

## 視線遷移と瞬きの時系列パターンに基づく調理動作識別手法<sup>※</sup>

井上 裕哉<sup>†\*a)</sup> 平山 高嗣<sup>†,††b)</sup> 道満 恵介<sup>†††,†c)</sup> 川西 康友<sup>†d)</sup>  
井手 一郎<sup>†,††e)</sup> 出口 大輔<sup>††††,†f)</sup> 村瀬 洋<sup>†,††g)</sup>

### A Classification Method of Cooking Operations Based on Temporal Patterns of Gaze Transitions and Blinks<sup>※</sup>

Hiroya INOUE<sup>†\*a)</sup>, Takatsugu HIRAYAMA<sup>†,††b)</sup>, Keisuke DOMAN<sup>†††,†c)</sup>,  
Yasutomo KAWANISHI<sup>†d)</sup>, Ichiro IDE<sup>†,††e)</sup>, Daisuke DEGUCHI<sup>††††,†f)</sup>,  
and Hiroshi MURASE<sup>†,††g)</sup>

あらまし 本研究では、調理者の視線遷移と瞬きの時系列パターンに基づいて調理動作を識別する手法を提案する。視線情報が人間の行動を理解するうえ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計測し、調理動作識別に用いる。視線パターンの表現には、視線情報に着目した動作認識において有用とされる *N*-gram を採用する。従来研究では、前フレームからの相対的な視線遷移のみが *N*-gram を構成する記号に用いられていた。これに対して本研究では、調理動作には調理対象への注視を持続させる振る舞いが頻繁に表れることから「停留」と、動作に対する集中を考慮するために「瞬き」を表現する記号を加えた。そして、高次元特徴を効率的に学習でき、特徴量のノイズに強いとされる Random Forest を利用し、視線パターンが各調理動作クラスに属する事後確率を出力する識別器を構築した。実験結果から各調理動作の性質と識別に対する寄与度が高い視線パターンとの関係性を明らかにしたうえで、調理動作識別において従来手法よりも高い識別精度が得られ、提案手法の有効性を確認した。

キーワード 調理動作識別、視線パターン、視線遷移、停留、瞬き

## 1. ま え が き

料理は、生命活動の一環として日常的に繰り返される行為であり、我々の生活において重要な役割を担っている。また、美味しい料理を作って食べることで、我々の生活はより豊かになると考えられる。このような背景もあり、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つつある。また、インターネットの普及により、Web 上の料理レシピや動画投稿サイトの調理映像を参考にして、調理を行う人も増えている。

このような時代背景の中で、情報通信技術を導入して調理を支援する様々なサービスが開発され、一般利用者に提供され始めている。例えば、2006年には携帯型ゲーム機向け調理支援ソフトウェア「しゃべる!DS お料理ナビ」[1]が発売され、100万本以上も販売された。このようなコンテンツを自動的に作成する方法に関する研究として、料理番組とその教材テキストの自動対応付けに関する研究 [2] や、食材や調理動作に関

<sup>†</sup> 名古屋大学大学院情報科学研究科, 名古屋市

Graduate School of Information Science, Nagoya University, Furo-cho, Chikusa-ku, Nagoya-shi, 464-8601 Japan

<sup>††</sup> 名古屋大学実世界データ循環学リーダー人材養成プログラム, 名古屋市

Graduate Program for Real-World Data Circulation Leaders, Nagoya University, Furo-cho, Chikusa-ku, Nagoya-shi, 464-8601 Japan

<sup>†††</sup> 中京大学工学部, 豊田市

School of Engineering, Chukyo University, 101 Tokodachi, Kaizu-cho, Toyota-shi, 470-0393 Japan

<sup>††††</sup> 名古屋大学情報戦略室, 名古屋市

Information Strategy Office, Nagoya University, Furo-cho, Chikusa-ku, Nagoya-shi, 464-8601 Japan

\* 現在, KDDI 株式会社

a) E-mail: inoueh@murase.m.is.nagoya-u.ac.jp

b) E-mail: hirayama@is.nagoya-u.ac.jp

c) E-mail: kdoman@sist.chukyo-u.ac.jp

d) E-mail: kawanishi@is.nagoya-u.ac.jp

e) E-mail: ide@is.nagoya-u.ac.jp

f) E-mail: ddeguchi@nagoya-u.jp

g) E-mail: murase@is.nagoya-u.ac.jp

<sup>※</sup> 本論文は食メディア研究専門委員会推薦論文である。

する映像データベースを構築し、料理レシピの各調理手順に対応する映像を自動的に付与することで、初心者が調理動作を視覚的に理解できるように支援する研究 [3] がなされている。

このように、様々な形で調理支援が模索されているが、これらは調理者側が調理状況に関する情報を積極的に要求することを想定している。一方で、調理中の初心者に有用な情報を適切なタイミングで与えるなど、システムが状況に応じて適応的に調理者を支援するインタラクティブな調理支援の方が有用な場合もあると考えられる。それを実現するためには、調理者が「何をしているのか」、あるいは「次に何をしようとしているのか」という調理行動の理解や予測を行う必要がある。そのためには、システムが「切る」や「混ぜる」といった基本的な調理動作を認識する必要がある。

従来、調理動作の自動認識に関する研究はいくつもある [4]～[6]。これらの研究では、調理者の身体運動に着目し、上方に設置した固定カメラから撮影された調理シーンの映像特徴に基づいて調理動作を認識するが、照明や調理器具などの環境的要因の変化に影響されやすいという問題がある。そこで本研究では、調理中の身体運動ではなく、視線情報に着目する。一般に、人間の行動は認知・判断・動作の過程を経てなされると考えられているため、調理状況の視覚的な注意や認知を重要視し、図 1 に示すような調理動作に付随する視線パターンの違いを分析する。ここで、視線情報を計測するためのウェアラブルセンサは、高性能化、小型化、低廉化により、一般の人が日常的に使用できる環境が整いつつある。

視線情報を用いた動作認識に関する研究として、



(a) 切る (Cut)

(b) 混ぜる (Mix)

図 1 異なる調理動作における視線運動の例。赤丸と赤線はそれぞれ、現在の視線座標と過去の視線座標の軌跡を表す。

Fig. 1 Example of eye movement in different cooking operations. The red circle and the red line represent the current gaze position and the trajectory of the past gaze positions, respectively.

Bulling らは視線遷移のパターンを利用して数種類のデスクワークを識別した [7]。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一人称視点映像の Optical flow 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た [8]。しかし、本研究で対象とする調理動作とデスクワークとは動作の性質が異なる。例えば、デスクワークでは視線が頻繁に動くのに対し、調理動作では調理対象への局所的な注視（視線の停留）が持続したり、集中を反映した瞬きが現れると考えられる。そのため、従来研究による視線運動データからの特徴抽出とそれに基づく視線パターンの表現では、このような性質を十分に分析することができないと考えられる。

そこで本研究では、調理動作識別に適した特徴抽出を導入することで従来手法 [7] を改良し、それによって表現された視線パターンに機械学習を適用することで、調理の基本動作の識別精度を向上することを目指す。本研究の主な貢献としては、(1) 視線の遷移だけではなく、停留及び瞬きを考慮した視線パターンの表現、(2) 視線パターンの頻度分布から、調理動作識別に寄与する特徴的なパターンの抽出、の 2 点が挙げられる。なお、視線パターンの表現には、従来手法 [7] と同様に  $N$ -gram を採用し、機械学習のアルゴリズムとしては、Random Forest [9] を適用する。

このようにして、視線情報に基づいて調理者の行動を理解することができれば、調理者支援のほか、料理レシピにおける調理手順画像の自動抽出や調理映像を要約するために「見どころ」を自動抽出することもできる。また、熟練者と初心者における視線パターンの違いを抽出することで、調理動作のコツだけではなく、認知・判断のコツを形式化することができる可能性もある。本研究では、これらを実現するための基盤技術を提案する。

以下、2. で停留と瞬きを考慮した視線パターンに基づく調理動作識別手法を提案し、3. で提案手法の有効性を確認するための評価実験について説明する。そして、4. で特徴的な視線パターンの分析を行い、5. で調理動作に関する多クラス識別の結果を示す。最後に、6. で今後の課題を挙げ、本論文を結ぶ。

## 2. 停留と瞬きを考慮した視線パターンに基づく調理動作識別手法

### 2.1 視線計測データの記号化

提案手法の処理手順を図 2 に示す. 本手法は, 学習フェーズと識別フェーズで構成され, それぞれにおいて視線計測データを入力として, 記号化, 解析窓分割, 特徴ベクトル作成, 識別器構築若しくは識別器適用が順に処理される. 本節では, 記号化について述べる. 図 3 は, その手順を示す.

視線計測データは,  $x$  座標系列,  $y$  座標系列の二つの時系列データからなる. まず, ノイズを除去するために, 各系列に対してメディアンフィルタを適用する. 次に, 各フレーム  $b$  について, 数フレーム前からの視線遷移ベクトル  $C(b) = (C_x(b), C_y(b))$  を計算する. ここで  $C_x(b), C_y(b)$  は, それぞれ  $x$  方向の視線遷移

量,  $y$  方向の視線遷移量を表す. そして, 従来手法 [7] と同様に式 (1)~(3) に示すスケールパラメータ  $\alpha$  を固定した Haar ウェーブレットによる連続ウェーブレット変換 (CWT-SD) を適用する. ここで,  $x(t), y(t)$  はそれぞれ  $t$  フレーム目において計測された  $x$  座標と  $y$  座標を表す.

$$C_x(b) = \frac{1}{\sqrt{\alpha}} \int \psi\left(\frac{t-b}{\alpha}\right) x(t) dt \quad (1)$$

$$C_y(b) = \frac{1}{\sqrt{\alpha}} \int \psi\left(\frac{t-b}{\alpha}\right) y(t) dt \quad (2)$$

$$\psi(\beta) = \begin{cases} 0 & (\beta < 0) \\ 1 & (0 \leq \beta < \frac{1}{2}) \\ -1 & (\frac{1}{2} \leq \beta < 1) \\ 0 & (1 \leq \beta) \end{cases} \quad (3)$$

そして, 大小 2 種類のしきい値  $H$  と  $L$  (負のしきい値を含めれば 4 種類) を設けて 5 段階に量子化を行ったうえで, 量子化した  $x$  座標系列  $\hat{C}_x(b)$ ,  $y$  座標系列  $\hat{C}_y(b)$  を統合して記号化する. なお, スケール  $\alpha$  は視線計測の標準化周波数に依存させる.

次に, 量子化された視線遷移ベクトル  $\hat{C}(b) = (\hat{C}_x(b), \hat{C}_y(b))$  の向きと長さに応じて, 図 4 に示す記号を各フレーム  $b$  に付与する. 図 4 中の記号において, 小文字が小さな視線遷移を, 大文字が大きな視線遷移を表す. これにより, 視線遷移量が小さいことが想定される「切る」などの動作と, 大きいことが想定される「混ぜる」などの動作の両方を扱えるようにする.

調理動作には食材や調理器具への注視を持続させる振る舞いが頻繁に現れるが, デスクワークを想定した従来研究 [7] で提案されている記号化方法ではそのよ

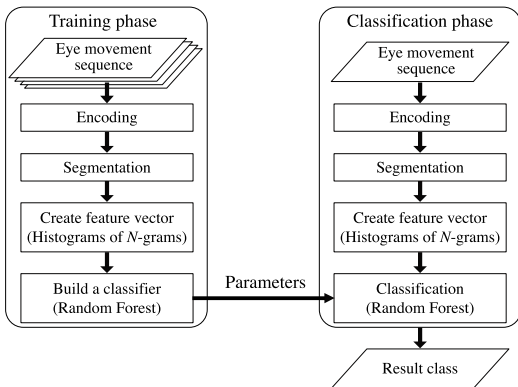


図 2 調理動作の識別手順  
Fig. 2 Classification steps of cooking operations.

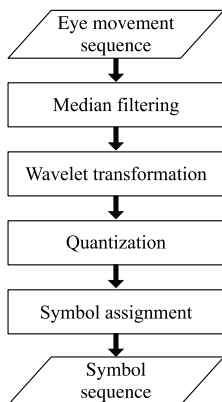


図 3 視線データの記号化手順  
Fig. 3 Symbolization steps of eye movement data.

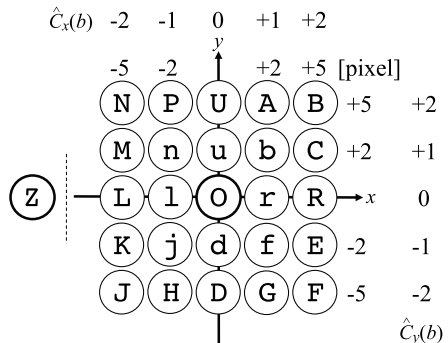


図 4 視線遷移及び瞬きに応じて付与する記号  
Fig. 4 Symbols that represent gaze transitions and eye blinks.

うな視線の停留が考慮されていない。そこで、視線遷移がないフレームを表現する記号「0」を図4の原点に追加する。また、調理動作には「切る」など、常に高い集中を要する動作と、「混ぜる」など、それほど集中を要さない動作がある。そこで、集中状態を反映する瞬き[10]を考慮することで、更なる識別精度の向上が見込めると考え、瞬き記号「Z」を追加する。これらの二つを追加し、合計26種類の記号のいずれかで各フレームでの視線の状態を表現する。

## 2.2 解析窓分割

以上の記号化によって作成した記号列から抽出される視線パターンが調理動作によって異なると想定する。本研究では、この記号列をある一定の時区間( $T$ フレーム)の解析窓で区切り、各々に一つの調理動作が含まれるものとし、区間ごとに調理動作を識別する。なお、この解析窓は、 $F$  ( $< T$ ) フレームずつ移動させる。ただし、解析対象の時区間から動作が継続し、次の $F$ フレームの間で動作が変わる場合は、その時点まで解析窓を拡げる。

## 2.3 $N$ -gram を用いた特徴ベクトルの作成

$N$ -gram モデルとは、ある記号列の中に連続して出現する $N$ 個の記号のパターン各々の出現頻度を分析するために用いられる言語モデルである[11]。例えば、記号一つの場合は1-gram (uni-gram)、二つの記号が連続して構成されるパターンは2-gram (bi-gram) と呼ばれる。 $N$ -gram は、記号の発生確率が直前の記号に依存すると仮定したモデルである。本研究における視線パターンも、直前の記号に依存すると考えられるため、上述の解析窓ごとにその時区間に含まれる $N$ -gram を抽出する。

従来研究[7]では $N$ -gram のヒストグラムから算出した五つの統計量(最大出現パターンの出現数、全出現パターンの平均出現数、出現パターン数、パターンの出現数の分散、最大出現数と最小出現数の差)を特徴量としている。しかし、これらはいずれも特定の視線パターンに依存した特徴量ではないため、調理動作識別のための表現力が低い可能性があり、識別に寄与する特徴的な視線パターンを分析することも難しい。そこで、本研究ではヒストグラムそのものを特徴量とし、1-gram から $N$ -gram までのヒストグラムを結合したものを、調理動作識別を行うための特徴ベクトルとして用いる。

## 2.4 識別器の構築

前節で作成した特徴ベクトルに対して、我々は従来、

Support Vector Regression (SVR) を適用した識別器を提案した[12]。そして、識別器が出力するゆう度が最大の調理動作クラスを識別結果とみなす多クラス識別で Bulling らの手法[7]より良い識別精度を得たが、十分な精度向上には至らなかった。2クラス識別の高精度な汎用モデルである Support Vector Machine (SVM) を多クラス識別に拡張する手法がいくつか提案され、その中で Platt の事後確率を用いる手法が最も高精度であると考えられる[13]が、データの性質や数に依存しないことが実験的に示されているわけではなく、その是非に対する議論が尽きない。そこで本研究では、高次元特徴を効率的に学習でき、特徴量のノイズに強いとされる Random Forest [9] を特徴ベクトルに適用し、標本に含まれる視線パターンが各調理動作クラスに属する事後確率を出力する識別器を構築する。

Random Forest は、複数の決定木を弱識別器として使い、その結果を統合して識別結果を得る。決定木のそれぞれのノードに、標本が進むべき子ノードを決定する分岐関数が与えられ、標本が到達した決定木の葉において各識別クラスに対する事後確率が算出される。決定木を構築する際に用いる標本は無作為に学習データから選択され、分岐関数を学習するための特徴次元としきい値の候補も無作為に選択される[14]。Random Forest の特徴として、特徴次元の重要度(識別寄与度)を算出できることも挙げられる[15]。具体的には、学習時に各ノードで標本集合が分割されるが、その際に減少する情報エントロピーの量に基づいて計算される。本研究においては、それに基づいて識別に有用な視線パターンを抽出することが可能となる。

## 3. 評価実験

### 3.1 評価方法

本論文では、まず、調理の基本動作と視線パターンとの関係性を検証する。そのために、解析窓で区切られた各時区間に含まれる視線パターンについて、それらが学習した調理動作に付随するものか否かを識別する2クラス識別を適用する。そして、本研究で新たに追加した記号と識別精度との関係を分析するとともに、決定木の構成に基づいて、調理動作ごとの特徴的な視線パターンについて分析する。次に、本研究の目的である視線パターンに基づく調理動作の識別について、調理動作を一意に特定する多クラス識別を適用し、提案手法の有効性を評価する。本節では以降、これらのための実験設定と実験データについて述べ、視線座標

系列の記号化に用いるしきい値と  $N$ -gram の結合範囲に関するパラメータ設定のために行った事前実験について説明する。

識別器の構築に用いた Random Forest については、決定木の数を 5、各木の最大深さを 10、分岐閾数の特徴次元候補の選択回数を 150、しきい値候補の選択回数を 5 とした。特徴次元候補の選択回数は、特徴ベクトルの次元数の平方根程度とする一般的な設定に倣った [9]。それ以外のパラメータは実験的に定めた。また、本研究では閉眼によって眼球を計測できずに視線座標が欠損した区間に瞬きが生じたと仮定した。

ベースラインとなる比較手法には、提案手法の元となった従来手法 [7] を採用する。この手法では「0」と「Z」を含まない 24 種類の記号に基づく  $N$ -gram のヒストグラム ( $N = 1 \sim 4$ ) を作成し、それから抽出した前述の五つの統計量を特徴ベクトルとする。

### 3.2 実験データ

提案手法の有効性を検証するために調理者の視線を計測するとともに、目視で調理動作をアノテーションするための一人称視点映像を撮影した。様々なレベルの調理経験をもつ調理者を起用することで、調理動作に個人差を含むようにした。日常的に調理を行っている者 1 名 (調理者 A)、時々行っている者 2 名 (調理者 B と C)、ほとんど行っていない者 1 名 (調理者 D) の計 4 名がハンバーグを一度ずつ調理し、この 4 回分の調理試行でデータを計測した。また、調理者 B<sup>(注1)</sup> はポテトサラダも 3 回調理し、そのデータも計測した。

これらの 7 回の調理試行のうち、1 回分をテストデータとして、その他を学習データとする Leave-one-out 交差検定を行うことで、調理者間での視線パターンの一般性を考察する。なお、調理者には料理レシピに記載された手順通りに調理を行うことを要求したが、調理器具や食材などの配置には制約を与えなかった。以下の実験では、Leave-one-out 交差検定の各テストデータに対して得られた識別精度を平均して評価する。更には、多クラス識別の各テストデータに対して得られた識別精度を個別に比較することで、調理者間での視線パターンの一般性を検証する。

調理の基本動作としては、従来研究 [5] による分類クラスに基づく「Cut」、「Mix」、「Crush」、「Peel」の 4 種類に注目し、これに「Wait」を加えた。「Cut」は包丁で食材を切っている状態、「Mix」は箸などで食材を

混ぜている状態、「Crush」はマッシャなどで食材を切碎している状態、「Peel」は皮むき器などで食材の皮をむいている状態、「Wait」は食材の焼き加減やゆで加減を観察するなど、動作を待機している状態と定義した。なお、調理者が同じ食材や調理器具を保持している間は同じ動作が継続しているものとした。これらの定義は、食材や調理器具、そしてそれらの状態には依存せず、調理者の動作の性質に基づくため、調理者数は少ないものの、各動作クラスに多様なデータが含まれることを想定できる。図 5 に、調理者 B の調理動作の遷移を示す。一人称視点映像を参考にして目視により、視線計測データの各フレームにこれらの動作ラベルを付与した。ただし、該当する動作ラベルがないフレームには何も付与せず、その視線計測データを取り除いた。

調理者の視線計測と一人称視点映像の撮影には nac Image Technology 社製のウェアラブル視線計測装置 EMR-9 [16] を用いた。計測範囲は装置装着者の頭部正面方向を中心にして、水平方向に  $\pm 40^\circ$ 、垂直方向に  $\pm 20^\circ$  で、標準化周波数は 60 Hz である。そして、水平方向分解能及び垂直方向分解能は  $0.1^\circ$ 、一人称視点映像の解像度は 640 pixel (水平)  $\times$  480 pixel (垂直) である。較正のために、9 個のマーカをまな板等の調理器具と同一の平面に設置し、実験者が指定する順番で調理者がそれらを注視した。

視線計測装置の標準化周波数に従って、式 (1) におけるスケールパラメータ  $\alpha$  を 6 とし、解析窓の幅  $T$  は従来研究 [7] に合わせて 900 フレーム (15 秒) とした。時間粒度の細かい調理支援を想定して、1 秒ごとの動作識別を行えれば必要十分であると考え、解析窓の移動幅  $F$  は 60 フレームとした。この解析窓で記号列を分割したところ、特徴ベクトルの標本数は 7,880 となった。それぞれに五つの基本動作クラスのいずれ

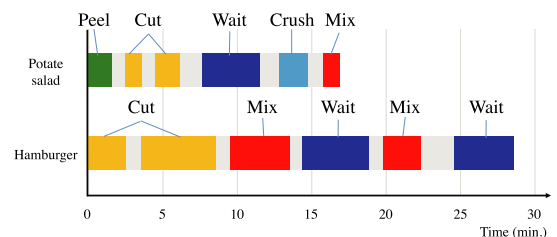


図 5 調理者 B による調理動作の遷移例  
Fig. 5 Example of transition of cooking operations by Subject B.

(注1)：調理者 4 名のうち、平均的な調理経験をもつ者を選んだ。

かが割り当てられる．また，本実験では視線座標が計測されなかった区間に瞬きが生じたと仮定したが，視線座標が欠損したまま CWT-SD を適用すると，記号化が不安定になるため，前後のフレームの視線座標を用いて線形補間を行った．

### 3.3 パラメータ設定のための事前実験

提案手法は，視線遷移ベクトルの向きと長さに応じて記号を付与する．図 4 中の記号において小文字と大文字，つまり小さな視線遷移と大きな視線遷移を適切に区別するためには，しきい値を適切に設定することが重要である．ここで小さな視線遷移を識別するためのしきい値  $L$  は視線運動の停留状態の定義 (100 ms の間に  $1^\circ$  以内の範囲に視線遷移が収まる状態) [17] に基づいて決定する．本実験で使用した EMR-9 で撮影された一人称視点映像において 480 pixel が視野角  $40^\circ$  に相当することから， $1^\circ$  あたり 12 pixel の解像度である．EMR-9 の標本化周波数が 60 Hz であることから，1 フレーム当たり 2 pixel までの視線遷移が停留状態として識別される必要条件となる．そのため，本研究ではしきい値  $L$  を 2 pixel とした．

大きな視線遷移を識別するためのしきい値  $H$  については，適切な値の仮説が存在しないため，実験的に求めた．3.2 で述べたデータを全て用い，しきい値  $H$  を変化させて 2 クラス識別を行ったところ，5 pixel と設定した際に平均 F 値の最大値が得られたため，以下の実験では，しきい値  $H$  としてその値を用いた．なお，F 値 (F-score) は識別器の正確性と網羅性の総合的な評価に用いられる一般的な指標で，再現率 (Recall) と適合率 (Precision) の調和平均として，次式によって求められる．

$$F\text{-score} = \frac{2 \times \text{Recall} \times \text{Precision}}{\text{Recall} + \text{Precision}} \quad (4)$$

$$\text{Recall} = \frac{TP}{TP + FN} \quad (5)$$

$$\text{Precision} = \frac{TP}{TP + FP} \quad (6)$$

ここで，TP (True Positive) は識別対象である調理動作クラスを正しく識別した標本 (解析窓) の数，FN (False Negative) はそれを誤って識別対象クラスではないと識別した標本 (解析窓) の数，FP (False Positive) は識別対象ではない調理動作クラスを誤って識別対象クラスであると識別した標本 (解析窓) の数である．

また，提案手法では 1-gram から  $N$ -gram のヒスト

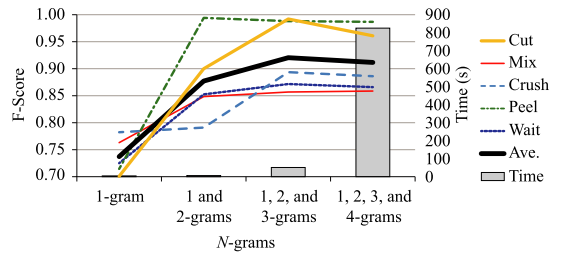


図 6 特徴ベクトルに用いる  $N$ -gram と 2 クラス識別の識別精度及び計算時間との関係

Fig. 6 Relation between 2-class classification accuracy and calculation time with  $N$ -grams used in the feature vector.

グラムを結合したものを特徴ベクトルとして用いるが，適切な  $N$  を決定する必要がある．図 6 に  $N$ -gram の結合範囲と F 値の関係を調理動作ごとに示す．おおむね  $N = 3$  までは  $N$  の値が大きいくほど識別精度が向上したが， $N = 4$  まで用いても向上はほとんど見られなかった．更に， $N = 4$  まで用いる場合には  $N = 3$  まで用いる場合と比べ，約 10 倍の計算時間を要した．これらより，以下の実験では  $N = 1 \sim 3$  の  $N$ -gram を結合したものを特徴ベクトルとして識別に用いた．

## 4. 2 クラス識別を利用した特徴的な視線パターンの分析

### 4.1 識別・分析方法

ある解析窓区間に含まれる視線パターンについて，それが学習した調理動作に付随するものか否かを識別する 2 クラス識別を適用する．これによって，追加記号と識別精度との関係について分析するとともに，決定木の構成に基づいて，調理動作ごとの特徴的な視線パターンについて分析する<sup>(注2)</sup>．識別器の構築に際しては，識別対象とする調理動作クラスの標本に正例，それ以外に負例のラベルを与えた 2 クラス識別を行う Random Forest 識別器を学習した．

### 4.2 識別精度に基づく追加記号の有効性の検証

提案手法では視線座標系列を記号化する際に，停留を考慮するために「0」，瞬きを考慮するために「Z」の二つの記号を従来手法 [7] の記号体系に追加した．本節では，Random Forest を用いた 2 クラス識別に対する追加記号の有効性を検証する．図 7 に，(1) 「0」も「Z」も含まない従来の記号化で作成した  $N$ -gram のヒ

(注2)：本論文では，調理動作識別に対する追加記号や視線パターンの貢献に注目して分析を行うが，それらの出現数が他の調理動作より少なくても，差が大きければ識別に対する寄与度が高くなる．

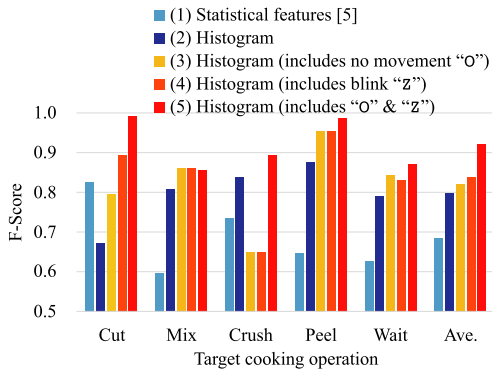


図 7 追加記号による識別精度の変化

Fig. 7 Change of classification accuracy according to additional symbols.

ストグラムから統計量を抽出 [7] し、Random Forest を適用した手法、(2) その統計量を抽出せずに提案手法と同様にヒストグラムそのものを特徴ベクトルとした手法、(3) それに加えて「O」だけを追加した手法、(4) 「Z」だけを追加した手法、(5) 「O」と「Z」の両方を追加した提案手法について、F 値を比較したものを示す。図中の右端に示した全調理動作の平均 F 値に着目すると、それぞれの記号が 2 クラス識別の精度向上に貢献していることが明らかである。以下では、各調理動作ごとに考察を行う。

「Cut」に関しては、停留と瞬きの両方を考慮することの有効性が最も高く表れた。切る動作は調理対象を注視する必要があることから停留「O」が、包丁を使うため動作に対する集中度が高まることから瞬き「Z」が、識別精度の向上に貢献したものと考えられる。これは、瞬きが集中するほど減少するという知見 [10] を考慮すると、他の 4 種類の調理動作より「O」を含むパターンが多く、「Z」を含むパターンが少ないことを意味する。

「Mix」に関しては、停留若しくは瞬きだけを考慮した場合の F 値が同程度で最も良く、両方を考慮した場合にわずかに F 値が低下した。これは、解析窓区間に「O」を含むパターンと「Z」を含むパターンの共起、若しくは「O」と「Z」を含むパターンが、他の 4 種類の調理動作と同程度に生じたためであると考えられる。

「Crush」に関しては、停留若しくは瞬きだけを考慮した場合に F 値が低下したが、両方を考慮した場合に最も高い F 値が得られた。これは、解析窓区間に「O」を含むパターンと「Z」を含むパターンの共起、若しくは「O」と「Z」を含むパターンが、他の 4 種類の調

理動作より多い、若しくは少ないことが考えられる。

「Peel」に関しては、ヒストグラムそのものを特徴ベクトルとすることによって大幅に F 値が向上し、追加記号によって更に向上した。皮むき中には、皮むき器が調理対象に接しているところを追従し、視線が一点に集中することは少ないことから停留「O」が、刃物を扱うため集中度が高まることから瞬き「Z」が、識別精度向上に貢献したと考えられる。これは、「O」を含むパターンも「Z」を含むパターンも、他の 4 種類の調理動作より少ないことを意味する。

「Wait」に関しては、他との差が小さかったものの、停留と瞬きの両方を考慮した場合に最も高い F 値が得られた。他の動作のように必ずしも集中する必要がないため、注視は分散し、瞬きが増加したと考えられる。

#### 4.3 調理動作ごとの特徴的な視線パターンの抽出

4.2 における考察の妥当性を検証し、調理動作ごとの特徴的な視線パターンを抽出するために、各調理動作の 2 クラス識別に対する各特徴次元の寄与度を求めた。寄与度は、決定木の各ノードにおける分岐前の情報エントロピーに対する分岐後の情報エントロピーの減少量に基づいて計算される。Leave-one-out 交差検定を行っているため、各調理動作において得られた F 値に近い値を示した学習データとテストデータの組を抽出し、その学習データを用いて構築された決定木から寄与度が高い視線パターンを抽出した<sup>(注3)</sup>。表 1～表 5 に、各調理動作に対して抽出された視線パターンを識別寄与度の上位 10 個を列挙した。なお、1-gram から 3-gram の視線パターンの合計は、18,276 個である。寄与度は情報エントロピーの減少量に基づくため、負の方向に大きい方が高い。

「Cut」に関しては、視線の上下方向への遷移 (Uf, Uu0, Dr) と右方向への遷移のパターン (uAR, rRR) の寄与度が高かった。前者は包丁の動きに誘目されているためであると考えられる。後者は、調理者が全員右利きで材料を左に置く傾向があり、右方向への視線遷移がほとんど起こらなかったためと考えられる。また、追加記号「O」を含むパターンは 2 個 (BBO, Uu0) が上位に入った。これは、調理対象を注視しながら、包丁の動きに誘目されている可能性がある。なお、図 7 で瞬き「Z」が識別精度の向上に貢献したことを確認

(注3)：抽出された視線パターンは学習データとテストデータの組に依存する。また、Random Forest を適用しているため、同じデータを繰り返し学習した場合でも、異なる視線パターンが抽出される可能性がある。

表 1 「Cut」における視線パターンと寄与度

Table 1 Eye movement pattern and contribution for “Cut”.

Rank	Pattern	Contribution
1	BBO	-0.8809
2	Uf	-0.8101
3	Uu0	-0.7903
4	uAR	-0.7903
5	GUG	-0.7903
6	BAP	-0.7903
7	nDb	-0.7903
8	rRR	-0.7503
9	bd	-0.5986
10	Dr	-0.5986

表 2 「Mix」における視線パターンと寄与度

Table 2 Eye movement pattern and contribution for “Mix”.

Rank	Pattern	Contribution
1	OPB	-1.0010
2	ubK	-0.9923
3	ZFJ	-0.9111
4	uAR	-0.9111
5	AGP	-0.7903
6	GUG	-0.7903
7	nDb	-0.7903
8	KE	-0.7868
9	JFF	-0.7068
10	D1	-0.7068

表 3 「Crush」における視線パターンと寄与度

Table 3 Eye movement pattern and contribution for “Crush”.

Rank	Pattern	Contribution
1	j	-0.9045
2	u00	-0.8579
3	BBb	-0.8338
4	rrr	-0.8148
5	BU	-0.7991
6	uu0	-0.7610
7	rA	-0.6881
8	Off	-0.6781
9	rLF	-0.5779
10	j11	-0.4579

したが、それを含むパターンは上位 10 個には入らなかった。

「Mix」に関しては、右回りに回転するような視線パターン (OPB, ubK, uAR, AGP) の寄与度が高くなった。これは、フライパンなどから食材がこぼれないように手の動きを追う視線遷移が他の動作に比べて特徴的であったことを示唆している。また、追加記号の「0」若しくは「Z」を含むパターンが 1 個ずつ (OPB, ZFJ) 上位に入っただけであり、他の調理動作ほどこれらの記号の寄与が高くなかった。

「Crush」に関しては、左下方向に小さく遷移するパ

表 4 「Peel」における視線パターンと寄与度

Table 4 Eye movement pattern and contribution for “Peel”.

Rank	Pattern	Contribution
1	BBB	-0.7647
2	AC	-0.7549
3	OR	-0.7054
4	duZ	-0.6794
5	d01	-0.6794
6	bE	-0.6256
7	jHd	-0.5638
8	OBb	-0.5247
9	fr0	-0.4879
10	Od	-0.1753

表 5 「Wait」における視線パターンと寄与度

Table 5 Eye movement pattern and contribution for “Wait”.

Rank	Pattern	Contribution
1	OZJ	-0.6908
2	CJ	-0.6764
3	ZC	-0.6764
4	dF	-0.6533
5	rCB	-0.6533
6	LR	-0.6198
7	Ed	-0.5795
8	uLL	-0.5380
9	ddH	-0.5154
10	Ujn	-0.4184

ターン (j) の寄与度が最高になったほか、一定の方向に 3 フレームとも遷移し続けるパターン (BBb, rrr) が上位に入った。これは、マッシャの一定方向への動きに誘目されることが多かったためと考えられる。また、追加記号「0」を含むパターンは 3 個 (u00, uu0, Off) が上位に入った。これは、「Cut」と同様に、調理対象を注視しながら、マッシャの動きに誘目されている可能性がある。「0」が連続するパターンが抽出されたことも考慮すると、「Cut」よりも停留が長かった、若しくはその頻度が多かったと考えられる。なお、図 7 で瞬き「Z」が識別精度の向上に貢献したことを確認したが、それを含むパターンは上位 10 個には入らなかった。

「Peel」に関しては、左上から右下方向に皮をむいていき、右下までむくとまた左上に戻すという一連の動作を繰り返すため、視線の左上方向と右下方向への遷移が多くなることが想定されるが、右上方向に遷移し続けるパターン (BBB, AC, OBb) の寄与度が高かった。これは、皮をむく際に、右上方向への視線遷移はほとんど起こらず、他の調理動作、特に「Crush」で頻繁に表れたため、識別に寄与したパターンとして抽



出されたと考えられる。また、追加記号の「O」を含むパターン5個（OR, dO1, OBB, frO, Od）と「Z」を含むパターン1個（duZ）が上位に入った。これは、図7に示された識別精度向上に対する「O」と「Z」の貢献を裏付ける。

「Wait」に関しては、特定の視線パターンの寄与度が高くならなかった。このように視線遷移が分散するときには瞬きの頻度が高くなり、「Z」を含むパターン2個（OZJ, ZC）が上位に入ったと考えられる。また、2-gramのパターンが上位に多く抽出され、図6と合わせてみると、2フレームで十分な識別精度を得られることが分かる。

## 5. 多クラス識別による調理動作識別

### 5.1 識別方法

ある解析窓区間に含まれる視線パターンに対して、それが付随する調理動作を一意に特定するためには、多クラス識別を行う必要がある。本研究では、五つの調理動作クラスのラベルをもつデータをまとめてRandom Forestを用いて学習して、5クラス識別器を構築し、テストデータに対して出力される事後確率が最も高い調理動作を識別結果とした。

### 5.2 識別結果

識別精度の混同行列（Confusion matrix）を表6に示す。この行列の各行のラベルは正解クラスを、各列のラベルは識別結果を表す。対角成分に示された識別率に着目すると、「Cut」、「Mix」、「Crush」、「Peel」に関しては、おおむね7割以上の標本を正しく識別することができたが、「Wait」に関しては半数弱の標本で正しく識別することができなかつたことが分かる。また、「Cut」に関しては再現率（recall）が高く、適合率（precision）が低いという結果になった。これは、「Cut」に含まれる視線パターンが他の動作においても共通して出現したためと考えられる。

表7に、手法ごとの全調理動作に対する平均識別率を示す。比較手法として「O」も「Z」も含まない従来の記号で作成した  $N$ -gram のヒストグラムから統計量を抽出したものの [7] に SVR を適用した手法、その統計量に Random Forest を適用した手法、提案手法と同様に、上記の統計量を抽出せず SVR を適用した手法 [12] を用いた。この表から、追加記号及び  $N$ -gram のヒストグラムをそのまま特徴ベクトルとすることの貢献は大きく、ヒストグラムから統計量を抽出する従来手法では、調理動作を識別する特徴を抽出すること

表6 多クラス識別結果（提案手法）

Table 6 Result of multiclass classification (Proposed method).

Actual class	Predicted class					Recall
	Cut	Mix	Crush	Peel	Wait	
Cut	<b>0.80</b>	0.08	0.00	0.04	0.08	0.80
Mix	0.26	<b>0.69</b>	0.00	0.00	0.05	0.69
Crush	0.04	0.00	<b>0.75</b>	0.20	0.01	0.75
Peel	0.15	0.00	0.01	<b>0.84</b>	0.00	0.84
Wait	0.32	0.16	0.00	0.00	<b>0.53</b>	0.53
Precision	0.51	0.74	0.98	0.78	0.79	—

表7 各手法における平均識別率

Table 7 Average classification rate by each method.

Method	Accuracy
Baseline (SVR version) [7]	0.401
Baseline (Random Forest version)	0.511
Proposed (SVR version) [12]	0.562
Proposed (Random Forest version)	0.662

が難しいと言える。更に、Random Forest で構築した識別器を用いることで精度向上を達成した。

### 5.3 調理者間での視線パターンの一般性に関する考察

前節までは、Leave-one-out 交差検定の各テストデータに対して得られた識別精度を平均したものを指標に提案手法を評価した。本論文で用いたデータセットの構成上、調理者 B の調理試行をテストデータにした評価では、学習データにも同じ調理者の他の調理試行が含まれ、過学習が生じている可能性がある<sup>(注4)</sup>。その他の調理者 (A, C, D) の調理試行をテストデータとした評価では、学習データに同じ調理者の試行が含まれない。そこで、各テストデータに対して得られた識別精度を個別に比較することで、調理者間での視線パターンの一般性を考察する。表8に、テストデータごとに求めた全調理動作に対する平均識別率を示す。

学習データにテストデータと同一人物が行った調理試行が含まれる四つの評価 (Test data 2, 5, 6, 7) に対するそれぞれの識別率を平均すると 0.676 であり、その他の三つの評価 (Test data 1, 3, 4) に対するそれが 0.702 であることから、個人差に基づく過学習が認められない<sup>(注5)</sup>。つまり、各テストデータの識別率に多少の分散があるものの、調理者間で視線パターンに一般性があると考えられる。一方で、調理者 B がボ

(注4)：提案手法の応用として、個人に特化した調理動作識別も想定されるため、この場合の過学習が必ずしも問題にはならない。

(注5)：学習データにもテストデータにも調理者 B の試行のみを含めて評価を行うことも考えられるが、本実験では学習データ数の平等性を重視して、その評価を行わなかった。

表 8 Leave-one-out 交差検定におけるテストデータごとの平均識別率

Table 8 Average classification rate of each test data on the leave-one-out cross validation.

Test data	Recipe	Subject	Accuracy
1	Hamburger	A	0.725
2	Hamburger	B	0.643
3	Hamburger	C	0.796
4	Hamburger	D	0.586
5	Potato salad	B	0.614
6	Potato salad	B	0.755
7	Potato salad	B	0.691

トサラダを調理した三つの試行に対する識別率にも分散が見られるため、調理経験が十分ではない調理者の視線パターンが安定しない可能性がある。なお、各テストデータの識別率が、表 7 に示す従来手法における平均識別率より高いことを確認することができる。

## 6. む す び

本研究では、インタラクティブな調理支援を実現するための基盤技術として、調理の基本動作を調理者の視線パターンに基づいて識別する手法を提案した。提案手法は、調理動作の性質を反映した視線の停留と瞬きを含む視線パターンを Random Forest によって抽出し、それを識別に寄与する特徴として用いた。実験によって、調理動作に付随する視線の遷移と停留、及び瞬きに基づく視線パターンを分析し、それに基づいて 5 種類 (「Cut」, 「Mix」, 「Crush」, 「Peel」, 「Wait」) の基本動作を識別することで、提案手法の有効性を検証した。

本研究で用いた実験データは、料理レシピの種類と調理者数が少なかったものの、各調理動作クラスを食材や調理器具に依存させなかったため、ある程度の多様性を含んでいると考えられるが、これらの基本動作を視線計測データだけを用いて高い精度で識別できることが示された。

本研究では、人間の内部状態 (認知・判断) を反映する視線情報のみに着目したが、調理者の注視対象を特定する手がかりとなる画像特徴を併用することで調理動作クラスを細分化して識別できるようになることが期待される。また、視線パターンが動作の開始、継続、終了の各区間で異なるという心理的知見 [18] が調理動作の分析にも適用できる可能性が高い。例えば、動作が切り替わる直前に、調理者の意識が次の動作に移り、その際に調理者の注意が次の動作に関連する調理器具や食材に移るなどである。システムが動作の切

り替わりを予測できると、適切なタイミングで調理者を支援することが可能になる。

本研究で扱った調理動作は 5 種類であり、実用には不十分である可能性がある。そのため、調理動作のクラス数を増加させたうえでより高い精度で多クラス識別を行えるように改良を検討する必要がある。また、本研究では、様々なレベルの調理経験をもつ調理者の視線計測データをひとまとめにして高い精度で基本動作識別を実現した。これは、調理経験のレベルによらず、動作間における視線パターンの差異が大きいのことを示唆するが、熟練者と未熟者との間に同一動作内での微妙な差異がある可能性もあるため、それを抽出したいと考えている。

**謝辞** 本研究の一部は、科学研究費補助金基盤研究 (A) (No.24240028) と若手研究 (B) (No.26730119) 及び名古屋大学実世界データ循環学リーダー人材養成プログラムの支援による。

## 文 献

- [1] 任天堂 (株), “しゃべる! DS お料理ナビ,” <https://www.nintendo.co.jp/ds/a4vj/>, 参照 March 25, 2016.
- [2] 三浦宏一, 高野 求, 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調理手順との対応付け,” 信学論 (D-II), vol. J86-D-II, no.11, pp.1647-1656, Nov. 2003.
- [3] 道満恵介, カイ承穎, 高橋友和, 井手一郎, 村瀬 洋, “マルチメディア料理レシピ作成のための料理レシピテキストと料理番組映像との対応付け,” 信学論 (A), vol. J94-A, no.7, pp.540-543, July 2011.
- [4] 久原 卓, 出口大輔, 高橋友和, 井手一郎, 村瀬 洋, “調理映像中の繰り返し調理動作の検出,” 信学論 (D), vol. J94-D, no.12, pp.1983-1985, Dec. 2011.
- [5] Y. Hayashi, K. Doman, I. Ide, D. Deguchi, and H. Murase, “Automatic authoring of a domestic cooking video based on the description of cooking instructions,” Proc. 5th Int. Workshop on Multimedia for Cooking and Eating Activities in Conjunction with ACM MM 2013, pp.21-26, Barcelona, Catalunya, Spain, Oct. 2013. DOI:10.1145/2506023.2506028
- [6] Y. Matsumura, A. Hashimoto, S. Mori, M. Mukunoki, and M. Minoh, “Clustering scenes in cooking video guided by object access,” Proc. 7th Int. Workshop on Multimedia for Cooking and Eating Activities in Conjunction with ICME 2015, pp.1-6, Torino, Piemonte, Italy, July 2015. DOI:10.1109/ICMEW.2015.7169812
- [7] A. Bulling, J. Ward, H. Gellersen, and G. Troster, “Eye movement analysis for activity recognition using electrooculography,” IEEE Trans. Pattern Anal. Mach. Intell., vol.33, no.4, pp.741-751, Feb. 2011.

DOI:10.1109/TPAMI.2010.86

- [8] K. Ogaki, K.M. Kitani, Y. Sugano, and Y. Sato, "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, Providence, RI, USA, June 2012. DOI:10.1109/CVPRW.2012.6239188
- [9] L. Breiman, "Random forests," Mach. Learn., vol.45, no.1, pp.5-32, May 2001. DOI:10.1023/A:1010933404324
- [10] T. Nakano, Y. Yamamoto, K. Kitajo, T. Takahashi, and S. Kitazawa, "Synchronization of spontaneous eyeblinks while viewing video stories," Proc. Royal Society B, vol.276, pp.3635-3644, July 2009. DOI:10.1098/rspb.2009.0828
- [11] P. Brown, P. Desouza, R. Mercer, V. Pietra, and J. Lai, "Class-based n-gram models of natural language," Computational Linguistics, vol.18, no.4, pp.467-479, Dec. 1992.
- [12] H. Inoue, T. Hirayama, K. Doman, Y. Kawanishi, I. Ide, D. Deguchi, and H. Murase, "A classification method of cooking operations based on eye movement patterns," Proc. 9th Biennial ACM Symposium on Eye Tracking Research & Applications, pp.205-208, Charleston, SC, USA, March 2016. DOI:10.1145/2857491.2857500
- [13] K.B. Duan and S.S. Keerthi, "Which is the best multiclass SVM method? An empirical study," N.C. Oza, R. Polikar, J. Kittler, and F. Roli eds., Multiple Classifier Systems, 6th Int. Workshop, MCS 2005, Seaside, CA, USA, June, 2005, Procs., Lecture Notes in Computer Science, vol.3541, pp.278-285, Springer-Verlag, Berlin, Heidelberg, 2005. DOI:10.1007/11494683\_28
- [14] P. Geurts, D. Ernst, and L. Wehenkel, "Extremely randomized trees," Mach. Learn., vol.63, no.1, pp.3-42, June 2006. DOI:10.1007/s10994-006-6226-1
- [15] 波部 齊, "ランダムフォレスト," 情処学コンピュータビジョンとイメージメディア研報, vol.2012-CVIM-182, no.31, pp.1-8, May 2012.
- [16] nac Image Technology, "NAC Image Technology introduces the latest in Eye Tracking Systems... The Eye Mark Recorder EMR-9," <http://www.nacinc.com/datasheets/archive/EMR9-Data-Sheet-June-09.pdf>, 参照 March 25, 2016.
- [17] D.E. Irwin, "Fixation location and fixation duration as indices of cognitive processing," J.M. Henderson and F. Ferreira eds., The Interface of Language, Vision, and Action: Eye Movements and the Visual World, pp.105-134, Psychological Press, New York, 2004.
- [18] M. Land and M. Hayhoe, "In what ways do eye movements contribute to everyday activities?," Vis. Res., vol.41, no.25-26, pp.3559-3565, Nov.-Dec. 2001.

(平成 28 年 3 月 27 日受付, 8 月 18 日再受付)



井上 裕哉

平 26 兵庫県立大・工・電気電子情報卒。  
平 28 名大大学院情報科学研究科博士前期課程了。修士(情報科学)。同年 KDDI 株式会社入社。視線パターンを用いた行動解析に興味をもつ。



平山 高嗣 (正員)

平 12 金沢大・工・電気情報卒。平 17 阪大大学院基礎工学研究科博士後期課程修了。博士(工学)。同年より京大大学院情報科学研究科特任助教。平 23 より名大大学院情報科学研究科特任助教。平 24 より同助教。平 26 より同特任准教授, 現在に至る。顔画像認識, 注視行動分析, 視覚的注意の計算モデルに関する研究に従事。平 21 ヒューマンインタフェース学会論文賞, ACM ETRA2014 Best Short Paper Award, IEEE ISM2014 Best Paper Award, 平 26 本会情報・システムソサイエティ論文賞受賞。情報処理学会, ヒューマンインタフェース学会, IEEE, ACM 各会員。



道満 恵介 (正員)

平 19 名大・工・情報卒。平 24 同大大学院情報科学研究科博士後期課程修了。博士(情報科学)。平 23 日本学術振興会特別研究員 DC2, 平 24 日本学術振興会特別研究員 PD, 平 24~25 まで中京大学情報理工学部講師, 平 25 より中京大学工学部講師。画像処理・パターン認識の技術開発とその ITS 応用及びマルチメディアコンテンツ生成技術に関する研究に従事。IEEE 会員。



川西 康友 (正員)

平 18 京大・工・情報卒。平 24 同大大学院情報科学研究科博士後期課程修了。博士(情報学)。平 24 同大学術情報メディアセンター特定研究員。平 26 名大未来社会創造機構特任助教。平 27 同大情報科学研究科助教, 現在に至る。防犯カメラ・車載カメラ映像を対象とした, 人物検出・追跡・検索を含む人物画像処理に関する研究に従事。平 23 年度 PRMU 研究奨励賞受賞。



井手 一郎 (正員：シニア会員)

平 6 東大・工・電子卒. 平 8 同大大学院工学系研究科情報工学専攻修士課程了. 平 12 同研究科電気工学専攻博士課程了. 博士 (工学). 同年国立情報学研究所助手. 平 16 名古屋大学大学院情報科学研究科助教授, 平 19 より准教授. この間, 平 14~16 総合研究大学院大学数物科学研究科助手併任, 平 16~22 情報・システム研究機構国立情報学研究所客員助教授・准教授兼任, 平 17, 18, 19 フランス情報学・統計システム研究所 (IRISA) 招聘教授. 平 22~23 オランダアムステルダム大学情報学研究所上級訪問研究員. パターン認識技術の実応用や映像メディア処理全般に興味をもっている. 情報処理学会シニア会員, 映像情報メディア学会, 人工知能学会, IEEE, ACM 各会員.



出口 大輔 (正員)

平 13 名大・工・情報卒. 平 18 同大大学院情報科学研究科博士後期課程修了. 博士 (情報科学). 平 16~18 まで日本学術振興会特別研究員. 平 18 名大大学院情報科学研究科研究員, 平 18 名大大学院工学研究科研究員, 平 20~24 まで同大大学院情報科学研究科助教, 平 24 より同大情報連携統括本部情報戦略室准教授, 現在に至る. 主に画像処理・パターン認識技術の開発とその ITS 及び医用応用に関する研究に従事. CARS2004 Poster Award, CADM2004 大会賞, 平 18 日本医用画像工学会奨励賞, 平 18 日本コンピュータ外科学会講演論文賞. IEEE 会員.



村瀬 洋 (正員：フェロー)

昭 53 名大・工・電気卒. 昭 55 同大大学院修士課程了. 同年日本電信電話公社 (現 NTT) 入社. 平 4 から 1 年間米国コロンビア大客員研究員. 平 15 から名大大学院情報科学研究科教授, 現在に至る. 文字・図形認識, コンピュータビジョン, マルチメディア認識の研究に従事. 工博. 昭 60 電子情報通信学会学術奨励賞, 平 6 IEEE-CVPR 最優秀論文賞, 平 7 情報処理学会山下記念研究賞, 平 8 IEEE-ICRA 最優秀ビデオ賞, 平 13 高柳記念奨励賞, 平 13 本会ソサエティ論文賞, 平 14 本会業績賞, 平 15 文部科学大臣賞, 平 16 IEEE Trans. MM 論文賞, 平 22 前島密賞, 平成 24 紫綬褒章, 他受賞. IEEE フェロー, 情報処理学会会員.