

複数人動画像からの異常動作検出

南里 卓也[†] 大津 展之^{† †}

概要 セキュリティ分野における映像監視や老人介護のモニタリングシステムなどへの応用を踏まえ、固定カメラによる複数人動画像からの異常動作検出手法を提案する。本手法では、画面内で頻繁に行われる動作を通常動作とし、異常動作をその通常動作の特徴分布から逸脱するものとして定義する。動作特徴として立体高次局所自己相関特徴を用いており、この特徴の加法性の性質と部分空間法を組み合わせることによって、画面内に複数人いる場合でも、個々の人物の切り出しやトラッキング無しで学習および検出することができる。実験では、複数人の歩行者の中での転ぶ動作を異常動作として検出し、手法の有効性を確認した。

Detection of Abnormal Motion from a Scene Containing Multiple Persons' Moves

Takuya NANRI[†] and Nobuyuki OTSU^{† †}

abstract This paper proposes an abnormal motion detection method from a scene containing multiple persons' moves for video surveillance. In this method, normal motions are defined as motions that frequently happen, and abnormal motions as motions that depart from normal motion distribution. We adopted cubic higher order local auto-correlation (CHLAC) features for motion features, because the additivity of CHLAC combined with a linear subspace method leads to learning of normal motion and detection of abnormal motion in even a scene containing multiple persons' moves without segmentation of each object or tracking. In this experiment, tumbling motion was detected as abnormal motion in a scene containing multiple persons' walking, and the validity of the method was confirmed.

1 はじめに

現在、カメラによる監視システムが多く利用されているが、人による動画像からの異常検出は労力も多大であり、これをコンピュータで代替できるならば大幅な労力削減につながる。また老人介護においても、発作などに対する自動通報システムがあるならば、介護者の負担を減らすことができるため、異常を知らせるなどのカメラによるモニタリングシステムが必要とされている。

異常動作検出についての先行研究としては、Sekiら [7] や Rao ら [8] が挙げられる。Seki らは、学習画像に対して SOM を適用してクラスタリングし、パラメトリック固有空間法を用いて動作の識別、また、異常動作の検出を行っている。この手法では、静止した画像からの特徴を用いているので、状態遷移のある動作において重要な、時間方向の相関特徴が得られておらず、画像内における位置や形のみに大きく依存していると考えられる。Rao らは、点のトラッキングにより位置と速度を特徴として抽出し、最尤推定により求めた確率密度を用いて異常動作を認識している。この手法では、点同士の空間的な相関関係は特徴として用いておらず、点のフローによる異常検出であると考えられる。

[†] 東京大学大学院情報理工学系研究科

Graduate School of Information Science and Technology,
The University of Tokyo

^{††} 産業技術総合研究所

National Institute of Advanced Industrial Science and
Technology (AIST)

そこで、対象に対して動作特徴を抽出するような動画からの動作認識が必要となる。動作認識の研究としては、オプティカルフローを用いて動物体の領域を切りだし、その領域の特徴から動作の認識をするもの [6][4] や、動作の変化を表現した MHI, MEI という画像を用いて動作を表現し、テンプレートマッチングにより動作を認識する方法 [5] などがある。特に、Kobayashi ら [9] は、顔画像認識などに有効な高次局所自己相関特徴 [2][3] を拡張し、時間方向の相関も加えた特徴である、立体高次局所自己相関特徴を用いた動作認識を行っている。立体高次局所自己相関特徴とはすなわち、画像を時系列に並べたボクセルデータに対して、その中の各点において 251 種類の局所的な自己相関特徴を計算し、この局所特徴をボクセルデータ全体にわたって積分することにより得られる統計的特徴であり、動作特徴であるといえる。この特徴に対して 4 つの動作を判別分析にかけた認識結果は、ほぼ 100 % と非常に高いものとなっている。

そこで、本論文では、動画からの特徴抽出である立体高次局所自己相関特徴を用いて異常動作を検出する統計的手法を提案する。

まず、異常動作の定義であるが、異常の事象を全て列挙することができないように、異常そのものを定義することはできない。しかし、異常は、「普通とはちがうこと。普通でないもの」(広辞苑) とあるように、そこで本論文では、異常動作を、通常動作ではないものとして定義する。通常動作は、動作特徴の統計的な分布を考えた場合に分布の集中する動作であると考えれば、容易に学習可能である。そして、その分布から大きく逸脱する動作を異常動作とする。例えば、通路や道路などに設置された防犯カメラでは、歩く動作のような一般的な動作は通常動作として学習され認識されるが、拳動不審のような動作は、歩く動作のような周期運動でもなく分布として少ないため、異常動作として認識されることになる。本論文では、「歩く」動作を通常動作、「走る」動作、「転ぶ」動作を異常動作として実験を行った。

具体的な手法としては、立体高次局所自己相関特徴による動作特徴空間の中で通常動作特徴の部分空間を構成し、その部分空間からの距離を異常値として異常動作を検出する。通常動作部分空間の構成に

は主成分分析を用い、累積寄与率 0.99 の主成分ベクトルにより主成分部分空間を構成する。

ここで、立体高次局所自己相関特徴には、対象の切りだしが不要で画面内において加法性があるという性質がある。この加法性により、通常動作部分空間を構成すると、画面内に何人の人が正常動作を行っても特徴ベクトルは通常動作部分空間の中に留まることになり、その中の一人でも異常動作を行うと、部分空間から飛び出し異常値として検出できることになる。個々に人を切り出して計算する必要があるため、計算量は人数に比例することなく一定となる。

以下ではまず、2 章で異常動作検出のための手法について詳しく述べ、次に、3 章で提案手法を用いた実験の結果を示す。最後に、4 章で、まとめと今後の課題について述べる。

2 異常動作検出手法

2.1 動きの検出

本手法では、動き部分の検出に、ピクセルごとの時間差分画像を用いる。このとき、画像はグレースケール (256 階調) である。ここで、時間差分画像は、背景の明るさの変化などによるノイズが含まれており、値のあるピクセルが全て対象物体というわけではない。そこでさらに、背景と動きを分離するために二値化処理を行う。この二値化には、大津の判別二値化手法 [1] を用いる。二値化した結果は、図 1 のようになる。

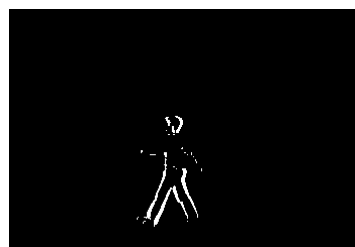


図 1: 二値化した画像

2.2 立体高次局所自己相関特徴

時系列二値画像からの動作特徴の抽出には、立体高次局所自己相関 (CHLAC) 特徴 [9] を用いる。この特徴は、二次元データからの特徴抽出である高次局所自己相関 (HLAC) 特徴 [2] に対して時間方向も加えて三次元に拡張したものである。定式化の形は HLAC 特徴とほとんど変わらず、 N 次自己相関関数は次のようになる。

$$x_f^N(a_1, \dots, a_N) \triangleq \int f(r) f(r+a_1) \dots f(r+a_N) dr \quad (1)$$

ここで、 f は時系列画像であり、変数 r 、 N 個の変位 a_i ($i = 1, \dots, N$) は画像内の二次元座標と、時間を成分として持つ三次元のベクトルである。さらに、時間方向の積分範囲は、どの程度の時間方向の相関を取るかのパラメータである。

変位ベクトルの組み合わせは、二値動画からの HLAC 特徴の場合、0 次は 1 個、1 次は 4 個、2 次は 20 個考えられ、計 25 次元のベクトルとして表されていた (図 2)。CHLAC 特徴の場合は、変位が 3 次元となるので図 3 のような立方体の格子で考えれば良く、二値画像からの CHLAC 特徴の成分の数は、0 次が 1 個、1 次が 13 個、2 次が 237 個であり、合計 251 次元のベクトルとして表されることになる。

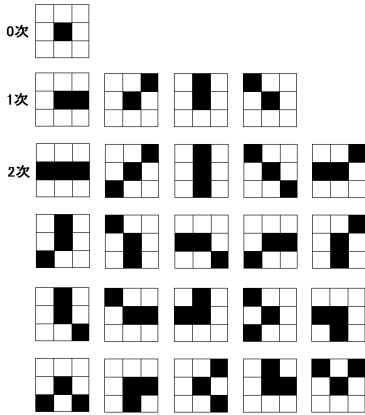


図 2: 高次局所自己相関 (HLAC) の変位パターン

ここで、HLAC 特徴、CHLAC 特徴の重要な性質は、加法性があるという点である。これは、この特

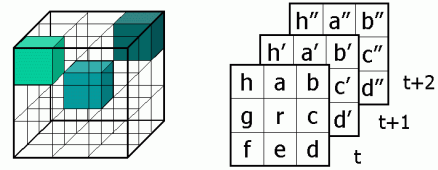


図 3: 立体高次局所自己相関 (CHLAC) の変位パターンの例 ($h \ r \ b$)

徴が画面全体を通して積分しているというところから得られており、例えば、 A という人の動作の特徴を f_A 、 B という人の動作の特徴を f_B としたときに、 A, B 二人が画面内で動作を行った場合、特徴は $f_A + f_B$ となるという性質である (図 4)。

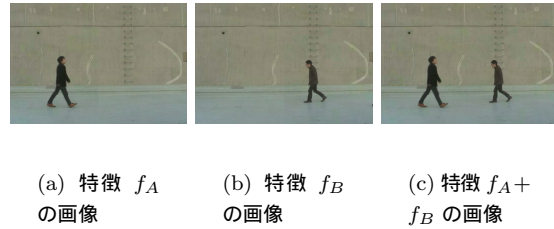


図 4: CHLAC 特徴の加法性

2.3 部分空間法

2.3.1 通常動作の部分空間

まず、通常動作の部分空間を構成するために、主成分分析により主成分ベクトルを求める。主成分ベクトル $U = [u_1, \dots, u_M]$ 、 $u_i \in V^M (i = 1, \dots, M)$ は、共分散行列

$$\Sigma_X \triangleq \frac{1}{N} \sum_{i=1}^N \{(x_i - \mu)(x_i - \mu)^T\} \quad (2)$$

を用いて、

$$\Sigma_X U = U \Lambda \quad (3)$$

の固有値問題より求まる。ここで、 x_i ($i = 1, \dots, N$) $\in V^M$ は M 次元の特徴ベクトル、 μ は、特徴ベクトル x の平均ベクトルである。

固有値行列 $\Lambda = \text{diag}(\lambda_1, \dots, \lambda_M)$ において, 累積寄与率 η_K は,

$$\eta_K \triangleq \frac{\sum_{i=1}^K \lambda_i}{\sum_{i=1}^M \lambda_i} \quad (4)$$

と表され, 累積寄与率 η_K が 0.99 となる次元までの固有ベクトル $u_1, \dots, u_K$ により張られる空間を, 通常動作の部分空間として適用している.

このようにして得られた主成分直交基底 $U_K = [u_1, \dots, u_K]$ によって張られた部分空間への射影子は, $P = U_K U_K'$ として表され, それに対する直交補空間への射影子は $P_\perp = I_M - P$ となる. 直交補空間への射影成分の距離 $d_\perp$ は,

$$\begin{aligned} d_\perp^2 &= \|P_\perp x\|^2 \\ &= \|(I_M - U_K U_K')x\|^2 \\ &= x'(I_M - U_K U_K')'(I_M - U_K U_K')x \\ &= x'(I_M - U_K U_K')x \end{aligned} \quad (5)$$

と表すことができる. 本論文では, 距離 $d_\perp$ を, 異常であるかどうかの指標としての異常値として定義する.

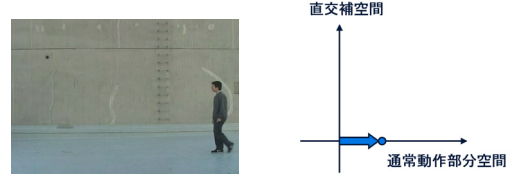
2.3.2 加法性との関係

この部分空間法では, 学習する特徴ベクトルの主成分ベクトルを求め, その主成分ベクトルを用いて部分空間を構成しているが, ここで重要なことは, 特徴ベクトル CHLAC の加法性の性質との相性が非常に良いということである. 部分空間に属するかどうかはベクトルの大きさに依存しない. すなわち, ベクトルの方向のみがどの部分空間に属するかの要素となる. 例えば, 一人の「歩く」という特徴ベクトルの構成する部分空間と, 二人の「歩く」という特徴ベクトルの構成する部分空間とは, 原理的に同じものとなる. したがって, 何人が画面内で通常動作を行ったとしてもそれは全て通常動作の部分空間に留まるが, その中で一人でも異常動作を行ったならば, 方向の異なるベクトルが加わるためその部分空間から飛び出し, 部分空間からの距離として異常を検出することができることになる. 式で表すと次のようにな

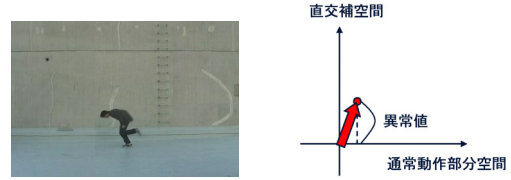
る. (N : normal, A : abnormal)

$$\begin{aligned} x &= x_1^N + \dots + x_n^N + x^A \quad \text{とすると} \\ P_\perp x &= P_\perp (x_1^N + \dots + x_n^N + x^A) \\ &= P_\perp (x_1^N + \dots + x_n^N) + P_\perp x^A \\ &= P_\perp x^A > 0 \end{aligned} \quad (6)$$

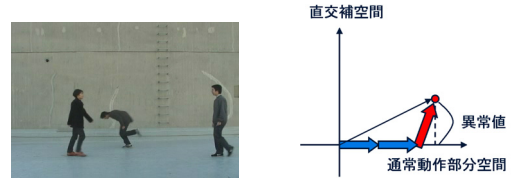
図による簡単な説明を図 5 に示す.



(a) 通常動作 (歩く)



(b) 異常動作 (転ぶ)



(c) 通常動作と異常動作

図 5: CHLAC 特徴の加法性と部分空間の性質

3 実験

実験では, フレーム内に一人存在する場合, フレーム内に複数人存在する場合に対して, 学習フェーズと認識フェーズに分けて本手法の実験を行った. さらに, フレーム内に複数人存在する場合に対して, フェーズを分けずにオンラインで学習する場合の実験を行った.

本実験に用いた特徴のパラメータとして、Kobayashi ら [9] の結果より、立体高次局所自己相関特徴は、 352×240 の解像度の画面に対して、画面内マスク幅 (変位幅) を 10, 時間方向のマスク幅 (変位幅) は 1 とし、積分範囲である時間幅は 20 として与えた。

3.1 フレーム内に一人存在する場合

フレーム内に一人存在する場合の実験では、異常動作検出の過程を学習フェーズと認識フェーズに分ける。学習フェーズでは通常動作のみのデータを用いて通常動作部分空間を学習し、認識フェーズでは通常動作と異常動作のデータを与えて結果の比較を行う。

学習フェーズには、通常動作として「歩く」動作を採用し、被験者 6 人が右方向、左方向にそれぞれ 4 回ずつ「歩く」動作を行ったデータ (計 48 回) を用いる。

認識フェーズでは、学習フェーズの 6 人とは別の一人が、「歩く (walk)」「走る (run)」「転ぶ (tumble)」の動作を行ったデータを用いる。動画の例を図 6、図 7 に示す。ここで、学習、認識に用いられる動画は、人が画面中にいない状態から、フレームイン、フレームアウトするまでのものとなっている。

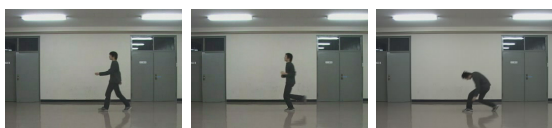
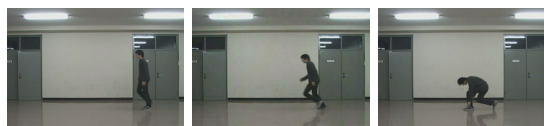


図 6: 認識データの例

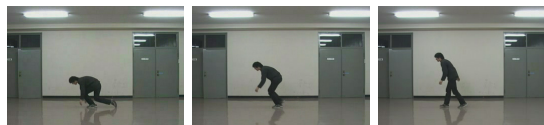
実験の結果を図 8 に示す。この図は、タイムステップ (フレーム) ごとの部分空間からの距離 (異常値) のグラフであり、通常動作である「歩く」は部分空間からの距離が小さく、異常動作である「走る」「転ぶ」は、異常値として大きな値が正しく検出されていることが確認された。ここで、特徴ベクトル 251 次元のうち、寄与率 0.99 となった主成分空間の次元は 12 次元であった。



(a) 20

(b) 30

(c) 40



(d) 70

(e) 80

(f) 90

図 7: 「転ぶ」動作の例 (数字はフレーム番号)

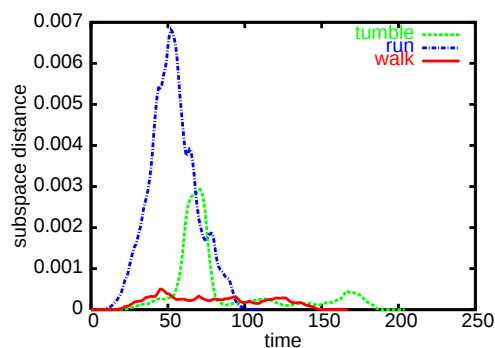


図 8: 異常値の例 (一人の場合)

3.2 フレーム内に複数人存在する場合

次に、フレーム内に複数人存在する場合を実験した。この場合も、学習フェーズと認識フェーズに分けて実験を行った。本実験では、フレーム内に最大三人存在するように設定し、三人とも「歩く」動作を行ってもらった。学習データでは、「歩く」のみのデータを用い、認識データには、「歩く」のみのデータと、一人だけ「転ぶ」データを用いる (図 9)。

実験の結果を図 10 に示す。実験の結果を示すグラフはフレーム内に一人存在する場合の実験のときと同様であり、normal が通常動作の動画、abnormal が異常動作を含んだ動画である。結果として、転んだ場合に正しく異常が検出された。



(a) 三人とも歩くデータ例 (b) 一人だけ転ぶデータ例

図 9: 複数人の例

人が複数人おり、特徴ベクトルの加法性を利用する場合、体の一部が隠れた状態になるフレームインとフレームアウトのときや、人が重なりあう状態のときには加法性は成り立たなくなる。しかし、結果としてはそのような条件下においても異常を正しく検出できたといえる。これは、二つの仮説が考えられる。まず、一つは体の一部が隠れた場合、画面全体に対する変化の割合が小さく、また、通常動作の持つ特徴ベクトルの方向とあまり変わらないということである。そのため、部分空間からの距離としては、大きな値を持たないというものである。

もう一つの仮説は、体の一部が隠れた状態や重なりあった状態の特徴量もある程度学習され、主成分空間に含まれたということである。しかし、体の一部が隠れたときや重なりあったときに小さな変化が出ることもあったことから、全ての重なりの特徴量が主成分空間に含まれているわけではないことが分かる。後者の仮説が正しいのであれば、もし、多くの動作を学習することによって全ての重なりの特徴量を学習した場合、そのような重なりの部分に反応しなくなる可能性がある。

また、寄与率 0.99 となる主成分空間の次元としては、特徴ベクトル 251 次元中 5 次元であった。これは、フレーム内に一人存在する場合よりも少ない結果となっており、自然な予想と異なる。この理由としては、一人の場合は、人に依る歩く動作の微妙な違い(歩様)が部分空間に反映されるが、複数人の場合は、それらの歩様成分が平均化されて、共通的な歩く動作成分が加算されて支配的になるため、部分空間の次元がむしろ小さくなっていると考えられる。

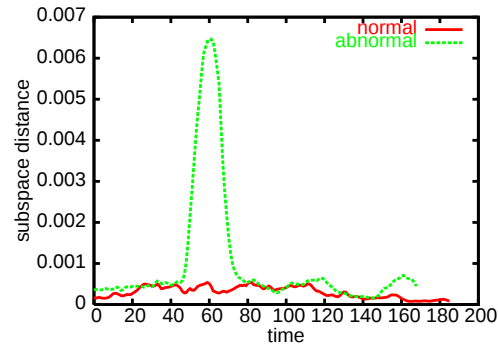


図 10: 異常値 (複数人の場合)

3.3 オンライン学習

これまでの実験では、学習フェーズで通常動作部分空間の学習を行い、その後、異常動作の認識を行った。本節では、学習フェーズ、認識フェーズの違いを設けず、逐次データから通常動作部分空間を構成し、現在の動作の評価をするというオンラインでの学習を行う。すなわち、最初のフレームから現在のフレームまでの共分散行列を逐次計算し、その固有値問題をステップごとに解き、その時点での部分空間を構成して、その部分空間に対する距離を測るというものである。

データは、前節の複数人が行き交う動画を連続的な動画として用いた。異常動作である「転ぶ」動作は、全体 2130 フレーム中で約 1950 フレーム目にあり、それ以外は複数人の「歩く」通常動作のみの動画である。

まず、図 11 は、次元の変化の結果である。total rank の方は、そのタイムステップまでの入力データ全体のランクを示しており、フルランクである 251 次元に到達するのは、629 フレーム目となっている。120 フレーム目までは、一つのサンプルが増えるにつれてランクも一つずつ増加しているが、その後はややなだらかな変化であり、全次元の特徴をとることは、「歩く」のみの特徴ではサンプルが多く必要であることが分かる。通常動作部分空間の次元は、常に 4 次元程度で変化しており、割合としては、一割程度と非常に小さいものとなっている。これは、前節において考察した、歩様成分の平均化が原因であると考えられる。

次に、図 12 は、通常動作部分空間からの距離の時間変化のグラフである。空間の次元が安定するまでは、大きな値が出ているように新しいサンプルは常に異なる次元に飛び出していると考えられる。中盤の変化は、画面内に人がいるか否かでやや時間変化をしている程度で安定しており、最後の「転ぶ」フレーム (1950 フレーム) では、正しく異常値を検出できている。

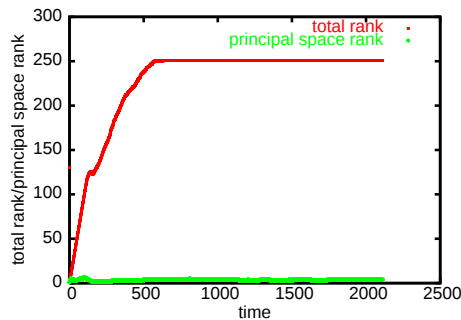


図 11: 入力データの次元と主成分部分空間の次元

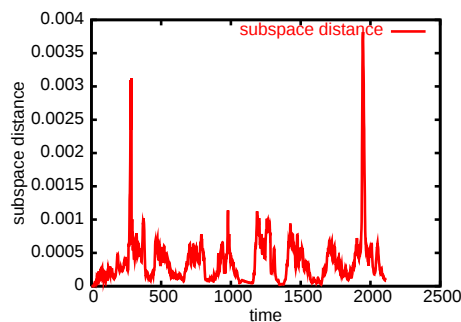


図 12: 異常値の変化 (オンライン学習)

4 おわりに

4.1 まとめ

本論文では、CHLAC 特徴の加法性と部分空間法を用いて、通常動作を統計的に部分空間として学習し、異常動作をそこからの逸脱として検出することができることを示した。この手法は、複数人の場合にも適用することができ、画面内で一人でも異常動作

を行えば、検出することができることも示した。しかも、対象の切り出しは不要で、計算量は人数に依らず一定であり、有効で実用性の高い手法となると思われる。また、この手法では、通常動作を陽に定義することなく統計的に学習しているため、設計の段階で通常動作とは何かについて定義する必要がなく、環境に則した自然な検出を行うことができる。さらに、モデルについての仮定を全くおいていない点で、人の動作に限らず、さまざまな環境および課題に対しても、通常・異常を判別できる汎用的な手法であるといえる。また、リアルタイムへの適用のため、学習フェーズを設定せずに、オンライン学習によっても同様に異常動作を検出することができることを示した。

4.2 今後の課題

まず、通常動作が複数存在する場合の問題がある。本論文の実験では、通常動作が一つの場合で行ったが、二つの場合、あまり正しく通常動作部分空間を構成することができていない。本手法では、一つの部分空間で通常動作部分空間を近似しようとしているため、分布が多峰性の場合には、大きな分布の間の谷に異常動作が含まれる場合など、やや限界があると考えられる。そのため、通常動作部分空間内においてもクラスタリングを行った後、そこからの距離を測るなど、多峰性の分布にも対応できるよう改善することが必要であり、今後の課題である。

また、主成分部分空間の構成における寄与率の決定の問題がある。本研究では寄与率 0.9, 0.99, 0.999 を比較し、経験的に最も異常値が判別される 0.99 で実験を行ったが、この値が適切かどうかの定量的評価は行っていない。そのため、寄与率の変化における主成分部分空間の構成についての検討も必要である。

最後に、この手法が人の動作に限らず、正常・異常を判別することができる汎用的な手法であると考えられることから、この手法の特性や可能性を調べるために、さまざまな状況設定での実験を行う必要がある。さらに、アプリケーションとしての実装を通して、本手法の定量的評価を行う必要がある。

参考文献

- [1] 大津 展之, “判別および最小 2 乗規準に基づく自動しきい値選定法”, 電子通信学会論文誌 D, J63-D-4, 349-356, 1980.
- [2] N. Otsu and T. Kurita, “A New Scheme for Practical Flexible and Intelligent Vision Systems”, in Proc. of IAPR Workshop on Computer Vision, 431-435, 1988.
- [3] T. Kurita, N. Otsu and T. Sato, “A Face Recognition Method Using Higher Order Local Autocorrelation and Multivariate Analysis”, in Proc. IAPR International Conference on Pattern Recognition, 213-216, 1992
- [4] S. X. Ju and M. J. Black and Y. Yacoob, “Cardboard people: A parameterized model of articulated motion”, International Conference on Automatic Face and Gesture Recognition, 38-44, 1996.
- [5] Romer Rosales, “Recognition of Human Action Using Moment-Based Features”, 1998-020, 1998.
- [6] Yaser Yacoob and Michael J. Black, “Parameterized Modeling and Recognition of Activities”, Computer Vision and Image Understanding, 73, 2, 232-247, 1999.
- [7] H. Seki and Y. Hori, “Detection of Abnormal Human Action Using Image Sequence”, in Proc. International Power Electronics Conference, pp.1272-1277, 2000.
- [8] Supriya Rao and P.S. Sastry, “Abnormal activity detection in video sequences using learnt probability densities”, TENCON, 2003.
- [9] T. Kobayashi and N. Otsu, “Action and Simultaneous Multiple-Person Identification Using Cubic Higher-Order Local Auto-Correlation”, International Conference on Pattern Recognition, 2004.