

料理映像における調理動作の解析

蒯 承穎[†] 志土地 由香[†] 高橋 友和[‡] 井手 一郎^{†*} 村瀬 洋[†][†]名古屋大学大学院情報科学研究科 〒464-8601 愛知県名古屋市千種区不老町 1[‡]岐阜聖徳学園大学経済情報学部 〒500-8288 岐阜県岐阜市中鶉 1-38

*国立情報学研究所 〒101-8430 東京都千代区一ツ橋 2-1-2

E-mail: [†] {kuai, shidochi, ttakahashi, ide, murase}@murase.m.is.nagoya-u.ac.jp

あらまし 近年、料理教材を電子化するなどの料理に関する新しい支援サービスへの期待が高まっている。我々は、調理者を支援する際に調理動作を視覚的に提示することを考慮し、調理動作映像データベースの構築を目指している。そのため、特に料理番組の映像を対象とし、特定の調理動作の映像を抽出することを考えた。料理映像中に現れる各種調理動作の中には、「切る・炒める・混ぜる」などの基本的な動作を繰り返すという特徴を持つものが存在する。本発表では、このような特徴に注目し、繰り返し動作の検出、ならびに動作特徴の解析により、これら3つの動作の分類方法を検討する。

キーワード 料理映像, 調理動作, 動作検出, 動作特徴

Analysis of Cooking Motions in Cooking Video

KUAI Cheng Ying[†] Yuka SHIDOCHI[†] Tomokazu TAKAHASHI[‡] Ichiro IDE^{†*}and Hiroshi MURASE[†][†] Graduate School of Information Science, Nagoya University

1 Furo-cho, Chikusa-ku, Nagoya-shi, Aichi, 464-8601 Japan

[‡] Faculty of Economics and Information, Gifu Shotoku Gakuen University

1-38 Nakauzura, Gifu-shi, Gifu, 500-8288 Japan

* National Institute of Informatics 2-1-2 Hitotsubashi, Chiyoda-ku, Tokyo, 101-8430 Japan

E-mail: [†] {kuai, shidochi, ttakahashi, ide, murase}@murase.m.is.nagoya-u.ac.jp

Abstract Recently, the demand of new cooking support service, such as digital cook textbooks, is rising. We consider to provide people with visual explanations for cooking support services, and therefore aim at building a video database of cooking motions by extracting cooking operations from cooking shows. In a cooking video, there are hundreds of cooking operations, most of which are repetition of basic motions, such as cutting, frying and mixing. In this presentation, we try to detect such repetitious motions, and also identify them by their features.

Keyword Cooking Video, Cooking Motion, Motion Detection, Motion Feature

1. はじめに

料理とは、食品や食材、調味料などを組み合わせて加工を行うこと、およびそれを行った結果の生成物の総称と定義できる。すなわち、料理は日常生活において大きな役割を果たすと同時に、一種の創作活動と考えることもできる。近年、マルチメディア技術を応用して、料理教材を電子化するなどの新しい料理に関する支援サービスへの期待が高まっている [8]。たとえ

ば、任天堂（株）が携帯ゲーム機向けに開発した「しゃべる！DS お料理ナビ」 [3] というソフトウェアには、調理者に対する画像と音声ガイダンスによる料理支援機能が搭載されている。本研究では、調理動作を視覚的に提示することによる料理支援を目的とし、料理番組映像（以降「料理映像」と呼ぶ）を対象として、特定の調理動作と対応する映像の抽出を行う。これにより、調理動作映像データベースの自動構築が可能となる。

本研究の先行研究として、テキスト教材が付随する料理番組に注目し、料理映像とテキスト教材の対応付けを行う手法 [4] が提案されている。しかしこのシステムでは、複数の調理動作を含む一つのシーンを対応付けの単位とするため、動作と動作の間に冗長な部分が含まれることと、必ずしも単一調理動作と対応する映像を抽出できない問題がある。一方、料理番組の映像の代わりに、自分で撮影した調理映像とレシピテキストを対応付けることによって調理コンテンツを自動作成する研究 [5] もある。しかし、この場合には、映像だけでなく、様々なセンサなどを組み合わせた専用環境で撮影する必要がある。

本研究では、料理映像中の調理動作に着目し、その中でも特に繰り返し動作特徴を解析することで、調理動作と料理映像の対応付けを行うことを目指す。

2. 料理映像の特徴

料理映像は一般的な映像とは異なる特徴を持つ。図 1 に示すように、料理映像は基本的に、人物ショットと手元ショットから構成される。人物ショットとは、人の全身または上半身が映っているショットであり、講師や補助者が料理についての説明やコツ等を説明する。これに対して、手元ショットでは料理の状態や調理動作が大きく映される場合が多いため、料理映像として重要な視覚情報が多く含まれると考えられる。

手元ショットはさらに、状態部分と調理動作部分に分けられる。状態部分の映像はほぼ静止した画面で構成されており、食材や料理の状態を示す。一方、調理動作部分の映像には一連の調理動作の様子が映っているが、多くの場合、動作の前後は比較的冗長である。

本研究では、手元ショット中の重要な調理動作部分に注目し、さらに、その中の繰り返し動作に対する映像解析の手法について検討する。



図 1: 料理映像の構成

3. 調理動作の分類

調理動作は数百種類にも及び、様々な分類方法が存在する。本研究では、Web 上で公開されている料理レシピ [1] 6,779 件を用いて、調理動作を分類した。調理動作を分類するためには、料理レシピから調理動作を抽出しなければならない。抽出するために、料理レシピの「作り方」を形態素解析する。形態素解析には日本語形態素解析ソフトウェア MeCab [2] を用いた。その結果、調理動作はのべ 67,532 件であり、語尾に基づいて分類したところ 226 種類に分類された。これらの調理動作を人手で大きく「混合」、「加熱」、「切碎」、「装飾」、「浸漬」、「冷却」、「分離・ろ過」、「その他」の 8 つの調理操作に分類したところ、各々の出現頻度は表 1 のようになった。

表 1: 料理レシピテキストから抽出された調理動作の分類

調理操作	出現頻度	調理動作例
混合	25.7%	加える, 混ぜる
加熱	17.7%	炒める, 焼く
切碎	9.6%	切る, 砕く
装飾	9.6%	盛る, 添える
浸漬	2.9%	浸す, 漬ける
冷却	2.3%	冷ます, 冷やす
分離・ろ過	1.4%	振るう, 絞る
その他	31.7%	包む, 溶かす

この結果、「その他」を除けば、「切碎」、「加熱」、「混合」が大きい割合を占めることが分かる。

本研究では映像情報を用いて料理映像を以下の 3 つの調理動作に分類して認識することを考えている。

1) 繰り返し動作

映像の局所領域上を対象物が往復するような周期的な動きが含まれる動作である。「千切りする」などの切り方の違いを表すものを全て「切る」とみなし、「和える」や「泡立てる」などは「混ぜる」とみなす。

2) 状態動作

状態動作の映像においては、画面上に大きな動きがなく、ほとんど静止した状態が続く。状態動作は開始時と終了時のみ動きが存在する。たとえば「茹でる」、「焼く」などである。

3) その他

繰り返し動作と状態動作以外の単一操作で完結する動作である。たとえば「絞る」、「盛る」などである。

本発表では、上の 3 つの分類の中で繰り返し動作に含まれる調理動作「切る」「炒める」「混ぜる」に注目し、次節においてこれらの映像特徴を解析する。

4. 繰り返し動作特徴の解析

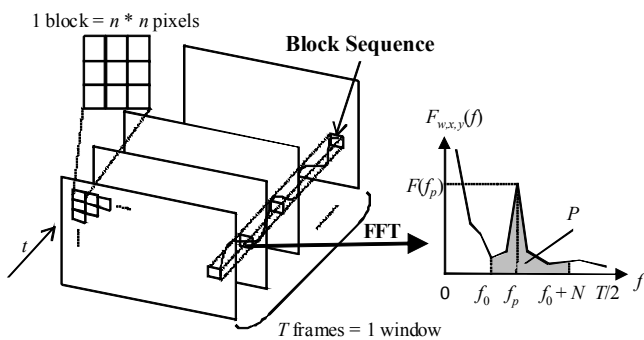
料理映像から繰り返し動作を検出するため、まず映像に対して、カット検出、ショット分類 [6] などの前処理を行う。次に、分類されたショットから手元ショットのみを選び、その中の調理動作部分を検出する。検出された調理動作に対して、フレーム画像中の各局所領域の輝度値の時間変化を周波数解析し、その周期性の有無から繰り返し動作を検出する。以下にその手順を説明する。

まず、図 2 (a) に示すように、各フレームを局所領域に分割する。ここでは、 16×16 ピクセルを一つの局所領域とする。次に、複数の連続するフレームにおける各局所領域の時間変化を観測するために、画像中の各局所領域の輝度値の時間変化に FFT を適用し、その周期性を調べる。

ここでは、図 2 (b) の FFT グラフに関する以下の 4 つの統計量から、ある周波数で明確なピークが存在するか調べる。

- ・対象周波数帯におけるパワーの総和： P
- ・対象周波数帯におけるパワーの最大値： f_p
- ・全体における f_p でのパワーの割合： F_{peak}
- ・ f_p におけるパワーのピークの鋭さ： R_{sharp}

そして、2 個以上の局所領域において、これらが全て閾値以上の値となるようなフレームにおいて、繰り返し動作を検出する。その際に、一般に想定される人間の繰り返し動作の速さから、考慮する周波数帯を $f_0 \leq f < f_0 + N$ と限定する (図 2 (b))。以上の特徴抽出手法は基本的に [7] による。また、フレームの上下左右の端は画像的に不安定であるため、映像の端から局所領域 2 つ分は考慮にしないことにする。



(a) 映像の分割 (b) FFT のグラフ

図 2: 映像の分割方法と局所領域への FFT の適用

繰り返し動作として検出された料理映像中の複数の連続するフレームに対して、図 3 に示すように、フレーム中の各局所領域位置に関して繰り返し局所領

域の出現回数を積算する。繰り返し局所領域と判定された回数が少ないほど白で、多いほど赤で表現されている。

次に、累積された繰り返し局所領域の位置に関する主成分分析を行い、局所領域分布の形の違いから、「切る」、「炒める」と「混ぜる」3 つの調理動作を分類することを検討する。実際に分類する際には、主成分分析の結果得られる第 1 固有値と第 2 固有値の比を動作の特徴とする。

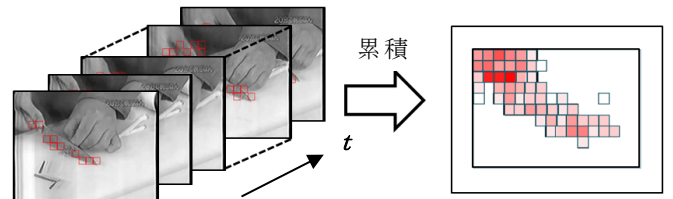


図 3: 局所領域の累積分布

5. 解析実験

5.1. 実験条件

カット検出、ショット分類などの前処理は人手により行い、繰り返し動作部分単独での解析を行う。

本実験では、NHK 番組「きょうの料理」から、手作業で「切る」30 区間、「炒める」27 区間、「混ぜる」30 区間を抽出して用いた。

前節で述べた手法に従い、はじめに繰り返し動作映像の各フレームを局所領域に分割した。各フレームから抽出された繰り返し局所領域の累積分布を主成分分析した。ここで、FFT を行う窓サイズ T は 32 フレーム (約 1 秒)、また窓の移動ステップは 16 フレーム (約 0.5 秒)。対象とする周波数帯は $f_0 = 3$ 、 $N = 12$ とした。本実験で周波数解析に用いた閾値を表 2 に示す。

表 2: 周波数解析に用いた閾値

P	500
f_p	3
F_{peak}	10
R_{sharp}	3

5.2. 実験結果

抽出された繰り返し局所領域の累積分布の例を図 4 に示す。各累積分布に対して主成分分析を行い、縦軸を第 1 固有値 (第 1 主成分 λ_1)、横軸を第 2 固有値 (第 2 主成分 λ_2) とする平面にプロットした結果を図 5 に示す。

「切る」の繰り返し局所領域は細長く分布していることがわかる。これは、動きの大きな局所領域が包丁

の周り，包丁の軌跡に沿って出現したためと考えられる．一方，「炒める」や「混ぜる」は，画面全体に繰り返し局所領域が分布していることがわかる．これに関して，原点を通る直線を分類境界とし，その傾きを少しずつ変化させたときに「切る」と他の動作とを最も良く分類する直線を調べたところ，本実験結果に対する最適な分類境界は傾き 1.9 の原点を通る直線であった．また，このとき分類成功率は，81.6% (71 / 87) であった．この結果から，「切る」と他の動作は，本手法で良好に判別可能であることが分かった．

また，「炒める」と「混ぜる」の判別はこの手法のみでは困難であることが分かった．その要因として，「炒める」と「混ぜる」の映像特徴は，画面全体に動きが存在するという点で類似していることが挙げられる．

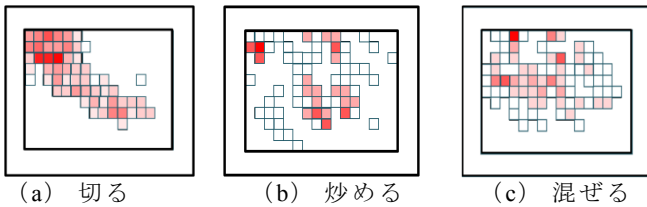


図 4: 局所領域の累積分布の例

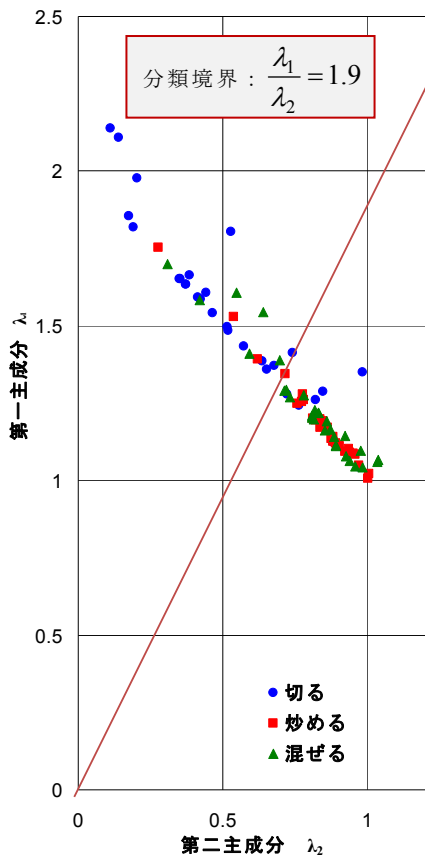


図 5: 各累積分布の形状特徴

6. おわりに

本発表では，料理映像を対象として，その中の繰り返し動作部分を解析により，「切る」などの3つの調理動作を分類する手法を検討した．本手法により「切る」に関しては良好に分類できることを実験によって確認した．

今後は，更に大規模な解析を行い，「炒める」と「混ぜる」の判別方法を検討する．ここで，繰り返し局所領域の累積分布だけではなく，局所領域の時系列分布も考慮する．

また，繰り返し動作以外の動作，特に「焼く」，「茹でる」などの状態動作に関して，動作の間はほぼ静止の料理の状態（焼いてる，茹でている）を示す部分であるが，実際に全てを一つの調理動作映像としたい．

謝辞

本研究の一部は国立情報学研究所の「評価用映像メディア DB」[9] を利用しました．

文 献

- [1] 味の素株式会社，“「味の素 KK」レシピ大百科”，<http://www.ajinomoto.co.jp/recipe/>
- [2] 京都大学，“日本語形態素解析システム和布蕪”，<http://mecab.sourceforge.net/>
- [3] Nintendo，“しゃべる！DS お料理ナビ”，<http://www.nintendo.co.jp/ds/a4vj/>
- [4] R. Hamada, K. Miura, I. Ide, S. Satoh, S. Sakai, H. Tanaka: “Multimedia Integration for Cooking Video Indexing”, *Proc. PCM2004, 5th Pacific Rim Conf. on Multimedia, Lecture Notes in Computer Science*, Vol.3332, pp.657–664 (Dec. 2004)
- [5] 山肩洋子, 角所考, 美濃導彦: “調理コンテンツの自動作成のためのレシピテキストと調理観測映像の対応付け”, *電子情報通信学会論文誌 (D)*, Vol. J90-D, No.10, pp.2817–2829 (Nov. 2007)
- [6] K. Miura, R. Hamada, I. Ide, S. Sakai, and H. Tanaka: “Motion Based Automatic Abstraction of Cooking Videos”, *Proc. ACM Multimedia 2002 Workshop on Multimedia Information Retrieval*, pp.29–32 (Dec. 2002)
- [7] R. Hamada, S. Satoh, S. Sakai, and H. Tanaka: “Detection of Important Segments in Cooking Videos”, *In Proc. IEEE Workshop on CBAIVL 2001*, pp.118–123 (Dec. 2001)
- [8] 椎尾一郎, 浜田玲子, 美馬のゆり: “Kitchen of the Future: コンピュータ強化キッチンとその応用”, *コンピュータソフトウェア*, Vol.23, No.4, pp.36–46 (Dec. 2006)
- [9] 馬場口登, 栄藤稔, 佐藤真一, 安達淳, 阿久津明人, 有木康雄, 越後富夫, 柴田正啓, 全炳東, 中村裕一, 美濃導彦, 松山隆司: “映像処理評価用映像データベースについて”, *電子情報通信学会技術研究報告, PRMU2002–30 (June 2002)*