

CHLAC 特徴量の錐制約部分空間表現による動作分類

佐藤 竜太[†] 亀田 能成^{†a)} 大田 友一[†]

Activity Recognition Based on Cone Restricted Subspace Expression of CHLAC Features

Ryuta SATO[†], Yoshinari KAMEDA^{†a)}, and Yuichi OHTA[†]

あらまし 本論文では、監視カメラに映る様々な物体の動作を、CHLAC 特徴量をもとに分類する方法を提案する。位置不変で加算可能な CHLAC は、映像解析における有力な画像特徴量の一つと考えられ、複数の対象の異なる動作をまとめて表現することができる。本論文では CHLAC 特徴量と相性の良い錐制約部分空間により動作を表現し、それにより単独で発生する動作と複数で発生する動作の分類を行えることを示す。歩行や自転車による移動、拾得動作を含む 6 種類の動作を組み合わせた動作に対する実験を通して、提案した単独の動作が発生している場合・複数種類の動作が発生している場合の分類器の性能評価を行い、提案手法の有効性を確認した。

キーワード 画像認識, 監視カメラ, サーベイランス, 複数動作認識

1. ま え が き

監視カメラにより撮影される映像を、その中に含まれる単独ないし複数の対象の総体的な動きによりアノテーション付けできれば、監視カメラ蓄積映像の検索や高次解析に役立つと考えられることから、監視カメラ映像に対する物体や動作認識への期待は高いものがある。

監視映像における主たる対象は人間や車両の行動や動作であるが、映像上での映り方や動作には様々な状況が想定されるため、その画像処理手法は汎用的である方が望ましい。そのため、我々は対象の形状や属性に依存しない特徴量に基づく、Cubic Higher-order Local Auto-Correlation (CHLAC) 特徴量 [4] に注目している。これは、画像全体の時空間情報を一度に記述できる有力な画像特徴量の一つである。

CHLAC 特徴量には、単純な時空間特徴量でありながら、画像全体に現れる時空間変動を一つの多次元ベクトルでユニークに記述できるという利点がある。また、複数物体の同時動作をその個々の動作に対応する

特徴ベクトルの和によって表現できるため、複数物体が出現しても表現形式を変える必要がないという、監視カメラでの動作認識に有用と思われる性質を有する。

本論文では、CHLAC 特徴量をもとに、種類の異なる物体が行う複数種類の動作が起こり得る場合に、一種類の動作が同時に発生した場合（単独発生動作）と、同時に複数種類の動作が発生した場合（複数発生動作）とについて分類方法を提案する。

単独発生動作においては錐制約部分空間法を用い、錐とクエリの間の角度による分類を行う。ここでは、加法に対して閉じている性質をもつ錐制約部分空間法が、同じく加法に対して閉じている CHLAC 特徴量と親和性をもつことを利用する。

一方、複数発生動作においては、単独発生動作分類と同様に、一つの動作を一つの錐制約部分空間で表現し、錐制約部分空間を複数用いて、入力映像から得られる CHLAC 特徴量のクエリベクトルを学習済み各動作の重み付きの和として表す。この動作の重みを既知の動作表現と比較することで、複数動作の分類を行う。

以降、**2.**では関連研究と、本研究で用いる CHLAC 特徴量について述べる。**3.**で単独発生動作に対する動作分類方法を述べ、それをもとにした複数発生動作に対する動作分類手法について**4.**で述べる。**5.**では、提案手法の実験結果を示しその有効性を考察する。最

[†] 筑波大学大学院システム情報工学研究科, つくば市
Graduate School of Systems and Information Engineering,
University of Tsukuba, Tsukuba-shi, 305-8573 Japan
a) E-mail: kameda@iit.tsukuba.ac.jp

後に 6. で本研究をまとめる。

2. 関連研究

監視カメラからの動作分類, 特に人間を対象とする動作分類の取組みは, 監視映像解析の取組みである TRECVID [7] の中でも Evaluation for Surveillance Event Detection として 2009 年度から継続して取り上げられているように, 盛んに研究が進められている分野である。

こうした監視カメラ映像内での人間行動の認識への方法論は多岐にわたる [8]. その中でも, Single-layered approaches に分類される手法は, 人間をはじめとするより様々な動作・運動の認識・分類に不可欠な研究分野である。

この分野は, 軌跡, 時空間解析, それに時空間特徴量に基づく解析の三つに大別される. ここでは, 軌跡を発生するとは限らない動作にも対応することを考え, 後者二つを取り上げる。

時空間解析を行う方法として, Efros ら [5] はカメラから距離をとって撮影されたシーンに合わせた認識手法を提案している. Ke ら [6] は映像から volumetric features を取り出して認識することで, 複雑で混雑した背景があっても人間の動作認識ができる手法を提案している. Wang ら [9] もまたスペクトルクラスタリングを用いた動作分類手法を提案している。

時空間特徴量に基づく解析としては, 局所特徴量ベースの手法 [10] や Wang ら [1] のように, 時空間に広がる局所特徴量を求めそれを BOF などの中間表現 [2] 等で統合した上で, 何らかの認識器 [3] と組み合わせることで動作分類を実現する方法が考えられている. [11], [12] 等も広い意味でこの方法論に含まれる研究として挙げられる。

これらの研究では単独動作のみを対象としており, シーンに同時に出現し得る複数の動作をそのまま扱うことはできない. 実際の監視カメラが撮影するシーンにおいては, 動作は単独で発生するとは限らないため, 複数発生動作の分類が可能であることが望ましい. しかし, このような物体ごとに記述を得るアプローチでは, 画像全体の記述に構造化が必要となり, その表現形式を一意に定めることは難しい。

これに対して本論文は, 画像全体の記述を直接得て動作分類する手法を提案する。

空間特徴量 HLAC [13] をベースに提案された時空間特徴 CHLAC [4] は, その特徴量を連続する画像全

体から一度に計算できること, 対象の形状や属性に対する予備知識が不要であることという性質の良さから今後の研究が期待される特徴量である. 時空間の局所の特徴量ベースでありながら, CHLAC はシーン中に出現する複数の動作の全てを含んだ形で表されるので, 画像全体の記述を得るための構造化が不要となることから, それに基づく提案動作分類法は [8] らの分類法に載らない新しい動作分類の方法論と考えられる。

CHLAC 特徴量の応用事例として, これまで, 動作認識 [14], [15] や異常検出 [16], [17] が行われてきている. 白木ら [14] は CHLAC を用いた動作識別を提案しているが, 複数動作識別には並列化した線形識別を用いており, CHLAC による直接的な複数動作分類の実現可能性には言及していない. 一方, 森下ら [15] は時間方向への伸縮などを考慮し CHLAC 利用の発展形について提案している. その中で 2 種類の動作についての認識実験も行っているが, その方法は一般化されていず, 複数種類の動作に拡張する方法は示されていない。

CHLAC を用いた異常検出のための手法は複数提案されている [16], [17] が, いずれも正常状態をクラスとしてそこからの逸脱を異常と定義しているため, 本研究目的のような複数発生動作分類には対応できない。

これらに対して, 本論文では CHLAC をもとに複数発生動作を分類する方法について提案する。

CHLAC はヒストグラム頻度に基づく特徴量ベクトルであることから, 複数の動作を観測した場合, それぞれの単独動作の重ね合わせである加算表現としてその特徴量ベクトルが得られる. CHLAC に組み合わせる複数発生動作の識別器は, この加算表現を解析できる必要がある. このことは, SVM [3] をはじめとする単純な多クラス識別器が利用できないことを意味する. これに対して, 部分空間表現であれば特徴量ベクトルの操作に様々な条件を設けることができる [19]~[21]. 本論文では特に, CHLAC の加法表現という特徴と各要素が正值であるという特徴に合致できる錐制約部分空間表現 [18] に注目する。

3. 錐制約部分空間表現に基づく単独動作分類

3.1 錐制約部分空間表現

本論文では, CHLAC 特徴量ベクトルに基づくクラス分類手法に, 部分空間法の拡張で小林らにより提案されている錐制約部分空間法 [18] を用いる. CHLAC

特徴量などのヒストグラム値を用いる特徴量は非負であり、錐形状に分布をなすため、錐制約部分空間法が利用可能である。

本論文においては、一つの動作を一つのクラスとし、一つの錐制約部分空間である凸錐で表現する。それぞれの動作に対応する錐制約部分空間は、学習により算出する。ある一つの動作を撮影した V 本の映像群を用意し、この映像群から算出された N_{CHLAC} 本の CHLAC 特徴量ベクトルを時系列にかかわらず、全てそのままその動作の学習データとして用いる。学習データから、錐制約部分空間である凸錐を計算する。

分類を行う際には、ある入力映像を構成するフレームから算出された複数のクエリベクトルをクエリベクトル集合 G_q とする。 G_q に属するすべてのベクトルについて分類を行い、結果とする。

3.2 厳密凸錐による錐制約部分空間の構成

錐制約部分空間法における錐制約部分空間の構成方法は厳密凸錐・包括凸錐・円錐近似の3種類があるが、本論文では分類性能が良いとされる厳密凸錐 [18] を用いる。

クラス C_i の厳密凸錐 S_i を構成するためには、 C_i に含まれる学習ベクトル群 \mathbf{x}_i を用いて、以下のステップに従い計算する。

- (1) C_i に属する学習ベクトル群の部分集合 C_{is} とその補集合 $C_{i\bar{s}}$ をランダムに作成
- (2) 学習ベクトル群の部分集合 C_{is} から、各 CHLAC 特徴量ベクトル $\mathbf{x}_j \in C_{is}$ に対して、その CHLAC 特徴量ベクトル \mathbf{x}_j を除いた凸錐への角度 θ_j を算出
- (3) しきい値 τ_θ に対して $\theta_j < \tau_\theta$ となる学習ベクトルを従属しているものとして、 C_i 及び C_{is} より除去
- (4) 学習ベクトル $\mathbf{x}_k \in C_{i\bar{s}}$ に対して、 C_{is} のなす凸錐 S_{is} への角度 θ_k を算出し、(3) と同様に $\theta_k < \tau_\theta$ となる学習ベクトルを従属しているものとして C_i から除去
- (5) C_i からの除去がなくなるまで (1~4) を複数回実行

これによって得られた学習ベクトル集合をクラス C_i の厳密凸錐 S_i として扱う。

3.3 錐制約部分空間法による分類

本節では、クエリベクトル \mathbf{x}_q を分類する方法について説明する。クエリベクトル \mathbf{x}_q とクラス C_i に属する凸錐 S_i の分類には、その二者のなす角 θ_i を評価

値として用いる。この評価値 θ_i に対してクラス C_i におけるしきい値 τ_i を設定し、それによるクエリベクトルの凸錐 S_i に対する判定を行う。クエリベクトル \mathbf{x}_q のクラス C_i での評価値 θ_i が以下の条件を満たすとき、そのクエリベクトル \mathbf{x}_q をクラス C_i に属していると判定する。

$$\mathbf{x}_q \in C_i \quad \text{if} \quad \theta_i = f(\mathbf{x}_q, S_i) < \tau_i \quad (1)$$

全てのクラスへの判定は並行して行われるので、一つのクエリベクトルは複数クラスに分類され得る。また、どのクラスにも分類されないという状況も存在し得る。

4. 複数発生動作分類

複数発生動作分類においては、一つのクエリベクトルは複数の動作ベクトルの加算形である。そのため、3. で述べた、錐制約部分空間の成す角度を用いた分類を直接適用することはできない。本論文では、クエリベクトルを、動作クラスを代表するベクトルに分解し、その加算係数を算出することにより、複数発生動作分類を行う。

動作クラスを代表する動作代表ベクトルの選定には錐制約部分空間を使用する。各動作クラスを表す凸錐を構成するベクトルを、動作代表ベクトルとして取り扱う。この方法により、学習データには、3. と同じく単独の動作を別々に撮影した映像群を用いることができる。

なお、動作代表ベクトルについては、学習時の CHLAC 特徴量ベクトルの全てをそれに相当させることも考えられるが、凸錐を構成するベクトルを用いる場合に比べて数が増えるため、動作推定時に数値不安定になる可能性がある。そのため、動作表現の範囲を狭めない限りにおいて凸錐を構成し、動作代表ベクトルの削減を図る。

以降、まず動作代表ベクトルが何らかの形で得られているという前提で複数動作発生時の加算表現とそれに基づく動作分類手順を 4.1 と 4.2 で示し、そのあとで本論文で提案する動作代表ベクトルの選択とそれに伴う動作近似について 4.3 で述べる。

4.1 複数発生動作の加算表現

M_{act} 個の単独発生動作を表現する CHLAC 特徴量ベクトル (以下、動作代表ベクトル) を $\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_{M_{\text{act}}}$ とすると同時発生した複数発生動作を観測して得られるクエリベクトルとなる CHLAC 特徴量 \mathbf{q} は、CHLAC

特徴量がヒストグラム特徴量であり加算可能であることから、動作代表ベクトルの加算で表されると期待できる。

$$\mathbf{q} = \sum_{i=1}^{M_{\text{act}}} k_i \mathbf{a}_i \quad (2)$$

ここで、 k_i は i 番目のクラスに対する重みであり、加算のみが可能なため、 $k_i \geq 0$ である。このことは、観測された複数動作が全て既知のものであれば成立する。言い換えれば、重み k_i をクエリベクトル \mathbf{q} から算出できれば、どの動作により構成された複数発生動作なのかが推定可能である。

以降、この重み k_i を求める方法について説明する。

4.2 誤差最小化による動作分類

今、動作代表ベクトル \mathbf{a}_i は互いに独立であると仮定しよう。もし M_{act} が CHLAC 特徴量ベクトルの次元数である 251 より小さいとすると、任意の \mathbf{q} に対しては、

$$\mathbf{q} = \sum_{i=1}^{M_{\text{act}}} k_i \mathbf{a}_i \quad (3)$$

を満たす k_i を求めることはできない。

しかし、 \mathbf{q} はもともと $\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_{M_{\text{act}}}$ の線形合成（加算合成）と期待されているため、この条件下においては、 k_i を一意に求めることができる。実際の観測においては、未知の動作やノイズなども含まれるので、ここでは最小化問題として k_i を求める。

本論文においては、ある重みにより得られたベクトルとクエリベクトルとの差を、最小化する重みを求める問題として取り扱う。以降で具体的な方法を説明する。

まず、重み k_i について重みベクトル $\mathbf{k} = [k_1, k_2, \dots, k_m]$ とする。ここで、ある重みベクトル \mathbf{k} に対する推定誤差 $e(\mathbf{k})$ を以下のように定義する。

$$e(\mathbf{k}) = \left\| \mathbf{q} - \sum_{i=1}^{M_{\text{act}}} k_i \mathbf{a}_i \right\| \quad (4)$$

この推定誤差 e を最小にする重みベクトル \mathbf{k} を求める。

$$\arg \min_{\mathbf{k}} e(\mathbf{k}) = \arg \min_{\mathbf{k}} \left\| \mathbf{q} - \sum_{i=1}^{M_{\text{act}}} k_i \mathbf{a}_i \right\| \quad (5)$$

ただし、 k_i は全て非負とする。

k_i の値が大きければ、動作 i が入力映像中に存在している可能性が高いと考えることができる。

4.3 錐制約部分空間表現による動作近似

4.2 にて説明した手法により、複数の動作によって構成された CHLAC 特徴量ベクトルを、それぞれの動作に分解することが可能となる。実際には、一つの動作に複数の代表動作ベクトルを用い、各動作の複雑さに対応する。

本論文では錐制約部分空間法により、ある動作の学習データに含まれる CHLAC 特徴量ベクトル群をよく近似する錐 C_i を求め、その錐をなす N_{cone}^i 個のベクトル群 \mathbf{a}_{ij} を一つの動作の動作代表ベクトル群とする。

この錐制約部分空間には、包括凸錐 [18] を用いる。これは、凸錐を少数の基底ベクトルからなる包括的な凸錐で近似するものである。

錐制約部分空間の表現として、これとは別に厳密凸錐があるが、ここでは、一つの CHLAC 特徴量ベクトルが複数の単独動作の重ね合わせによって構成されており、その際には多少の変動要因が含まれ得ることから、包括凸錐で十分であると考えられる。また、これによって、厳密凸錐に比べて動作代表ベクトル群の算出にかかる計算コストの低減が見込まれる [18]。

まず、学習データベクトル群 \mathbf{x}_i に対して正規化を行い、単位超球面上に射影し、その自己相関行列の固有値・固有ベクトルを求める。固有値 λ_i とそれに対応する固有ベクトル \mathbf{e}_i を降順に並べたとき、第 1 固有ベクトルは凸錐の原点からの方向を示すベクトルとなり、それ以降のベクトルは超球面上の分布の広がりを示すと考え、2 番目以降の固有軸上で分布を包括する。よって、一つの固有ベクトル \mathbf{e}_i ($i \geq 2$) はスケーリングのパラメータである。また、次元数 d_i の決定に

$$\eta(i) = \frac{\sum_{j=2}^i \lambda_j}{\sum_{j=2}^{251} \lambda_j} \quad (6)$$

累積寄与率 η_i が、しきい値 τ_η に対して条件

$$\eta(d_i) < \tau_\eta < \eta(d_i + 1) \quad (7)$$

を満たすような d_i を選択する。また、そのような条件をもつ d_i が選択できない場合、第一主成分ベクトル \mathbf{e}_1 のみを動作代表ベクトル群として用いる。

ある瞬間の単独発生動作 i から算出された CHLAC 特徴量ベクトル \mathbf{s}_i は、包括凸錐 S_i の内部に存在するため、その動作代表ベクトル群 \mathbf{a}_{ij} の重み付き加重和

によって表すことができる.

$$\mathbf{s}_i = \sum_{j=1}^{N_{\text{cone}}^i} k_{ij} \mathbf{a}_{ij} \quad (8)$$

このため, 複数の動作から算出された CHLAC 特徴量ベクトル \mathbf{q} は, \mathbf{s}_i の重み付きの加重和で表せることが期待できる.

$$\mathbf{q} = \sum_{i=1}^{M_{\text{act}}} k_i \mathbf{s}_i \quad (9)$$

$$= \sum_{i=1}^{M_{\text{act}}} \sum_{j=1}^{N_{\text{cone}}^i} k_{ij} \mathbf{a}_{ij} \quad (10)$$

ただし, $k_{ij} \geq 0$ である.

この複数ベクトルからなる動作代表ベクトル群による推定誤差 e_2 は以下のように定義できる.

$$e_2(\mathbf{k}) = \left\| \mathbf{q} - \sum_{i=1}^{M_{\text{act}}} \sum_{j=1}^{N_{\text{cone}}^i} k_{ij} \mathbf{a}_{ij} \right\| \quad (11)$$

誤差関数 e の代わりに e_2 を用いて, 式 (5) による最小化を行うことで, 動作代表ベクトル群による動作推定を行う.

このとき, k_{ij} が一意に定まるためには, \mathbf{a}_{ij} が全て独立であること, \mathbf{a}_{ij} の加算和で \mathbf{q} が構成されていることが条件となる.

ここで, 代表ベクトルが一つの動作に対し複数存在するため, 推定される重み k_{ij} も一つの動作に対し複数存在する. そのため, クエリベクトルにおけるそれぞれの動作の存在を知るためには, 一つの動作に対する評価値を求める必要がある. 本論文においては, それぞれの動作の動作代表ベクトルに対する重みを合計することで, 入力クエリベクトルの動作 \mathbf{s}_i に対する評価 J_i とする.

$$J_i = \sum_{j=1}^{N_{\text{cone}}^i} k_{ij} \quad (12)$$

ここに, しきい値 τ_J を設定することで, 分類を行う. 評価値 J_i がしきい値 τ_J を超えた場合に, クエリベクトル中に動作 \mathbf{s}_i が存在しているものとする.

5. 実験と考察

提案したそれぞれの手法について性能評価実験を

行った. 実験映像は本学内の 4 階屋上の高さから周囲の駐車場を撮影したものである. 監視カメラとして SONY BRC-H700 を用いた. 映像は横 640 画素縦 360 画素の解像度であり, グレースケールで 30 [fps] である.

5.1 評価用データセット

実験で取り上げた動作 (実験対象動作) は, 自転車の左方向への移動 (自転車左), 自転車の右方向への移動 (自転車右), 人間の左方向への歩行 (歩行者左), 人間の右方向への歩行 (歩行者右) への歩行, ものを拾う動作 (拾得動作) の体を上へ動かす動作 (拾得上), 拾得動作の体を下へ動かす動作 (拾得下) の 6 種類の動作である. ただし, ここでの左・右は映像中における左方向・右方向である. それぞれの動作は, 同じカメラからほぼ同一の撮影条件で撮影された.

学習用映像データとして, 歩行者・自転車の動作を 10 回, 拾得下・拾得上動作は一つひとつの動作が短いため 20 回撮影した. 歩行動作左のあるフレームを図 1 に示す. なお, 学習データセットに含まれる動作を行っている人物は全て一人の人物である (図 2). CHLAC 特徴量を算出する際に参照するフレーム間差分画像の例を図 3 に示す.

学習データとしての各動作における処理対象フレームの数, すなわち CHLAC 特徴量ベクトルの本数



図 1 歩行動作左 (入力)
Fig.1 Walking action, left (input).

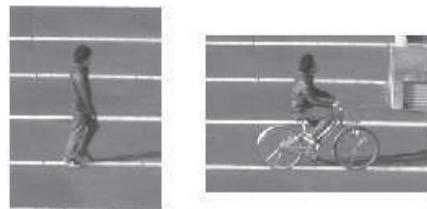


図 2 学習データ中の人物
Fig.2 A person in learning dataset.

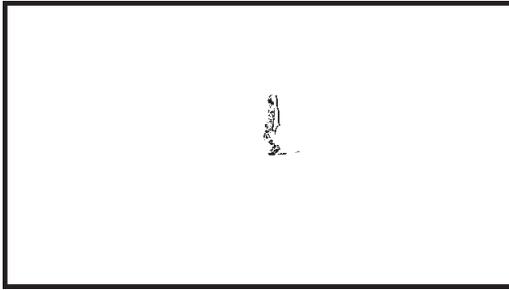


図3 歩行動作左のフレーム間差分
Fig. 3 Walking action, left (input).

表1 各動作の学習データにおけるフレーム数
Table 1 Frames of learning dataset of actions.

歩行左	歩行右	自転車左	自転車右	拾得上	拾得下
3131	3326	778	887	525	513

表2 各単独動作クエリにおけるフレーム数
Table 2 Frames of each single action for query.

歩行左	歩行右	自転車左	自転車右	拾得上	拾得下
321	581	119	110	131	94

表3 それぞれのクエリデータにおけるフレーム数
Table 3 Frames of combined actions for query.

クエリ名	フレーム数	クエリ名	フレーム数
歩行左+歩行右	226	歩行左+自転車左	182
歩行左+自転車右	111	歩行左+拾得上	81
歩行左+拾得下	91	歩行右+自転車左	204
歩行右+自転車右	215	歩行右+拾得上	135
歩行右+拾得下	93	自転車左+拾得上	75
自転車左+拾得下	152	自転車右+拾得上	98
自転車右+拾得下	56		

N_{CHLAC} は表1のとおりである。

5.1.1 単独発生動作クエリ

単独発生動作のクエリとして、実験対象動作が単独で発生しているものを用意した。それぞれの動作クエリにおいては、歩行左・歩行右・自転車左・自転車右の4動作が2回、拾得上と拾得下が10回繰り返されている。動作を行っている人物は1人で、学習データの人物とは別の人物である。それぞれの動作におけるフレーム数、つまりクエリ数を表2に示す。

5.1.2 複数発生動作クエリ

複数発生動作のクエリは、実験対象動作を表3のように2種類組み合わせたものである。それぞれの動作を1人の人物が行い、計2人の人物がクエリ中に存在する。クエリ中の動作は全て同じ2人の人物が行っているが、いずれも学習データとして撮影された人物と

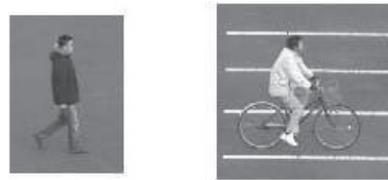


図4 クエリ中の人物
Fig. 4 People in query.

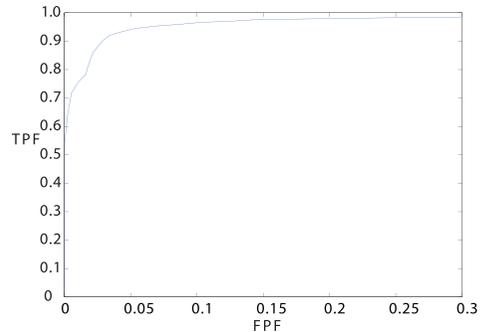


図5 単独動作分類におけるROC曲線
Fig. 5 ROC of single action classification.

以下で説明するクエリデータとして撮影された人物は別の人物である(図2・図4)。

5.2 単独発生動作分類法の評価

クエリには5.1で説明したものを用い、学習データからは、厳密凸錘を $\tau_\theta = 0.05$ [rad] として構成し用いた。

図5にROC曲線を示す。縦軸は真陽性率 (TPF) であり、正解クラスを正解として出力できた割合を示す。横軸は偽陽性率 (FPF) で不正解クラスを誤って正解として出力してしまった割合である。グラフ中の曲線は、表2中の1356本の全ての単独発生動作クエリにおいて、角度に対するしきい値 τ_i を0.00から1.00まで変化させた場合の疑陽性率・真陽性率の軌跡である。軌跡が左上にあれば良い性能を示している。真陽性率を求めるためにはそれぞれのクエリを用い、偽陽性率を求めるためには残りのクエリを全て用いた。ROC曲線から、偽陽性率が5%のとき、94.1%の認識率を得られていることが分かる。

5.3 複数発生動作分類法の評価

4.で説明した複数発生動作分類手法による分類結果を図6にROC曲線として示す。それぞれの直線は、凡例で示されている数値を累積寄与率に対するしきい値 τ_j として用いた場合に、評価値 J_i に対するしきい

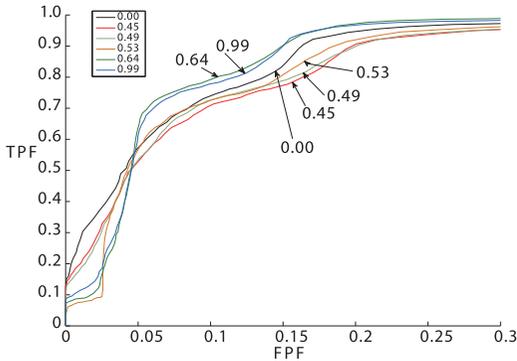


図 6 複数発生動作分類における ROC 曲線
Fig. 6 ROC of combined action classification.

表 4 累積寄与率と代表ベクトルの数
Table 4 Number of representative vectors w.r.t. τ_η .

τ_η	歩行左	歩行右	自転車左	自転車右	拾得上	拾得下
0	1	1	1	1	1	1
0.45	1	1	1	4	1	1
0.49	1	2	1	4	2	2
0.53	2	2	1	6	2	2
0.64	4	4	4	12	6	6
0.99	272	288	274	310	280	256

値 τ_J を 0.00 から 1.00 まで変化させた場合の偽陽性率・真陽性率の軌跡である。真陽性率・偽陽性率の算出に際しては、どちらも先述した複数発生動作クエリデータをすべて用いた。また、それぞれの累積寄与率に対するしきい値 τ_J において動作代表ベクトルが用いられた数を表 4 に示す。 $\tau_\eta = 0.00$ の場合は全ての動作クラスにおいて 1 本の動作代表ベクトルしか用いられていないため、これは 3. で述べた手法に等しい。しきい値 τ_η が 0.53 程度までは、0.00 の場合と比べて性能が悪化しているが、0.64 程度まで高くなると、偽陽性率が 0.05 を超える部分において性能改善が認められる。これは、高いしきい値 τ_η により動作代表ベクトルの数が増加し、動作代表ベクトル群が学習データの分布をよく近似できるようになったことが理由と考えられる。

5.4 結果

$\tau_\eta = 0.99$ の場合のそれぞれのクエリの推定誤差 e_2 を表 5 に示す。どのクエリにおいても、 e_2 はある程度小さな範囲に収まっている。これは、各クエリが実験対象 6 動作の錐制約部分空間により、よく表せていることを示唆している。

図 6 の偽陽性率が 0.05 より小さい部分で表されるように、偽陽性率が低い場合に、累積寄与率に対する

表 5 $\tau_\eta = 0.99$ でのそれぞれのクエリにおける推定誤差 e_2 の平均・分散

Table 5 Average and variance of estimation residual e_2 at $\tau_\eta = 0.99$.

クエリ名	平均	分散 [$\times 10^{-4}$]
歩行者左+歩行者右	0.0696	26.0
歩行者左+自転車左	0.0497	2.98
歩行者左+自転車右	0.0572	2.84
歩行者左+拾得上	0.0905	6.08
歩行者左+拾得下	0.1153	74.0
歩行者右+自転車左	0.0469	8.19
歩行者右+自転車右	0.0618	2.27
歩行者右+拾得上	0.1320	42.0
歩行者右+拾得下	0.1048	9.66
自転車左+拾得上	0.0961	0.199
自転車左+拾得下	0.1034	66.00
自転車右+拾得上	0.0879	0.813
自転車右+拾得下	0.0852	0.677

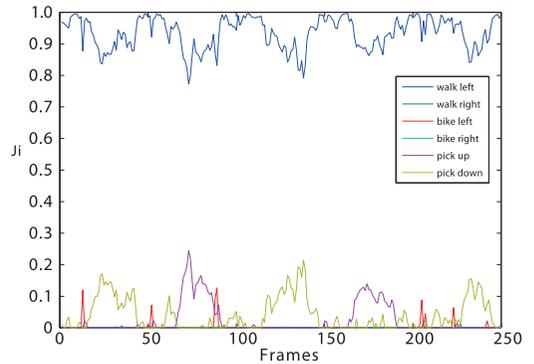


図 7 歩行左・拾得動作が同時発生しているクエリの評価値
Fig. 7 Query evaluation of “walking action, left” and “picking-up” actions.

しきい値 τ_η を高くとった場合でも認識率が低いところがあったため、原因を調査し、これは後に述べるように主に拾得動作が原因と判明した。

この確認のため、歩行左と拾得上・拾得下動作が同時に発生している動作のクエリを入力とし、それぞれの動作における評価値 J_i を時系列に並べたグラフを図 7 に示す。拾得動作は、体を下げて上げる動作のため、拾得下と拾得上が交互に現れている。ここで、拾得動作の評価値が歩行左動作の評価値と比べて低い。この原因を考察する。

学習データにおける拾得動作・歩行左動作の平均 CHLAC 特徴ベクトルのノルムは、それぞれ歩行左動作が 1099、拾得上動作が 235、拾得下動作が 219 となっている。そのため、歩行左動作は拾得上・拾得下動作に比べ 5 倍程度の大きさをもっている。それに対し、代表ベクトル群を算出する際に用いる包括凸錐で

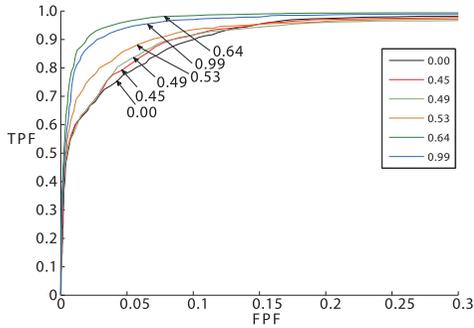


図 8 歩行・自転車のみ複数発生動作分類による ROC 曲線

Fig. 8 ROC of combined action classification only of walking and bicycling.

は、超球に射影する際すべてのベクトルの大きさを正規化するため、この大きさの差が反映されない。そのため、重み k_{ij} はクエリベクトル中に含まれる各動作のノルムの違いを反映してしまう。このため、拾得動作の評価値が低くなってしまふ原因は、拾得動作のノルムが他の動作に比べ小さいことためだと考えられる。

そこで、この考えが正しいかどうかを確認するために、比較的ノルムが同じ大きさの歩行左・歩行右・自転車左・自転車右の4動作による組合せのみをクエリとして用い、拾得上・拾得下を使用した組合せのクエリを取り除いた。他の条件は全て先の実験と同一である。その場合の ROC 曲線を図 8 に示す。累積寄与率に対するしきい値 τ_η を高くとった場合に認識率が改善されている。このことから、図 6 で示した実験での偽陽性率が低い原因として、ノルムが他と比べて小さい拾得動作の影響が大きかったということがうかがえる。

次に、4.での提案手法の前提である、学習データに対する a_{ij} の独立性を考察する。独立性を考察するために、動作代表ベクトル群内での推定誤差を参照すること考える。これは、本実験で用いられた動作代表ベクトル群 a_{ij} から、ある一つのベクトルを取り除き、取り除かれたベクトルをクエリベクトルとして、推定誤差を次に述べる評価関数を用いて算出したものである。具体的に、 α 番目の動作の β 本目のベクトルを取り除いた場合の推定誤差 $c_{\alpha\beta}$ は次の式により算出される。

$$c_{\alpha\beta} = \min_k \left(\left\| \mathbf{a}_{\alpha\beta} - \sum_{i=1}^{i \neq \alpha} \sum_{j=1}^{j \neq \beta} k_{ij} \mathbf{a}_{ij} \right\| \right) \quad (13)$$

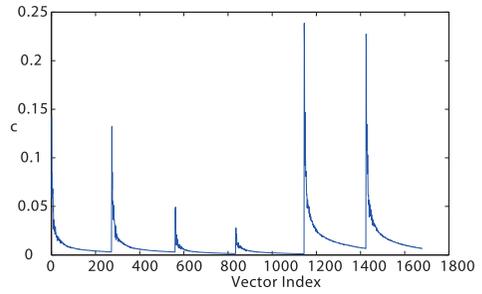


図 9 動作代表ベクトルに対する $c_{\alpha\beta}$

Fig. 9 $c_{\alpha\beta}$ of action representative vectors.

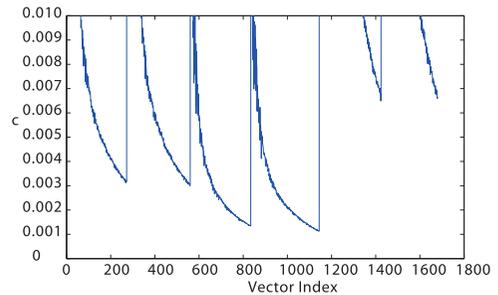


図 10 図 9 の $c_{\alpha\beta} = 0.00$ 近傍の拡大

Fig. 10 Enlarged graph of Fig. 9 around $c_{\alpha\beta} = 0.00$.

実際に本研究で用いたしきい値 $\tau_\eta = 0.99$ の場合の 1680 本の動作代表ベクトル群における推定誤差 $c_{\alpha\beta}$ を図 9 に示す。また、図 9 において $c_{\alpha\beta} = 0.00$ 近傍を拡大したものを図 10 に示す。1680 本の動作代表ベクトル群中、最小の $c_{\alpha\beta}$ は約 0.0011 であった。

このことは、 $c_{\alpha\beta} = 0.000$ となるようなベクトルが存在していなかったことを意味する。すなわち、他のベクトルの重み付き加重和で表せてしまうようなベクトルは動作代表ベクトル群 a_{ij} には含まれず、本実験における推定結果は不定に陥りにくいことが分かる。

最後に、クエリベクトル \mathbf{q} を本実験で使用した 6 種類の動作による動作代表ベクトル群 a_{ij} の重み付き加法で表現できていることの検証を行う。この 6 種類の動作による重み付き加重和による表現が不可能な、本実験に含まれない動作を分類させた場合の推定誤差 e_2 は比較的大きな値となることが予想される。

重み付き加重和で表現できない場合の推定誤差 e_2 に比べて、重み付き加重和で表現できる本実験で使用したクエリの推定誤差 e_2 が十分に小さければ、予想が裏づけられることになる。そこで、本実験で使用した 6 種類の動作とは異なる動作として画面中で上から下へ走行する車を撮影 (図 11) し、その動作による



図 11 自動車動作
Fig. 11 Vehicle action.

CHLAC 特徴量ベクトルを自動車動作クエリベクトルとして推定誤差 e_2 を算出し比較した。自動車動作クエリのフレーム数は 84 である。

この結果は、平均 0.255, 分散 0.0017 であった。この自動車動作クエリの推定誤差 e_2 は表 5 で示した実験で用いたクエリに対して大きい。そのため、自動車動作は学習した 6 動作の重み付き加重和により表現不能であることが示されているといえる。

ここで、複数動作の種類数と、提案手法の性能との関係について考察する。

本提案手法では、動作の種類が増えた場合、それに対応する凸錐を構成するベクトルの演算は、動作の種類の数とその学習用ベクトルの数によって時間がかかることになるが、厳密凸錐・包括凸錐のいずれも動作の種類に対して指数爆発する計算量にはならない。

クエリベクトルからの動作分類時には、その加算表現の分解にかかる計算コストが動作の種類数の影響を受けるが、4. のアルゴリズムに示すように、動作種類数に対して著しい計算コスト増になることはない。

もう一つの懸念は、動作種類数が増えるにつれて動作近似における各ベクトルの独立性が失われ分解が不能になることであるが、見かけが異なる動作である限り CHLAC の 251 次元の特徴量ベクトルでは同じ表現に陥る可能性は低いと考えられる。

よって、動作種類数の増加に対しても本手法は適応可能であるといえる。

5.5 比較対照実験

提案手法の有効性を示すために、映像から Wang らの提案する局所特徴量 [1] を求め、そこから BOF [2] による映像全体の表現を得て、マルチクラス SVM [3] で認識を行った結果を本節で示す。以後、この組合せによる手法をここでは STIP 法と呼ぶ。

これらの手法はいずれも検証実験のためのプログラムが公開されており、本論文で特に断らない限り、これらの手法の利用において必要なパラメータ値は公開されている状態の値を利用するものとする。

学習に用いた映像データは表 1 と同じである。全ての映像について局所特徴量を求めたあと、BOF を構

表 6 STIP 法による単独発生動作の分類正答率
Table 6 Result of single action by STIP method.

動作	学習データ [%]	クエリデータ [%]
歩行左	99.9	98.4
歩行右	69.0	64.6
自転車左	33.8	35.9
自転車右	37.5	32.6
拾得上	8.8	0.0
拾得下	22.0	0.0
総計	74.2	57.1

表 7 STIP 法による複数発生動作の分類正答率
Table 7 Result of combined action by STIP method.

動作	正答率 [%]
歩行左	71.6
歩行右	35.0
自転車左	12.5
自転車右	13.2
拾得上	0.0
拾得下	6.8
総計	38.9

成する。このとき、あとで用いる映像特徴量ベクトル長を CHLAC 特徴量ベクトルの 251 次元に合わせるため、BOF の構成要素数を初期値の 100 個から 251 個に変更した。その上で、六つの動作について、マルチクラス SVM で分類器を構成した。

性能評価には学習データそのもののほか、表 2 の単独発生動作クエリと、表 3 の複数発生動作クエリを用いた。

表 6 に単独発生動作に対する分類結果を示す。拾得動作は上下とも 1 動作の映像時間が短いため、局所特徴量自体の検出が行えていないことから、低い分類率となった。また、全般に学習データに対しても正答率が低い傾向にある。これらについては、パラメータチューニングにより改善される可能性がある。しかし、一方で CHLAC 特徴量に錐制約部分空間表現を組み合わせる提案手法では、そもそも調整すべきパラメータ数が少なく、結果も図 5 に見られるような分類成功率を示していることに注意されたい。

表 7 では、複数発生動作に対する分類結果を示す。この表では、複数発生動作クエリに対してマルチクラス SVM から得られる第 1 位の回答がクエリ中の指定した側と一致した場合を正答としているため、第 2 位までを正答と解釈する場合に比べて評価が低くなっていることを断っておく。この場合の複数発生動作クエリ全体に対する正答率は 38.9% であった。一方、複数発生動作クエリにおいて、回答がそのクエリに含まれる二つの動作のうちどちらかである場合を正答と

した場合、複数発生動作クエリ全体に対する正答率は88.5%であった。このことから、第2位までを正答とする場合は、この二つの正答率の間にSTIP法の性能が見込めると考えられる。

6. む す び

本論文では、監視カメラ画像全体を一つのシーンと捉え、そのシーン中で種類の異なる物体が行う複数種類動作を総体的に記述するCHLAC特徴量を用いて、単独ないし複数発生動作の分類方法を提案した。

本提案手法により、発生した動作を分類できるようになるので、監視カメラ映像のシーン記述の手段として使われていくことが期待される。

本手法では、単独の種類動作が同時に発生している場合に分類する手法と、複数の種類の動作が同時に発生している場合に分類する手法の2種類を提案しているが、これらは前処理の時点でCHLAC特徴量の総量から単独か複数物体かを判別して適宜適用することが可能である。

今後の課題として、より多数種類・多数の動作に対する性能評価を行うこと、拾得動作のようなCHLAC特徴量のノルムが小さな動作が含まれている状況に自動的に対応することなどが挙げられる。

謝辞 筑波大学大学院システム情報工学研究科修士課程の並木重哲氏には、比較提唱実験とデータ解析に協力を頂いた。また、本研究の一部は、日本学術振興会基盤研究(B)「環境カメラ群映像の安心かつ効率的に見える化の為の時空間解析と複合現実感的可視化」(課題番号23300064)の支援を受けて行われた。ここに謝意を表する。

文 献

- [1] H. Wang, M.M. Ullah, A. Klaser, I. Laptev, and C. Schmid, "Evaluation of local spatio-temporal features for action recognition," British Machine Vision Conference, 2009.
- [2] M. Aly, M. Munich, and P. Perona, "Indexing in large scale image collections: Scaling properties and benchmark," IEEE Workshop on Applications of Computer Vision, (WACV), 2011.
- [3] C.C. Chang and C.J. Lin, "LIBSVM: A library for support vector machines," ACM Trans. Intelligent Systems and Technology, vol.2, pp.27:1-27:27, 2011.
- [4] N. Otsu, "Towards flexible and intelligent vision systems - From thresholding to CHLAC," IAPR Conference on Machine Vision Application, pp.430-439, 2005.
- [5] A.A. Efros, A.C. Berg, G. Mori, and J. Malik, "Recognizing action at a distance," 9th IEEE International Conference on Computer Vision, pp.726-733, 2003.
- [6] Y. Ke, R. Sukthankar, and M. Hebert, "Volumetric features for video event detection," Int. J. Comput. Vis., vol.88, no.3, pp.339-362, 2010.
- [7] A.F. Smeaton, P. Over, and W. Kraaij, "Evaluation campaigns and TRECVID," 8th ACM International Workshop on Multimedia Information Retrieval, pp.321-330, 2006.
- [8] J.K. Aggarwal and M.S. Ryoo, "Human activity analysis: A review," ACM Comput. Surv., vol.43, issue.3, 2011.
- [9] X. Wang, L. Wang, and A. Wirth, "Pattern discovery in motion time series via structure-based spectral clustering," Comput. Vis. Pattern Recognit., pp.1-8, 2008.
- [10] 村井泰裕, 藤吉弘亘, 金出武雄, "Space-Time Patchを用いた物体の移動方向識別とセグメンテーション," MIRU講演論文集, pp.389-396, 2007.
- [11] J.C. Niebles, H. Wang, and F. Li, "Unsupervised learning of human action categories using spatial-temporal words," Int. J. Comput. Vis., vol.79, no.3, pp.299-318, 2008.
- [12] 木谷クリス真実, 岡部孝弘, 佐藤洋一, 杉本晃宏, "視覚的文脈を考慮した人物動作カテゴリの教師無し学習," MIRU講演論文集, pp.28-33, 2008.
- [13] N. Otsu and T. Kurita, "A new scheme for practical, flexible and intelligent vision systems," Proc. IAPR Workshop on Computer Vision, pp.431-435, 1988.
- [14] 白木孝義, 石黒勝彦, 深野 亮, 鴨志田良和, 白井達也, 斉藤秀雄, 田浦健次郎, 大武美保, 佐藤知正, 大津展之, "CHLAC特徴とGridコンピューティングを併用したリアルタイム動作認識," 信学技報, PRMU2005-218, 2006.
- [15] 森下雄介, 小林 匠, 森崎巧一, 大津展之, "時間重みと外的規準を用いた動作評価手法," CVIM, vol.2008, no.27, pp.47-52, 2008.
- [16] 南里卓也, 大津展之, "複数人動画からの異常動作検出," CVIM, vol.2004, no.91, pp.173-186, 2004.
- [17] 村井泰裕, 藤吉弘亘, 数井誠人, "時空間特徴に基づくエスカレータシーンにおける人の異常行動検知," 信学技報, PRMU2008-87, 2008.
- [18] 小林 匠, 大津展之, "パターン識別のための錐制約部分空間法," 信学論(D), vol.J92-D, no.1, pp.104-111, Jan. 2009.
- [19] 福井和広, 山口 修, "一般化差分部分空間に基づく制約相互部分空間法," 信学論(D), vol.J87-D, no.8, pp.1622-1631, Aug. 2004.
- [20] 福井和広, 山口 修, "カーネル非線形制約相互部分空間法による物体認識," 信学論(D), vol.J88-D, no.8, pp.1349-1356, Aug. 2005.
- [21] P. Favaro and A. Ravichandran, "A closed form solution to robust subspace estimation and clustering," Comput. Vis. Pattern Recognit., pp.1801-1807, 2011.

(平成23年6月6日受付, 10月4日再受付)



佐藤 竜太

2011 筑波大学大学院システム情報工学研究科博士前期課程了。現在日本電気(株)勤務。在学中は時系列画像処理の研究に従事。



亀田 能成 (正員：シニア会員)

1991 京都大学大学院博士後期課程認定退学。京都大学助手，2001～2002 米国 MIT 客員研究員(併任)，筑波大学講師を経て，2004 同大学院システム情報工学研究科助教授，現准教授。京都大学博士(工学)。知的画像処理，マルチメディア処理，複合現実感技術等の研究に従事。情報処理学会，バーチャルリアリティ学会，画像電子学会，IEEE 各会員。



大田 友一 (正員：フェロー)

1977 京都大学大学院博士課程了。京都大学情報工学科助手，筑波大学電子・情報工学系講師，カーネギーメロン大学計算機科学科客員研究員，筑波大学電子・情報工学系助教授を経て，1992 同教授。2004 同大学院システム情報工学研究科教授。2009 より，同研究科研究科長。工博。コンピュータビジョン，視覚情報メディア，複合現実感の研究に従事。2008～2009 年度本会理事。2009 年度日本 VR 学会論文賞受賞。パターン認識国際連盟フェロー，情報処理学会フェロー。