

研究速報

調理映像中の繰返し調理動作の検出

久原 卓^{†a)} (学生員) 出口 大輔[†] (正員)

高橋 友和^{††} (正員)

井手 一郎^{†b)} (正員：シニア会員)

村瀬 洋[†] (正員：フェロー)

Detection of Repetitive Cooking Motions in a Cooking Video

Taku KUHARA^{†a)}, Student Member, Daisuke DEGUCHI[†],

Tomokazu TAKAHASHI^{††}, Members,

Ichiro IDE^{†b)}, Senior Member, and Hiroshi MURASE[†], Fellow

[†] 名古屋大学大学院情報科学研究科, 名古屋市

Graduate School of Information Science, Nagoya University,
Nagoya-shi, 464-8601 Japan

^{††} 岐阜聖徳学園大学経済情報学部, 岐阜市

Faculty of Economics and Information, Gifu Shotoku Gakuen
University, Gifu-shi, 500-8288 Japan

a) E-mail: tkuhara@murase.m.is.nagoya-u.ac.jp

b) E-mail: ide@is.nagoya-u.ac.jp

あらまし 本論文では、調理映像中の「切る」や「混ぜる」といった繰返し動作を検出する手法について検討した結果を報告する。繰返し動作の検出には、映像フレーム中の動作位置に依存しない特徴 (CHLAC 特徴) と依存する特徴 (2 値特徴) を用いる。これにより、繰返し動作領域の位置が大きく移動する動作及び一定である動作の両方に対応する。実際の料理番組映像に本手法を適用した結果、F 値 0.86 の精度で検出できることを確認した。

キーワード 調理映像, 動作解析, CHLAC 特徴, フーリエ解析

1. まえがき

近年、画像や映像を用いたマルチメディア技術が我々の日常生活支援に導入されつつある。本研究で注目する調理活動に対する支援としては、Web 上の料理レシピや動画投稿サイト上の調理映像などが挙げられる。例えば、大手料理レシピサイト COOKPAD^(注1) は、ユーザによる新たなレシピの投稿により、日々その情報量を増やしている。しかし、レシピ自体はテキストと数枚の画像が付加されたものが主流である。このような我々が普段よく目にするレシピに対し、各調理過程に対応した調理映像を付加することができれば、レシピの利便性が高まると考えられる。この映像付加には、調理動作と調理食材が索引付けされた大量の調理映像が必要である。そのため、動作や食材に対応する調理映像を自動的に抽出し、索引付けする手法が必要である。

調理動作に基づく索引付けの従来手法として、浜田

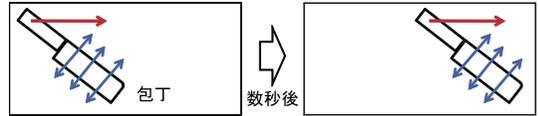


図1 動作位置が移動する繰返し動作 (「繰り返し切る」調理映像)

Fig.1 A repetitive motion whose location moves. (A cooking video of “Cut repeatedly with a knife”)

ら [1] は、重要な調理動作である可能性が高い繰返し動作の検出を行った。具体的には、映像中の局所領域に対する時間方向の周波数解析を利用した。一般的に、図 1 に示すような千切りやフライパンで食材を炒めている様子をカメラを移動させながら撮影する場合、繰返し動作の位置が移動する。浜田らの手法では、このような映像中の動作位置が移動するような繰返し動作は正しく検出できない。一方、カイラ [2] は、画像全体の見えの時間的変化を解析することにより、繰返し動作の検出を行った。この手法では、動作速度が速い場合やズームインされている場合のような、動作領域が大きい非繰返し動作を誤って検出する問題があった。そこで本論文では、動作領域の大きさに依存せず、かつ動作位置が移動する繰返し動作も検出可能な手法を提案する。

2. 提案手法

提案手法は、動作位置に依存しない特徴 (CHLAC 特徴) と、依存する特徴 (2 値特徴) を組み合わせることで用い、入力映像の繰返し動作を検出する。CHLAC (Cubic Higher-order Local Auto Correlation; 立体高次局所自己相関) 特徴 [3] は、画面全体に現れる時空間変動、具体的には映像中の動作が起きている領域の「形状」及び「動き」を表現する画像特徴である。CHLAC 特徴は、映像中の対象物体の位置に不変な性質をもつため、図 1 のような動作位置が移動するような繰返し動作の周期性を捉えられると考えられる。一方、2 値特徴は対象物体の位置に依存する性質をもち、動作位置がほぼ一定の繰返し動作の周期性を解析する際に有効と考えられる。この性質は従来手法と共通であるが、従来手法が抱えていた「動作の大きさ」に対する依存性をもたない。この点を考慮して、位置に依存する 2 値特徴を採用することにする。これら 2 種類の異なる性質をもつ特徴を用いることにより、動作位

(注1): クックパッド (株), “毎日の料理を楽しみに COOKPAD.” <http://cookpad.com/>

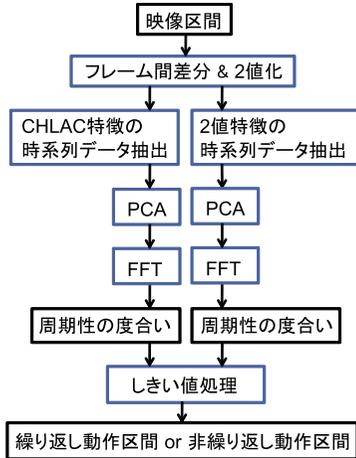


図2 解析窓の処理の流れ

Fig. 2 Process flow of the analysis window.

置が移動する動作及び一定である動作の両方を検出する。次に、各特徴の周期性の度合をフーリエ解析により求め、繰返し動作を検出する。

2.1 繰返し動作区間の検出

入力映像中の繰返し動作検出手法について述べる。その流れを図2に示す。まず、入力映像に対してフレーム間差分と2値化により、2値画像列を得る。次に、この画像列からCHLAC特徴及び2値特徴の時系列データを抽出する。CHLAC特徴の時系列データは、画像3フレームごとに一つのCHLAC特徴を抽出し、1フレームずつずらしながら抽出を繰り返して得る。2値特徴とは、画像の各画素値をラスタスキャン方式でベクトル化したものである。これら2種類の時系列データに対して、以下同様の処理を行う。

まず、CHLAC特徴と2値特徴の各時系列データを主成分分析により次元削減する。そして、 M 次元の特徴ベクトルの時系列データを得る。図3は、繰返し動作の各繰返しの位置が時間経過とともに大きく移動する映像から求めた、各特徴の第1主成分の時系列データの例である。CHLAC特徴の位置不変性により、CHLAC特徴の時系列データは2値特徴の時系列データよりも、周期性が明瞭に観察できる。次に、各次元の値だけを取り出した時系列データに対してFFTを適用し、 M 個のワースペクトルを得る。一般に、映像中の動作領域が大きい場合、CHLAC特徴と2値特徴は大きな値となる。そのため、FFTにより得られるワースペクトルも大きくなる。そこで、映像中の動作の大きさの影響を減らすため、各ワースペクトル

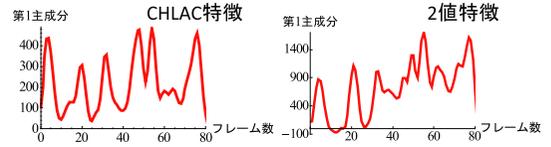


図3 各特徴の第1主成分の時系列データの例

Fig. 3 Example of the time series data of the first principal component of each feature.

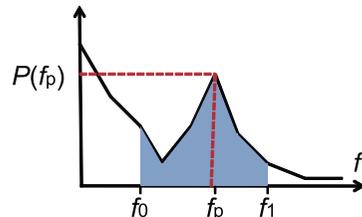


図4 パワースペクトルの形状からの特徴抽出

Fig. 4 Feature extraction from the shape of a power spectrum.

ルに対し、パワーの総和が1になるように正規化する。この処理により、以降の特徴抽出において、映像中の動作領域の大きさに影響されない特徴抽出が可能となる。

続いて、文献[1]で採用された周期性の度合を表す五つの特徴量を抽出する。そのために、図4に示すような、人の繰返し動作の速さを考慮した周波数帯 $f_0 \leq f \leq f_1$ を設定する。そして、この範囲内のパワーのピークに対応する周波数 f_p を求め、スペクトルにおける $P(f_p)$ の突出度、範囲内のパワーの総和など五つの特徴量を抽出する。文献[1]では、これらの特徴量に加え、繰返し動作の条件として周波数 f_p に対するしきい値処理を行っている。しかしながら、本手法では様々な速さの繰返し動作を対象とするため、この特徴を用いない。

そして、各スペクトルの各特徴量に対し正の重み付け和をとり、「周期性の度合」と定義する。最後に、CHLAC特徴と2値特徴のそれぞれから計算した周期性の度合の重み付け和をとることで一つの実数値を求め、その値に対するしきい値処理により繰返し動作を検出する。

3. 実験

3.1 実験映像及び実験条件

実験映像として、NHKの料理番組「きょうの料理」10本（総計250分）から323個の手元ショット区間（最短2秒、最長70秒、平均17秒）を人手で切り出

表 1 繰返し動作検出の実験結果 (F 値)

Table 1 Experiment result for detecting repetitive motions (F-measure).

提案手法	CHLAC 特徴 & 2 値特徴	0.86
	CHLAC 特徴	0.81
	2 値特徴	0.79
従来手法	局所解析 [1]	0.67
	大域解析 [2]	0.71

表 2 動作位置の移動, 及び動作の大きさに関する実験結果

Table 2 Experiments result about shift of motion location or magnitude of motion space.

	データ数	適合率	再現率	F 値
位置の移動あり	1,092	0.83	0.82	0.83
位置の移動なし	2,483	0.87	0.90	0.88
動作領域大	1,972	0.90	0.84	0.87
動作領域小	1,603	0.85	0.82	0.84

した. 手元ショットとは手元が拡大されて撮影された映像であり, 調理上重要な調理過程であることが多いため, 索引付け調理映像として有用である. これらの区間から, 10 フレーム (約 0.3 秒) ずつずらしながら, 80 フレーム (約 2.7 秒) の映像区間を複数人手で抽出した. その結果, 10,726 個の区間が得られた. 次に, 繰返し動作が行われている区間か否かのラベル付けを行ったところ, 繰返し動作区間は 3,575 個, それ以外の区間が 7,151 個であった. また, 繰返し動作の判定はこの 80 フレームの区間それぞれに対して行った. また, 主成分分析のパラメータ M は, CHLAC 特徴においては $M = 4$, 2 値特徴においては $M = 7$ とし, 各特徴の累積寄与率はともに 70% とした.

3.2 実験結果

まず, 提案手法と従来手法を上述の実験データに適用し, 評価を行った. 具体的には, CHLAC 特徴と 2 値特徴の両方を用いた手法, CHLAC 特徴及び 2 値特徴のいずれか一つの周期性の割合により検出する手法, 及び従来手法 [1], [2], を比較した. その結果を表 1 に示す. 結果は, いずれの提案手法も従来手法を上回り, 特に CHLAC 特徴と 2 値特徴の両方を用いた手法が最も良い結果を得た.

次に, 繰返し動作を「位置移動の有無」及び「領域の大小」といった要因に手作業で分類したデータに対して, CHLAC 特徴と 2 値特徴の両方を用いた手法の性能を評価した. 表 2 のデータ数は, 各要因をもつ映像区間の個数であり, これに 7,151 個の繰返し動作でない区間を加え, 実験を行った. 結果を表 2 に示す.

4. 考 察

表 1 より, CHLAC 特徴と 2 値特徴の両方を用いる提案手法が最も良い結果を示した. これは, 動作位置の影響を受けにくい CHLAC 特徴と動作領域の大きさの影響を受けにくい 2 値特徴を組み合わせた効果であり, これによりどちらの繰返し動作に対しても高い検出性能が得られたと考えられる. 一方, 2 値特徴のみを用いた手法においても, 従来手法よりも良い検出結果が得られることが確認された. これは, 提案手法で導入したパワースペクトルの正規化処理により, 従来手法で問題であった, 動作領域の大きさの影響を軽減できたためであると考えられる. また, 動作の大小で分類した際の実験結果 (表 2) からこの正規化処理の効果がうかがえる. 位置の移動の有無に関しては, 大きく移動しない動作の方が F 値で 0.05 程度高いという結果となった. しかし, 位置の移動を伴う場合においても F 値で 0.83 を実現しており, 提案する特徴が有効であったことが確認できる. しかしながら, 一部の位置の移動を伴う繰返し動作に対して, CHLAC 特徴から計算される周期性が高い値を示す傾向が見られた. 今後, 動作領域の移動距離に基づき, 提案手法の各特徴の周期性の割合に重み付けを行うことで, より高い精度で検出が可能になると考えられる.

5. む す び

本論文では, 調理映像中の繰返し動作を検出する手法を提案した. 提案手法は, 位置に依存しない CHLAC 特徴及び依存する 2 値特徴の周期性を解析して組み合わせることで, 多様な繰返し調理動作に対応した. 実験では F 値 0.86 が得られ, 提案手法の有効性を確認した. 今後は, 繰返し動作の動きの性質に応じて各特徴の周期性の割合に重み付けを行い, 更に検出精度を向上させる手法の検討を行う.

謝辞 本研究の一部は科学研究費補助金による.

文 献

- [1] R. Hamada, S. Satoh, S. Sakai, and H. Tanaka, "Detection of important segments in cooking videos," Proc. IEEE Workshop on Content-Based Access of Image and Video Libraries, pp.118-123, Dec. 2001.
- [2] カイ承頼, 高橋友和, 井手一郎, 村瀬 洋, "画像特徴の時間変化に基づく調理映像の分類," 2009 信学総大, A-16-2, March 2009.
- [3] N. Otsu, "Towards flexible and intelligent vision systems — From thresholding to CHLAC," Proc. 9th IAPR Conf. Machine Vision Applications, pp.430-439, May 2005.

(平成 23 年 2 月 28 日受付, 7 月 6 日再受付)