

特集 色情報と知識処理による車載カメラ映像の インデキシング*

Indexing of On-board Captured Images Based on Color Information and Knowledge Processing

笹木美樹男
Mikio SASAKI

In this paper, we propose a method that generates image descriptions relatively easily by using color information and knowledge processing. At first, for each block of pixels of on-board captured images that are stored in MPEG format, we generated object indices triggered by color information from decoded DCT (Discrete Cosine Transform) coefficients. Secondly, by concurrently using a three dimensional model based method and rule based reasoning, we extracted three dimensional relative locations between major objects and scene features. The results were used to automatically generate final image descriptions in XML format. Moreover, through the experiments we have confirmed that it is possible to conduct image retrieval using natural language like queries, on the basis of matching operations between the image descriptions above and query profiles.

Key words: Indexing, DCT, Model based, On-board captured images, Image retrieval

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。¹⁾²⁾³⁾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ーズにマッチしたもの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。⁴⁾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、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（離散コサイン変換）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。

2. 車載カメラ映像のインデキシング

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、ドライバモニタ、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。⁵⁾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たり、車が獲得した映像記憶を検索したりするといった応用をめざしている。

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、対向車両、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”などは一つのシーンに付与できる。

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語列を対応

* (社)情報処理学会の了解を得て、「第54回オーディオビジュアル複合情報処理研究会」(2006.9)より、一部加筆して転載

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。

Fig. 1に今回のシステム構成を示す。インデキシング部ではMPEG-1形式⁶⁷⁾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ごとに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 (Table 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。

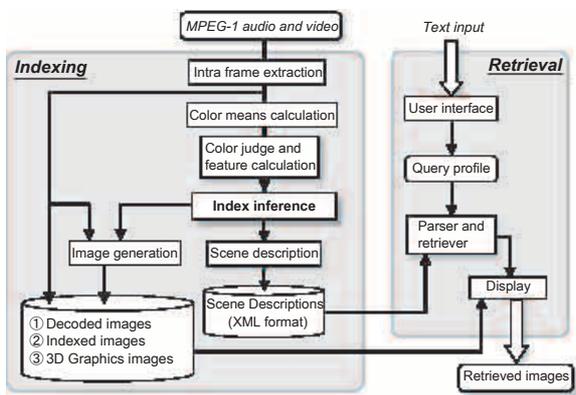


Fig. 1 System architecture

3. 色インデキシング

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成分の各々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

Table 1 List of color index

No.	Word	Objects to be judged on color								
		On road			Not specified			Out of road		
1	Blue	Blue sign	Blue vehicle	Blue signal	Sky	-	-	-	-	Blue building
2	Green	Green sign	Green vehicle	Green light	-	-	-	Green area	Street trees	Green building
3	Red	Red sign	Red vehicle	Red signal	-	-	-	-	Street trees	Red building
		Stop light	Street light	Headlight	-	-	Red light	Red poster	Red flower	-
4	Orange	Orange sign	Orange vehicle	Median divider	Sky	Construction board	Orange light	-	-	Orange building
5	Yellow	Yellow sign	Yellow vehicle	Yellow signal	-	Yellow signboard	-	Yellow flower	Street trees	Yellow building
		-	-	Headlight	-	-	-	Yellow flower	-	-
6	White	Guardrail	White vehicle	White line	Sky	-	-	White flower	-	White building
		Marked crosswalk	-	Headlight	-	-	-	-	-	-
7	Gray	Road	Gray vehicle	-	Sky	-	-	-	-	Gray building
8	Black	Shadow on road	Black vehicle	Shadow of vehicle	Night sky	-	Wiper	-	-	Black building
-	(unknown)	Sign	Oncoming vehicle	Preceding vehicle	Pedestrian	-	Parking vehicle	-	-	Building

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Fig. 2のようになる。

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、Table 1に示す八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Fig. 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、次の二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。

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

（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Aとみなす。一方、k番目の色に対応する代表ベクトル $A_R(k)$ ($k=1, \dots, K$)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、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。

$$D_A(A, A_R(k), w(k)) = \frac{1}{W(k)} \sum_{i=1}^3 w_i(k) |A_i - A_{Ri}(k)| \quad (2)$$

ただし、

$$A = (A_1, A_2, A_3) = (Y, U, V)$$

$$A_R(k) = (A_{R1}(k), A_{R2}(k), A_{R3}(k)) = (Y_R(k), U_R(k), V_R(k))$$

$$w(k) = (w_1(k), w_2(k), w_3(k)), W(k) = w_1(k) + w_2(k) + w_3(k)$$

であり、 $w(k)$ はk番目の色に対応する荷重係数ベクトルである。

3.3 車載カメラ映像の色インデキシング例

まず、Fig. 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。Fig. 3にその結果を抜粋して示す。目視の評価では、

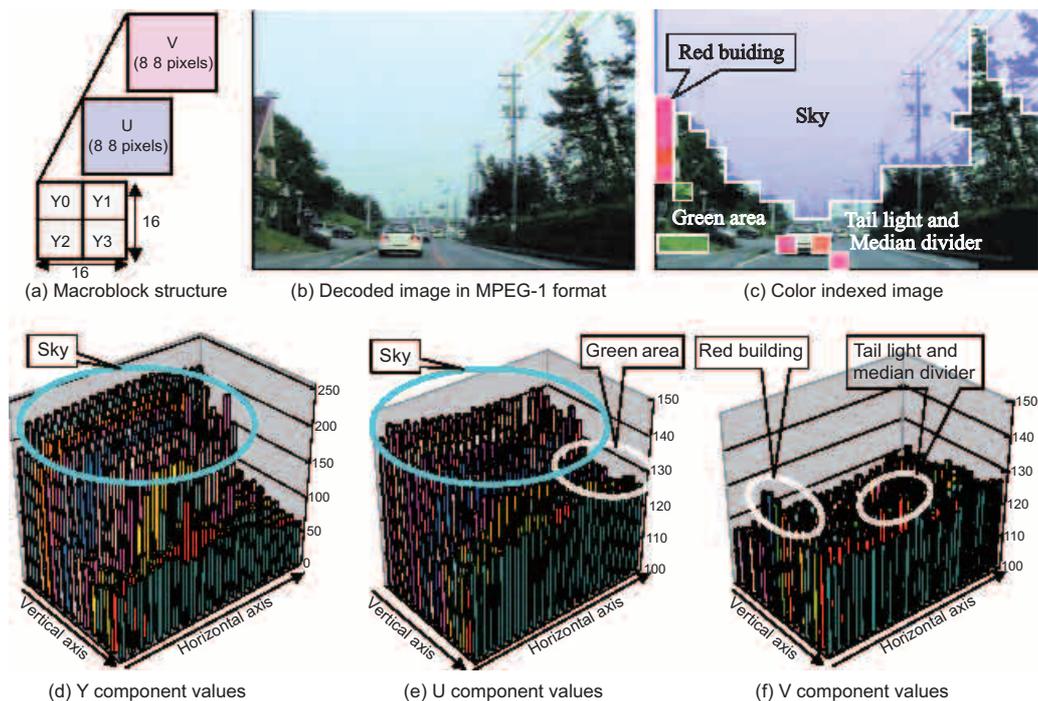


Fig. 2 Analysis of YUV component

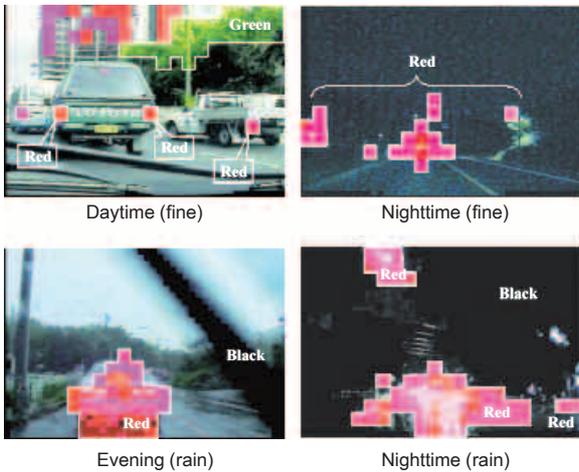


Fig. 3 Examples of color indexed images

前方車両，対向車線の車両，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分かった。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，対向車両などは昼間の画像よりも抽出されやすいことが分かった。

また，雨天ではワイパ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分かった。なお，画像中のワイパ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。

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⁹⁾，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。⁹⁾¹⁰⁾¹¹⁾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。

4.1 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Fig. 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。¹²⁾¹³⁾¹⁴⁾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。Fig. 5はFig. 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，Fig. 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Fig. 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

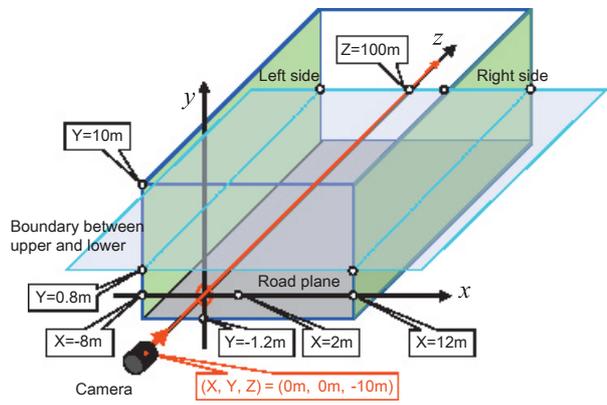


Fig. 4 Solid structure model of road scene

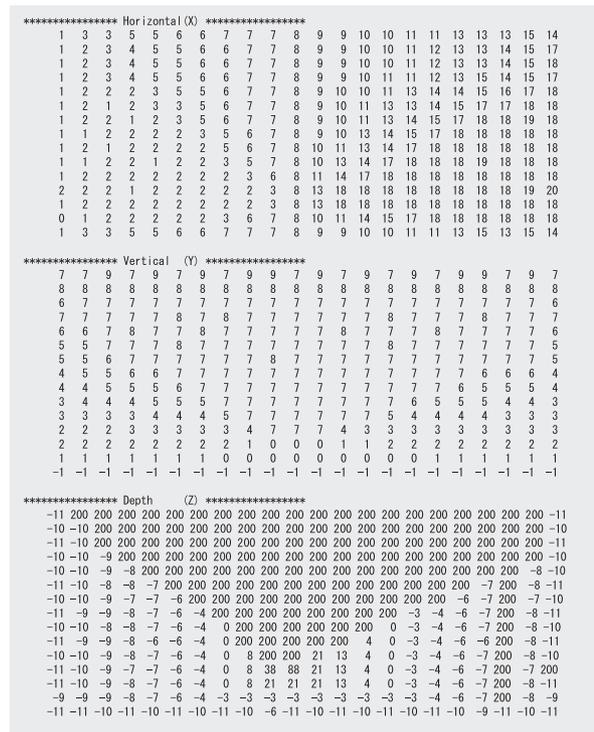


Fig. 5 Distance information estimated from road structure model

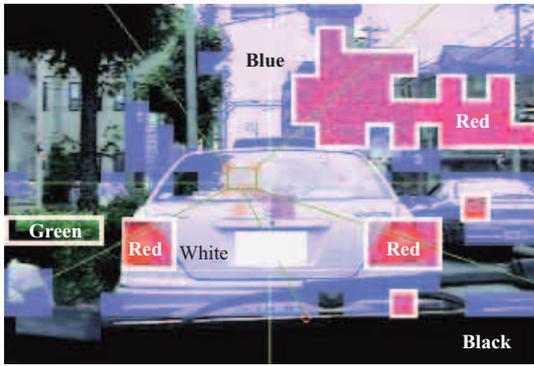


Fig. 6 Color indexed image

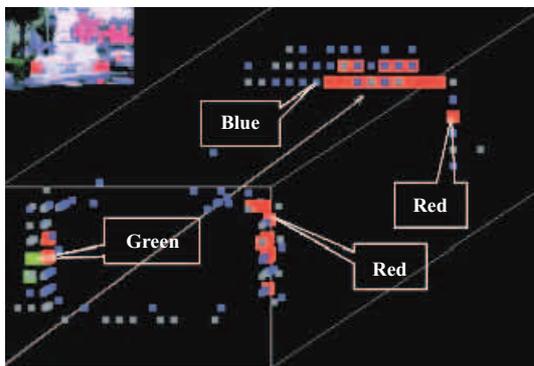


Fig. 7 Three dimensional display of depth

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。

4.2 空仮説によるモデルの修正

そこで、Fig. 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。

$$Y_{MBK_mean}(m_r, m_c) \geq Y_{th} \quad (3)$$

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 の四つのYブロック(8×8画素)が存在するため、次式で計算する。

$$Y_{MBK_mean}(m_r, m_c) = \frac{1}{4} \sum_{i=0}^3 Y_{BLK_mean}(m_r, m_c, i) \quad (4)$$

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($i=0, \dots, 3$)の平均画素値である。

この空仮説をFig. 5の距離情報マップに適用すると、Fig. 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Fig. 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。

5. オブジェクトの判別

5.1 確信度ベクトル

笹木¹⁵⁾は車載環境における N 個のアプリケーションやコンテンツジャンルの各々について要求度を表現する要求ベクトルという概念を導入し、制約充足型の推論システムを提案した。本稿では、この要求ベクトルをインデックスの妥当性を表現する確信度ベクトルという概念に置き換えて推論システムを構築する。

いま、あるマクロブロックにインデックス A_i を付与することに対する確信度を C_i として、

$$C = (C_1, C_2, \dots, C_i, \dots, C_N)^T \quad (5)$$

を確信度ベクトルと定義する。

この確信度ベクトルは色やブロック位置、他のオブジェクトとの関係など様々な判定規則で個別に算出した部分的な確信度ベクトルの和として次式で表現する。

$$C = c_0 + \sum_{m=1}^M c_m \quad (6)$$

ただし、 c_0 : 初期条件として与えられる確信度ベクトル(N次元列ベクトル)

c_m : m 番目($m=1, \dots, M$)にヒットした判定規則に対応する部分的な確信度ベクトル(N次元列ベクトル)

である。この確信度ベクトルの最終値において、最大成分に相当するインデックスをマクロブロックに該当するオブジェクトであると判断する。

この手法の利点は従来のオブジェクト認識で用いら

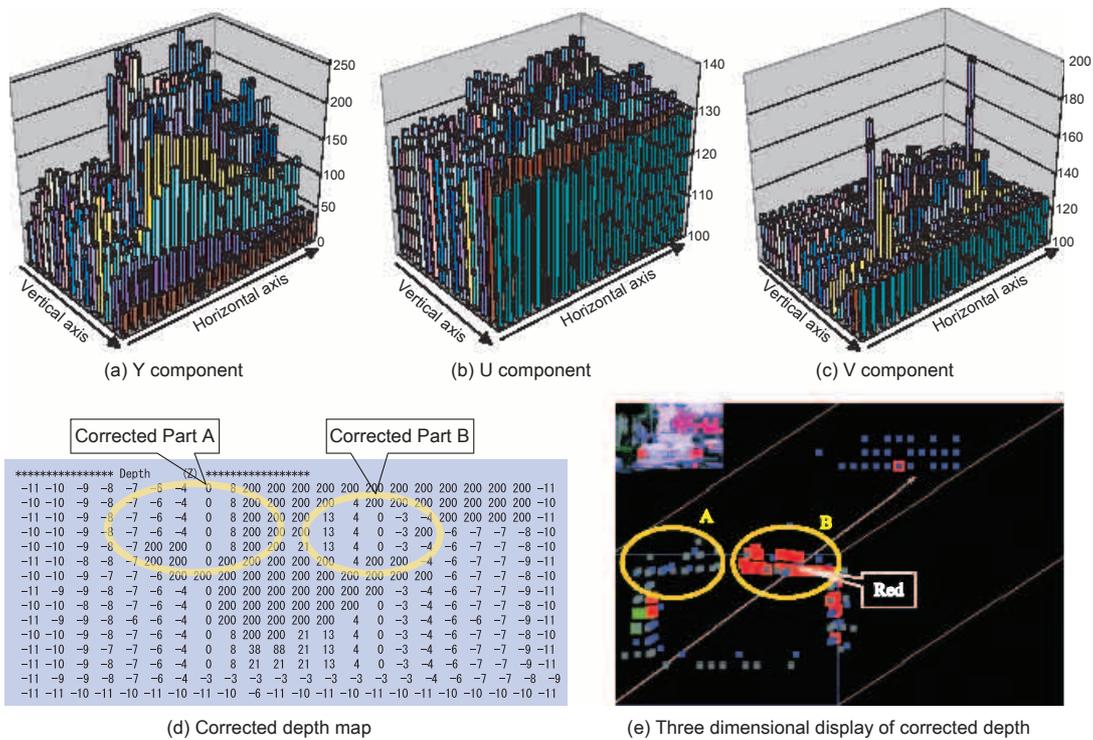


Fig. 8 Correction of depth on the basis of sky information

れてきた特徴空間の分割に基づく手法に比べて、考えられる複数の解釈に対応する確信度という形で認識の状態を保持し、随時ある基準の元で唯一の解釈を動的に選択できるところにある。

5.2 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、Fig. 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 (Fig. 9)。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。

5.3 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。

5.3.1 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。

5.3.2 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二つのテールランプ位置のマクロブロック (赤) を検出し、二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

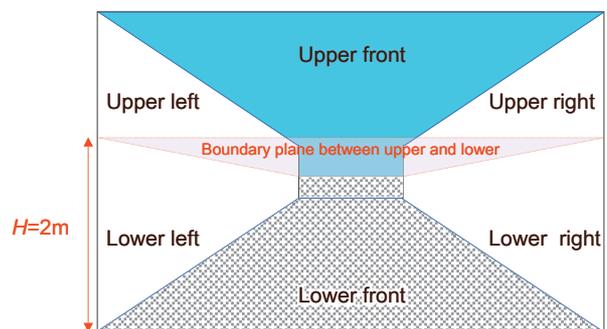


Fig. 9 Area definition on image plane

き位置 $Z(m)$ ($m=1, \dots, 5$)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_{zh} 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。

5.3.3 ビルと街路樹

ビルの判別の際には、上述のようなブロック単位の色判定に加え、Fig. 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。すなわち、Fig. 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。

$$Q_{AC} = \sum_{l=0}^5 S_{AC}[l] \quad (7)$$

ただし、 $S_{AC}[l]$ はブロック別のAC成分発生量であり、 $l=0, 1, \dots, 5$ 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、

$$S_{AC}[l] = \sum_{h=0}^7 \sum_{v=0}^7 |B_l(h,v)| - |B_l(0,0)| \quad (8)$$

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、

$$Q_{AC} \geq Q_{AC_tree} \quad (9)$$

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。

$$G_{AC} = \sum_{l=0}^5 \frac{G_V[l]}{G_H[l]} \quad (10)$$

$$G_H[l] = \frac{1}{64} \sum_{h=0}^7 \sum_{v=0}^7 (h+1) |B_l(h,v)| \quad (11)$$

$$G_V[l] = \frac{1}{64} \sum_{h=0}^7 \sum_{v=0}^7 (v+1) |B_l(h,v)| \quad (12)$$

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Fig. 10に示す。この結果から経験的に、

$$G_{AC} \geq 8.0 \quad (13)$$

のとき樹木の確信度を高め、それ以外るときビルの確信度を高める操作を行った。

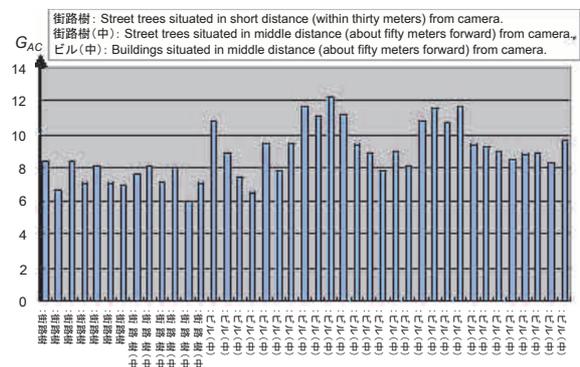


Fig. 10 Relative intensity of vertical components in AC coefficients

6. 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。

6.1 時系列ヒストグラムの判定

Fig. 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ごと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 (Fig. 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。すなわち、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高い。
- 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高い。
- 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高い。

同様にまた、雨の日の夕刻の走行映像についても赤

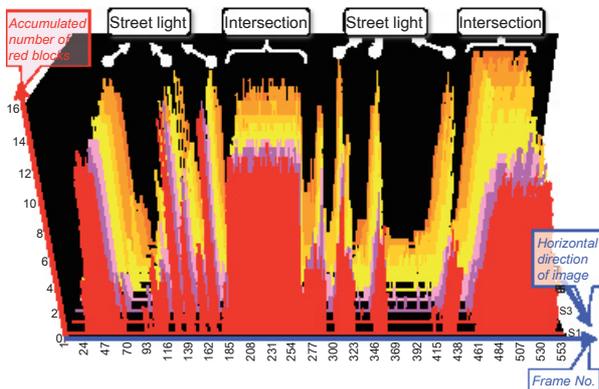


Fig. 11 Time series histogram regarding red regions (nighttime)

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高い。
- 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高い。
- 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高い。

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

6.2 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。

$$D_n = \sum_{k=0}^{k-1} \left\{ W_k \sum_{i=0}^5 |Z_0 [i, Nd(k), n] - Z_0 [i, Nd(k), n-1]| \right\} \quad (14)$$

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

$Nd(k)$: 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。

7. 実験

以上にに基づき、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。

7.1 実験システム

市販デ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、フロントガラス越しに前方の道路状況を撮影した (Fig. 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Fig. 1のシステムの入力とした。

7.2 認識結果の視覚表現

本システムは、色インデックス画像 (Fig. 6) をトリガーとしてシーン構造モデル (Fig. 4)、空仮説 (Fig. 8)、車仮説、領域情報 (Fig. 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 (Fig. 13)。

7.3 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Fig. 14(a)に示す。これに対し、Fig. 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Fig. 14(d)に示す。ここでは「赤いビル」に対応するオブジェクトも同時に表示されているが、誤りではない。

メタデータが事前に作成されている場合、上記の検索タスクそのものは実質的にテキストファイル中のキーワードマッチングになるため、ほとんど時間を要しない。したがって検索の効果は検索結果をブラウズす



Fig. 12 On-board camera

る画像の枚数と内容の妥当性で判断するのが適切である。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

$$S = N_R / N_A \tag{15}$$

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¹⁰⁾。

$$\alpha_r = \frac{N_P}{N_P + N_Q} \tag{16}$$

$$\alpha_p = N_P / N_R \tag{17}$$

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¹⁰⁾を用いれば、

$$F = \frac{2}{\alpha_r^{-1} + \alpha_p^{-1}} \tag{18}$$

となる。ヒューマンインタフェースの観点からは F ($0 \leq F \leq 1$) を最大化し、 S ($0 \leq S \leq 1$) 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Table 2とFig. 15に示す。

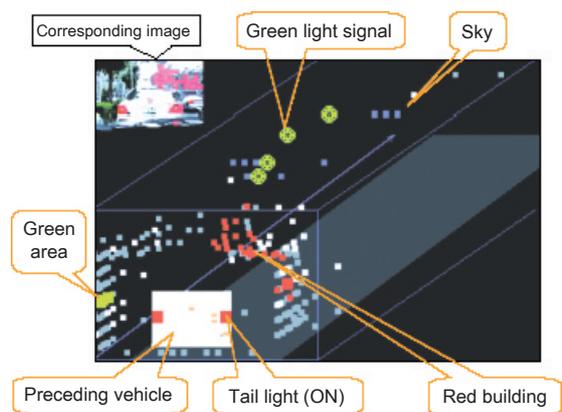


Fig. 13 Three dimensional expression of estimated results

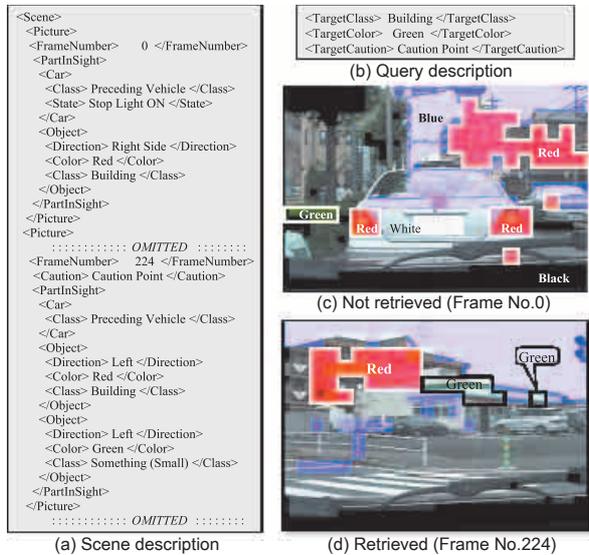


Fig. 14 Description and retrieval of scene

8. 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、ユーザがブラウザ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 (Table 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と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統計的推論手法の併用が考えられる。既に我々は動きや

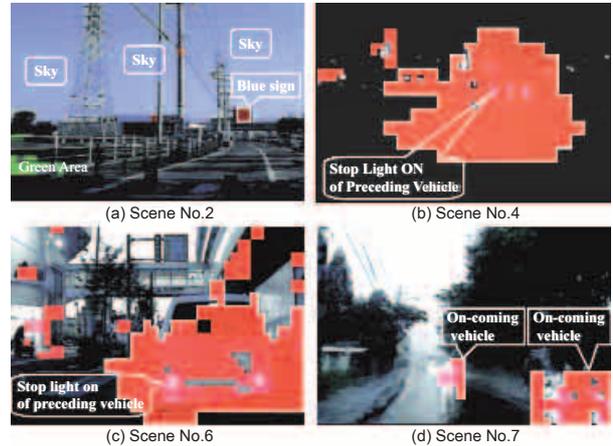


Fig. 15 Examples of retrieved results

AC成分を含む多変量回帰分析を用いて景観認識実験を開始した。¹⁷⁾¹⁸⁾¹⁹⁾ その詳細は次回に報告する。

9. おわりに

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。

Table 2 Experimental results of image retrieval

No.	Time zone	Weather	Source Images			Query words	Retrieved results						
			Road class	Date of capture	N_A		N_R	N_P	N_Q	S	α_r	α_p	F
1	Morning	Fine	General road in residential area	2004/6/10	300	Green building and caution point	12	10	0	0.04	1.00	0.83	0.91
2	Morning	Fine	General road in residential area	2005/11/21	202	Blue sign	22	11	6	0.11	0.65	0.50	0.56
3	Day	Fine	Suburban toll road	2005/10/20	277	Blue sign	34	26	6	0.12	0.81	0.76	0.79
4	Night	Fine	By-pass road with median divider	2005/10/20	277	Stop light ON of preceding vehicle	7	5	0	0.03	1.00	0.71	0.83
5	Night	Fine	By-pass road with median divider	2005/10/21	60	Passing under street light	9	9	1	0.15	0.90	1.00	0.95
6	Evening	Fine	Main street near by residential area	2005/10/20	277	Stop light ON of preceding vehicle	3	3	3	0.01	0.50	1.00	0.67
7	Evening	Rain	Suburban toll-free road	2004/6/25	315	On-Coming vehicle	48	48	31	0.15	0.61	1.00	0.76
Average	—	—	—	—	244.0	—	19.3	16.0	6.7	0.09	0.78	0.83	0.78

<参考文献>

- 1) ISO/IEC 15938-3 Information technology-Multi-media Content Description Interface Part 3: Visual.
- 2) ISO/IEC 15938-5 Information technology-Multi-media Content Description Interface Part 5: Multimedia Description Schemes.
- 3) B. S. Manjunath et al., "Introduction to MPEG-7", WILEY (2002).
- 4) 進藤, 浅川, 「一家に一台, 映像ナビ」, 日経エレクトロニクス, 2004年6月21日号, pp. 105-129.
- 5) 狩集, 浅川, 「ドライブレコーダ あなたの運転, 記録します」, 日経エレクトロニクス, 2005年10月24日号, pp. 101-119.
- 6) "Information technology-Coding of moving pictures and associated audio for digital storage media at up to about 1,5 Mbit/s-Part 1: Systems", ISO/IEC11172-1 First edition 1993-08-01.
- 7) "Information technology-Coding of moving pictures and associated audio for digital storage media at up to about 1,5 Mbit/s-Part 2: Video", ISO/IEC11172-2 First edition.
- 8) 金出武雄, コンラッド・ポールマン, 森田俊彦, "因子分解法による物体形状とカメラ運動の復元", 電子情報通信学会論文誌 D-II, J74-D-II-8 (Aug.1993), pp. 1497-1505.
- 9) 笹木, "モデルベース顔領域抽出と領域別量子化による顔画質向上", 1997年映像情報メディア学会年次大会 ITE 97 (1997年7月), pp. 381-382
- 10) M. Sasaki, A. Toyoda, "3D Motion Estimation of a Moving Object using Model-Based Techniques", ITS95 (November 1995), pp. 1075-1081.
- 11) 浅田, 白井, "マルチセンサ情報を動的に統合することによる道路シーンの解釈とモデリング", 情報処理学会論文誌, Vol.31, No.12, (Dec.1990), pp. 1743-1754.
- 12) Youichi Horry, Ken-ich Anjyo and Kiyoshi Arai, "Tour into the picture", Proc. SIGGRAPH'97, Los Angels, California, August 3-8 (1997), pp. 225-232.
- 13) 杉山, 目黒, 金子, "少数の透視投影画像に基づいて生成された仮想空間における3次元移動表現", 信学技報 MVE 2001-148 (2002-3).
- 14) 石川, 相澤, 羽鳥, "簡易モデルのあてはめによる3次元仮想空間の構築手法", 画像符号化・映像メディア処理シンポジウム (PCSJ・IMPS98) I-1.4, Oct. 26-28 (1998), pp. 9-10.
- 15) 笹木, 難波, 粉川, 小川, "車載情報通信におけるメディア適応化手法の提案 - 制約充足問題を用いた適応化手法の検討 - ", 信学技報 ITS2004-22, Vol.104 No.325 (2004年9月), pp. 35-42.
- 16) 北, 津田, 獅々堀, "情報検索アルゴリズム", 共立出版 (2002).
- 17) 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 - ", 2006年度人工知能学会全国大会 (JSAI 2006) 1E2-2 (2006年6月).
- 18) 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-54-9 (2006).
- 19) 笹木, "統計的推論に基づく景観認識の検討", 情報処理学会研究報告, 2006-AVM-54-8 (2006).



<著者>



笹木 美樹男
(ささき みきお)
基礎研究所
知能情報通信の研究開発に従事