

料理レシピの記述に従った家庭内調理映像の要約

林 泰宏[†] 道満 恵介^{††} 井手 一郎[†] 出口 大輔^{†††} 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

^{††} 中京大学情報理工学部 〒470-0393 愛知県豊田市海津町床立 101

^{†††} 名古屋大学情報連携統括本部 〒464-8601 愛知県名古屋市千種区不老町

E-mail: [†]hayashiy@murase.m.is.nagoya-u.ac.jp, ^{††}kdoman@sist.chukyo-u.ac.jp,

^{†††}{ide,murase}@is.nagoya-u.ac.jp, ^{††††}ddeguchi@nagoya-u.jp

あらまし 近年, Web 上で個人が公開する料理レシピが増えている. しかし, これらの料理レシピは, 文章と数枚の画像のみで構成されており, 具体的な調理動作を把握しにくかったり, 調理手順全体の流れを把握しにくいことがある. そこで, 個人が調理過程を撮影した映像 (家庭内調理映像) を付与することで, 料理レシピを視覚的に理解しやすい形で公開できることが望ましい. しかし, 家庭内調理映像は未編集でカメラワークがない映像であるため, 長時間であり, かつ細かい操作が見にくい. そのため, 家庭内調理映像に含まれる時間的冗長性と空間的冗長性の両方を除去する必要がある. 本報告では, これらのうち, 家庭内調理映像の時間的冗長性の除去を目的とした映像要約手法を提案する. 提案手法では, 対応する料理レシピの記述を利用し, 調理操作の種類に基づいて重要な映像区間を抽出する. また, 被験者実験により要約規則が妥当であることを確認した.

キーワード 映像要約, 調理映像, 調理動作, 料理レシピ

Summarizing a Domestic Cooking Video According to the Descriptions of a Cooking Recipe

Yasuhiro HAYASHI[†], Keisuke DOMAN^{††}, Ichiro IDE[†],

Daisuke DEGUCHI^{†††}, and Hiroshi MURASE[†]

[†] Graduate School of Information Science, Nagoya University Japan

^{††} School of Information Science & Technology, Chukyo University Japan

^{†††} Information and Communications Headquarters, Nagoya University Japan

E-mail: [†]hayashiy@murase.m.is.nagoya-u.ac.jp, ^{††}kdoman@sist.chukyo-u.ac.jp,

^{†††}{ide,murase}@is.nagoya-u.ac.jp, ^{††††}ddeguchi@nagoya-u.jp

Abstract Recently, the number of cooking recipes published by an individual is increasing on the Web. However, since they are mostly composed of only text and a few images, it is sometimes difficult to understand specific cooking operations or the entire cooking process. Therefore, it is preferable to publish a recipe with a video that records the actual cooking process (domestic cooking video). However, since a domestic cooking video is an unedited video with no camera work, it is too lengthy and also difficult to see the details of operations. Therefore, it is necessary to remove both temporal and spatial redundancies from it. Among the two, this paper proposes a video summarization method to remove temporal redundancy in a domestic cooking video. In the proposed method, the description of a corresponding cooking recipe is used to extract important video segments based on the type of cooking operations. The validity of the summarization rules was confirmed by a subject experiment.

Key words Video summarization, cooking video, cooking operation, cooking recipe

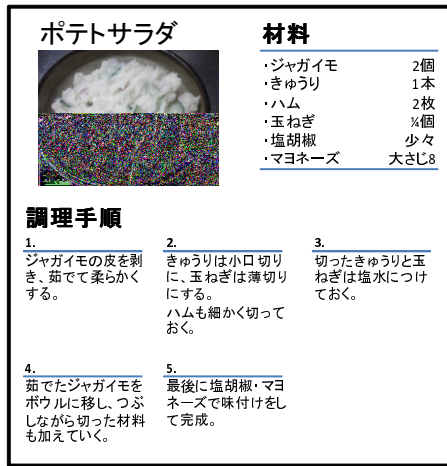


図 1 料理レシピの例

1. はじめに

調理は、我々が生きるうえで重要な活動である。最近では、料理を作るために必要な調理方法を得るために、Web 上の料理レシピを参照することが増えている。例えば、料理レシピ投稿サイトである COOKPAD^(注1)は、個人ユーザが自由に料理レシピを投稿できる Web サービスであり、多数の料理レシピを収録している。

料理レシピの例を図 1 に示す。投稿される料理レシピは、一般に材料と手順の文章と数枚の画像で構成されており、ユーザはこれらの情報を参照しながら調理する。しかし、これらの情報のみでは、具体的な調理方法を理解しにくい場合や、料理レシピの全体像を把握しにくい場合がある。そこで本研究では、料理レシピに映像を付与することで視覚的に理解しやすくすることを考えた。その方法として、(1) 調理手順ごとに対応する調理操作の映像を個別に付与する方法、(2) 料理レシピ全体にその流れを説明する 1 本の映像を付与する方法がある。前者については、既に山肩ら [1] の料理レシピと映像の対応付け手法や道満ら [2] の調理映像データベースを利用したマルチメディア料理レシピに関する研究、浜田ら [3] の調理中にリアルタイムに映像を提示するシステムに関する研究がある。そのため、本研究では後者を実現する方法を検討する。

料理レシピに付与する映像は、料理レシピの各手順を可能な限り短時間で把握でき、調理操作が見やすいことが望まれる。しかし、個人で調理の様子を撮影する場合には、少数の固定カメラで始終記録するような撮影方法が現実的である。そのため、個人撮影の調理映像は、時間的冗長性と空間的冗長性が高い。つまり、調理過程を始終記録する場合、作業の合間の時間や待ち時間などの冗長なシーンが多いため、料理レシピに付与するための調理映像にすべてを含める必要はない。また、少数の固定カメラで調理の状況を撮影するためには広範囲を視野に収める必要があり、個々の調理操作が大きく映らないことが多い。そこで、個人撮影の調理映像を有効利用するためには、(a) 時

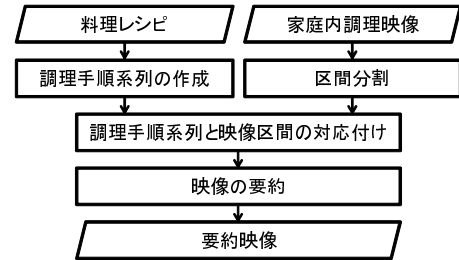


図 2 提案手法における処理の流れ

間的冗長性を除去して調理映像に必要なフレームを抽出し、(b) 抽出されたフレームから空間的冗長性を除去して必要な領域を切り出す必要がある。このうち、空間的冗長を除去する手法 [4] は既に提案されており、有効性が確認されている。

このような背景をうけ、本報告では個人撮影の調理映像（家庭内調理映像）の時間的冗長性の除去を目的とした、映像要約手法を提案する。提案手法では、料理番組映像を対象とした要約手法 [5] のように単純に動きなどの画像特徴を用いるだけでなく、料理レシピを積極的に利用する。

具体的には、まず、動きに基づいて映像を区間に分割・分類し、料理レシピの調理手順と映像区間に対応付ける。この結果を利用することで、各調理手順に基づいて映像を要約する。例えば、「切る」調理操作は、動作の様子を提示する必要があり、「炒める」調理操作は、状態の変化を提示する必要があると考えられる。そのため、提案手法では調理操作の種類によって適切な要約規則を設定することで、各調理操作の様態を把握するために必要なシーンを適切に抽出する。

2. 提案手法

2.1 手法の概要

料理レシピ中の調理手順と家庭内調理映像を対応付け、調理手順中の各調理操作に基づいて映像を要約する。図 2 に提案手法における処理の流れを示す。以下、各処理について説明する。

2.2 料理レシピと家庭内調理映像

2.2.1 料理レシピ

図 1 に示すように、料理レシピは使用する材料が記述された部分と調理手順が記述された部分から構成される。提案手法では、これらのうち、後者を対応付けに用いる。調理手順は、調理の順序に従って大まかに分割されているため、複数の調理操作がまとめて書かれていることが多い。そのため、個々の調理操作と家庭内調理映像を正確に対応付けるためには、調理手順を調理操作に分割する必要がある。提案手法では、文献 [6], [7] を参考に「材料」と「調理操作」の組を構成要素として、調理手順を分割する。これにより、調理手順から調理操作系列を作成する。

2.2.2 家庭内調理映像

家庭内調理映像は、料理レシピに従って個人が調理を行う様子を撮影した映像である。本研究では、「流し」、「調理台」、「コンロ」のような調理場所ごとにカメラを設置し、調理の様子を撮影する状況を想定する。ここで、各カメラは撮影領域が重複

(注1) : <http://cookpad.com>

せず、設置場所は既知とする。以下の処理に先立って、これらのカメラにより撮影された映像から、各時刻で調理が行われた場所の映像を選択してつなぎ合わせることで、調理過程を示す1本の家庭内調理映像を作成しておく。

2.3 動き情報による家庭内調理映像の区間分割

家庭内調理映像の各フレームを (a) 静止, (b) 繰り返し動作, (c) その他の動作の3つに分類し、連続する同じ分類のフレームをまとめて1つの区間とする。提案手法では、まず、静止区間を検出し、静止区間として検出されなかった区間から繰り返し動作区間を検出する。最後にこれらのいずれにも検出されなかった区間をその他の動作区間とする。これにより、家庭内調理映像を区間単位に分割・分類する。なお、この処理の詳細については、文献[8]を参照されたい。

2.4 家庭内調理映像と調理手順の対応付け

家庭内調理映像から得られる映像区間の系列と、料理レシピから得られる調理操作系列を対応付けることで、家庭内調理映像の区間と個々の調理操作を対応付ける。

対応付けには DP マッチングを利用し、特徴量には調理操作の動作分類と調理場所（カメラの設置場所）の情報を用いる。調理操作の動作分類は、区間検出と同じ3種類とする。映像区間系列の要素である映像区間と調理手順系列の要素である調理操作から、それぞれ動作分類と調理場所の情報を求める。映像区間の場合、区間の分類結果をそのまま動作分類とする。また、調理場所は撮影時に分かっているため、その情報を利用する。一方、調理操作は、表1のような14クラスに分類するための辞書を作成しておき^(注2)、各クラスに応じた動作分類、調理場所とする。

以上により、映像区間と調理操作から動作分類と調理場所に関する情報が得られる。これを用いて、以下の式により各情報の一致・不一致に基づいて映像区間 i と調理手順 j の距離 $d(i, j)$ を決定する。

$$d(i, j) = \begin{cases} 0 & \text{調理場所と動作分類が一致する} \\ g(i, j) & \text{動作分類のみが一致しない} \\ \infty & \text{調理場所のみが一致しない} \\ \infty & \text{どちらも一致しない} \end{cases} \quad (1)$$

ここで、 $g(i, j)$ は、映像区間 i と調理手順 j から得られる動作分類が一致しない場合に、以下の式で定義される関数である。

$$g(i, j) = \begin{cases} 10 & \text{一方が「静止」である} \\ 1 & \text{どちらも「静止」でない} \end{cases} \quad (2)$$

式(1)、式(2)の距離計算式を用いて DP マッチングにより対応付けることで、調理手順と映像区間が対応付く。

2.5 家庭内調理映像の要約

2.3 及び 2.4 で説明した処理により、料理レシピ中の各調理操作に対応する映像区間が得られる。次に、各調理操作の種類や性質を考慮して、対応する映像区間を要約する。そのために、各調理操作に対する抽出規則を定めた。本研究では、「調理操作

表1 調理操作のクラス分類

クラス	動作分類	調理場所	例
計量	その他	流し・調理台	計る
洗浄	繰り返し・その他	流し	洗う
浸漬	静止	流し・調理台	漬ける
熟成	静止	調理台・コンロ	冷ます
切碎	繰り返し	調理台	擂る
	その他	調理台	割る
混合	繰り返し	調理台	混ぜる
	その他	調理台・コンロ	加える
濾過	その他	流し・調理台	絞る
伸展	繰り返し	調理台	伸ばす
成形	その他	調理台・コンロ	整える
装飾	その他	調理台・コンロ	添える
加熱	繰り返し	コンロ	炒める
	静止	コンロ	焼く

の動作分類」と「状態の変化」に注目した。

動作分類のうち、「静止」は動きがないため、他の動作分類に比べて優先度が低いと考えられる。しかし、「静止」は、「焼く」のように状態を変化させる調理操作であることが多い。また、その調理操作の継続時間は、調理において重要である。この情報は要約後に字幕で補足することができる。そのため、「静止」の調理操作は、状態の変化が分かるように、開始直後と終了直前の短時間を要約映像に含めるのが効果的と考えた。

「繰り返し動作」には、「切る」のように、動きの様態が重要な動作と、「炒める」のようにその調理操作による状態の変化が重要な動作の2種類があると考えられる。そこで、動きの様態が重要なものは動きが分かるシーンを抽出し、状態の変化が重要なものは変化後の状態が分かるシーンを抽出するのが効果的と考えた。また、類似する動作を繰り返すため、連続して長時間の映像を抽出する必要はないと考えた。

「その他の動作」は、「繰り返し動作」と同様に動きの様態が重要であるか、状態の変化が重要であるかで抽出するシーンを分ける。しかし、「繰り返し動作」のように単調な動作ではないため、短時間で動作の全容を把握するのは難しい。そのため、他の動作分類よりも長時間の映像を要約映像に含めるのが効果的と考えた。

一方、調理動作分類とは別に、「洗浄」の調理操作は、調理手順を理解するうえではあまり重要でないと考えられる。また、「計量」の調理操作は、映像で把握するよりも料理レシピの材料欄の記述を参考にする方が良い。従って、これらの調理操作を行っている映像は、相対的に必要性が低いと考えた。

以上のことを考慮し、表1の各調理操作クラスに対して、以下の規則により対応する映像区間からフレームを抽出する。

(1) 「浸漬」、「熟成」、「加熱（静止）」：映像区間の開始直後と終了直前の3秒を抽出する。

(2) 「切碎（繰り返し）」、「伸展」：映像区間の中間6秒を抽出する。

(3) 「混合（繰り返し）」、「加熱（繰り返し）」：映像区間の終了直前の6秒を抽出する。

(注2)：この分類は食事学における文献[9]、[10]に基づいて定めた。

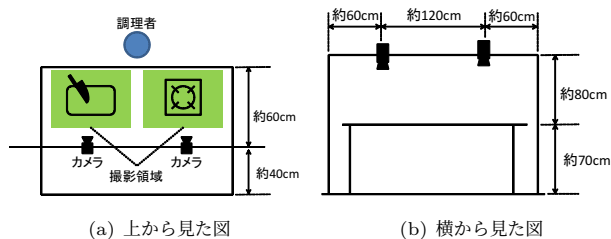


図 3 撮影環境の見取り図

表 2 各手法でのフレーム一致率による対応付け精度

料理名	提案手法	比較手法 1	比較手法 2
ハンバーグ	0.741	0.250	0.417
豚の生姜焼き	0.300	0.057	0.080
ポトフ	0.400	0.391	0.959
ガトーショコラ	0.812	0.110	0.015
ピーマンの肉詰め	0.569	0.001	0.058
ポテトサラダ	0.205	0.203	0.016
平均	0.606	0.301	0.218

(4) 「切碎 (その他)」, 「混合 (その他)」, 「濾過」: 等間隔に 5 秒の区間を 4 区間まで抽出する。

(5) 「成形」, 「装飾」: (4) と同様の区間に加えて, 区間終了直前も抽出する。

(6) 「計量」, 「洗浄」: 要約映像に含めない。

3. 実験と考察

3.1 対応付け手法に関する評価実験

2.4 で述べた家庭内調理映像と調理手順の対応付け手法の精度を評価した。

3.1.1 実験方法

実験では, 図 3 に示すような撮影環境で撮影した家庭内調理映像を使用した。家庭内調理映像は, 「ハンバーグ (10 本)」, 「豚の生姜焼き (3 本)」, 「ポトフ」, 「ガトーショコラ」, 「ピーマンの肉詰め」, 「ポテトサラダ」の 6 種類, 合計 17 本を使用した。実験では, 以下の 2 つの手法と比較した。

- 比較手法 1: 調理場所の情報を利用しない対応付け手法
- 比較手法 2: 動作分類の情報を利用しない対応付け手法

対応付け精度は, あらかじめ人手で各調理操作を映像区間に対応付けたものを正解として, 以下の式により実験結果との一致率を評価した。

$$\text{一致率} = \frac{\text{全調理手順の一致フレーム数}}{\text{全調理手順のフレーム数}} \quad (3)$$

3.1.2 実験結果

提案手法と比較手法の対応付け結果を表 2 に示す。表中, ハンバーグと豚の生姜焼きは, 全映像の平均精度である。全対応付けの平均精度を見ると, 提案手法が最も高く, 調理場所と動作分類の情報を共に用いることの有効性が確認できた。

3.1.3 考察

ハンバーグの家庭内調理映像については, 提案手法が他の手法よりも高い精度で対応付けできた。実験で使ったハンバーグの料理レシピは, 調理操作数とその種類が多かった。このよ

うなデータに対しては, 様々な手順や含まれる調理操作が多いデータでは, 調理場所と動作分類を両方用いることで, 高精度に対応付けできたと考えられる。

一方, 豚の生姜焼き, ポトフ, ポテトサラダは, 対応付け精度の絶対値が著しく低かった。この原因として, 区間検出時の誤検出が挙げられる。実際の調理では, 「切る」などの繰り返し動作であっても, 一時的に手を休めたり, 「焼く」のような静止の調理操作であっても, 焼き加減の確認を行うために動作が発生する。このような調理手順に記述されていない短い区間は, 対応付けの際に違った区間に対応付けられてしまうことがある。さらに, このような区間が多いと, 全体の対応付けにも大きく影響する。調理手順には関係ない短い区間は長さに基づいて除去することもできるが, 「塩を振る」のように本来短時間の調理操作もあるため, 単純に区間の長さだけで判断することはできない。そのため, 区間検出の段階から, 調理手順を考慮して調理手順と関係ない区間を除去できれば精度の改善が見込める。

3.2 映像要約規則に関する評価実験

2.5 で述べた要約規則の有効性を確認するために, 被験者実験により評価した。

3.2.1 実験条件

ハンバーグの家庭内調理映像について各調理手順に対応する区間から, 提案手法により抽出された区間のほか, 開始直後, 終了直前, 中間をすべて選択候補として 62 個の映像断片候補を用意した。そして, 5 人の被験者にこれらの映像断片候補から, 要約映像を作成する際に必要と思われるものを 16 個～20 個の範囲で選択することで, 要約映像を作成してもらった。

3.2.2 実験結果

図 4 に, 各映像断片候補を選択した被験者の人数を示す。●を付した選択候補は, 提案手法の要約規則により抽出されたものである。過半数以上の被験者が選択した区間は要約映像に必要であると考え, 提案手法により抽出された区間が全て含まれ, 要約規則の有効性が確認できた。

参考までに, 提案手法により要約した映像のフレーム系列を図 5 に示す。

3.2.3 考察

「静止」, 「繰り返し動作」, 「その他の動作」の抽出規則についてそれぞれ考察する。

a) 「静止」の抽出規則に関する考察

被験者は, 「荒熱をとる」, 「味をなじませる」, 「生地を焼く」のような静止区間をあまり選択しなかった。将来, 字幕の挿入により情報を提示できるため, 抽出規則では, 区間の開始直後と終了直前を抽出したが, 本実験では, 字幕を挿入していない映像を用いたため, やむを得ない結果である。今後, 字幕を挿入する必要があることを確認した。

b) 「繰り返し動作」の抽出規則に関する考察

本実験で, 「繰り返し動作」に属する調理操作は, 「長葱を切る」, 「椎茸を切る」, 「材料を炒める」, 「材料を混ぜる」があった。

まず, 前 2 者について考察する。これらの「切碎」に属する調理操作は, 動きを示すために, 抽出規則では, 中間の区間を抽出した。これは, 開始直後と終了直前では, 動作の様子を提

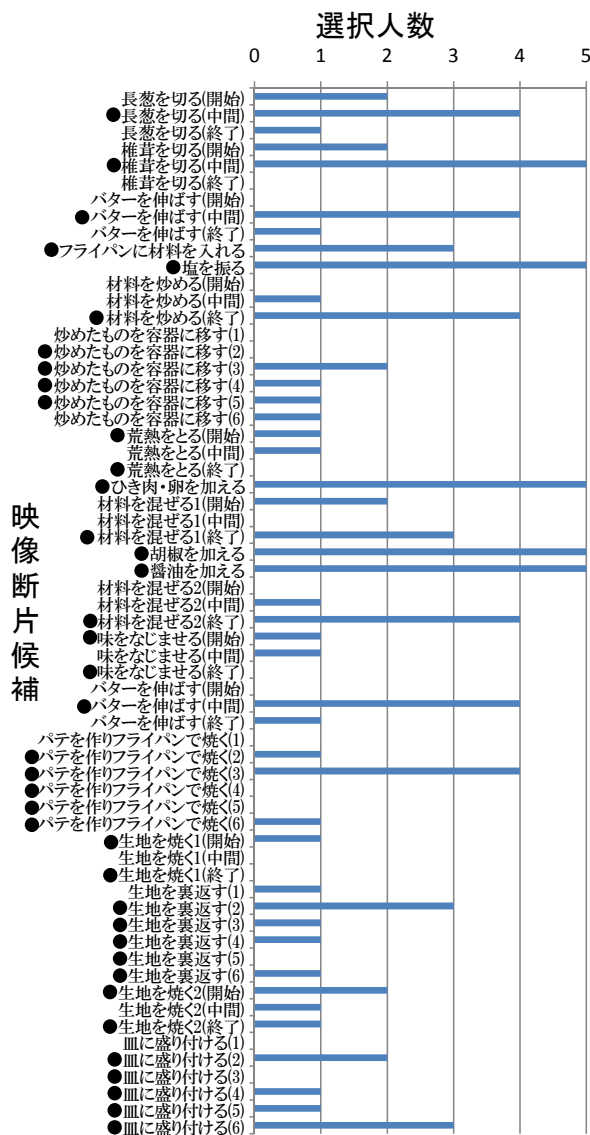


図 4 各映像断片候補を選択した被験者数。●を付したものは提案手法で選択される映像断片を示す。

示するには不十分と考えたためである。全ての実験結果で、過半数以上の被験者が中間の区間を選択した。

次に、後 2 者について考察する。これらの調理操作は、状態の変化を提示するために、抽出規則では、終了直前の区間を抽出した。実験結果では、「材料を炒める」と 2 回出現する「材料を混ぜる」はいずれも、過半数以上の被験者が、終了直前の区間を選択した。

以上から、繰り返し動作に関する抽出規則の有効性を確認できた。

c) 「その他の動作」の選択規則に関する考察

ここでは、複数の選択候補が存在した「炒めたものを容器に移す」、「パテを作り焼く」、「生地を裏返す」、「皿に盛り付ける」について考察する。これらの選択結果を見ると、多くの被験者が特定の区間のみ選択した場合と、分散して選択したものがある。これは、複数の選択候補から、調理操作を把握するのに最

適な区間があったり、最適な区間がなく個人差により選択区間が分散したためと考えられる。「その他の動作」は、「繰り返し動作」のように単調な動作ではないため、調理操作を理解しやすいシーンを適切に抽出することは難しかったと考えられる。そのため、提案手法のように調理操作を含む映像を分割して複数の区間を抽出するのは、無難ではあるが、理想的な要約規則ではなかった。より適切な要約を行うには、映像内容も解析し、調理操作を把握するうえでより適切なシーンを抽出する必要がある。

4. む す び

本報告では、料理レシピを用いた家庭内調理映像の要約手法を提案した。提案手法では、映像区間の系列と調理手順系列を対応付けることで、各調理手順が行われている映像区間を求め、調理手順の内容を考慮した映像要約を可能にした。

評価実験により、提案した対応付け手法と映像要約規則を評価し、それぞれの有効性を調査した。要約規則については、おおむね妥当な規則だったことが確認できた。今後は、映像内容の解析によるシーン抽出を行うことで、より良い要約映像を生成できると考えられる。また、多数の家庭内調理映像を用いて評価する必要がある。

5. 謝 辞

被験者実験に協力していただいた名古屋大学村瀬研究室諸氏に深く感謝致します。また、本研究では画像処理に MIST ライブラリ (<http://mist.m.m.is.nagoya-u.ac.jp/>) を使用した。

文 献

- [1] 山肩洋子, 角所考, 美濃導彦, “調理コンテンツの自動生成のためのレシピテキストと調理観測映像の対応付け,” 信学論 (D), vol.J90-D, no.10, pp.2817–2829, Oct. 2007.
- [2] 道満恵介, カイ承穎, 高橋友和, 井手一郎, 村瀬洋, “調理動作に着目した料理レシピのマルチメディア化の提案,” 情処学全大講演論文集, no.5, pp.189–190, Mar. 2010.
- [3] R. Hamada, J. Okabe, I. Ide, S. Satoh, S. Sakai, and H. Tanaka, “Cooking Navi: Assistant for daily cooking in kitchen,” Proc. 13th ACM Int. Multimedia Conf., pp.371–374, Nov. 2005.
- [4] 兵庫渉, 林泰宏, 野田雅文, 出口大輔, 井手一郎, 村瀬洋, “調理手順に従った撮影対象領域の決定に基づく調理映像を対象としたデジタルカメラワーキング,” 信学技報, MVE2011-100, Mar. 2012.
- [5] 三浦宏一, 浜田玲子, 井手一郎, 坂井修一, 田中英彦, “動きに基づく料理映像の自動要約,” 情処学論 CVIM, vol.44 no.SIG9, pp.21–29, Jul. 2003.
- [6] 浜田玲子, 井手一郎, 坂井修一, 田中英彦, “料理テキスト教材における調理手順の構造化,” 信学論 (D-II), vol.J85-D-II, no.1, pp.79–89, Jan. 2002.
- [7] 刈米志帆乃, 藤井敦, “料理レシピテキストを対象とした構造解析の高精度化,” 信学技報, DE2012-8, Jun. 2012.
- [8] 林泰宏, 道満恵介, 出口大輔, 井手一郎, 村瀬洋, “時間的冗長性の除去による調理履歴映像の要約,” 信学技報, DE2012-10, Jun. 2012.
- [9] 南出隆久, 大谷貴美子, “栄養学シリーズ NEXT 調理学,” 講談社, Sep. 2000.
- [10] 山崎清子, 渋川祥子, 下村道子, 杉山久仁子, 市川朝子, 島田キミエ, “NEW 調理と理論,” 同文書院, Apr. 2011.



図 5 提案手法により要約した映像のフレーム系列。枠で囲まれたフレームは 1 つの調理操作から抽出されたことを示し、左上の番号は適用した要約規則を示す。また、<http://www.youtube.com/watch?v=GAV3rF2yBN8> で映像を公開している。