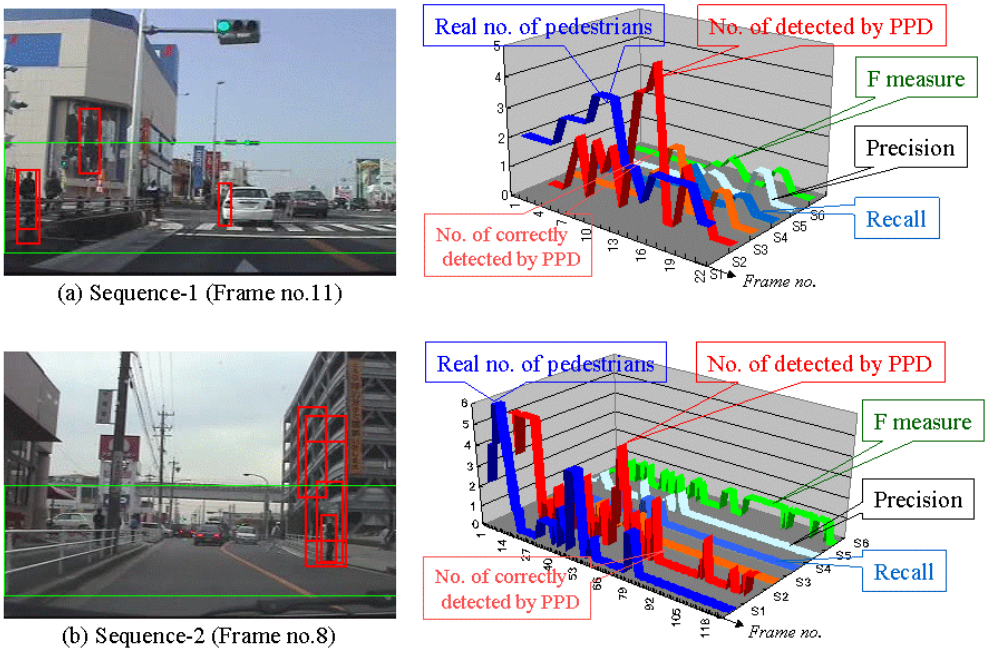


図 5.23 PPD の実験結果

Fig. 5.24. Experimental results of PPD.

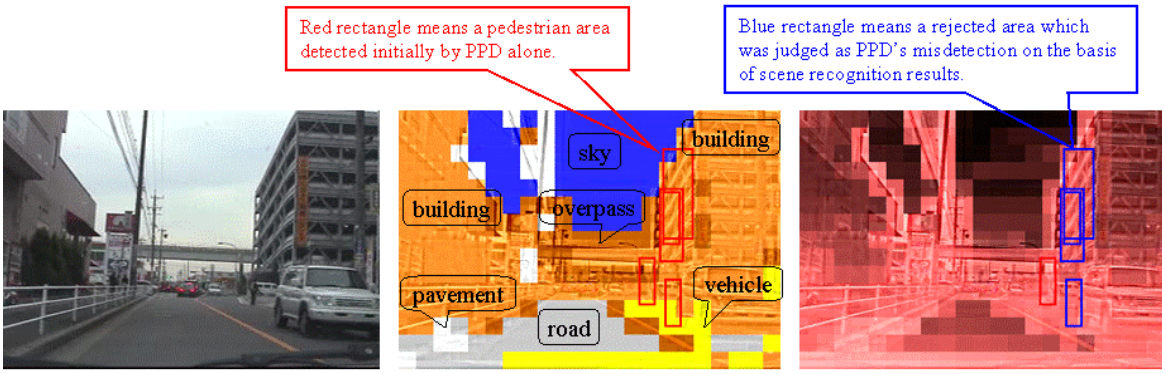


Fig. 5.7. Experimental results of performance improvement.

そこで，歩行者検出タスクの性能を定量評価するために，F 尺度を検出率の計算に採用した．3 章での議論と情報検索問題からの類推により，再現率，適合率，F 尺度はそれぞれ次式で表現される．

$$\alpha_r = \frac{N_P}{N_P + N_Q}, \quad \alpha_p = \frac{N_P}{N_R}, \quad F = \frac{2}{\alpha_r^{-1} + \alpha_p^{-1}} \tag{5.17}$$

ここで，N_P は正しく検出できた歩行者の数，N_Q は検出できなかった歩行者の数（未検出），そして N_R は PPD が検出したすべての歩行者の数であり，ここには誤検出も含まれている．

図 5.6 は道路シーンに関する 2 つの代表的なシーケンスの実験結果である．各シーンはフレームレート 30fps で DV 形式(720×480 pixels, 30 fps)で撮影されているがそれを MPEG-1 (352×240 pixels, 30 fps)にダウンコンバートした．そして，扱いやすくするために 2fps のフレームレートで INTRA 画像だけを処理した． 各サンプル画像中に描画した赤い長方形は PPD で検出した歩行者領域を図形的に示すものである．ここで，PPD による最大検出数を 5 に設定した．Ground truth は人間のオペレータによりマニュアル入力され，車載カメラから約 100m 離れた範囲内の歩行者について定義した．また，この操作は，ビデオシーケンスを処理フレームの前後を含む時間範囲でブラウズしたときに，ターゲットが少なくとも部分的に見えており，知覚されていることを条件として行った．

その結果，シーケンス 1 については，23 フレームの連続する INTRA 画像について F 尺度の全平均が 0.14 であった．また，シーケンス 2 については，連続する 124 フレームの INTRA 画像についての F 尺度の全平均は 0.49 であった．

これらの結果は，正解の規定（特にカメラから認識対象とする歩行者までの距離を何 m までとるか、陸橋や高架、ビル内の人を対象とするかなど），シーンの特徴（背景や周囲にどのようなオブジェクトがあるか、交差点や横断歩道の前であるか、など），および歩行者の見え方の特性（隠れ、後姿、横向き、屈んだ姿勢など）に密接に依存する．このような歩行者にまつわる様々な状況要因については，7 章（移動体検出）や 10 章｛予測と認識の相互作用｝においてさらに詳しく考察することとし、本章では代表的なシーンに関する実験の程度にとどめる．

B. シーン認識による改良

そこで，これらの誤認識結果を除去すべく，シーン認識の結果（図 5.5）で駆動されるルールを適用することを試みた．

第 1 のルールは，人間らしいオブジェクトのそれぞれについて位置と高さの制約を加えることである．ここで，各ブロック画素はおおまかな 3 次元位置を 3D モデルベース手法により持っていると考える．4 章（図 4.4　　道路シーンの立体構造モデル）で用いた 3D 道路シーン構造モデルをここでも用いる．このモデルにおける各量子化点は，消失点の 2 次元位置に基づいて，透視変換により画像上の対応位置に写像される．これは平均色や動きベクトルといった MPEG-1 ベースのブロック属性に基づいて計算される．

したがって，高さ，幅，そして 3D 位置を画像中のすべての赤い矩形領域について計算でき，人間の体の観点でそれらの妥当性を評価できる．

第 2 のルールは赤い矩形領域とその背景との間の意味的関係である．シーン認識の結果，それらの領域がどのカテゴリに属するかを判別できる．例えば，もし赤い矩形領域が“ビル“領域によって完全に取り囲まれているならば，その背景領域もまた,”ビル“である．上記で定義した第 1 のルールは，現在処理中の赤い矩形領域を棄却することが正しいのかどうかを判定する．この”ビル“が”樹木“や”歩道“など，車道の外にある他のオブジェクトの単語に置き換えられるときも同様のことが言える．しかし，もし背景が”空“であるならば，カメラの近くに位置する人間は，透視投影の性質により，巨大なオブジェクトとして現れる．この場合，この解釈は幾分難しく，時間と空間の連続性に関するいくつかの観測が必要になる．

図 5.7 に実験例を示す．ここで，青い矩形は上記で記したルールに基づいて棄却した赤い矩形を意味する．上述の第 1 のルールを不等式で表現し、その閾値を，足の位置，高さ，歩行者を表す矩形領域の幅，のそれぞれについて，3.0m，3.0m，5.0m とした．これらの棄却操作の結果，誤検出の 82%を削減することができた．図 5.8 に 124 個の連続する INTRA フレームにわたって，F 尺度の改善度に関する時間的性能を示す．結論として，本手法により F 尺度は全フレーム平均で 0.49 から 0.59 に改善された．

もうひとつ考慮すべき要因は歩行者の確信度そのものであり，これは学習データによっても非常に影響を受ける．図 5.7(c)では歩行者に関する確信度の 2 次元分布を可視化した．この図では各ブロック画素の輝度が高いほど投影されたオブジェクトが歩行者である確信度が大きいことを意味する．もともと，この確信度は本手法のシーン認識プロセスでは常に計算され，評価されている．それは“歩行者”がオブジェクト辞書中の 64 語彙に含まれているからである．しかし，本章で用いた CVSCR は 2fps の INTRA 静止画のみ処理対象とし，しかも対象画像中で歩行者が占める面積が小さいため，歩行者認識そのものは CVSCR では行っていない．DCT 係数と動きベクトルを用いた歩行者認識については 7 章で述べることにする．

5.5 安全性予測に向けて

歩行者検出における誤検出を削減するため，可能な方法をいくつか提示した．しかし，いまだなお，重要でありながら未解決の問題が未検出のケースについて残されており，それらのケースは主として不明瞭または雑音の多い形状，隠れ，そして突然の形状変化などによって引き起こされる．大体 100m ほど離れた歩行者についてのいくつかの簡単な事例では，HDV カメラ(1920×1080 pixels, 30 fps)とズーミング操作は有効であるらしい．しかし，実際は，非常に複雑なケースが存在する．無線技術との統合がより有望な方法へと導くと確信している．コンピュータビジョンや人間の挙動理解という観点では，時系列解析と高分解能の動き推定技術を導入していく必要がある．

5.6 おわりに

交通事象の予測を可能にするプローブカーシステムのコンセプトを紹介し，その中で画像符号化ベースの統計的景観認識によって従来の歩行者検出性能の限界を克服する方式を提案した．本章のコンセプトは 10 章「予測と認識の相互作用」においてその応用例である「人予報」に継承される．したがって，より定量的な議論は 10 章で行うことにする．

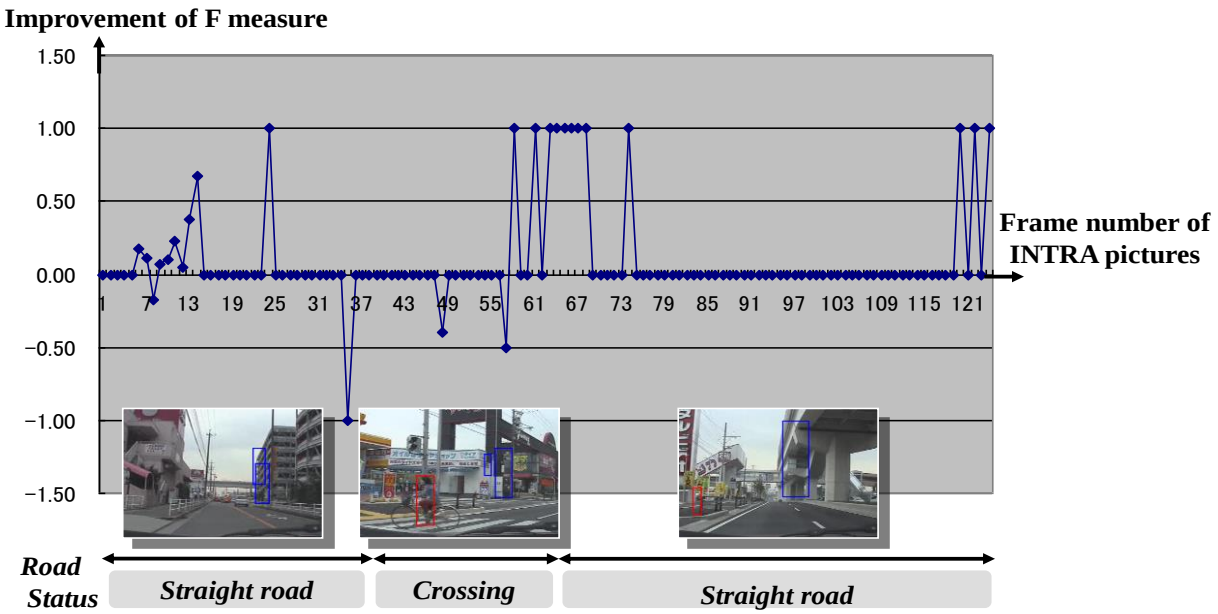


図 5.8 F 尺度の改善

Fig. 5.8. Improvement of F measure.

6. 統計的推論に基づく景観認識

本章では車載カメラの映像から周囲の景観を認識するための統計的手法を提案する．特に，車での記録や無線通信に適した方式をめざして MPEG 映像に着目し，その符号化特徴量と状況情報からブロック画素単位で特徴ベクトルを形成する．この特徴ベクトルからオブジェクトインデックスに関する確信度ベクトルを推定し，最大確信度のインデックスを認識結果とした．この推定には，過去の学習結果を利用した逐次更新や分散協調が必要であるという応用上の観点から，多変量回帰分析を用いる．５つの異なる場所で 2 年間に渡って撮影した走行映像群をもとに提案手法で交叉学習実験を行った．特に今回，シーン全体と主要オブジェクト 10 種類について各々の認識率を算出し，教示データ量，場所，時刻，季節，天候，およびフレーム時間との関係を考察した．
キーワード：景観認識，MPEG，確信度，多変量回帰分析

6.1 はじめに

近年，車が周辺の景観を記憶し，そこから得た知識を他の車両と共有することにより，運転時におけるドライバの利便性や安全性を飛躍的に高めるような情報提供サービスが提案されている[A6][A7][A8]．ここでは，センシングした映像情報をプローブカーに代表される無線通信ネットワークでやりとりすることが想定される．そこで，インターネットやマルチメディアの分野で普及している MPEG 系の画像圧縮技術との親和性が高い映像処理方式が，相互運用性の観点から強く望まれる．すなわち，圧縮データを解凍してから全画素に認識処理を施すのではなく，極力，MPEG 符号化データの持つ信号属性で構成した画像特徴を有効利用して認識することが演算量削減と通信帯域の有効利用の鍵となる．さらに，解凍結果をフレームメモリに展開する必然性もなくなる．また，5 章でも示したように，歩行者認識に景観認識を組み合わせることで，従来のパターン認識手法では回避困難であった誤認識を削減できる．

このような観点から，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が提案されている[A7][A8][A9]．これは従来の画像理解手法が主として 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は応用上の課題として切り分ける傾向にあったことに対し，新しい方向性を示している．筆者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．この色と画面上の 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など特定の対象については行われているが，その大半がルールベースによる判断であった．しかし，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で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．さらには，時々刻々と変化する実際の交通環境に存在する様々な条件を，すべてルールのみで組み合わせて判定することはその膨大な事象数から考えて不可能に近い．また，そのような交通環境をあたかも実験室か計算機上であるかのごとく随時，人為的に設定して検証することも不可能である．

そこで本稿では，MPEG 動画像を事前に学習して車載カメラ映像から周辺環境を認識する手法と，通信や記録によりその認識性能を逐次改善する方法を提案する．

6.2 オブジェクトの認識

すでに2章で紹介した関連研究[C2][C3][C4][C6]ではSVMやクラスタリング木，HMMなどの統計的手法が用いられている．同様に，画像に限らず一般の推論システムでは適応ネットワークベースのファジイ推論システムANFIS[G1]などの優れた研究がある．しかしながらこれらは静的な教示データには適するが，教示データの追加更新に関してはあまり効率的ではない．ANFISでは多変量線形回帰分析を取り入れ，部分的に逐次学習を可能にしたが，ファジイ推論部では逐次学習が行えない[G1]．一方，筆者のめざす景観認識では，複数の車両にわたって分散的に獲得した個々の学習特性を効率よく合成し，最良の学習特性を獲得することが望まれる．そこで，筆者はそれを可能にするて変量線形回帰分析に着目した．

6.2.1 実験システム

市販ディ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，フロントガラスを通して前方の道路状況を撮影した（図 4.12　車載カメラ）．それぞれで撮影した DV 形式の映像音声を PC 上で MPEG-1 形式に変換し，実験システムの入力とした．

6.2.2 認識率

3 章での定義に基づき，たとえば安全性という視点 *safe* から見た認識率 Q_{safe} はインデックス群 O_{safe}=｛先行車，対向車，信号，標識｝に基づいて計算することができる．同様に利便性という観点 *convenience* から見た認識率 $Q_{conenience}$ はインデックス群 O_{conv}=｛ビル，緑地，空，先行車，対向車，信号｝に基づいて計算できる．今回，道路画像の景観上最も重要と思われる 10 個のオブジェクト｛空，街路樹，ビル，建物，歩行者，先行車，対向車，道路，高架，青の標識｝の個々についてオブジェクト認識率の時系列変化を考察する．

6.2.3 認識実験の概要

2006 年から 2007 年にかけて同じ車両で撮影した走行映像のデータベースから，異なる主観的

表 6.3 交叉実験の結果．

Table 6.3 Experimental results of cross learning.

番号	学習画像	特徴	季節	時間帯	天気	学習量	認識画像および認識率（%）				
							1a	2a	3a	4a	5a
1e	060510-0750_知立高架A	1aと天気、開始フレームの景観が異なる（大樹なし）	春	朝	曇	1039	34.2	-	-	-	-
1f	060510-0750_知立高架B	1aと季節・天気が異なる	春	朝	曇	1592	75.7	-	-	-	-
1g	061124-0809_知立高架	1aと季節が異なる	秋	朝	晴	821	77.9	-	-	-	-
1h	061213-0804_知立高架	1aと季節が異なる	冬	朝	晴	1723	74.2	-	-	-	-
1i	070305-0807_知立高架	1aと季節・天気が異なる	春	朝	雨	1636	72.0	-	-	-	-
1j	070312-1120_知立高架	1aと時間・天気が異なる	春	昼前	曇	1737	76.3	-	-	-	-
1a	070629-0826_知立高架	場所1の代表シーケンス	夏	朝	晴	3446	74.1	74.0	71.3	51.1	90.2
1b	070705-0819_知立高架	1aと近い日時	夏	朝	晴	670	70.8	70.7	67.4	61.8	92.6
1c	070927-0820_知立高架	1aと季節・景観が異なる（道路工事中）	秋	朝	晴	699	55.6	-	-	-	-
1d	071023-0830_知立高架	1aと季節が異なる	秋	朝	晴	1993	71.0	-	-	-	-
2a	060726-1713_EXPO	場所2の代表シーケンス	夏	夕	晴	1429	39.3	74.0	67.0	75.9	67.8
3a	070628-1734_Aoki	場所3の代表シーケンス	夏	夕	晴	1375	42.4	65.8	77.3	48.9	74.9
4a	070628-1744_APITA	場所4の代表シーケンス	夏	夕	曇	789	41.9	70.8	76.8	76.2	60.5
4b	070628-1305_APITA	4aと時間帯が異なる	夏	昼	曇	1134	-	-	-	67.3	-
5a	060514-1748_健康の森P	場所5の代表シーケンス	春	夕	晴	1453	60.7	70.6	62.4	41.8	94.0

シーンクラスに属する合計 5 個の場所（知立高架，AOKI，APITA， EXPO，健康の森 P）を選択し（図 6.1），各場所について異なる日時で撮影した映像を複数個用意した．これらを用いて，学習映像と評価映像の組み合わせに関して下記の条件で実験を行った．

（実験１）：同じ場所かつ同じ時間帯の場合

（実験２）：同じ場所かつ異なる時間帯の場合

（実験３）：異なる場所かつ同じ時間帯の場合

（実験４）：異なる場所かつ異なる時間帯の場合

なお，今回は状況ベクトルを用いず，画像特徴量のみで認識を行った．

6.3 教示データの作成

式(12)で定義した計測行列では，各列の教示データの選択に関して特に制約はない．ところが多変量回帰分析を含む統計的認識手法では一般に，教示量が同じであっても教示データそのものの選択が個々のオブジェクトの認識率を大きく左右することが多い．そこで，教示データの作成においてはできるだけ実験者の作為が介入しない定型的な手順を確立する必要がある．今回，フルフレームの MPEG-1 動画を学習対象とするにあたり，フレーム内の個々のオブジェクトの面積比率に沿った教示を行うため，教示データの基本単位を 1 マクロブロック（以後，イベントと呼ぶ），教示量の基本単位を 1 フレーム（最大 330 イベント）とした．また，MPEG-1 動画の復号順序に合わせて教示量を増大させるほうが定型的手順としては都合がよいため，イントラフレーム（I），予測フレーム（P），補間フレーム（B）の 3 種類のフレームについて，基本的には下記手順で段階的に教示量を増やした．ここで，フレーム番号 n でピクチャタイプが’X’であるフレームを n’X’と表記した．

第 1 段階：1I, 4P

第 2 段階：1I, 4P,2B,3B

第 3 段階：2 番目の I フレームを第 2 段階に追加

第 4 段階以降：前段階の学習に用いた最後尾のフレームより後のフレームから適宜教示データを選択して前段階に追加する

(1) 知立高架　(2) EXPO　(3) AOKI　(4) APITA　(5) 健康の森P

図 6.1 実験に用いた映像

Fig. 6.1. Sequences for experiments.

特に動画の特徴として(3 章 3.24)式の動きベクトルを含めるため，第 1 段階で 4P の教示データを入れた．この結果，第 1 段階で 2 フレーム分の教示データ（最大 660 マクロブロック）を持つことになり，このままでは教示量がゼロに近い場合の認識特性を考察できない．そこで，第 1 段階よりもさらに教示量が小さい場合を作成する際は，第 1 段階の教示データをイベント単位で無作為にサンプリングした．次に，第 2 段階では 1I と 4P の間の補間フレーム 2B と 3B を入れることで IPB の連関の最小構成を学習させた．また，第 3 段階は少し時間をおいた教示データを獲得するため 2 番目の I フレームを学習させた．第 4 段階以降は特に強い規定はないが，ほぼ PBBI のフレーム順序に沿って教示量の増大を行った．

6.4 差分フレームへの対応

本章では 5 章までの I フレームのみの認識[A8][A9]に比べて，新たに差分信号ブロックを含む P

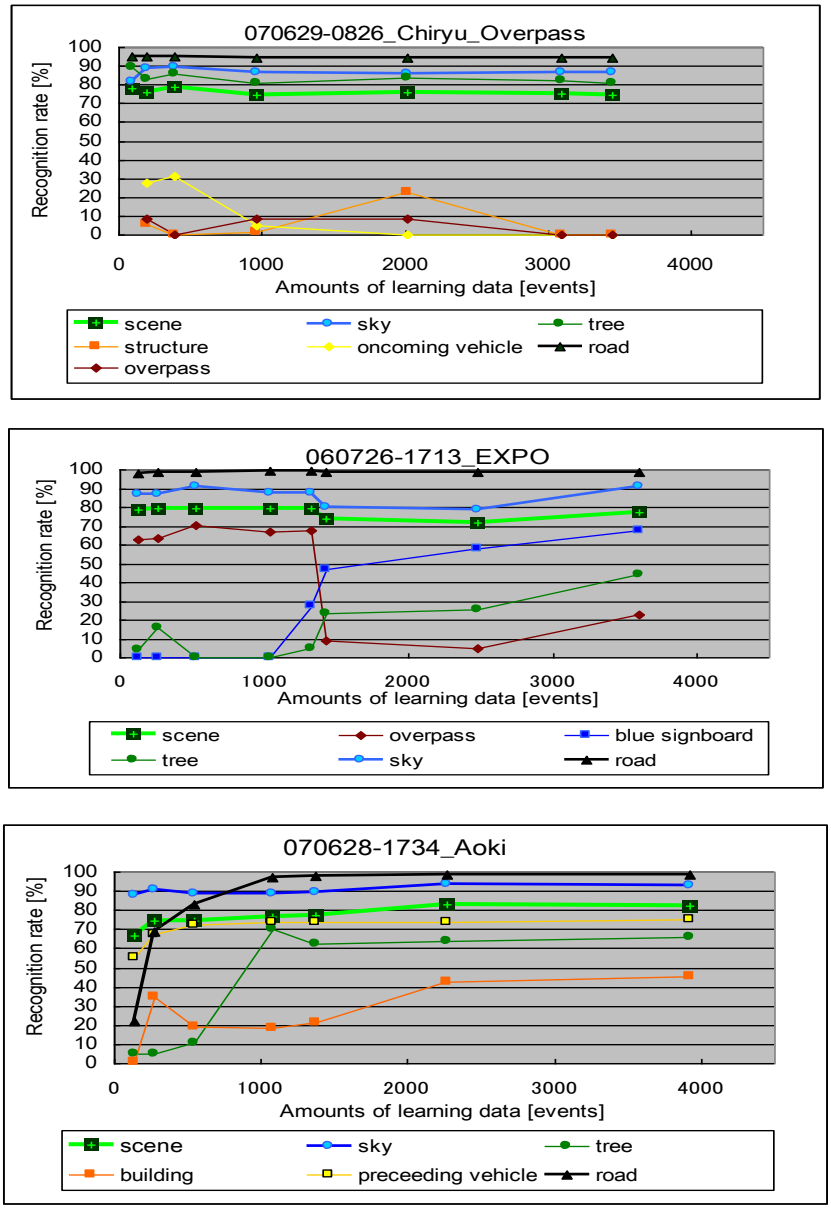


図 6.2 学習量による認識率の変化

Fig.6.2 Recognition rates vs. learning amounts.

フレームと B フレームも認識対象に加えて動画認識を行った．これによる期待効果は大きく分けて次の３つである．

1) MPEG-1 動画では I フレームのみだと 2 frame/s となり，0.5 秒間隔の認識しかできない．それに比べて P,B フレームを加えれば 30frame/s（約 33ms 間隔）の認識が行え，走行時の道路画像など動的変化が激しいシーンに追従できる．

2) P,B フレームを加えることで動きやフレーム間差分を用いた認識が可能になり，I フレームのみでは行えなかった移動車両など動物体の認識が可能になる．

3) I フレームのみの認識では教示データがイントラブロックのみに限られていたが，P,B フレームを認識対象に加えれば教示データを自由に選択できる．

しかし，差分信号ブロックの持つ情報のみではオブジェクトの色に関する特徴を十分に認識できない．そこで，$\boldsymbol{S}_{color}$ を拡張し，マクロブロック単位の符号化属性の識別子（イントラ，予測，補間の３種類）と前フレームのブロック内平均色の計 6 次元を追加した．これによって差分信号ブロックとイントラブロックを同じ次元構成の特徴ベクトルで表現できるため，式(3)を個別に計算する必要がなくなる．

6.5 認識単位の拡張

以上，マクロブロック（16×16 画素）単位の教示と認識を基本として実験内容を説明した．画像データ構造から言えば，MPEG-1 形式ではマクロブロックをさらに 4 分割したブロック（8×8 画素）単位で DCT 演算を行っている．そこで本稿では，教示データは作成コストの観点からマク

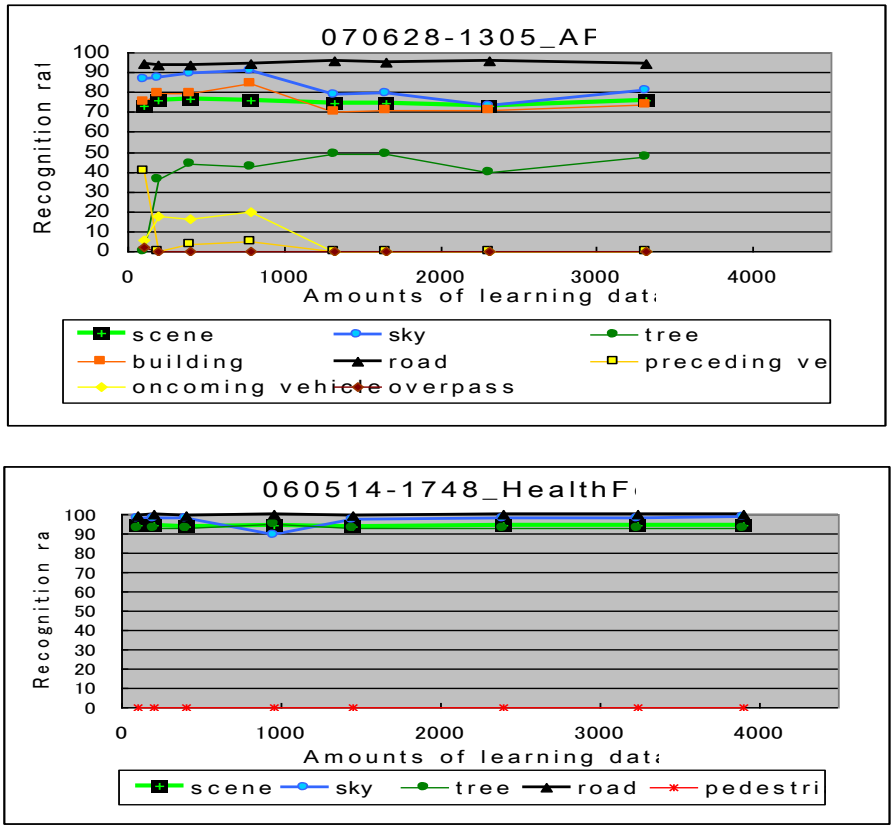


図 6.16　学習量による認識率の変化

Fig.6.3 Recognition rates vs. learning amounts.

ロブロック単位とし，認識時は式(3.21)，式(3.22)でブロック単位の画像特徴ベクトルを計算した．式(3.23)，式(3.24)については教示単位に合わせ，4つのブロックすべてにマクロブロックのデータを転写した．この結果，本稿の実験ではインデックス $X(m_r,m_c,k)$ を各ブロックについて個々に算出できる．認識率については，4ブロックのいずれかが正解であればマクロブロックとして正解とみなし，5章の議論をそのまま適用した．

6.6 考察

6.7 実験1：同じ場所かつ同じ時間帯の場合

6.7.1 自己交叉学習の場合（実験 1A）

まず，表 6.1 の各場所における代表シーケンス Seq.1a, 2a,3a,4a,5a について上記手順で自己交叉の教示量を変化させ，シーン認識率とオブジェクト認識率を算出した（図 6.2, 図 6.3）．簡単のため，シーン認識率の *view* は *all* に設定した．全般的に教示データ数 1000 イベント前後でシーン認識率は飽和傾向を示し，2000 イベント以上では逆に認識率が低下するケースもある．一方，オブジェクト認識率では，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が教示量の増大に対して飽和傾向を示す．他方，画像面中で占有面積の小さいオブジェクトでは，Seq.2a（AOKI）のビルや Seq.4a（EXPO）における青の標識と高架および街路樹などで教示量に対する増加が認められる．しかし，歩行者は全シーケンスにおいてまったく認識できなかった．図 6.4 に Seq.1a の実験画像を示す．インデックス画像の同じ色領域で高輝度のマクロブロックはイントラブロックを示し，それ以外は差分ブロックを示す（以下同様に表示）．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（図 6.5）．これに対応して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．

6.7.2 異なる映像の場合（実験 1B）

Seq.1a に対して時間帯，天候，季節が同じで日時も近い Seq.1b を学習画像とする．実際は人通りや交通量，先行車の車体色などの動的変化に起因して両者の画像内容が変わるため，認識率の低下が予想される．実験結果ではシーン後半で空や街路樹のオブジェクト認識率の低下とともにシーン認識率が低下した．この原因としては，同じ「晴」でも雲の配置が異なること，Seq.1b は Seq.1a よりも対向車の出現頻度が圧倒的に多いことなどが考えられる．

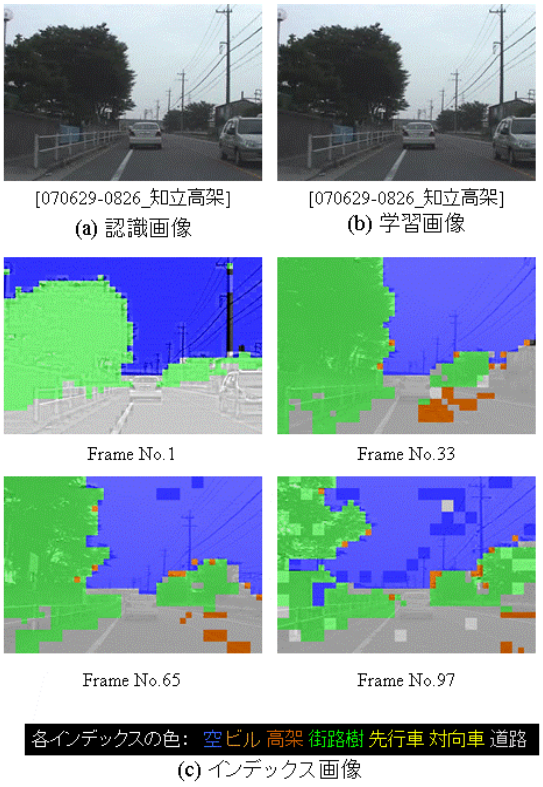


図 6.4 入出力画像（実験 1A）

Fig. 6.4 Input and output images (Experiment 1A).

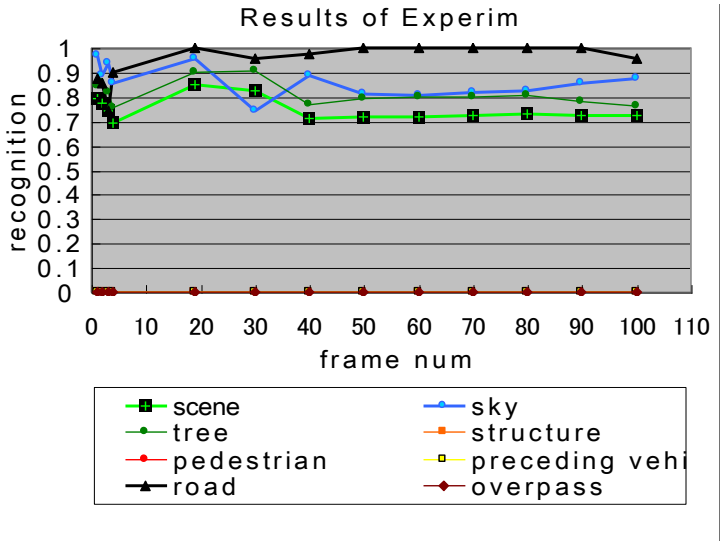


図 6.5 認識率の時間変化（実験 1A）

Fig. 6.5 Temporal changes of recognition rates.(Experiment 1A).

6.7.3 季節が異なる場合（実験 1C）

Seq.1g（11 月）や Seq.1h（12 月）を学習画像として Seq.1a（6 月）を認識する場合，同一時間帯であるにもかかわらず影の長さや日射の強さ，日照時間などがシーンに影響を与えるため，学習効果は低減すると予想される．しかし実験結果では空，街路樹，道路などの大領域のオブジェ

クト認識率は比較的安定しており，シーン認識率は低下していない（表 6.1）.

一方，Seq.1c や Seq.1e を学習画像として Seq.1a を認識する場合ではシーン認識率が大きく低下した．Seq.1c では道路工事による景観変化，Seq.1e では撮像範囲の違いによる景観変化（Seq.1a 画面左側の大樹が写っていない）が各々の劣化の原因と考えられる．

以上のことから，早朝や夕方のように季節によっては夜同様に日照がなくなる時間帯を除けば，季節に起因する変動よりも景観変化による変動のほうが学習に大きな影響を与えると推測される．

6.7.4 天気が異なる場合（実験 1D）

天気が異なる場合は空の色とシーン全体の明るさに影響を与えるため，晴れのシーンで作成した学習データを曇りのシーンに適用した場合やその逆の場合，認識性能が著しく劣化することが予想される．Seq.1i（雨天）を学習画像とした場合，Seq.1a の街路樹や道路の領域でワイパーと誤認識したブロックが多く出現し，認識率は大幅に劣化した（表 6.1）.

6.8 実験２：時間帯が異なる場合

同じ場所でも時間帯が異なると照明や車両などの発生状態が変化するため，同じ天候でもオブジェクトの明るさに差が生じる．Seq.4b（昼）を学習させて Seq.4a（夕方）を認識させると，空またはビルと道路との間の境界領域で誤認識が多く発生した．これは画面垂直方向での判別性能が劣化していることを示す．一方で，道路両側のビルと空の開口の狭さが水平方向の判別に寄与するため，平均シーン認識率はさほど劣化していない．

6.9 実験３：場所が異なる場合

Seq.3a を認識対象画像とし，Seq.4a を学習画像とした場合の実験結果を図 6.7 と図 6.8 に示す．Seq.3a では空と道路のオブジェクト認識率が高い値を保つためシーン認識率も安定し，平均 76.8%という良好な値を示した．

一方，シーン中に占める面積の小さいオブジェクトや出現，接近，消失の頻度が高いオブジェ

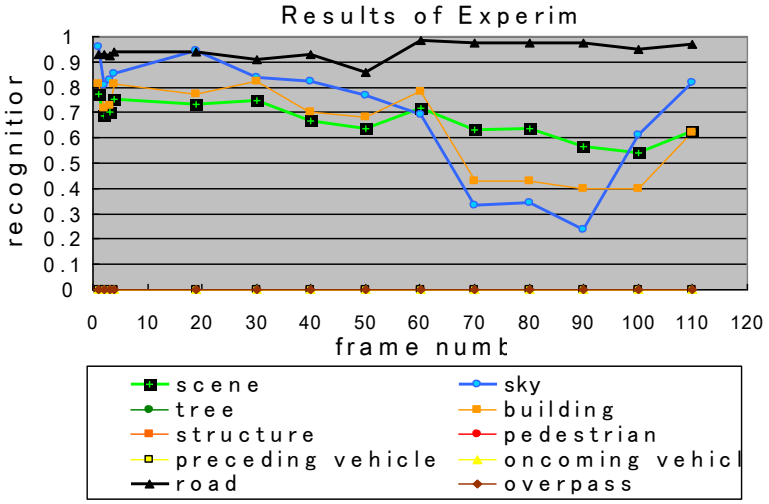


図 6.6 認識率の時間変化（実験 2）

Fig.6.6 Temporal changes of recognition rates.

(Experiment 2).

クト（標識，看板，対向車，歩行者など）が視点の対象になっている場合ではシーン認識率の変動は大きくなる．Seq.1a，Seq.2a，Seq.5a で大領域であった街路樹は Seq.3a では小領域に近くなり，街路樹の平均オブジェクト認識率は 39.0%程度に留まった．

他方，今回の主要オブジェクト(10 種類)に入っていない「赤い看板」や「停止ランプ」などが良好に検出された．これは学習画像のオブジェクトの構成が認識画像と類似するためと考えられる．

6.10 実験４：場所と時間が異なる場合

場所と時間の両方が異なる場合の平均シーン認識率は 40%台の低い値を示し，常識的な予想に一致する．また，交叉学習の認識率において，必ずしも画像の役割（学習と認識）を相互に交換した場合の対称性は成立しない．

6.11 P フレームと B フレームの効果

図 6.5，図 6.6，図 6.8 に示した認識率の時間変化は，開始部分の｛1I,2B,3B,4P,10P｝以降は 10 フレーム間隔で作成した Ground Truth との比較に基づく．そのため，これらの特性は特定のピクチャタイプに偏ったものではない．したがってピクチャタイプに起因する目立った変動は特に見られず，時間的に連続した認識が 30fps で行われていることがわかる．また，図 6.3，図 6.8 から I フレームのみでは原理的に困難であった移動車両の認識も一部可能になっていることがわかる．以上により，今回 P フレームと B フレームを認識対象に加えたことで，I フレームのみではなしえなかった期待効果が確認された．

6.12 MPEG でない場合との比較

冒頭で述べたように，本稿では記録や通信および普及度の観点から MPEG 形式の画像データを認識の出発点とした．その前提条件のもとで，MPEG 動画を学習して他の MPEG 動画を認識するという手法は他に研究例がない．そこで近年，一般物体認識[F1]で用いられている SIFT（Scale-Invariant Feature Transform）特徴量[F3]に着目し，本方式と比較する．まず MPEG-1 復号画像を

表 6.4 他方式との比較．

Table 6.4 Comparison with other methods.

認識方式	認識画像とフレーム平均のシーン認識率(%)*				
	知立高架	EXPO	AOKI	APITA	健康の森
SIFT (LoG)	39.9	52.6	68.3	52.9	57.3
SIFT (DoG)	53.8	55.2	69.6	53.9	54.1
本方式	77.8	72.3	78.4	69.6	92.5

認識方式	認識画像と特定オブジェクト(抜粋)に関するフレーム平均のオブジェクト認識率(%)*									
	知立高架		EXPO		AOKI		APITA		健康の森	
	道路	高架	街路樹	高架	ビル	先行車	ビル	道路	空	街路樹
SIFT (LoG)	0.0	70.8	34.9	61.4	79.7	66.6	90.8	0.0	86.5	92.2
SIFT (DoG)	15.3	74.2	54.8	81.8	83.0	65.6	89.9	0.0	86.4	89.8
本方式	100	0.0	0.0	76.5	28.1	67.1	82.0	98.0	100	98.5

* 認識対象フレーム: ｛1I, 2B, 3B, 4P｝
教示データ: ｛1I, 4P｝(自己交差)

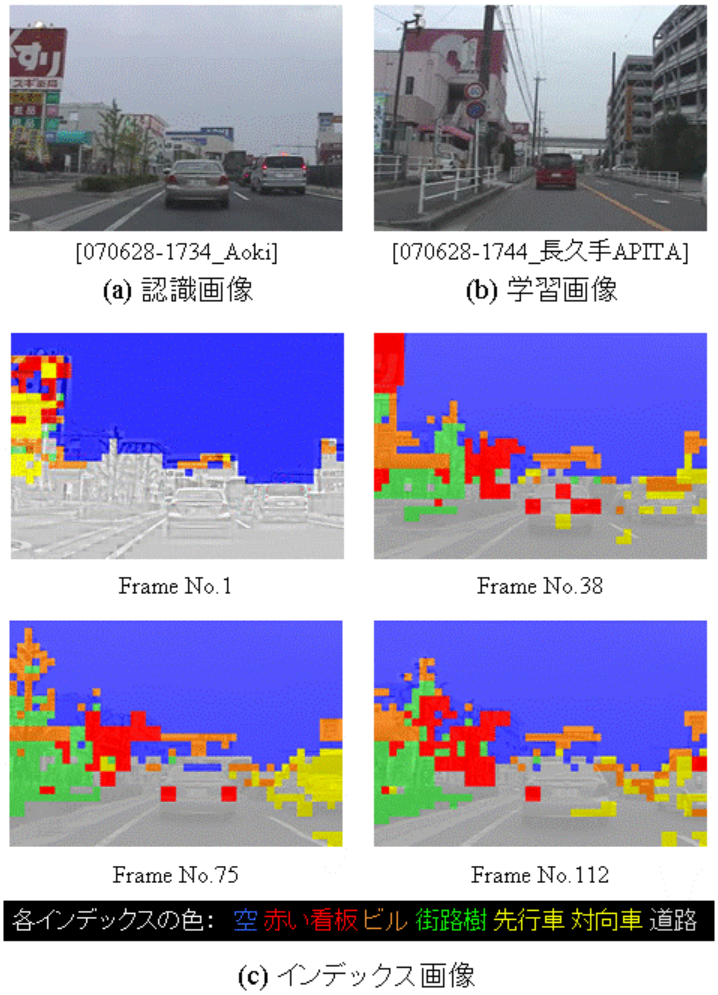


図 6.7　入出力画像（実験 3）

Fig. 6.7　Input and output images (Experiment 3).

720×480 画素の bmp ファイルに変換し，代表的な SIFT 特徴量として LoG（Laplacian of Gaussian）と DoG（Difference of Gaussian）の 2 種類を個別に抽出する．ここで，MPEG-1 の符号化復号化および bmp への変換には SONY 製 DVgate Plus2.2.01.12150 を使用した．また，SIFT 特徴量の抽出には Web で公開されている lip-vireo[F3]を用いた．次に，抽出した特徴点の二次元位置座標に基づき，対応する特徴ベクトル（128 次元）を MPEG-1 形式のマクロブロック単位で配列する．この構成のもとで本方式と同じ教示データ（マクロブロック位置と教示語彙の対）と同じ認識器（多変量回帰分析）を適用し，自己交差で認識性能を比較した（表 6.2）．

この結果，本稿の MPEG ベースの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 ベースの最高値を最大 35.2%上回った．この主たる原因は SIFT ベースの手法における道路の認識率が極端に低いことにある．特に道路を空と誤認識するケースが大半であり，一方で空やビル，街路樹については SIFT ベースでも比較的良好に認識している．これらは 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 [12]反面，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．逆に，テクスチャに特徴がある小領域については MPEG ベースよりも SIFT ベースの方が良好に認識できて

いるが，小領域であるためにシーン認識率への寄与度は低い．

また，SIFT はフレーム内の特徴抽出に適用されるケースが大半であり，上記の SIFT ベースの手法でもフレーム間の特徴抽出はまったく行っていない．これに対し MPEG では，今回の実験の映像ソースである符号化データが生成された時点で，動きやフレーム間の変化を抽出完了している．以上により，動画のシーン全体を捉えるような景観認識においては，演算量と認識性能の両面で MPEG ベースの提案方式が SIFT ベースよりも優位であると判断される．

6.13 むすび

MPEG 映像に適した景観認識手法として，有限個のインデックスに対応した確信度ベクトルをブロック画素単位で定義し，状況要因と画像特徴を観測量とする多変量線形回帰分析で推定する方法を提案した．また，実験により効率的に認識率を確保できる学習条件を模索した．教示データ量を変化させる実験ではほぼ 1000 イベント程度で認識率に飽和傾向が見られた．また，状況要因（場所，時間帯，季節，天候）を変化させる実験では，構成オブジェクトや状況要因に大きな変化がない限り，平均シーン認識率は概ね良好な性能を示した．

結論として，提案手法はフルフレームの走行映像をひとまとまりの景観として認識する際には有効となりうる．この結果，8 章で述べるように景観のシーンクラスを自動判別できる可能性がある[A11]．しかし，遠方の歩行者や車両および特定の構造物といった画像中の小領域で表されるオブジェクトをターゲットとする場合は十分な性能を期待できない．そこで，次章で述べるように，近距離から中距離の歩行者認識や交通状況の認識，あるいは娯楽や利便を想定した場の認識などの具体的応用を想定して景観認識を位置づける．

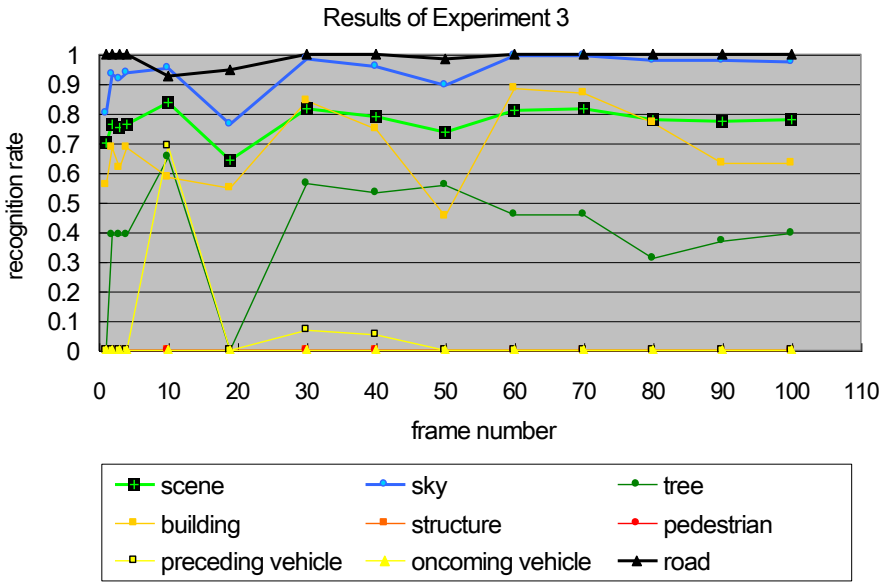


図 6.8 認識率の時間変化（実験 3）

Fig. 6.8 Temporal changes of recognition rates.

(Experiment 3).

7. 移動体検出

本章では MPEG ビデオ処理に基づく移動体検出とトラッキングに関する手法を提案する．まず，フレーム差分の AC 電力を各ブロック画素について DCT 係数から算出し，それを歩行者の確信度として評価する． 次に，近隣の MPEG マクロブロックとの間で動きの類似度を算出し，移動物体の分類規範として評価する．最後に，それぞれの確信度を統合して特定のルールベースによって判定し，認識結果を出力する．この第 1 段階の方法は 1 秒あたりで最大 90%を超える認識率を達成し，移動カメラを対象とする第 2 段階のアルゴリズムとして，PCA（主成分分析）を用いた多変量時系列解析を試行した．また，適応化と挙動モデルに関する将来の研究についても考察する．

キーワード　歩行者トラッキング，Motion, PCA，確信度，挙動，時系列解析

7.1 はじめに

本章では，異なる複数の認識エージェントによる認識プロセスおよびその認識結果として得られる情報の共有，知識交換を想定し，その一例としてプローブカーによる歩行者と車両の認識を考える．車載装置による歩行者の検出や挙動理解は近年ようやく実用化がなされつつある．しかし，多くの自動車メーカーが期待する完全さと実際の技術レベルの間には相当ギャップがあり，現在実用化されているものはまだ限られた仕様条件の範囲内で（例えば，全身が適度な大きさで隠れなく撮影され，しかも車道上に出てきている場合など）成立するものにすぎない．

現在，もっとも一般的な歩行者認識はフレーム独立に形状情報と画像特徴に関するパターン認識を行うものである．だが，静止画認識では動きや時間的連続性に関する人間の認識能力を達成できない．そこで本章では動画像認識としての MBV に焦点を当て，また同様に無線通信技術との組み合わせについても注力する．本章ではその技術を，車載無線通信装置を介して伝送された圧縮ビデオデータから歩行者と車両を検出するシステ

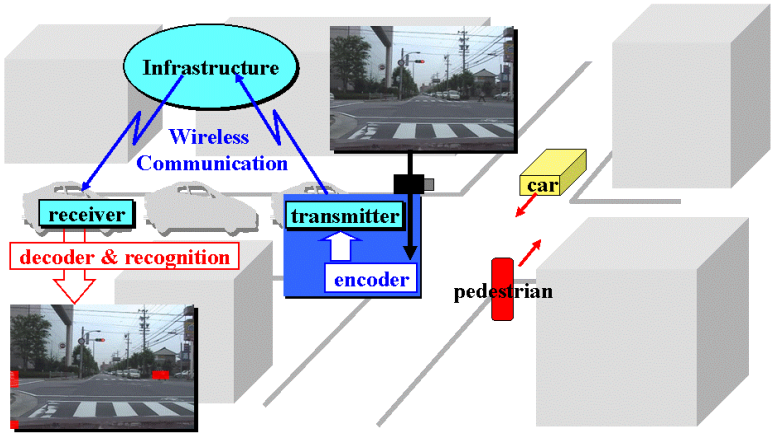


Fig. 7.25. Our Targeted System..

ムに応用することを考える．

7.2 目指すシステム

システムの観点から言うと，最適な手法はターゲットアプリに応じて決まる．一般的には安全に関しては３つのレベル，即ち，衝突安全，予防安全，そして危険予知がある．キャプチャしたシーンは衝突時間 TTC (Time To Collision)に基づいてこれらのカテゴリのいずれかに分類される．現在，多くの自動車関連企業が衝突安全を志向したシステムの開発に力を注いでいる．しかしながら，TTC が小さくなればなるほど，衝突は避けられない．そこで，我々は TTC が 5 秒より大きくなると思われる衝突予測フェーズに焦点をあてる．他方，そのようなフェーズでは画像サイズや様々な隠れの問題があり，車載カメラでは歩行者を十分にキャプチャできない．したがって無線通信インフラが役に立ち，図 7.1 に示すようなターゲットシステムを考えた．これは死角にいる歩行者の接近を事前に検出しドライバに警告することを目的としている．この場合，システム提供者はセンサと通信インフラから認識性能に関する製造物責任（PL）の問題を分離できるという利点がある．

5 章ではスナップショットベースの歩行者検出に景観認識を組み合わせることで 80%を超える誤検出が削減可能であることを示した[A9]．だが，静止画ベースの手法は原理的に人間の動的な行動を把握できないし，動きの急激な変化にも追従できない．そこで本章では MPEG 動画像に含まれる動きに関する特徴量を導入する．

7.3 第 1 段階のアルゴリズム

MPEG ベースのビデオ符号化においては一般に，フレーム差分が大きく分けて２つの主要なビデオ符号化のカテゴリに変換される．１つめは DCT (Discrete Cosine Transform) 係数であり，ビデオ信号の統計的性質を仮定するとき，この DCT は KLT (Karhunen-Loeve

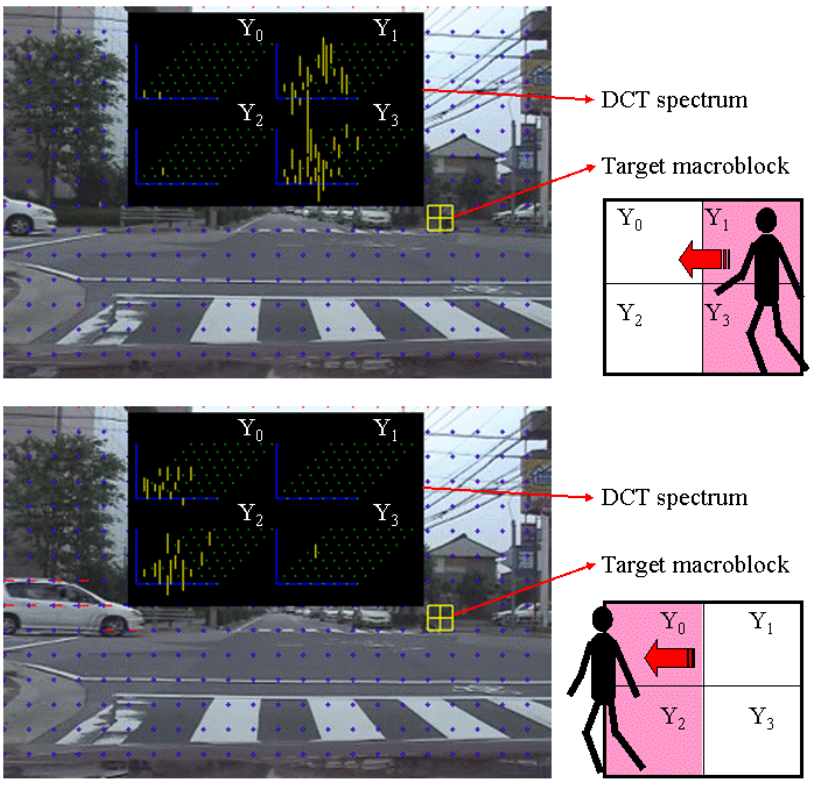


Fig. 7.2. Pedestrian detection by using DCT spectrum.

Transform) の最良近似となることが知られている．そして，第 2 のカテゴリは動きベクトルであり，ブロック画素上の 2 次元のフレーム間の位置差分を表現する．この２つの特徴を判定することにより移動物体とシーン変化を識別することができる．．

7.3.1 動きに関する予備実験

当初，洗練された動き検出アルゴリズムであればどのようなオブジェクトの動きも検出できると仮定した．そこで，KLT (Kanade-Lucas-Tomasi)法をベースバンドのビデオ信号に適用して予備実験を行った．その結果によれば，MPEG-1 の標準フレームサイズに相当する 320×240 画素の動画像について，カメラから約 20m 程度離れたところを歩行する人の動きを KLT 法で検出することは可能である．だが，さらに多くの誤検出が背景部分で発見される．結論として，遠く離れた人間の移動を検出することは KLT 法だけでは困難である．これは背景部の誤検出を除去できる有効な手法が確立されていないためである．また，人間の動作の検出については，20m 以上離れた位置の歩行者を撮影するとき，MPEG-1 エンコーダで人間の動作をブロック画素の動きとして明確にとらえることは難しいことがわかった．

7.3.2 DCT 係数

そこで，DCT 係数の時間変化特性に着目する．これは，MPEG-1 形式においては，もしブロック画素の動きベクトルが検出できないならば，DCT 係数以外にオブジェクトの変化を表現できる信号はないからである．

まず，観測窓をサイズと位置が 1 個の MPEG-1 マクロブロックに整合するように設定しする．MPEG-1 マクロブロックには 8×8 画素の Y 成分ブロックが 4 個含まれる．図 7.2 の上の絵は一人の人間がこの観測窓に右側から入ろうとしているシーンを示している．上部真中のエリアにある 4 つの図形は，観測窓に含まれる 4 つのブロック画素の各々に対する２次元 DCT 係数を示す．簡単のため，その観測窓を BEF (Block Edge Filter)と呼ぶことにする．容易にわかるよに，BEF の右半分は人間の動きを表す著しいレベルの DCT 係数が発生している．この特徴は背景テクスチャに依存しない．それは MPEG ベースの DCT 係数がブロック画素のフレーム差分に関する DCT スペクトルを示しているからである．図 7.2 の下の絵は人間が BEF から出ようとする様子を示しており，対応する DCT スペクトルは BEF の左半分において顕著である．図 7.2 の性質から，各 BEF における DCT 係数の電力評価が移動する人間を発見するための実際的方法となる．

各ブロック画素の内部において，AC 電力に関する基本的な評価関数は次式で定義される：

$$L_{AC}(l_{blk}) = \sum_{m=0}^{7}\sum_{n=0}^{7}\left|DCT(l_{blk}, m, n)\right| - \left|DCT(l_{blk}, 0, 0)\right| \tag{7.18}$$

ただし，$DCT(l_{blk}, m, n)$ は２次元 DCT 係数を意味し，l_{blk} は MPEG マクロブロック（以後，MBK）中のブロック番号を示す．m と n はそれぞれ水平座標と垂直座標を示す．

さて，これで一群の閾値を DCT 係数の絶対値和を判定するために設定でき，それは ISO/MPEG 標準において定義されている picture coding type によって決まる．MPEG には基本的に少なくとも３つの picture coding type があり，I (intra), P (predictive), B (interpolational)がそれにあたる． P または B フレームにおいては，フレーム間差分また

は動き補償差分を 16×16 画素について行って符号化した MBK が存在する．その MBK の内部で BEF は移動物体を検出できる．それは移動物体の領域では 256 次元の DCT 係数による時空間波が際立った成分として伝播するからである．

7.3.3 動きの評価

図 7.3 から容易にわかるように，距離が 20m 未満で画像面を水平に横切るような場合，移動車両は MPEG-1 形式の動画像において MBK ベースの動きベクトルを生成する．この図では青い矢印が過去のフレームと現在フレームの比較により抽出した前向き動きベクトル $\boldsymbol{v}_f$ を表す．同様に，赤い矢印は未来フレームと現在フレームの比較により抽出した後ろ向き動きベクトル $\boldsymbol{v}_b$ を表す．簡単のため，これら 2 つの動きベクトルを合成しトータルの動きベクトル $\boldsymbol{v}$ を対応する MBK について考えると以下のように定義される：

$$\boldsymbol{v}(l_{mbk}) = \boldsymbol{v}_b(l_{mbk}) - \boldsymbol{v}_f(l_{mbk}) \tag{7.19}$$

ただし，l_{mbk} は 1 枚の画像中での MBK の番号を示す．この定義により，車両領域に含まれる各 MBK は類似する大きさと類似する方向をもった動きベクトルを生成する傾向にあることが容易にわかる．このような傾向は，歩行者がカメラに対して非常に近い場合でない限り，歩行者エリアでは起こりえない．このような経験的結果は判断の規則として定式化できる．

1) 動きの大きさ: 一般に，普通の歩行者は車両よりも非常にゆっくりと歩く．そして，フレーム間で変化する画素の数はかなり小さい．遠方にある歩行者ではまったく動きベクトルを発生しないかあるいはたかだか 2，3 画素の大きさの動きベクトルとなる傾向がある．したがって，次の不等式により車両と人間を判別することができる．

$$|\boldsymbol{v}| \geq M_{car} \tag{7.20}$$

ただし，M_{car} は判定閾値であり，|.| はユークリッド距離あるいは絶対値距離である．本実験では約 20m 離れた位置を 30km/h 未満の速度で水平に移動する車両の場合について $M_{car} = 4$ とした．

動きの大きさに関する類似度は次式で算出される：

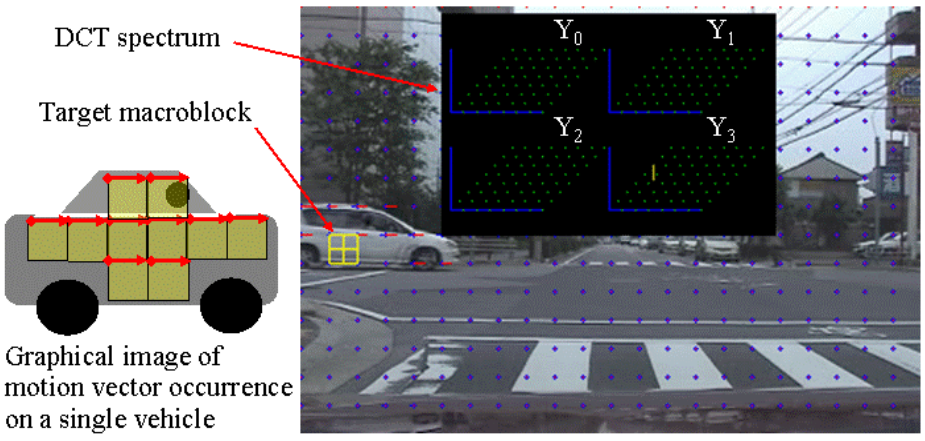


Fig.7.3. Pedestrian detection by using DCT spectrum.

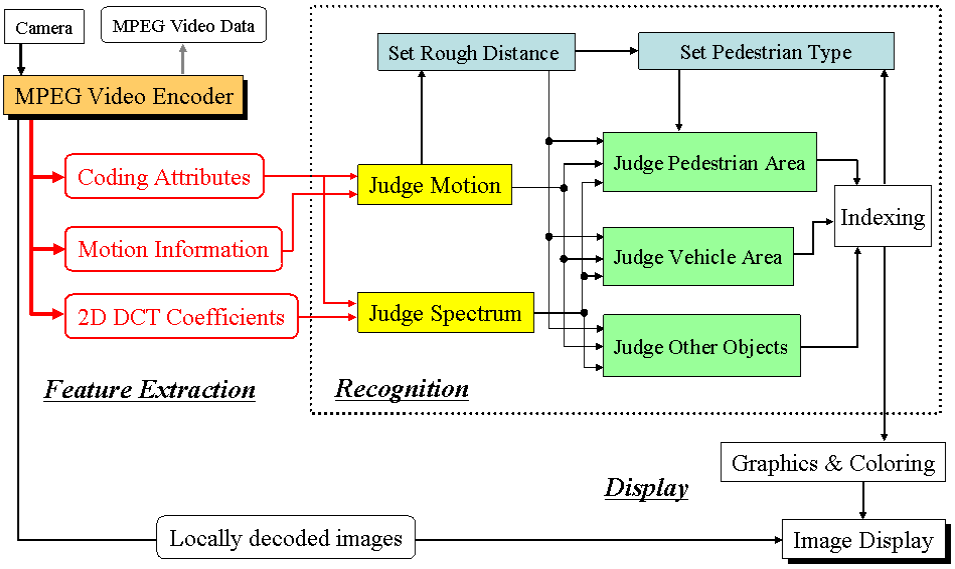


Fig. 7.4. Architecture of MPEG-based video recognition.

$$SAMV = 1 - \frac{\left\| |\boldsymbol{v}_1| - |\boldsymbol{v}_2| \right\|}{\left(|\boldsymbol{v}_1| + |\boldsymbol{v}_2| \right)} \tag{7.21}$$

これは 0 と 1 の間に正規化された値をとる.

2)動きの方向: 動きの方向は, もし車両の概略形状と道路平面が事前に既知であれば, 制約をかけることができる. 動きの方向の類似性は以下で定義される :

$$SDMV = \frac{(1+\cos\theta)}{2} = \frac{1}{2}\left(1 + \frac{\boldsymbol{v}_1 \boldsymbol{v}_2^{\,T}}{|\boldsymbol{v}_1||\boldsymbol{v}_2|}\right) \tag{7.22}$$

ただし, $\boldsymbol{v}_i$ (i=1,2)は 2 次元動きベクトルであり, θ は $\boldsymbol{v}_1$ と $\boldsymbol{v}_2$ の間の角度を示す. *SDMV* は正解化されて 0 と 1 の間の値をとる.

3) 動きに関するトータルの類似性:i 番目の MBK ライン上で隣接する 2 つの MBK について, トータルの動きに関する類似性 SMV は次式で定義される :

$$SMV(i,j) = \frac{\boldsymbol{v}(i,j)\boldsymbol{v}(i,j+1)^T}{|\boldsymbol{v}(i,j)||\boldsymbol{v}(i,j+1)|} \tag{7.23}$$

ただし, $\boldsymbol{v}(i,j)$ は画像中で i 番目の MBK ラインで j 番目の MBK の動きを表現する 2 次元行ベクトルを表す.

7.3.4 ルールベースの決定

移動物体の最終判定では, 上述の簡単な数値判定が十分に動作しないケースも多い. 様々な経験的結果から考えると, 被写体としての道路シーンとセンサとしての車載カメラの双方について, 応用特定のルールが多く存在する. 中でも以下に示すルール群は特に効果的で比較的実践が容易なものである :

R1) 車は車体の表面上にほとんどテクスチャがない場合が多い.

R2) 車体とタイヤの境界部分はしばしば AC 成分を生成し, それは BEF における DCT 係数の急激な変化を引き起こす. それは人間の誤検出につながる.

R3) 人間の映像には多くの影部分を持つ傾向がある.

R4) 人間の映像には複雑なテクスチャを有することが多い.

R5) 車両の動きは連続的であり（運動方程式に従う），上下の運動が比較的少ない．他方，人間はいくつかの特有かつ周期的な動きを有する.

R6) 一般的に，車は人間よりもはるかに速く移動する.

R7) 車両の一部が投影された BEF ではしばしば完全な隠れを形成し，背景を隠す.

R8) 人が投影された BEF では，人がよほどの近距離（数 m 以内）にいない限り，背景テクスチャが見え隠れする.これは人間が完全にマクロブロックを覆うような形状ではなく、しかも手足が動くことにより背景がマクロブロックに映りこむからである。

R9)- 人が投影される BEF の範囲と各 BEF における DCT スペクトルの発生パターンは人の着衣に依存する．これは人の着衣が人のシルエットと各 BEF 内での DCT スペクトル（色、模様、時間変化）に影響を与えるためである.

R10)- 人間はしばしば非 Gaussian の動き（突然の停止，方向転換など）を生成する.

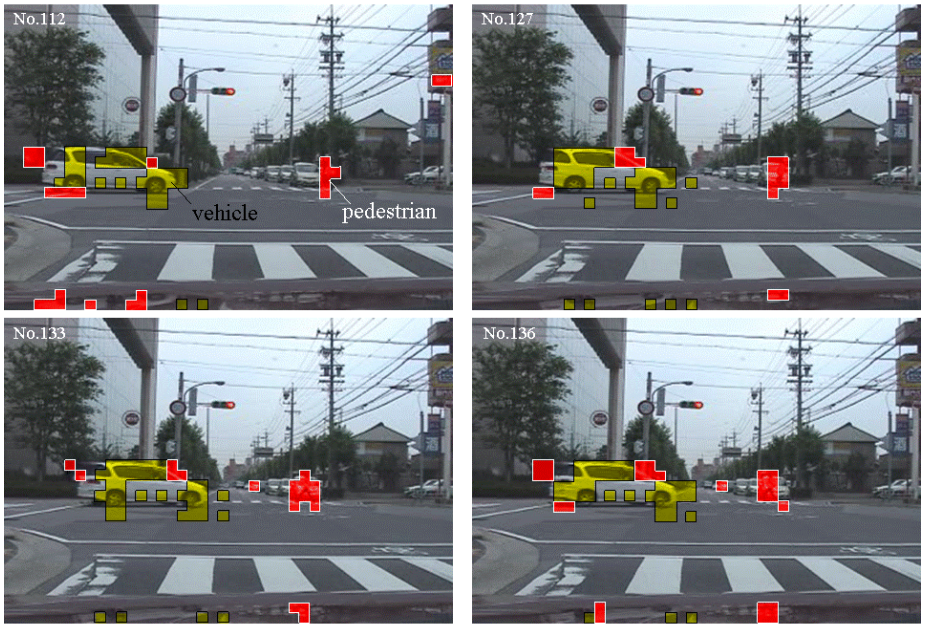


Fig. 7.5. Experimental results.

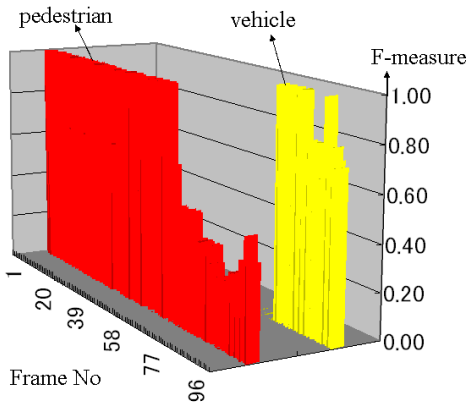


Fig. 7.6. Temporal characteristics of F-measures.

7.3.5 実時間システムに向けて

図 7.4 は実時間処理が可能な MPEG ベース認識システムのアーキテクチャである．ここでは，たくさんの種類の電子機器に広く用いられている MPEG ビデオエンコーダ LSI は画像特徴抽出を行う最も信頼性の高い実時間処理プロセッサと考えることができる．もし，データインタフェースを少し変更できるならば，認識プロセスにおける MPEG ストリームのパーシング操作が必要なくなるであろう．それは，MPEG エンコーダプロセッサから通信路符号化を行う前の MPEG 特徴量を直接抽出できるからである．以上により既存の MPEG エンコーダ LSI の設計資産が効率的に利用でき，実現コストを劇的に下げることができる．

7.4 実験

7.4.1 実験結果と考察

静止画像中の物体検出では一般に SVM やニューラルネットワークなどのパターン認識手法が用いられている．しかし，静止画のパターン認識はそのままではフレーム間処理を行わないため，トラッキングに関しては良好な性能を示さない．また，目標対象物に関してフレームごとに特定の代表的形状を学習することは，歩行者の動的な挙動を認識する際に良好ではない．それらを鑑みて，前節で提案した第 1 段階の MPEG ベースの手法を用いて移動体認識の実験を行った．

図 7.5 に実験結果を示す．歩行者の一部として判定された各ブロック画素は赤で示される．同様に，移動車両の一部として判定されたブロック画素は黄色で示される．図 7.6 は図 7.5 の実験結果に対応する F 尺度の時間的特性を示す．最初の 2 秒間（60 フレーム）の間，歩行者トラッキング率（赤）は 90%/秒を超え，最大平均は 94%/秒である．その 2 秒間の後の後半部分では，その性能は劣化している．これは主として，左側から歩行者

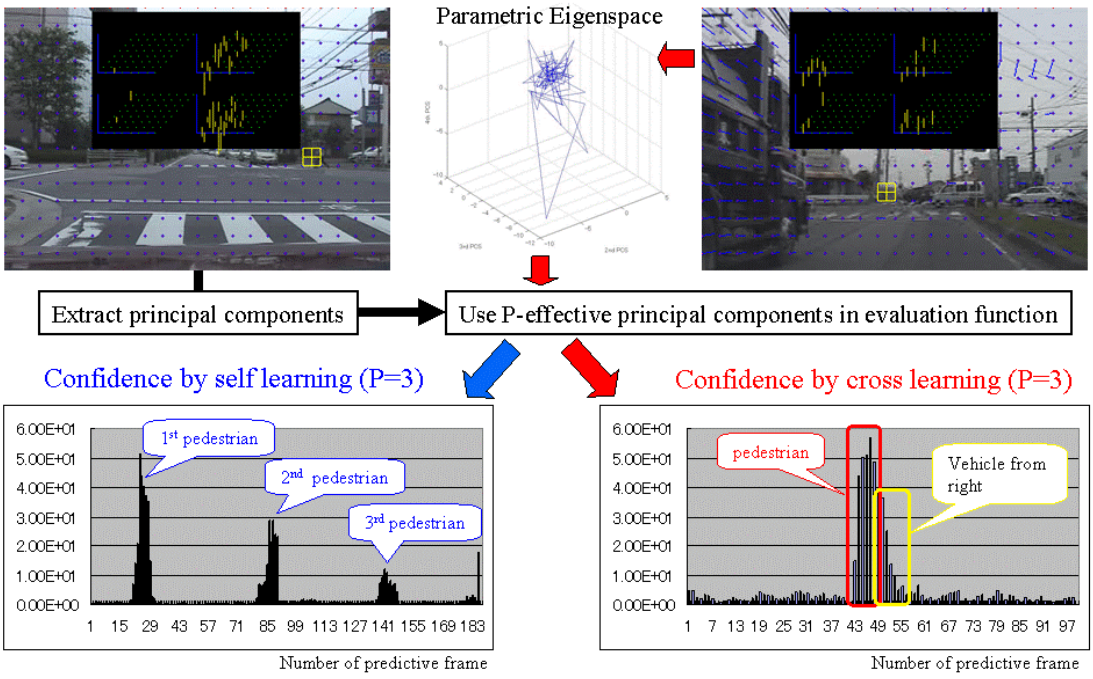
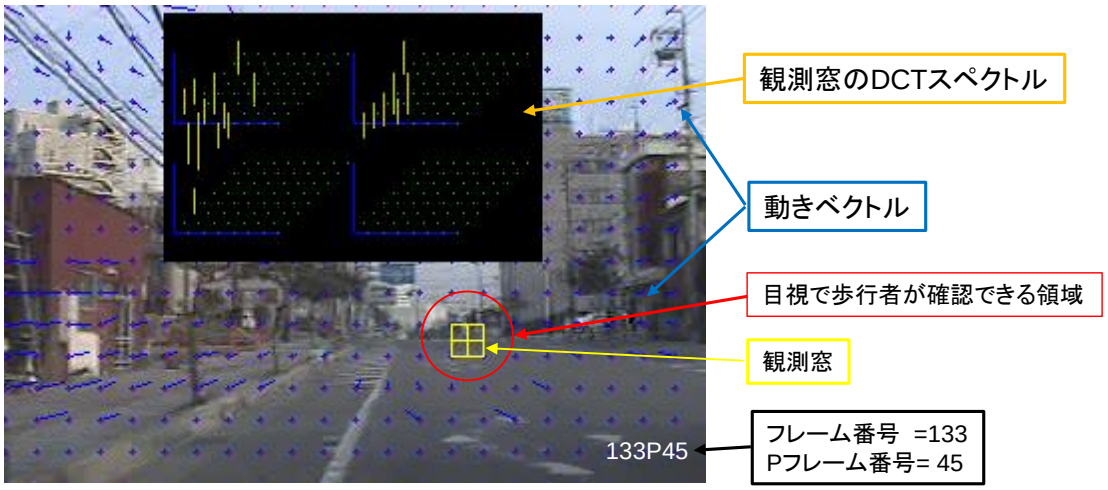


Fig. 7.7. Approach from multivariate time series analysis using LPCA.

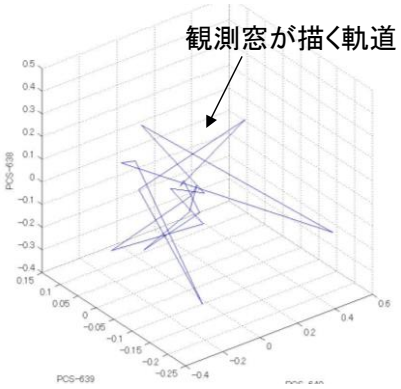
に接近している白い車が検出されているためであり，それは黄色のブロックで示される．そして，いくつかの誤認識した赤のブロック（歩行者）は車両のエッジ上に現れている．同時に，車両トラッキング率（黄色）は最初の 2 秒間の後は急激に立ち上がり，平均で 90%/秒以上の値を示す．

誤認識したブロックに関連して，少なくとも 5 種類のタイプがある．車両のエッジ上に見つかる第 1 のタイプは形状情報を用いたより洗練された決定論理を導入することで低減できると考えられる．第 2 のタイプの誤検出は空やビルなどの領域上に見つかるもので，5 章ではそれらを景観認識によって低減できる可能性があることを示した．第 3 のタイプは 2，3 フレームの間に明滅するもので，それらはよほどの高速移動物体でないかぎり，通常は人間の知覚上意味を持たない雑音として処理される．第 4 のタイプは第 3 のタイプよりも長い時間間隔で見え隠れするものであり，{視点，対象物，遮蔽物} の 3 者の間の位置関係が，それら 3 つのうちの少なくとも一つが動的に変化することによって引き起こされる．一般にこの問題は時系列処理と知識処理によって解決が試みられることが多い．その他のタイプは非常に遠く離れた移動車両を観測したときなどに含まれるもので，道路構造や地理的要因、景観を考慮した多元的な認識処理が必要である．

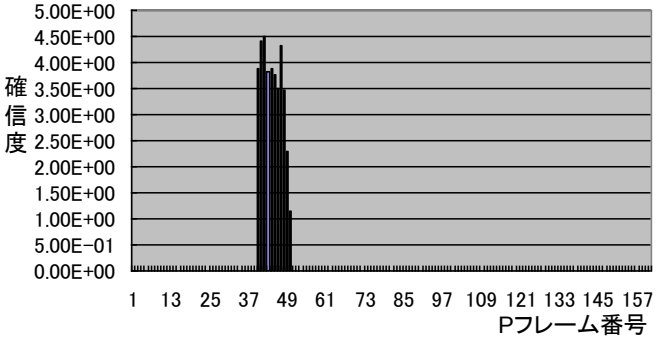
7.4.2 評価



(a) 時速40km/h走行時の画像と観測窓



(b) パラメトリック固有空間



(c) 観測窓における歩行者確信度の時間変化

Fig. 7.8. Experimental results of multivari ate time series analysis using KPCA

従来の歩行者検出手法（ニューラルネットによる静止画パターン認識）を同じシーケンスに適用したところ，F 尺度で 29%/秒未満のトラッキング率であった．同様にビデオ出力の主観評価から，この従来手法では移動物体を十分にトラッキングできないことが分かった．この結果，提案する MPEG ベースのビデオ認識は交差点のケースにおいて，従来のフレーム独立処理によるパターン認識を遥かに凌ぐという結論を得た．

7.4.3 第 2 段階のアルゴリズム

第 2 段階では移動カメラによる認識を行う．それを実現するには AC 電力に基づく基本式(7.1)を改良し，人間の映った画像の特性を扱うことができるようにせねばならない．まず第 1 の方策は統計的な時系列解析を導入することである．第 2 の方策はブロック画素 1 個あたり最大 64 個存在する DCT 係数から有効な成分を抽出することである．そこで，この統合のための予備実験として，3 章の理論に基づき，線形主成分分析（LPCA）をはじめとする以下の多変量時系列解析を試行した．

（１）線形主成分分析（LPCA）

図 7.7 は上記の計算に基づく認識プロセスの実際のフローを示す．ここではパラメトリック固有空間を用いて最大から上位 3 個の主成分の有効性を歩行シーン期間について視覚化した．それにしたがって，固有空間における原点と軌道上の各点との絶対距離を新しい評価関数として定義し，確信度を算出した．図 7.7 では，２つの異なるシーンについてかなり良好な認識性能であることがわかる．

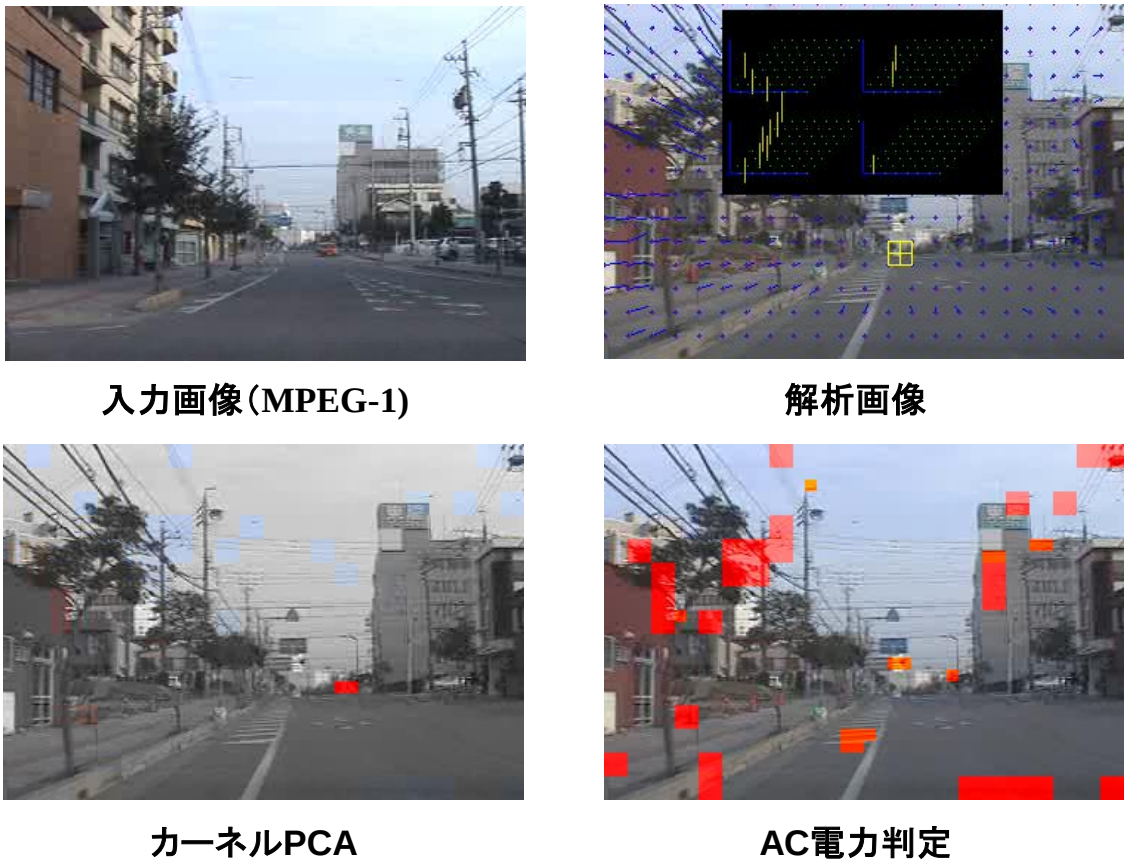


Fig. 7.9. Approach from multivariate time series analysis.

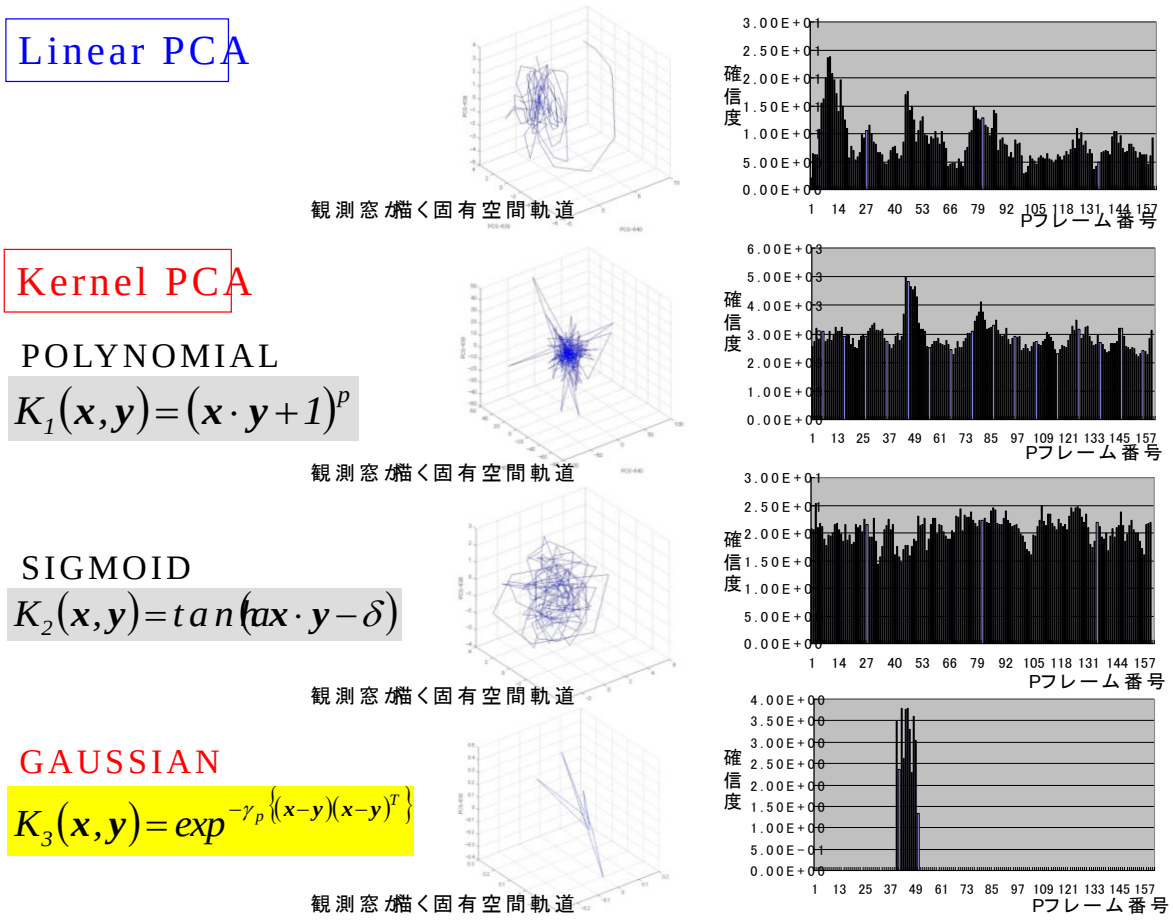


Fig. 7.9. Comparison of performances in various kernels..

（２）カーネル主成分分析（KPCA）

図 7.7 の歩行者までの距離は約 20m 程度であり，画像面上の人の大きさがマクロブロック 1 個分にほぼ相当し，人の形状も十分確認できる。これに対し，図 7.8 では約 50m 程度遠方の不鮮明な歩行者を認識対象として走行速度 40 ㎞/h で接近するシーンを用いて実験を行った．この場合，移動に伴う背景の変化で AC 電力の発生が全画面にわたり，歩行者が観測窓を通過する時間範囲において，LPCA で算出した確信度では明瞭な極大値の発生を観測することができなかった．そこで，3 章のカーネル PCA（KPCA）を導入し，検出を試みたところ，図 7.8 の結果を得た．設定した観測窓において歩行者が通過する時間範囲で確信度のピークが出ており，良好に検出できていることがわかる．なお，カーネルの選定に際しては図 7.9 の予備実験を踏まえ，Gaussian カーネルを採用した．

（３）時空間画像

さらに，BEF 単位で時空間画像を生成し，これによって人の歩行動作を検知できるかどうかを確認する実験を行った（図 7.10）．この手法の魅力的な特徴は比較的シンプルな画像処理技術によって歩容（gait）情報を容易に抽出できる点にある．画像上で観測される交差する複数の波形は歩行者の歩行動作を表している．実験の結果，対応する波の形状は差分信号ベースの時空間画像においても観測され，カメラが移動する場合も観測可能

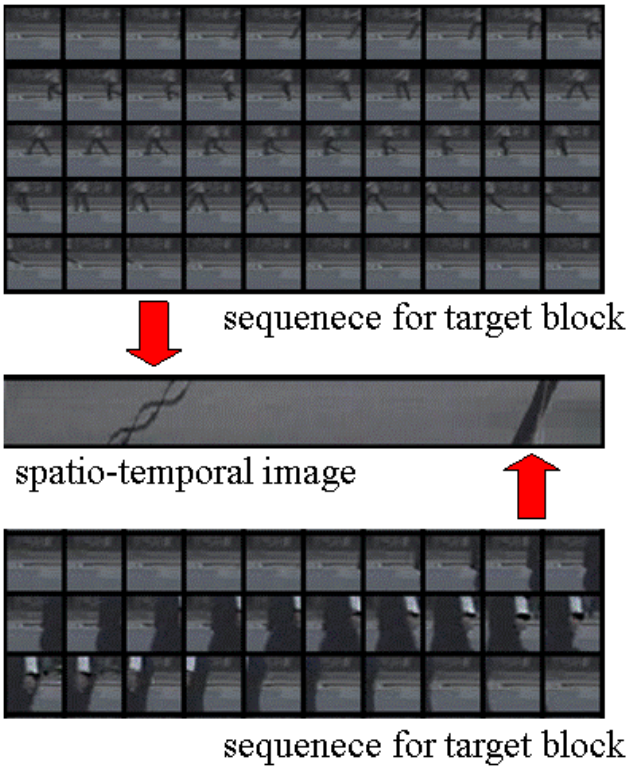


図 7.10 時空間画像.

Fig. 7.10. Spatio-temporal image.

であることがわかった.

7.5 様々なシーンへの適応化に向けて

上記の結果によれば，KPCA が最良の手法に見えるが，KPCA に限らず非線形手法では設定パラメータによって性能が左右されることが多いため，どのような画像についても常に同じ設定条件で LPCA や AC 電力判定を凌ぐというわけではない．8 章で示すようなシーンクラスに応じて適宜使い分けていくことが実際の戦略となる．そこで，Fig. 7.9 に示すような適応化機構を考える．これはカメラが移動速度や認識対象物との距離が動的に変化しているときでもシーンと対象物をロバストに認識することをめざしている．図中の判定パラメータを動的に適応化するには 2 つの戦略がある．

第 1 の戦略は 6 章の景観認識を併用することであり，環境や運転状態の推定を可能にする．この具体的な解決手段は 8 章「シーンクラスの生成」に継承される．

第 2 の戦略は歩行者モデリングの導入であり，ここでは図 7.10 で示すような状態遷移ダイアグラムで基本挙動が記述される．これについて以下に考察する．

7.5.1 歩行者の挙動モデル

挙動理解はドライバが遭遇しうる危険を予測する際に重要な役割を果たす．運転の安全性という観点から歩行者の挙動は正常，危険，異常という 3 つのクラスに分類される．図 7.10 に示すように，これらのクラスは互いに密接に連関している．これらの関係を用いることで，図 7.9 中の適応化メカニズムは良好に動作し，高い認識率を保持できるであ

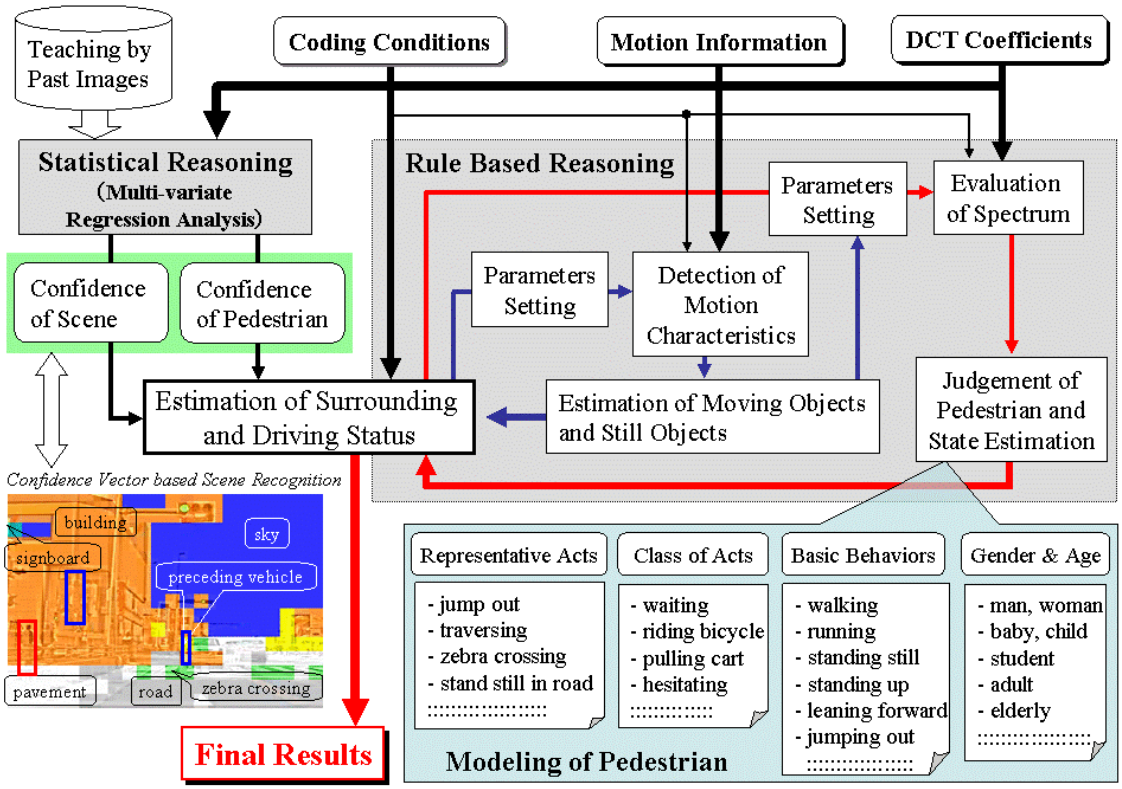


図 7.9 適応メカニズム

Fig. 7.9. Adaptation mechanism.

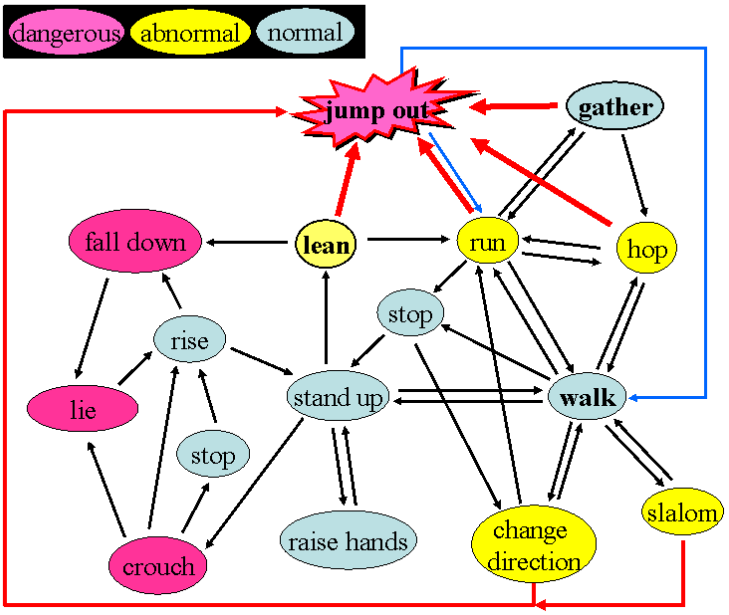


図 7.10 歩行者挙動のモデリング

Fig.7.10. Modeling of pedestrian's behaviors.

ろう．また，仮に信号源が曖昧なセンシング情報であったとしても，システムは歩行者プロファイルをより正確に識別できるようになるであろう．このプロファイルには行動，姿勢の状態，挙動クラス，年齢と性別，着衣，ヘアスタイル，荷物などが含まれる．ま

た，ここで獲得されるプロファイルには時間的連続性や周期性，あるいはルールベースによる記述が容易な属性が多く存在する．この性質は人出現の予測（3 章，10 章）や 3S モデル（3 章）とも密接に関係する．さらには観測するカメラ側の統計的挙動も考えるならば(3.53)式中に示したユーザプロファイル（ここではカメラの時空間移動に関するプロファイル）がシステムの最終的な予測動作において大きくかかわってくる．これは歩行者の挙動モデルを，観測するカメラの統計的挙動と最新の観測値によってデータ同化させることに相当する．

7.5.2 異常行動

まず，歩行者挙動の３つのクラスのうち，異常のクラスについて考察する．筆者の運転経験などから得られる事例を列挙すると，たとえば以下のケースが異常と考えられる．

1)飛び出し: 一般的に，歩行者は歩道に位置し，垂直に立ち，傾くことはほとんどない．そして，歩行者は隠れがある場合以外では，ドアや階段を通じて自ら出現し，移動し，消失する．もし，最初の出現が観測可能であるならば，以下のように，５つのケースの“飛び出し”があるであろう：

a) 隠れのない開放的な歩道を歩くまたは走ったのちに直線的に飛び出してくる

b)樹木や車両またはほかの歩行者のためにいくぶん隠れのある歩道を通じて歩いたり走ったりしたのちにまっすぐに飛び出してくる

c) 隠れがなく見通しの良い歩道上でしばし立ち止まり，その後，まっすぐに飛び出してくる

d) 樹木や車両またはほかの歩行者のためにいくぶん隠れのある歩道でしばし立ち止まり，その後，まっすぐに飛び出してくる

e) 車道の真ん中でしばし立ち止まった後，まっすぐに飛び出してくる

これらのケースで，a)と e)は上述の手法で検出可能である．一方，b)と c)の検出可能性は隠れの程度に大きく左右されるが，断片的な観測結果をもとに隠れ状態のトラッキングを行うことで検出率を上げることが試みられている[E14]．ケース d)の場合は，図 7.10 における挙動遷移モデルがかなり効果的であり，人間がとりうる生態的挙動の観点から“前傾姿勢”が事前に観測されるであろう．従ってもしこのような姿勢変化が DCT 係数における垂直方向の低周波成分として検出できるならば，それに続く挙動“飛び出し”は予想可能であり検出可能となる．

2)立ちすくむ：:歩道においては“たちすくむ”は危険ではないが，道路の真ん中ではそれはしばしば危険な行動に転じる．DCT 係数の観点からいうと，“たちすくむ”によって(1)式の AC 電力が劇的に減少することになる．したがって，下記で定義される AC 電力のフレーム間差分を評価することが考えられる：

$$D(n)=L_{AC}(n)-L_{AC}(n-1) \tag{7.7}$$

ただし，n はあるフレームについての離散時間番号を表す．
これは次に示す２つの不等式を含む論理式で達成できる:

$$A=\{D(n)<0\}AND\left\{\left[\sum_{k=0}^{K}L_{AC}(n+k)\right]\le E_{th\_stop}\right\} \tag{7.8}$$

ただし，‘AND’は論理積を表し，E_{th_stop} は経験的に決定した閾値である．K は適当な

長さの観測期間の長さを離散時間で表す整数である．これにより，連続的な観察を通じて，もし論理値 *A* が真であるならば，それは“立ちすくんでいる”と判定される．

7.5.3 予兆の検出

(1) 大きな動き

経験的に言えば，車道周辺での大きな動きは異常行動の前兆であると考えるのが自然である．特にそれが周囲と比べて特異である場合はとりわけそのように判定できる．以下のケースが考えられる：

a) 大きな移動度 (走る，速く歩く，他)

この場合，たとえ認識ターゲットである歩行者が約 20m 離れていたとしても，動きベクトルは検出可能である．しかし，その大きさが自動車より大きくなることはまれである．なお，このケースではときどき動き補償サブブロックではなく，イントラブロックが生成されことがある．

b) 腕や脚の大きな動き

この場合，BEF においても大きな差分が発生し，通常の歩行時よりも高い AC 電力を背生成する．

c) 旗，ポスター，標識などを振り回す場合

この場合，BEF においてさらにフレーム差分を発生させ，それは高いレベルの DCT 係数を発生する．

(2) 垂直方向の動き

垂直方向に隣接した複数の BEF において際立った DCT 係数が周期的に発生するならば，歩行者が上下運動を繰り返している（飛んだり跳ねたりなど）可能性がある．このクラスの垂直方向の挙動は異常行動の前兆として判定され，対応する確信度はそれまでよりも高く設定される．

7.6 おわりに

MPEG ベースの歩行者認識手法を提案し，車載カメラが交差点で停止するとき，最大で 90%以上のトラッキング率を達成することを確認した．これはあきらかに従来のスナップショットベースの手法を凌いでいる．しかし，カメラが高速移動する場合や歩行者の挙動など高度な認識への応用には，シーンクラスの判別（8 章）と挙動のモデリング（9 章）を導入することが課題となる．一方で本手法は無線通信に適しており，プローブカーのように異なる時空間に存在する複数の認識器を連携させる際にも有効である．9 章以降ではこの連携に立脚した予測と推定のプロセスにより上記の課題解決を模索する．