

MPEG Based Vision

2014年12月

笹木 美樹男

1章 序論①

beachの例



人は静止画でもシーンを予想できる

時空間をわかりやすい単位で伝える

1章 序論②

シーンの自動分類ができると・・・

- 分類の手間が省ける
- ためこんだ映像が容易に取り出せる
- 機械の概念学習が容易になる

- ユーザの目的は？
鑑賞？ 検索？ 計測？ ……

- ・アプリによってクラス定義が異なる
- ・アプリによって同じクラス名でも分類規範が異なる
- ・クエリによって対象クラスは異なる(辞書が変わる)
- ・ユーザプロファイルの影響を受ける



BGS (Back Ground Scene)

- 音楽のBGS
- アメニティのBGS

漠然とした検索・ブラウズ



2. 関連研究

景観認識

一般物体認識

MPEG画像の解析

歩行者認識

ヒューマントラッキング

構造復元

シーンの予測

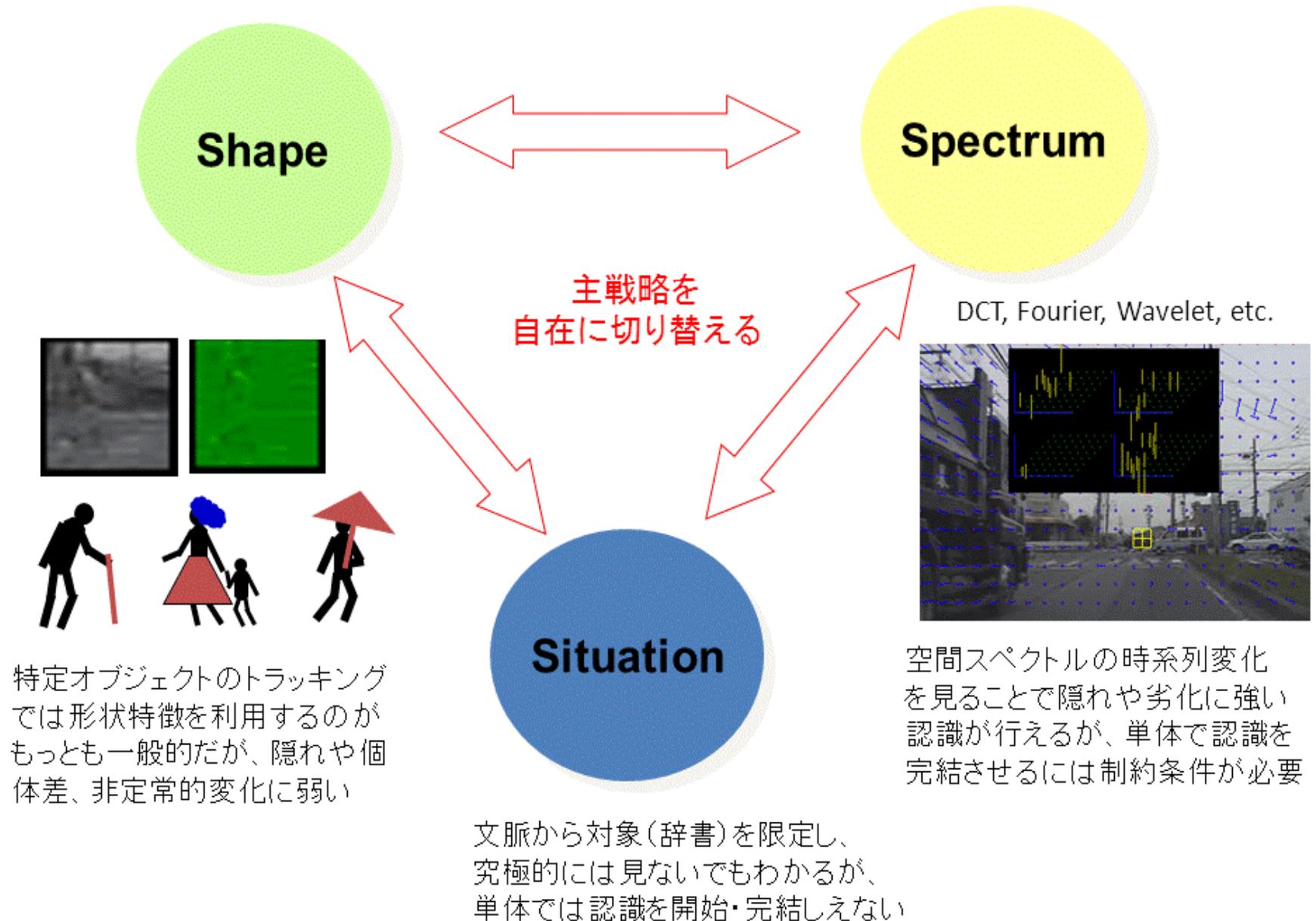
アクティブセンシング

3. 理論①

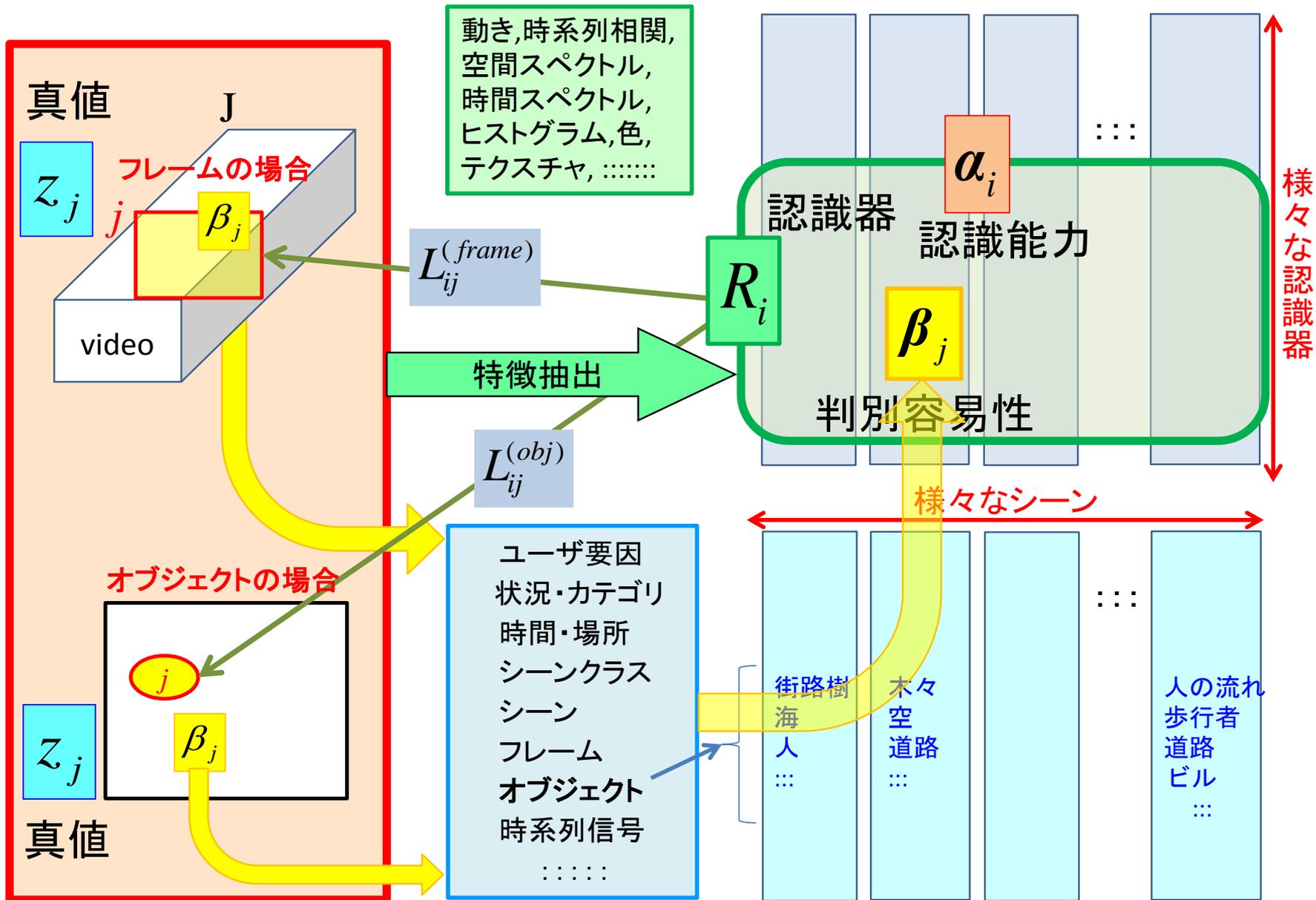


プローブ協調と知識処理により符号化画像を認識する

3. 理論②

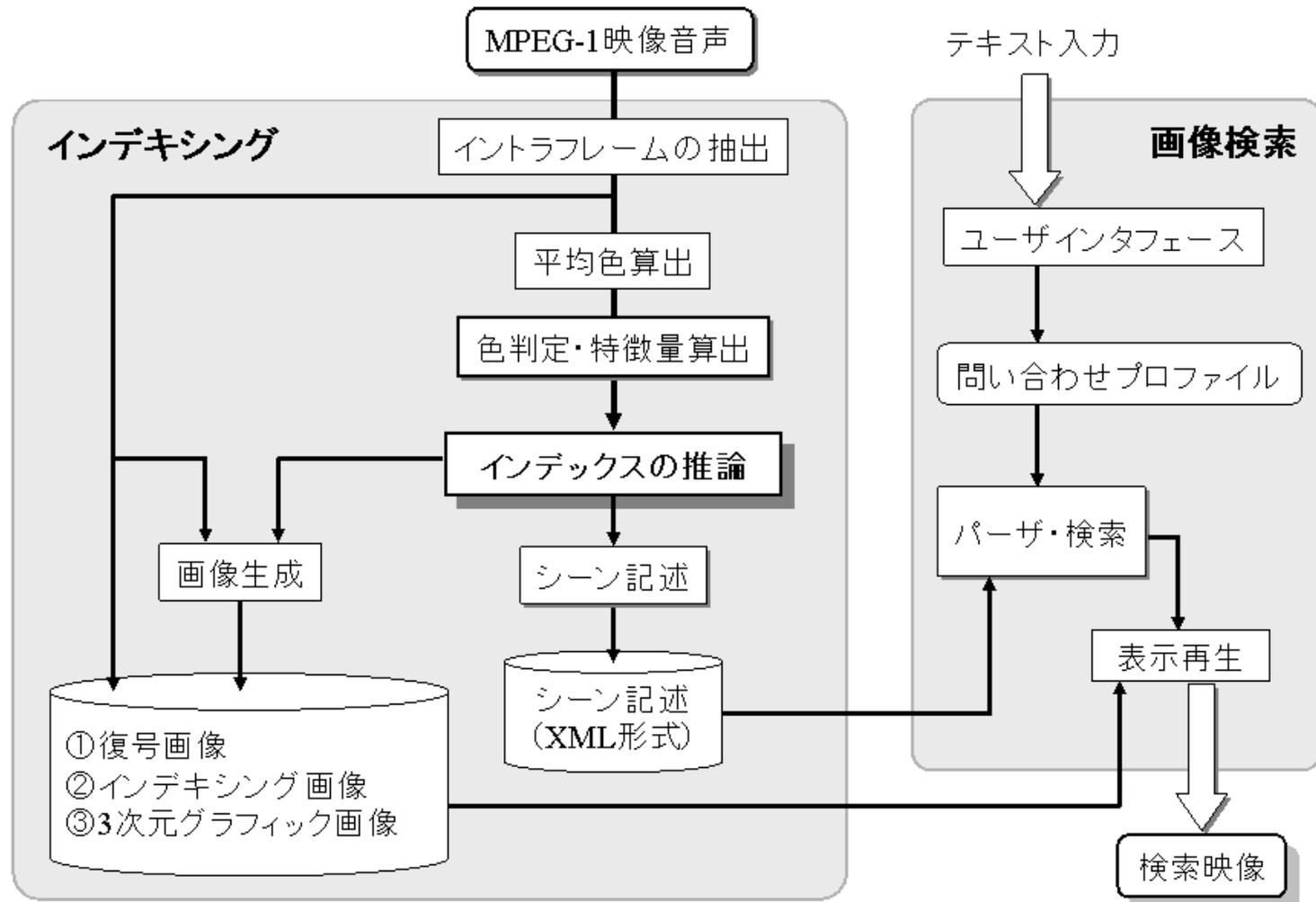


3. 理論③

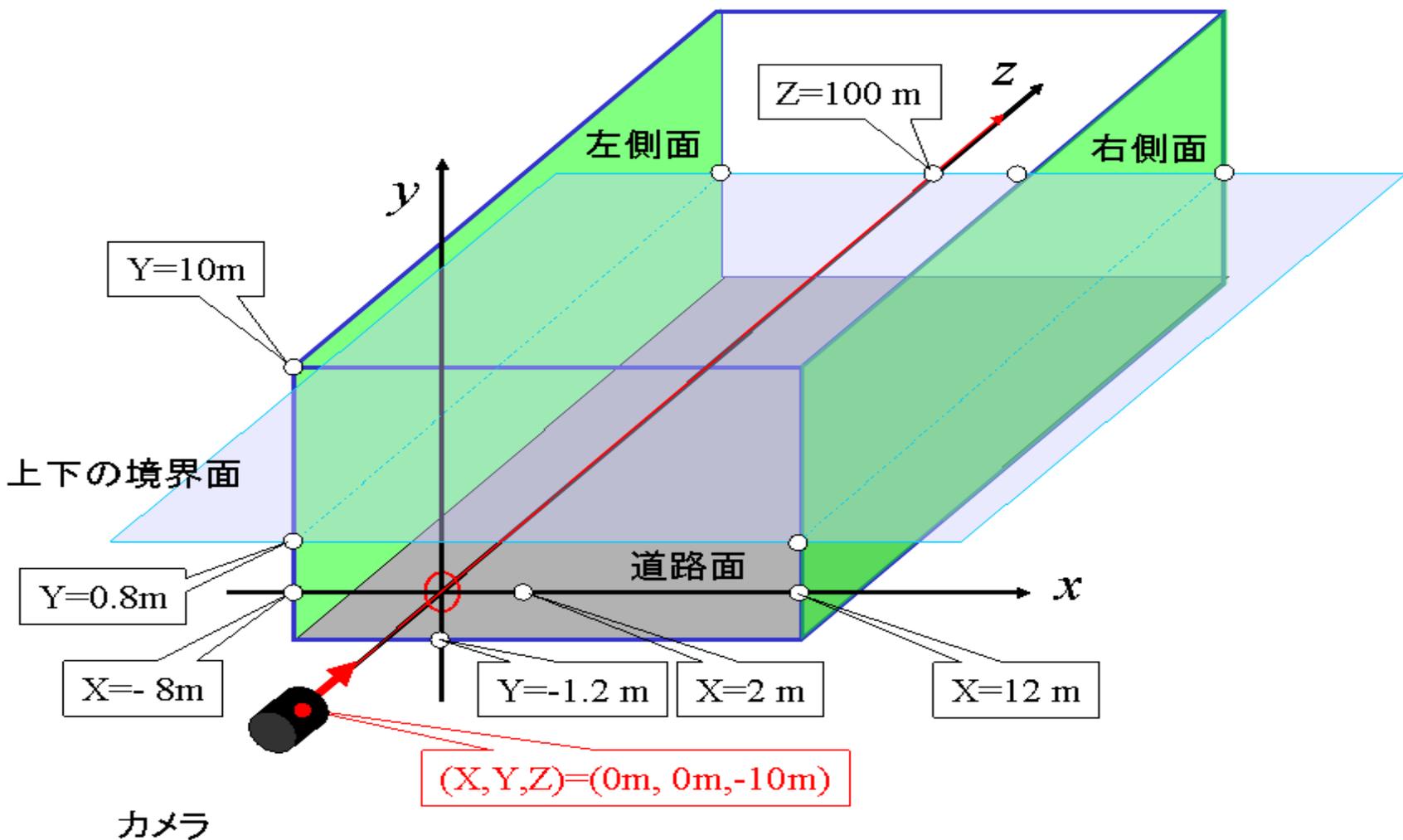


4章 ルールベースによる景観認識①

景観をひとことで表現する



4章 ルールベースによる景観認識②



4章 ルールベースによる景観認識③

```
<Scene>
  <Picture>
    <FrameNumber> 0 </FrameNumber>
    <PartInSight>
      <Car>
        <Class> 前方車両 </Class>
        <State> 停止ランプ点灯 </State>
      </Car>
      <Object>
        <Direction> 右手 </Direction>
        <Color> 赤い </Color>
        <Class>ビル? </Class>
      </Object>
    </PartInSight>
  </Picture>
  <Picture>
    :::::::::::::: 途中省略 ::::::::::::::
    <FrameNumber> 224 </FrameNumber>
    <Caution> 注意箇所 </Caution>
    <PartInSight>
      <Car>
        <Class> 前方車両 </Class>
      </Car>
      <Object>
        <Direction> 左手 </Direction>
        <Color> 赤い </Color>
        <Class>ビル? </Class>
      </Object>
      <Object>
        <Direction> 左手 </Direction>
        <Color> 緑色 </Color>
        <Class> 何か(小さい) </Class>
      </Object>
    </PartInSight>
  </Picture>
  :::::::::::::: 以下省略 ::::::::::::::
```

(a) Scene description

```
<TargetClass> ビル? </TargetClass>
<TargetColor> 緑色 </TargetColor>
<TargetCaution> 注意箇所 </TargetCaution>
```

(b) Query description

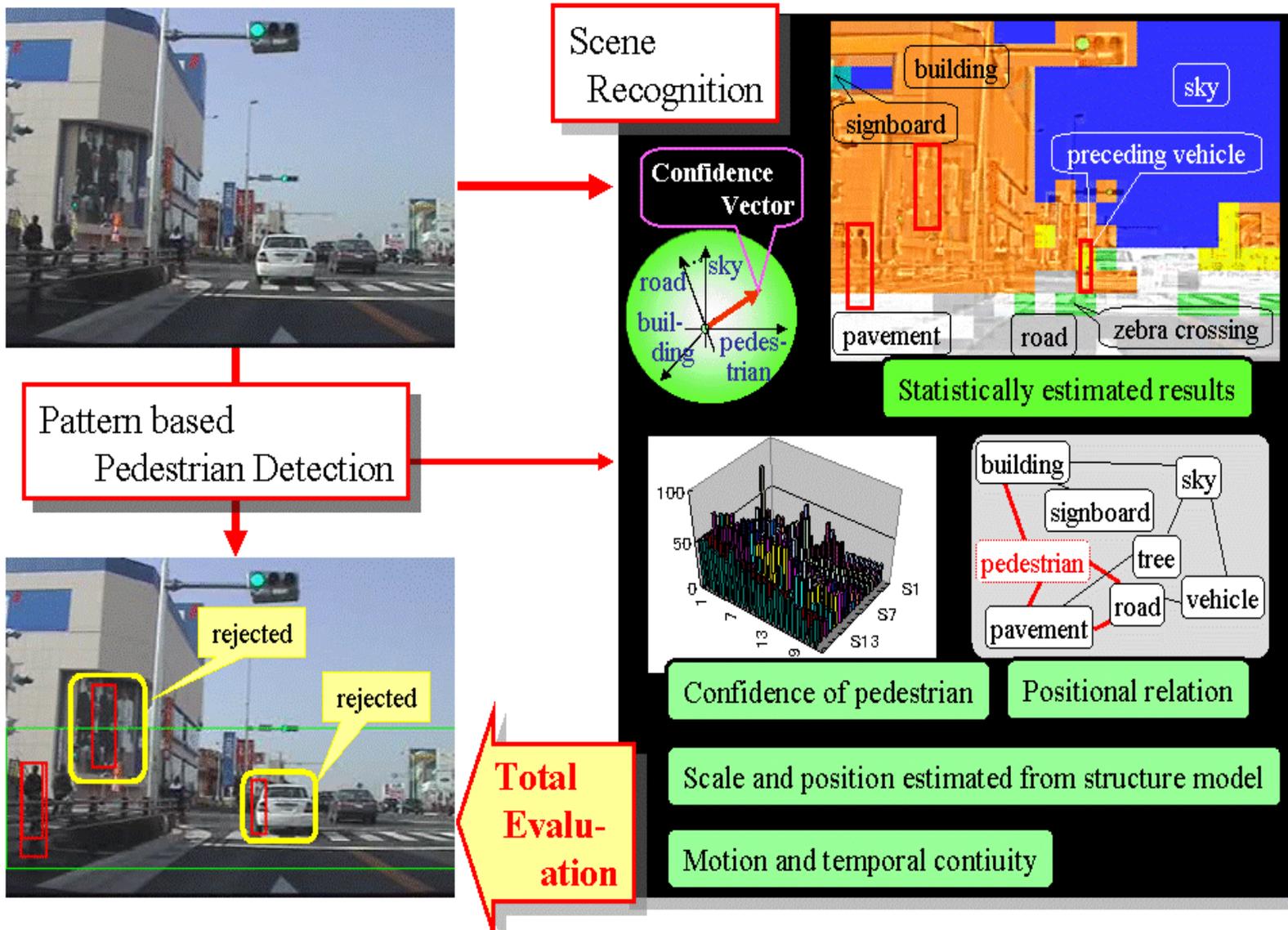


(c) Not retrieved (Frame No.0)



(d) Retrieved (Frame No.224)

5章 景観認識を用いた歩行者認識の改善



6章 統計的推論に基づく景観認識



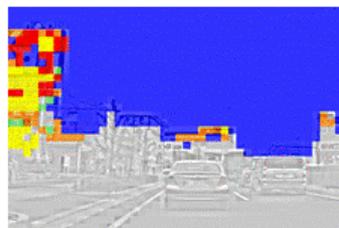
[070628-1734_Aoki]

(a) 認識画像

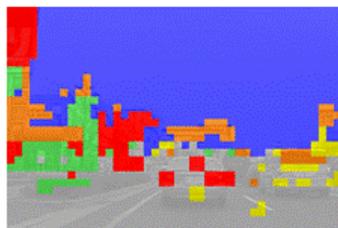


[070628-1744_長久手APITA]

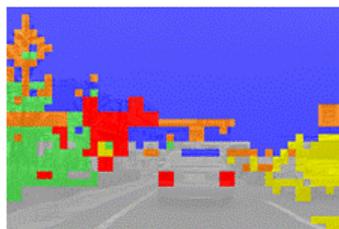
(b) 学習画像



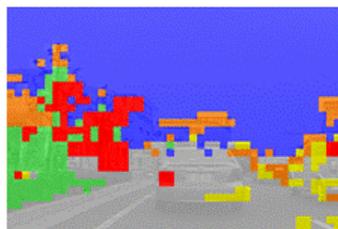
Frame No. 1



Frame No. 38



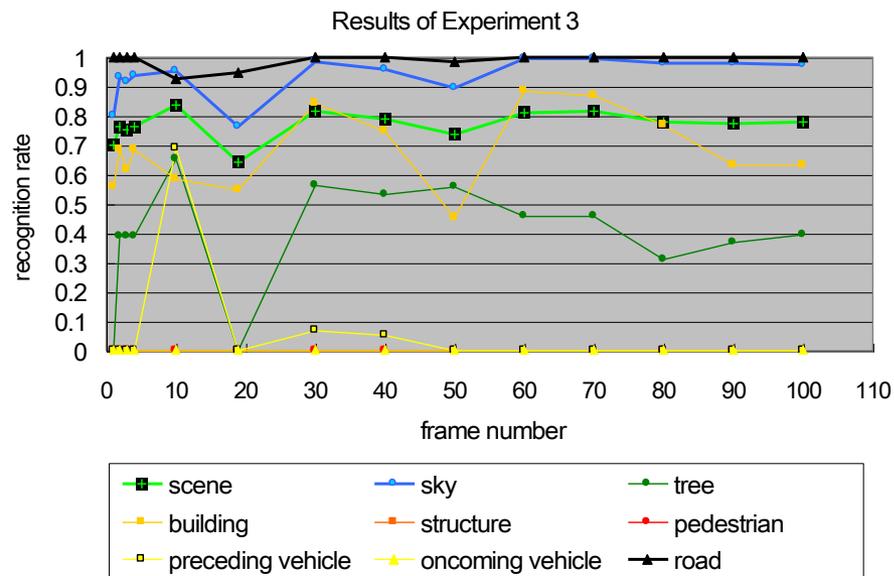
Frame No. 75



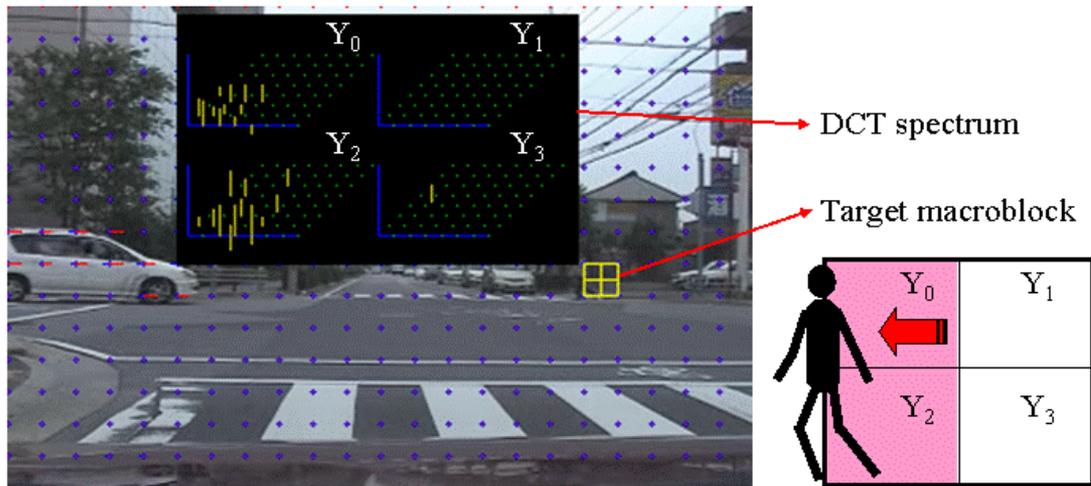
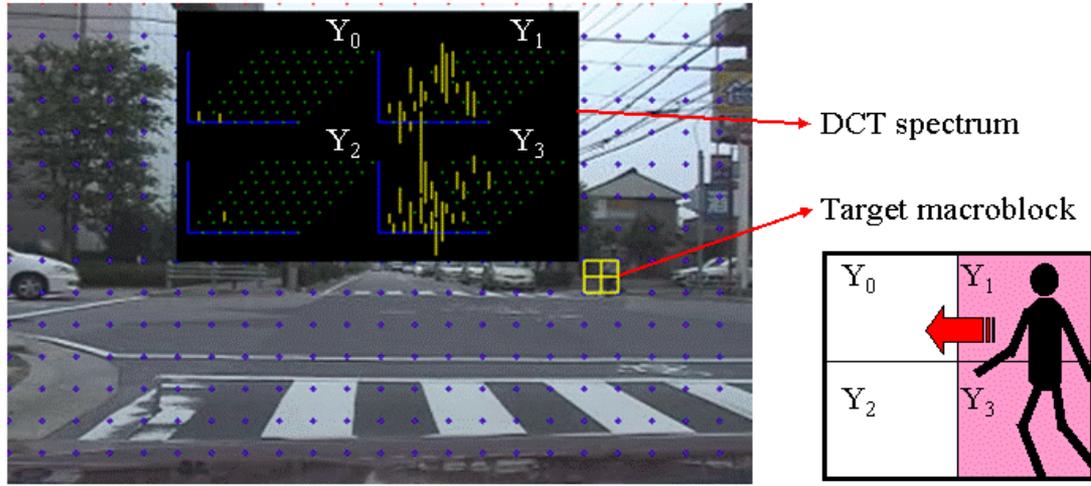
Frame No. 112

各インデックスの色: 空 赤い看板ビル 街路樹 先行車 対向車 道路

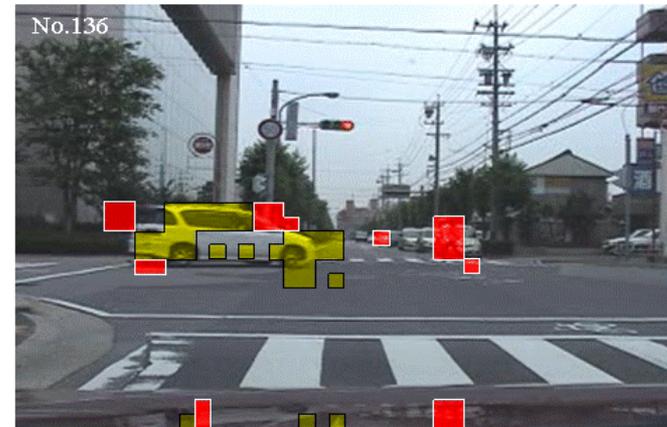
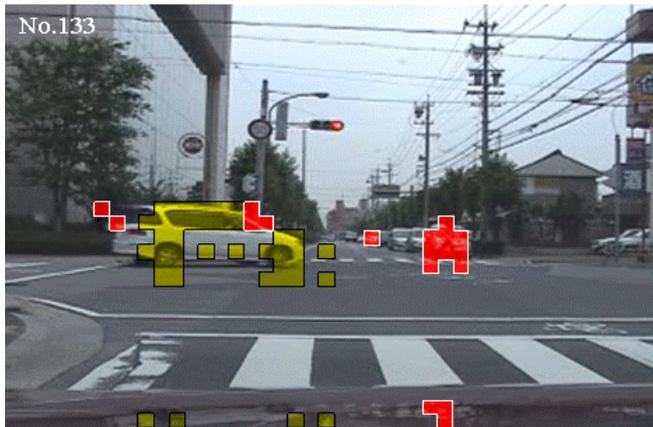
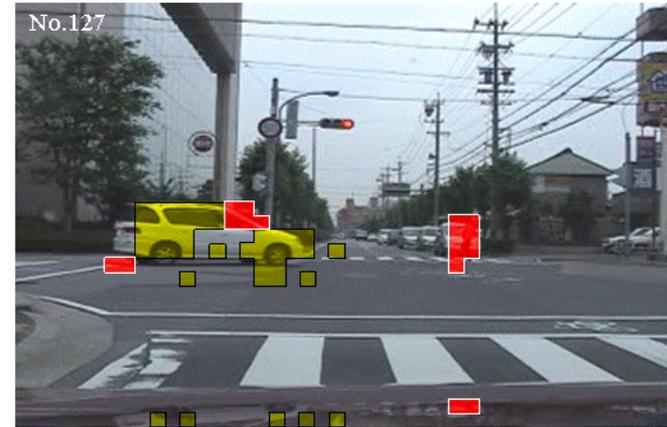
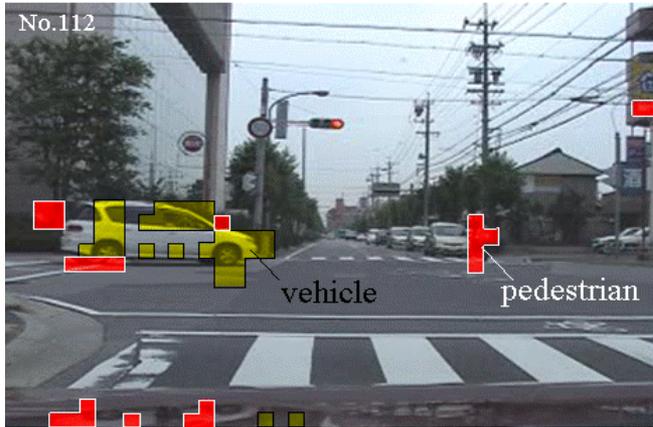
(c) インデックス画像



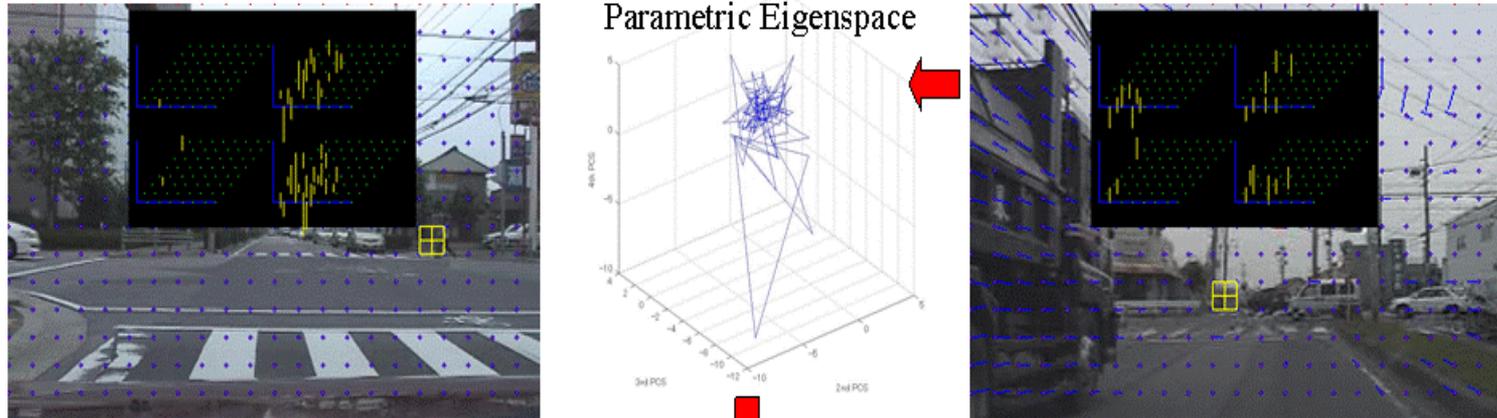
7章 移動体検出①



7章 移動体検出②



7章 移動体検出③

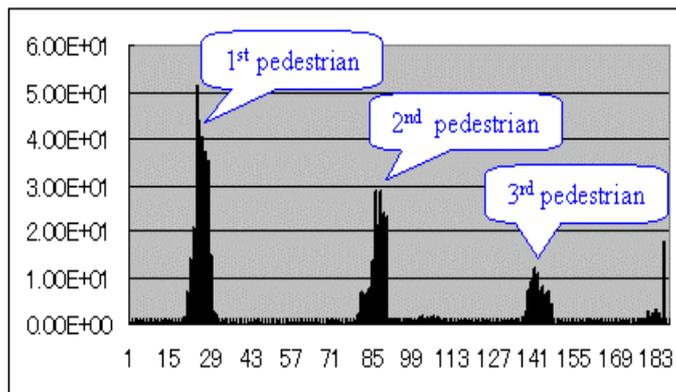


Extract principal components

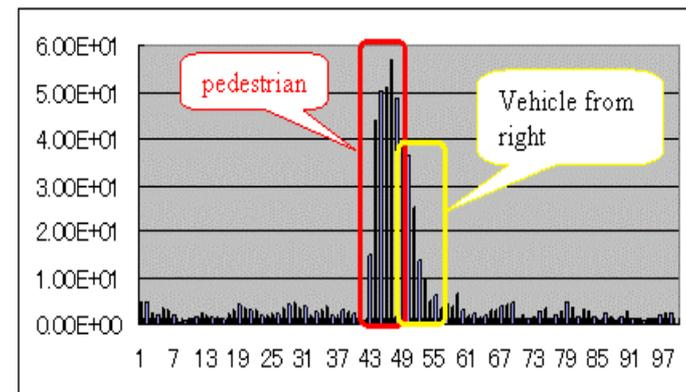
Use P-effective principal components in evaluation function

Confidence by self learning (P=3)

Confidence by cross learning (P=3)

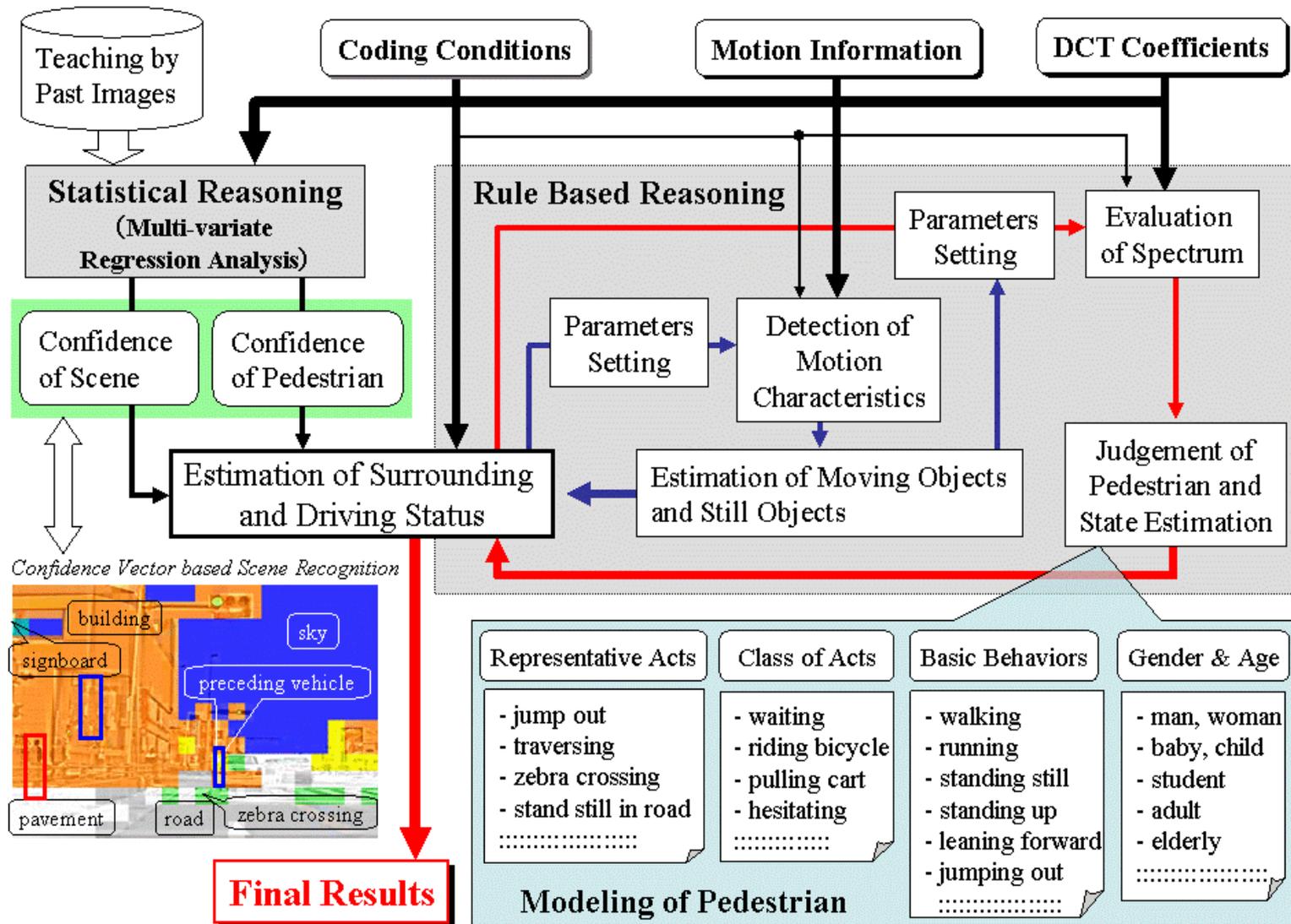


Number of predictive frame



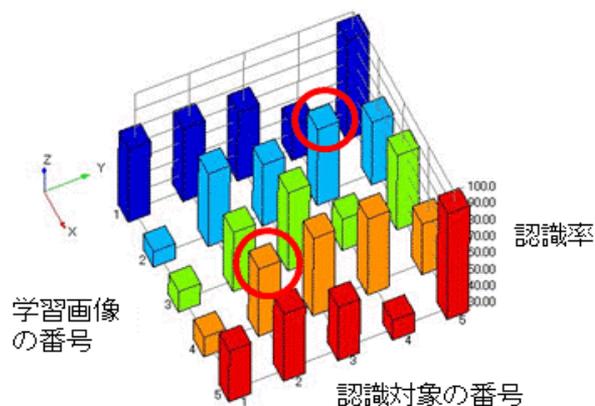
Number of predictive frame

7章 移動体検出④



8章 シーンクラスと景観認識①

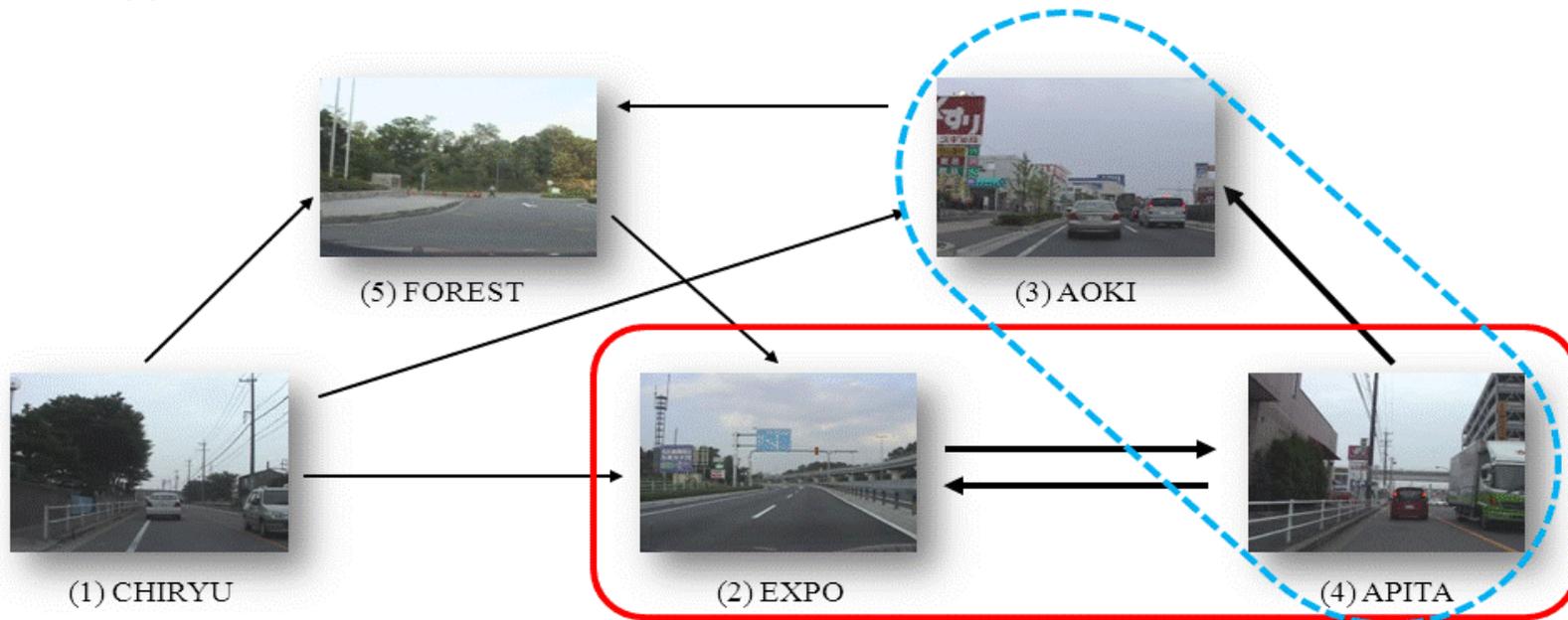
分類型のラベルでシーンを早期に判定



(a) 交叉学習によるシーン認識率

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

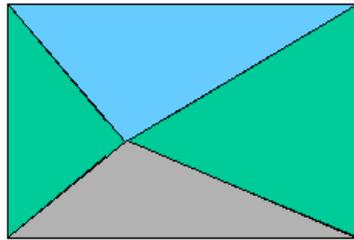
(b) 各シーンの主観的シーンクラス (SSC)



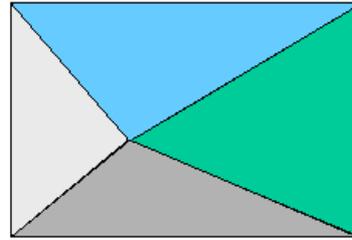
(c) 各シーン間の認識能力を示す有向グラフ

8章 シーンクラスと景観認識②

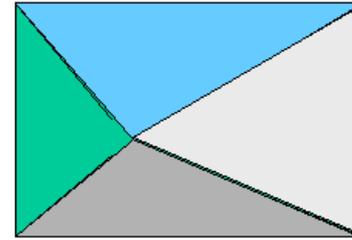
道路構造を用いたシーンクラス



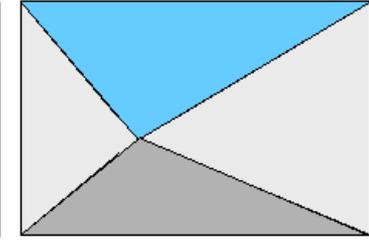
(1) Greens on both sides



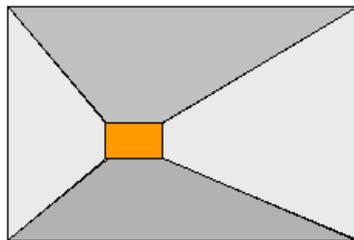
(2) Greens on the right side



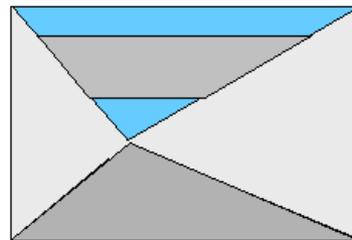
(3) Greens on the left side



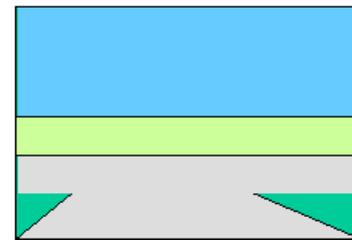
(4) Buildings on both sides



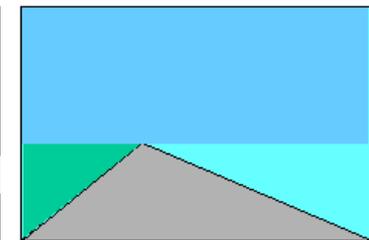
(5) Indoor



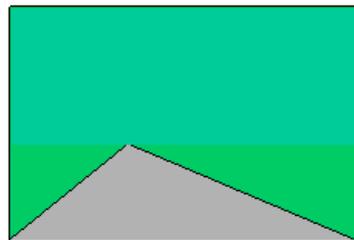
(6) Passing under overpass



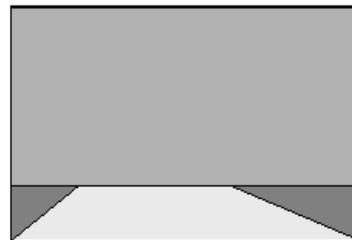
(7) T junction, greens forward



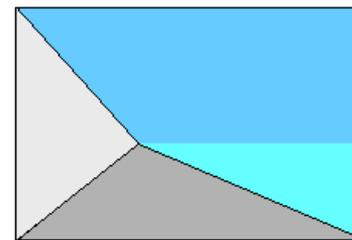
(8) Green planes on the left side, waters on the right side



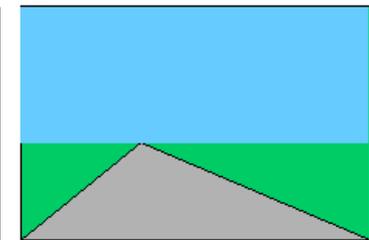
(9) Greens in all directions



(10) Walls forward



(11) Buildings on the left side, waters on the right side

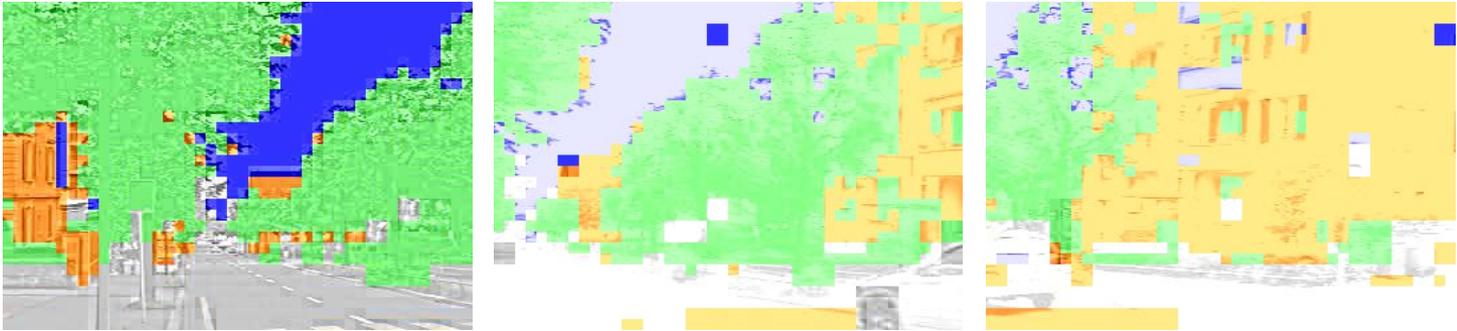


(12) Green planes on both sides

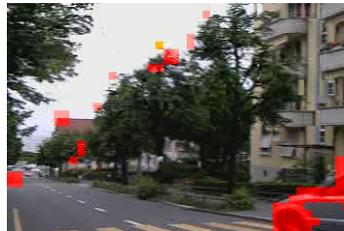
8章 シーンクラスと景観認識③



(a) Decoded image (070612_Lausanne, Frame No.1, 51, 74, respectively).



(b) Recognized results (070612_Lausanne, Frame No.1, 51, 74, respectively).



(c) Regions of high AC power
(070612_Lausanne, Frame No.51)

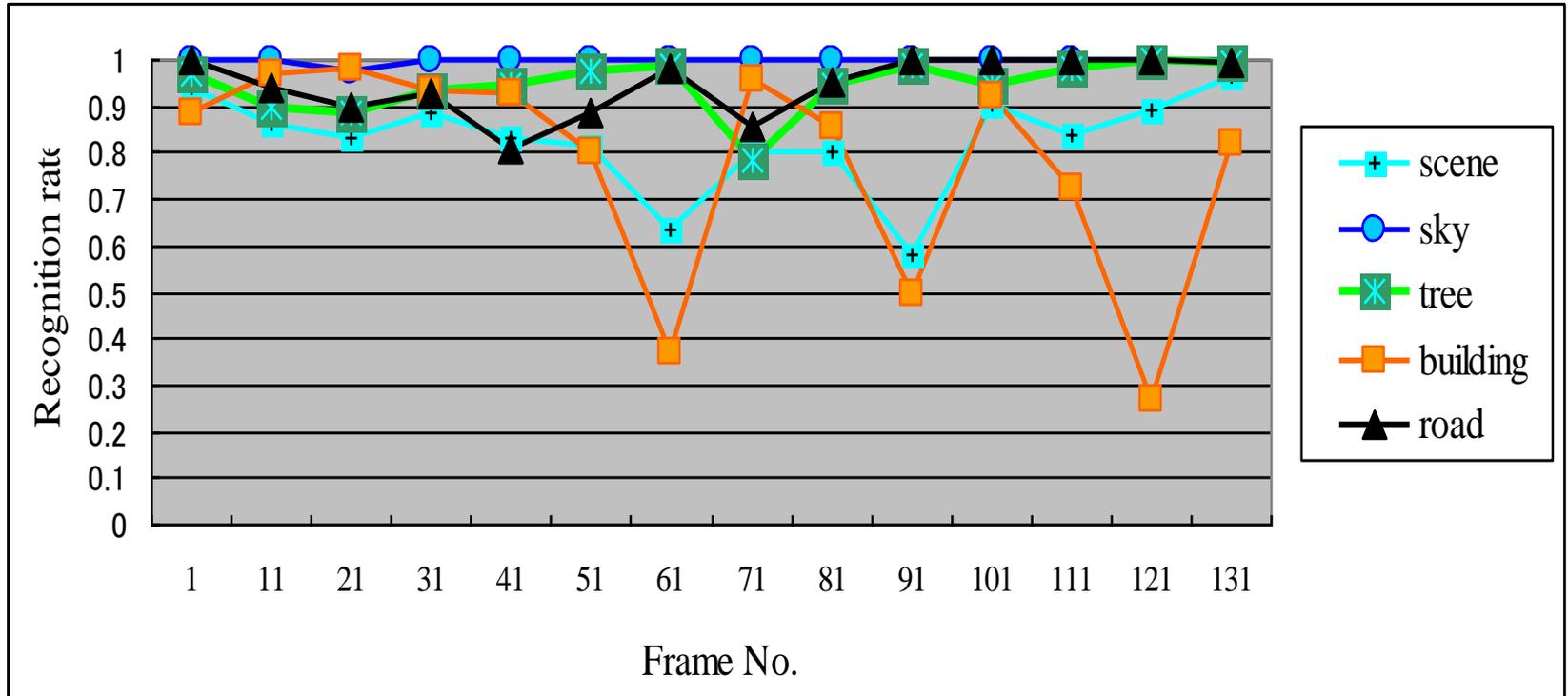


(e) Learned scene No.1
(060514-1748_FOREST)

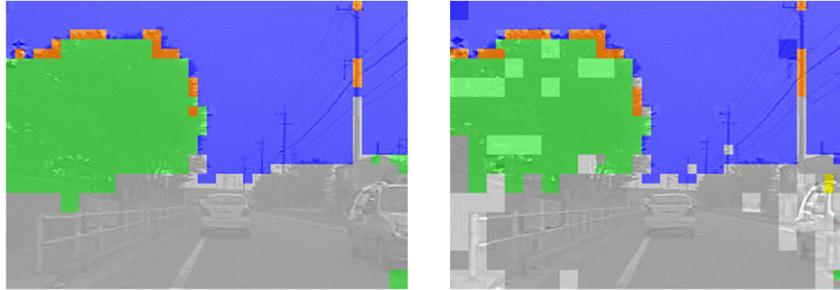


(f) Learned scene No.2
(070628-1744_APITA)

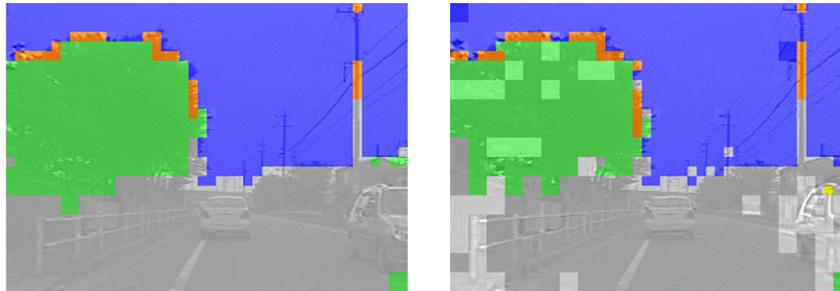
8章 シーンクラスと景観認識④



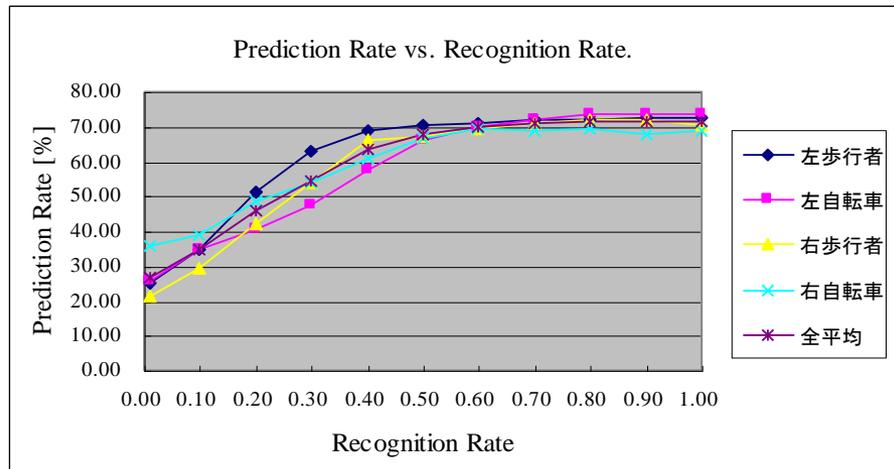
9章 予測と認識の相互作用



(a) 回帰係数を合成した場合

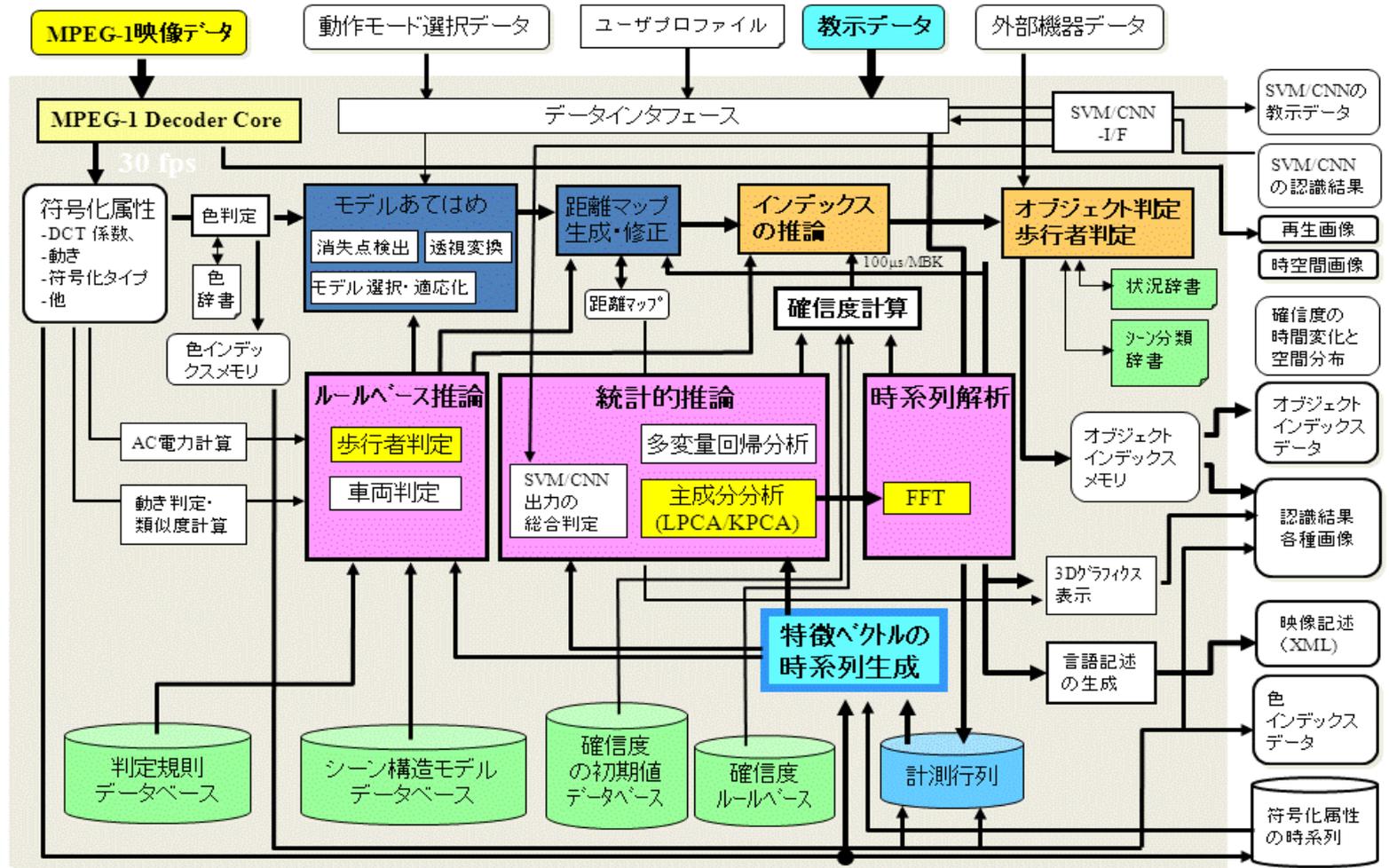


(b) 教示データを合成した場合



認識率と予測率

10章 カメラ移動に基づく認識①



10章 カメラ移動に基づく認識②

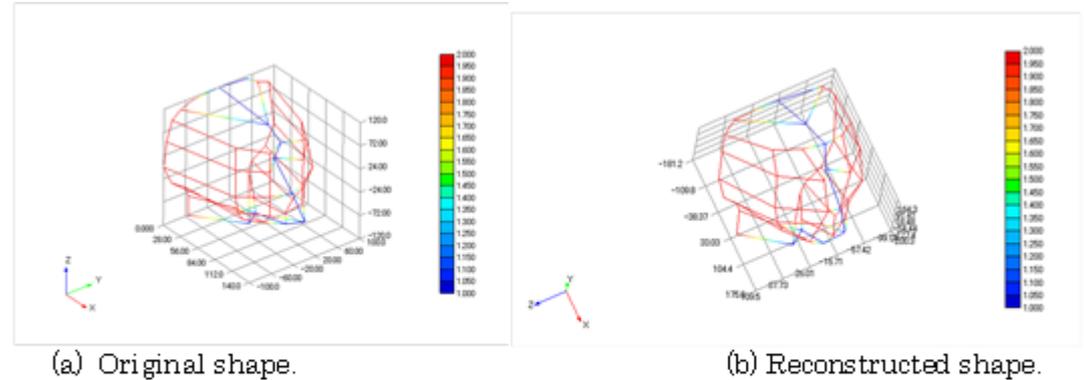
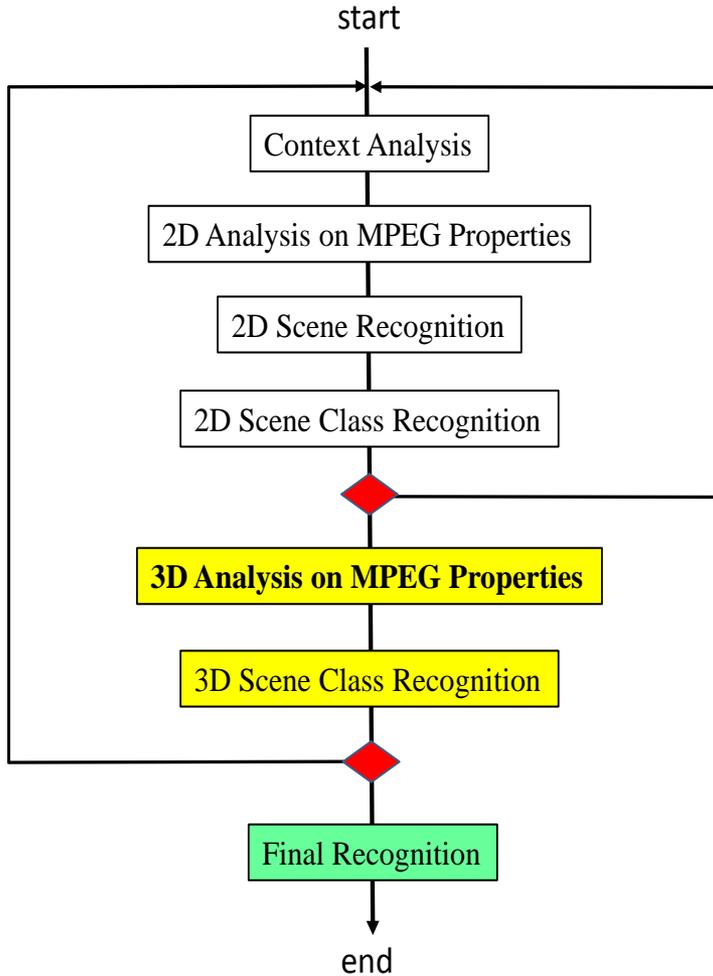
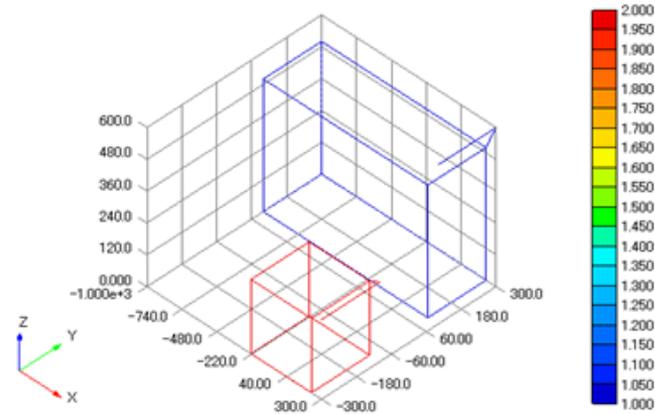


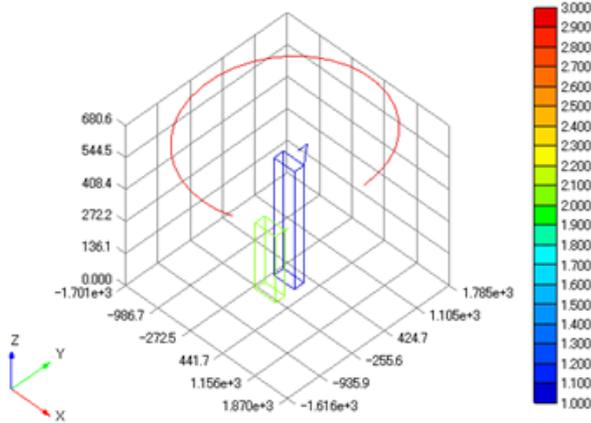
図 10.4 因子分解法の基本性能
Fig.10.4 Basic performance of factorization method.

10章 カメラ移動に基づく認識③

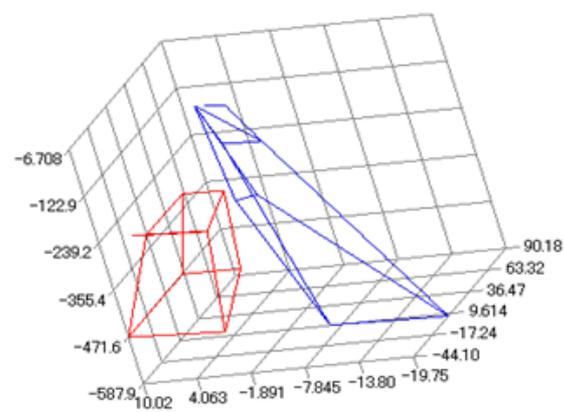


(1) Original shape.

カメラ軌道によって
3次元復元状況が
異なる

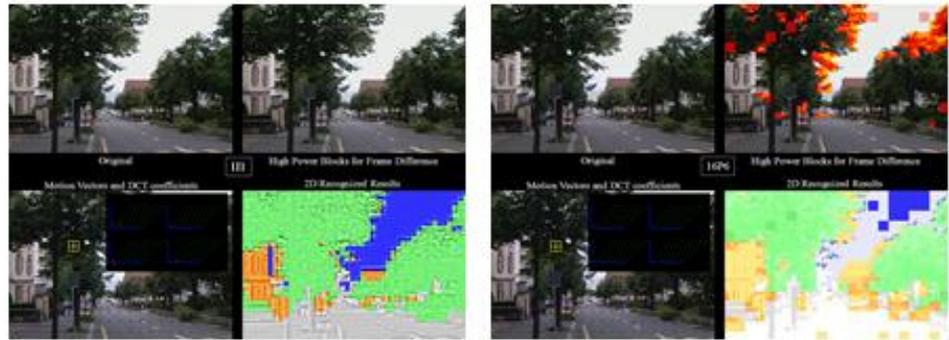


(b) Camera trajectory No.1



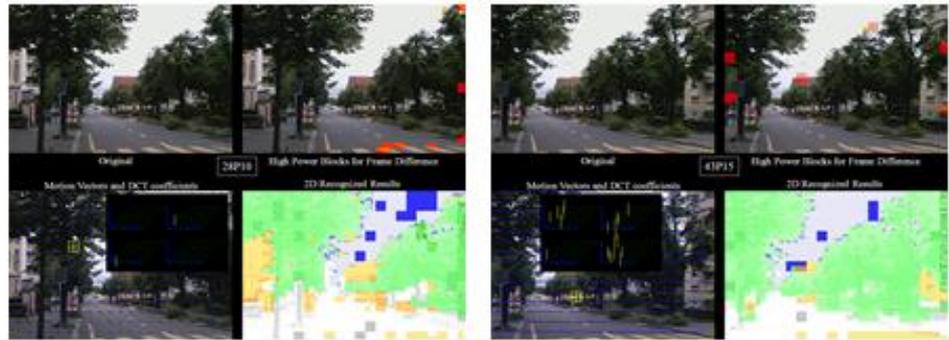
(c) Reconstructed shape.

10章 カメラ移動に基づく認識④



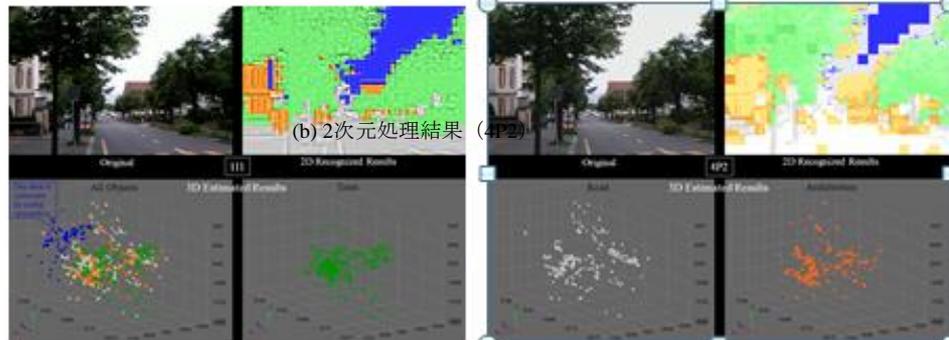
(a) 2次元処理結果【111】

(b) 2次元処理結果【16P16】



(c) 2次元処理結果【28P10】

(d) 2次元処理結果【43P15】



(e) 因子分解の結果【111】

(f) 因子分解の結果【4P2】

図 10.10 実験シーン No.1【Lausanne】

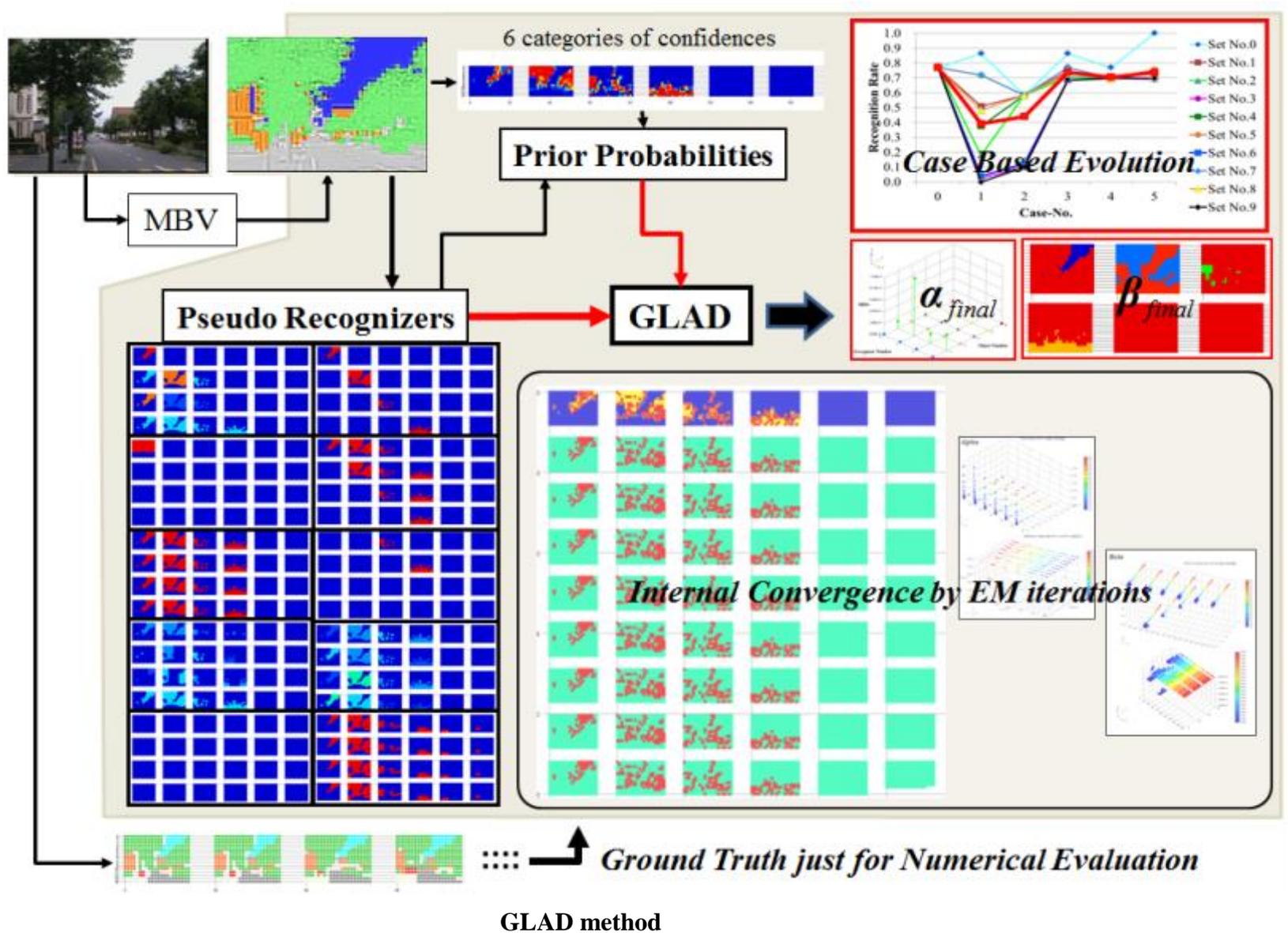
Fig. 11.10. Experimental Scene No.1 (Lausanne).

(a) 2次元処理結果 (111)

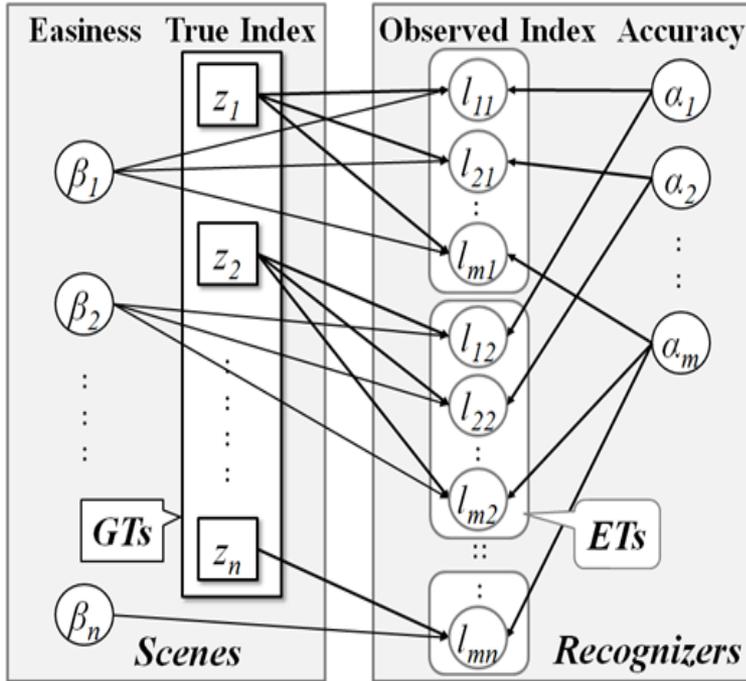
(b) 2次元処理結果 (4P2)

11章 複数の認識器の協調①

Experimental Flow of GT-less Evolution



11章 複数の認識器の協調②



(a) Graphical model.

Model:
$$p(l_{ik} = z_k | \alpha_i, \beta_k) = \frac{1}{1 + e^{-\alpha_i \beta_k}}$$

E-step:
$$p(z_k | l, \alpha, \beta) = p(z_k | l_k, \alpha, \beta_k) \propto p(z_k | \alpha, \beta_k) p(l_k | z_k, \alpha, \beta_k)$$

M-step:
$$Q(\alpha, \beta) = E \left[\ln \left\{ p(l, Z | \alpha, \beta) \right\} \right]$$

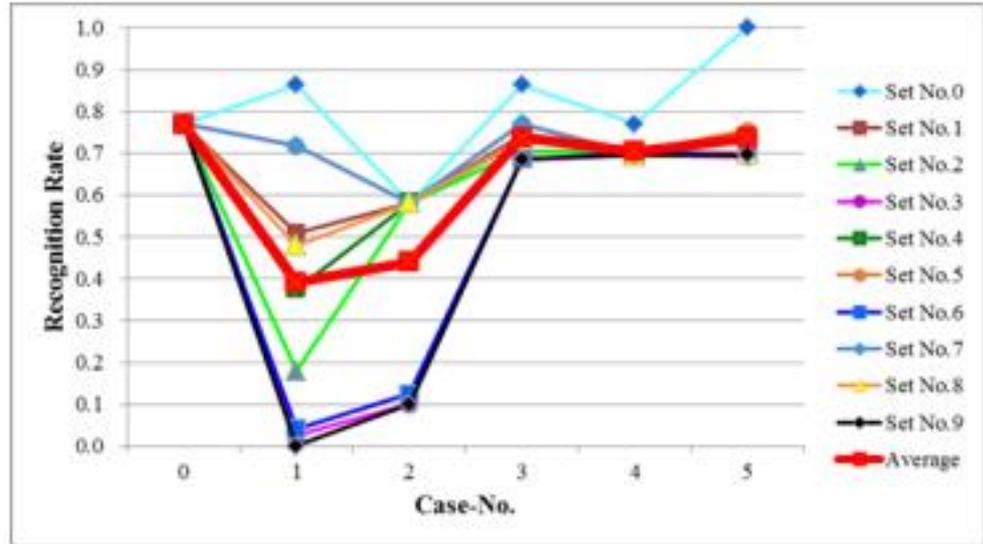
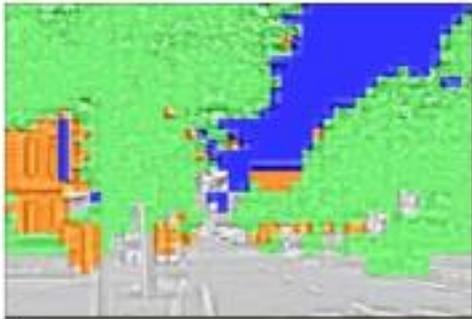
$$= E \left[\ln \left\{ \prod_k p(z_k) \prod_i p(l_{ik} | z_k, \alpha_i, \beta_k) \right\} \right]$$

$$= \sum_k E \left[\ln \{ p(z_k) \} \right] + \sum_{ik} E \left[\ln \{ p(l_{ik} | z_k, \alpha_i, \beta_k) \} \right]$$

Locally maximize $Q(\alpha, \beta) \Rightarrow (\alpha, \beta)$

(b) Probabilistic model and EM based evolution.

11章 複数の認識器の協調③



		Case-0	Case-1	Case-2	Case-3	Case-4	Case-5
Collective Processing		No	No	Yes	No	Yes	Yes
A Priori Probabilities		No	No	No	Yes	Yes	Yes (w=0.5)
Recognizers Set No.	0	0.769697	0.863636	0.581818	0.863636	0.769697	1.000000
	1	0.769697	0.509091	0.581818	0.743939	0.693939	0.693939
	2	0.769697	0.179545	0.581818	0.706061	0.700000	0.696970
	3	0.769697	0.025000	0.100000	0.684848	0.696970	0.690909
	4	0.769697	0.375000	0.581818	0.729545	0.700000	0.736364
	5	0.769697	0.718182	0.581818	0.769697	0.700000	0.754545
	6	0.769697	0.040909	0.124242	0.687121	0.700000	0.700000
	7	0.769697	0.718182	0.581818	0.769697	0.693939	0.696970
	8	0.769697	0.480303	0.581818	0.732576	0.693939	0.700000
	9	0.769697	0.000000	0.100000	0.684848	0.696970	0.696970
Average		0.769697	0.390985	0.439697	0.737197	0.704545	0.736667

11章 複数の認識器の協調④

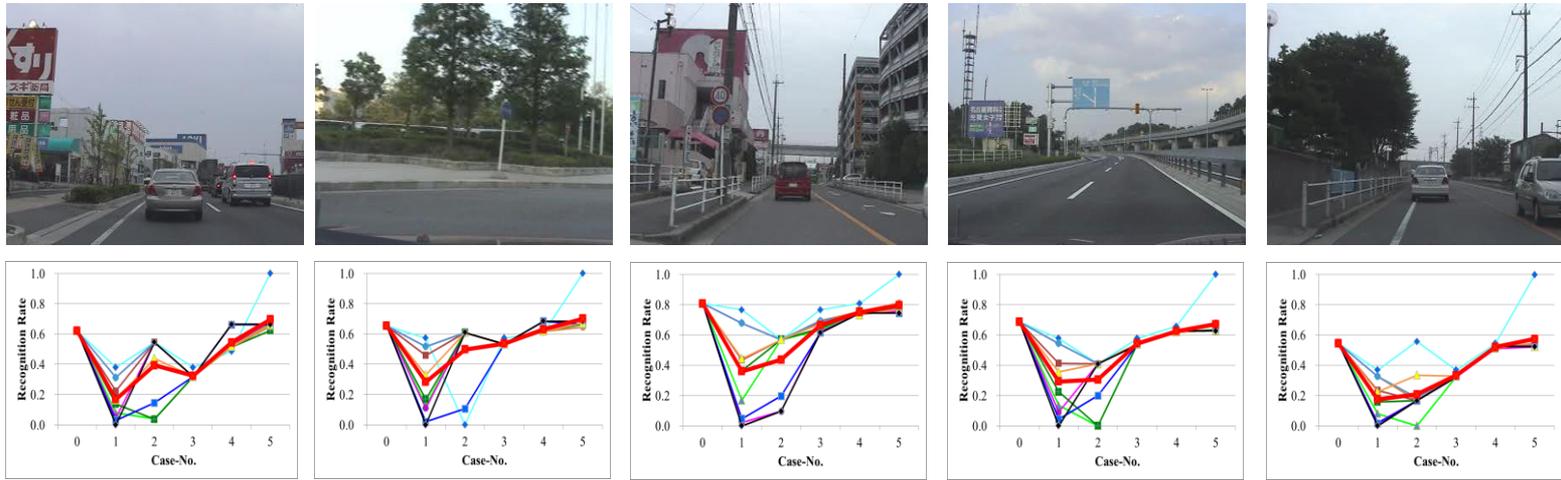
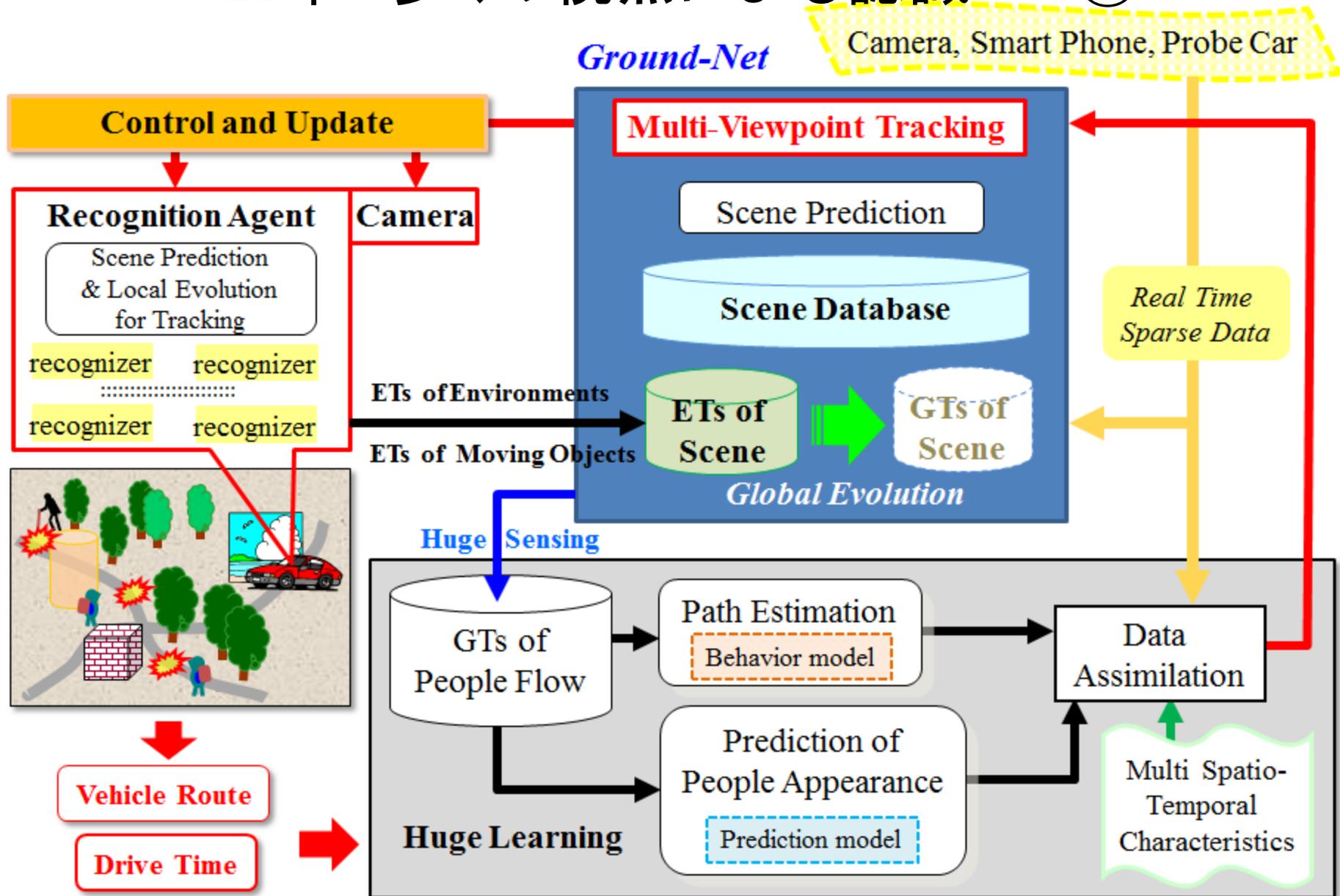


Figure 11.5: Evolution patterns extracted from other scenes.

12章 多くの視点による認識へ ①



12章 多くの視点による認識へ ②

