

物体検出のための Relational HOG 特徴量とワイルドカードを用いたバイナリーのマスクング*

松島 千佳^{†a)} 山内 悠嗣^{†b)} 山下 隆義^{†,††c)} 藤吉 弘亘^{†d)}

Relational HOG Feature and Masking of Binary by Using Wild-Card for Object Detection*

Chika MATSUSHIMA^{†a)}, Yuji YAMAUCHI^{†b)}, Takayoshi YAMASHITA^{†,††c)},
and Hironobu FUJIYOSHI^{†d)}

あらまし 本論文では、物体検出に有効な HOG 特徴量のメモリ量を削減するために、Relational HOG 特徴量とワイルドカードを用いたバイナリーのマスクングを提案する。HOG は、人検出に有効な特徴量であるが、局所領域に着目しているため高次元な特徴量である。そこで、本論文では特徴量の情報量を削減するために、二つの局所領域から抽出した HOG 特徴量の大小関係によりバイナリーパターン化する R-HOG 特徴量を提案する。これにより、局所領域間の関係性を捉えたバイナリーパターンを作成することが可能であるが、R-HOG 特徴量には識別に不必要なバイナリーが含まれる。そこで、Real AdaBoost を用いて学習する際に、“0” と “1” の二つのバイナリーを許容するワイルドカード (*) を導入することにより、識別に悪影響を及ぼす一部のバイナリーを観測しないようにマスクングする。評価実験の結果より、提案手法はメモリ量を削減したにもかかわらず、従来法である HOG 特徴量の検出性能と同程度以上であることを確認した。

キーワード 物体検出, 2 値化, Histograms of Oriented Gradients, ワイルドカード, マスクング

1. ま え が き

近年、デジタルカメラや車載カメラの普及に伴い、撮影画像の画質改善や運転者のアシストを実現するために人検出の実用化の要望が高まり、FPGA 等によるハードウェア化の研究が行われている [1] ~ [5]。ハードウェア化の上で重要な点は、高い精度を維持したまま高速かつ低メモリで動作することである。

近年、提案されている人検出法は、画像局所特徴量と統計的学習手法の組合せによる手法 [6] ~ [10] が多く、特徴量としては局所領域のこう配に着目した特徴

量が多数提案されている [11], [12]。これらの特徴量は、物体の形状を捉えることが可能であるが、局所領域ごとに特徴量を求めるため非常に高次元となる。これは実装メモリが少ない小規模なハードウェアでは大きな問題となる。この問題を解決するためには、特徴量の情報量を削減する必要がある。情報量の削減は二つの利点がある。一つ目はメモリ量の削減、二つ目は共通の性質を表現する特徴量を類型化できることである。

情報量を削減する手法には、特徴量の空間を圧縮して特徴量数を削減する手法と、個々の特徴量の情報量自体を削減する二つの手法が考えられる。前者のアプローチは、特徴量の数を削減するベクトル量子化 [13] や特徴量の次元圧縮を行う主成分分析などがある。これらの手法は、元の情報量を維持しながら特徴量の次元数を削減することができる。しかし、人検出では膨大な数の検出ウィンドウを処理するため、計算コストが高いこれらの手法では非効率である。

後者のアプローチは、特徴量を低ビットに量子化する手法である。例えば、スカラ量子化は問題に応じてビット数を設定し、必要となる情報量を表現すること

[†] 中部大学大学院工学研究科, 春日井市
Department of Computer Science, Chubu University, 1200
Matsumoto, Kasugai-shi, 487-8501 Japan

^{††} オムロン株式会社, 草津市
OMRON Corporation, 2-2-1 Nishikusatsu, Kusatsu-shi,
525-0035 Japan

a) E-mail: matsu@vision.cs.chubu.ac.jp

b) E-mail: yuu@vision.cs.chubu.ac.jp

c) E-mail: takayosi@omm.ncl.omron.co.jp

d) E-mail: hf@cs.chubu.ac.jp

* 本論文は第 13 回画像の認識・理解シンポジウム推薦論文である。

が可能である．また，2 値化も情報量を削減するために有効な方法の一つであり，最も少ないビット数で表現することができるほか，ノイズに対する頑健性や扱いやすいという利点がある．その一手法としてしきい値処理による 2 値化が考えられ，簡単に処理することが可能であるため，計算コストが低いという利点がある．しかし，多数のサンプルに対して最適なしきい値を一意に決定することは難しい．このほかの 2 値化手法として大小関係を用いた 2 値化がある．Ojala らにより提案された Local Binary Pattern (LBP) [14] を用いた手法 [15], [16] や，これを発展させた手法 [17] は，ある二つの値の大小関係から 2 値化するため，しきい値が不要というメリットがある．また，しきい値処理による 2 値化と大小関係を用いた 2 値化では，一つのバイナリーが含む情報量が異なる．しきい値処理では値の大きさのみを表現するのに対して，大小関係を用いた 2 値化は，更に二つの値の関係性の情報を含むことができる．

本研究では，後者の情報量を削減する手法の一つである大小関係を用いた 2 値化に着目し，物体検出に有効な HOG 特徴量のメモリ量を削減することを目的とする．特徴量の情報量を削減しつつ高精度な人検出を実現するために，二つの局所領域から抽出した HOG 特徴量 [6] の大小関係を用いた Relational HOG 特徴量 (R-HOG) を提案する．R-HOG 特徴量は，二つの HOG 特徴量の大小関係を用いるためしきい値が不要であり，局所領域間の関係性を表現することが可能である．しかし，R-HOG 特徴量は複数のバイナリーを組み合わせた特徴量であるため，識別に不必要なバイナリーが含まれる．そこで，学習する際に“0”と“1”の二つのバイナリーを許容するワイルドカード(*)を導入する．これにより，Real AdaBoost [18] を用いて識別に悪影響を及ぼす一部のバイナリーを観測しないようにマスキングする．

2. HOG 特徴量と 2 値化

本章では，本研究で用いる HOG 特徴量と HOG 特徴量の情報量を削減する 2 値化について述べる．

2.1 HOG 特徴量

Dalal らによって提案された Histograms of Oriented Gradients 特徴量 (HOG) [6] は，局所領域における輝度のこう配方向をヒストグラム化した特徴量であり，物体の形状を捉えることが可能である．この特徴量は，近接画素のこう配を局所領域ごとにヒストグ

ラム化するため，照明の影響を受けにくく，局所的な幾何学変化に頑健という特徴がある．

HOG 特徴量を算出するためには，はじめに入力画像をあらかじめ定義したパッチサイズにリサイズし，各ピクセルの輝度 I からこう配強度 m とこう配方向 θ を次式より算出する．

$$m(x, y) = \sqrt{I_x(x, y)^2 + I_y(x, y)^2} \quad (1)$$

$$\theta(x, y) = \tan^{-1} \frac{I_y(x, y)}{I_x(x, y)} \quad (2)$$

$$\begin{cases} I_x(x, y) = I(x+1, y) - I(x-1, y) \\ I_y(x, y) = I(x, y+1) - I(x, y-1) \end{cases} \quad (3)$$

次に，算出したこう配強度 m とこう配方向 θ を用いて，式 (4) よりセル領域 c ($p \times p$ ピクセル) における量子化こう配方向 θ' のこう配強度の和を算出する．

$$v_c(\theta') = \sum_x \sum_y m(x, y) \delta[\theta', \theta(x, y)] \quad (4)$$

δ は Kronecker のデルタ関数であり，こう配方向 $\theta(x, y)$ が量子化こう配方向 θ' と同じヒストグラムの要素である場合は 1 を返す．式 (4) より算出した量子化こう配方向 θ' におけるこう配強度の和の集合を， N 方向のこう配方向ヒストグラム $V_c = \{v_c(1), v_c(2), \dots, v_c(N)\}$ とする．最後に，式 (5) を用いてブロック領域 ($q \times q$ セル) ごとに正規化し，特徴量を抽出する．

$$v'_c(n) = \frac{v_c(n)}{\sqrt{\left(\sum_{k=1}^{q \times q \times N} v_c(k)^2 \right) + \epsilon}} \quad (\epsilon = 1) \quad (5)$$

ブロック領域は，1 セルごとに正規化対象のセル領域が重なるように移動しながら正規化する．正規化後のこう配方向ヒストグラム V'_c は， $V'_c = \{v'_c(1), v'_c(2), \dots, v'_c(B \times N)\}$ となる．ここで， B はブロック領域に含まれるセル領域の数である．

2.2 Binarized HOG 特徴量

抽出した HOG 特徴量の情報量を削減するために，HOG 特徴量をしきい値処理により 2 値化した Binarized HOG 特徴量 (B-HOG) を得る．この特徴量は，セル領域における N 方向分のバイナリーを一つの特徴量 (バイナリーパターン) として観測することにより，セル領域におけるこう配方向の関係性を捉える．

B-HOG 特徴量は，セル領域における 8 方向のこ

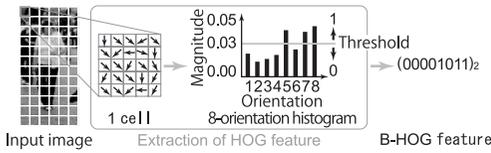


図1 B-HOG 特徴量の算出法
Fig.1 B-HOG feature calculation method.

う配方向ヒストグラム V'_c を式 (6) に示すようにしきい値処理することにより、2 値化した特徴量 $B_c^{BH} = \{b_c^{BH}(1), b_c^{BH}(2), \dots, b_c^{BH}(8)\}$ となる。文献 [6] では 9 方向の量子化こう配方向を用いているが、1 方向を 1 ビットで表現した場合、9 方向では 2 バイト必要とする。この量子化こう配方向を 8 方向にすることで、特徴量を 1 バイトで表現することが可能である。そのため、本研究では量子化こう配方向を 8 方向とし、メモリ量を 1/2 に削減する。

$$b_c^{BH}(n) = \begin{cases} 1 & \text{if } v'_c(n) \geq th \\ 0 & \text{otherwise} \end{cases} \quad (6)$$

ここで、 th はしきい値を表す。例えば、図 1 のような入力画像より得られる HOG 特徴量を 2 値化した場合、 $B_c^{BH} = (00001011)_2$ となる。

2.3 B-HOG 特徴量の利点と問題点

HOG 特徴量と B-HOG 特徴量には、特徴量に含まれる情報量に違いがある。式 (5) により求めた HOG 特徴量は、浮動小数点型 (8 バイト) で表現しなければならないが、B-HOG 特徴量は符号なし整数型 (1 バイト) で表現できる。そのため、B-HOG 特徴量は、HOG 特徴量よりメモリ量を 1/8 に削減することができる。しかし、人検出を行う環境が学習サンプルを収集した環境と異なる場合、その環境に合わせた最適なしきい値 th を求めなければならない問題がある。

3. 提案手法

本章では、提案する Relational HOG 特徴量 (R-HOG) と Real AdaBoost を用いた学習について述べる。提案手法による学習の流れを図 2 に示す。2.3 に示した 2 値化の利点を維持したまま問題点を解決するために、まず、二つの局所領域から抽出した HOG 特徴量を大小関係により 2 値化する。このとき、2 値化ではなく多値化することにより、多くの情報を表現することが可能であるが、最適な値の選定法や多値化手法が必要となる。また、多値化することにより必要な

メモリ量が増加するため、本論文では 2 値化のみを対象とする。

次に、統計的学習手法である Real AdaBoost を用いて学習する。R-HOG 特徴量は、局所領域間の関係性を捉えることができるが、識別に不必要なバイナリーが含まれる。そこで、学習する際に “0” と “1” の二つのバイナリーを許容するワイルドカード (*) を導入することにより、識別に悪影響を及ぼす一部のバイナリーを観測しないようにマスキングする。

3.1 Relational HOG 特徴量 (R-HOG)

R-HOG 特徴量は、二つの局所領域から得られる HOG 特徴量の大小関係を用いることにより 2 値化する。そのため、R-HOG 特徴量を求める際にはしきい値が必要ない。B-HOG 特徴量は、一つのバイナリーでこう配の強さしか表現できないのに対して、R-HOG 特徴量は二つの特徴量の関係性も表現することができる。また、提案手法である R-HOG 特徴量は HOG 特徴量の大小関係により 2 値化するため、HOG 特徴量の正規化処理を必要としない。正規化処理は、HOG 特徴量を計算する上で最も高い計算コストを占めているため、提案手法は処理コストを大幅に削減することが可能である。

3.1.1 二つのセル領域の HOG 特徴量を用いた 2 値化

R-HOG 特徴量は式 (7) に示すように、二つのセル領域から得られる 8 方向のこう配方向ヒストグラム V_{c_1} 、 V_{c_2} の大小関係により、2 値化した特徴量 $B_{c_1 c_2}^{RH} = \{b_{c_1 c_2}^{RH}(1), b_{c_1 c_2}^{RH}(2), \dots, b_{c_1 c_2}^{RH}(8)\}$ である。

$$b_{c_1 c_2}^{RH}(n) = \begin{cases} 1 & \text{if } v_{c_1}(n) \geq v_{c_2}(n) \\ 0 & \text{otherwise} \end{cases} \quad (7)$$

図 3 に示すように、二つのセル領域における特徴量の大小関係を用いることにより、局所領域間の関係性を捉えるバイナリーパターンを作成できる。このとき、R-HOG 特徴量は全てのセル領域の組合せから算出する。しかし、図 4 に示すように、抽出した特徴量が類似している場合、特徴量の大小関係が曖昧であるため、明確に 2 値で表現することが困難となる。

3.1.2 こう配方向のシフトを導入

R-HOG 特徴量の問題点を解決するために、図 4 の (b) や (c) に示すように、一方のセル領域から抽出される 8 方向のこう配方向ヒストグラム V_{c_2} のこう配方向を s 方向分シフトし ($s = 1, 2, \dots, 8$)、8 通りのこう配方向ヒストグラム $V_{c_2 s}$ を作成する。その後、

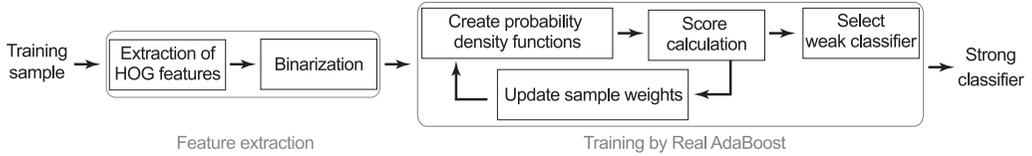


図 2 学習の概要

Fig. 2 Overview of training.

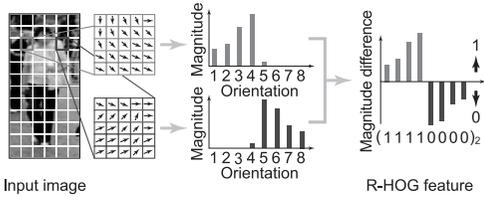


図 3 二つのセル領域の HOG 特徴量を用いた 2 値化
Fig. 3 Binarization using HOG features of two cell regions.

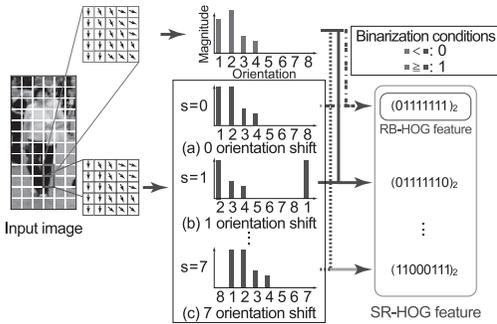


図 4 こう配方向のシフトを導入

Fig. 4 Introducing a shift in the orientation.

式 (7) と同様に、式 (8) よりこの配方向をシフトしていないこの配方向ヒストグラムと、 s 方向シフトしたこの配方向ヒストグラムのこの配強度の大小関係により 2 値化した特徴量 $B_{c_1 c_2 s}^{SRH}$ を算出する。

$$b_{c_1 c_2}^{SRH}(n, s) = \begin{cases} 1 & \text{if } v_{c_1}(n) \geq v_{c_2}(l) \\ 0 & \text{otherwise} \end{cases} \quad (8)$$

$$l = \text{mod}((n + s - 1), 8)$$

この配方向をシフトしたヒストグラムとの大小関係を算出することにより、抽出した特徴量が類似している場合でも、大小関係を明確に表現できる。本論文では、この配方向をシフトして R-HOG 特徴量を抽出したものを Shifted Relational HOG 特徴量 (SR-HOG) と呼ぶ。

3.2 ワイルドカード (*) によるバイナリーのマスクング

R-HOG 特徴量を抽出した後、統計的学習手法である Real AdaBoost を用いて学習する。識別に有効なセル領域の位置とバイナリーパターンを選択すると同時に、バイナリーパターンにワイルドカード “*” を導入し、識別に悪影響を及ぼす一部のバイナリーを観測しないようにマスクングする。これにより、検出精度の向上が期待できる。

3.2.1 “*” の導入

提案する R-HOG 特徴量は、図 5 (a) に示すような二つの HOG 特徴量を抽出した場合、図 5 (b) のようなバイナリーパターンが得られる。このとき、ポジティブサンプルの同じ位置のセル領域から抽出したこの配方向ヒストグラムでも、サンプルごとに二つのヒストグラムの大小関係が異なるため、バイナリーが反転してしまう可能性がある。そのため、2 値化した全てのバイナリーが識別する際に有効な情報を含んでいるとは限らず、中には誤識別の原因となるバイナリーも存在する。そこで、本論文では図 5 (c) に示すように、“0” と “1” の二つのバイナリーを許容する “*” を導入することにより、識別に悪影響を及ぼす一部のバイナリーを観測しないようにマスクングする。これにより、図 5 (d) のように、類似した複数のバイナリーパターンを同時に表現することが可能となる。“*” の数と何ビット目のバイナリーに適用するかは Real AdaBoost により選択する。

そこで、各セル領域の全組合せと、図 6 に示すようなバイナリーと “*” の全組合せを考慮したパターンを観測する弱識別器候補を用意する。例えば、一つの検出ウィンドウ内に存在するセル領域の数が 8×16 セルの場合、弱識別器候補数は、セル領域の全組合せ数 $8,128 (=_{128} C_2)$ と、バイナリーと “*” の全組合せ数 $6,561 (= 3^8)$ を掛け合わせた約 5000 万 $(= 8128 \times 6561)$ 個存在する。

また、SR-HOG 特徴量の場合は、一方のセル領域

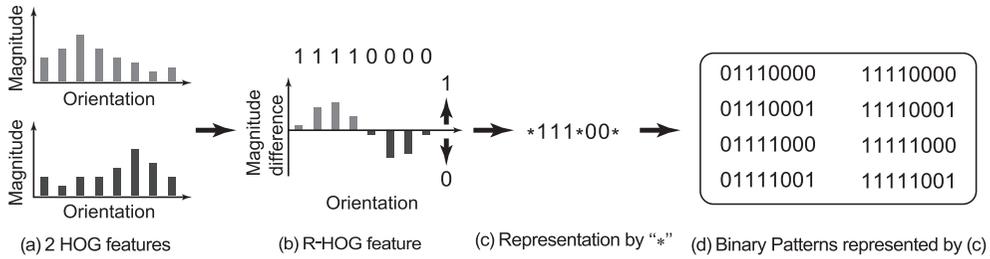


図 5 “*”を導入したバイナリーパターンによる表現の例

Fig. 5 Example of a representation by binary patterns to which “*” is introduced.

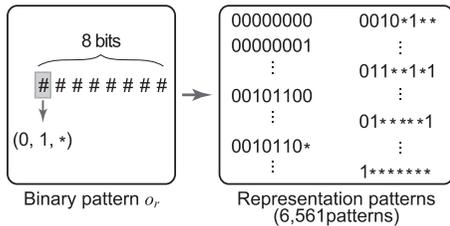


図 6 一組のセル領域における “*”を用いたパターン
Fig. 6 Patterns using wild-card “*” in a pair of cell regions.

から抽出した HOG 特徴量のこう配方向をシフトすることによりバイナリーパターンを求める。そのため、弱識別器候補数は R-HOG 特徴量の 8 倍である約 4 億 (= 8128 × 6561 × 8) となる。これらの弱識別器候補を用いて Real AdaBoost により識別器を学習する。

本論文では、R-HOG 特徴量や SR-HOG 特徴量を求める際に、各セル領域の全組合せを考慮している。そのため、弱識別器候補を限定する手法として、Sequential forward selection (SFS) を用いる手法 [19] や、Boosting により組み合わせる特徴量を限定する手法 [20] などを利用することにより、学習時間を短縮することが可能である。また、バイナリーパターンの全バイナリーが識別に有効であるとは限らないため、“*”の数を制限することにより学習時間を短縮することも考えられる。

3.2.2 学習手法

提案する学習手法は、3.2.1 にて述べた多数の弱識別器候補の中から Real AdaBoost により識別に有効な弱識別器を選択する。学習の流れを図 7 に示す。

はじめに、図 7 の 1、に示すように、前処理としてラベルの付いた I 枚の学習サンプル $(x_1, y_1), \dots, (x_I, y_I)$ を用意する。 x_i は画像であり、 y_i はクラスラベルである。検出対象のクラスは $y_i = +1$ 、非検出対象のクラ

```

1, Input: Labeled training samples  $I$ .
2, Initialization:
   Initialization of sample weights  $D_1(i)$ .
3, Training:
   For  $t = 1, \dots, T$  //Number of weak classifiers.
     //Combination number of cell regions.
     3.1, For  $r = 1, \dots, R$ 
       //Number of shifts.
       For  $s = 1, \dots, S$ 
         //Binary pattern with wild-card “*”.
         For  $o = 1, \dots, O$ 
           3.1.1, Calculate the probability density
             functions  $W_+, W_-$ .
           3.1.2, Calculate the score  $Z$ .
         End for
       End for
     End for
     3.2, Select weak classifier  $h(x)$ .
     3.3, Update sample weights  $D_t(i)$ .
   End for
4, Output: Strong classifier  $H(x)$ .
    
```

図 7 学習アルゴリズム
Fig. 7 Training algorithm.

スは $y_i = -1$ とする。また、図 7 の 2、に示すように各サンプルの重みを $D_t(i)$ とし、式 (9) により初期化する。

$$D_1(i) = 1/I \tag{9}$$

次に、図 7 の 3、に示すように、サンプルの重み付き頻度の算出から学習サンプルの重み更新までの処理を一定の弱識別器数 T 、または一定の識別率が得られるまで繰り返し、学習をする。まず、図 7 の 3.1.1、に示すように、弱識別器候補を観測するための各パターンにおけるサンプルの重み付き頻度を算出する。提案手法で用いるサンプルの重み付き頻度 W_+ 、

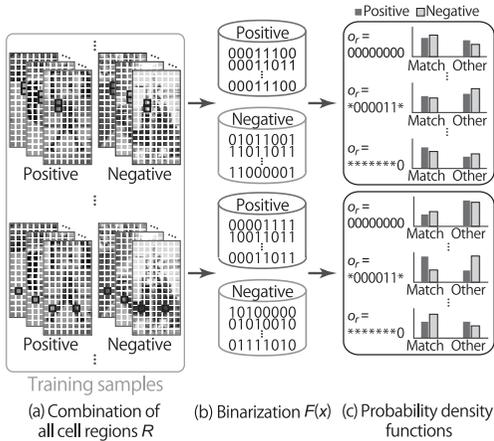


図 8 サンプルの重み付き頻度の算出
Fig. 8 Calculation of weighted histogram of training samples.

W_- は、図 8 に示すように、 N 個のセル領域の全組合せ $R = \{c_i, c_j\}_{i=1,2,\dots,(N-1),j=i+1,i+2,\dots,N}$ から抽出したバイナリーパターン $F(x)$ と、バイナリーと “*” を組み合わせたパターン $o_{r,s}$ が一致する頻度を表し、式 (10), (11) より算出する .

$$W_+ = \sum_{i:y_i=+1} D_t(i) \delta[F(x_i), o_{r,s}] \quad (10)$$

$$W_- = \sum_{i:y_i=-1} D_t(i) \delta[F(x_i), o_{r,s}] \quad (11)$$

ここで、+ は検出対象のクラス、- は非検出対象のクラスを表す . また、 $F(x)$ は入力画像 x からバイナリーパターンを観測する関数であり、 $o_{r,s}$ は図 6 に示すように、二つのセル領域の組合せ r において存在するバイナリーと “*” を組み合わせたパターンであり、 s はこう配方向のシフト数である . R-HOG 特徴量を用いる場合は、HOG 特徴量を 2 値化する際に、こう配方向をシフトをしないため $s = 1$ となる . SR-HOG 特徴量の場合は、こう配方向のシフトを考慮するため、 $s = \{1, 2, \dots, 8\}$ となる . δ は Kronecker のデルタ関数であり、入力する二つのパターンが一致する場合は 1 を返す .

サンプルの重み付き頻度 W_+ , W_- を算出後、図 7 の 3.1.2, に示すように式 (12) より分離度を表す評価値 Z を算出する .

$$Z_{o_{r,s}} = |W_+ - W_-| \quad (12)$$

評価値 Z は大きいほど、ポジティブサンプルとネガ

ティブサンプルの頻度の差が大きいことを意味する . そのため、図 7 の 3.2, に示すように、各セル領域の全組合せを考慮して “*” を用いたバイナリーパターンから、式 (13) より評価値 Z が最大となるパターン $o_{r,s}$ を観測する弱識別器候補を、 t ラウンド目において識別に有効な弱識別器 h_t として選択する .

$$h_t = \arg \max_{o_{r,s} \in (R \times S \times O)} Z_{o_{r,s}} \quad (13)$$

弱識別器を選択後、図 7 の 3.3, に示すように、誤識別した学習サンプルが次のラウンドで正しく識別されるために、式 (14) より学習サンプルの重みを更新する .

$$D_{t+1}(i) = D_t(i) \exp(-y_i h_t(x_i)) \quad (14)$$

$$h(x_i) = \begin{cases} \frac{1}{2} \ln \frac{W_+ + \epsilon}{W_- + \epsilon} & \text{if } F(x_i) = o_{r,s} \\ \frac{1}{2} \ln \frac{(1 - W_+) + \epsilon}{(1 - W_-) + \epsilon} & \text{otherwise} \end{cases} \quad (15)$$

このとき、選択された弱識別器のサンプルの重み付き頻度 W_+ , W_- を用いて、式 (15) より弱識別器の出力 $h(x)$ を算出する . ここで、 ϵ は分母が 0 になることを防ぐ係数 ($\epsilon = 1/I$) である . これにより、R-HOG 特徴量を用いた場合は、二つのセルの位置とバイナリーパターン、“*” の位置と数が選択される . また、SR-HOG 特徴量の場合は、こう配方向のシフト数も選択される .

最後に、図 7 の 4, に示すように、これまでの処理を一定の学習回数、または一定の識別率が得られるまで繰り返し、式 (16) に示す強識別器 $H(x)$ を得る .

$$H(x) = \text{sign} \left[\sum_{t=1}^T h_t(x) \right] \quad (16)$$

3.3 関連研究と提案手法の関係性

こまでに示した手法により特徴量を 2 値化することでメモリ量を削減し、“*” を用いて識別に不要なバイナリーをマスクングすることにより識別精度の低下を抑制することが期待できる . ここでは、他の物体検出手法との関係性について説明する .

まず、Ojala らにより提案された LBP 特徴 [14] は、近傍画素と注目画素の輝度値の大小関係を符号で表現することにより、テクスチャを表現することができる . また、複数の特徴量を組み合わせる手法として、文献 [19] で提案された Joint Haar-like 特徴や文献 [20], [21] で提案された共起特徴がある . これらの手法は、識別した結果をもとにして特徴量を組み合わせ

るため、識別結果を誤ってしまった場合や、検出対象に隠れが生じた場合は、組み合わせた特徴量にも悪影響を及ぼす。そこで、提案手法の一つである“*”を応用することで、識別に不要なバイナリーをマスクすることにより、識別制度の低下を抑制することが期待できる。

4. 評価実験

提案手法の有効性を評価するために、二つの評価実験をする。一つ目は R-HOG 特徴量と SR-HOG 特徴量の有効性を確認するために、HOG 特徴量や B-HOG 特徴量と精度を比較する。二つ目は、ワイルドカード(*)を用いて識別に悪影響を及ぼす一部のバイナリーを、マスクングすることの有効性を確認するために“*”の導入前と導入後の精度を比較する。

4.1 データセット

実験には、人のデータセットである INRIA person dataset [6] と、文献 [21] で使用されている車両のデータセットを用いる。各データセットの一部を図 9 に示す。INRIA person dataset は、学習用ポジティブサンプル 2,416 枚、ネガティブサンプル 12,180 枚であり、評価用ポジティブサンプル 1,126 枚、ネガティブサンプル 453 枚である。評価用ネガティブサンプルは、人が存在しない 1 枚の背景画像であり、検出ウィンドウをラスタスキャンして評価に用いる。また、車両のデータセットは、学習用ポジティブサンプル 710 枚、ネガティブサンプル 8,800 枚であり、評価用ポジティブ

サンプル 1,230 枚、ネガティブサンプル 3,880 枚である。

4.2 評価実験の概要

評価実験では、四つの手法を比較する。

- HOG 特徴量 (HOG)
- Binarized HOG 特徴量 (B-HOG)
- Relational HOG 特徴量 (R-HOG)
- R-HOG 特徴量 + こう配方向のシフト (SR-HOG)

各データセットのパラメータを表 1 に示す。評価には、Detection Error Tradeoff (DET) カーブを用いる。DET カーブは、横軸に False Positive Per Window (FPPW)、縦軸に Miss rate を表し、左下の原点に近いほど検出精度が高いことを示す。FPPW と Miss rate は式 (17)、(18) より算出される。

$$FPPW = \frac{(\# \text{ of false positive samples})}{(\# \text{ of windows})} \quad (17)$$

$$\text{Miss rate} = \frac{(\# \text{ of false negative samples})}{(\# \text{ of positive samples})} \times 100 \quad (18)$$

また、B-HOG 特徴量を算出する際に用いるしきい値 th は、予備実験にて決定し、人 (INRIA) の場合は $th = 0.09$ 、車両 (Vehicle) の場合は $th = 0.10$ を用いる。

4.3 実験 1: R-HOG 特徴量と SR-HOG 特徴量の有効性

実験 1 では、提案手法である R-HOG 特徴量と SR-HOG 特徴量の有効性を確認する。各データセットの実験結果の DET カーブを図 10 に示す。

まず、B-HOG 特徴量と R-HOG 特徴量を比較する。FPPW が 1.0×10^{-2} のときの検出率を比較すると、図 10 (a) より R-HOG 特徴量の検出率が B-HOG 特徴量より約 8.5% 向上することが分かる。また、図 10 (b) より人の場合と同様に、R-HOG 特徴量の検出率が約 1.7% 向上することが分かる。

次に、R-HOG 特徴量と SR-HOG 特徴量を比較する。FPPW が 1.0×10^{-2} のときの検出率を比較す

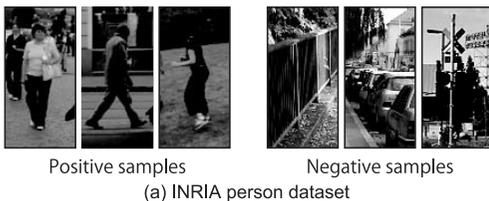


図 9 学習に用いるデータセットの一部
Fig. 9 Examples of training dataset.

表 1 データセットごとの実験に使用するパラメータ
Table 1 Parameters used in the experiment of each dataset.

| Dataset | Image size [pix.] | Cell size [pix.] | Block size [cell] | Orientation |
|--------------|----------------------|---------------------|----------------------|-------------|
| INRIA [6] | 64 × 128 | 8 × 8 | 2 × 2 | 8 |
| Vehicle [21] | 72 × 54 | 9 × 9 | 2 × 2 | 8 |

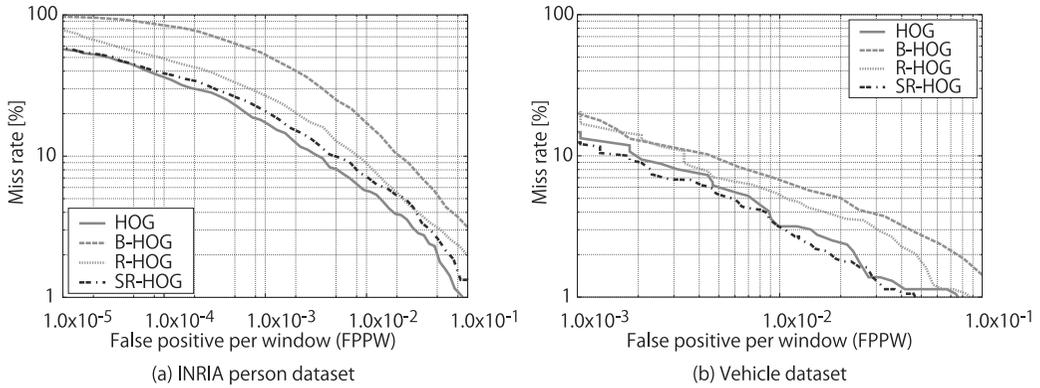


図 10 R-HOG 特徴量の有効性の実験結果
Fig. 10 Effectiveness of R-HOG feature.

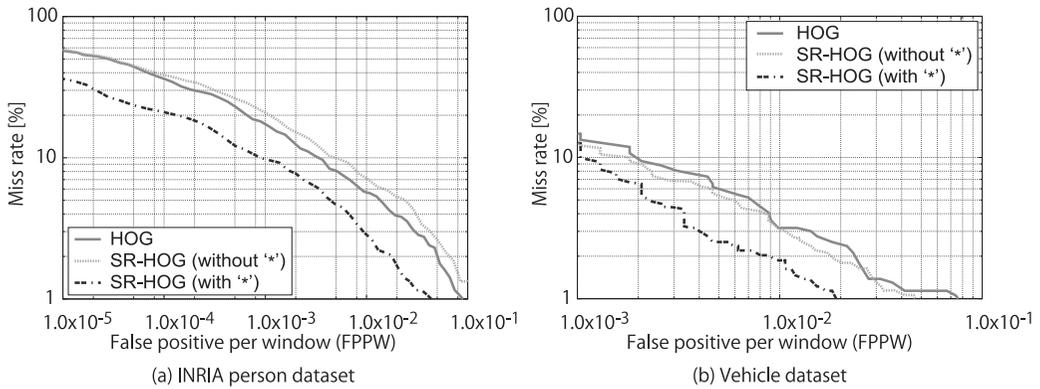


図 11 “*” の有効性の実験結果
Fig. 11 Effectiveness of the wild-card “*”.

ると、図 10 (a) より SR-HOG 特徴量の検出率が R-HOG 特徴量より約 1.7%向上することが分かる。また、図 10 (b) より人の場合と同様に SR-HOG 特徴量の検出率が約 2.1%向上することが分かる。これにより、しきい値処理を用いてバイナリーパターンを求める B-HOG 特徴量よりも、二つのセル領域から抽出した HOG 特徴量を比較してバイナリーパターンを求める R-HOG 特徴量の方が、セル領域間の関係を捉えることができるため検出率が向上することが分かる。更に、SR-HOG 特徴量のように一方のセル領域から抽出した HOG 特徴量のこの配方向をシフトしてバイナリーパターンを求めることにより、大小関係が明確になるため、R-HOG 特徴量よりも検出精度が向上する。

最後に、車両のデータセットに関しては HOG 特徴量と比較すると、ほぼ同程度の検出精度であることが分かる。

4.4 実験 2：“*” によるマスクングの有効性
実験 2 では、“*” を用いて識別に悪影響を及ぼす一部のバイナリーを、マスクングすることの有効性を確認する。

4.4.1 評価実験の結果

本実験では、実験 1 の結果より提案手法の中で最も検出精度がよい SR-HOG 特徴量に着目し、学習時に “*” を導入した手法と導入しない手法を比較する。各データセットの実験結果の DET カーブを図 11 に示す。

図 11 より、学習時に “*” を導入することにより精度が向上し、HOG 特徴量と同程度以上の検出精度であることが分かる。FPPW が 1.0×10^{-2} における検出率を比較すると、図 11 (a) より、“*” を導入した SR-HOG 特徴量が、“*” を導入しない手法より約 4.1%向上し、HOG 特徴量より約 1.5%向上することが分かる。また、図 11 (b) より、“*” を導入した

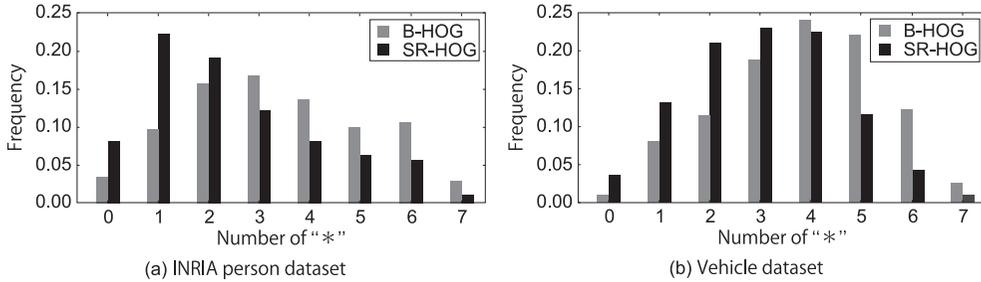


図 12 選択された "*" の数の割合
Fig.12 Proportion of number of selected "*" .

SR-HOG 特徴量が, "*" を導入しない手法と HOG 特徴量より約 1.2%向上することが分かる. これは, "*" を導入して反転しやすいバイナリーを観測しないことにより, 識別に悪影響を及ぼす一部のバイナリーをマスキングできているためであると考えられる.

4.4.2 識別に有効なバイナリー

識別に有効なバイナリー数を考察するために, 各データセットにおいて選択された "*" の数の割合を比較する. 比較結果を図 12 に示す.

図 12 より各特徴量の最も選択頻度が高い "*" の数は, 人の場合, B-HOG 特徴量が 3 個, SR-HOG 特徴量が 1 個である. また, 車両の場合, B-HOG 特徴量は 4 個, SR-HOG 特徴量は 3 個である. この結果より, B-HOG 特徴量よりも SR-HOG 特徴量の方が選択される "*" の数が少ないことが分かる. "*" は, "0" と "1" の二つのバイナリーを許容するため, しきい値との差や大小関係が曖昧な反転しやすいバイナリーが "*" として表現される. SR-HOG 特徴量は, 二つの局所領域から抽出した HOG 特徴量のこう配方向をシフトしながら大小関係を算出するため, 大小関係が明確なバイナリーが生成され, B-HOG 特徴量よりも "*" の数が減少したといえる.

4.4.3 2 値化した特徴量による識別結果の傾向

B-HOG 特徴量と SR-HOG 特徴量を用いた際の識別結果の傾向を考察する. B-HOG 特徴量を用いた場合は, 未検出してしまうのに対して, SR-HOG 特徴量を用いることにより, 検出可能な人のサンプルを図 13 に示す. 図 13 に示すように, HOG 特徴量をしきい値処理により 2 値化した B-HOG 特徴量は, 全サンプルに対して一定のしきい値を用いているため, 極端に明るい画像や暗い画像に対して未検出しやすい傾向がある. それに対して, SR-HOG 特徴量は, 二つのセル領域から抽出した HOG 特徴量の大小関係により 2



図 13 SR-HOG 特徴量を用いることで識別可能なサンプルの例
Fig.13 Examples of sample that can be classified by using the SR-HOG feature.

表 2 メモリ量の比較
Table 2 Comparison of memory.

| Feature | HOG | B-HOG | R-HOG | SR-HOG |
|-------------|------|-------|-------|--------|
| Memory [KB] | 3.91 | 0.50 | 0.49 | 0.98 |

値化しているため, 照明変動に対して頑健であり, 誤識別を抑制できた.

4.5 メモリ量の比較と処理時間

まず, 特徴量の情報量を削減することにより, どの程度のメモリ量が削減できるかを確認するため, HOG 特徴量と提案手法に必要なメモリ量を比較する. 弱識別器数 T を 500 個としたときの, 1 検出ウィンドウ当りのメモリ量を表 2 に示す. メモリ量 M の計算には次式を用いる.

$$M^{HOG} = T \times \text{Type}_d \tag{19}$$

$$M^{B-HOG} = T \times \text{Type}_{uc} + th \tag{20}$$

$$M^{R-HOG} = T \times \text{Type}_{uc} \tag{21}$$

$$M^{SR-HOG} = T \times (\text{Type}_{uc} + S) \tag{22}$$

ここで, Type はデータの型を表し, Type_d は浮動小数点型 (8 バイト), Type_{uc} は符号なし整数型 (1 バイト) である. また, th はしきい値であり浮動小数点型 (8 バイト), S はシフトするこう配方向数であり符号なし整数型 (1 バイト) である. 表 2 より,

SR-HOG 特徴量は明確な大小関係を得るためにこう配方向をシフトするため、B-HOG 特徴量や R-HOG 特徴量よりもメモリ量が増加する。しかし、HOG 特徴量と比較すると、HOG 特徴量を 2 値化することによりメモリ量を約 75.0%削減することができる。

次に、1 検出ウィンドウ (64×128 ピクセル) から、最終的な識別結果を出力するまでに必要な処理時間について述べる。このとき、各特徴量の識別器を学習する際に使用した弱識別器数は 500 個である。まず、1 検出ウィンドウから 500 個の特徴量を算出するための処理時間は、HOG 特徴量の場合、 5.39×10^{-7} ms、SR-HOG 特徴量の場合、 2.70×10^{-7} ms であった。また、抽出した特徴量を識別するための処理時間は、HOG 特徴量の場合、 1.51×10^{-14} ms、SR-HOG 特徴量の場合、 1.96×10^{-14} ms であった。この結果より、特徴量を識別するための処理時間は同程度であるが、SR-HOG 特徴量を抽出するための処理時間は、HOG 特徴量と比べて約 50.0%に削減することができた。これは、SR-HOG 特徴量を算出する際に、処理コストが高い HOG 特徴量の正規化処理をする必要がないためである。

5. む す び

本論文では、物体検出に有効な HOG 特徴量のメモリ量を削減するために、Relational HOG 特徴量とワイルドカードを用いたバイナリーのマスキングを提案した。本論文の貢献は 2 点ある。一つ目は、二つのセル領域から抽出した HOG 特徴量の大小関係により 2 値化した Relational HOG 特徴量である。この特徴量は、大小関係により 2 値化するためしきい値が不要となり、こう配強度の大きさと局所領域間の関係性を同時に捉えることが可能である。

二つ目は、“*”を導入することにより識別に悪影響を及ぼす一部のバイナリーを観測しないようにマスキングすることである。ポジティブサンプルの同じ位置のセル領域から抽出したこう配方向ヒストグラムでも、サンプルごとに二つのヒストグラムの大小関係が異なるため、バイナリーが反転してしまう可能性がある。そのため、2 値化した全てのバイナリーが識別する際に有効な情報を含んでいるとは限らず、中には誤識別の原因となるバイナリーも存在する。そこで、“0”と“1”の二つのバイナリーを許容するワイルドカード(*)を導入することにより、識別に悪影響を及ぼすバイナリーの反転を抑制し、特徴量の情報量を削減しても高

精度な検出を実現した。

本手法で用いている弱識別器候補は、二つのセル領域の全組合せに対してバイナリーと“*”を組み合わせたパターンを考慮しているため膨大な数となり、学習に多くの時間が必要となる。例えば、最も学習に時間がかかる SR-HOG 特徴量の場合では、約 1 か月である。そのため、今後は弱識別器候補を限定することにより、学習時間を削減する方法を検討する予定である。

文 献

- [1] V. Nair, P.O. Laprise, and J.J. Clark, “An FPGA-based people detection system,” *EURASIP J. Applied Signal Processing*, vol.2005, pp.1047–1061, 2005.
- [2] A. Ess, B. Leibe, K. Schindler, and L.V. Gool, “A mobile vision system for robust multi-person tracking,” *IEEE Conference on Computer Vision and Pattern Recognition*, pp.1–8, Anchorage, Alaska, June 2008.
- [3] K. Khattab, J. Dubois, and J. Miteran, “Cascade boosting-based object detection from high-level description to hardware implementation,” *EURASIP J. Embedded Systems*, vol.2009, pp.1–12, 2009.
- [4] R. Kadota, H. Sugano, M. Hiromoto, H. Ochi, R. Miyamoto, and Y. Nakamura, “Hardware architecture for HOG feature extraction,” *Intelligent Information Hiding and Multimedia Signal Processing*, pp.1330–1333, Kyoto, Japan, Sept. 2009.
- [5] 山中悠歩, 山崎俊彦, 相澤清晴, “FPGA による Histogram of Oriented Gradients アルゴリズムの高速化,” *情報処理学会創立 50 周年記念全国大会*, pp.1330–1333, no.2Y-7, March 2010.
- [6] N. Dalal and B. Triggs, “Histograms of oriented gradients for human detection,” *IEEE Conference on Computer Vision and Pattern Recognition*, pp.886–893, CA, USA, June 2005.
- [7] B. Wu and R. Nevatia, “Detection of multiple, partially occluded humans in a single image by bayesian combination of edgelet part detectors,” *IEEE International Conference on Computer Vision*, pp.90–97, Beijing, China, Oct. 2005.
- [8] B. Leibe, E. Seemann, and B. Schiele, “Pedestrian detection in crowded scenes,” *IEEE Conference on Computer Vision and Pattern Recognition*, pp.878–885, CA, USA, June 2005.
- [9] O. Tuzel, F. Porikli, and P. Meer, “Human detection via classification on riemannian manifolds,” *IEEE Conference on Computer Vision and Pattern Recognition*, pp.1–8, MN, USA, June 2007.
- [10] T. Watanabe, S. Ito, and K. Yokoi, “Co-occurrence histograms of oriented gradients for human detection,” *IPSN Trans. Computer Vision and Applications*, vol.2, pp.39–47, 2010.
- [11] C. Hou, H.Z. Ai, and S.H. Lao, “Multiview pedestrian detection based on vector boosting,” *Asian*

Conference on Computer Vision, pp.210–219, Tokyo, Japan, Nov. 2007.

- [12] A. Bosch, A. Zisserman, and X. Munoz, “Representing shape with a spatial pyramid kernel,” Proc. 6th ACM International Conference on Image and Video Retrieval, pp.401–408, Amsterdam, Netherlands, July 2007.
- [13] Y. Linde, A. Buzo, and R. Gray, “An algorithm for vector quantizer design,” IEEE Trans. Commun., vol.C-28, no.1, pp.84–95, 1980.
- [14] T. Ojala, M.P. Ainen, and D. Harwood, “A comparative study of texture measures with classification based on featured distributions,” Pattern Recognit., vol.29, pp.51–59, 1996.
- [15] X. Wang, T.X. Han, and S. Yan, “An HOG-LBP human detector with partial occlusion handling,” IEEE International Conference on Computer Vision, pp.1–8, Kyoto, Japan, Sept. 2009.
- [16] Y. Mu, S. Yan, Y. Liu, T. Huang, and B. Zhou, “Discriminative local binary patterns for human detection in personal album,” IEEE Conference on Computer Vision and Pattern Recognition, pp.1–8, Anchorage, Alaska, June 2008.
- [17] A. Hadid, M. Pietikainen, and T. Ahonen, “A discriminative feature space for detecting and recognizing faces,” IEEE Conference on Computer Vision and Pattern Recognition, pp.797–804, DC, USA, June 2004.
- [18] R.E. Schapire and Y. Singer, “Improved boosting algorithms using confidence-rated predictions,” Mach. Learn., vol.37, no.37, pp.297–336, 1999.
- [19] 三田雄志, 金子敏充, 堀 修, “顔検出に適した共起に基づく Joint Haar-like 特徴,” 信学論 (D), vol.J89-D, no.8, pp.1791–1801, Aug. 2006.
- [20] 山内悠嗣, 山下隆義, 藤吉弘亘, “Boosting に基づく特徴量の共起表現による人検出,” 信学論 (D), vol.J92-D, no.8, pp.1125–1134, Aug. 2009.
- [21] 高木雅成, 山内悠嗣, 藤吉弘亘, “Geometric Context を用いた特徴量間の共起による物体検出の高精度化,” 信学論 (D), vol.J93-D, no.8, pp.1429–1438, Aug. 2010.
(平成 22 年 10 月 15 日受付, 23 年 2 月 23 日再受付)



山内 悠嗣 (学生員)

2009 中部大学大学院博士前期課程了。現在同大学院博士後期課程に在籍。2010 独立行政法人日本学術振興会特別研究員。コンピュータビジョン, 動画像処理の研究に従事。



山下 隆義

2002 奈良先端科学技術大学院大学情報科学研究科修士課程了。同年オムロン(株)入社。顔及び人画像センシング技術の研究に従事。IEEE-CS, 情報処理学会各会員。



藤吉 弘亘 (正員)

1997 中部大学大学院博士後期課程了。1997~2000 米カーネギーメロン大学ロボティクス研究所 Postdoctoral Fellow。2000 中部大学講師を経て 2004 より同大准教授。2006 米カーネギーメロン大学ロボティクス研究所客員研究員。2010 中部大学教授。工博。計算機視覚, 動画像処理, パターン認識・理解の研究に従事。2005 年度ロボカップ研究賞, 2009 年度情報処理学会論文誌コンピュータビジョンとイメージメディア優秀論文賞, 2009 年度山下記念研究賞, 情報処理学会, 電気学会, IEEE 各会員。



松島 千佳 (学生員)

2009 中部大・工・情報工学卒。現在同大学院博士前期課程に在籍。コンピュータビジョン, 動画像処理の研究に従事。2010 年度 MIRU 学生賞受賞。