6N-04

動き検出による料理映像の解析 *

三浦 宏一[†], 浜田 玲子[‡], 井手 一郎[††], 坂井 修一[†], 田中 英彦[†]

[†,‡]{miura,reiko,sakai,tanaka}@mtl.t.u-tokyo.ac.jp, [††]ide@nii.ac.jp

[†] 東京大学大学院情報理工学系研究科 [‡] 東京大学大学院工学系研究科

[††] 国立情報学研究所

1 はじめに

近年の映像技術の進歩にともない、放送やインターネットなどを通じて発信されるマルチメディアデータは増大の一途をたどっている。このような大量のデータを整理し、効率良く保存・検索するため、マルチメディアデータの解析はますます重要な技術となりつつある。そこで我々は、料理映像に着目した映像解析、索引付けなどの研究を行っている [1]。また最近では特に料理映像の自動要約を目標とし、映像中の動きに着目した研究を進めている [2]。料理映像の自動要約が実現すれば、映像による料理レシピの閲覧の他に、要約料理映像データベースの構築や検索など、様々な応用が考えられる。

これまで、映像の自動要約に関しては様々な研究がなされてきたが、一般的に、要約された映像は見づらいという研究結果も報告されている [3]。この原因は音声が断続的に途切れるためであるが、料理映像では視覚的な情報から動作や手順を知ることができるため、音声がなくても概要を理解することができる。また、冗長な映像も多く含むため、料理映像は要約に適した素材であると考えられる。そこで我々は、音声を含まない要約映像を作成することを前提としている。

本稿ではこれまでの研究 [2] をふまえ、映像中から検出された動きを基にした料理映像の解析手法について検討する。

2 動き検出による料理映像の解析

2.1 料理映像の特徴

料理映像には、一般的に対応するテキスト教材が存在することが多い。従って料理映像においては、テキスト教材では表現しきれない視覚的な情報を示す部分が特に重要であると考えられる。すなわち、調理動作や、料理や食材の状態に関する映像は特に重要である。

ここで、料理映像のショットは、図 1に示すように、大きく (A) 人物ショット、(B) 手元ショット の 2 つに分類でき、主に手元ショットに重要な映像を含む。

(A) 人物ショット　　(B) 手元ショット

図 1: 料理映像におけるショットの分類

しかし多くの場合、図 2に示すように、手元ショットはその中にさらに構造があり、調理において重要な映像を含む一方で、動作と動作の間などは比較的冗長である。したがって料理映像の要約を作成する際には、映像の構成を解析し、元の映像から手元ショットにおける重要部分を取り出すことが必要となる。

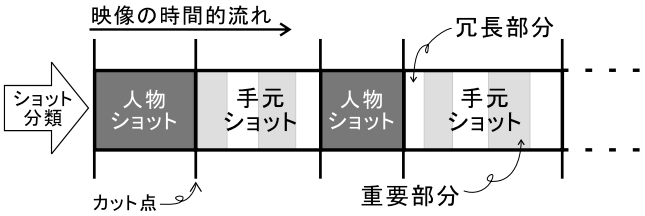

図 2: 料理映像の構成例

また前述のように、料理映像における重要部分は、調理動作や料理や食材の状態に関する映像であるが、これらは表 1のような特徴をもつ。

表 1: 料理映像の重要部分とその特徴

重要部分	特徴
調理動作	映像の動きが大きい (激しい)
料理や食材の状態	映像の動きはほぼない

2.2 映像中の動き検出

表 1の特徴により、料理映像における重要部分を検出するためにまず、映像中の動きを検出する。本研究では映像中から動きを検出する手法として、後に動きの方向や速度などを利用することも考慮し、オプティカルフローを利用する。オプティカルフローを検出する手法は数多く提案されているが、現時点では基本的な手法である Horn らの手法 [4] を用いている。具体的には、30frame/sec の料理映像中の手元ショットに対しオプティカルフローを検出し、その大きさの和を計算することで、図 3に示すようなグラフを作成する。我々のこれまでの研究では、このグラフを基に、90%程度の精度

---

* "Analysis of Cooking Video by Motion Detection"
Koichi Miura[†], Reiko Hamada[‡], Ichiro Ide[††], Shuichi Sakai[†], Hidehiko Tanaka[†]
[†]Graduate School of Information Science and Technology, The University of Tokyo
[‡]Graduate School of Engineering, The University of Tokyo
7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8656, Japan
[††]National Institute of Informatics
2-1-2 Hitotsubashi, Chiyoda-ku, Tokyo 101-8430, Japan

で映像の重要部分検出に成功している [2]。また、新たにショット内でのオプティカルフローの大きさの分散 σ^2 を閾値として導入することによって、検出精度の向上を図った。

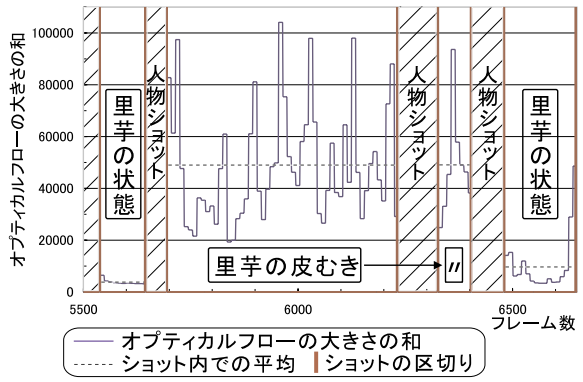

図 3: フレーム毎のオプティカルフローの大きさ

なお、誤検出や検出洩れの主な原因は、

- カメラワーク
- 調理動作とは関係のない人の動き

などの、重要ではない動きを誤検出したことであった。

2.3 カメラワークの検出

誤検出を減らすため、検出されたオプティカルフローを利用してカメラワークを検出する。今回は、カメラワークの中でもその多くが誤検出の原因となる、カメラの平行移動 (パン) の検出手法について検討した。具体的な手順を次に示す。

1. 1 フレーム中の全画素において、それぞれオプティカルフローベクトルの向き (角度) を計算する。ベクトルの大きさで重み付けをし、角度の分布をとる。
2. 一続きの動きと見なせる範囲のフレームについて、角度分布の平均をとる。

以上により、パンがある場合には角度分布は図 4のように正規分布に近い形になり、カメラワークがなく、調理動作のみの場合には図 5のように明確なピークがないことがわかった。したがってこれを利用し、角度分布のピークの値 (頻度)F_p がある適当な閾値 F_{th} 以上であり、かつピークが 1 つのみであるものをパンとして検出することとした。

3 単純な要約映像の作成

ここで、今まで検出してきた重要部分を取り出すことにより、簡単な要約映像を作成した。今回の手法としては、重要部分として検出された部分のうち、調理動作部分に関しては最初の 2 秒間を、料理や食材の状態部分に関しては最後の 2 秒間を取り出し、それらを単純に時系列にそって結合している。また、カメラワーク (パン) に関しても自動検出し (F_{th}=0.025 に設定)、その部分は要約から除外するようにした。なお、ショット分類

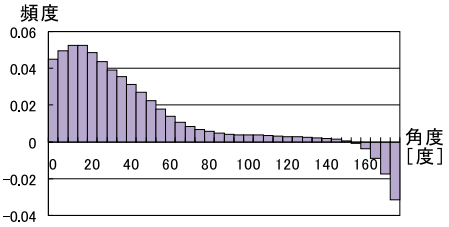

図 4: カメラワーク (パン) がある場合の角度分布

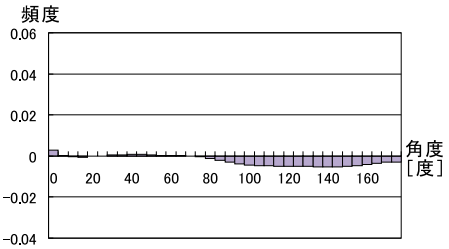

図 5: カメラワークがない場合の角度分布

は理想的に行われたものとし、それ以降の処理に関しては自動化して行った。

このようにして作成された要約映像は時間的には元の映像の 1/8 ∼ 1/10 程度であったが、調理手法の概要はほぼ理解できるものであった。またカメラワーク (パン) を検出したことにより、誤検出の約 4 割を削減できた。

4 おわりに

本研究では、料理映像の要約を目標に、映像中の動きに着目することで、映像構成を解析する手法を検討している。本稿においては、主にこれまでの手法では誤検出の原因となっていたカメラワークの検出手法について検討した。また、簡単な要約映像を作成することにより、手法の有効性を確認した。

今後は、動きの分類をすることなどによる映像解析の精度向上、及びそれを利用した料理映像の要約手法の詳細について検討する。また、動き検出とは別に、映像を要約する上で重要と考えられる映像中の字幕の検出についても検討する。

参考文献

[1] R. Hamada, I. Ide, S. Sakai, and H. Tanaka: “Associating Cooking Video with Related Textbook”, Proc. ACM Multimedia 2000 Workshops, pp.237-241, Nov. 2000.

[2] 三浦宏一, 浜田玲子, 坂井修一, 田中英彦: “料理映像の要約のための動き検出”, 第 63 回情処学全大, No.6L-5, Vol.2, pp.61-62, Sep. 2001.

[3] M. Christel, M. Smith, C. Taylor, and D. Winkler: “Evolving Video Skims into Useful Multimedia Abstractions”, Proc. ACM CHI’98 Conference on Human Factors in Computing Systems, pp.171-178, April 1998.

[4] B. K. P. Horn and B. Schunck: “Determining optical flow”, Artif. Intel., Vol.17, pp.185-203, Aug. 1981.