† †† ††† †††† †

† ††
464–8601
††† 470–0393 101
†††† 464–8601
E-mail: †inoueh@murase.m.is.nagoya-u.ac.jp, ††{hirayama,kawanishi,ide,murase}@is.nagoya-u.ac.jp

あらまし

N-gram

N-gram

SVR (Support Vector Regression)

0.168 0.856 F

キーワード N-gram SVR

Cooking Operation Classification Based on Analysis of Eye Movement Patterns

Hiroya INOUE†, Takatsugu HIRAYAMA††,†, Keisuke DOMAN†††,†, Yasutomo KAWANISHI†,
Ichiro IDE†, Daisuke DEGUCHI††††,†, and Hiroshi MURASE†

† Graduate School of Information Science, Nagoya University
†† Graduate Program for Real-World Data Circulation Leaders, Nagoya University
Furo-cho, Chikusa-ku, Nagoya-shi, Aichi, 464–8601 Japan
††† School of Engineering, Chukyo University
101 Tokodachi, Kaizu-cho, Toyota-shi, Aichi, 470–0393 Japan
†††† Information Strategy Office, Nagoya University
Furo-cho, Chikusa-ku, Nagoya-shi, Aichi, 464–8601 Japan
E-mail: †inoueh@murase.m.is.nagoya-u.ac.jp, ††{hirayama,kawanishi,ide,murase}@is.nagoya-u.ac.jp

Abstract In this presentation, we propose a classification method of cooking operations by analyzing eye movement patterns. Since gaze information is important in understanding human behavior, we obtain it by a head–mounted device, and use it to classify cooking operations. We use the N-gram model known as effective in action recognition that focuses on gaze information. Conventionally, only relative movement from the previous frame was used as symbols for the N-gram. However, since in cooking, users pay attention to cooking ingredients and equipments, we consider fixation as a component of the N-gram. We also consider eye blinks which may reflect concentration. The proposed method estimates the likelihood of the cooking operations by Support Vector Regression (SVR) using frequency histograms of N-grams as explanatory variables. The effectiveness of the proposed method was confirmed through an experiment, which obtained the average F-score of 0.856, 0.168 higher than the conventional method.

Key words Classification of cooking behaviors, gaze analysis, eye movement pattern, fixation, blink, N-gram

1. は じ め に

[1]

[2]

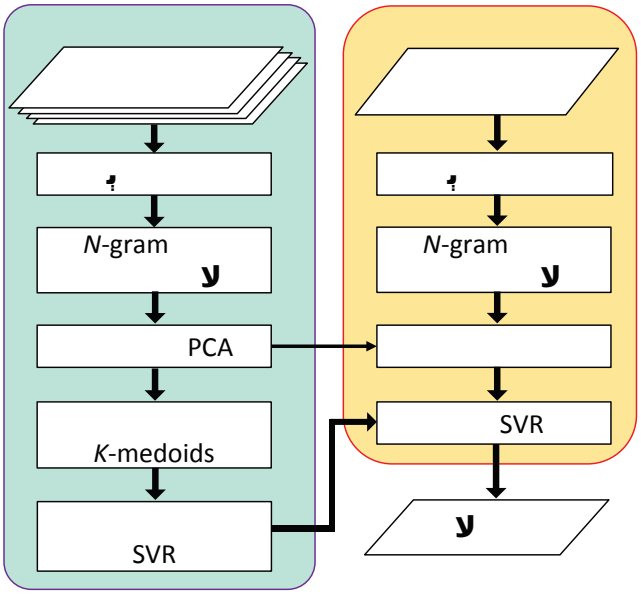
1

[3]

[4]

(a) Cut (b) Mix

1



2

[5]

Bulling

[6]

[7]

1

Optical flow

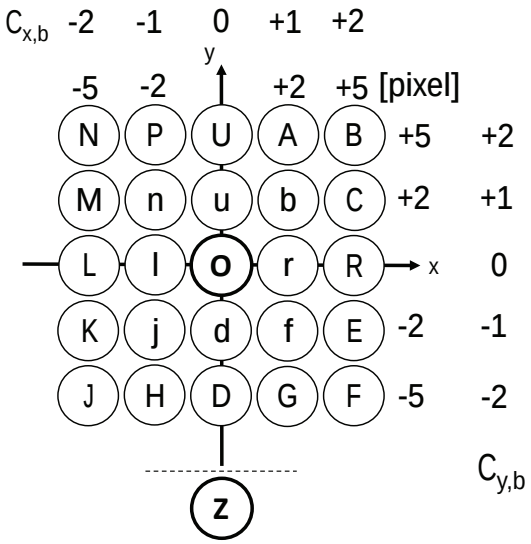
[8]

3

2.

3.

4.



4

2. 視線運動パターンの分析に基づく調理動作識別手法

2 N-gram 3

x y

$C = \{C_{x,b}, C_{y,b}\}_{b=1}^{T}$ [6]

(1), (2) α Haar (CWT-SD: Continuous Wavelet Transform for Saccade Detection)

$$C_{x,b} = \frac{1}{\sqrt{\alpha}} \int \psi\Big(\frac{t-b}{\alpha}\Big) x_t \, dt \tag{1}$$

$$\psi(\beta) = \begin{cases} 1 & (0 \leqq \beta < \frac{1}{2}) \\ -1 & (\frac{1}{2} \leqq \beta < 1) \end{cases} \tag{2}$$

$C_{x,b}$ b x x_t t x α 100 ms $\alpha = 6$

y 2 4 H L (3) 5 x y

$$\hat{C}_{x,b} = \begin{cases} 2 & (H \leqq C_{x,b}) \\ 1 & (L < C_{x,b} \leqq H) \\ 0 & (-L < C_{x,b} \leqq L) \\ -1 & (-H < C_{x,b} \leqq -L) \\ -2 & (C_{x,b} \leqq -H) \end{cases} \tag{3}$$

4 4

Cut Mix

[6] [6]

O

Cut Mix

[9]

Z 26

3 3

1 [6] 900 15 1 60 60

N-gram [6]

1-gram N-gram

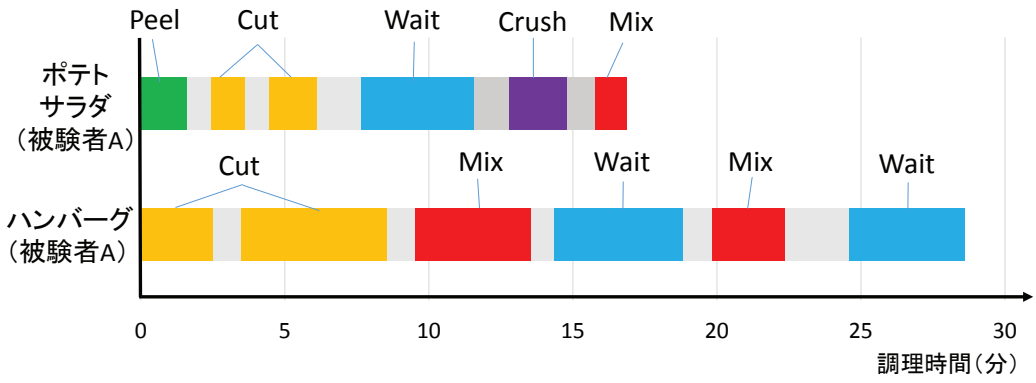
PCA: Principle Component Analysis

SVR [1] (Support Vector Regression) [10]

-1 1 one-against-all 0

2

1 lib-SVM epsilon-SVR



5

k-medoids [11]

k

3. 実　　験

7

4 3

Cut Mix Wait

Crush Peel 5 [2]

5

NAC EMR-9 [12]

$\pm 40^\circ$ $\pm 20^\circ$ 60 Hz

0.1° 1°

640 H 480(V) 2.

7,880

5

SVR

2

F F

precision

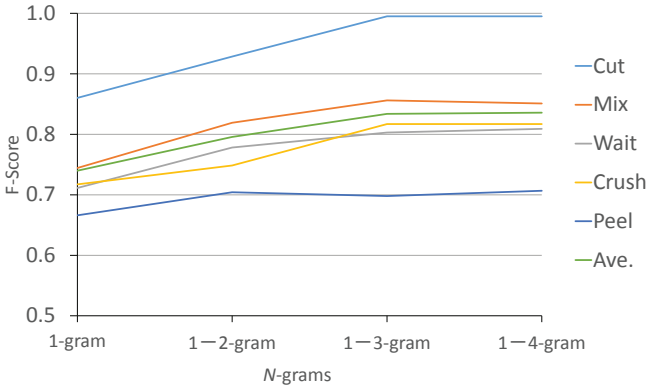
recall

$$F = \frac{2 \cdot \mathrm{precision} \cdot \mathrm{recall}}{\mathrm{precision} + \mathrm{recall}} \quad (4)$$

3.1 事 前 実 験

4

L [13]



6 N-gram

EMR-9 480 pixel 40°

1° 12 pixel

100 ms 1°

60 Hz 1

2 pixel

L 2 pixel H

$H = 5$ pixel

N-gram

N

6 N-gram F

$N = 3$ N-gram N

4-gram

4-gram

3-gram

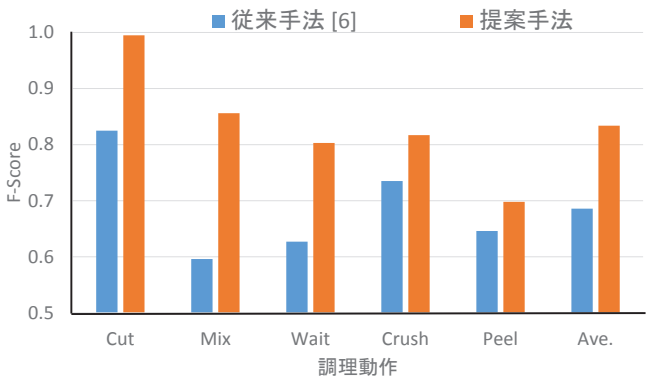
$N = 3$ N-gram

3.2 比 較 手 法

[6]

O Z

24 N-gram



7

8

1-gram 4-gram

4 Wordbook 5

(1) (2)

(3) (4) (5)

3.3 識 別 結 果

7

[6]

F [6]

0.168 0.854 Wait 9

9

3.4 特徴量間の比較

O Z

[6] O Z *N*-gram

O Z

8

Mix , Wait

Peel

3.5 多クラス分類による比較

2

Confusion Matrix

Cut , Mix , Wait

Crush , Peel

Mix recall

precision

Mix

4. む す び

5

2

N-gram

5. 謝　　辞

文　　献

[1] K. Doman, C. Kuai, T. Takahashi, I. Ide, and H. Murase, “Video CooKing: Towards the synthesis of multimedia cooking recipes,” Proc. 17th Int. Conf. on Multimedia Modeling (MMM2011), pp. 135–145, Jan. 2011.

[2] 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, pp.21–26, Oct. 2013.

[3] , “ ,” , 2013-CVIM-188-27, Aug. 2013.

[4] 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.

[5] , , , , , “ ,” , MVE), Vol. 114, No. 487, pp.47–48, Mar. 2015.

[6] 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.

[7] K. Kitani, T. Okabe, Y. Sato, and A. Sugimoto, “Fast unsupervised ego-action learning for first-person sports videos,” Proc. 24th IEEE Computer Society Conf. on Computer Vision and Pattern Recognition (CVPR2011), pp.3241–3248, June 2011.

[8] 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.

[9] T. Nakano, Y. Yamamoto, K. Kitajo, T. Takahashi, and S. Kitazawa, “Synchronization of spontaneous eyeblinks while viewing video stories,” Proc. R. Soc.B, Vol.276, pp.3635–3644, July 2009.

[10] R. Collobert and S. Bengio, “Support vector machines for large-scale regression problems,” J. Machine Learning Research, Vol.1, pp.143–160, Feb. 2001.

[11] D. Vinod, “Integer programming and the theory of grouping,” J. American Statistical Assoc., Vol.64, No.326, pp.506–519, June 1969.

[12] `http://eyemark.jp/product/emr_9/index.html`

[13] D. Irwin, “Fixation location and fixation duration as indices of cognitive processing,” The Interface of Language, Vision, and Action: Eye Movements and the Visual World, pp.105–134, Psychology Press, New York, 2004.