

CHLAC 特徴量と錘制約部分空間法による動作分類

佐藤 竜太[†] 亀田 能成[†] 大田 友一[†]

[†] 筑波大学大学院システム情報工学研究科

〒 305-8573 茨城県つくば市天王台 1-1-1

E-mail: †rsato@image.iit.tsukuba.ac.jp, ††{kameda,ohta}@iit.tsukuba.ac.jp

あらまし 本稿では、監視カメラに映る様々な物体の動作を、CHLAC 特徴量をもとに識別する方法を提案する。位置不変で加算可能な CHLAC は、監視目的での映像解析における有力な画像特徴量の一つと考えられる。本稿では CHLAC 特徴量と相性の良い錘制約部分空間法を動作認識器として用い、それによって複数の動作識別を行えることを示す。歩行者・自転車・自動車の 3 種類の物体による動作に対する実験を通して、本提案手法の性能評価を行った。キーワード 監視カメラ, 多クラス, パターン認識

Action Classification Based on CHLAC Features and Cone-Restricted Subspace Method

Ryuta SATOH[†], Yoshinari KAMEDA[†], and YUICHI OHTA[†]

[†] Graduate School of Systems and Information Engineering,

University of Tsukuba

1-1-1 Tennodai, Tsukuba, Ibaraki, 305-8573 Japan

E-mail: †rsato@image.iit.tsukuba.ac.jp, ††{kameda,ohta}@iit.tsukuba.ac.jp

Abstract This paper describes a cone restricted subspace method for multi-classification of CHLAC features that are taken from both human and non-human motions. CHLAC features are obtained from binarized frame-difference image sequences. Because CHLAC features are position invariant and addable, they should be suitable for generic motion analysis in surveillance. We show that the cone restricted method can be applied to action classification. We have evaluated the method by the experiment of actions of person, bicycle, and automobile.

Key words Surveillance Camera, Multi Class, Pattern Recognition

1. はじめに

シーン中で様々な物体が起こす動作を分類・識別することには、監視目的を含め様々なアプリケーションが想定されるため、その社会的貢献は大きい。それゆえ、一般動作認識は画像認識における大きな研究分野の一つである。

監視目的でカメラが設置されている場合、映像中で監視対象となる動作は多様であるので、特徴量の選定や認識器の構成方法には慎重な配慮が求められる。Cubic Higher-order Local Auto-Correlation (CHLAC) [1] は、このような監視目的の一般物体認識における有力な画像特徴量の一つである。

CHLAC 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コンパクトに表現できるという利点を持つ。また、物体の出現位置に不変であることから、シーン中への侵入物体の位置に関わらず同じ記述が得られることや、複数の同

時動作をその各々の動作に対応する特徴量の和によって表現できることなど、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。計算コストも低い部類に入るため、様々な用途にも応用しやすい。

そこで、本稿では、この CHLAC 特徴量をもとに、種類の異なる物体が行う様々な動作を識別する方法を提案する。

2. 関連研究

2.1 画像特徴量の検討

人間を対象とした一般動作認識では、対象の形状や属性に依存しない特徴量に基づくアプローチが、その応用性の広さから注目を集めている（例えば [2] [3] [4] 等）。それに対して、CHLAC [1] に基づく動作認識の報告は、その特徴量の性質の良さから今後増加が期待される。

SIFT(Scale Invariant Feature Transform) [8] は本研究での

求められる条件である，対象に対する非依存性・位置への非依存性に対応できる優れた手法である．しかし，時間方向の微分に反応しないため，動作認識に対応するためには，さらなる工夫が必要である．

動作へ対応できる手法として，時間方向の畳み込みを行う特徴量である，ST-patch [9] がある．これも良い手法であるが，計算コストが高いという難点がある．

また，SIFT や ST-patch のようなインタレストオペレータベースの局所特徴量では，あるフレーム上での 1 つの動作が含まれる画像全体での表現を得ようとすると，それらの局所特徴量を統合して扱う処理が必要となるという問題点がある．

これに対して，HLAC [10]，CHLAC [1] は動作全体を表す特徴量の構成が容易であるという点で注目されつつある．

HLAC [10] は 2 次元の局所領域における相関パターンの出現を全体で数え上げることで特徴の算出を行う．そのため，画像全体からの特徴を記述することが可能である．さらに，画像全体から相関を取るというその性質上，位置の影響を受けない．

2 次元の局所領域を用いる HLAC に対して，3 次元の局所領域を用いる特徴量が CHLAC [1] である．3 次元を画像平面の 2 次元と時間軸の 1 次元として用いることで，時間方向の情報も得ることができる．

CHLAC 特徴量の応用事例として，これまで，動作認識 [7] [11] や異常検出 [5] [6] が行われてきている．

白木ら [7] は CHLAC および Grid コンピューティングを用いた動作認識を提案しているが，研究の主題が高速化であるため，動作認識には並列化した線形認識を用いており，動作の種類が増えた場合の認識精度向上には言及していない．

一方，森下ら [11] は時間方向への伸縮などを考慮した CHLAC 利用の発展形について提案し，2 種類の動作について実験も行っているが，その主眼は時間的に変動を含む動作への適応であり，複数動作認識が目的ではない．

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 [5] [6] が，いずれも正常状態をある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動作認識には対応していない．

そこで，本稿では CHLAC を複数動作認識に用いる方法について提案する．また，実際の実験結果を示すとともに，その結果から考察を行う．

2.2 識別器の検討

CHLAC のような高次元データベクトルに対するパターン識別手法としてさまざまな手法が提案されている．

ニューラルネットワークはバックプロパゲーション [12] を利用することで，線型分離不可能な学習データに対しても学習を行うことができるが，汎化性能を保証するものではなく，未知データに対する問題がある．

一方，例えば SVM [13] はマージン最大化することで，ニューラルネットワークにおける入力された未知データに対する挙動の問題を低減することができる．しかしながら，多クラスへの対応は向いておらず，対応したとしても，複雑なアルゴリズムや高い計算コストが必要である．

それに対し，部分空間法 [14] は，特徴ベクトルを高次元である特徴空間の部分空間に射影することで識別する手法であり，高次の特徴が一部の次元に偏って存在する場合に適している．さらに，小林らの提案する錘制約部分空間法 [15] は，CHLAC のような非負値のみをとる特徴量に向けて設計されており，よりよい性能が望める．

そこで，本稿では CHLAC 特徴量を入力ベクトルとして，それらを錘制約部分空間法により各動作に識別する．

3. CHLAC 特徴量

本節では CHLAC について簡単に説明する．

3.1 HLAC 特徴量

(1) 式をある関数 $f(x)$ に対する N 次自己相関という．

$$\int f(x)f(x+a_1)f(x+a_2)\cdots f(x+a_N)dx \quad (1)$$

二値化された画像上のある点 p における高次自己相関について考えるとき，注目画素と近傍の領域は相関が高いと予想できることから，すべての領域で相関をとるのではなく，局所領域内からのみ相関を算出することを考える．これを HLAC(Higher-order Local Auto-Correlation) [10] 特徴量とよぶ．

3.2 CHLAC 特徴量

3.1 節で述べた HLAC は 2 次元の局所領域での相関を求めるものであったが，局所領域に時間軸を導入し， $3 \times 3 \times 3$ の局所領域で相関を求めるものが CHLAC(Cubic Higher-ordered Local Auto-Correlation) [1] である．HLAC と同じく平行移動に対してユニークな相関を 2 値差分画像の連続 3 枚の中の画像上で求めると 251 通りとなる．ゆえに，CHLAC 特徴量 x_D は 251 次元のベクトルである．

4. 錘制約部分空間法

本稿では，特徴量ベクトルの多クラス分類に CHLAC に対して適していると考えられる錘制約部分空間法を用いる．これに対する説明と，識別方法についての説明を以下の各節で行う．

錘制約部分空間法は部分空間法の拡張で小林ら [15] により提案された．部分空間法は部分空間への射影を用いることで分類を行うが，錘制約部分空間法は部分空間として錘形状の部分空間を用いる．CHLAC などのヒストグラム値を用いる特徴量は非負であり，錘形状に分布をみるため，効率的に制約を加えることができる．

錘制約部分空間法においてクラス C_i の凸錘 S_i は学習ベクトル群 X_i により計算される．また識別は，凸錘 S_i とクエリベクトル x_q の角度 θ_i を用いる．この角度 θ_i は以下のようにして求めることができる．

$$\theta_i = \arcsin(\epsilon_i) \quad (2)$$

$$= \arcsin\left(\frac{\min_{x \in X_i} \|x_q - x\|}{\|x_q\|}\right) \quad (3)$$

4.1 クラス構成

錘制約部分空間法における錘制約部分空間の構成方法は厳密凸錘・包括凸錘・円錐近似の 3 種類があるが，性能が良いとさ

れる [15] 厳密凸錘について検討を行った。これは、一つのクラスにおける学習ベクトル群すべてからなる凸錘から、取り除いても凸錘全体に影響がないベクトルを取り除くことで、凸錘をなすベクトルの数を減らし、計算量を削減するものである。クラス C_i の厳密凸錘 S_i を構成するためには、学習ベクトル群 $x_i \in C_i$ 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 $C_{i\bar{s}}$ をランダムに作成する

(2) 各学習ベクトル $x_i \in C_i$ に対して、その学習ベクトルを除いた凸錘 S_{s_i} への角度 θ_i を算出する

(3) 閾値 τ_θ に対して $\theta_i < \tau_\theta$ となる学習ベクトルを従属しているものとして、 C_i および C_{is} から除去する

(4) 学習ベクトル $x_j \in C_{i\bar{s}}$ に対して、 C_{is} の成す凸錘 S_{is} への角度 θ_j を算出し (3) と同様に $\theta_j < \tau_\theta$ となる学習ベクトルを従属しているものとして C_i から除去する

(5) (1) ~ (4) を複数回繰り返す、 C_i からの除去がなくなれば終了。そうでなければ繰り返す

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

4.2 錐制約部分空間法による分類

本節では、クエリベクトル x_q を分類する方法について説明する。先に述べたとおり、クエリベクトル x_q とクラス C_i に属する凸錘 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 における閾値 $\tau_{\theta,i}$ を設定し、それによるクエリベクトルの凸錘 S_i に対する判定を行う。クエリベクトル 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 x_q はクラス C_i に属していると判定する。

$$x_q \in C_i \quad (\forall i: \tau_{\theta,i} > \theta_i) \quad (4)$$

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

5. 多クラス分類器の構成

前節までで説明した手法を実際の動作分類に適用する方法について述べる。

カメラで撮影された映像はフレーム間差分される。時系列で最新である3枚のフレーム間差分画像をしきい値 τ_I を用いて二値化したものに CHLAC を適用し、251次元のベクトルを算出する。さらに、学習データから凸錘を構成し、得られた凸錘を用いて4.2節のとおり分類を行う。

このとき、ある動作の映像シーケンス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。

なお、(4)式において必要とされる、それぞれのクラスに分類する閾値 $\tau_{\theta,i}$ は、以下のように決定する。

まず、クラス C_i の N_i 個の学習データ自体をそのクラスの分類器で分類し、その評価値の大きさにより、降順に並べる。

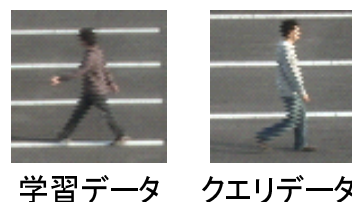


図1 実験に用いた人物の例

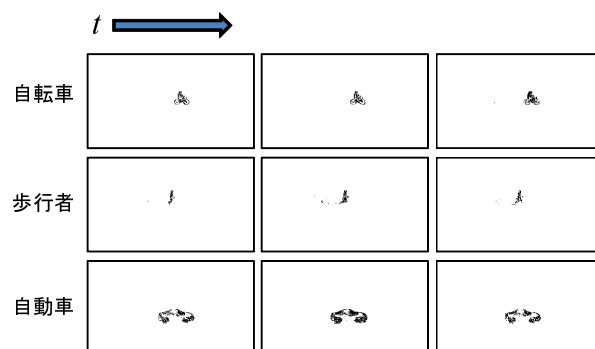


図3 CHLAC 算出の例

$$\theta_{i,1} > \theta_{i,2} > \dots > \theta_{i,N_i-1} > \theta_{i,N_i} \quad (5)$$

その上で、クラス C_i の学習データの個数 N_i に対する割合 r_i により、閾値 $\tau_{\theta,i}$ を決定する。

$$\tau_{\theta,i} = \theta_{i,j} (j = r_i \cdot N_i) \quad (6)$$

この、 r_i は別途定めるものとする。

6. 実験

実験として実際に動物体の動作进行分类する。ここでは、BRC-H700 カメラを監視カメラとして使用した。カメラから得られる映像は 1920×1080 のハイビジョン映像である。

実験で使用した動作は、自動車での左右移動・自転車での左右移動・人間の左右への歩行である。それぞれの動作は、同じ画角にセットされた同じカメラから撮影された。撮影は学習データとして20回、クエリデータとして2回の動作について行った。自動車の左右移動を除き、学習データとして撮影された人物と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1)。撮影された映像は1人・1台の人物・自動車による動作である。

それぞれの動作を図2に示す。すべての映像はCHLACを算出する前処理として、1/6に縮小した。CHLACを実際に参照する3つのフレーム間差分画像の例を図3に示す。

撮影した映像は、グレイスケール256階調のフレーム間差分画像を閾値 $\tau_I = 15$ を用いて2値化し、CHLAC特徴量を抽出し、凸錘を $\tau_\theta = 0.01[\text{rad}]$ を用いて構成した。クエリデータは別途撮影したそれぞれの動作を使用した。また、5.節で記述した方法によりクエリデータを分類させた。その際に、閾値 $\tau_{\theta,i}$ を決定するパラメータ r_i は全てのクラス i について $r_i = 0.005$ とした。

結果を表1に示す。上段に並んでいるクエリが左列に示され

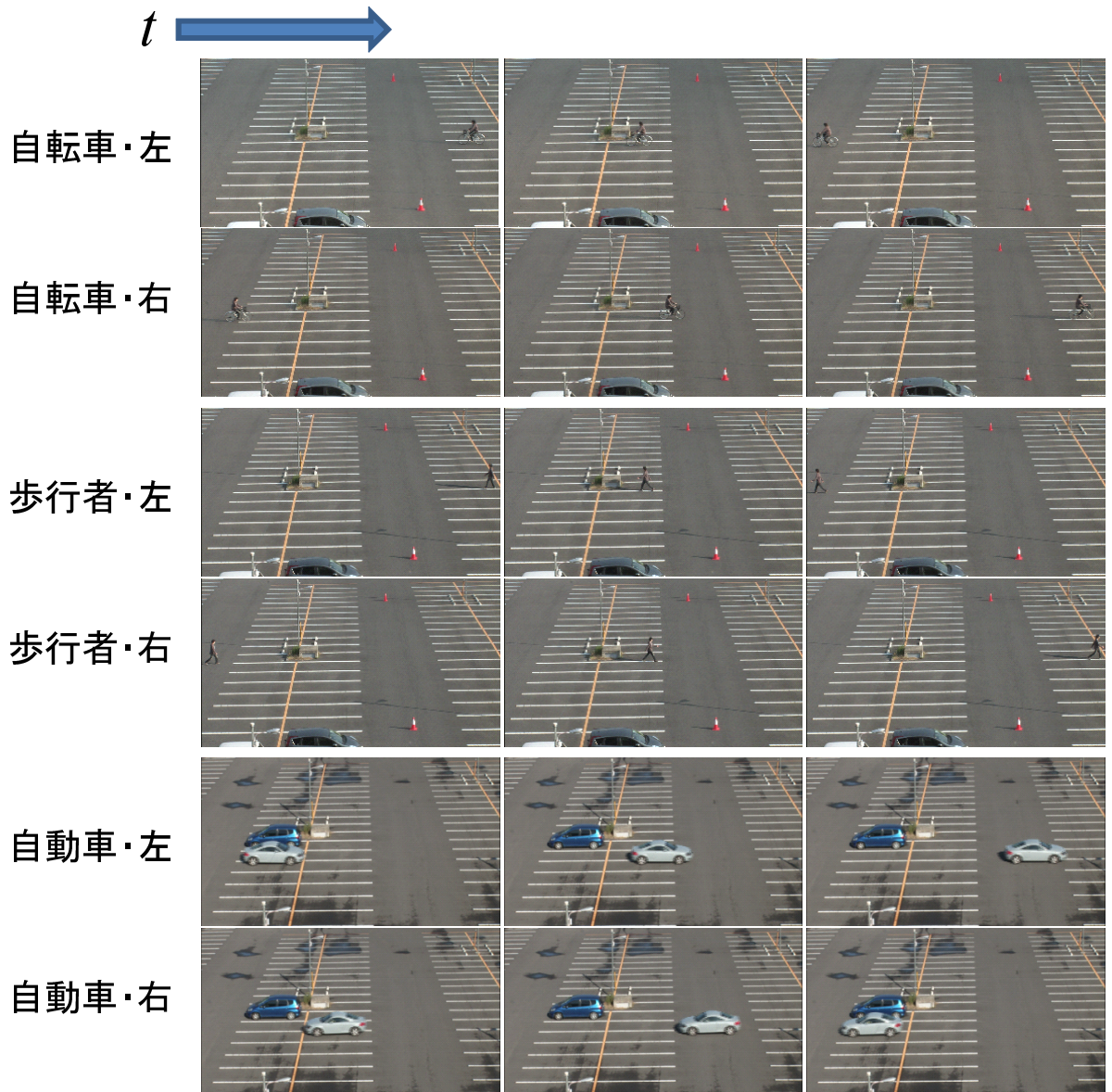


図 2 実験に用いた動作の例

表 2 ユニークに識別されたクエリ

クラス	自転車左	自転車右	歩行左	歩行右	車前進	車後進
フレーム数	87	91	99	123	42	24
割合 [%]	50.9	53.2	48.8	63.7	73.7	41.4

ているクラスに何フレーム分類されたかを示す．クラスは自動車の前進・後進および左方向および右方向への歩行・自転車の移動の 6 クラスである．それぞれの識別器を独立に評価しているため，複数の識別器が反応した場合は，割合の合計は 100%とはならない．また，一つの正解クラスのみユニークに識別されたフレーム数・割合を表 2 に示す．太字で示した数値が正しい識別器が反応したフレーム数・割合である．

また，それぞれのクラスにおける r_i を 0 から 1 まで変化させた際の ROC 曲線を図 4 に示す．横軸が False Positive，縦軸が True Positive となっている．

クエリ集合を構成する CHLAC 特徴量ベクトルについて，41.4～73.7%の割合でクエリを正解クラスのみで分類することができた．また，正解クラスに分類できた割合は 48.8～93.0%となる．正解クラスのみで分類できた割合をすべてのクラスについて平均すると 55.3%となる．最も成績の良かったクラスは自動車・左クラスの 73.7%であり，最悪値は自動車・右クラスの

表 1 動作分類結果 (単位:フレーム数・括弧内は全体に対する割合 [%])

		クエリ					
		自転車左	自転車右	歩行左	歩行右	自動車左	自動車右
フレーム数		171	171	203	193	57	58
結果	無識別	46 (26.9)	37 (21.6)	104 (51.2)	70 (36.3)	4 (7.0)	3 (5.2)
	自転車左	122 (71.3)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
	自転車右	0 (0.0)	122 (71.3)	0 (0.0)	0 (0.0)	0 (0.0)	15 (25.9)
	歩行左	38 (22.2)	0 (0.0)	99 (48.8)	0 (0.0)	0 (0.0)	0 (0.0)
	歩行右	0 (0.0)	43 (25.1)	0 (0.0)	123 (63.7)	0 (0.0)	0 (0.0)
	自動車左	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	53 (93.0)	23 (39.7)
	自動車右	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	11 (19.3)	45 (77.6)

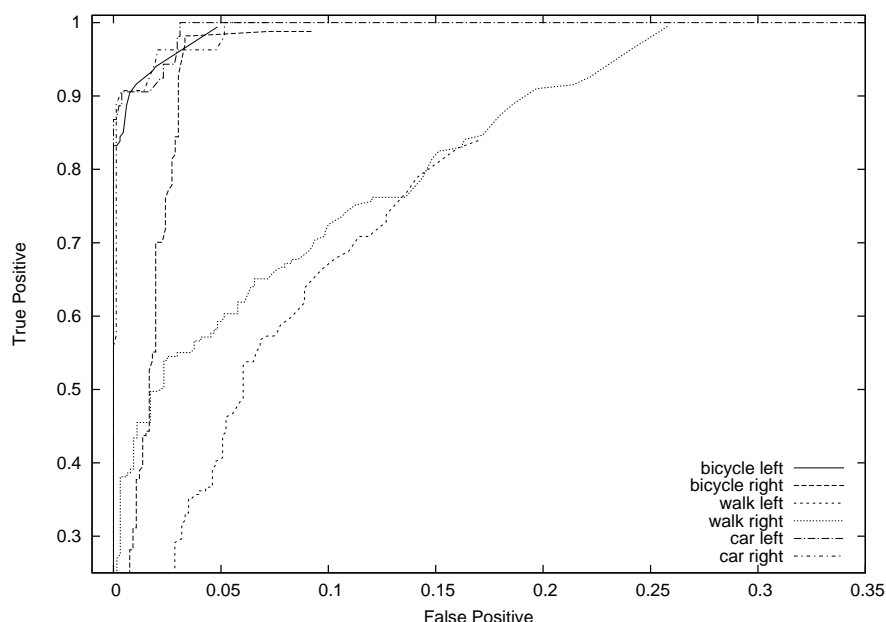


図 4 それぞれのクラスにおける ROC

41.4%となった。

ROC においては、自動車・自転車・歩行者の順に良い結果を得ることができた。これは、それぞれのクエリが持つ自由度が影響していると思われる。

また、自動車クラスが方向によりユニークに識別された数が極端に変化した原因は次のように考えている。図 4 から、ROC は自動車左方向・右方向で大きな変化はない。そのため、自動車クラス単独での性能は問題ないとみなせる。しかし、他クラスの識別器が反応したため、単独での識別率が低くなっている。このことは、CHLAC 表現上において自動車右クラスとその他のクラスが重なっている可能性が高いといえる。今後自動車右クラスの特徴空間中での広がりを調査していきたいと考えている。

最後に、今回の実験においては、学習データが 1 人・1 台の

みから構成されている。このことは、学習データの分布が狭い可能性が高く、より大規模なデータセットを用いることで、より高い性能を期待することができる。

7. おわりに

本研究では、監視目的で設置された固定カメラの映像による、多種類の物体の動作を識別する手法を提案した。

本手法では、映像からフレーム間差分し得られた 2 値画像系列から CHLAC 特徴量を算出し、識別に用いる。

方法としては、まず事前に用意した複数物体の複数動作それぞれについての学習データについて、部分空間法の一種である錘制約部分空間法により錘を構成した。その上で、入力 CHLAC 特徴量に対して、クラスごとに錘との角度における閾値を設定し、独立した識別を行える識別器を構成した。

実験では、歩行者、自転車、自動車の3種類の物体による合計6種類の動作に対して実験を行い、55.3%の識別結果を得た。

錐制約部分空間法はCHLAC特徴量のように加法に対して閉じている制約が加えられているため、CHLAC特徴量の加法であらわすことのできるような同時発生の複数動作に対して良い識別性能を示す可能性がある。今後この方向について、検討していきたい。

文 献

- [1] N. Otsu, "Towards Flexible and Intelligent Vision Systems - From Thresholding to CHLAC -", IAPR Conference on Machine Vision Applications, pp.430-439, 2005.
- [2] J. C. Niebles, H. Wang, and Li F. "Unsupervised Learning of Human Action Categories Using Spatial-Temporal Words", International Journal of Computer Vision, pp. 299-318, 2008.
- [3] 村井泰裕, 藤吉弘巨, 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.
- [4] 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.
- [5] 南里卓也, 大津展之, "複数人動画像からの異常動作検出", CVIM, vol.2004, no.91, pp.173-186, 2004.
- [6] 村井泰裕, 藤吉弘巨, 数井誠人 "時空間特徴に基づくエスカレーターシーンにおける人の異常行動検知" 電子情報通信学会 パターン認識・メディア理解研究会 (PRMU), pp.247-254, 2008.
- [7] 白木 孝義, 石黒 勝彦, 深野 亮, 鴨志田 良和, 白井 達也, 齊藤 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 PRMU vol.105, pp.97-102, 2006.
- [8] D. G. Lowe, "Object recognition from local scale invariant feature", Proc. of IEEE International Conference on Computer Vision (ICCV), pp. 1150-1157, 1999.
- [9] E. Skechtman, M. Erani, "Space-time behavior based correlation", Proc. CVPR, pp.405-412, 2005.
- [10] N.Otsu and T.Kurita, "A new scheme for practical, flexible and inteligent vision systems", Proc. IAPR Workshop on ComputerVision, pp.431-435, 1988.
- [11] 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.
- [12] D.E. Rumerlhar, G.E. Hilton and R.J. Williams, "Learning representation by back-propagating errors", Nature, 323, pp. 533-536, 1986
- [13] C. J. C. Burges, "A tutorial on support vector machines for pattern recognition", Data Mining and Knowledge Discovery 2, pp.121-167, 1998
- [14] 石井健一郎, 上田修功, 前田英作, 村瀬洋, "わかりやすいパターン認識", オーム社出版局, 1998.
- [15] 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", 電子情報通信学会論文誌.D, 情報・システム, pp.104-111, 2009.