

HOG 特徴の n -gram 表現を用いた人物行動による物体認識

小島 篤博[†] 笹野 泰正^{*††} 黄瀬 浩一^{††}

[†] 大阪府立大学 高等教育推進機構 〒 599-8531 大阪府堺市中区学園町 1-1

^{††} 大阪府立大学大学院 工学研究科 〒 599-8531 大阪府堺市中区学園町 1-1

E-mail: [†]ark@las.osakafu-u.ac.jp, ^{††}sasano@m.cs.osakafu-u.ac.jp, ^{†††}kise@cs.osakafu-u.ac.jp

あらまし 従来、物体認識の分野では物体の外観的特徴に基づく手法が主流であった。これに対し、物体を扱う際の人物の動作特徴に基づき、対象物体を識別する手法が試みられている。本研究では、HOG 特徴で表した人物姿勢の変化を n -gram で表現し、対象物体の学習・認識を行う手法を提案する。まず、入力動画の各フレームにおける人物の姿勢を HOG 特徴により表現し、連続フレームの姿勢変化を n -gram として抽出する。そして、認識対象の物体を人物動作の n -gram の集合とみなすことで学習・認識を行う。本論文では、室内物体について実験を行い、提案手法の有効性を確認した。

キーワード 人物動作による物体認識, n -gram 表現, HOG 特徴, Bag of Features

1. はじめに

従来、物体認識の研究では、色や形状、テクスチャなどの物体の外観的特徴を認識の手がかりとする研究が数多く提案されている。しかしながら、同じカテゴリとして扱われる物体の中にも様々なデザインのものがあるため、外観的特徴を用いる手法では学習に必要なデータ数が膨大になるという問題がある。例えば“イス”を学習するためには、パイプイスや安楽イス、ソファなどの様々な種類のイスを用意しなくてはならない。

一方で、人物が物体を扱う際の挙動は、物体のデザインによらず類似している。例えば、安楽イスでも、パイプイスでも、イスであるならば“座る”という動作は共通である。これはそれぞれのイスが“座ることができる”という同じ機能を提供しているためであり、物体に対する人物の行動はその物体の機能を表していると言える。このため、物体の外観的特徴を用いる代わりに、その物体を使う人物の行動を観察することで対象物の機能および属性を認識し、対象物の認識を行うという新しい考え方が提案されている [1]。Moore らは、モデルベースの物体認識と物体に対する人物の手の動きを相補的に用いることで、人物の行動と物体とを推定する手法を提案している [2]。また、樋口らは人物の動作と物体の機能や用途といった概念を階層モデル化することで、人物動作と物体の形状や機能の統合的な認識を試みている [3]。さらに物体と動作との関係を確認ネットワークで学習させる結び付けることで、より柔軟な認識を目指した手法も提案されている [4]。また、確率的に動作と物体との関係を学習させる手法で、物体の外観情報を用いることなく物体カテゴリを識別する研究もなされている [5], [6]。本研究

でもこの考え方にに基づき、人物の行動から物体を認識する手法を提案する。

物体と人物行動との関連性に注目すると、一つの物体に対する動作のパターンは、時間軸上において局所的な挙動の集合であると考えられることができる。例えば、イスに関連する動作としては、座る、立つなどの動作があるが、これらの動作は手を動かしたり、足を曲げたりなどの前後数フレーム分の人物の挙動の集合であると考えられることができる。そこで本研究では、物体を扱う際の人物の行動を動作特徴の集合で表現し、物体を学習・認識する手法を提案する。

本研究では、まず人物が物体を扱う際の動作の特徴として、姿勢特徴の n -gram 表現を用いる。 n -gram とは主に文書解析で用いられる特徴表現の一つで、記号列中に含まれる長さ n の部分列のことである。人物行動の特徴表現に n -gram を用いた事例として、Hamid らは人物の行動を n -gram モデルで表現し、異常行動を監視する方法を提案している [7]。ただし、Hamid らの手法は高レベルな動作に対するアプローチであり、認識済みの低レベルの動作の系列に対して n -gram モデルを適用したものである。また、姿勢特徴として HOG 特徴量を用い、動作特徴として n -gram モデルを適用したもとしては、Thureau らによるものが挙げられる [8]。本研究でも、人物の姿勢特徴として幾何学的変化や照明変化に頑健な HOG 特徴 [9] を用いる。そして、各物体を扱う人物の行動を Bag-of-Features の考え方にに基づき n -gram の集合で表し、学習・認識を行う [10]。Bag-of-Features は、画像検索などにおいて、大量の特徴量に基づく手法として広く用いられている。

以下、2. で本研究の概要について説明する。そして、3. では人物の動作を HOG 特徴量として抽出し、記号に置き換え、4. では記号の n -gram 表現により動作特徴を

(注*): 現在 オムロンソフトウェア株式会社

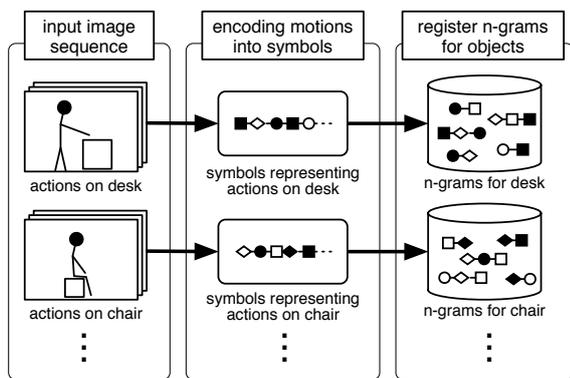


図1 提案手法の流れ

表現し、Bag-of-Features の考え方にに基づき動作特徴の出現頻度の分布を用いて物体カテゴリを学習・認識する手法について説明する。そして、実験結果を 5. に示し、6. と 7. で考察しまとめる。

2. 提案手法の概要

ここでは、提案手法の概要について説明する。図 1 に人物動作から物体カテゴリを学習する手順を示す。まず、人物が物体を扱う動作を動画像で撮影し、1 フレームごとに人物領域を抽出する。そして、その領域から人物の姿勢特徴として HOG 特徴を抽出する。さらに、それぞれの HOG 特徴ベクトルを一意な記号に割り当てることにより、動画像に対応する記号系列を得る。次に、この記号系列に対して動作特徴の表現として n -gram を抽出する。そして、Bag-of-Features の考え方にに基づき、物体カテゴリごとに n -gram の出現頻度の分布を集計したものを学習する。また、認識過程では、認識対象の動画像に対して同様に n -gram の出現頻度の分布を求めた後、各物体カテゴリの分布と照合して物体を識別する。

まず、各物体カテゴリについて、人物が物体を扱う動作を動画像で撮影する。例えば、棚の場合、人物が棚の扉を開け、物を取り出し、扉を閉めて棚から離れるまでの一連の動作を記録する。次に、得られた画像系列から人物姿勢の特徴列を求める。動画像から 1 フレームごとに画像を取得し、画像から人物の領域を求める。そして、その人物領域に対して人物姿勢を表現する HOG 特徴量を抽出し、その HOG 特徴に対し記号化という処理を行うことで一意な記号に置き換える。次に、人物姿勢を表現する記号系列から、動作特徴の表現として n -gram を抽出する。 n -gram とは記号列中に含まれる長さ n の部分列のことで、 n フレーム分の姿勢特徴の並びを用いることで、動作特徴を表現する。そして、これらを Bag-of-Features の考え方にに基づき、物体カテゴリごとに n -gram の出現頻度の分布を学習する。以上の処理により、人物の動作を用いて各物体カテゴリを学習することができる。

一方、物体を認識する際は、認識対象の画像系列に対

して同様に n -gram の出現頻度の分布を求めた後、各物体カテゴリの分布と照合して物体を識別する。

3. 人物姿勢の記号化

続いて、人物の動作からその人物の姿勢の連続的な変化を表す記号系列を抽出する手順について説明する。まず、人物動作を動画像で撮影し、そこから 1 フレームごとに人物の姿勢特徴として HOG 特徴を抽出する。更にその HOG 特徴に対し記号化という処理を行うことで一意な記号に置き換え、人物動作に対応する記号系列を取得する。

本手法では、人物の動作特徴として HOG 特徴の n -gram を用い、この頻度の分布により物体を表現する。しかしながら、HOG 特徴は多様な値を取り、そのままでは HOG 特徴間の対応付けが難しいため、ある程度類似した姿勢を表す HOG 特徴を同一の記号で表現することとする。この記号化の処理により、以後は人物動作を記号系列として扱うことができる。

3.1 人物姿勢からの HOG 特徴の抽出

まず、人物が物体に対して行う一連の動作を動画像で撮影する。例えば、棚に対する人物動作であれば、人物が棚の扉を開け、物を取り出し、扉を閉めて棚から離れるまでの一連の動作を記録する。本研究の前提条件として、各動画像において、人物が複数の物体を同時に扱っていないものとする。

次に、1 フレームずつ動画像から姿勢特徴を抽出する。先行研究 [6] においては、人物の姿勢を表現するために、手や顔の位置を特徴量として用いている。それに対して、本研究では人物の全身の輝度勾配から特徴を抽出することによって、より詳細に人物姿勢を記述できる HOG 特徴を用いる。HOG 特徴は、局所領域における輝度勾配の分布に基づく特徴量であり、人物領域を検出するなどの目的で用いられる。HOG 特徴を用いることで、人物を撮影する方向や照明の明るさの変動に対して頑健な認識が可能となる。

特徴抽出の前処理として、画像から HOG を抽出する対象となる人物領域を求める。まず、背景差分法によって背景を除去し、差分として人物のシルエットの二値画像を得る。次に、このシルエットの二値画像に対してモルフォロジー変換を加えることによってノイズを除去したあと、最も大きい連結成分を抜き出し、外接する矩形により切り出す。そして、これをマスクとして元画像に適用したものを人物領域とする。このとき、同時に人物領域のサイズの正規化を行う。これは画像サイズによって、得られる HOG 特徴量の次元数が変化するためである。本手法では 40×40 画素にリサイズすることで正規化を行う。

まず、各画素 (i, j) の輝度値 L に対して、局所領域における勾配強度 m_{ij} と勾配方向 θ_{ij} を式 (1),(2) を用いて

算出する．ここで， f_x および f_y はそれぞれ x 方向および y 方向における輝度変化である．

$$m_{ij} = \sqrt{f_x(i, j)^2 + f_y(i, j)^2} \quad (1)$$

$$\theta_{ij} = \tan^{-1} \frac{f_y(i, j)}{f_x(i, j)} \quad (2)$$

次に， 5×5 画素のセルに分割し，各セルにおいて輝度勾配の分布を求める．まず，勾配方向 θ_{ij} を 9 等分して区切り，それぞれピンとすることでヒストグラムを作成する．そして，ヒストグラムの各ピンについて，勾配強度 m_{ij} を積算することで分布を求める．さらに，セルを複数まとめた領域をブロックとし，式 (3) によりブロックごとの正規化を行う．ここで， h_{ij} はブロック内の i 行 j 列におけるヒストグラムの値であり， ϵ は分母が 0 になるのを防ぐための係数である．また， \sum_{block} はブロック内のすべてのヒストグラムの値の総和を取ることを表す．

$$h'_{ij} = \frac{h_{ij}}{\sum_{block} h_{ij}^2 + \epsilon} \quad (3)$$

このようにして得られたブロックごとの正規化ベクトル h'_{ij} を連結することにより，HOG 特徴ベクトルが得られる．

3.2 HOG 特徴の記号化

このようにして得られた HOG 特徴量の次元数は (勾配方向の数 \times ブロックあたりのセルの数 \times ブロックの数) となり，一般に高次元になる．したがって，HOG 特徴量の n -gram の頻度は非常に疎な分布になる．そこで，記号化という処理を行うことにより得られる姿勢特徴の数を絞り込む．

記号化というのは，特徴を一意的な記号に置き換える処理のことである．すなわち，記号化の処理では，類似した姿勢には同じ記号を，異なる姿勢には異なる記号を割り当てる．

図 2 に具体的な記号化の処理を示す．前処理として，記号化に用いる対応表，すなわちコードブックを作成する．まず，学習用の画像から得られたすべての姿勢特徴ベクトルに対してクラスタリングの処理を行うことにより，各クラスタの重心ベクトルを求める．本手法ではクラスタリングに K-means を使い， K 個の重心ベクトルを求める．そして，この重心ベクトルそれぞれに対して一意的な記号を割り当て，対応表を作成する．

次に，この対応表を用いて姿勢特徴を記号化する．まず，記号化の対象となる姿勢特徴ベクトルと各重心ベクトルに対して最近傍探索を行う．そして，最近傍重心ベクトルに割り当てられた記号を，対象ベクトルに対応する記号とする．

ここまでの処理により，画像系列の各フレームに対応する HOG 特徴を抽出し，記号化の処理により HOG 特徴に対応する記号が得られる．したがって，これを

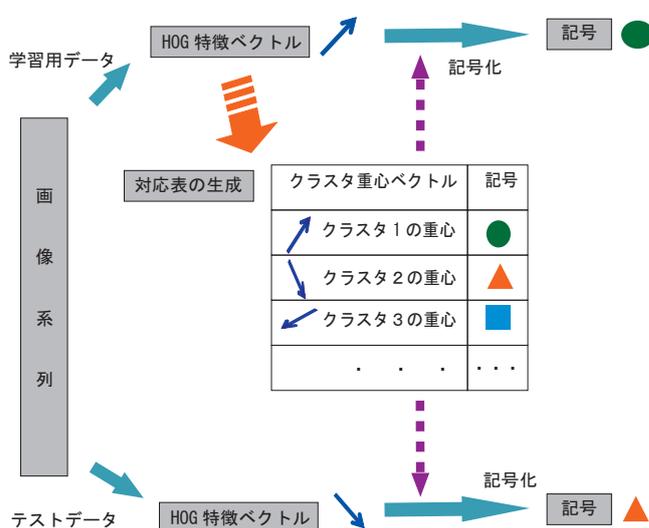


図 2 姿勢特徴の記号化

時系列順に並べることで，動画像に対応した記号系列 $x_0x_1 \dots x_t \dots$ が得られる．

4. 物体カテゴリの学習と認識

ここでは，人物動作特徴の記号系列 $x_0x_1 \dots x_t \dots$ を用いて，物体カテゴリに対応する人物の行動を学習・認識する方法について説明する．本研究では，動作特徴として，人物動作特徴の記号系列から n -gram を抽出し，これらを Bag-of-Features の考え方にに基づき，物体カテゴリごとに n -gram の出現頻度の分布を学習する．認識過程では，認識対象の画像系列に対して同様に出現頻度の分布を求めた後，各物体カテゴリの分布と照合して物体を識別する．

4.1 人物動作の n -gram 表現

n -gram とは，図 3 のように，一連の記号系列から連続する n 個の記号を抽出したものである．例えば， $abcd$ という記号系列から得られる n -gram は $n = 2$ の場合， ab, bc, cd であり， $n = 3$ の場合， abc, bcd である．一般に，人物動作特徴の記号系列 $A = x_0x_1 \dots x_t \dots$ に対して，任意の長さの n -gram x_t^n を以下のように定める．

$$x_t^n = \underbrace{x_t x_{t+1} \dots x_{t+n-1}}_n \quad (4)$$

本研究では，各記号は各フレームにおける姿勢を表している．したがって記号系列から抽出した n -gram は連続した n フレームの姿勢変化を表しており，同時に人物の動作特徴を表している．また， $n = 1$ の場合は，そのまま 1 フレームの姿勢特徴を表していると考えられることができる．

また本研究では， n -gram 抽出の前処理として， A に連続して同じ記号が続く場合，重複した記号 (すなわち $x_k = x_{k+1}$ になる場合は x_{k+1}) を除去する．この処理は動作の速度の変化を吸収し，時間伸縮性を与える効果が

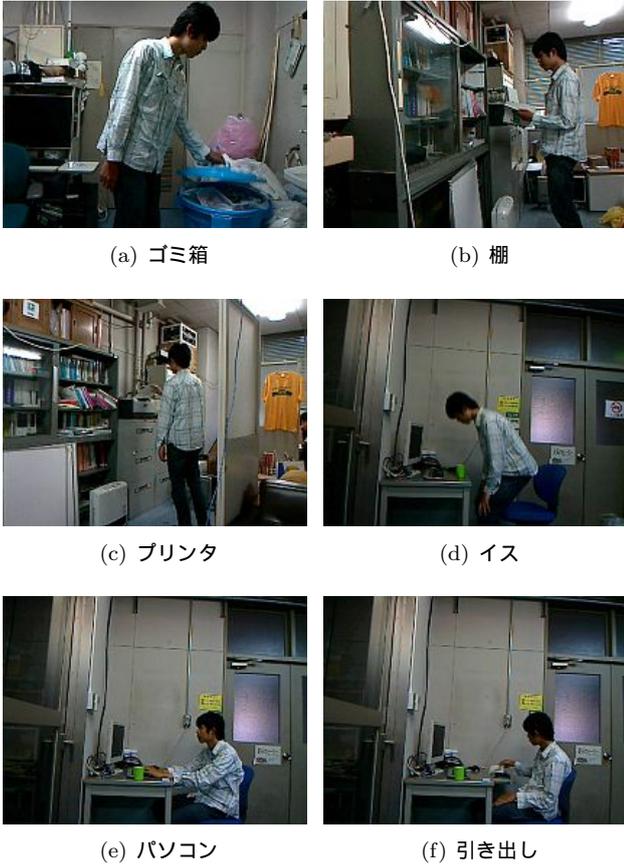


図 4 実験に用いた物体（一部）

5.1 実験条件

実験に用いる 11 物体のうち、図 4(a)~(c) は立った状態で扱う物体、(e)~(f) が座った状態で扱う物体である。また、(d) のイスは立った状態と座った状態の両方の状態を遷移する物体である。各物体に対して、物体を使用する人物の行動を撮影し、1 回の撮影につき 150 フレームの画像系列を得た。この動作の中には、人物が物体を使用する動作の他に物体とは無関係の動作が含まれる。このうち、前者は識別性の高い動作であるから、4.2 で示したように大きな重みを与えられる。逆に、物体とは無関係の動作は、複数の物体間で共通して出現するため、小さな重みを与えられる。結果として、これは人物が物体を使用する動作を抽出していることと類似していると言える。また、撮像条件を変化させるため、一つの物体に対し、人物行動が最もよく観測できる方向から 0° 、 30° 、 45° と角度を変化させて撮影した。

また、1 つの物体カテゴリに対して 1 人あたり 30 回の実験を行い、5 段階の交差検定を用いて認識率を求めた。物体の撮影には解像度 640×480 の一般的な USB カメラを用い、撮影のフレームレートは 5fps で行った。

本実験の結果を評価するため、各物体のテストデータに対し、それぞれの物体として認識されるか確認した。次に、 n の値を変化させた際の全物体の平均認識率、および各物体の認識率の変化についての評価を行った。ま

Drawer	0.0	0.7	0.0	0.0	0.0	1.3	6.0	0.0	2.0	2.7	31.3	56.0
Pen	0.0	0.0	0.0	0.0	0.7	0.0	0.7	4.0	0.0	9.3	0.0	85.3
Book	0.0	0.7	0.0	0.0	0.0	0.0	3.3	2.0	14.0	2.0	1.3	76.7
PC	0.0	0.0	0.0	0.0	0.0	0.0	0.7	8.7	0.7	0.7	0.7	88.7
Cup	0.0	0.0	0.0	1.3	0.7	0.7	30.7	0.7	2.0	4.0	4.7	55.3
Chair	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0
Door	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Hanger	0.0	0.0	6.0	90.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	4.0
Printer	1.3	8.0	82.7	2.7	0.0	0.0	0.0	0.0	0.0	0.0	1.3	4.0
Shelf	0.0	70.7	12.7	6.0	0.7	0.0	0.0	0.0	0.0	0.0	0.0	10.0
Trash	90.7	0.7	0.0	0.0	0.0	0.7	0.0	0.0	0.0	0.0	0.7	7.3
	Trash	Shelf	Printer	Hanger	door	Chair	Cup	PC	Book	Pen	Drawer	Reject

図 5 $n = 5$ の場合の認識結果

た実験では参考のため、 n -gram を単独で用いる場合と、 $n = 1$ から N までを混合して用いる場合についての比較を行った。なお、この 2 つの実験では、姿勢を表現する記号の数 K を 150 に設定している。次に、物体を学習する動作の個人差をみるため、認識対象となるテストデータと同じ人物のデータを学習用データから取り除いた際の認識率について評価を行った。ここで、今回学習に用いた物体は 11 物体であるが、この物体数を増やすと物体の表現に必要な姿勢の数が増えていくため、学習する物体数によって最適な記号の数 K を探す必要がある。そこで、姿勢を表現する記号の数 K を変化させた際の平均認識率を確認した。

5.2 実験結果

図 5 に $n = 5$ の場合の認識結果を示す。このとき、各物体の認識率は平均して 57.1% となった。次に、図 6 に $n = 1$ から 5 までの n -gram を混合した場合の認識結果を示す。このときの各物体の平均認識率は 72.3% となった。2 つの図を比較すると、図 5 では座って扱う物体の認識率、特に PC の認識率は 8.7% と低くなっていることがわかる。また、同時に座って扱う物体のリジェクト率が高くなっている。これは、 n の値が大きいく n -gram のほうが相対的に出現頻度が低く、テストデータと学習用データのヒストグラムに共通する n -gram が現れなくなるためである。図 6 では、リジェクトがなくなり、PC の認識率も 62.7% と大幅に向上したことがわかる。他の物体に対しても同様に認識率が向上していることから、 n -gram を混合することが認識率の上昇につながることがわかる。

また、類似した姿勢で扱う物体は誤認識される傾向がみられた。例えば、棚とプリンタは互いに 10.0%、14.7% の誤認識があったが、これはプリンタに給紙する動作と棚に本を入れる動作が類似しているためだと考えられる。

図 6 において、立って扱う物体の認識率の平均 (90.5%)

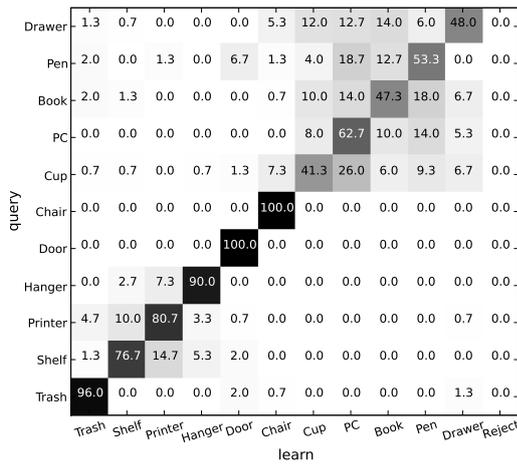


図6 $n = 1$ から 5 まで混合した場合の認識結果

は比較的良好であったのに対し、座って扱う物体の認識率の平均 (50.5%) は低下した傾向がみられた。これは、座った状態で物体を扱う場合、複数の異なる物体を扱う動作が類似した姿勢で行われることが原因であると考えられる。特に、座って扱う物体はしばしば PC として誤認識された (コップ 26.0%, 本 14.0%, ペン 18.7%, 引き出し 12.7%)。

図4で示したような立った状態で扱う物体と座った状態で扱う物体との間の誤認識は、相互に少ない傾向にある。これは立って扱う時に現れる姿勢特徴と座って扱う時に現れる姿勢特徴が大きく異なるためと思われる。

次に、 n の値を 1 から 20 まで変化させた際の各物体カテゴリの認識率および全カテゴリの平均認識率を評価した。図7に認識結果を示す。まず、全カテゴリの平均認識率について、 n -gram 単独でヒストグラムを作成した場合、 n の値を大きくすると認識率が $n = 2$ の場合を除き、単調に減少していることがわかる。これは、 n の値が大きい n -gram は出現頻度が相対的に低く、テストデータの中に含まれる全ての n -gram がどの物体カテゴリのヒストグラムにも出現せず、リジェクトが発生するためである。一方、 n -gram を混合してヒストグラムを作成した場合、 n の値が大きくなった場合でも必ずいずれかのカテゴリに識別されるため、 n の値が大きくなるほど認識率は高くなった。また $n = 14$ で認識率 (74.2%) が変化しなくなり、飽和していることが分かる。

次に、物体を学習する動作の個人差を確認するため、認識対象となるテストデータと同じ人物の学習用データを含む場合、および含まない場合の認識率について比較したものを図8に示す。また、ここでは姿勢を表現する記号の数 K を変化させた際の平均認識率を確認した。テストデータと同一人物の学習データを取り除いていない場合、 K の値を大きくすれば大きくするほど物体の認識率は高くなった。これは、 K の値が大きくなると、表現できる姿勢のバリエーションが多くなり、細かな動きの

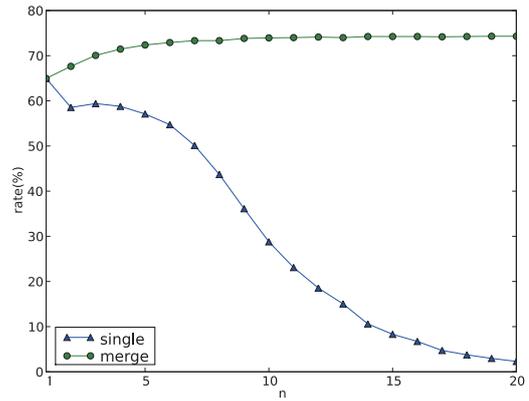


図7 n の値を変化させた際の認識率の変化

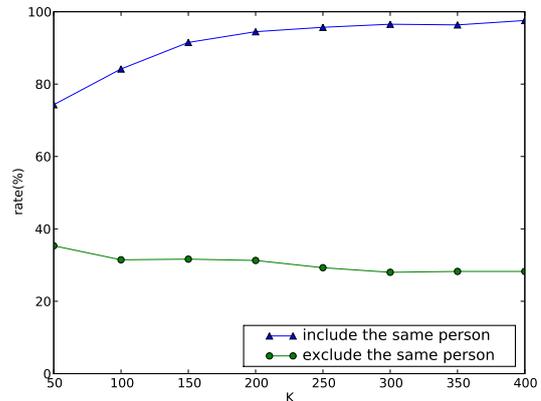


図8 テストデータと同じ人物を含む場合と含まない場合の認識率の差異

違いが表現できるよになるためであると考えられる。一方、テストデータと同一人物の学習データを取り除くと、先ほどとは逆に K の値を大きくするにつれて認識率が低下する。これは複数人の中でみられる共通の姿勢が、 K の値を大きくすることによって、人物ごとに異なる記号に割り当てられるためであると考えられる、また、テストデータと同一人物の学習データを取り除いた時の認識率は、30%前後と低い値にとどまっている。これは本手法で人物姿勢の個人差を吸収しきれていないことを示しており、人物姿勢の表現に改良が必要であると考えられる。

6. 考察

前節で示した実験結果について5つの観点から考察する。

(1) 類似した姿勢で扱う物体について

実験では、類似した動作で扱う物体は互いに誤認識される傾向がみられた。例えば、 n -gram を $n = 1$ から 5 まで混合してヒストグラムを作成した場合、棚とプリンタの間には相互に 10.0%, 14.7% の誤認識があった。これは、本手法では動作特徴として人物の全身の姿勢特徴

である HOG 特徴に基づく特徴を用いるため、全身の姿勢としては、棚に本を収納する姿勢とプリンタに給紙する姿勢が類似しているためであると考えられる。一方で、手の動きに着目すると、棚に本を収納する際とプリンタを給紙する際の手の動きは互いに異なると考えられる。そこで今後の課題として、全身の特徴だけでなく、体の一部から得られる特徴も抽出できるよう、人物の動作特徴の表現を改良することなどが挙げられる。

(2) 姿勢の変化が少ない物体について

実験結果に示した通り、立って扱う物体において高い認識率を得ることができた。これは、立って扱う物体は動作がより大きく、その物体固有の動きをよく表現しているためであると考えられる。一方で、PC などのように、物体を扱う際、人物の姿勢がほとんど変化しない物体の認識率は比較的低くなっている。そこで、人物姿勢を表現する記号の数のパラメータを調節することにより、人物の姿勢をより正確に取り出すことが考えられる。また、本手法において記号数はクラスタリングに K-means を用いているため固定となっているが、X-means などのクラスタ数を推定する手法を用いることで、より適切な記号の数を自動で推定するなど考えられる。

また、本研究では、 n -gram 抽出の前処理として、記号系列に連続して同じ記号が続く場合、重複した記号を除去している。しかしながら、連続する記号の長さに関わりなく重複した記号を除去することにより姿勢の変化の特徴が失われてしまうという問題がある。そこで、連続した記号を全て除去するのではなく、連続した記号の数がある程度反映されるように重複した記号を残す方法が考えられる。これは例えば、連続した記号 n 個分が 1 つの記号になるように除去するなどである。

(3) 物体を扱っている時区間について

本手法では、人物が物体を扱う動作の他に物体とは無関係な動作も撮影される。この物体とは無関係な動作区間は物体を識別する際ノイズとなるため、可能な限り除去すべきである。これに対して本研究では、人物が物体を扱う動作に重みを与えることでこの問題に対処している。まず、人物が物体を扱う動作はその物体固有の動作であることが比較的多いため、4.2 で示したように大きな重みが与えられる。また、物体とは無関係な動作は複数の物体に重複して現れることが比較的多く、小さな重みを与えられる。結果として物体を扱っている時区間に含まれる動作がより強調され、これにより正しい物体カテゴリが識別できていると考えられる。

(4) 物体領域の検出について

一般的に外観的特徴を用いた物体認識手法では、物体から特徴量を抽出するために画像のセグメンテーションなどを行い、物体の領域を検出する必要がある。しかし

ながら本研究では、人物の動作から特徴を抽出し物体のカテゴリを識別することを目的としているため、物体領域そのものの検出は特に行っていない。今後の課題として、人物の姿勢と物体の位置の関係性に着目することで、人物行動を用いて物体の領域を検出することが挙げられる。そこで、人物の姿勢特徴に人物と物体の相対的な位置関係の情報を付加することで、物体の位置を推定することが可能なのではないかと考えられる。

更に今後の課題として、小型ロボットを用いて室内を探索し、人物の行動を用いて物体のカテゴリと位置情報を認識し、室内の環境マップを作成するなどのアプリケーションに応用することが考えられる。

(5) 動作認識との違いについて

1. において、関連研究として物体と人物行動を関連付けることにより物体を認識する手法を紹介した。これらの人物行動に基づく物体認識手法は、単に動作認識によって動作カテゴリを求め、それを物体カテゴリに置き換えたものではない。例えば、イスのような物体に対して「座る」「立つ」といった個別の動作以外にも、その間の過渡的な動作が含まれているはずであるが、このような過渡的な動作のなかにも、イスに固有の何らかの特徴が含まれている可能性がある。これに対して本研究では、物体に対して行われる一連の動作から得られる動作特徴の集合を用いることで対象物体の認識を行なっている。

7. ま と め

本研究では、人物の姿勢特徴の n -gram 表現を動作特徴として用い、Bag-of-Features の考え方により動作特徴の集合で物体を表現することにより、物体を学習・認識する手法を提案した。従来、外観情報を用いた物体認識には、様々なデザインを持った物体をすべて登録しなければならないという問題点があった。これに対して、本手法では、物体の外観情報を用いることなく、人物行動を用いて物体カテゴリを推定する手法を提案している。人物の姿勢特徴として HOG 特徴として抽出し、人物が物体を扱う際の動作特徴として姿勢特徴の n -gram を作成し、Bag-of-Features の考え方に基づいて、認識対象の物体を動作特徴の集合とみなすことで、物体カテゴリと人物動作を対応付けている。そして実験により、11 種類の物体を異なる 5 人の人物の動作により学習・認識することで本手法の有効性を示した。今後の課題として物体を扱う姿勢と物体カテゴリの対応をより詳細に分析し、特徴の表現を改良することなどが挙げられる。

文 献

- [1] M. Kitahashi, A. Kojima, M. Higuchi, and K. Fukunaga: "Toward a Cooperative Recognition of Human Behaviors and Related Objects", Proc. 15th European Japanese Conf. on Information Modelling and

- Knowledge Bases, pp.321–330, May. 2005.
- [2] D. Moore, I. Essa, and M. Hayes: “Exploiting Human Actions and Object Context for Recognition Tasks,” Proc. Int. Conf. on Computer Vision 1999 (ICCV1999), pp.80–86, 1999.
 - [3] M. Higuchi, S. Aoki, A. Kojima, and K. Fukunaga: “Scene Recognition based on Relationship between Human Actions and Objects,” Proc. 17th Int. Conf. on Pattern Recognition (ICPR2004), Vol.3, pp.73–78, Aug. 2004.
 - [4] A. Gupta, and L. S. Davis: “Objects in Action: An Approach for Combining Action Understanding and Object Perception,” Proc. IEEE Computer Society Conf. on Computer Vision and Pattern Recognition (CVPR 2007), pp.1–8, Jun. 2007.
 - [5] 三木博史, 小島篤博, 宮本貴朗, 黄瀬 浩一: “DBN を用いた人物の動作パターンに基づく物体認識,” 画像の認識・理解シンポジウム (MIRU2008) 論文集, IS3–3, pp.877–884, Jul. 2008.
 - [6] A. Kojima, H. Miki, and K. Kise: “Object recognition based on n-gram expression of human actions,” Proc. 20th Int. Conf. on Pattern Recognition (ICPR2010), pp. 372–375, Aug. 2010.
 - [7] R. Hamid, A. Johnson, S. Batta, A. Bobick, C. Isbell, and G. Coleman: “Detection and explanation of anomalous activities: Representing activities as bags of event n-grams,” Proc. IEEE Computer Society Conf. on Computer Vision and Pattern Recognition (CVPR 2005), Vol.1, pp.1031–1038, 2005.
 - [8] C. Thureau, and V. Hlavac, V: “Pose primitive based human action recognition in videos or still images,” Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR 2008), pp.1–8, 2008.
 - [9] N. Dalal, and B. Triggs: “Histograms of Oriented Gradients for Human Detection,” Proc. IEEE Computer Society Conf. on Computer Vision and Pattern Recognition (CVPR 2005), Vol.1 pp.886–893, 2005.
 - [10] 笹野泰正, 小島篤博, 黄瀬浩一: “姿勢特徴の n -gram 表現を用いた人物行動による物体認識,” 信学技報, Vol. 110, No. 296, PRMU2010-114, pp.7–12, Nov. 2010.