6L-05

料理映像の要約のための動き検出 *

三浦 宏一†, 浜田 玲子‡, 坂井 修一†, 田中 英彦†

{miura,reiko,sakai,tanaka}@mtl.t.u-tokyo.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

1 はじめに

近年の映像技術の進歩にともない、テレビや WWW 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, 2]。本稿では、特に料理映像の自動要約を目標とし、そのために重要である映像中の動きの検出について検討する。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様々な研究がなされてきた。しかし一般的に、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、音声がなくても理解することができる。一方で、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、本研究では音声を含まない要約映像を作成することを前提とする。

2 料理映像中の動き検出

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 つに分類できる。

人物ショットは、人物を中心にスタジオ全体が映されるか、もしくは人物の上半身がアップに映されるショットであり、調理人やその助手が調理法などに関して説明していることが多い。しかし、調理動作などは映されていないか、部分的に小さく映されているのみであり、映像から調理に関して視覚的な知見を得ることはできない。一方、手元ショットは材料やそれを調理する手元が大きく映され、重要なショットである。しかし多くの場合、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。

(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

料理映像の構成例を図 2に示す。図 2のように、料理映像においては人物ショットと手元ショットがほぼ交互に出現し、重要であると考えられる手元ショットの中には、さらに重要な部分と比較的冗長な部分が含まれる。

したがって、料理映像の要約を作成する際には、このような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。

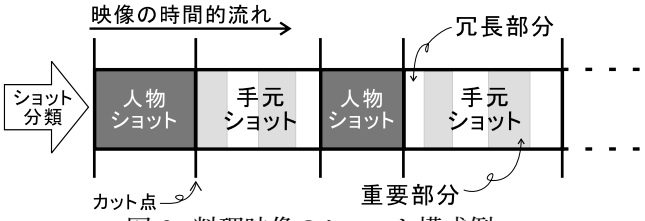

図 2: 料理映像のショット構成例

料理映像における重要部分は、先程も述べたように、調理動作や、料理や食材の状態に関する映像であるが、実際の料理映像を参照したところ、これらは表 1のような性質をもつことがわかった。 これにより、本稿では

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

映像中の動きに着目することで、料理映像における重要部分の検出を検討する。なお、我々は、料理映像において特に重要な調理動作の検出手法として、繰り返し動作に着目した研究も行っている。しかし本稿では、そのよ

---

* "Motion Detection for Cooking Video Abstraction"
Koichi Miura†, Reiko Hamada‡,
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

うな特定の重要動作のスポッティングではなく、映像全体の動きを利用した映像構成の解析を目的とする。

2.2 オプティカルフローの検出

映像中から動きを検出する手法として、本研究では、後に動きの方向や速度などを利用することも考え、オプティカルフローを利用する。これまでに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。しかし今回は映像全体の動きを出すことが目的であり、厳密な動きの解析は必要でないと考えられるため、基本的な手法である Horn らの手法 [4] を用いることとした。

2.3 動き検出による映像構成の解析

オプティカルフローを利用して、映像の要約のための重要部分を検出することを考える。具体的には、30frm/s の料理映像に対し、以下の手法を適用する。

1. カット検出及びショット分類を行い [2]、人物ショットを取り除く。
2. 残りの手元ショット中の 1 枚 1 枚の画像に対し、オプティカルフローの検出を行う。
3. 1 枚の画像 (1frame) 中の全画素 (320×240) において求められたオプティカルフローのベクトルの大きさを計算し、それらの和を取る。
4. 10frames 毎にベクトルの大きさの和の平均をとり、グラフ化する。

そのようにして描いたグラフの 1 部分を図 3に示す。

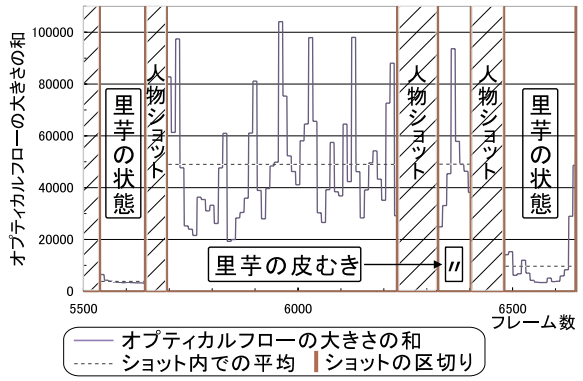

図 3: フレーム毎のオプティカルフローの大きさ

2.4 予備実験

オプティカルフローのグラフを元に、映像中の調理動作および食材の状態を示す部分を検出する予備実験を行った。まず、ある時点でのオプティカルフローの大きさの和を S、またショット内での平均値を S_{ave} とおく。調理動作については、$S_{ave} > S_{move}$ であるショットのうち、S が S_{ave} の α 倍以上の部分を検出する。料理や食材の状態については、$S < S_{state1}$ が F frm 以上続く部分、または、$S_{ave} < S_{state2}$ であるショットのうち $S < S_{state2}$ の部分を検出した。

2 レシピ分 (約 13 分) の料理映像に対して、本手法を用いて重要部分を検出した結果を表 2に示す。なお、今回の実験ではショット分類は理想的に行われたものとし、$S_{move} = S_{state2} = 10000$, $S_{state1} = 7000$, $\alpha = 1.0$, $F = 90$ (3 秒間) とした。

表 2: 検出結果

重要部分	正解	正検出	誤検出	洩れ	再現率	適合率
調理動作	41	40	9	1	98%	82%
状態	18	16	1	2	89%	94%

この結果から、映像中の動きにより有効に、料理映像の構成を解析できることが示された。

なお、誤検出や検出洩れは、

- テレビカメラの動き
- 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が主な原因であった。

2.5 今後の方針

今後は、より精密に映像の構成を解析するために、カメラ移動の検出や、また、映像中の重要な動きと重要でない動きを分類・検出する手法を検討する。これらも、オプティカルフローの大きさや向き、またはその分布を利用することで実現できることが期待される。また、それらの結果を利用した料理映像の要約映像作成システムの構築を検討する。

3 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料理映像の構成を解析する手法を検討した。具体的には、手元ショットからオプティカルフローの検出を行い、その結果を利用して映像の構成を解析する。本稿においては予備実験を通してその可能性を示した。

今後は、動き検出手法の精度向上、及びそれを利用した料理映像の要約について検討する。

参考文献

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

[2] 三浦宏一, 浜田玲子, 井手一郎, 坂井修一, 田中英彦: “料理映像の構造解析による手順との対応づけ”, 第 62 回情処学全大, No.6R-9, Vol.3, pp.31-32, Mar. 2001.

[3] M. Christel, M. Smith, C. Taylor, and D. Winkler: “Evolving Video Skims into Useful Multimedia Abstractions”, Proc. of 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.