

安全・安心のための動画像認識技術

小林 匠^{1,a)}

概要：近年、人々の安全を確保するセキュリティ目的、また生活の安心を保証する見守り目的で、監視カメラ等からの動画像の利活用が進んでいる。カメラ設置数の増加から、人手による動画像の処理は膨大な人的コストがかかるため、動画像を用いて人の動きを自動解析する省力化されたシステムの実現が求められている。そのような自動解析のためには、まず動画像から人の動作情報を効果的に抽出する特徴抽出手法が必要となる。本稿では、立体高次自己相関特徴（CHLAC）をはじめとする動画像特徴抽出法を説明し、安全・安心分野への動画像認識技術の適用事例を紹介する。

1. はじめに

街頭での監視カメラや病院・介護施設における見守り目的でのカメラ利用など、様々な場面で映像（動画像）の利活用が進んでいる。しかしながら、動画像においては画像の空間情報に加えて時間情報をも含んだデータとなるため、得られた映像の内容を調べる作業量はカメラの普及と共に増大し、人的コストの観点から必ずしも全ての状況で十分に映像が活用されているとは言い難い。そこで、コンピュータビジョンやパターン認識における動画像認識技術を適用して、自動でそれら映像内容を認識処理するシステムが求められている。これにより、映像の確認作業が大幅に省力化され、少ない人的コストで多数の映像を効率よく扱うことが可能となり、見落としなどのエラーも少なくなることが期待される。

本稿では、そのような安全・安心分野で重要となる動画像認識技術に関して、特に動作特徴抽出の観点から説明する。さらにそれら特徴量を利用した応用事例についても紹介する。多量の情報を含む動画像を効率よく処理し、実世界の多様な変動に対応するためには、動作特徴量の果たす役割は非常に大きい。また、映像利用の際にはプライバシー保護についても熟慮する必要があるが、カメラで撮影する段階で特徴抽出をすることで、プライバシー情報を極力削除した形式での認識システムの構築も期待できる。

2. 動画像特徴抽出手法

動画像認識では大きく2つのアプローチが考えられる。1つは、動画像を時間方向に並ぶ静止画像フレーム列と

捉え、各フレーム抽出した（静止）画像特徴量を Hidden Markov Model (HMM) などを用いて時間解析するものである [1]。もう1つは、動画像を空間 XY 軸と時間 T 軸から成る 3-way データ（図 1）と見なすことで、ここから時空間情報を直接的に抽出するものである。前者では空間情報と時間情報を別々に扱っているが、後者では抽出した特徴量自体に動作などの時間情報が含まれており、その後の処理は静止画像認識と同様に様々なベクトル識別器を適用できるという利点がある。ここでは後者の枠組みに基づく動画像特徴抽出法について述べる。3-way 動画像における対象の動作は、XYT 空間において3次元曲面を成し、これは motion surface [2] と呼ばれる（図 1）。つまり、ここでの動画像特徴抽出とは、その motion surface の幾何学的情報を抽出することに帰着される。

入力動画像には照明変動といった認識対象（動作など）には無関係な環境雑音が含まれている。そのため、前処理によりそのような変動を軽減させるが、どのような前処理を適用するかにより、動画像は様々な形式の 3-way データとなり、特徴抽出法も変わる。本稿では、前処理としてフレーム間差分と2値化処理を適用した 3-way 2値スカラー場に対する立体高次局所自己相関特徴（cubic higher-order auto-correlation: CHLAC）[3]、また時空間勾配処理を適用した 3-way ベクトル場に対する時空間勾配の局所自己相関特徴（space-time auto-correlation of gradients: STACOG）[4] について説明する。いずれも、局所近傍での自己相関関数を用いて、各々のデータ表現における局所共起性を効果的に特徴として抽出していることが大きな特長である。

¹ 独立行政法人 産業技術総合研究所
Umezono 1-1-1, Tsukuba 305-8568, Japan

^{a)} takumi.kobayashi@aist.go.jp

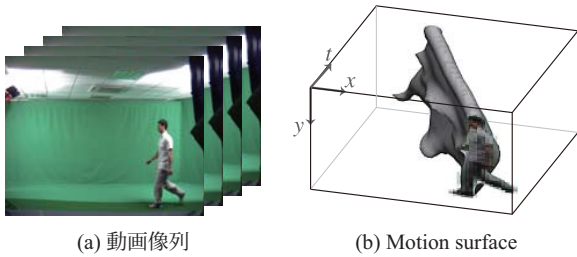


図1 動画データ表現

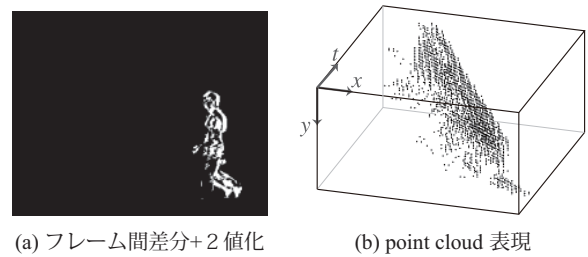


図2 CHLAC での前処理

2.1 立体高次局所自己相関: Cubic Higher-order Local Auto-Correlation (CHLAC) [3]

$f(\mathbf{r}), \mathbf{r} = [x, y, t]^T$ を時空間 XYT 空間にて定義されるスカラー関数とすると, N 次の CHLAC は一般には次式により定義される.

$$\int f(\mathbf{r})f(\mathbf{r} + \mathbf{a}_1) \cdots f(\mathbf{r} + \mathbf{a}_N)d\mathbf{r}. \quad (1)$$

ここで, $\mathbf{a}_i (i = 1, \dots, N) \in \mathbb{R}^3$ は変位ベクトルを示す.

動画画像特徴としての CHLAC では, 前処理として動画画像列に対するフレーム間差分及び 2 値化処理 (図 2a) を適用し, 動作情報を効果的に表現する. 2 値化処理においては, 環境変動への頑健性を高めるために, 差分値に対して大津の 2 値化法 [5] を適用するのが一般的である. 2 値化後のピクセル値 ($f(\mathbf{r}) \in \{0, 1\}$) には, 動き情報のみが表現されている (“動いた” ピクセルのみが 1). つまり, この前処理により前述の motion surface は point cloud として表現されることになるため (図 2b), CHLAC 特徴とは時空間での point cloud の特性を記述していることになる. 相関の次数を 2 次 (3 点相関) 以下 ($N \leq 2$), 変位ベクトルを (隣接する) 26 近傍点に限定すると, 図 3 に示すようなマスクパターン 251 種を用いて局所相関値が算出される. 対象が 2 値データであるため, これは

$$\sum_{\mathbf{r}} f(\mathbf{r}) \wedge f(\mathbf{r} + \mathbf{a}_1) \wedge f(\mathbf{r} + \mathbf{a}_2) \quad (2)$$

の論理積となり, point cloud においてマスクパターンが合致する頻度が特徴となる. これはまさに局所領域内での “動き” ピクセルの共起頻度を測定していることになり, 幾何学的観点からは 2 点関係では勾配の, 3 点関係では曲率の方向・種類をパターン化しているとも考えられる. 論理積 (2) は計算コストも小さく, 動画画像から高速に特徴を抽出することが可能となっていることにも注意されたい [6].

以上から, CHLAC とは前処理 (フレーム間差分 + 2 値化) により動画画像内の motion surface を point cloud として表現し, そのような離散表現に対して離散マスクパターンを適用することで, 結果的に motion surface の幾何学的特性 (勾配・曲率) をパターン化し特徴づけていると言える.

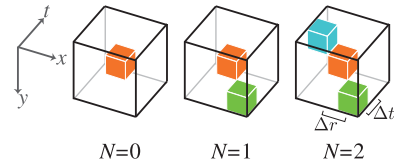


図3 CHLAC マスクパターン. $\Delta r, \Delta t$ はそれぞれ XY 空間内, 時間軸での相関幅を示す.

2.2 時空間勾配の局所自己相関: Space-Time Auto-Correlation of Gradient (STACOG) [4]

静止画像においては HOG [7] に代表されるように勾配情報を利用することで高い認識性能が得られている. そこで, 3-way 動画画像データにおいても前処理として微分フィルタ処理を適用して時空間 (3 次元) 勾配を抽出する. これにより, 動画画像は時空間 XYT における勾配ベクトル場として表現される (図 4). ここでは, 時空間勾配ベクトルそのものではなく, その 3 次元角度及び勾配ベクトルの大きさを 3-way データの基礎表現として用いる (図 5). さらに, 3 次元角度を量子化方向表現すると, 例えば, 図 5a に示すように 1 つの勾配ベクトル (の方向) は高々 4 つの近傍量子化方向で表現され, これを量子化数 d の次元をもつベクトル $\mathbf{h} \in \mathbb{R}_+^d$ で表す. 勾配ベクトルのノルムを $w(\mathbf{r}) = \|\nabla I\|$ とすると, STACOG 特徴は一般には次のように定義される.

$$\sum_{\mathbf{r}} \min[w(\mathbf{r}), w(\mathbf{r} + \mathbf{a}_1), \dots, w(\mathbf{r} + \mathbf{a}_N)] \times \mathbf{h}(\mathbf{r}) \otimes \mathbf{h}(\mathbf{r} + \mathbf{a}_1) \cdots \otimes \mathbf{h}(\mathbf{r} + \mathbf{a}_N) \quad (3)$$

テンソル積を用いることで量子化方向の相関を算出しており, 勾配方向の共起性を抽出している. さらに, 勾配の大きさ, 特に $\min$ を重み関数に採用することで雑音への頑健性を高めている. ここでは, 相関次数は 1 次 (2 点相関) 以下 ($N \leq 1$) とし, CHLAC と同様に変位ベクトルを (隣接する) 26 近傍に限定する. 1 次相関での変位パターン $\mathbf{a}_1$ は 13 種となり, 方向量子化数を例えば $d = 21, \mathbf{h} \in \mathbb{R}_+^{21}$ (図 5b) とすると, $21 \times 21 \times 13 + 21 = 5754$ 次元の高次元の特徴ベクトルとなる. しかしながら, (3) の表現はスパース (高々 4 つの非ゼロ成分) であるため, 特徴量算出に必要な計算量は非常に少なく, 得られる特徴ベクトルとしてもスパースになることが期待される.

上記処理では, 対象の “動き” には特に注目せず, 空間

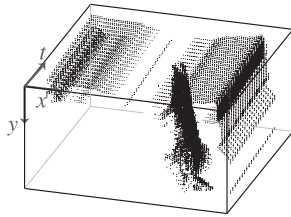


図4 動画像での時空間勾配ベクトル場

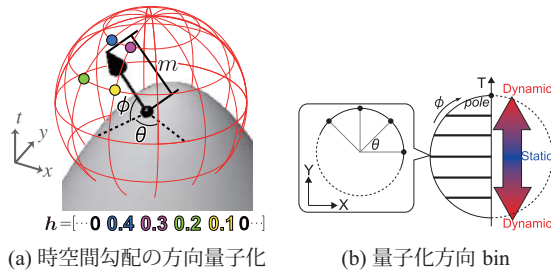


図5 量子化時空間勾配

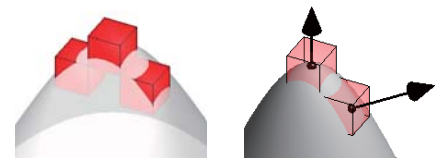
表1 STACOG 特徴での組み合わせ

	Accuracy (%)	92.8	95.6	94.6
0th	dynamic	✓	✓	✓
	static			✓
	dynamic × dynamic	✓	✓	✓
1st	dynamic × static		✓	✓
	static × static			✓

XY と時間 T 軸を均等に扱っている．そのため，得られる STACOG 特徴には動作情報だけでなく，静止情報も含まれている．実際に図4では対象人の動きだけでなく，背景の静止物体に対しても時空間勾配が得られていることが分かる．時空間勾配の向きを T 軸に沿って考えると，図5bに示すように，静止対象では T 軸でのピクセル値変動が無いために勾配は XY 平面内を向いており，動きの大きな対象ほど T 軸に沿った方向を向く．静止～動作まで一括して特徴抽出できることは，状況理解などの応用には有用と考えられるが，動画認識では背景などの静止物体は認識対象から外したいことが多い．そのため，静止状態を意味する XY 平面内の量子化方向を STACOG 特徴から除外することで，動作のみに着目した特徴量が得られる．実際に KTH データセット [8] を用いた動作認識性能の比較を表1に示す．1次相関（共起）では動きと静止情報の組み合わせが3種あることに注意されたい．これを見ると，静止情報のみを STACOG 特徴から除外することで性能が向上し，興味深いことに静止 × 動きの共起は採用することで性能が大きく向上する．これは，人の動作を記述する上では，動く部位と静止部位の関係性も重要であることを示している．

2.3 CHLAC と STACOG

STGLAC 特徴では，0次相関の場合には，時空間勾配方向ヒストグラムとなり，1次相関の場合には勾配方向の共起ヒストグラムとなる．つまり，0次では単純に勾配を，



(a) CHLAC 2次相関 (b) STACOG 1次相関

図6 CHLAC と STACOG

1次では勾配ペアの共起，つまり曲率の特徴を抽出していることに相当する．CHLAC では1次及び2次の相関によりこれらの幾何特徴を抽出していたが，ここでは3-way データ表現として時空間勾配方向を用いたために，低い次数の相関で同等の幾何特性を表現できる（図6）．これら二つの特徴抽出法の比較を表2にまとめる．

表2 CHLAC と STACOG の比較

	CHLAC	STACOG
前処理	フレーム間差分+2値化	微分フィルタ
基礎表現	動きピクセル点 $\in \{0, 1\}$	量子化時空間勾配 $\in \mathbb{R}_+^d$
勾配	1次相関（2点共起）	0次相関
曲率	2次相関（3点共起）	1次相関（2点共起）

以上，相関に基づく2つの動画特徴抽出手法を説明した．これらはいずれも動画を3-way データと見なし，その2値化や勾配方向量子化により，質的（カテゴリーカルな）表現へと変換した後に自己相関関数を適用している．質的データの自己相関により共起性を効率良く算出している点に大きな長がある．また，これらの局所共起性は局所幾何特性（勾配・曲率）としても解釈できるため，これらの特徴量は動画3-way データ，つまり motion surface の幾何特徴を反映していると言える．

2.4 その他の特徴抽出法

動画を3-way データと見なすことで，2-way データである静止画像の自然な拡張となる．そのため，静止画像における様々な画像特徴抽出法も3次元拡張を行うことで，動画特徴抽出法として提案されている [2], [9], [10], [11], [12]．上述した2つの特徴抽出法もそれぞれ静止画像における特徴抽出法 [13], [14] を3次元へと自然に拡張したものとなっている．

また，一般物体認識や様々な画像認識応用において高い性能を示す Bag-of-Feature の枠組み [15] も動画認識へ適用されている [9], [11], [16]．

3. 識別手法

上記特徴抽出により，入力動画内の動き情報を1つの特徴ベクトルとして表現することができる．そのため，判別分析などの多変量解析 [17] や SVM [18] などのベクトル識別器を適用して，動画の認識を実現することができる．動画シーケンスにおいては，ある一定幅の時間窓を用意

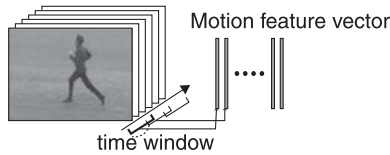


図7 時間窓

し、その区間のデータ（部分動画）から特徴ベクトルを時々刻々と抽出し識別することも多い（図7）。特に、オンラインでは1枚ずつフレームが入力されるため、それに対応する特徴量のみを計算することでインクリメンタルに特徴ベクトルを更新でき、計算時間の削減に効果がある。また、時間幅を全区間とすれば動画シーケンス全体の特徴が抽出される。具体的な識別手法はタスクに応じて様々であるため、次章で応用事例とともに述べる。

4. 応用

人の動作を対象とした動画認識課題へ適用する。ここでは安全・安心の観点から、動作（歩行）行為者である個人の同定、状況理解のための動作分類、状況に基づく動作の意味理解としての異常検出、の3つの応用を紹介する。

4.1 個人同定 (Gait Recognition)

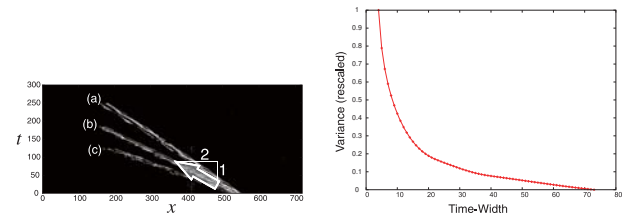
歩行動作（歩容）からの個人同定を行う。監視カメラなどの固定カメラから歩行者の同定を行うことで、対象に非接触で離れた距離からも個人同定ができるという利点がある。“歩行の様”である歩容は一見すると個人間の差異が小さいように思われるが、歩容からの個人同定の可能性は心理学的観点からも様々に議論されている[19], [20]。ここではCHLAC特徴を歩容認識へ適用する[21]。

CHLACを用いた歩容の識別は次のように行われる。CHLAC特徴におけるパラメータは相関幅 $\Delta r, \Delta t$ 及び、時間窓幅 ΔT となる。基本的には、各パラメータの組において、訓練データから得られるCHLAC特徴ベクトルと個人を示すラベル c_i を用いて判別分析を行い、特徴ベクトルを判別空間へ射影した後、その空間内で k NN識別を行う。 k NNによって、各個人への事後確率が求まるため（ k NNサンプルの属性クラス分布から）、それを全パラメータ組の中で最大の事後確率を示す個人へ識別する。

$$\hat{c}(t) = \arg \max_i \max_{\{\Delta r, \Delta t, \Delta T\} \in \mathbb{P}} p_{kNN}(c_i | \mathbf{v}_t; \Delta r, \Delta t, \Delta T) \quad (4)$$

$\mathbb{P}$ はパラメータ範囲を示し、 $p_{kNN}(c_i | \mathbf{v}_t; \Delta r, \Delta t, \Delta T)$ はパラメータ $\Delta r, \Delta t, \Delta T$ の下で得られる時刻 t でのCHLAC特徴 $\mathbf{v}_t$ に判別空間内で k NNを適用して得られる各クラスへの事後確率を示す。最終的にはシーケンス内で上記識別結果 $\hat{c}(t)$ を投票し、最も多く支持された個人を識別結果として出力する。

ここではパラメータの定義域 $\mathbb{P}$ が重要となる。パラメー



(a) XT スライス

(b) CHLAC 特徴の分散

図8 歩行の解析



図9 USF-NIST gait データセット

タの組は膨大であるが、歩行動作をXYT空間で解析することでその範囲を限定していく。まず、相関幅 $\Delta r, \Delta t$ については、歩行の速度と大きく関係する。つまり、 $\frac{\Delta r}{\Delta t}$ は主に対象とする歩行速度を示しており、訓練データから得られる平均的な歩行速度によって $\frac{\Delta r}{\Delta t}$ を規定する。2値化3-wayデータをXT平面でスライスした画像を図8aに示す。ここから平均的な歩行速度は $\frac{\Delta r}{\Delta t} = 2$ であることが分かる。また、画面内での人の平均的な大きさから $\Delta r \leq 16$ と規定される。一方、時間窓幅 ΔT は、歩行の時間周期と関係している。歩行周期以上の時間窓を設定することで、CHLAC特徴量の分散は小さくなることから、各時間窓幅でのCHLAC特徴の分散を解析することで時間幅を決定する。特徴量の分散は図8bのようになり、時間幅が $\Delta T \geq 30$ で安定している。そのため、それ以上の時間幅を適用することは冗長であるため、ここでは $\Delta T \leq 30$ と規定する。以上をまとめると、パラメータの定義域は次式となる。

$$\mathbb{P} = \left\{ \Delta r, \Delta t, \Delta T \mid \frac{\Delta r}{\Delta t} = 2, \Delta r \leq 16, \Delta T \leq 30 \right\} \quad (5)$$

USF-NIST gait データセット[22]（図9）に上述の歩容認識手法を適用すると、認識性能は図10のようになった。データセットには訓練セットとテストセットが含まれ、テストセットは訓練環境との差異の度合いから難易度の異なる7種のセット（A～G）から成る。他の手法と比較しても、良好な認識性能が得られている。当然のことながら、困難なテストセットでは性能が低下するが、比較的その低下度合いも少ないため、環境に対する頑健性も示されている。

4.2 動作識別

前節では動作の行為者を識別対象としていたが、ここでは各人に依らず動作そのものの識別・分類を目的とする。対象者の動作の種類を認識することで、対象者の意図や状況を理解することにつながる。また、動作としてジェスチャーなどを対象とすることで、非接触インタフェースと

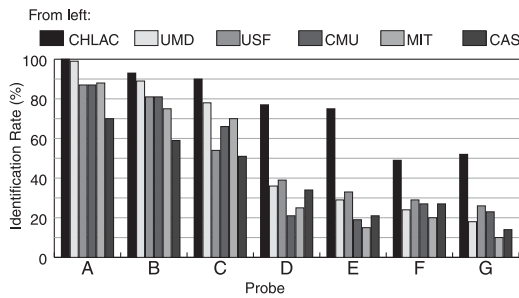


図 10 USF-NIST gait データセットでの個人同定性能

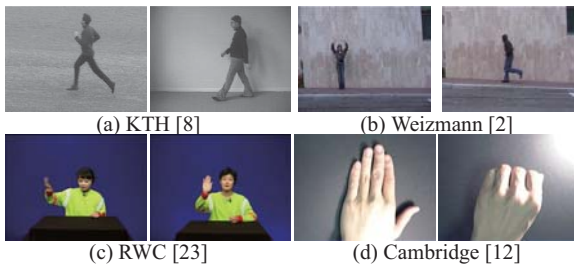


図 11 動作識別のデータセット

しても利用することが出来る。

ここでは STACOG 特徴を動作識別へ適用する [4]。前節と同様に判別分析を利用しつつ、Bag-of-Features の枠組みで識別を行う。まず、訓練データから（時間窓を用いて）抽出した STACOG 特徴に対して判別分析を適用し判別空間を構築した後に、その空間内で k -means クラスタリングを行いクラスターを抽出する。各クラスターはある動作全体から見た一部の動きを意味するため、motion word と呼ぶ。各時刻での（時間窓の）STACOG 特徴に対して判別空間内で近傍の motion word を割り振り、シーケンス全体で motion word のヒストグラムを算出する。最終的にはそのヒストグラム特徴を linear SVM [18] で各動作へと識別する。判別空間でクラスタリングを適用することで、各 word クラスターは動作を記述する上で弁別性の高い表現となる。

動作認識のベンチマークとして用いられる KTH データセット [8] と Weizmann データセット [2]、及びジェスチャー認識の RWC データセット [23] と Cambridge データセット [12]（図 11）において動作認識性能の評価を行い、表 3 に結果を示す。全てのデータセットにおいて他の手法と比較しても非常に高い性能が得られている。

4.3 異常動作検出

最後に異常動作検出への適用例を示す。監視カメラにおいて、異常動作の検出は主要なアプリケーションの一つである。複数台の監視カメラから人手で異常動作を見張り続けることは人的コストがかかるため、その自動化は非常に有意義となる。例えば高齢者の見守りシステム等においては、異常動作は転倒などが挙げられる。

異常な動作とは事前に規定され得るものではなく、状況に応じて決まる。つまり、ある状況で頻繁に行われる動作

表 3 動作識別結果

(a) KTH dataset		(b) Weizmann dataset	
Method	acc. (%)	Method	acc. (%)
Schuldt et al. (2004)	71.7	Scovanner et al. (2007)	82.6
Laptev et al. (2008)	91.8	Dollar et al. (2005)	86.7
Wang et al. (2009)	92.1	Ballan et al. (2009)	92.4
Ballan et al. (2009)	92.1	Jhuang et al. (2007)	98.8
STACOG	95.6	Wang and Mori (2009)	100
		STACOG	100
(c) RWC dataset		(d) Cambridge dataset	
Method	acc. (%)	Method	acc. (%)
Ishihara and Otsu (2004)	95.7	Wong et al. (2005)	44
Kobayashi and Otsu (2009)	95.9	Niebles et al. (2006)	66
STACOG	98.1	Kim et al. (2007)	82
		STACOG	89

が正常なものであり、滅多に起こらない動作が異常なものと解釈されるように、動作の統計分布からその異常性が測られる。Nanri & Otsu [24] は CHLAC 特徴に部分空間法を適用することで、教師無し学習の枠組みでの異常検出法を提案した。頻繁に行われる動作（CHLAC 特徴）の分布を部分空間によりモデル化しているが、その他の 1 クラス学習法、例えば 1-class SVM [25] や錐制約部分空間法 [26] などにも同様に適用することができる。ここでは簡便な Nanri & Otsu [24] の異常検出法を説明し、その適用例を見る。

動作特徴ベクトルを $v \in \mathbb{R}^d$ とすると、部分空間は

$$U^* = \arg \min_{U|U^T U = I} \sum_i \|v_i - U U^T v_i\|^2 \quad (6)$$

として求め、次式の固有値問題を解くことで得られる。

$$\left(\sum_i v_i v_i^T \right) U^* = U^* \Lambda \quad (7)$$

ここで、 $\Lambda = \text{diag}(\lambda_1, \dots, \lambda_d)$ は固有値を対角に並べた対角行列である。実際にはある状況での動作は全空間に渡って分布することは無く、ある限られた領域（部分空間）に存在し、それが正常動作を示す部分空間（正常部分空間）となる。そのような正常部分空間は誤差最小の観点から累積寄与率 $\eta_l = \frac{\sum_{j=1}^l \lambda_j}{\sum_j \lambda_j}$ に基づいて決定される。ある閾値 $0 \leq \tau \leq 1$ を設定し、 τ 以上で最小の累積寄与率を与える次元 $l^* = \arg \min_{\eta_l \geq \tau} l$ が正常部分空間の次元となる。このようにして求めた正常部分空間の基底ベクトル $U^* \in \mathbb{R}^{d \times l^*}$ を用いて、動作の異常の度合いは正常部分空間からの逸脱度（最小 2 乗距離）として算出される。

$$\sigma = \|v - U^* U^{*T} v\| \quad (8)$$

滅多に起こらない動作は正常 CHLAC 特徴分布からは逸脱するため、上記の異常指標 σ は大きくなる。

Montreal fall detection データセット [27] を用いて、異常検出の性能を評価した。これは部屋に設置された固定カメラで、部屋での人の行動を録画した動画データである。



図 12 Montreal Fall データセット

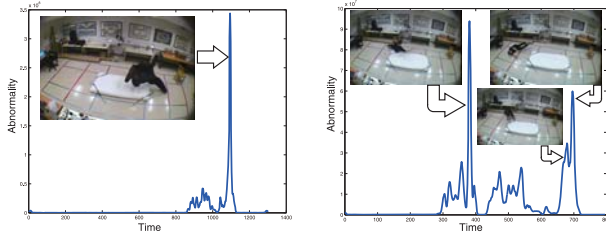


図 13 異常動作の検出結果

通常は歩き回ることが多いが、まれに人が転倒する様子も記録されている。結果を図 13 に示す。転倒の瞬間に異常指標は高い値を示しており、その他の動作状況（歩行など）では低い値に留まっていることから、転倒検出が正しく行われていることが分かる。ここでは、正常部分空間には全ての動画シーケンスを学習に用い、同じ学習シーケンスに対する異常指標を示している。そのため、異常動作として“転倒”動作を事前に教示することなく、統計的に転倒動作が異常であることを示している。

5. おわりに

本稿では、安全・安心のための動画認識システムを実現する上で重要な動作特徴抽出法として、CHLAC [3] 及び STACOG [4] を説明した。これらの手法は、動画 3-way データ内での局所相関に基づき、motion surface の勾配や曲率といった幾何学的特徴を効果的に抽出している。これらの動作特徴を用いて、動画認識への適用事例として、歩容からの個人同定、動作の識別、さらに異常動作検出といった応用を紹介した。これらの動画認識手法は、見守りや監視システムの根幹を成す技術として広く社会への応用が期待される。

謝辞 産業技術総合研究所 大津展之氏には動画認識に関する様々なご助言を頂き、深く感謝する。

参考文献

- [1] Kale, A., Sundaresan, A., Rajagopalan, A., Cuntoor, A., Roy-Chowdhury, N., Kruger, A. and Chellappa, R. V.: Identification of humans using gait, *IEEE Transactions on Image Processing*, Vol. 13, No. 9, pp. 1163–1173 (2004).
- [2] Blank, M., Gorelick, L., Shechtman, E., Irani, M. and Basri, R.: Actions as space-time shapes, *ICCV*, pp. 1395–1402 (2005).
- [3] Kobayashi, T. and Otsu, N.: A Three-Way Auto-Correlation Based Approach to Motion Recognition, *Pattern Recognition Letters*, Vol. 30, No. 3, pp. 185–192 (2009).
- [4] Kobayashi, T. and Otsu, N.: Motion Recognition Using Local Auto-correlation of Space-time Gradients, *Pattern Recogni-*

- tion Letters*, Vol. 33, No. 9, pp. 1188–1195 (2012).
- [5] Otsu, N.: A Threshold Selection Method From Gray-level Histograms, *Systems, Man and Cybernetics*, Vol. 9, pp. 62–82 (1979).
- [6] Iwata, K., Satoh, Y., Kobayashi, T., Yoda, I. and Otsu, N.: Application of the Unusual Motion Detection Using CHLAC to the Video Surveillance, *ICONIP* (2007).
- [7] Dalal, N. and Triggs, B.: Histograms of Oriented Gradients for Human Detection, *CVPR*, pp. 886–893 (2005).
- [8] Schudt, C., Laptev, I. and Caputo, B.: Recognizing human actions: A local SVM approach, *ICPR*, pp. 32–36 (2004).
- [9] Dollar, P., Rabaud, V., Cottrell, G. and Belongie, S.: Behavior recognition via sparse spatio-temporal features, *IEEE International Workshop on Visual Surveillance*, pp. 65–72 (2005).
- [10] Jhuang, H., Serre, T., Wolf, L. and Poggio, T.: A Biologically Inspired System for Action Recognition, *ICCV* (2007).
- [11] Laptev, I., Marszaek, M., Schmid, C. and Rozenfeld, B.: Learning Realistic Human Actions from Movies, *CVPR* (2008).
- [12] Kim, T.-K., Wong, S.-F. and Cipolla, R.: Tensor Canonical Correlation Analysis for Action Classification, *CVPR* (2007).
- [13] Otsu, N. and Kurita, T.: A new scheme for practical flexible and intelligent vision systems, *IAPR Workshop on Computer Vision*, pp. 431–435 (1988).
- [14] Kobayashi, T. and Otsu, N.: Image Feature Extraction Using Gradient Local Auto-correlations, *ECCV* (2008).
- [15] Csurka, G., Bray, C., Dance, C. and Fan, L.: Visual categorization with bags of keypoints, *ECCV Workshop on Statistical Learning in Computer Vision*, pp. 1–22 (2004).
- [16] Wong, S.-F. and Cipolla, R.: Extracting Spatio-temporal Interest Points using Global Information, *ICCV* (2007).
- [17] Duda, R. O., Hart, P. E. and Stork, D. G.: *Pattern Classification*, Wiley-Interscience, 2 edition (2001).
- [18] Vapnik, V.: *Statistical Learning Theory*, Wiley (1998).
- [19] Johansson, G.: Visual perception of biological motion and a model for its analysis, *Perception and Psychophysics*, Vol. 14, pp. 201–211 (1973).
- [20] Cutting, J. and Kozlowski, L.: Recognizing friends by their walk: Gait perception without familiarity cues, *Bulletin of the Psychonomic Society*, Vol. 9, pp. 353–356 (1977).
- [21] Kobayashi, T. and Otsu, N.: A Three-Way Auto-Correlation Based Approach to Human Identification by Gait, *IEEE Workshop on Visual Surveillance*, pp. 185–192 (2006).
- [22] Sarker, S., Philips, P., Liu, Z., Vega, I. R., Grother, P. and Bowyer, K.: The HumanID Gait Challenge Problem: Data Sets, Performance, and Analysis, *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 27, pp. 162–177 (2005).
- [23] Hayamizu, S., Hasegawa, O., Itou, K., Sakaue, K., Tanaka, K., Nagaya, S., Nakazawa, M., Endoh, T., Togawa, F., Sakamoto, K. and Yamamoto, K.: RWC Multimodal Database For Interactions By Integration Of Spoken Language and Visual Information, *International Conference on Spoken Language Processing*, pp. 2171–2174 (1996).
- [24] Nanri, T. and Otsu, N.: Unsupervised Abnormality Detection in Video Surveillance, *MVA*, pp. 574–577 (2005).
- [25] Schölkopf, B., Platt, J., Shawe-Taylor, J., Smola, A. and Williamson, R.: Estimating the support of a high-dimensional distribution, *Neural Computation*, Vol. 13, No. 7, pp. 1443–1471 (2001).
- [26] Kobayashi, T. and Otsu, N.: Cone-Restricted Subspace Methods, *ICPR* (2008).
- [27] Auvinet, E., Rougier, C., J.Meunier, St-Arnaud, A. and Rousseau, J.: Multiple cameras fall dataset, Technical Report 1350, Université de Montréal (2010).