

食品消費モニタリングのための切断音を用いた食品認識

小島 嵩道 井尻 敬 片岡 秀公 平林 晃

概要. 食品消費のモニタリングは、食糧廃棄の削減や生活習慣病の予防など、食に関する社会的問題を解決する上で有効である。本研究では、小型マイクを装着したナイフによる食品認識システムを提案する。具体的には、切断音をマイクにより取得し、この音信号をスペクトル変換した特徴量を学習することで切断対象を認識する。提案システムの利点は、安価で小型のマイクセンサから得られる音信号のみで消費食品の認識が可能な点である。提案法の精度を検証するため3種類の食品（キャベツ・鶏肉・葱）の切断音データを用いた交差検証を行い、80%以上の正解識別率を確認した。

1. はじめに

食品消費モニタリングは、食糧廃棄の最適化、生活習慣病の予防など、食に関する社会的問題を解決する上で有効である。例えば、使用した食品量と食品在庫量をモニタリングすることで食品の買い過ぎを防止ができ[1]、消費履歴を記録することで、健康維持のために摂取すべき食品の提示や健康状態への影響を提示できる[2]。

近年、食とITを関連付ける研究は増加している。フライパンに取り付けたセンサにより調理を支援するPanavi[3]や調理時の動作や食品の識別を行う手法[4]が知られる。Kranz et.al. [4]は食品認識も行っているが、複数のセンサを用いたことにより実用性が低いという問題がある。

本研究では、食品消費モニタリングのための切断音を利用した食品認識システムを提案する。具体的には、まず、ナイフに装着したマイクを利用して食品の切断音を取得する(図1左)。さらに、この音信号を周波数解析に基づく特徴空間に射影し、この特徴空間でk-Nearest Neighbor (k-NN)法を適用することで切断対象をクラス分けする。提案法の精度を検証するため、3種類の食品(キャベツ・鶏肉・葱)の20個の切断音データを準備し、交差検証を行った。その結果80%以上の正解識別率を確認した。

2. 提案手法

本研究では、正解データ(特定の食品の切断音データ)が複数利用可能なもので、新たな食品の切断音を識別する、教師付き学習を行う。認識するための特徴量には、周波数解析を利用した特徴空間を用い、識別にはk-NNを利用する。

Copyright is held by the author(s).

立命館大学 情報理工学部メディア情報学科

1.1 正解データの収集

このステップでは、野菜の切断音の収集を行う。我々は、1音源当たり、5回の切断を行い、切断時の間隔は0.5~1.0秒程度になるように調節した。また、切断音の収集は、環境音の小さい空間で行った。

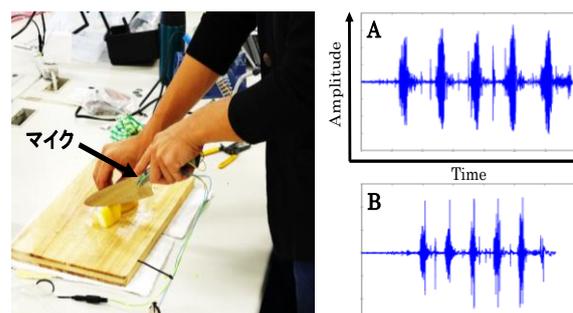


図1. 提案システム(左)、キャベツの切断音(A)と葱の切断音(B)

図1A・Bにキャベツと葱の切断音を示す。

1.2 特徴抽出

得られた切断音データを、以下の手順で特徴ベクトルに変換する。

前処理. 事前の調査において、人は切断音から対象食品をある程度認識できることを確認できたため、識別のために十分な情報が可聴域に含まれると考えられる。そこで、4.4kHzで録音された切断音を1.6kHzにダウンサンプリングする。また、プリエンファシスによる高周波成分の強調を行う。

離散フーリエ変換(DFT). 前処理された切断音からスペクトログラムを取得する。このとき、窓幅は4000(0.25秒)、オーバーラップは400(10%)とし、ハニング窓を利用した。図3左に得られたスペクトログラムを示す。

閾値処理. 得られたスペクトログラムの各"列"をひとつの特徴ベクトルとして利用する。しかし、スペ

クトログラムの各列は、ある瞬間のスペクトル情報を表現するものであり、切断音を含まない”列”も存在する。食品切断に関わる部分を抽出するため、スペクトログラムから音量の大きな列を抽出する。具体的には、各列をベクトル \mathbf{v} とみなし、ベクトルの長さ $\|\mathbf{v}\|$ が指定した閾値以上の列のみを後の計算に利用する。本研究では、閾値を 2000 とした(図 3 右)。

主成分分析による特徴次元圧縮。 処理の高速化のため主成分分析により、特徴ベクトルの次元数を削減する。本研究では、前述の閾値処理により抽出された特徴ベクトル \mathbf{v} の次元数が約 2000 次元であるのに対し、主成分分析によりこれを 30 次元まで圧縮した。

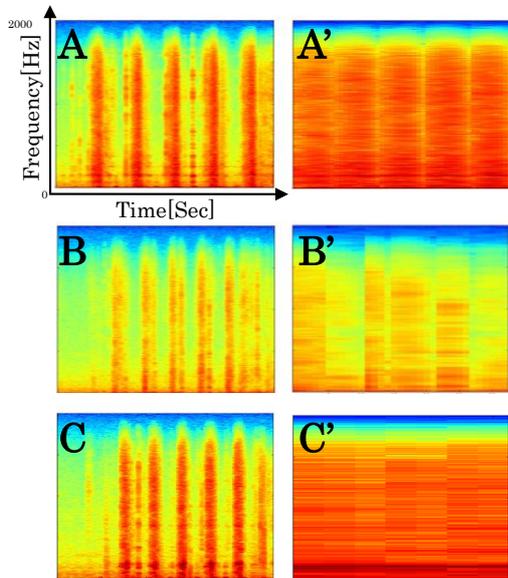


図 3. スペクトログラム 特徴抽出前(左)特徴抽出後(右)
キャベツ (A) 鶏肉 (B) 葱 (C)

3. 結果と考察

提案法を評価するため交差検証を行った。具体的には、録音された 20 音源 (キャベツ 4 音源、鶏肉 8 音源、葱 8 音源) のうち、任意の 1 音源をテストデータとし、その他を学習データとして扱い、テストデータの正解識別率を算出した (表 1)。結果、80% を超える平均正解識別率を確認した。

評価の結果、キャベツや鶏肉に関しては、高い識別率を得ることができたが、葱の識別率が低い結果となった。これは、葱の音源自体の音量が低いため、閾値処理を行った際に十分なデータ数が得られなかったと考えられる。

result(%)	80.26
-----------	-------

file number	cabbage(%)	chicken(%)	leek(%)
1	98.53	84.85	35.71
2	89.36	83.05	85.71
3	97.73	62.86	84.85
4	72.73	96.88	68.97
5	N/A	84.62	34.09
6	N/A	96.55	91.30
7	N/A	86.05	62.50
8	N/A	84.29	88.24
mean	88.14	84.39	65.45

表 1. 識別結果

4. 今後の課題

提案法の識別率は、取得するデータの質に強く影響を受ける。今回はマイクの集音面をナイフ表面に装着しデータを取得したが (図 1 左)、録音時に音が割れる現象を確認された。切断音を良好な状態で取得するには、マイク装着位置の再検討が必要である。

本稿では、単純な k-NN を用いても 80%以上の正解識別率を得られた。より強力な Support Vector Machine, Random Forest, Deep Neural Network を適用することで、より高い識別率が得られる可能性がある。また、特徴抽出時に食品切断の瞬間に着目した (スペクトログラムの各列を特徴ベクトルとして利用した) ため、時系列を無視することとなった。今後、スペクトルの時間的な変化を考慮した特徴ベクトルを利用することで、識別率の向上が見込まれる。

謝辞。 本研究は、JSPS 科研費 (15H05924) の助成をうけた。

参考文献

- [1] M. Fan, et. al. SoQr: Sonically Quantifying the Content Level inside Containers. *UbiComp'15 3-14,2015*
- [2] J. Mankoff, et. al. Using Low-Cost to Support Nutritional Awareness. *UbiComp'02 371-378,2002*
- [3] D. Uriu, et. al. panavi: Recipe Medium with a Sensors-Embedded Pan for Domestic Users to Master Professional Culinary Arts. *CHI'12 129-138,2012*
- [4] M. Kranz, et. al. Sensing Technologies and the Player Middleware for Context-Awareness in Kitchen Environments. *INSS'07 179-186,2007*