

## 調理作業手順の効率的提示方法

市場彩花<sup>†</sup> 齋藤 隆文<sup>‡</sup>東京農工大学 工学部情報工学科<sup>†</sup>東京農工大学大学院 生物システム応用科学府<sup>‡</sup>

## 1. 背景と目的

調理手順の説明が必要とされる場合には、一般的に文章と静止画または、動画と音声の組み合わせが用いられる。静止画を用いた場合には、一見ただけで内容がわかり全体像を把握しやすいが、動きの具体的な説明が難しい場合がある。一方、動画を用いた場合には、動きは理解しやすいが、すべての手順を閲覧するには時間がかかり、見たい部分の検索も容易ではない。

そこで本研究では、両者の長所を組み合わせ、文章と短時間の動画を組み合わせることで調理手順を提示することを目的とする。具体的には、調理映像をもとに、動きの抽出、繰り返し動作の抽出などを適用し、作業手順ごとに短時間動画の作成を行い、既存の料理レシピより効率的な方法で提示する方法を考える。

## 2. 関連研究

カイらは、料理映像中の調理動作の解析を行った。Web上で公開されている料理レシピの調理動作を分類した結果、「混合」が動作の約4分の1を占めているという結果が得られた。この結果から、「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」繰り返し動作を重要な動作とし、解析を行った[1]。Domanらは、料理レシピのマルチメディア化に利用する調理動作映像データベースを自動構築することを目的に、料理レシピテキストと料理番組映像との対応付け手法を提案した。その中で、料理番組映像中の動作を分類するために、繰り返し動作・状態提示・その他の3つに動作を分類し解析を行った。[2]

## 3. 提示方法

本研究では、料理を対象としているが、“料理全般”では調理動作の種類が非常に多いため、現段階では、初心者には特に分かりづらい表現

の多い“洋菓子”に範囲を限定している。提示に用いるのは、実際に調理している様子を撮影した動画をもとに作成した GIF アニメーションである。関連研究より、調理動作において特に重要と考えられる、“繰り返し動作”を抽出・提示する。

繰り返し動作を抽出した際に、そのつなぎ目が不自然であると、見づらいものになってしまう。そこで、より自然なつなぎ目で動画を切り出せるように、2枚の画像の差をとる方法、画像の色相と彩度によるヒストグラムを比較する方法、の2つの抽出方法による抽出を試みた。

## 3.1. 差分法

差分法は、もともとなる画像を $I_{back}(x, y)$ 、比較したい画像を $I_{cmp}(x, y)$ 、差分画像を $I_{diff}(x, y)$ とすると、

$$I_{diff}(x, y) = |I_{cmp}(x, y) - I_{back}(x, y)| \quad \dots (1)$$

という式で表すことが出来る。 $I_{diff}$ を閾値により二値化することで、どこが変化したのかを把握することが出来る。図1の画像2枚に対する差分を二値化画像で示したものが図2である。

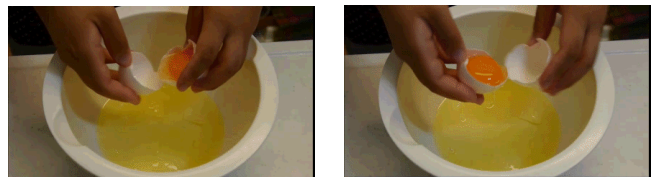


図1 差分をとる画像



図2 差分画像

差分により閾値以下の画素を数え、その数が大きいほど2枚の画像が類似しているとする。

Efficient presentation method of cooking process

<sup>†</sup>Department of Computer and Information Sciences, Tokyo University of Agriculture and Technology.<sup>‡</sup>Graduate School of Bio-Applications and Systems Engineering, Tokyo University of Agriculture and Technology.

### 3.2. ヒストグラムを比較する方法

本研究内でのヒストグラムは、色相と彩度を用いた2次元ヒストグラムとしている。2画像のヒストグラムを比較することで似ているかを比べる。比較には、Bhattacharyya 距離を用いる。画像に対してヒストグラムを作成すると、図3のようになる。ヒストグラムの横軸は色相、縦軸は彩度を示している。

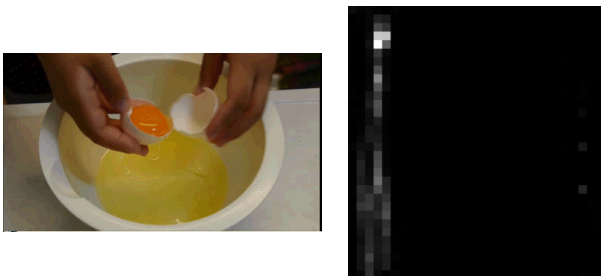


図3 画像のヒストグラム

2つの方法を用いて、1フレーム目とそれ以降のフレームを1枚ずつ比較していく。その際、1フレーム目とほとんど変化のない直近の1秒分(29フレーム)は比較対象として用いない。一番類似していた画像を抽出する最終フレームとし、その部分のみ GIF アニメーションに変換する。

### 3.3. 抽出結果の比較

繰り返し動作として、かき混ぜる、切るように混ぜるといった動作に対して抽出を行った。用いた動画は 30fps である。

1フレーム目と抽出した最終フレームを示す。

- ・かき混ぜる



図4 差分法による抽出結果



図5 ヒストグラムによる抽出結果

元とした動画は 253 フレームであったのに対し、差分法では 33 フレーム、ヒストグラムによる方

法では 32 フレームが抽出され、ほとんど差異のないものであった。用いた動画中の動作は、動作の1回転が速いため、どちらの手法でも、再生するとつなぎ目がわからないような結果となっていた。

- ・切るように混ぜる

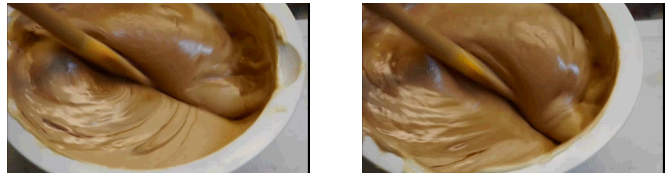


図6 差分法による抽出結果

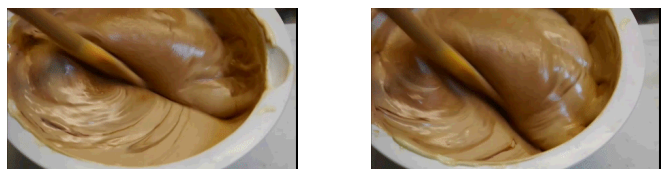


図7 ヒストグラムによる抽出結果

元とした動画は 152 フレームであったのに対し、差分法では 58 フレーム、ヒストグラムでは 34 フレームが抽出された。フレーム数からわかるように、再生すると長さの違いが見られる。この結果の場合、差分法による抽出結果の方がよりなめらかであった。

## 4. 評価項目

本手法の評価のため、抽出結果におけるつなぎ目のなめらかさについての定性的評価を行う。また、3章に示した方法で抽出した結果を文章と共に示すことで、既存の文章と画像で示されている料理レシピとの比較・評価を行うことを考えている。

### 参考文献

- [1]カイ 承穎, 志土地 由香, 高橋 友和, 井手 一郎, 村瀬 洋, “料理番組における調理動作の解析”第4回デジタルコンテンツシンポジウム講演予稿集, No.8-2, 2008.
- [2]Keisuke Doman, Cheng Ying Kuai, Tomokazu Takahashi, Ichiro Ide, Hiroshi Murase : Video Cooking: Towards the Synthesis of Multimedia Cooking Recipes, Proc. 17th Int'l Multimedia Modeling Conference (MMM 2011), Part II, LNCS 6524, pp. 135-145 (2011).