

視線情報の分析に基づく調理行動理解に向けて

井上 裕哉[†] 平山 高嗣[†] 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

^{††} 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

あらまし 本報告では調理行動の理解に向けて、視線情報の分析によって調理動作の識別可能性を検証する。視線情報が人間の行動を理解する上で重要であることから、頭部に装着した視線計測装置により視線運動データを取得し、調理動作識別に用いる。視線の移動パターンの表現手法には視線情報の分析に有用とされる N -gram の頻度ヒストグラムを採用する。これを SVM 識別器で学習した結果、対象の調理動作かそうでないかの 2 クラス識別において平均 F 値が 87.7% と高い精度が得られた。

キーワード 調理動作識別, 視線分析

Toward understanding cooking activities based on gaze analysis

Hiroya INOUE[†], Takatsugu HIRAYAMA[†], Ichiro IDE[†],

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

[†] Graduate School of Information Science, Nagoya University

Furo-cho, Chikusa-ku, Nagoya-shi, Aichi, 464-8601 Japan

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

Furo-cho, Chikusa-ku, Nagoya-shi, Aichi, 464-8601 Japan

Abstract In this report, we evaluate the possibility of classifying cooking actions by means of gaze analysis for the understanding of cooking activities. Since gaze information is important in understanding human behavior, we obtain it by a device mounted on the head, and use it to classify cooking actions. To analysis gaze information, we use the N -gram model known as effective for representing gaze transition. As a result of learning this by SVM classifiers, in case of a 2-classes problem where the input action was judged as a target action or not, an average F score of 87.7% was obtained.

Key words Classification of cooking actions, gaze analysis

1. はじめに

近年、料理教室が人気を博しており、調理技術の向上を目指す人が増えている。調理状態や調理者の技術に応じた支援を行うためには、調理者の調理行動をシステムが理解する必要がある。従来研究では、台所上の固定カメラで撮影された調理シーンの映像特徴に基づいて調理動作が認識されている [1]。本研究では、行動が認知・判断・動作の過程からなると考え、調理状況の視覚的な注意や認知を重要視し、調理者の注視行動を分析対象とする。視線運動は人間の内部状態を反映し [2]、その視野の周辺には行動に関連する情報が含まれる [3] とされている。また、これらの視線情報を計測できるウェアラブルセンサの高性能化、小型化、低廉化により一般的な人が日常的に使用できる環境が整いつつある。視線情報の分析に基づいて調理者の行

動を理解できれば、たとえば料理レシピへの掲載や調理映像の要約のために見どころを抽出できる。また、注視行動の個人差を抽出することで、調理動作のコツだけではなく、判断のコツを形式化することができる可能性もある。そこで、本研究ではそのような調理行動理解の前段階として、視線情報に基づいて調理の基本動作を識別することを目標とする。関連研究としては Bulling らが視線運動を利用して数種類のデスクワークの識別を行っている [4]。また大垣らは注意を向ける方向の小さな変化に関係する視線運動と、その大きな変化に関係する頭部運動を反映した 1 人称視点映像の Optical flow とを組み合わせて更に高精度にデスクワークを識別できることを示している [5]。そこで本報告では、これらの従来手法を応用して、視線情報の分析に基づいて調理の基本動作を識別できる可能性を検証する。

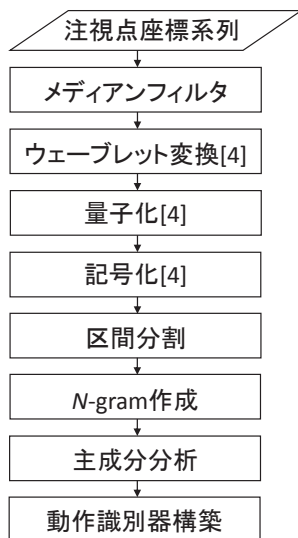


図1 調理動作識別の手順

2. 視線情報の分析に基づく調理動作識別手法

図1に本手法の処理手順を示す。計測した視線座標系列にはノイズや瞬き等を含むため、まず x 座標系列、 y 座標系列それぞれに対してメディアンフィルタを適用した後、Haar ウェーブレットによる連続ウェーブレット変換を行う。次に大小2種類（負の閾値を含めれば4種類）の閾値を設けて5段階に量子化を行う。そして量子化した x 座標系列、 y 座標系列を統合して記号化する。現フレームに付与された記号は前フレームからの相対的な注視点移動を表す。注視点の移動に応じて付与する記号を図2に示す。調理行動には調理対象への注視を継続させる振る舞いが頻繁に現れるため、視線移動のない原点を加えた25通りに変換する。以上のように作成した記号列を視線情報として利用する。この視線情報から抽出される特徴が調理動作によって異なると考えられる。本研究ではこの記号列をある一定の時間幅の解析窓で区切り、各々に1つの調理動作が含まれるものとする。解析窓はその幅を900フレーム（15秒）とし、60フレーム間隔で移動させる。ただし解析対象の時区間に続く次の60フレームの間で動作が変わる場合は、その動作が変わる時点までその解析窓を拡げる。そして解析窓ごとに N -gramの頻度ヒストグラムを作成する。本研究では1-gramから3-gramまでのヒストグラムを作成し、結合する。ここで、この結合ヒストグラムをそのまま特徴ベクトルとすると、次元が16,275次元と高次元になってしまうため、主成分分析により次元を削減する。最後に学習データ全ての解析窓で作成した特徴ベクトルを用いてSVM識別器を学習する。

3. 実験

本手法で調理の基本動作が識別可能かを検証するために、調理映像7本に対して調理動作の識別を行った。これらの調理映像は、4本がハンバーグの、3本がポテトサラダの調理過程を撮影したもので、調理動作として「Cut」、「Mix」、「Wait」、「Crush」、「Peel」の5種類を含む。前述の解析窓の幅で視線記

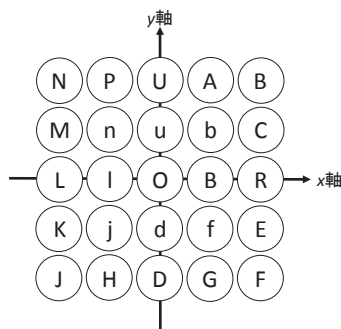


図2 注視点の移動に応じて付与する記号 [4]

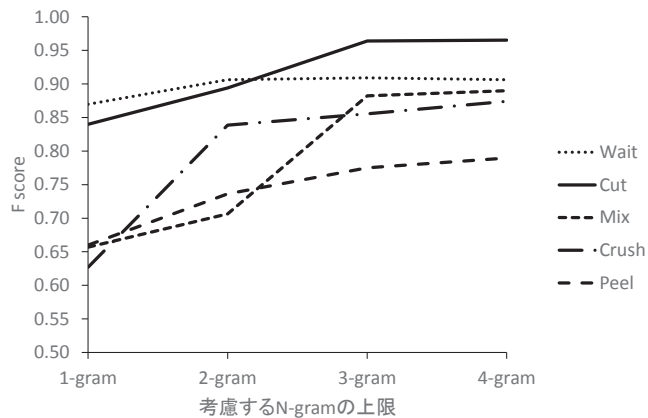


図3 実験結果

号列を分割したところ、サンプル数は7,880となった。それぞれに上記5つの動作のいずれかが割り当てられる。このサンプルに対して8分割交差検定を行った。視線計測にはNAC製EMR-9を用いた。評価実験では動作ごとに学習した識別器を用いて、対象動作がその動作かそれ以外の動作のいずれかという2クラス識別を行い、F値で評価を行った。考慮する N -グラムの上限とF値の関係を図3に示す。例えば2-gramまで考慮する場合は、識別に1-gramと2-gramの特徴量を使用する。3-gramまでは考慮する N の値を大きくするほど精度が上がる事が分かる。そしてその精度は3-gramまで考慮すると平均87.7%であった。なお最も精度が低いPeelについては、調理者が皮むきを手元に引き付けて行ったため、注視対象が視線計測装置のキャリブレーション平面から離れてしまい、視線を正確に計測できなかったことが原因と考えられる。

4. むすび

本報告では視線情報を用いて5種類の基本的な調理動作の識別可能性を検証した。実験では動作が混ざらない時区間を対象に2クラス識別で各動作の識別を行ったため高い精度が得られたが、今後は動作区間検出や多クラス識別を行うなどの改良を検討していく。

謝辞 本研究の一部は、科学研究費補助金による。

文献

- [1] Y. Hayashi, K. Doman, I. Ide, D. Deguchi, and H. Murase, "Automatic authoring of domestic cooking video based on the description of cooking instructions," Proc. 5th Int. Workshop on Multimedia for Cooking and Eating Activities (CEA2013), pp.21–26, Oct. 2013.
- [2] 平山高嗣, "人間の内部状態を顕在化する視覚的インタラクション," 情報研報, 2013-CVIM-188-27, Aug. 2013.
- [3] Y. Li, A. Fathi, and J. Rehg, "Learning to predict gaze in egocentric video," Proc. 2013 IEEE Int. Conf. on Computer Vision (ICCV2013), pp.3216–3223, Dec. 2013.
- [4] A. Bulling, J. Ward, H. Gellersen, and G. Troster, "Eye movement analysis for activity recognition using electrooculography," Proc. IEEE Trans. on Pattern Analysis and Machine Intelligence, Vol. 33, No. 4, pp.741–751, Feb. 2011.
- [5] K. Ogaki, K. Kitani, and Y. Sugano, "Coupling eye-motion and ego-motion features for first-person activity recognition," Proc. IEEE Workshop on Egocentric Vision in conjunction with CVPR2012, pp.1–7, June 2012.