

画像特徴の時間変化に基づく料理映像の分類

Classification of Cooking Video Segments based on Sequences of Image Feature

カイ 承穎¹
KUAI Cheng Ying

高橋 友和²
Tomokazu TAKAHASHI

井手 一郎^{1,3}
Ichiro IDE

村瀬 洋¹
Hiroshi MURASE

名古屋大学¹
Nagoya University

岐阜聖徳学園大学²
Gifu Shotoku Gakuen University

国立情報学研究所³
National Institute of Informatics

1 はじめに

近年、放送映像に対する高度な検索を可能にするための意味構造解析技術が注目されている。とりわけ、料理番組は作業を視覚的に教示するという意味で資料的価値が高く、特に手元ショットは、調理動作などの詳細な解説が行われるため重要である。本講演では、料理映像における手元ショットを「繰り返し動作」(e.g. 切る, 炒める), 「状態提示」(e.g. 料理, 食材), 「その他の動作」(e.g. 揚げる, 焼く) の3つの映像区間に分類する手法を提案する。

関連研究として繰り返し動作の検出手法 [1] があるが、局所領域の画素のみに注目するため、画像全体の動作特徴を反映することが困難である。それに対して提案手法では、画像全体を特徴空間上の点で表し、その時間的変化を解析することにより料理映像区間の分類を行う。

2 提案手法

「繰り返し動作」, 「状態提示」, 「その他の動作」のショット内の連続するフレーム画像を固有空間に投影したものを図1に示す。この図において、連続するフレーム画像は、固有空間上で軌跡をなし、それぞれの調理動作で特徴空間上の軌跡が異なる。本研究では、このような特徴空間上の軌跡の違いに注目し、手元ショット内の各料理映像を「繰り返し動作」, 「状態提示」, 「その他の動作」の3つに分類する。処理の流れを以下に示す。

1. カット検出, ショット分類により, 料理映像を手元ショットと人物ショットに分類する。
2. 手元ショットのうち, W 枚の連続するフレームから固有空間を作成する。ここで, 固有空間上で軌跡をなす各点の第一固有成分のみを用いる。
3. ノイズ低減のため, 軌跡の平滑化を行う。
4. 軌跡のピークの数 m , 最大値と最小値の差 Δr を計算する。 $m > \theta_m$ であれば「繰り返し動作」, そうでなく, かつ $\Delta r < \theta_{\Delta r}$ であれば「状態提示」とする。いずれにも当てはまらない場合は「その他の動作」とする。

以上の処理を1区間とし, W_{step} フレームずつずらしながら手元ショットの映像全体に適用する。

3 評価実験

前節で述べた手法を用いて, 料理映像の分類評価実験を行った。実験には, NHK「きょうの料理」の番組映像25分を用いた。なお, 本実験では, 料理映像のカット検出, ショット分類は人手で行った。また, 予備実験の結

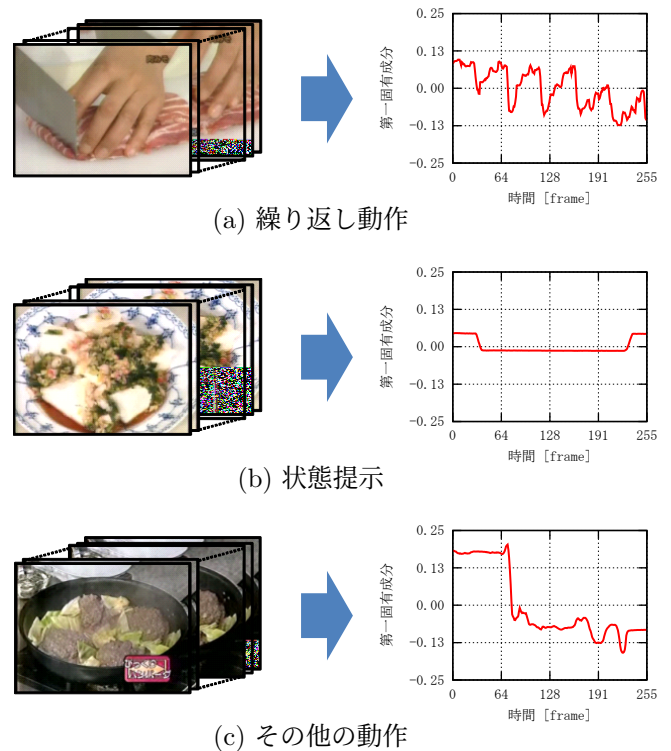


図1 料理映像の固有空間への投影

果を基に, 窓幅 W は256 フレーム (約8秒), 窓の移動ステップ W_{step} は64 フレーム (約2秒), 閾値 θ_m は18, 閾値 $\theta_{\Delta r}$ は0.23とした。

実験の結果, 手元ショット内の映像区間のうち71.4% (115/161) を正しく分類することができた。

典型的な成功例はいずれも高速かつ規則的な「繰り返し動作」であった。一方, 「状態提示」が「その他の動作」と誤分類される例が多く見られた。これは大きなカメラワークなどが原因と考えられる。

4 むすび

本講演では, 画像特徴の時間変化に基づく料理映像の分類手法を提案した。評価実験により, 提案手法の有効性を確認した。今後の課題としては, 料理映像に付随したクローズドキャプションを利用して, より具体的な調理動作 (e.g. 切る, 炒める) の分類を検討する。

参考文献

- [1] 三浦ら, “動きに基づく料理映像の自動要約”, 情処学 CVIM 研究会論, Vol.44, No.SIG9, pp.21–29, 2003