

テクニカルノート

# 類似スーパーピクセルに基づく画像内の物体認識

染谷 昌史<sup>1</sup> 丸山 稔<sup>1</sup> 宮尾 秀俊<sup>1,a)</sup>

受付日 2014年9月18日, 採録日 2015年5月9日

**概要:** 本研究ではエネルギー最小化に基づく画像内の物体認識・領域抽出手法において, 類似スーパーピクセルを利用した項を導入することにより性能向上を図る. このために, まず, 認識対象のスーパーピクセルと類似するスーパーピクセルをラベル付された学習用画像から取得することで, スーパーピクセル内のクラスを限定する手法を検討する. さらに色差が大きく異なる隣接スーパーピクセルが存在しうるクラスに関する認識率を向上させるための類似スーパーピクセルペアによるペナルティ項の導入手法について提案し, MSRC-21 を用いた実験により, 認識精度向上を示す.

**キーワード:** 領域抽出, 物体認識, スーパーピクセル, エネルギー最小化

## Object Recognition in Images based on Similarity of Superpixels

MASASHI SOMEYA<sup>1</sup> MINORU MARUYAMA<sup>1</sup> HIDETOSHI MIYAO<sup>1,a)</sup>

Received: September 18, 2014, Accepted: May 9, 2015

**Abstract:** We propose methods to improve energy-based image segmentation by introducing term which reflect similarity of superpixels. We consider restricting possible set of class labels for each superpixel depending on the similarity of superpixels. With ordinary methods, if two adjacent superpixels have different colors, they tend to be classified into different classes. It is not always true depending on the class they belong. In this paper we modify ordinary smoothness potential based on the distribution of color-pairs via similarity search of superpixels. We examine improvement of recognition accuracy using MSRC-21 dataset.

**Keywords:** segmentation, object recognition, superpixel, energy minimization

### 1. はじめに

画像内の物体認識・領域抽出は, 局所領域に対し, 認識物体が所属するクラスのラベル付けを行う処理であると考えられる. このために, 通常, ラベル付けを評価するエネルギー関数を構築し, 最小化する手法が用いられる [1]. 通常, エネルギー関数は局所領域単位で認識を行う Unary Potential と局所領域間の色の平滑性を考慮した Smoothness Potential より構成され, 使用する局所領域の単位にはピクセル, または, 複数のピクセルを1つの単位として扱うスーパーピクセルが用いられる [2], [3].

Shotton [4] は画像中のピクセルを局所領域とし, 多数の

フィルタの出力をテキストチャとして扱う認識手法を提案した.

より豊富な情報を含む基本単位として, ピクセルの集合であるスーパーピクセルを単位として認識を行う手法として Boix [5] の手法があげられる. スーパーピクセルを使用することでピクセルと比較し, 局所領域の範囲が増加するため, 局所領域から取得可能な特徴量が増加するという利点がある. また, 彼らは局所領域のみの情報では領域ラベルを誤認識することが多いという問題を解決するために, 画像にどのようなクラスが出現しているかを推測・制限する項をエネルギー関数に導入する手法を提案した. これにより, 局所領域の誤認識の低減を実現している.

Shotton [4], Boix [5] はともに, 物体を識別するための識別関数を学習し, これをエネルギー関数の一部に用いて認識を行う手法である. また, 物体の出現位置制約等も導入

<sup>1</sup> 信州大学工学部・情報工学科  
Department of Computer Science & Engineering, Faculty of Engineering, Shinshu University, Nagano 380-8553, Japan  
<sup>a)</sup> miyao@cs.shinshu-u.ac.jp

し、画像中の位置に基づくクラス制限を設けている。

識別関数を構築する手法のほかに、識別関数を構築しないで認識を行う手法も存在する。Tighe [6] は類似画像から画像内の出現クラスを限定し、類似スーパーピクセルを探索することで、物体認識・領域抽出を行った。これは、取得できた類似スーパーピクセルの個数に基づいてスーパーピクセルに対するクラスの尤度を計算するというものである。また、物体の共起性を導入し、隣接領域間でともに発生しやすいクラスが認識されやすくなるクラス制限を設けている。

本研究では、エネルギー最小化による認識・領域抽出方式に、類似スーパーピクセルによるクラス制限を行う手法を提案する。本手法では Tighe [6] の手法のように画像単位の類似スーパーピクセル検索を中心に認識されるクラスを決定するのではなく、識別関数と類似スーパーピクセルを複合してエネルギー関数の最小化によって認識されるクラスを決定する。さらに識別精度向上のために、類似スーパーピクセルペアによるペナルティ項導入を提案する。Shotton [4], Boix [5], Tighe [6] の Smoothness Potential では隣接領域で同一クラスの認識を期待しても、色差が顕著な場合には異なるクラスラベルが選択されやすいという問題が存在した。類似スーパーピクセルペアによるペナルティ項の導入はこの問題を低減するための手法である。具体的には類似画像から隣接関係のスーパーピクセルで、かつ、同一ラベルをとるようなスーパーピクセルペアに対し、類似検索を行い、このとき検索された類似1位の類似度をペナルティ項に利用する。認識対象画像のスーパーピクセルペアが異なるラベリングがされる時、類似度が高いものが存在すれば、異なるラベリングを行うことは不適切と判断し、大きなペナルティを働かせる。

以下では2章でエネルギー関数の説明を行い、3章で提案手法の説明を行う。4章で実験により提案手法の認識精度を示し、5章でまとめを述べる。

## 2. エネルギー関数による認識

物体認識、領域抽出はピクセル値が類似する領域を1つの領域とするスーパーピクセル単位 [2], [3] で認識を行う(図1)。物体認識・領域抽出はエネルギー関数の最小化を通して行う。エネルギー関数はスーパーピクセル単位で認識を行うため、最初にスーパーピクセルでクラス識別を行う識別器を構築する。物体認識時に作成した識別器の出力値に基づきエネルギー項を計算し、認識を行う。エネルギーの最小化には  $\alpha$ -expansion [7], [8] を使用する。

### 2.1 識別器の構築

領域のラベル付けのために、各クラスごとに識別器を構築する。このために、学習画像集合からピクセルごとに周辺の特徴量を取得し、k-means法を適用することで bag

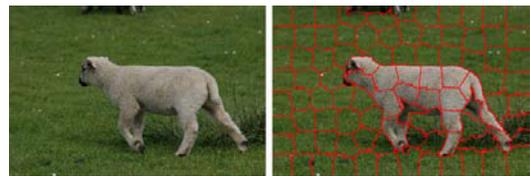


図1 画像のスーパーピクセル表現

Fig. 1 Superpixels representation of image.

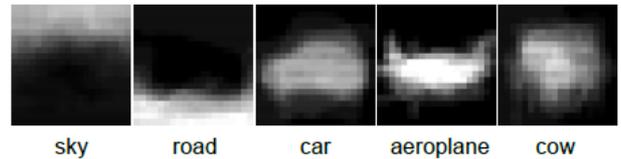


図2 位置情報に基づく確率分布

Fig. 2 Probability distribution based on location.

of visual words 型の特徴量を構築し、識別関数への入力とする。その元となる画像特徴量としては SIFT [9], Self Similarity Descriptor [10], RGB を用いた。スーパーピクセル内の特徴量を k-means クラスタリングによって量子化を行い、ヒストグラムを作成する。このヒストグラムをスーパーピクセルの特徴ベクトルとして使い、クラスごとに SVM [11] によって識別器を構築する。本研究においては SVM のカーネル関数として RBF を用いた。

### 2.2 エネルギー関数

スーパーピクセルに対してクラスラベルを割り当てるためにエネルギー関数  $E(\mathbf{x})$  を式 (1) で定義する。

$$E(\mathbf{x}) = \sum_{i \in V} \varphi_i(x_i) + \sum_{(i,j) \in \varepsilon} \psi_{ij}(x_i, x_j) + \sum_{i \in V} g_i(x_i) \quad (1)$$

ここに  $V$  はスーパーピクセルの集合、 $\varepsilon$  は隣接するスーパーピクセルの組の集合、 $\mathbf{x} = (x_1, \dots, x_{|V|})$  はスーパーピクセル列に割り与えられるクラスラベル列を表す。 $\varphi_i(x_i)$  はスーパーピクセル  $i$  にラベル  $x_i$  を与えたとき、コストが発生する Unary Potential である。 $\psi_{ij}(x_i, x_j)$  は隣接するスーパーピクセル  $i, j$  に異なるラベルが与えられたときのコストを表す Smoothness Potential である。 $g_i(x_i)$  は類似画像によるクラス制限項である。

Unary Potential  $\varphi_i(x_i)$  は式 (2) で定義する。

$$\varphi_i(x_i) = -K_i \left( \log \frac{1}{1 + \exp(-a(x_i) \text{score}(x_i; \mathbf{z}_i) - b(x_i))} + \log \text{loc}(\hat{\mathbf{i}}, x_i) \right) \quad (2)$$

$K_i$  はスーパーピクセル  $i$  内のピクセル数である。 $a(x_i)$ ,  $b(x_i)$  は学習により最適化されるパラメータである。 $\text{score}(x_i; \mathbf{z}_i)$  はスーパーピクセル  $i$  から生成される特徴ベクトル  $\mathbf{z}_i$  を、クラス  $x_i$  の識別器に入力したときの出力値である。 $\text{loc}(\hat{\mathbf{i}}, x_i)$  はクラス  $x_i$  がスーパーピクセルの中心座標  $\hat{\mathbf{i}}$  に出現する確率である。図2に  $\text{loc}$  を可視化し

た例を示す．本研究では， $score$  は RBF-カーネルによる SVM を用いた [11]．また，Unary Potential 中のパラメータ  $a$ ， $b$  は最急降下法により求めた．

Smoothness Potential  $\psi_{ij}(x_i, x_j)$  は式 (3) で定義する．

$$\psi_{ij}(x_i, x_j) = \left\{ a(x_i, x_j) K_{ij} \exp\left(\frac{-|c_i - c_j|^2}{b}\right) \right\} T[x_i \neq x_j]$$

$$a(x_i, x_j) = \begin{cases} \lambda & f[x_i][x_j] > 0 \\ \mu & f[x_i][x_j] = 0 \end{cases} \quad \frac{\mu}{2} < \lambda \leq \mu \quad (3)$$

$c_i$ ， $c_j$  はスーパーピクセル内の RGB 値の平均を表し， $K_{ij}$  は隣接するスーパーピクセル  $i$ ， $j$  間で，境界を共有するピクセル数を表す． $f$  は隣接スーパーピクセル間でクラスの共起性を表すパラメータであり， $f[x_i][x_j] > 0$  であれば，共起性ありを示し，共起性なしのパラメータ  $\mu$  より低い値を  $\lambda$  に設定する．これにより隣接領域間では共起関係にあるクラスが選択されやすくなることが期待できる．

類似画像によるクラス制限項  $g_i(x_i)$  は式 (4) で定義する．

$$g_i(x_i) = KT[x_i \notin X_g] \quad (4)$$

類似画像によるクラス制限項は，画像内に出現している物体のクラスを類似画像より推測し，推測されたクラスを選択的に出現しやすくする働きを持つ項である． $X_g$  は類似画像により許容されるクラスの集合である．許容クラス外のクラスがラベリングされるときペナルティ  $K$  が発生するため，許容クラス内のクラスがラベリングされやすくなる．なお，類似画像の検索は 2.1 節で述べた画像の bag of visual words 型表現に基づき，ヒストグラム間の重なり度合 (histogram intersection) によって類似度を定義して行った．

### 3. 類似スーパーピクセルを利用した認識

Tighe らの手法 [6] による類似画像検索に基づくクラス制限では，画像単位の制限となるため，すべてのスーパーピクセルに対して許容されるクラスが同一であった．そのため，スーパーピクセルが許容クラス内のクラスで誤認識される問題が存在する．そこでスーパーピクセルごとに許容クラスを変化させる類似スーパーピクセルによるクラス制限を提案する．また，既存の Smoothness Potential では隣接領域間では同一クラスとしての認識を期待しても，色差が大きいときは異なるラベル付けが行われやすい問題が存在した．この問題を低減させる類似スーパーピクセルペアによるペナルティ項の提案を行う．

#### 3.1 類似スーパーピクセルによるクラス制限

認識対象のスーパーピクセルに対して類似するスーパーピクセルを類似画像から取得し，スーパーピクセルごとに許容クラスを設定することを考える．概略を図 3 に示す．

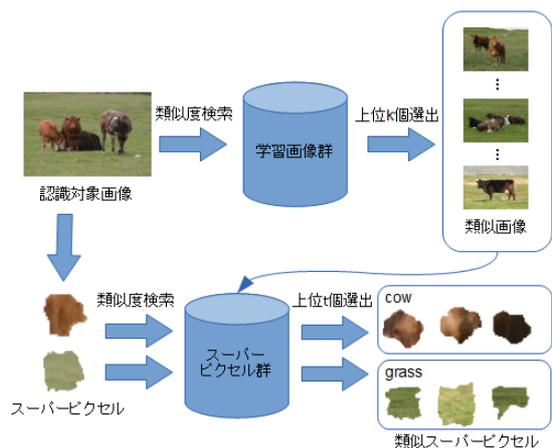


図 3 類似スーパーピクセルによるクラス制限

Fig. 3 Class constrain based on similarity of superpixels.

類似画像の場合と同様に，類似スーパーピクセルもスーパーピクセル内の特徴ベクトルに基づいて得られた bag of visual words に基づいて，ヒストグラムの重なり度合 (histogram intersection) によって類似度を算出する．

これにより，類似画像による画像単位のクラス制限よりきめ細かなクラス制限の実現が期待できる．類似スーパーピクセルによるクラス制限によりエネルギー関数は式 (5) のようになる．

$$E(\mathbf{x}) = \sum_{i \in V} \varphi_i(x_i) + \sum_{(i,j) \in \varepsilon} \psi_{ij}(x_i, x_j) + \sum_{i \in V} g_i(x_i) + \sum_{i \in V} s_i(x_i) \quad (5)$$

$s_i(x_i)$  が類似スーパーピクセルによるクラス制限項である． $s_i(x_i)$  は式 (6) で定義される．

$$s_i(x_i) = RT[x_i \notin X_i] \quad (6)$$

$X_i$  はスーパーピクセル  $i$  の類似スーパーピクセルにより許容されたクラスの集合である．許容クラス外のクラスがラベリングされるときペナルティ  $R$  が発生する．

#### 3.2 類似スーパーピクセルペアによるペナルティ項

スーパーピクセルのカテゴリ認識に色情報は大きな役割を果たし，多くの場合 Smoothness Potential の形式より，隣接するスーパーピクセル対の色差が顕著なとき，これらは異なるクラスと認識される傾向が強くなる．これは，複数の異なる色が発生しうるクラスの場合，必ずしも望ましいとはいえない．本研究ではこの問題を解決するためのペナルティ項の導入を考える．

認識対象のスーパーピクセルペアと類似するスーパーピクセルペアを類似画像から取得する．このとき，類似画像から検索されるスーパーピクセルペアは隣接関係のスーパーピクセルで，かつ，同一ラベルのものである．類似度が最大となるスーパーピクセルペアを取得し，このとき計算された類似度をペナルティ項に利用する．概略を図 4 に示す．

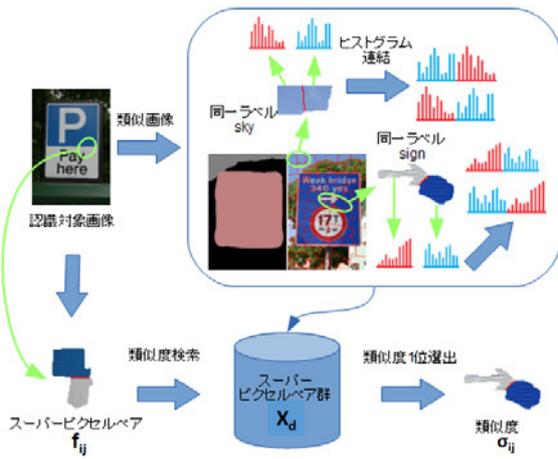


図 4 類似スーパーピクセルペアによるペナルティ項

Fig. 4 Penalty method based on similar superpixel pair.

す. 類似スーパーピクセルペアによるペナルティ項の導入により Smoothness Potential は式 (7) のように変更する.

$$\psi_{ij}(x_i, x_j) = \left\{ \begin{array}{l} a(x_i, x_j)K_{ij} \exp\left(\frac{-|c_i - c_j|^2}{b}\right) + d(f_{ij}) \\ T[x_i \neq x_j] \end{array} \right. \quad (7)$$

$$d(f_{ij}) = -w \log(1 - \sigma_{ij})$$

$$\sigma_{ij} = \max_{f_{pq} \in X_d} \text{sim}(f_{ij}, f_{pq})$$

$d(f_{ij})$  が類似スーパーピクセルペアによるペナルティ項である.  $f_{ij}$  はスーパーピクセル  $i, j$  のそれぞれのヒストグラムを連結したものである.  $X_d$  は類似画像のスーパーピクセルのうち, 隣接関係にあるスーパーピクセルペアで, かつ, 同一ラベルをとるスーパーピクセルのヒストグラムを連結したものの集合である. 類似度  $\text{sim}$  は, 2つのヒストグラムの重なり度合から算出した. ペナルティ項  $d(f_{ij})$  は隣接スーパーピクセルのラベル  $x_i, x_j$  が異なるラベルのとき作用するため, 同一ラベルを取るようなスーパーピクセルペア集合  $X_d$  内に  $f_{ij}$  と類似するものが存在すれば,  $x_i, x_j$  に異なるラベリングを行うことは不適切と判断し, ペナルティ  $d(f_{ij})$  は大きくなる. 逆に,  $X_d$  内に類似するものが存在しなければ, ペナルティ  $d(f_{ij})$  は小さくなる.

#### 4. 実験と考察

提案手法を導入した場合の実験を行った. 画像は MSRC-21 dataset [12] を利用し, 学習に 296 枚, 認識に 295 枚, クラス数は 21 とした. 類似画像の選出数  $k = 5$ , 類似スーパーピクセルの選出数  $t = 10$  とした. 最初に式 (1) のエネルギー関数で認識を行った場合を baseline として実験を行った. 次に類似スーパーピクセルによるクラス制限, 類似スーパーピクセルペアによるペナルティ項を導入して実験を行った. 2章で述べた識別器構築の特徴量抽出において, SIFT, Self Similarity Descriptor, RGB の k-means クラスタリングは  $k = 300$  として行った. 式 (3) の Smoothness

表 1 実験結果

Table 1 Experimental results.

	baseline	s	d	s+d
building	78.4	79.4	79.8	<b>81.0</b>
grass	90.7	92.2	90.8	<b>92.5</b>
tree	<b>88.0</b>	87.9	87.7	87.9
cow	62.9	63.9	64.3	<b>64.4</b>
sheep	68.1	69.2	69.1	<b>70.2</b>
sky	<b>91.6</b>	91.3	88.9	90.4
aeroplane	69.5	69.1	72.2	<b>73.4</b>
water	70.3	73.8	66.4	<b>76.9</b>
face	81.4	81.1	81.4	<b>81.5</b>
car	71.9	77.7	74.4	<b>82.0</b>
bicycle	76.2	<b>79.0</b>	78.4	<b>79.0</b>
flower	78.7	84.3	85.3	<b>86.6</b>
sign	21.7	22.5	18.9	<b>28.7</b>
bird	35.3	<b>36.8</b>	33.6	<b>36.8</b>
book	89.3	90.5	<b>93.9</b>	93.6
chair	56.8	60.6	<b>68.0</b>	67.9
road	<b>80.8</b>	80.5	80.4	80.4
cat	33.3	<b>33.7</b>	<b>33.7</b>	28.6
dog	61.2	61.5	61.4	<b>65.2</b>
body	54.4	53.0	<b>54.7</b>	49.7
boat	30.5	35.6	26.8	<b>38.9</b>
mean	66.2	67.8	67.2	<b>69.3</b>

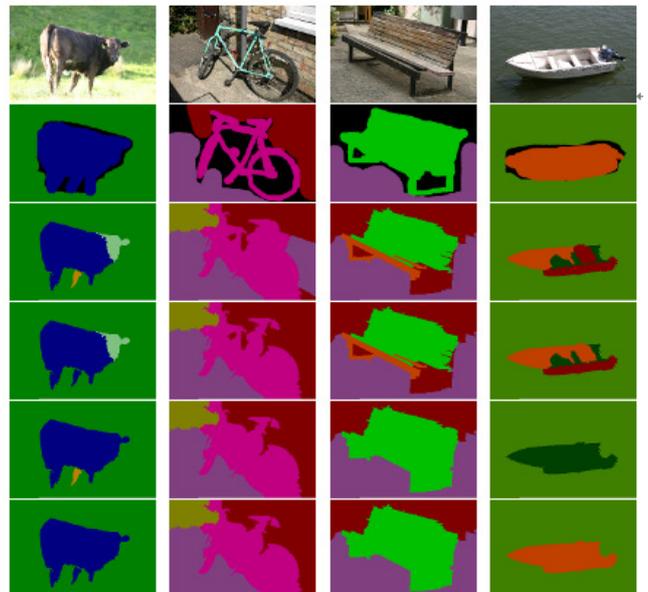


図 5 認識結果例

Fig. 5 Recognition result example.

Potential におけるパラメータ  $b, \lambda, \mu$ , 式 (4) のクラス制限項におけるパラメータ  $K$ , 式 (6) の類似スーパーピクセル項におけるパラメータ  $R$ , 式 (7) のペナルティ項における

パラメータ  $w$  は、いずれもいくつかの予備実験を行って決定し、実験では  $b = 2000$ ,  $\lambda = 140$ ,  $\mu = 270$ ,  $K = 1200$ ,  $R = 600$ ,  $w = 1500$  を用いた。実験結果を表 1, 図 5 に示す。表 1 より  $s$  が類似スーパーピクセルによるクラス制限,  $d$  が類似スーパーピクセルペアによるペナルティ項を示す。また、図 5 より上から入力画像, Ground Truth, baseline,  $s$ ,  $d$ ,  $s + d$  の画像を示す。制限を追加することで平均認識精度 (mean) が向上し、両制限を追加することで最も認識精度が向上した。しかし、精度の向上は小さなものとなった。これはクラスによっては認識精度が逆に下がってしまったものが存在するためである。類似スーパーピクセルは類似画像から取得するため、類似画像が正しく取得できていない場合、認識を期待しないクラスが許容され、認識されやすくなるためと考えられる。

## 5. まとめ

本研究では類似スーパーピクセルを利用し、画像内の物体認識精度を向上させる手法を提案した。提案手法の類似スーパーピクセルによるクラス制限, 類似スーパーピクセルペアによるペナルティ項のいずれも認識精度向上を実現した。類似スーパーピクセルは類似画像から取得するため、類似画像を正しく取得できればさらなる精度の向上が期待できると考えられる。

## 参考文献

- [1] Szeliski, R.: *Computer Vision: Algorithms and Applications*, Springer (2011).
- [2] Achanta, R. et al.: SLIC Superpixels, EPFL Technical Report 149300 (2010).
- [3] Achanta, R. et al.: SLIC Superpixels Compared to State-of-the-art Superpixel Methods, *IEEE Trans. PAMI*, Vol.34, No.11, pp.2274–2282 (2012).
- [4] Shotton, J. et al.: Textonboost for image understanding: multi-class object recognition and segmentation by jointly modeling texture, layout, and context, *International Journal of Computer Vision*, Vol.81, No.1, pp.2–23 (2009).
- [5] Boix, X., Gonfaus, J.M., van de Weijer, J., Bagdanov, A.D., Serrat, J. and Gonzalez, J.: Harmony Potentials, *International Journal of Computer Vision*, Vol.96, pp.83–102 (2012).
- [6] Tighe, J. and Lazebnik, S.: SuperParsing: Scalable Nonparametric Image Parsing with Superpixels, *International Journal of Computer Vision*, Vol.101, No.2, pp.329–349 (2013).
- [7] Boykov, Y., Veksler, O. and Zabih, R.: Fast Approximate Energy Minimization via Graph Cuts, *International Conference on Computer Vision*, Vol.1, pp.377–384 (1999).
- [8] Boykov, Y. and Kolmogorov, V.: An Experimental Comparison of Min-Cut/Max-Flow Algorithms for Energy Minimization in Vision, *IEEE Trans. PAMI*, Vol.26, No.9, pp.1124–1137 (2004).
- [9] Lowe, D.: Distinctive Image Features from Scale-invariant Keypoints, *International Journal of Computer Vision*, Vol.60, No.2, pp.91–110 (2004).

- [10] Shechtman, E. and Irani, M.: Matching Local Self-Similarities across Images and Videos, *Proc. Computer Vision and Pattern Recognition* (2007).
- [11] Joachims, T.: Making large-Scale SVM Learning Practical, *Advances in Kernel Methods - Support Vector Learning*, Scholkopf, B., Burges, C. and Smola, A. (Eds.), MIT-Press (1999).
- [12] Microsoft Research Cambridge Object Recognition Image Database available from (<http://research.microsoft.com/en-us/projects/objectclassrecognition/>).



染谷 昌史

2012年信州大学工学部情報工学科卒業。2014年信州大学大学院理工学系研究科修士課程修了。在学中、画像認識の研究に従事。



丸山 稔 (正会員)

1982年東京大学工学部計数工学科卒業。同年三菱電機(株)入社。1990年から1991年MIT人工知能研究所客員研究員。現在、信州大学工学部情報工学科教授。博士(工学)。コンピュータビジョン、機械学習等の研究に従事。電子情報通信学会, ACM, IEEE 各会員。



宮尾 秀俊 (正会員)

1989年長岡技術科学大学電子機器工学課程卒業。1991年長岡技術科学大学電気電子システム工学専攻修了。現在、信州大学工学部情報工学科准教授。博士(工学)。パターン認識・学習, HCIの研究に従事。電子情報通信学会, ACM, 日本データベース学会各会員。