

H-043

シルエット特徴量のテンソル分解に基づく動画像から人物の動作認識手法の検討

Study of Recognizing Human Actions on Silhouette Feature from a Video Sequence by a Tensor Decomposition Based Approach

堀 隆之[†] 大谷 淳[†] 榎 順[‡]
Takayuki HORI[†] Jun OHYA[†] Jun KURUMISAWA[‡]

1. はじめに

画像処理技術において動画像から対象の動きや動作を認識することは重要な研究テーマである。店舗や人の往來の激しい場所には防犯などの目的で監視カメラが設置され、画像処理を用いた人物の行動認識技術の発展は監視システムの自動化やシーンの定常性評価などへの応用が期待されている。人物の行動などが映された動画像には非常に有益な情報が含まれているが、画像処理技術を用いてそれらを理解するためには見え隠れに対する問題や人の姿勢の多様性に関わる問題など、基本的で重要な課題が残されている。

Vasilescu ら[1]によると、人物の顔や筆跡から個人を特定し認証ができるように、人物の動作にも特有の表象が存在するため、それを利用することで個人や動作を特定することが可能とある。Vasilescu の取り組みはベクトル空間で扱われてきた人物における動作の情報を、より多次元的にテンソル空間へと拡張したものであり、多次元テンソル空間におけるテンソル分解[2]によって求められる。Vasilescu の手法は N-mode SVD によって、人の動作からその表象を抽出した。ただし彼らは認識過程においてデータベースに含まれない未知の人物を対象にしていなかった。

本研究は Vasilescu の手法を用いて画像処理に基づく動画像からの未知の人物の動作の認識を行う。その際、人物・動作・時間における特徴量をテンソル空間に格納し、認識過程において同様に撮影された未知人物の動作の判定に利用する。本論文ではテンソル分解に基づく動画像の認識法を検討し、その認識率と PCA (Principal Component Analysis), NN 法 (Nearest Neighbor Method) を用いた認識率を比較することでその有効性を検証する。

2. 未知人物の動作認識アルゴリズム

2.1 コアテンソルの算出

テンソルとはベクトルの概念を拡張したものである。テンソルでは多次元空間を扱うことができ、多変量に独立して変動する値に対して多次元に空間を与えることが可能である[1]。3次以上のテンソル空間に対して SVD は適用されないため、ここでは HOSVD (High Order SVD : N-mode SVD) を利用する。これはテンソル \mathcal{A} において SVD によって分解された N 次対角空間を $\mathbf{U}_1, \mathbf{U}_2, \dots, \mathbf{U}_N$, コアテンソルを \mathcal{Z} とすると、HOSVD は式 (1) によって求められる。

$$\mathcal{A} = \mathcal{Z} \times_1 \mathbf{U}_1^T \times_2 \mathbf{U}_2^T \dots \times_n \mathbf{U}_n^T \dots \times_N \mathbf{U}_N^T \quad (1)$$

このコアテンソル \mathcal{Z} は行列によって定義される点から一般

[†] 早稲田大学大学院国際情報通信研究科
Graduate School of Global Information and
Telecommunication Studies, Waseda University

[‡] 千葉商科大学政策情報学部
Dept. of Policy Informatics, Chiba University of Commerce

的な SVD の結果と類似しているが、多次元を有している点で異なる。HOSVD のアルゴリズム[1]は以下のとおりである。

$n = 1$ から N までにおいて

1. テンソル \mathcal{A} について n 次元の行列 $\mathbf{A}_{(n)}$ を構築する。
2. 行列 $\mathbf{A}_{(n)}$ において SVD を求める。
3. \mathbf{U}_n に SVD の結果を代入する。

ここで $P \times A \times T$ (P は人物, A は動作の種類, T は時系列データ) のサイズを持つテンソル \mathcal{D} を定義すると、コアテンソル \mathcal{Z} は式(2)によって与えられる

$$\mathcal{D} = \mathcal{Z} \times_1 \mathbf{P}_1 \times_2 \mathbf{A}_2 \times_3 \mathbf{T}_3 \quad (2)$$

2.2 認識アルゴリズム

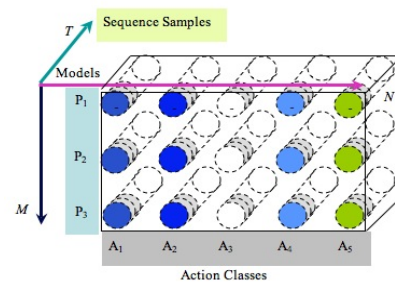


図 1. 動作におけるテンソルの構造

テンソル \mathcal{Z} が求まった後、未知人物の動作データ \mathcal{D}_{pa} が入力された際、各動作についてこの未知人物のデータをデータベース内の一人のデータと置き換え、その未知人物の動作の表象は次の式 (3) で求められる

$$\mathcal{D}_{pa} = \mathcal{B}_a \times_1 \mathbf{P}^T \quad (3)$$

ここで $\mathcal{B}_a = \mathcal{Z} \times_2 \mathbf{A}_a^T \times_3 \mathbf{T}$ であり、 \mathcal{D}_{pa} は $1 \times 1 \times T$ のテンソルである。認識過程において、この \mathcal{D}_{pa} がデータベース内の既存のデータと最も類似しているものを探すことになる。そのための評価手段として、本論文ではコアテンソル \mathcal{Z} を用いることにする。ただし入力されるデータのサイズに対してデータベースのサイズが大きいため、未知人物の動作のテンソル \mathcal{D}_p とデータベースのテンソル \mathcal{D} を単純に比較することはできない。よってデータベース内の一人と入力データを入れ替えることで、データベースとサイズの等しいテンソル \mathcal{D}_p を算出した。この作業は人物、各動作のすべてに対し行われ、人物と動作の種類の組み合わせの数だけ繰り返される。ここで元のデータと入力データが似ている場合、互いの値を入れ替えてもコアテンソルの値は類似する。つまり元のデータベース内の各動作に関するコアテンソルの値と入力データによって新しく算出されたコアテンソルの差が最も小さくなる場合を探し出し、その結果を未知人物の動作の判定に利用する。

3. 未知人物の動作認識実験

3.1 シルエット画像特徴量

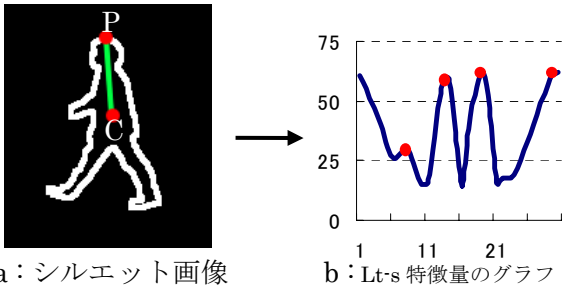


図2. シルエット画像と Lt-s feature

未知人物の動作を認識するために、動画画像から背景差分法により人物のシルエット画像を抽出し、そのシルエット画像の特徴量を利用することでコアテンソルが求められる。本論文ではこのシルエット画像の特徴量に関する記述子として図形の形状認識手法として主に知られている Lt-s feature[3]を使用した。Lt-s 特徴量とは人物のシルエット画像の各輪郭画素に対して図形の中心と最頂点との距離の和によって求められる特徴量である。ここで人物のシルエット画像の重心を C、上半身の主軸と輪郭線が交わる点を P、現在の輪郭画素を A とすると、Lt-s 特徴量の算出方法は、式 (4) で求められる。

$$Lt-s = CA + AP \quad (4)$$

3.2 認識実験

動作の認識実験に際し、互いに類似した歩行系の基本的動作を 7 種類選び、33 名の大学生を対象に撮影を行った。ここで対象にした動作は「歩く」「走る」「スキップ」の動作に対し、歩幅にバリエーションを持たせた計 7 種類である。ただし認識に際して動作間での運動周期をそろえておく必要があるため、実際に実験で使用した動画画像では動作の 1 周期を 10fps にそろえて運動周期の正規化処理を行っている。

認識実験で提案手法と PCA[4]、NN 法を利用した結果を比較する。ここでの PCA (Principal Component Analysis) とはシルエット特徴量に対して主成分分析から求められる動作の固有ベクトルを用いて行う認識手法であり、NN 法 (Nearest Neighbor Method) とは動作についてシルエット特徴量を平均化しその距離から動作を認識する手法である。ここではデータベースから 10 人のデータを用いてテンソルを構築し、残りのデータを認識用に利用した。そしてデータベースに用いるサンプルの組み合わせを変えながら、各動作に対して認識実験を行った。

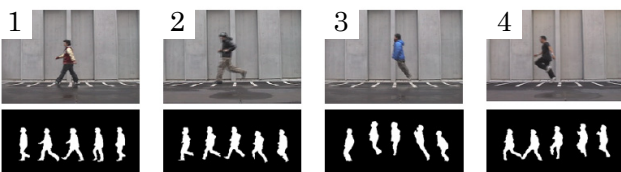


図3. 実験で利用した動作の動画画像とシルエット画像 (1: walk 2: run 3: jump 4: skip)

3.3 認識結果

未知人物の動作の認識に対してテンソル分解に基づく手法を用いた場合と PCA, NN 法に基づく認識結果を図 4. に示す。提案手法では 7 つの動作について認識結果は 80% を上回っておりその有効性を伺える。特に「歩く」動作以外では認識率の平均は 91.7% でありその有効性が確認できる。一方 PCA を用いた場合の認識率をみると 73%, 「歩く」動作以外で平均 85.5% である。PCA を用いた場合において、「歩く」動作の認識率が低い理由として、バリエーションとして用意した walk1, walk2 の動作の違いが判別できないため、互いの動作を誤検出したために認識率が下がったと考えられる。また NN 法の場合、認識率は 70% 前後であるが、各動作に対して安定した認識率が得られている。ところで各動作について比較するとすべての動作について提案手法は PCA, NN 法と比べて高い認識率を示しており、優位性を確認できる。これらの結果より、提案手法は PCA, NN 法を用いた場合と比較して各動作について安定的に認識が行え、高い認識率が得られる手法であると言える。

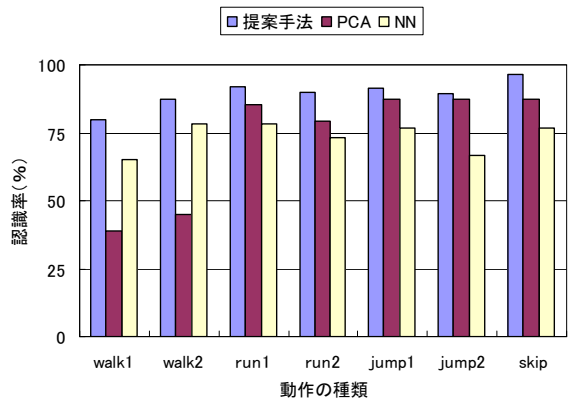


図4. 各認識手法による認識率の比較

4. まとめ

本論文ではテンソル分解を用いた未知人物の行動認識手法について提案し、7 つの動作に対する認識結果を PCA, NN 法と比較することで、提案手法の有効性を確認した。今後は対象とする動作のバリエーションと被験者の数を増やすことで、信頼性の高い認識結果を求めたい。

参考文献

- [1] M.A.O. Vasilescu, "Human Motion Signatures: Analysis, Synthesis, Recognition", International Conference on Pattern Recognition (ICPR'02).
- [2] T. G. Kolda. Orthogonal tensor decompositions. SIAM J. on Matrix Analysis and Applications, 23(1):243-255, 2001.
- [3] K. Takahashi, T. Sakaguchi, and J. Ohya, Remarks on a Real-Time 3D Human Body Posture Estimation Method using Trinocular Images, 15th International Conference on Pattern Recognition, Vol. 4, pp.693-697, Barcelona, Spain, September 2000.
- [4] M.A. Turk and A.P. Pentland, "Face recognition using eigenfaces," Proceedings of Computer Vision and Pattern Recognition, pp.586-591, 1991
- [5] Bradski G., Pisarevsky V., Intel's computer vision library. In Proceedings of Computer Vision and Pattern Recognition, vol. 2, pp. 796-797, 2000.