ISSN 2278-3091
Volume 8, No.1.4, 2019
International Journal of Advanced Trends in Computer Science and Engineering
Available Online at http://www.warse.org/IJATCSE/static/pdf/file/ijatcse1181.42019.pdf
https://doi.org/10.30534/ijatcse/2019/1181.42019

Improving Invisible Food Texture Detection by using Adaptive Extremal Region Detector in Food Recognition

Mohd Norhisham bin Razali [1,2], Noridayu Manshor [1], Norwati Mustapha [1], Razali Yaakob [1], Mohammad Noorazlan Shah Zainudin[3]

[1] Faculty of Computer Science and Information Technology, Universiti Putra Malaysia, Malaysia
[2] Faculty of Computing and Informatics, Universiti Malaysia Sabah, Malaysia
[3] Faculty of Electronics and Computer Engineering, Universiti Teknikal Malaysia, Melaka, Malaysia

ABSTRACT

The advancement of mobile technology with reasonable cost has indulge the mobile phone users to photograph foods and shared in social media. Since that, food recognition has become emerging research area in image processing and machine learning. Food recognition provides an automatic identification of the types of foods from an image. Then, further analysis in food recognition is performed to approximate the calories and nutritional information that can be used for health-care purposes. The interest region-based detector by using Maximally Stable Extremal Region (MSER) may provides distinctive interest points by representing the arbitrary shape of foods through global segmentation especially the food images with strong mixture of ingredients. However, the classification performance on food categories with less diverse texture food images by using MSER are obviously low compared to the other food categories that have more noticeable texture. The texture-less food objects were suffered from small number of extremal regions (ER) detection beside having low image brightness and small resolutions. Therefore, this paper proposed an adaptive interest regions detection by using MSER (aMSER) that provide a mechanism to choose appropriate MSER parameter configuration to increase the density of interest points on the targeted food images. The features are described by using Speeded-up Robust Feature Transform (SURF) and encoded by using Bag of Features (BoF) model. The classification is performed by using Linear Support Vector Machine and yield 84.20% classification rate on UEC100-Food dataset with competitive number of ER and computation cost.

Key words : Food recognition, MSER, Local features, Bag of Features

1. INTRODUCTION

There is strong correlation between obesity and overweight with the occurrence of so-called diet-related chronic diseases such as diabetes, heart disease, kidney diseases and even cancers. Dietary assessment is a treatment undertaken by medical practitioner and dietitian to combat obesity and overweight problems. However, the traditional dietary assessment is a tedious process that often lead to inaccuracy in making evaluation to describe the information about the foods consumed [1]. An adequate information is compulsory to be deliberated such as the preparation methods, portion size, brand, calories and nutritional contents that must be recorded in daily basis. Furthermore, the traditional dietary assessment tends to lead under-reporting problem [2].

An automatic dietary assessment via food recognition algorithm has become active research area under the umbrella of image processing and machine learning field [3]–[5]. By using food recognition, the calories estimation of foods can be calculated precisely. The mobile technology nowadays has been equipped with good imaging quality and at reasonable costs have provide the ubiquity way in acquiring images. Capturing food images have also become a phenomenon with the popularity of social media network. In fact, the explosive amount of food images in social media has potential to provide useful and real information about eating habits and food preferences in our society that can be benefited by the food and health-care industry.

Foods have complex appearance as food objects have non-rigid deformation and high variations that widen the intra-class variability and narrow the inter-class inter-similarities [6], [7]. Thus, feature representation method plays an important role in transforming the raw pixels of food images into higher semantic of representation. The feature representation by using local features are the common practices in food recognition. This is because of the complex appearance nature of foods can be effectively captured through the properties of local feature that invariant to illuminations, rotations, scale and orientation [6], [8] and a compact and discriminative features can be produced [9]. There are numerous types of local features in the literature. A research conducted by [10] employed the interest region detector by using Maximally Stable Extremal Region (MSER) to detect food interest points, considering MSER as among the best interest region detector in term of effectiveness and efficiency [11]. MSER detects a set of connected regions from an image to define the extremal regions (ER). In food recognition, MSER has capability to deal with arbitrary shape of foods and detects the grainy food objects via ER detection by using global segmentation.

The common problem of any interest points detector such as DoG and Hessian is its tendency to detect denser features only on the textured surface [12], [13]. In contrast, low number of interest points were detected for texture-less foods that affect its classification performance. MSER encountered the similar problem as the other interest points detector. In fact, according to [11], [14], MSER detect even smaller number of interest points among the interest points detector. Therefore, this study proposed an adaptive approach for ER detection in MSER (aMSER) that choose appropriate MSER parameter configuration to increase the density of ER on the texture-less food images.

The rest of the paper is organized as follows. In section 2, we provide related works on food recognition with adaptive approach. Section 3 describes the aMSER extraction mechanisms and the MSER parameter configuration. Section 4 and 5 presents the feature representation, dataset and performance measurement. Section 6 presents the experimental results and section 7 conclude this paper with the recommendation for future works.

2. RELATED WORKS

In general, a recognition process is composed by feature extraction, feature encoding and classification. Each process has its own components and configurations that impact the classification performance. For instance, the feature extraction required a rigorous evaluation to identify the types of feature, the sampling techniques, descriptor size and so on. The same case goes to feature encoding and classification stage that required certain extend of evaluation on the components used. The initial idea of an adaptation model in object recognition was exposed by [15] where an adaptive configuration of components in object recognition need to be designed to cater the diversity appearance of the objects that probably required different kind of component or configuration in order to perform effectively. Due to different nature of food objects with the other types of objects, food recognition required different kind of methods adaptation [6].

The use of various type of features are inevitable to cater the high variability of food objects. However, different foods might require different features, for instance, the colour feature provide a better description for ‘Potage’ and the shape feature may provide better description for ‘hamburger’. Concerning on this matter, an adaptive feature extraction by using Multiple Kernel Learning (MKL) have been proposed [16], [17] to measure the significance of the variety of features for food objects. The overall classification accuracy reported is 62.5% with poor recall rate on food category simmered pork, ginger pork saute, toast, pilaf and egg roll due to less diverse surface of these food objects.

The research conducted by [13] performed a technical investigation on the components within BoF model to determine the optimal sampling technique, descriptor size, types of local features, clustering for generate visual dictionary and the classifier. The classification accuracy was reported as 78%. However, the evaluation of the local features is only performed within the family of Scale Invariant Feature Transform (SIFT). SIFT has problem at describing the image with complex background [18]. As a result, the proposed BoF model still unable to create a discriminative feature to distinguish different class of foods.

In summary, the recognition performance on food categories that consist many texture-less food objects need to be improved by using suitable features and at reasonable computation cost in feature detection, feature description and feature encoding.

3. ADAPTIVE MAXIMALLY STABLE EXTREMAL REGION (AMSER) EXTRACTION

An adaptive system can be described as the capability of a system to react in a way according to the responses received from its surrounding. An adaptive system is incorporated with MSER (aMSER) in detecting the extremal regions in food images. By using aMSER, the selection of MSER parameter configuration can be performed based on the pre-defined conditions. There are certain conditions of food images that led to insufficient number of interest points and even worst, resulted null interest points. The invisible texture such as the liquid-based foods, small images and low images brightness contribute to the low number of interest points detection.

Indeed, in any interest points detector including MSER, the density level of interest points are governed by its parameter configuration [19]. For this reason, tweaking on MSER parameter has become necessary. However, we believed that the food images with denser interest points will probably not benefit much from the parameter configuration. In fact, the overwhelming number of interest points will in turn drag the timeframe for feature detection and feature encoding [20]. Thus, the detection of extremal region (ER) via aMSER will be executed in more flexible and sensitive based on the food images condition. Apparently, the aMSER will be expected to increase the number of ER on the targeted food images by configuring the intensity threshold (IT) value and maximum area variation (MAV) value. The following section explain on the MSER parameter configuration and flowchart to the development of aMSER.

3.1 MSER Parameter Configurations

The low number detection of ER occurred due to inability of ER to grow by using current intensity function as there are less sensitive towards the existence of ridgelines in the images. Samples of food with low ER detection are showed in Figure 1.

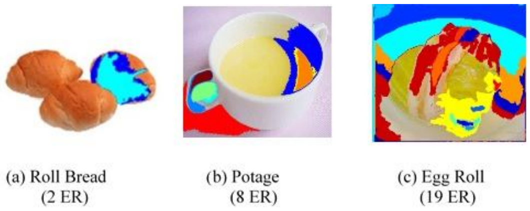
(a) Roll Bread (2 ER)
(b) Potage (8 ER)
(c) Egg Roll (19 ER)

Figure 1 : Samples of Food Image with Low Volume of ER

The sensitivity of MSER towards the invisible ridgelines as shown in Figure 1 can be increased by manipulating of IT and MAV value [21]. The MAV and IT are the parameters in MSER that control the region density and uniformity. A suitable threshold for MAV and IT should be determined to produce a stable region [21]. The evaluation of the parameters are necessary as the optimal value of IT and MAV are subject to the test data [22]. With this concern, an evaluation on the ITV and MAV have been conducted based on the parameter configuration as shown in Table 1.

Table 1: MSER Parameter Configurations

Stages	Configurations	IT	MAV
Stage 1	MSER 1	5	0.25
	MSER 2	3	0.25
	MSER 3	1	0.25
Stage 2	MSER 4	Optimum IT	0.5
	MSER 5	Optimum IT	0.75
	MSER 6	Optimum IT	1

The parameter evaluation is divided into two stages. The first stage is to find the optimum value of IT and the second stage is to find the optimum value of MAV. The MSER 1 is the original parameter configuration of MSER. For each run, the quantity of ER, the time taken for extraction and classification performance were recorded. The range of IT and MAV value are based on the recommendation in [23]. As the IT value decreased and MAV value increased, it will detect more ridgelines that produced a greater number of extremal regions.

3.2 Flowchart of aMSER algorithm

Figure 2 shows the flowchart of the execution of aMSER algorithm.

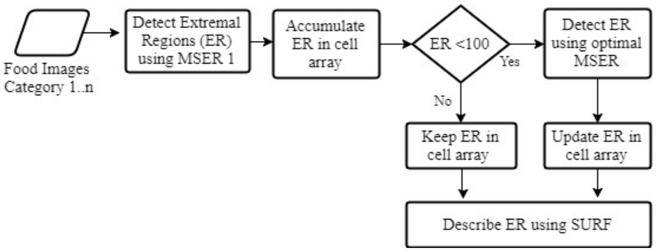
Food Images Category 1..n
Detect Extremal Regions (ER) using MSER 1
Accumulate ER in cell array
ER <100
Yes
No
Detect ER using optimal MSER
Keep ER in cell array
Update ER in cell array
Describe ER using SURF

Figure 2 : Flowchart of aMSER algorithm

The aMSER is executed in food category basis from category 1 to category n. Initially, the food images within a category were accessed. The images were converted from RGB (Red, Green, Blue) format into gray-scale format to reduce the complexity [24]. Then, the ER detected by using MSER 1, followed by counting the total of ER for each image and stored in a cell array. After that, a re-evaluation on the quantity of ER for each image is performed. During the re-evaluation, a condition is set. The condition states if the quantity of ER in an image is less than a 100, the image will need to repeat the ER detection stage where an optimal parameter of MSER will be applied. The observation indicates that the food categories that consist lot of foods with ER below than 100 has yield low classification rate. Then, all the new set of the quantity of ER will be updated in cell array before the features are extracted by using Speeded Up Robust Feature (SURF) descriptor. SURF is chosen to be paired with MSER due its balanced performance between accuracy and efficiency, less sensitive to noise and more practical for real time application [10], [11], [14]. Furthermore, SURF generates shorter length of feature vector that was reasoned for a speedy feature encoding process, produced a distinctive feature and robust to the geometric and photometric deformation.

4. FEATURE REPRESENTATION

The aMSER generate a huge and diverse amount of interest points. Literally, local feature can be represented as $X = [x_1, x_{2,\dots\dots,} x_n] \in R^{dxn}$ from n dimensional features from an image. For instance, hundreds or even thousands of interest points were generated per image and the amounts of interest points for all images may reach up to hundred thousand of interest points. With this condition, it was not feasible to feed the feature descriptions into machine learning classifier as it may incur lot of computational cost. Eventually, the representation of local feature needs to be transformed into another level of representation by using certain feature encoding technique.

Hard assignment technique encodes local feature by assigning each descriptor to the nearest visual word with indication of response 1 and the rest of visual words with response 0. The visual words are generated by using unsupervised learning algorithm that provide a model of local feature interest points distribution of X. Specifically, the clustering algorithm such as k-means is chosen due to its simplicity and it was commonly used in previous researches. The visual word is the terminology used in BoF which is referring the cluster centroid that was defined with cluster size of K or vocabulary size. Let a set of interest points are described as $X_1, \dots\dots, X_n \in \mathbb{R}^d$. Every interest point are assigned to visual words $q_{1,\dots\dots\dots} q_2 \in \{1, \dots. k\}$. In hard assignment, for each interest points X_m is assigned to cluster k, then $r_{mk} = 1$ and $r_{mj} = 0$ for $j \neq k$ and objective function can be defined as:

$$min\ j\ (\{r_{mk}, d_k\}) = \sum_{m=1}^{m} \sum_{k=1}^{k} r_{mk} \parallel x_m - d_k \parallel^2 \tag{1}$$

Then, the coding representation v for the local feature x is described as:

$$v(i) = \begin{cases} 1 & if\ i = arg\ min_y(\|x - b_j\|_2, i = 1,2, \ldots\ldots, m \\ 0, & otherwise \end{cases} \tag{2}$$

5. DATASET AND PERFORMANCE MEASUREMENT

The experiments are conducted by using UEC-Food100 dataset [25]. The UEC-Food100 dataset consists of 100 food categories with total of 14,467 images. Each image is having a different pixel dimensions and on average, there are around 150 images per category. These images are collected from the World Wide Web from real world settings with multiple classes of food types, great differences in image contrast, lighting and appearance. Figure 3 shows the samples of image from the dataset.

Figure 3 : Samples of UEC100-Food dataset

There are four performance measurement that were used to measure the classification performance which are classification rate, error rate, precision and recall. There are calculated by using the following formula:

$$\text{Classification Rate (CR)} = \frac{\text{Total of correctly classified images}}{\text{Total of test image}}$$

$$\text{Error Rate (ERT)} = \frac{\text{\% of incorrectly classified images}}{\text{Total of test image}}$$

$$\text{Precision (PREC)} = \frac{\text{Total of correctly classified images}}{\text{Number of predicted images}}$$

$$\text{Recall (REC)} = \frac{\text{Total of correctly classified images}}{\text{Ground truth number of images}}$$

In addition to that, the performance of detector and descriptor are measured based on the ideal properties recommended by [26] which are the quantity of interest points and the execution time. Both are mentioned as the most practical performance measurement for the real-time applications. The descriptor is also measured based in the compactness to describe the size of representation.

6. EXPERIMENTAL RESULTS

The results of the experiments can be divided into three sections. Section 6.1 presents the evaluation on MSER parameter. Section 6.2 presents the performance comparisons of aMSER with the other methods. Section 6.3 provide the performance on the texture-less food categories.

6.1 Evaluation of MSER parameter configuration

This section provides the experiments results of the MSER parameter configurations. The Intensity Threshold (IT) and Maximum Area Variation (MAV) value have been configured to increase the quantity of extremal region (ER) detection on food images with the number of ER below than 100. Figure 4 shows the classification rate and the quantity of ER for each MSER configuration. The effect of IT configuration can be referred in MSER 2 and MSER 3. While the effect of MAV configuration can be referred in MSER 4, MSER 5 and MSER 6. MSER 1 refers to the original parameter configuration.

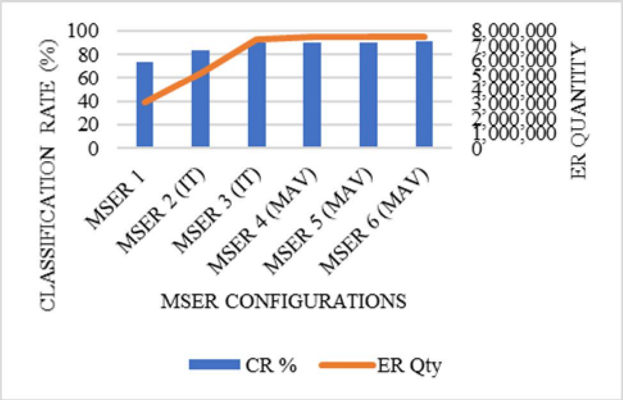

Figure 4 : Classification rate and ER quantity of MSER

The graph in Figure 4 shows the significant improvement of the classification rate through the IT configuration from 73.89% by using MSER 1 to 89.75% by using MSER 3. The quantity of ER has also increased dramatically which is about 140%. However, the configuration of MAV has showed little effect on both classification rate and ER quantity. Figure 5 shows the time taken in seconds for detection and encoding.

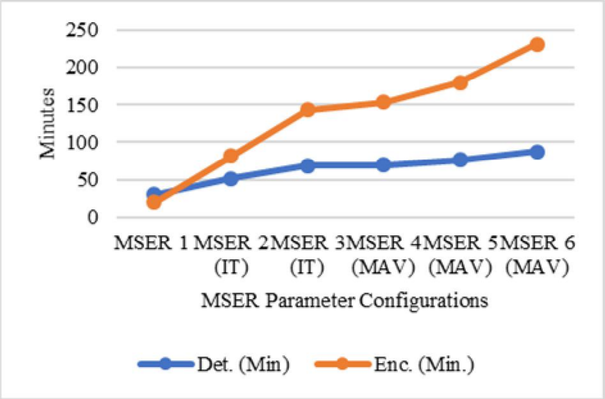

Figure 5 : Detection and encoding time of MSER parameter configuration

The graph in Figure 5 showed regardless IT or MAV configurations, both have consistently extended the detection and encoding execution time. The detection time has been increased by about 200%. The effect of the configurations is even more obvious on the encoding time which has spike for more than 100 times from the initial configurations in MSER 1. This is because this treatment (parameter configuration) has been applied to all food images regardless their interest points quantity. This problem has led to the idea of implementing adaptive mechanism in the MSER extraction where only certain food images are selected to undergo this treatment. Figure 6 shows the effect of IT and MAV configuration on a sample of food image. Figure 6 (c) and (d) show the effect of IT configuration and Figure 6 (e), (f) and (g) show the effect of MAV configuration.

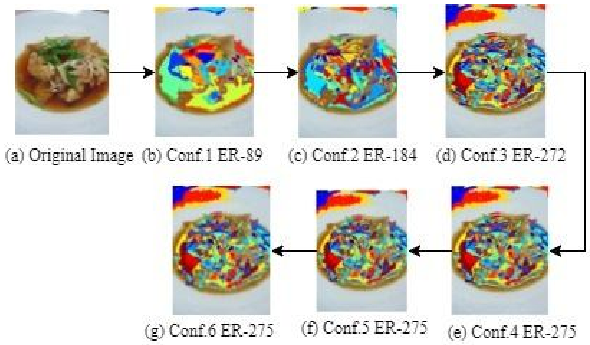

Figure 6 : Sample of ER detection by using different MSER parameter configurations

As aforementioned previously, the configuration of MAV has little effect on ER density. In fact, the configuration of MAV in Figure 6 (e), (f) and (g) have null effect on the ER quantity. While, the IT configuration has increased the ER quantity from 89 to 272. The ER from the background has increase as well. Also, the ER detection have become grainier as it was more sensitive towards region intensity. This finding shows the capability of MSER to detect regions from the fine-grained type of foods.

6.2 Evaluation of aMSER

This section presents the performance results of aMSER. By using aMSER, the extremal region quantity on the targeted food images have been increased by using the configuration MSER 6. The food images that have the number of ER below than 100 are usually food images with texture-less, small images and low contrast. Table 2 shows the comparisons of classification performance and extraction time between aMSER, MSER 1 and MSER 6.

Undeniably, the MSER 6 yield the best classification performance as the number of extremal regions or interest points are much denser. Indeed, dense interest points sampling tend to produce an informative feature representation that lead to better classification accuracy [27].

Table 2: Classification Performance and Extraction Time of aMSER

Performance Measurement	**MSER 1**	**aMSER**	**MSER 6**
Detection (min.)	**30.04**	113.7	87.33
Encoding (min.)	**19.73**	41.01	231.51
Classification Rate %	73.89	84.20	**90.96**
Error Rate %	0.40	0.20	**0.10**
Precision %	74.00	84.30	**91.00**
Recall %	73.90	84.20	**91.00**
Extremal Regions	**3,087,664**	3,576,594	7,610,852

In the flipside, it has also increased the quantity of ER by about 146% and 10 times for encoding time from the MSER 1. The proposed aMSER has also improved significantly the classification rate from 73.89% to 84.2% with only about 15% rise in ER quantity. The encoding time also demonstrated just a slight increase. Figure 7 and 8 showed a graph of performance comparisons of aMSER with the features that were used in previous food recognition including Histogram of Gradient descriptor by using Different of Hessian detector (DoH-HOG) [28], Speeded Up Robust Feature by using Different of Hessian detector (DoH-HOG) [29] and Scale Invariant Feature Transform by using Different of Gaussian detector (DoG-SIFT)[6].

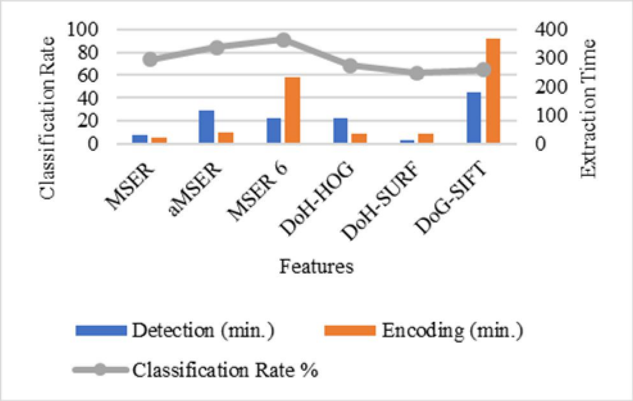

Figure 7 : Performance of features

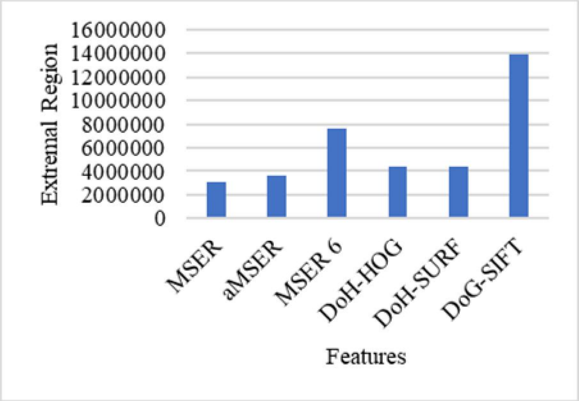

Figure 8 : No of ER between the features

The results showed the classification rate of aMSER and even MSER, has outperformed the DoH-HOG, DoH-SURF and DoG-SIFT. The extraction time of DoG-SIFT is the lengthiest and even more than MSER 6. Figure 8 shows the graph of the quantity of ER generated by the features.

DoG-SIFT generates the highest amount of ER. The aMSER produced lesser amount of ER than DoH-HOG, DoH-SURF and DoG-SIFT but still managed to get better classification rate as shown in Figure 7.

6.3 Evaluation of texture-less foods

This section presents the classification rate of food categories that contained many texture-less foods images as shown in Figure 9.

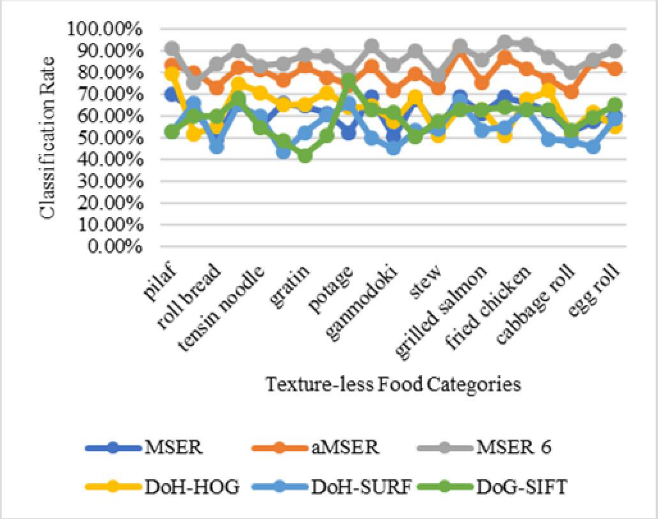

Figure 9 : Classification rate of texture-less food categories

Based on classification rate showed in Figure 9, the proposed method aMSER has improved the classification rate of texture-less food category by obtaining an average of 79.36%. Meanwhile, the average of classification rate by using MSER, HOG, SURF and SIFT are 61.07%, 63%, 54.9% and 58.97% respectively. Figure 10 shows the improvement of ER detection by using aMSER on few samples of texture-less food images. The configuration of IT and MAV value and in MSER has managed to increase the ER detection in the respective food images. Thus, more features can be captured and represented.

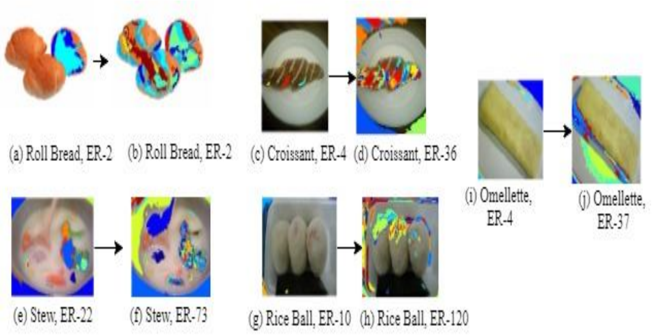
(a) Roll Bread, ER-2 (b) Roll Bread, ER-2 (c) Croissant, ER-4 (d) Croissant, ER-36 (i) Omellette, ER-4 (j) Omellette, ER-37 (e) Stew, ER-22 (f) Stew, ER-73 (g) Rice Ball, ER-10 (h) Rice Ball, ER-120

Figure 10 : Samples of extremal region detection on texture-foods by using aMSER

7. CONCLUSIONS AND FUTURE WORKS

The proposed aMSER has provides a flexibility in detecting interest regions to overcome the problem of interest points scarcity on food images with smooth texture such as the liquid-based foods, tiny images and low level of brightness. Furthermore, the datasets are built from the real-world setting or uncontrolled condition that make food images characterized by the inconsistency and variability of image quality. The aMSER has improved the classification rate of the texture-less food categories to 79.36% from 61.07% by using the traditional MSER as well as the other previous methods. In overall, the aMSER has improved the classification accuracy from 73.89% to 84.20%. The findings also highlighted the efficiency aspect of the recognition where a reasonable speed of interest points detection and feature encoding as well as compact number of interest points have been generated by using aMSER. In the future work, aMSER can be extended to self-adaptive ER detector where a learning algorithm can be incorporated to identify the most optimal parameter for each individual food images. The problem even can be modularized beyond using optimal parameter for the lack of interest region density since there is cases where small set of interest regions are already informative enough to describe the characteristic of foods. There are might be the other factors that can be considered other than tuning the MSER parameter to improve the recognition performance. By using self-adaptive detector, an optimal way in selecting parameter tuning and an optimum of interest points for each image can be performed.

ACKNOWLEDGMENT

The authors acknowledge the financial supported by the Putra Grant (Cost Center: 9569000) funded by the Universiti Putra Malaysia (UPM).

REFERENCES

[1] A. M. Coulston, C. j. Boushey, and M. Ferruzzi, G., **Nutrition in the Prevention and Treatment of Disease,** in *Nutrition in the Prevention and Treatment of Disease*, 3rd ed., Academic Press, 2013, pp. 5–30.

[2] F. Ragusa, V. Tomaselli, A. Furnari, S. Battiato, and G. M. Farinella, **Food vs Non-Food Classification**, in *2nd International Workshop on Multimedia Assisted Dietary Management*, 2016, pp. 77–81. https://doi.org/10.1145/2986035.2986041

[3] T. Ege and K. Yanai, **Image-Based Food Calorie Estimation Using Knowledge on Food Categories, Ingredients and Cooking Directions**, *Proc. Themat. Work. ACM Multimed. 2017*, pp. 367--375, 2017. https://doi.org/10.1145/3126686.3126742

[4] G. M. Farinella, D. Allegra, M. Moltisanti, F. Stanco, and S. Battiato, **Retrieval and classification of food images**, *Comput. Biol. Med.*, vol. 77, pp. 23–39, 2016. https://doi.org/10.1016/j.compbiomed.2016.07.006

[5] M. N. Razali and N. Manshor, **Object Detection**

Framework for Multiclass Food Object Localization and Classification, *Adv. Sci. Lett.*, vol. 24, no. 4, pp. 1357–1361, 2018.
https://doi.org/10.1166/asl.2018.10749

[6] F. Kong, H. He, H. A. Raynor, and J. Tan, **DietCam: Multi-view regular shape food recognition with a camera phone**, *Pervasive Mob. Comput.*, vol. 19, no. C, pp. 108–121, 2015.

[7] H. Kagaya and K. Aizawa, **Highly Accurate Food/Non-Food Image Classification Based on a Deep Convolutional Neural Network**, in *International Conference on Image Analysis and Processing*, 2015, vol. 9281, pp. 350–357.

[8] F. Zhu, M. Bosch, N. Khanna, C. J. Boushey, and E. J. Delp, **Multiple Hypotheses Image Segmentation and ClassificationWith Application to Dietary Assessment**, *IEEE J. Biomed. Heal. Informatics*, vol. 19, no. 1, pp. 377–388, 2015.
https://doi.org/10.1109/JBHI.2014.2304925

[9] Z. Zong, D. T. Nguyen, P. Ogunbona, and W. Li, **On the combination of local texture and global structure for food classification**, *Proc. - 2010 IEEE Int. Symp. Multimedia, ISM 2010*, pp. 204–211, 2010.

[10] M. N. Razali, N. Manshor, A. A. Halin, R. Yaakob, and N. Mustapha, **Food Category Recognition using SURF and MSER Local Feature Representation,** in *International Visual Informatics Conference*, 2017, pp. 212–223.
https://doi.org/10.1007/978-3-319-70010-6_20

[11] M. H. Lee and I. K. Park, **Performance evaluation of local descriptors for maximally stable extremal regions,** *J. Vis. Commun. Image Represent.*, vol. 47, pp. 62–72, 2017.

[12] S. Krig, **Local Feature Design Concepts, Classification, and Learning,** *Comput. Vis. Metrics*, pp. 131–189, 2014.

[13] M. M. Anthimopoulos, L. Gianola, L. Scarnato, P. Diem, and S. G.Mougiakakou, **A Food Recognition System for Diabetic Patients Based on an Optimized Bag-of-Features Model,** *IEEE J. Biomed. Heal. Informatics*, vol. 18, no. 4, pp. 1261–1271, 2014.
https://doi.org/10.1109/JBHI.2014.2308928

[14] P. Ma, M. Seeland, M. Rzanny, N. Alaqraa, and J. Wa, **Plant species classification using flower images — A comparative study of local feature representations,** *PLoS One*, pp. 1–30, 2017.

[15] X. Zhang, Y.-H. Yang, Z. Han, H. Wang, and C. Gao, **Object Class Detection: A Survey**, *ACM Comput. Surv.*, vol. 46, no. 1, pp. 1–46, 2013.

[16] T. Joutou and K. Yanai, **A food image recognition system with Multiple Kernel Learning**, in *2009 16th IEEE International Conference on Image Processing (ICIP)*, 2009, pp. 285–288.

[17] H. Hoashi, T. Joutou, and K. Yanai, **Image Recognition of 85 Food Categories by Feature Fusion**, in *IEEE International Symposium on Multimedia*, 2010.
https://doi.org/10.1109/ISM.2010.51

[18] J. Yu, Z. Qin, T. Wan, and X. Zhang, **Feature integration analysis of bag-of-features model for image retrieval**, *Neurocomputing*, vol. 120, pp. 355–364, 2013.

[19] E. Nowak, F. Jurie, and B. Triggs, **Sampling strategies for bag-of-features image classification,** in *9th European Conference on Computer Vision*, 2006, vol. 3954 LNCS, pp. 490–503.

[20] E. Salahat and M. Qasaimeh, **Recent Advances in Features Extraction and Description Algorithms : A Comprehensive Survey**, in *IEEE International Conference on Industrial Technology (ICIT)*, 2017.

[21] N. Takeishi, A. Tanimoto, T. Yairi, Y. Tsuda, F. Terui, N. Ogawa, and Y. Mimasu, **Evaluation of Interest-region Detectors and Descriptors for Automatic Landmark Tracking on Asteroids,** *Trans. Jpn. Soc. Aeronaut. Space Sci.*, vol. 58, no. 1, pp. 45–53, 2015.
https://doi.org/10.2322/tjsass.58.45

[22] S. Krig, **Interest Point Detector and Feature Descriptor Survey,** in *Computer Vision Metrics*, no. 1, Apress, Berkeley, CA, 2014, pp. 217–282.

[23] J. Matas, O. Chum, M. Urban, and T. Pajdla, **Robust Wide Baseline Stereo from**, *Br. Mach. Vis. Conf.*, pp. 384–393, 2002.

[24] S. Jabeen, Z. Mehmood, T. Mahmood, T. Saba, A. Rehman, and M. T. Mahmood, **An effective content-based image retrieval technique for image visuals representation based on the bag-of-visual-words model**, *PLoS One*, pp. 1–24, 2018.
https://doi.org/10.1371/journal.pone.0194526

[25] K. Yanai and Y. Kawano, **Food Image Recognition Using Deep Convolutional Network with Pre-Training and Fine-Tuning,** in *IEEE International Conference on Multimedia & Expo Workshops (ICMEW)*, 2015.

[26] A. Ziomek and M. Oszust, **Evaluation of Interest Point Detectors in Presence of Noise,** *Int. J. Intell. Syst. Appl.*, vol. 8, no. 3, pp. 26–33, 2016.
https://doi.org/10.5815/ijisa.2016.03.03

[27] B. Ionescu, J. Benois-Pineau, T. Piatrik, and G. Quenot, **Fusion in Computer Vision**, *Adv. Comput. Vis. Pattern Recognit.*, p. 272, 2014.
https://doi.org/10.1007/978-3-319-05696-8

[28] Y. Kawano and K. Yanai, **FoodCam: A real-time food recognition system on a smartphone**, *Multimed. Tools Appl.*, vol. 74, no. 14, pp. 5263–5287, 2015.
https://doi.org/10.1007/s11042-014-2000-8

[29] H. Pooja and P. S. A. Madival, **Food Recognition and Calorie Extraction using Bag-of- SURF and Spatial Pyramid Matching Methods**, *Int. J. Comput. Sci. Mob. Comput.*, vol. 5, no. 5, pp. 387–393, 2016.