

人物認識のためのHOGをベースとした低次元特徴量の検討

高橋 秀和^{1,a)} 佐藤 真平^{1,b)} 高橋 篤司^{1,c)}

概要: 機械学習を用いた人物認識において、高い認識精度を実現できる特徴の一つにHOGがある。HOGは、画像の輝度の勾配方向と勾配強度の関係をヒストグラムで表現する。画像から抽出したHOGは高次元であり、人物認識システムのHW実装に伴うメモリ量の削減が課題となる。本研究では、従来のHOGで用いる輝度の勾配方向の中から、認識に有効な勾配方向を選んでヒストグラム化し、さらに二値化処理を施すことで得られる低次元特徴量を提案する。各特徴量の出現頻度が高まり、認識への貢献が増すと考えられる。また、特徴量の格納に伴うメモリ量は、従来のHOGを用いた場合と比較して約99%削減される。実験では、提案手法を用いてどの程度の認識精度が得られるか検証する。ヒストグラム化する勾配方向の選び方は複数パターン実験する。その結果、いくつかのパターンでは、従来のHOGを用いた場合と比較して最大4.3ポイントの認識精度向上が見られる。メモリ量削減と高い認識精度が同時に実現可能であると結論付けられる。

Considering low-dimension features based HOG for Human Recognition

HIDEKAZU TAKAHASHI^{1,a)} SHIMPEI SATO^{1,b)} ATSUSHI TAKAHASHI^{1,c)}

Abstract: In human recognition using machine learning, HOG is one of the features that can achieve high recognition accuracy. HOG expresses the relationship between the gradient directions and the gradient strength of the luminance of the image with histograms. The features extracted from the image has a high dimension, and reduction of the memory amount accompanying HW implementation of the person recognition system becomes a problem. In this research, we propose the lower dimensional features which are obtained by selecting gradient directions effective for recognition from the gradient directions used in the conventional HOG, converting it into histograms, and then binarizing it. The frequency of appearance of each feature, and the contribution to recognition increases. In addition, we can reduce the amount of memory associated with the storage of the features by 99% compared with the case of using the conventional HOG. In the experiment, we verify how much recognition accuracy can be obtained by using the proposed features. Selection of the gradient directions to form histograms was done in multiple pattern experiments. As a result, in some patterns, we obtained recognition accuracy equal to or higher than that obtained by using a conventional HOG. We conclude that high recognition accuracy and memory reduction can be realized at the same time.

1. はじめに

近年、自動車の自動運転等に向けて画像から自動的に人物を認識する技術の確立が求められている。人物認識は、コンピュータビジョンにおいて最も盛んに研究が行われて

いる分野の一つである。高い認識精度実現に向け、提案されたアルゴリズムは数多い。

機械学習を用いた物体認識において、物体の特徴を定量化した特徴量が認識精度を左右する。例えば、物体の輪郭という特徴を定量化して得られた特徴量は、背景と物体の境界を識別する際の助けとなりうる。物体認識を高精度に行うために、対象になる物体によって適した特徴量が異なる。

Histograms of Oriented Gradients(HOG)は、人物認識に有効な特徴として知られている [1]。HOGは画像の輝度

¹ 東京工業大学
Tokyo Institute of Technology, Meguro, Tokyo 152-8550,
Japan

a) h_takahashi@eda.ict.e.titech.ac.jp

b) satos@ict.e.titech.ac.jp

c) atsushi@ict.e.titech.ac.jp

の勾配方向と勾配強度の関係をヒストグラムで表現する。画像の局所領域に着目し、物体の形状を詳細に捉えることが可能である。HOG の特徴量算出では、画像をセルと呼ばれる領域で分割し、セルごとにヒストグラムを作成する。入力画像が大きくなるほど、HOG は高次元な特徴ベクトルを扱う。

人物認識のリアルタイム処理を低消費電力で実現するため、Field Programmable Gate Array(FPGA) を用いた HW 実装が数多く研究されている [2] [3]。高次元な特徴ベクトルを扱う HOG の特徴量算出に向けた HW 実装は、メモリ量の削減が課題となる。文献 [4] では HOG に二値化処理を施した特徴量が提案されている。ヒストグラムに閾値を設け、特徴量ベクトルを二値で表現することで大幅なメモリ量削減を実現できる。

HOG を用いた識別には多くのアルゴリズムが適用されている。特に文献 [2] [3] では二値化された HOG の特徴量とブースティング [5] を用いて、人物認識を FPGA 上で行う実装を提案している。ブースティングは単体では性能の低い識別器を複数用いることで強識別器を構成する。また、文献 [6] では、新たなブースティング手法が提案されている。特徴量の出現頻度を基に、入力画像から得られた特徴量とパターンマッチングを行うことで識別を行う。

本研究では、より低次元な特徴量と文献 [6] の手法を用いた人物認識システムを提案する。文献 [4] [6] でそれぞれ提案された特徴量、ブースティング手法を用いた場合、出現頻度の低い特徴量は認識への貢献が小さい。各特徴量の認識への貢献度を高め、認識精度を改善することを目的とする。提案する低次元特徴量は、従来の HOG で用いる輝度の勾配方向の中から、認識に有効な勾配方向をいくつか選んでヒストグラム化し、さらに二値化を施すことで得られる。提案する特徴量を用いた場合、それぞれの出現頻度が高くなり、認識への貢献度が増すと考えられる。また、特徴量の格納に伴うメモリ量は、従来の HOG の表現精度に倍精度浮動小数点数を用いた場合と比較して約 99%削減される。

実験では、提案手法を用いてどの程度の認識精度が得られるか検証する。ヒストグラム化する勾配方向の選び方は複数パターン実験した。その結果、いくつかのパターンでは、従来の HOG を用いた場合と比較して最大 4.3 ポイントの認識精度向上が見られた。メモリ量削減と高い認識精度が同時に実現可能であると結論付けられる。

2. Histograms of Oriented Gradients

HOG は、画像の輝度の勾配方向と勾配強度の関係をヒストグラムで表現する。画像の局所領域に着目し、物体の形状を詳細に捉えることができる。本章では、本研究で必要とする HOG の算出手順の中から重要なポイントに絞って説明する。

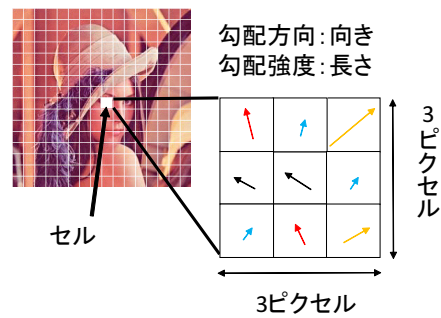


図 1 勾配方向と勾配強度

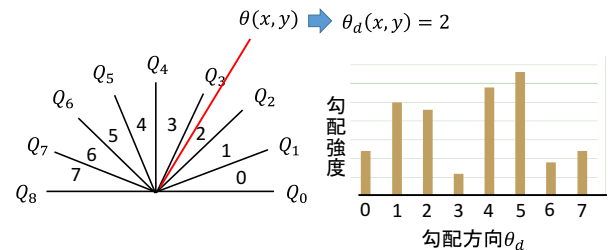


図 2 ヒストグラム作成

2.1 勾配方向と勾配強度

上下左右で隣り合うピクセルの輝度 L から着目ピクセル (x, y) の勾配強度 $m(x, y)$ と勾配方向 $\theta(x, y)$ を求める。

$$m(x, y) = \sqrt{\Delta L_x^2 + \Delta L_y^2} \quad (1)$$

$$\theta(x, y) = \tan^{-1} \frac{\Delta L_y}{\Delta L_x} \quad (2)$$

ピクセル (x, y) における輝度を $L(x, y)$ と表す。次に ΔL_x と ΔL_y を以下のように定義する。

$$\Delta L_x = L(x + 1, y) - L(x - 1, y) \quad (3)$$

$$\Delta L_y = L(x, y + 1) - L(x, y - 1) \quad (4)$$

2.2 ヒストグラム作成

ヒストグラムは、入力画像を $c \times c$ ピクセルずつ分割した局所領域 (セル) ごとに作成される。ヒストグラム作成にあたって、勾配方向 $\theta(x, y)$ を量子化する。図 2 を参考に、例えば $Q_2 \leq \theta(x, y) < Q_3$ である場合、 θ は 2 にラベルづけられる。図 2 の例では、 θ を 0 から 7 の 8 段階に量子化している。量子化後の勾配方向 $\theta(x, y)$ を $\theta_d(x, y)$ とする。ヒストグラムの横軸は量子化後の勾配方向 θ_d であり、縦軸は各ピクセルの勾配方向 θ_d に対応する勾配強度 m の総和である。図 1 は、 $c = 3$ の場合の例を示す。

3. 二値化 HOG を用いた識別器

2.2 節で作成したヒストグラムに対して、二値化処理を施すことで情報量を削減することが提案されている [4]. この章では HOG の二値化と、二値化 HOG を用いた識別器の構成を説明する. 今回はブースティングと呼ばれる学習法を用いて識別器を構成する.

3.1 二値化 HOG : 2 値勾配強度コード

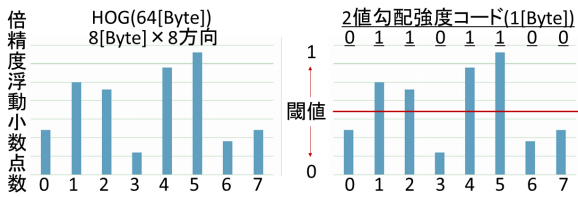


図 3 2 値勾配強度コードの作成

正規化後のヒストグラムに閾値を設け、勾配強度と閾値を比較して二値化を行う. 図 3 の例では、ヒストグラムが 8 ビット列 $(01101100)_2$ で表されることを示している. 勾配強度を二値化しているため、以後このビット列を (8 方向) 2 値勾配強度コードと呼ぶ. つまり各ヒストグラムは、 $(0)_{10}$ から $(255)_{10}$ の範囲で表現できる. 二値化処理を施すことによって、ヒストグラムを表す情報は、 $1/64$ に削減できる.

3.2 ブースティング

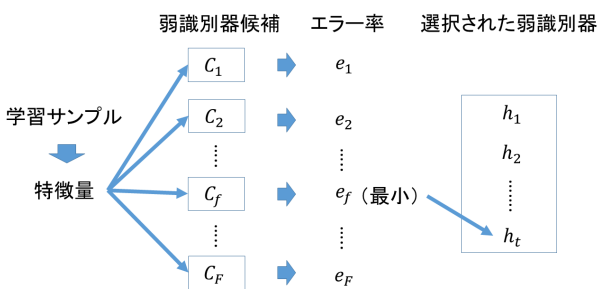


図 4 t ラウンド目における弱識別器選定の様子

ブースティングは単体では識別性能の低い識別器 (弱識別器) を複数連結させて強識別器を構築する手法である [5] [6]. 事前に弱識別器になりうる識別器 (弱識別器候補) を定義しておく必要がある. 学習時、人物が写っている画像サンプル (ポジティブサンプル) と写っていない画像サンプル (ネガティブサンプル) に対し、定義した識別器を適用する. 最も性能のよい識別器を実際に用いる識別器として選択する. その後、画像の重みを変化させて同様の操作を行う. 一定回数 (ラウンド) 後、選択された複数の識別器の出力を線形和として全体の出力とする.

3.3 2 値勾配強度コードとブースティング

最も識別に有効なセルはどれかを考えることで、3.2 節で述べた弱識別器を定義できる. この弱識別器の選定を学習段階で行う.

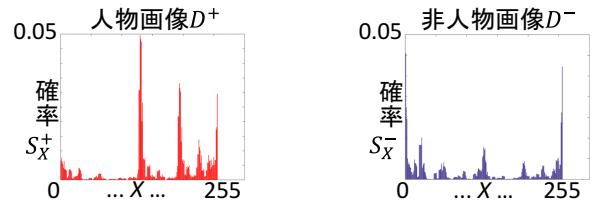


図 5 全サンプルから得られた 2 値勾配強度コードの出現頻度分布

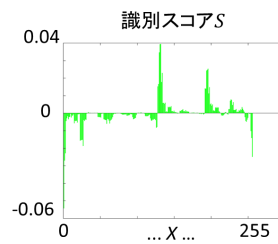


図 6 全サンプルから得られた出現頻度分布差

3.3.1 識別の原理

ポジティブサンプルから得られる傾向の高いコード、ネガティブサンプルから得られる傾向の高いコードは識別に有効である. 図 5 は、あるセルに着目した場合に全サンプルから得られるコードの出現頻度分布である. また、図 6 は、図 5 の両出現頻度分布の差を表す (識別スコア). この出現頻度分布は学習段階で作成する. 弱識別器はそれぞれの着目セルに関する出現頻度分布情報を保持しており、入力 X に対して、識別スコアを出力とする. 出力値が正であれば、その識別器は入力画像をより人物画像らしいと判断したといえる.

3.3.2 2 値勾配強度コードによるブースティング

(1) 頻度分布の作成

- 入力画像のセル数 : $C_w \times C_h$
- 学習画像の重み
 - (M 枚の) ポジティブサンプルの重み $W^+ = \{w_1^+, w_1^+, \dots, w_m^+, \dots, w_M^+\}$
 - (N 枚の) ネガティブサンプルの重み $W^- = \{w_1^-, w_1^-, \dots, w_n^-, \dots, w_N^-\}$
- 各学習画像のセル (i, j) から得る 2 値勾配強度コード
 - ポジティブサンプル $X^+ = \{X_1^+(i, j), \dots, X_m^+(i, j), \dots, X_M^+(i, j)\}$
 - ネガティブサンプル $X^- = \{X_1^-(i, j), \dots, X_n^-(i, j), \dots, X_N^-(i, j)\}$
- 全学習画像のセル (i, j) における出現頻度分布

- ポジティブサンプル : $D^+(i, j, X^+)$
- ネガティブサンプル : $D^-(i, j, X^-)$

```

1: procedure CREATE POSSIBILITY DISTRIBUTION
2:   for  $i = 1$  to  $C_h$  do
3:     for  $j = 1$  to  $C_w$  do
4:       for  $m = 1$  to  $M$  do
5:          $D^+(i, j, X^+(i, j, m)) += w_m^+$ 
6:       end for
7:       for  $n = 1$  to  $N$  do
8:          $D^-(i, j, X^-(i, j, n)) += w_n^-$ 
9:       end for
10:    end for
11:  end for
12:  Normalize  $D^+$  and  $D^-$ 
13: end procedure

```

(2) 弱識別器の選定

以下に定義する Z が, 図 4 におけるエラー率にあたる.

$$Z = 1 - \sum_i \sum_j \sum_x \sqrt{D^+(i, j, X)D^-(i, j, X)} \quad (5)$$

Z の値が小さければ小さいほど D^+, D^- の類似度は小さく, Z が最も小さくなるセル (i, j) が最も識別に有効であるといえる. つまり, セル (i, j) から得られる 2 値勾配強度コード X を入力とし, 対応する識別スコア S を出力とする弱識別器が最も有効であると判断する. 弱識別器決定後, 学習画像の重みを更新し, 再び頻度分布の作成に戻る. この操作を繰り返し, ユーザーが設定したラウンド回数だけセル (i, j) が決定される.

(3) 強識別器

得られた弱識別器の出力の線形和を強識別器の出力とする. 入力画像から, 各弱識別器が着目するセルの 2 値勾配強度コードを抽出し出力を得る. 得られた出力の線形和がある閾値を超えているかどうかで入力画像の識別を行う.

4. 低次元特徴量の検討

3 章において, 各弱識別器の出力は特徴量の出現頻度に依存することを述べた. 出現頻度が限りなく低い 2 値勾配強度コードは, 実際の識別に与える影響が少ないといえる. つまり 3.1 節において, 8 方向 2 値勾配強度コードを用いたが, 必ずしも 8 方向を選ぶ必要はない. 方向を選ぶ際, より少ない方向数でコードを作成することができればさらなる情報量の削減につながる. 少ない方向数でコードを作成することで, 各コードの出現頻度が高まり, 認識への貢献が増すと考えられる. この操作は, 8 方向 2 値勾配強度コードのいくつかを同一とみなすことに等しい. このとき, 出現頻度が低いコードを他のコードと同一とみなすことができれば, 認識精度を落とすことなく情報量の削減

が行えると考える.

4.1 提案手法 : 6 方向 2 値勾配強度コード

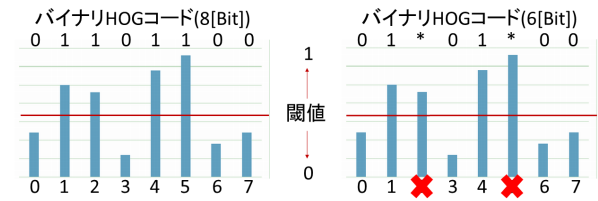


図 7 8 方向 2 値勾配強度コードと 6 方向 2 値勾配強度コードの比較

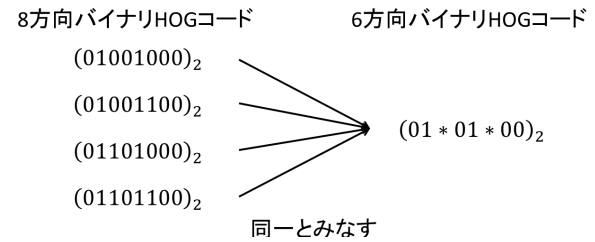


図 8 8 方向 2 値勾配強度コードと 6 方向 2 値勾配強度コードの関係

図 7 は, 一部の勾配方向 (6 方向) のみを用いて 2 値勾配強度コードを作成した例である. この例の 6 方向 2 値勾配強度コードでは, ヒストグラムが 6 ビット列 $(010100)_2$ で表されている. 元の 8 方向 2 値勾配強度コード $(01101100)_2$ の下から 3 ビット目, 6 ビット目の値によらずヒストグラムを表現する. 言い換えると, 量子化後の勾配方向 $\theta_d(x, y) = 0, 1, 3, 4, 6, 7$ を用いてコードを作成している. 以後このようなコードの構成を $H_b(013467)$ と表記する. これは図 8 のように, いくつかの 8 方向 2 値勾配強度コードを同一とみなすことに等しい.

4.2 出現頻度分布の変化

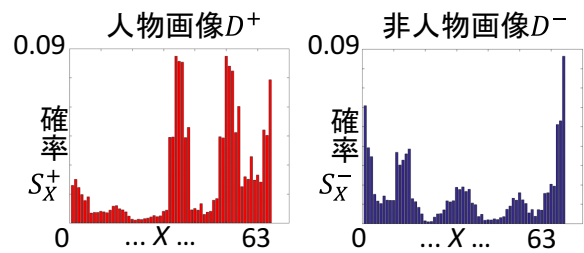


図 9 出現頻度分布 (6 方向 2 値勾配強度コード)

3.3 節で考えた出現頻度分布を同様に作成する. 図 9 は提案手法を用いて作成した出現頻度分布である. また, 図 10 は図 9 における両出現頻度分布の差を表す. 各セルのヒストグラムが 6 ビットで表されるため, 横軸の次元数は 64 となる. 弱識別器の定義や学習段階での識別器の構築方法は同様である.

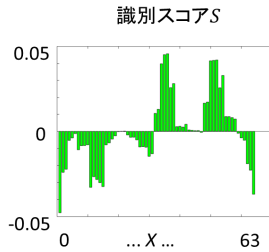


図 10 出現頻度分布の差 (6 方向 2 値勾配強度コード)

5. 評価

4.1 節で述べた提案手法を用いて認識精度の評価を行う。

4.1 節において、少ない勾配方向で 2 値勾配強度コードを作成することを述べた。このとき、どの勾配方向を選ぶべきか考慮する必要がある。また、8 方向の中からいくつの勾配方向を選んでコードを作成すべきか考慮することも必要である。

初めにある画像データセットを用いて、実際に少ない勾配方向でコードを作成する。作成したコードを用いて実際に識別を行い、認識精度を評価する。このとき従来の HOG, 8 方向 2 値勾配強度コードと比較して認識精度にどのような変化が起きるか観察する。そして同等以上の認識精度を得ることができるコードがあるか探す。

次に、人物認識に有効であるコードと異なるデータセットを用いて識別実験を行う。このときも従来の HOG, 8 方向 2 値勾配強度コードと同等以上の認識精度を得ることができれば、このコードは人物認識において汎用的に用いることができると考える。

今回、二値化処理に用いた閾値は文献 [3] を参考に 8/128 を用いている。

5.1 実験 1

4.1 節において、ヒストグラムの一部の勾配方向に二値化を施すことを述べた。このとき、どの勾配方向に処理を施すべきか考慮する必要がある。また、8 方向の中からいくつの勾配方向に対して二値化を施すべきか考慮することも必要である。

- 実験方法

データセットに CU dataset [7] を用いる。このデータセットは、屋外環境下で撮影された画像から一定区画を切り出した画像データベースである。初めに実際に少ない勾配方向数で 2 値勾配強度コードを作成する。例えば 8 方向から 7 方向を選んで作成できるコードは 8 パターン、6 方向を選んで作成できるコードは 28 パターン存在する。今回はこれら 6 方向と 7 方向で作成されるコードの全 36 パターンを用いて識別実験を行った。

- 結果

本論文では、認識精度が優れていた 2 パターン、悪化したパターンの結果を従来の HOG, 8 方向 2 値勾配強度コードを用いた場合と共に表 1 に提示する。認識精度の評価として、誤認識率 5.0% 時の認識精度を提示する。認識率、誤認識率の定義を以下に示す。

- 認識率

$$100 - \frac{\text{人物画像を非人物と認識した枚数}}{\text{人物画像総枚数}}$$

- 誤認識率

$$\frac{\text{非人物画像を人物と認識した枚数}}{\text{非人物画像総枚数}}$$

表 1 CU dataset を用いた場合の認識率

手法	認識精度
HOG	73.2
8 方向 2 値勾配強度コード	68.8
6 方向 2 値勾配強度コード $H_b(013467)$	77.5
6 方向 2 値勾配強度コード $H_b(013457)$	75.7
6 方向 2 値勾配強度コード $H_b(123456)$	63.5

提案手法の 6 方向 2 値勾配強度コードは、従来の HOG と比較して最大 4.3 ポイント、8 方向 2 値勾配強度コードと比較して最大 8.7 ポイント認識精度の改善が見られた。また、方向の選び方によっては認識精度の低下が見られた。認識精度が向上するパターンは限られているといえる。

5.2 実験 2

実験 1 で有効に評価した 6 方向 2 値勾配強度コードを用いて、異なるデータセットに対して識別実験を行う。今回、データセットに INRIA person dataset [8] を用いる。認識精度の評価は実験 1 同様、誤認識率 5.0% 時の認識率を用いる。結果を従来の HOG, 8 方向 2 値勾配強度コードを用いた場合と共に表 2 に提示する。

表 2 INRIA person dataset を用いた場合の認識率

手法	認識精度
HOG	96.6
8 方向バイナリ HOG	94.5
6 方向 2 値勾配強度コード $H_b(013467)$	94.4
6 方向 2 値勾配強度コード $H_b(013457)$	94.8
6 方向 2 値勾配強度コード $H_b(123456)$	92.9

提案手法の 6 方向 2 値勾配強度コードは、従来の HOG と比較して最大 3.7 ポイントの低下、8 方向 2 値勾配強度コードと比較して最大 0.3 ポイントの改善が見られた。8 方向 2 値勾配強度コードと、実験 1 で有効とされた 2 パターンはほぼ同程度の認識精度が得られた。一方、実験 1 で有効ではないと考えられたコードを適用した場合、実験 1 同様に認識精度は最も低い数字が得られた。

5.3 考察

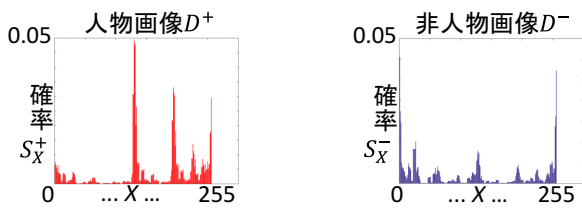


図 11 8 方向 2 値勾配強度コードを用いた場合の出現頻度分布

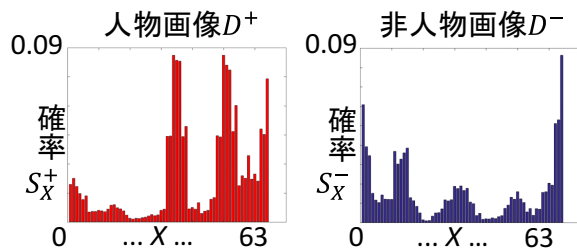


図 12 6 方向 2 値勾配強度コードを用いた場合の出現頻度分布

図 12 は、実験 1 において最も精度の良かった 6 方向 2 値勾配強度コード $H_b(013467)$ を用いた場合の出現頻度分布の差を示す。また、図 11 は、8 方向 2 値勾配強度コードを用いた場合の出現頻度分布の差を示す。先に各コードの出現頻度が高まることで、認識への貢献が高まると述べた。図 12 と図 11 を比較すると、6 方向 2 値勾配強度コードを用いた場合、コードがより多く出現している。各コードの出現頻度がある程度の大きさになっており、識別への貢献が高まっているといえる。

6. 結論

本研究では、従来の HOG で用いる輝度の勾配方向の中から、認識に有効な勾配方向を選んでヒストグラム化し、さらに二値化処理を施すこと得られる低次元特徴量を提案した。8 方向 2 値勾配強度コードとブースティングを用いた識別は、出現頻度の低い特徴量が存在する問題がある。提案した特徴量を用いることで各特徴量の出現頻度が高め、認識精度を改善することを目的とする。提案した特徴量を用いて識別実験を行ったところ、従来の HOG、二値化 HOG を用いた場合と比較して同等以上の認識精度を得ることができた。異なるデータセットを用い、有効とされた特徴量を用いた場合、認識精度は従来法とほぼ同等であった。一方、有効でないとされた特徴量を用いた場合、先の結果と同様に認識精度は最も低くなった。一連の結果から、一つのデータセットに対してその有効性が決定された特徴量は、異なるデータセットに対してもある程度の汎用性があるといえる。提案した特徴量は、それぞれ出現頻度が高まり、認識への貢献が増すと考えられる。また、特

微量の格納に伴うメモリ量は、従来の HOG の表現精度に倍精度浮動小数点を用いた場合と比較して約 99%削減される。メモリ量削減と高い認識精度が同時に実現可能であると結論付けられる。

参考文献

- [1] N. Dalal and B. Triggs: *Histograms of oriented gradients for human detection*, Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on. IEEE, 2005 .
- [2] K. Negi, K. Dohi, Y. Shibata, and K. Oguri: *Deep pipelined one-chip FPGA implementation of a real-time image-based human detection algorithm*, Field-Programmable Technology (FPT), 2011 International Conference on , vol., no., pp.1,8, 12-14 Dec. 2011.
- [3] J. Rettkowski, A. Boutros, and D. Göhringer: *Real-time pedestrian detection on a xilinx zynq using the HOG algorithm*, ReConfigurable Computing and FPGAs, 2015 International Conference on. IEEE, 2015.
- [4] W. Sun and K. Kise: *Speeding up the Detection of Line Drawings Using a Hash Table*, in Proceedings of The 1st China Japan Korea Joint Workshop on Pattern Recognition, vol. 2, nov 2009, pp. 896–900.
- [5] Y. Freund and R. E. Schapire: *A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting*, Journal of Computer and System Sciences, 37, p297-p336, 1999.
- [6] R. E. Schapire and Y. Singer: *Improved Boosting Algorithms Using Confidence-rated Predictions*, Machine Learning, 37, p297-p336, 1999.
- [7] CU-dataset
入手先 <<http://www.vision.cs.chubu.ac.jp/CVTutorial/>>
- [8] INRIA Person Dataset
入手先 <<http://pascal.inrialpes.fr/data/human/>>