

動作・外観特徴を用いた段階的クラスタリングによる機械作業映像の教師なし分節化
 Unsupervised Temporal Segmentation of Operation Video of Industrial Equipment
 based on Gradual Clustering using Motion and Appearance Features

河森 大樹[†]
 Kawamori Daiki

中村 和晃[‡]
 Nakamura Kazuaki

新田 直子[‡]
 Nitta Naoko

馬場口 登[‡]
 Babaguchi Noboru

1. はじめに

工場等の生産現場では、作業実態の把握や作業スケジュールの効率化等を背景として、作業内容の自動解析（どのような作業がどの日時・時間帯に行われていたかを自動的に解析すること）に対する需要が高まっている。本稿では、生産現場でも比較的導入の容易なカメラからの映像を用いて作業解析を行う手法を検討する。

映像からの作業解析は、入力映像を複数の分節へ分割したのち、各分節に対し作業ラベルを付与することにより実現される。このような処理は調理等の日常作業を対象に広く検討されている[1,2]が、その多くは教師あり学習に基づいている。教師あり学習では、分節化・ラベル付与済みの映像が教師データとして多数必要となるが、その収集を実際の機械作業現場で行うことは人的コストの観点から現実的でない。そこで本稿では、教師データを必要としない、教師なし学習に基づく作業解析手法を提案する。

2. 機械作業映像の分節化手法

提案手法では、日時・場所・機体・操縦者等に多様性のある多数の機械作業映像が利用できるものと想定する。この条件下では、同一の作業に対応する区間には映像によらず共通した特徴が含まれるものと考えられる。このことを踏まえ、本手法では、各映像を初期的な分節に分割したのち、類似する初期分節同士を単一のクラスタにまとめることにより作業解析を実現する。クラスタリングに際し利用する特徴量として、映像には一般に動作に関する特徴と外観に関する特徴の二種類が含まれるが、機械作業には動作・外観特徴のいずれか一方にしか共通性を持たない作業も存在すると考えられる。このような作業は、動作・外観特徴を単純に併用するクラスタリングでは過度に多くのクラスタに細分化される可能性が高い。そこで本手法では、動作・外観特徴を併用せず、各々を段階的に用いることによりこの問題の解決を図る。具体的には、図1の概要図に示す3ステップにより機械作業の自動解析を実現する。以下、各々の詳細を2.1～2.3節で述べる。

2.1 映像の固定長分節への分割

まず、入力映像を長さ L の固定長分節 s_1, \dots, s_n に分割する。これらは初期的な分節であり、最終的な分節化結果とは異なる。すなわち、隣り合う初期分節同士が後続のクラスタリングにより同一のクラスタに割り当てられた場合、それらは単一の分節に統合される。 L は、その値が大きいほど多くの情報が各分節に含まれ、クラスタリングの精度が向上すると期待されるが、短時間で終了する作業の抽出は困難となる。従って本手法では L の値を実験的に定める。

2.2 分節ごとの特徴量の抽出

前節で取得した初期分節の各々から動作特徴と外観特徴を抽出する。具体的には、フレームごとの特徴量をまず抽出し、その時系列として分節ごとの特徴量を定める。

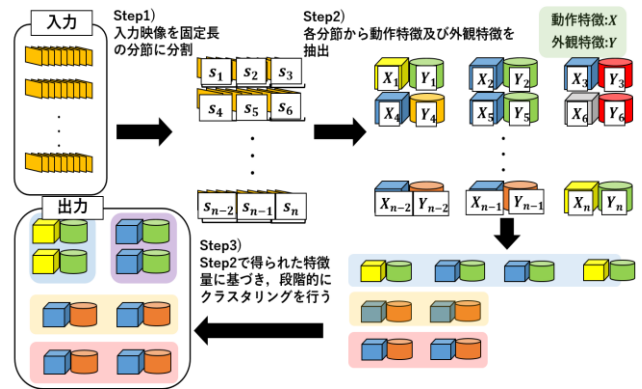


図1 提案手法の概要

2.2.1 動作特徴

本手法では、DeepFlow [3]により推定したオプティカルフローを用いてフレームごとの動作特徴を抽出する。

初期分節 s_i 中の各フレーム j ($j = 1, \dots, L$)について、対象フレームとその直後のフレームからオプティカルフローを計算する。これを H_{ij} とする。次に、 H_{ij} から特徴量を抽出する。本手法では、 H_{ij} に含まれる情報をできるだけ保存した特徴量を得るため、自己符号化器(Autoencoder;AE)に H_{ij} を入力し、AEの中間層から得られるベクトル x_{ij} を特徴量とする。以上により得られたベクトルの系列 x_{i1}, \dots, x_{iL} を初期分節 s_i の動作特徴 X_i とする。

2.2.2 外観特徴

初期分節 s_i 中の各フレーム j ($j = 1, \dots, L$)について、その画像を I_{ij} とする。動作特徴の場合と同様、 I_{ij} に含まれる情報をできるだけ保存した特徴量を得るため、 I_{ij} をAEに入力し、その中間層からベクトル y_{ij} を抽出する。このベクトルの系列 y_{i1}, \dots, y_{iL} を初期分節 s_i の外観特徴 Y_i とする。

2.3 分節集合の段階的クラスタリング

動作特徴および外観特徴に基づいて、初期分節の集合をクラスタリングする。前節の方法で抽出した X_i と Y_i は何れもベクトルの系列であり、その分布は複雑なものになると考えられることから、本手法ではそのような場合にも対応可能なスペクトラルクラスタリング[4]を採用する。スペクトラルクラスタリングにより分節集合 $\{s_i | i = 1, \dots, n\}$ を K 個のクラスタに分割する手順は次の通りである。

- (1) s_i と s_j の類似度をDPマッチングにより算出し、その値を i 行 j 列成分とする重み行列 W を作成する。
- (2) W に基づいてグラフラプリアンを計算し、その固有値・固有ベクトルを上位 K 個分求める。
- (3) K 個の固有ベクトルを列ベクトルとして並べた行列を作成し、その i 行目の行ベクトルを分節 s_i の新たな特徴量とみなす。その上で K -meansを実行する。

本手法では、まず動作・外観特徴の何れか一方のみに基づいて上記の手順を適用する。これにより得られたクラスタのうちサイズが大きいものを対象として、他方の特徴に基づき上記手順を再度適用する。二段階目のクラスタリン

表 1 建設機械作業ラベルの種類と出現割合

正解ラベル	作業項目	フレーム数	割合(%)
0	停止	21022	24.3
1	掘削	4059	4.7
2	積荷ホイスト旋回	5056	5.8
3	ダンプ待ち	3889	4.5
4	ダンプ積み込み	4624	5.4
5	空荷旋回	3245	3.8
6	その他排土	239	0.3
7	その他作業	2626	3.0
8	分類なし	41640	48.2

グにおいて追加分割の対象とするクラスタは、以下の条件に基づいて設定する。

一、二段階目のクラスタリングにおけるクラスタ数をそれぞれ K_1 , K_2 としたとき、二段階目のクラスタリングにおいて、 $K_1/2$ 個のクラスタが追加分割対象になると想定した場合、追加分割後の一クラスタあたりの要素数の期待値は

$$\Phi = \frac{n}{\left(\frac{1+K_2}{2}\right)K_1} = \frac{2n}{(1+K_2)K_1}$$

となる。これを踏まえ、本手法では、一段階目のクラスタリングの結果得られたクラスタのうち、要素数が 2Φ より大きいクラスタを追加分割の対象とする。

3. 評価実験

実際の作業現場で撮影された建設機械作業の映像を対象として実験を行った。映像の総数は 12 本であり、その各々で作業日時・場所・操縦者は異なっている。計算時間短縮のため、各映像から 240 秒の区間をランダムに抽出し分節化の対象とした。なお、評価のため、対象区間にはフレーム毎に表 1 に示す作業ラベルを手で付与した。上記の映像群に対し、提案手法を用いた場合と、動作・外観特徴を単純に併用した場合とでそれぞれクラスタリングを行い、分節化結果を比較した。初期分節の長さ L は、1, 3, 5, 10[秒] の 4 段階を設定して各々の条件下で実験を行い、その結果最も良い性能を示した 10[秒] を採用した。クラスタ数は、一段階目を 5, 6, 7, 8 の 4 段階、二段階目を 2, 3, 4 の 3 段階に設定した。また、分節化結果の評価尺度としては F 値を用いた。

実際に付与した作業ラベルの種類数が 9 であったことを考慮し、最終的なクラスタ数が 9 となった場合の比較結果を図 2 に示す。一段階目でどちらの特徴を用いるか、などの条件にもよるものの、提案手法が比較対象に比べ高い F 値を示すことが分かる。比較対象では F 値が 0.4 弱であるのに対し、提案手法では最大で 0.47 となっており、20%程度の性能向上がみられる。このことから、提案手法の有効性がある程度確認できた。提案手法による分節化結果の具体例として、一段階目に $K_1 = 7$ として外観特徴でクラスタリングを行った後、二段階目に $K_2 = 3$ として動作特徴でクラスタリングを行った場合のクロス表を表 2 に示す。表 2 の結果では、「静止」「分類なし」といった動きの少ない作業に対応するクラスタが大半を占め、動きのある作業が十分に抽出できていない。この原因としては、各フレームの情報ができるだけ保存されるような特徴量を用いたため、背景など一部の作業にしか有効でない情報がクラスタリン

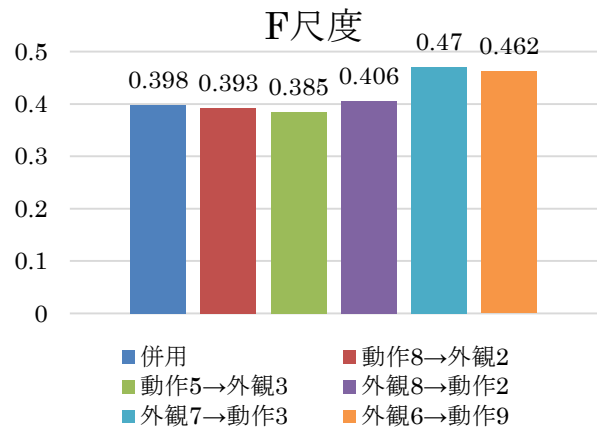


図 2 分節化結果の比較

表 2 分節化結果の具体例 ($K_1 = 7$ (外観), $K_2 = 3$ (動作))

推奨クラスタ番号	評価用正解ラベル								
	停止	掘削	積荷旋回	ダンプ待	ダンプ積	空荷旋回	排土	その他	分類なし
0	7	0	0	11	2	0	0	4	10
1	0	1	1	0	7	7	1	9	53
2	1	0	8	0	11	3	1	13	30
3	35	0	0	0	0	0	0	0	0
4	0	0	0	0	0	0	0	0	15
5	24	0	0	0	0	0	0	0	0
6	0	0	0	0	0	0	0	0	5
7	0	0	0	0	0	0	0	0	19
8	0	0	0	0	0	0	0	0	10

グの際の支配要因となった可能性が考えられる。従って、この問題の解決には特徴量の再検討が必要である。

4. おわりに

本稿では、動作・外観特徴を段階的に利用したクラスタリングによる機械作業映像の分節化手法を提案した。実験の結果、提案手法の有効性がある程度示唆されたが、特徴量の問題などもあり、有意義な分節化結果が得られたとは言いがたかった。教師なし学習では適切な特徴量を適応的に設計することは困難であることから、今後は、一部に人の手を介した半教師あり学習に基づく手法へと提案手法を拡張することを検討する。

謝辞

データ提供等の面で多大なご協力を賜りました(株)小松製作所 生産技術開発センタに深く感謝申し上げます。

参考文献

- [1] J. Lei et al.: "Fine-Grained Kitchen Activity Recognition Using RGB-D," Proc. of ACM Conf. on UbiComp, pp.208-211, 2012.
- [2] S. Bansal et al.: "Kitchen Activity Recognition Based on Scene Context," in Proc. of IEEE ICIP, pp.3461-3465, 2013.
- [3] P. Weinzaepfel et al.: "DeepFlow: Large Displacement Optical Flow with Deep Matching," in Proc. of IEEE ICCV, pp.1385-1392, 2013.
- [4] A. Y. Ng et al.: "On Spectral Clustering: Analysis and an Algorithm," in Proc. of Int'l Conf. on NIPS, pp.849-856, 2001.

‡ 大阪大学工学研究科

Graduate School of Engineering, Osaka University