料理映像における繰り返し動作検出とその応用

6L−6

浜田 玲子[†], 佐藤 真一[‡], 坂井 修一[††], 田中英彦[††]
{reiko,sakai,tanaka}@mtl.t.u-tokyo.ac.jp,satoh@nii.ac.jp
[†] 東京大学大学院 工学系研究科, [‡] 国立情報学研究所, [††] 東京大学大学院 情報理工学系研究科 *

1 はじめに

映像技術の進歩に伴い、テレビやWWWなどを通じて様々な映像が発信され、大量に蓄積されつつある。そこで近年は、映像の索引付けや検索に関する研究が盛んに進められている。我々は、料理映像に着目した映像解析および索引付けの研究を行なっている[2]。

料理映像には、多くの場合付随するテキスト教材が存在するが、映像にはテキストでは表現しきれない様々な情報を含んでおり、特に料理手順の理解のためには視覚情報が非常に有効である。料理映像に索引付けを行なうことにより、映像の要約や検索など、様々な実用的なアプリケーションへの応用が可能となる。今後は、家庭内への計算機の進出に伴い、このような索引付けされた料理映像に対する需要は高まっていくものと考えられる。

本稿では、料理映像への索引付けの手がかりとして、繰り返し動作に着目した重要部分抽出について検討する。

2 料理映像の特徴

料理映像におけるショットは大きく人物ショットと手元ショットに分類され、これらがほぼ交互に出現する。人物ショットは台所のほぼ全体が映されるが、手元は小さく映るのみであり、調理に関する視覚的な情報は少ない。一方、手元ショットでは材料を調理する手元が大映しにされ、視覚的にも重要である。しかし、手元ショットの中にはさらに調理の中心となる重要な映像と、動作と動作の間などの比較的冗長な映像が含まれる。このような構造の料理映像において、一般的な手法[1]のように映像をショットに分割してそれぞれに索引をつけるだけでは、手元ショットに含まれる重要な映像を抽出することはできない。そこで、本稿では重要な部分を解析して直接抜き出すことを考える。

料理映像において特に重要なのは、調理動作部分である。そこで、実際の料理映像を参照して検討した結果、調理の中心となる動作の多くは繰り返しの動作であるということがわかった。具体的には、「切る」「あえる」「泡立てる」など、様々な対応する動詞がある。図1に繰り返し動作の例をいくつか示す。本研究では、このような動作の時間方向の周期性に着目した重要部分検出を行なう。

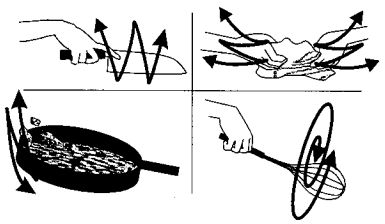

図 1: 調理中の繰り返し動作の例

3 繰り返し動作の検出

一般的には、色情報により手領域を認識・追跡して動作の識別などを行なうが[3]、料理映像においては道具の形状・色などが一定ではないため、これは困難である。そこで本研究では、時間周波数解析によって微少領域の輝度値の時間変化を解析し、その周期性の有無から繰り返し動作の検出を行なう。まず、各フレームを小さなブロックに分割する。図2に示す通り、各ブロックは3 × 3ピクセルから成る小さな正方形である。各ブロックに含まれるピクセルの平均輝度値を$V_{x,y}(t)$とする。なお、x,yは画像におけるブロックの空間座標、またtはそのブロックが属するフレームの時間座標である。

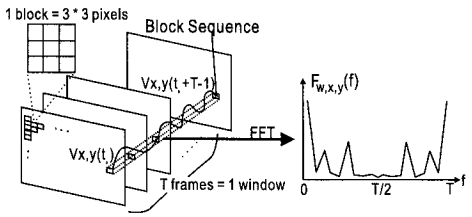

図 2: 映像の分割方法とFFTの適用

x,yを一定の位置に固定してtを動かすと、$V_{x,y}(t)$は位置x,yにおける輝度値の時間変化を示す。そこで、画像中のすべてのx,yにおけるブロック列$V_{x,y}(t)$にそれぞれFFTを適用し、その周期性を調べる。FFTを適用

---

* "Detection of Repetitious Motions in Cooking Video and its Applications"
Reiko Hamada[†], Shin'ichi Satoh[‡], Shuichi Sakai[††], Hidehiko Tanaka[††],
[†]Graduate School of Engineering, The University of Tokyo,
[‡]National Institute of Informatics,
[††]Information Science and Technology, The University of Tokyo

する時間方向の範囲は、大きさ（フレーム数）$T = 2^n$ の時間窓内とする。この窓を t 方向に T_{step} フレーム単位で移動していくことにより、各時点での振動の有無を調べる。ここで、$V_{x,y}(t)$ に明確な周期性がある場合、結果の FFT グラフにはある周波数で明確なピークができる。これを検出するため、FFT のパワーおよびピークの明確さを表わす、以下に示す 6 つのパラメータを利用する。

- $Power$: パワーの総和
- f_p : $F(f)$ が最大となる周波数
- $F_{peak}(f_p)$: $F(f_p)$ と $F(f \neq f_p)$ の平均値との比
- R_1 : $F(f_p)$ と $F(1)$ の比
- R_2 : $F(f_p)$ と $F(2)$ の比
- R_{sharp} : $F(f_p)$ と $F(f_p$の周辺) の比

2 点以上のブロックにおいて、6 パラメータがいずれも閾値以上の値を持つとき、その時の時間窓において繰り返し動作の検出を行なう。なお、人間の繰り返し動作の早さから、各パラメータにおいて考慮する周波数帯を $f_0 \leq f < f_0 + N$ の範囲に限定した。

4 評価実験

前章で述べた方法に従い、評価実験を行なった。実験には、1 つの料理番組から取得した料理映像 16 レシピ分（合計約 69 分）を利用した。映像は圧縮方式が mpeg2、画像サイズは 360×240 ピクセル、フレームレートは 15frm/s である。解析の際には、窓の大きさ $T = 32frm$、窓の移動の大きさ $T_{step} = 16frm$、またパラメータを計算する際の周波数の範囲を $f_0 = 3, N = 12$ とした。なお、閾値については実験に際して手動で適切に設定した。

実験結果の評価においては、対象の料理映像から、振動動作部分の映像を人手で抜き出し、これを正解とした。そして自動解析結果と人手による正解を照合することで、手法の評価を行なった。その結果を表 1に示す。なお、人手による結果を Ans_H、自動解析による結果を Ans_M、両者が一致した答を Ans_C とすると、再現率は Ans_C/Ans_H、適合率は Ans_C/Ans_M である。

表 1: 実験結果

Ans_H	Ans_M	Ans_C	再現率	適合率
33	26	24	72.7%	92.3%

表 1に示す通り、本手法では誤検出が少なく、90% 以上の適合率で繰り返し動作を検出することがわかった。一方、再現率は 73% 程度である。これは、ゆっくりとした繰り返し動作は周期性があいまいなことが多く、検出漏れが多かったためである。

5 料理映像の要約

本手法の実用性を確認するため、簡単な料理映像の要約アプリケーションを作成した。図 3に、実装されたアプリケーションの例を示す。

この例においては、要約から人物ショットを除くためにまずカット検出を行ない、肌色を利用して人物ショットと手元ショットに自動分類する。そして、各手元ショットにおいて、本手法によって振動が検出された部分、およびショットの最初と最後を 2 秒ずつ拾って要約とする。手元ショットに振動のない場合は、ショットの真ん中をとる。

このように要約された料理映像によって、レシピのおおまかな手順やかかる手間などがほぼ理解できることがわかった。将来は、圧縮された料理映像をデータベース化することにより、映像による直感的なレシピの検索や選択などに利用することができる。

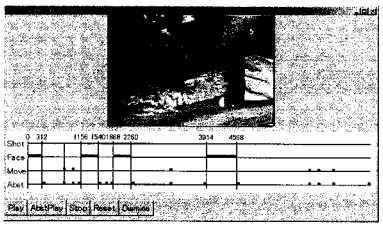

図 3: 料理映像要約アプリケーションの例

6 まとめ

我々は、料理映像の索引付けについて検討している。本稿では、料理映像における代表的な重要動作として繰り返し動作に着目し、その自動検出手法を提案、評価実験を通して本手法の有効性を示した。また本手法の応用として料理映像の要約アプリケーションを実装し、紹介した。今後の課題として、繰り返し動作の検出精度の向上、および本手法を利用した応用の検討などがあげられる。

参考文献

[1] Y. Ariki, "Multimedia Technologies for Structuring and Retrieval of TV News," News Generation Computing, Vol.18, No.4, pp.341-358, 2000.

[2] 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.

[3] C. Wren, A. Azarbayejani, T. Darrell, and A. Pentland, "Pfinder: Real-Time Tracking of the Human Body," IEEE Trans. PAMI, Vol. 18, No. 7, pp. 780-785, July 1997.