EXTREMAL REGION DETECTION AND SELECTION WITH FUZZY ENCODING FOR FOOD RECOGNITION

MOHD NORHISHAM BIN RAZALI @ GHAZALI

FSKTM 2019 48

EXTREMAL REGION DETECTION AND SELECTION WITH FUZZY ENCODING FOR FOOD RECOGNITION

By

MOHD NORHISHAM BIN RAZALI @ GHAZALI

Thesis Submitted to the School of Graduate Studies, Universiti Putra Malaysia, in Fulfillment of the Requirements for the Degree of Doctor of Philosophy

June 2019

DEDICATION

To my loving parents and family.

To my other half, Rozita

and my kids, Lisha & Idlan Afi

Abstract of thesis presented to the Senate of Universiti Putra Malaysia in fulfilment of the requirements for the degree of Doctor of Philosophy

EXTREMAL REGION DETECTION AND SELECTION WITH FUZZY ENCODING FOR FOOD RECOGNITION

By

MOHD NORHISHAM BIN RAZALI @ GHAZALI

June 2019

Chair : Noridayu Manshor, PhD
Faculty : Computer Science and Information Technology

This study proposes the improvement of feature representation by using Maximally Stable Extremal Region (MSER) detector in Bag of Features (BoF) model which incorporates an interest points detection and selection, and fuzzy encoding for food recognition. Three algorithms were used to accomplish the task of feature representation. The first algorithm locates interest points in food images using an MSER. Dense sampling and Difference of Gaussian (DoG) have been used in previous studies but were unable to detect salient interest points due to complex appearance of food images. MSER provide discriminative features via global segmentation. The arbitrary shape of regions produced by the global segmentation is suitable to detect interest points from mixed food objects which are known to be characterised by non-rigid deformations and very large variations in appearance. However, the traditional MSER detects very few interest points on texture-less food images. Thus, an Extremal Region Detection (ERD) algorithm in MSER is improved by finding optimum configuration of MSER parameters, allowing the quantity of interest points for certain food images to be increased appropriately.

The second algorithm reduces the quantity of interest regions by using the Extremal Region Selection (ERS) algorithm. A high number of interest regions does not guarantee outstanding classification performance as redundant interest regions as well as interest regions from food images with complex background were detected. Consequently, computational effort should be used to execute the feature encoding process in the Bag of Features model. By decreasing the quantity of interest regions, the time efficiency of feature encoding can thus be improved without sacrificing classification accuracy. The ERS algorithm is performed using unsupervised learning to determine the spatial information of the interest regions detected, indicating whether they are from the image background, and can thus be removed as noise.

In the third algorithm, a soft assignment technique using fuzzy encoding is used to transform low-level features into a higher-level feature representation. The fuzzy encoding approach adopts fuzzy set theory (FST) to minimise the uncertainty and plausibility problems in feature encoding arising from hard assignment and fisher vector approaches used in previous studies. The uncertainty and plausibility problems have led to confusion in assigning feature descriptions to visual words, and they occur due to the high intra-class variability of food appearances due to high diversity in color and texture. By adopting FST, a thorough evaluation is performed in each assignment of feature description to visual words, which is translated into a membership value that indicates the relevance of that assignment.

The proposed methods have been evaluated using two image datasets: UECFOOD-100 and UNICT-FD1200. The performance of algorithms was measured based on classification accuracy, error rate, and precision and recall. The quality of the interest region detector was evaluated based on the quantity of interest regions. Classification was performed using a Support Vector Machine (SVM) with a linear kernel. The experimental results demonstrate the superior classification performance of the proposed methods over the previous methods. Specifically, the proposed method achieved 99.95% and 100.00% classification accuracy on the UECFOOD-100 and UNICT-FD1200 datasets, respectively, whereas previous methods have only been able to achieve 79.20% and 85.01% on the same datasets.

Overall, the propose method generates a compact and discriminative visual dictionary for food recognition using only a single feature type, small numbers of interest regions, and low-dimensional feature vectors. Moreover, it provides a holistic feature representation able to give outstanding classification performance on foods with great variation in appearance.

Abstrak tesis yang dikemukakan kepada Senat Universiti Putra Malaysia
sebagai memenuhi keperluan untuk ijazah Doktor Falsafah

PENGESANAN DAN PEMILIHAN KAWASAN EXTREMAL DENGAN PENGEKODAN KABUR UNTUK PENGECAMAN MAKANAN

Oleh

MOHD NORHISHAM BIN RAZALI @ GHAZALI

Jun 2019

Pengerusi : Noridayu Manshor, PhD
Fakulti : Sains Komputer dan Teknologi Maklumat

Kajian ini mencadangkan pembaikan perwakilan ciri dengan menggunakan *Maximally Stable Extremal Region* (MSER) di dalam model *Bag of Features* (BoF) yang menggabungkan pengesanan dan pemilihan titik minat, dan pengekodan ciri untuk pengecaman makanan. Tiga algoritma digunakan untuk mencapai tugas perwakilan ciri. Algoritma pertama mencari titik minat dalam imej makanan menggunakan MSER. Persampelan padat dan pengesan *Difference of Gaussian* (DoG) telah digunakan di dalam kajian terdahulu tetapi tidak dapat mengesan titik minat penting kerana penampilan kompleks imej makanan. MSER menyediakan ciri-ciri yang diskriminatif melalui pensegmenan global. Bentuk kawasan sembarangan yang dihasilkan oleh pensegmenan global adalah sesuai untuk mengesan titik minat dari objek campuran makanan yang diketahui mempunyai ciri ubah bentuk tidak tegar dan variasi yang sangat besar di dalam penampilan. Walau bagaimanapun, MSER tradisional mengesan titik minat yang sangat sedikit pada imej makanan tanpa tekstur. Oleh itu, algoritma Pengesanan Wilayah Extremal (ERD) di dalam MSER telah ditambah baik dengan menentukan konfigurasi parameter MSER yang optimum, yang membolehkan kuantiti mata minat untuk imej makanan tertentu ditingkatkan dengan sewajarnya.

Algoritma kedua mengurangkan kuantiti kawasan minat dengan menggunakan algoritma Kawasan Extremum (ERS). Kawasan minat yang banyak tidak menjamin prestasi pengelasan yang cemerlang memandangkan kawasan minat yang berlebihan serta kawasan minat dari imej makanan dengan latar belakang yang kompleks turut dikesan. Akibatnya, usaha pengkomputeran harus digunakan untuk melaksanakan proses pengekodan ciri di dalam model BoF.

Dengan mengurangkan kuantiti kawasan yang minat, kecekapan masa pengekodan ciri dapat ditingkatkan tanpa mengorbankan ketepatan pengelasan. Algoritma ERS dilakukan menggunakan pembelajaran tanpa penyeliaan untuk menentukan maklumat ruangan kawasan minat yang dikesan, untuk menunjukkan sama ada ia berasal dari latar belakang imej, dan dengan itu boleh singkirkan sebagai gangguan.

Di dalam algoritma ketiga, teknik umpukan yang lembut menggunakan pengekodan kabur digunakan untuk mengubah ciri peringkat rendah ke dalam perwakilan ciri tahap tinggi. Pendekatan pengekodan kabur mengamalkan teori set kabur (FST) untuk meminimumkan masalah ketidakpastian dan kemungkinan dalam pengekodan ciri yang timbul daripada pendekatan umpukan keras dan *fisher vector* yang digunakan dalam kajian terdahulu. Masalah ketidakpastian dan kemungkinan telah menyebabkan kekeliruan dalam memberikan perihalan ciri kepada kata-kata visual, dan ia berlaku disebabkan oleh variasi keterampilan antara intra kelas yang tinggi kerana kepelbagaian warna dan tekstur yang tinggi. Dengan mengadaptasi FST, penilaian yang teliti dilakukan dalam setiap tugasan perihalan ciri kepada kata-kata visual, yang diterjemahkan ke dalam nilai keanggotaan yang menandakan perkaitan di dalam tugasan itu.

Kaedah yang dicadangkan telah dinilai menggunakan dua data imej: UECFOOD-100 dan UNICT-FD1200. Prestasi algoritma diukur berdasarkan ketepatan pengelasan, kadar kesilapan, dan ketepatan dan penarikan balik. Kualiti pengesan kawasan minat dinilai berdasarkan kuantiti kawasan minat. Pengelasan dilakukan menggunakan *Support Vector Machine* (SVM) dengan kernel linear. Hasil eksperimen menunjukkan prestasi pengelasan superior di dalam kaedah yang dicadangkan berbanding kaedah sebelumnya. Secara khusus, kaedah yang dicadangkan mencapai ketepatan pengelasan 99.62% dan 100.00% pada dataset UECFOOD-100 dan UNICT-FD1200, manakala kaedah sebelumnya hanya dapat mencapai 79.20% dan 85.01% pada dataset yang sama.

Secara keseluruhannya, kaedah yang dicadangkan menghasilkan kamus visual yang kompak dan diskriminatif untuk pengecaman makanan menggunakan hanya satu jenis ciri, bilangan kawasan minat kecil, dan vektor ciri rendah dimensi. Tambahan pula, ia memberikan perwakilan ciri holistik yang dapat memberikan prestasi pengelasan yang luar biasa terhadap makanan dengan variasi penampilan yang hebat.

ACKNOWLEDGEMENTS

In the name of Allah, the most beneficent, the most merciful.

Firstly, I would like to express my deep and sincere gratitude to my research supervisor Dr.Noridayu Manshor for providing invaluable guidance throughout this research. Her dynamisms, sincerity and motivation have deeply inspired me.

Besides, I would like to thank the co-supervisors: Dr.Alfian Abdul Halin, Associate Professor Dr.Razali Yaakub and Associate Professor Datin Dr. Norwati Mustapha for a thorough comments in the research progress meeting, and reviewing the articles and thesis.

I thank my fellow members in UPM especially the merepek meraban members: Dr.Shah, Sufri, Dr.Raihani, Nizam, Dr.Rafiez, Kak Ana, Dr. Hazrina, Kak Nani, Arzila and Liana, from the stimulating academic discussions to emotional and motivational supports, advices and for all the fun and joys : *makan-makan* and *jalan-jalan* we have in the last four years. Also, I thank my friends in Universiti Malaysia Sabah.

Finally, to my loving and supportive wife, Rozita and the kids. Your encouragement when the times got rough are much appreciated, to all my family, the parents for supporting me spiritually throughout study and my life in general.

This thesis was submitted to the Senate of Universiti Putra Malaysia and has been accepted as fulfilment of the requirement for the degree of Doctor of Philosophy. The members of the Supervisory Committee were as follows:

Noridayu Manshor, PhD
Senior Lecturer
Faculty of Computer Science and Information Technology
Universiti Putra Malaysia
(Chairman)

Datin Norwati Mustapha, PhD
Associate Professor
Faculty of Computer Science and Information Technology
Universiti Putra Malaysia
(Member)

Razali Yaakob, PhD
Associate Professor
Faculty of Computer Science and Information Technology
Universiti Putra Malaysia
(Member)

Alfian Abdul Halin, PhD
Senior Lecturer
Faculty of Computer Science and Information Technology
Universiti Putra Malaysia
(Member)

ROBIAH BINTI YUNUS, PhD
Professor and Dean
School of Graduate Studies
Universiti Putra Malaysia

Date: 17 October 2019

Declaration by graduate student

I hereby confirm that:

- this thesis is my original work;
- quotations, illustrations and citations have been duly referenced;
- this thesis has not been submitted previously or concurrently for any other degree at any other institutions;
- intellectual property from the thesis and copyright of thesis are fully-owned by Universiti Putra Malaysia, as according to the Universiti Putra Malaysia (Research) Rules 2012;
- written permission must be obtained from supervisor and the office of Deputy Vice-Chancellor (Research and Innovation) before thesis is published (in the form of written, printed or in electronic form) including books, journals, modules, proceedings, popular writings, seminar papers, manuscripts, posters, reports, lecture notes, learning modules or any other materials as stated in the Universiti Putra Malaysia (Research) Rules 2012;
- there is no plagiarism or data falsification/fabrication in the thesis, and scholarly integrity is upheld as according to the Universiti Putra Malaysia (Graduate Studies) Rules 2003 (Revision 2012-2013) and the Universiti Putra Malaysia (Research) Rules 2012. The thesis has undergone plagiarism detection software.

Signature: ________________________ Date: __________________

Name and Matric No.: Mohd Norhisham bin Razali, GS43882

Declaration by Members of Supervisory Committee

This is to confirm that:

- the research conducted and the writing of this thesis was under our supervision;
- supervision responsibilities as stated in the Universiti Putra Malaysia (Graduate Studies) Rules 2003 (Revision 2012-2013) are adhered to.

Signature: ______________________
Name of Chairman of Supervisory Committee: Dr. Noridayu Manshor

Signature: ______________________
Name of Member of Supervisory Committee: Associate Professor Datin Dr. Norwati Mustapha

Signature: ______________________
Name of Member of Supervisory Committee: Associate Professor Dr. Razali Yaakob

Signature: ______________________
Name of Member of Supervisory Committee: Dr. Alfian Abdul Halin

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

BGP	Binary Gabor Pattern
BoF	Bag of Features
CR	Classification Rate
DCNN	Deep Convolutional Neural Network
DoG	Difference of Gaussians
ER	Extremal Region
ERS	Extremal Region Selection
ERT	Error Rate
FAST	Features from Accelerated Segment Test
FCM	Fuzzy C-Means
FST	Fuzzy Set Theory
FV	Fisher Vector
GFD	Gabor-based image decomposition, and fractal dimension estimation
GMM	Gaussian Mixture Model
GSURF	Gauge Speeded Up Robust Feature
HOG	Histogram of Gradients
HSV	Hue Saturation Value
ITV	Intensity Threshold Value
LoG	Laplacian of Gaussian
LBP	Local Binary Pattern
LCP	Local Correlation Pattern
LPQ	Local Phase Quantization
MAV	Maximum Area Variation

MKL	Multiple Kernel Learning
MSER	Maximally Stable Extremal Region
OF	Occurrence Frequency
PCA	Principal Component Analysis
PDD	Probability Density Distribution
PDF	Probability Density Function
PHOG	Pyramid Histogram of Gradient
Prec	Precision
PRICoLBP	Pairwise Rotation Invariant Co-Occurrence Local Binary Pattern
Rec	Recall
RGB	Red Green Blue
RFDg	Gaussian receptive fields descriptor
SIFT	Scale Invariant Feature Transform
SURF	Speeded Up Robust Feature
SVM	Support Vector Machine
PFI	Pittsburgh Fast-food Image

CHAPTER 1

INTRODUCTION

1.1 Research Background

The long-standing goal of visual object recognition is to develop algorithms able to distinguish objects from one another in real-world scenes and cluttered environments. Object recognition is important enough to specialise into various domains such as face recognition, pedestrian recognition, vehicle recognition, and many others. Image processing and machine learning are the heart of most tasks in object recognition. Recently, technological advancements in smartphone price and image quality, as well as explosive growth in the number of images on social media has attracted more interest from researchers to further explore this exciting research field. Social media users are more inclined to use images instead of relying only on textual content to share their activities and interactions. Images provide a highly-expressive medium and are easy to capture, store, and share (Song et al., 2015).

The advancement of mobile technology at a reasonable cost has allowed people to photograph their food intake and to share their excitement when having a meal on social media. This indulgence has become a worldwide phenomenon (Rich et al., 2016). Food recognition is an emerging research area in object recognition which has grown substantially in the era of the smartphones and social media services (Kagaya & Aizawa, 2015; Xu et al., 2015). Food recognition provides automatic identification of the category of foods from an image and can estimate the caloric and nutritional content in order to assist dietary assessment in treating diet-related chronic diseases. Diet-related chronic diseases such as diabetes, hypertension, and heart disease are strongly linked with obesity and are caused by an imbalanced nutritional intake and a lack of physical activity. This epidemic has serious consequences worldwide with 1.9 billion adults categorised as overweight and 650 million of these classed as obese (WHO, 2018). Food recognition to provide self-nutrition services is important to prevent and overcome the obesity problem. A daily record of food intake via dietary assessment may provide a measure of caloric and nutrient intake, allowing personalised diet and food intake balance information. However, traditional dietary assessment has been reported to be less accurate in measuring the amount of food consumed due to the under-reporting of food intake (Coulston et al., 2013). Hence, there is demand for novel tools able to provide an automatic, personalised, and accurate dietary assessment through food recognition algorithms (Anthimopoulos et al., 2013).

In general, food recognition is a challenging task (Zhang et al., 2013) due mainly to very small inter-class similarities which make make foods from different categories look identical, and large intra-class differences of food objects which make foods in the same category look different. The natural appearance of food objects is complex as they have deformable structure with large variations. Furthermore, state-of-the-art object recognition methods are not necessarily robust enough for food recognition (Farinella et al., 2016; Kong et al., 2015; Min et al., 2018). Hence, it has become highly challenging to achieve noteworthy food recognition performance in real-world applications.

The various types of real-world foods are highly diverse and thus are grouped into many categories. Indeed, the recent focus in object recognition research is to improve recognition performance on large-scale object categories (Li et al., 2016), which can be performed by extracting image features to represent the unique visual characteristics of the respective food categories. Ideally, features that can maximise inter-class discrimination and minimise intra-class robustness are sought (Bosch et al., 2011). Local feature-based representations that identify interest points from images are an effective technique in describing features suitable for the complex appearance of food. Such representations have the capability to capture more detailed properties of food images and are robust towards variations in illumination, scale, and orientation (Kong et al., 2015; Zhu et al., 2015).

However, local features are low-level and have little semantic content since the descriptions they yield are too large and are highly diverse. Thus, local features need to be encoded into higher-level representations before machine learning algorithms can be employed for the classification of food categories. Commonly, local features are encoded using a Bag of Features model (BoF) (Csurka et al., 2004; Huang & Tan, 2014) which consists of three main stages: local feature extraction, local feature encoding, and classification.

In previous research, local features have been described using Scale Invariant Feature Transform (SIFT) (Giovany et al., 2017; Kong et al., 2015; Zheng, Wang, & Zhu, 2017; Zhu et al., 2015), Histogram of Gradients (HOG) (Kawano & Yanai, 2015), Speeded-Up Robust Features (SURF) (Pooja & Madival, 2016; Zhu et al., 2015), colour (Kawano & Yanai, 2015; Martinel, Piciarelli, & Micheloni, 2016; Zheng et al., 2017), and texture (Farinella et al., 2016). Dense sampling and Difference of Gaussians (DoG) are the two most common methods of sampling interest points used in previous studies in food recognition (Kawano & Yanai, 2015; Martinel et al., 2016; Sasano et al., 2016).

Local feature extraction generates millions of interest points describing the features that need to be transformed into a simpler form via feature encoding before they can be fed into a machine learning classifier. Commonly, local features are encoded using the k-means algorithm, a hard assignment approach, to generate visual words, and then BoF is constructed by counting the number of interest points assigned to each cluster center, or centroid (Farinella et al., 2016; Kong et al., 2015). In a more recent approach, local features are encoded by using a technique known as Fisher Vector (FV) (Kawano, 2015; Zheng et al., 2017) to overcome issues of efficiency and encoding error found in hard assignment approaches. In FV, a Gaussian Mixture Model (GMM) is used to generate visual words which are further processed to obtain a final feature representation. Since they have proven to be successful in image classification, Support Vector Machines (SVMs) with a linear kernel have frequently been used for classification in food recognition (Cui et al., 2015). Furthermore, the sparse features generated by local features are also more compatible with, and separable by, a linear kernel in SVM (Fan, Wang, & Lin, 2015; J. Yang, Yu, Gong, & Beckman, 2009).

There is always room to improve the existing local-feature-based representations in providing a holistic and compact representation that can also handle the enormous diversity in food appearance. Notably, certain food categories have proved challenging to classify at anything higher than average accuracy. This is probably due to the lack of discriminative features or insufficient information when a small number of interest points have been detected. Despite the merit of local features in localising interest points on a target object, they often struggle with images that have a background with more visible texture and higher contrast than the foreground. In this context, local segmentation is not the best option to rectify this problem as food objects have an arbitrary shape, non-homogenous structure, and fewer visible edges (Zhu et al., 2015).

Existing feature encoding techniques based on hard assignment and FV do not handle the problem of uncertainty and plausibility in constructing visual words. Uncertainty and plausibility are problems in food recognition because food appearance is highly variable, leading to the occurrence of error or information loss in constructing a visual dictionary.

The aim of this study is to propose a food recognition algorithm that can cope with a highly diverse set of foods, regardless of their appearance. This is to achieved by enhancing the capability of a recognition algorithm by improving interest points detection and feature encoding in BoF representation. In addition to improving recognition performance on food categories, a more compact local feature representation will result from the interest point selection procedure, without losing classification accuracy. This is followed by the improvement of feature encoding via fuzzy set theory (FST) approach, which can successfully reduce the effects of uncertainty and implausibility, and to produce a highly discriminative visual dictionary.

1.2 Research Problems

MSER detector provide a good alternative interest points sampling for foods. Nevertheless, food categories with texture-less surfaces and low contrast images are poorly recognised due to the low number of interest points detected (Ma et al., 2017; Takeishi et al., 2015). Indeed, sparse interest points such as MSER are one of the drawbacks of local feature detectors as they tend to detect denser features on textured surfaces as compared to texture-less surfaces (Krig, 2014; Anthimopoulos et al., 2014). The density of interest points detected by using MSER can be increased by configuring certain parameter values (Takeishi et al., 2015). However, the parameter configuration will resulted to a significant increase of unnecessary interest points number as well that may increase the computations time.

Inevitably, features will be extracted from irrelevant interest points (i.e from the background, especially if it is complex) (Altintakan et al. 2015) and will generate less informative descriptions regardless of the sampling techniques being used. Interest region-based detectors using Maximally Stable Extremal Region (MSER) use global segmentation and take into account regions from images with complex backgrounds as well. In fact, detectors based on DoG also unavoidably detect interest points within complex and noisy backgrounds (Yu et al., 2013). Furthermore, the number of interest points is still very high for real-time applications and the irrelevant interest points increase the computational cost of the feature encoding process (Lin et al., 2016; Mukherjee et al., 2016; Xu et al., 2015).

In the many food recognition studies, hard assignment strategy using k-means (Farinella et al., 2016; Giovany et al., 2017; Hassannejad et al., 2017; Martinel, Piciarelli, & Micheloni, 2016) and Fisher Vector (FV) (Yoshiyuki Kawano, 2015; Zheng et al., 2017) were used generate visual words is popular due to its simplicity and efficiency. There are two main problems when using hard assignment and FV: visual word uncertainty and visual word plausibility (Umit L. Altintakan & Yazici, 2015). Visual word uncertainty is the condition of assigning a feature description to one visual word without evaluating the other visual words that may be more suitable. Visual word plausibility is the condition of assigning a feature description to a visual word even though none of the visual words are suitable. These problems of uncertainty and plausibility are of concern for food images due to their huge variety of appearance (Ge et al., 2013). This problem has also been mentioned in (Kong et al., 2015; Martinel et al., 2016; Pouladzadeh et al., 2014) where food images are themselves the source of the uncertainty problem due to their visual characteristics of color and texture and the irregularities in appearance of foods images that include high deformation, complex backgrounds, and high intra-class variations.

Figure 1.1 shows three examples images from the food category sushi in the UECFOOD-100 dataset (Kawano & Yanai, 2015) which demonstrate the uncertainty and plausibility problems in generating a visual dictionary using hard assignment.

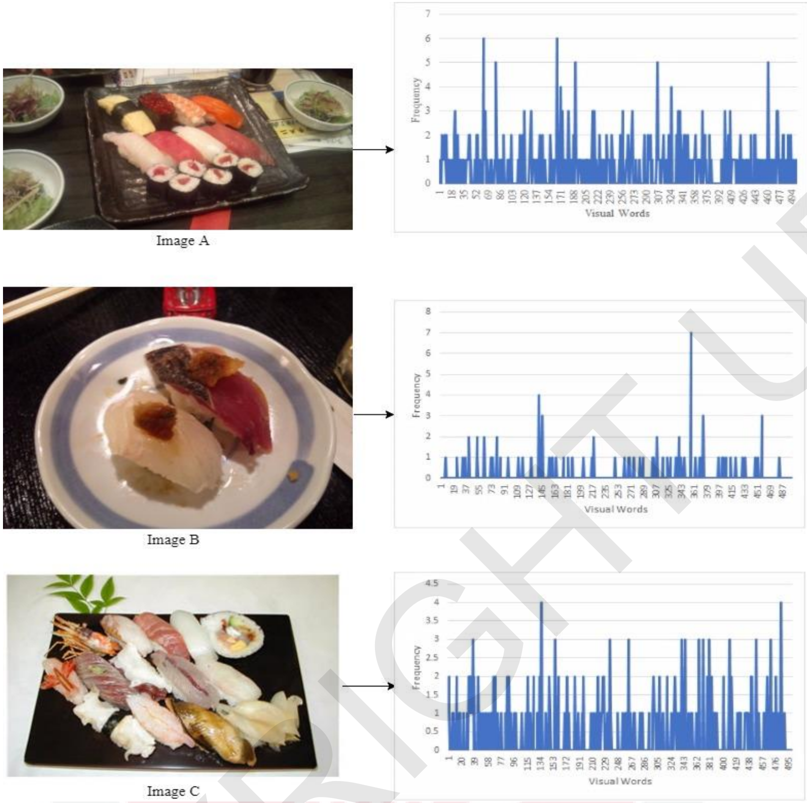

Figure 1.1 : Examples of the uncertainty and plausibility problems in generating a visual dictionary using hard assignment

As shown in Figure 1.1, images A, B, and C are from the same category but have a very different appearance. Consequently, different patterns in the visual dictionary are produced as the visual words are generated solely by the respective image without any comparisons with other images. The uncertainty and plausibility problems (as well as differing numbers of interest points between the images) have resulted in large intra-class variation. Similarly, the same problems of uncertainty and plausibility are present in FV, as shown in Figure 1.2.

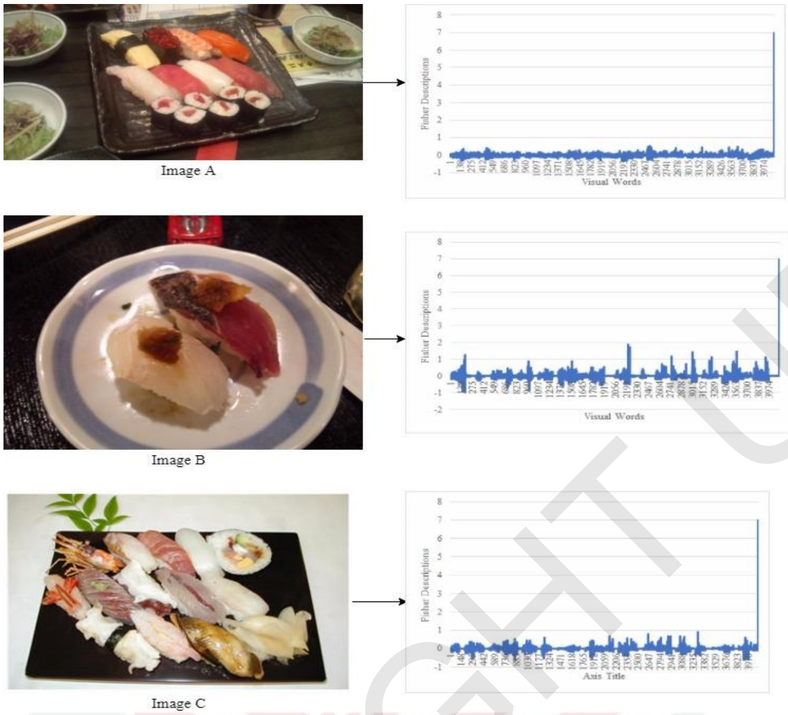

Figure 1.2 : Examples of the uncertainty and plausibility problems in generating a visual dictionary using Fisher Vector

According to Figure 1.2, three different patterns of feature representation are generated using FV on three sushi images A, B, and C. In addition, the computation in FV produces an extremely long feature vector with a high computational cost for classification, making it less suitable for large-scale applications (Wang & Deng, 2015; Xie et al., 2016).

1.3 Research Objectives

This study aims to improve the performance in food category recognition. To achieve this, the objectives are defined as follows:

1. To propose an Extremal Region Detection (ERD) technique in MSER to increase the density of interest points detection on foods with little texture.The sub-objectives to achieve this objective are as follows:

   a. To evaluate and choose the optimum parameters configuration in MSER in order to increase the number of extremal regions.
   b. To evaluate the optimal number of extremal regions of food images that required the MSER parameter configuration.

2. To propose an Extremal Region Selection (ERS) in MSER to reduce the quantity of interest regions. The sub-objectives to achieve this objective are as follows:

   a. To evaluate the optimal number of extremal regions that required extremal regions selection.
   b. To evaluate the optimum parameters used in ERS.

3. To devise fuzzy encoding technique to reduce the effects of uncertainty and plausibility in constructing visual dictionary.

1.4 Research Scope

The scope of this study is defined in the following:

1. This study focuses on the recognition tasks that consist of two main stages which are feature representation and classification to identify the category of foods.

2. This study emphasises the feature representation aspect within the Bag of Features (BoF) model which focuses on the aspects of interest point detection, feature description and the feature encoding process.

3. The classification task of the proposed methods are benchmarked using a Support Vector Machine with a linear kernel and by adopting a one-versus-rest training strategy.

4. The performance evaluation of the proposed methods is conducted primarily on classification performance in both individual and overall food categories. The interest point detection performance is measured based on the number of interest points produced. The length of feature vector produced by feature encoding is used to measure the performance of feature encoding techniques.

5. The datasets used in this study are food images in real-world settings obtained from the World Wide Web (Farinella et al., 2016; Kawano & Yanai, 2015). The labels represent the food category, and only a single label is considered in this research even if food images consist of multiple food categories.

1.5 Research Contributions

The main contribution of this study is the improvement of Bag of Feature (BoF) model in food recognition performance via Extremal Region Detection (ERD), Extremal Region Selection (ERS), and fuzzy encoding. This study has made six individual contributions. The first two concern extremal region detection technique. The third contribution is related to the extremal region selection technique. The fourth contribution is on the overall BoF model and the last three contributions pertain to the feature encoding technique. Explanations of each contribution are as follows:

1. This study has improved recognition performance on foods with a strong mixture of ingredients (such as pizza, raisin bread, sirloin cutlets) by adopting an interest region detector based on Maximally Stable Extremal Region (MSER) and Speeded Up Robust Feature (SURF) descriptor (MSER-SURF). Remarkably, MSER-SURF has the lowest computational complexity for both interest region detection and description. This is due to the low number of interest regions generated by MSER and also the shorter length of the feature vectors yielded by the SURF descriptor.

2. By evaluating the MSER parameters, a thresholding mechanism which is to determine suitable threshold has been used that has increased the density of interest regions on food images with smooth or less diverse textures. This mechanism also provides an alternative way to exclude the pre-processing task of increasing the brightness and contrast on dull food images as well as resizing very small images in order to detect an appropriate number of interest regions.

3. The extremal region selection procedure aims to reduce the amount of redundant and noisy interest regions in MSER, especially those that have been detected in the image background. Therefore, a compact, informative, and discriminative set of interest regions was generated.

4. This study has simplified the stages in the food recognition process since no image segmentation or local feature dimensionality selection has been performed. Even though an extremal region detection and selection procedure has been introduced, the computational complexity of extraction was not affected significantly.

5. This study has incorporated fuzzy set theory (FST) to the construction of a visual dictionary in order to minimise the problems of uncertainty and plausibility posed by the high variability and high intra-class differences of food images. As a result, visual word assignment during feature

encoding is performed in a clear and concise manner that decreases sparsity as well as promoting greater discriminability of visual words.

6. Evaluation of the effect of vocabulary sizes in feature encoding shows that the proposed fuzzy encoding technique can construct a visual dictionary capable of outstanding recognition accuracy using a smaller vocabulary size than traditional hard assignment. Also, this technique does not increase feature vector length with vocabulary size as is the case for Fisher Vector (FV). The computation of FV, even when using even a small vocabulary size, will vastly increase the feature dimensions.

1.6 Organisation of the Thesis

Chapter 2 provides a literature review on food recognition and its associated techniques in object recognition. This is followed by Chapter 3 which describes the overall methodology undertaken for this study. Chapter 4 presents the proposed framework, elaborating the improved algorithms for food recognition. Chapter 5 presents experimental results and discussion. Chapter 6 concludes this study with remarks about the achievements made and possible future research.

REFERENCES

Abbirami, R. S., Abhinaya, A., Kavivarthini, P., & Rupika, T. (2015). Large Scale Learning for Food Image Classification. *Journal on Recent and Innovation Trends in Computing and Communication*, *3*(March), 973–978.

Ahmed, A. (2015). Food Image Recognition by Using Bag-of-SURF Features and HOG Features. In *Proceedings of the 3rd International Conference on Human-Agent Interaction* (pp. 179–180).

Aizawa, K., Maruyama, Y., Li, H., & Morikawa, C. (2013). Food Balance Estimation by Using Personal Dietary Tendencies in a Multimedia Food Log. *IEEE Transaction on Multimedia*, *15*(8), 2176–2185.

Altintakan, U L, & Yazici, A. (2016). A novel fuzzy feature encoding approach for image classification. *2016 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*, 1134–1139. https://doi.org/10.1109/FUZZ-IEEE.2016.7737815

Altintakan, Umit L., & Yazici, A. (2015). An improved BOW approach using fuzzy feature encoding and visual-word weighting. *IEEE International Conference on Fuzzy Systems*, *2015-Novem*(114). https://doi.org/10.1109/FUZZ-IEEE.2015.7338108

Altintakan, Umit Lutfu, & Yazici, A. (2015). Towards effective image classification using class-specific codebooks and distinctive local features. *IEEE Transactions on Multimedia*, *17*(3), 323–332. https://doi.org/10.1109/TMM.2014.2388312

Anthimopoulos, M., Dehais, J., Diem, P., & Mougiakakou, S. (2013). Segmentation and recognition of multi-food meal images for carbohydrate counting. *13th IEEE International Conference on BioInformatics and BioEngineering, IEEE BIBE 2013*, 1–4. https://doi.org/10.1109/BIBE.2013.6701608

Ao, S., & Ling, C. X. (2015). Adapting New Categories for Food Recognition with Deep Representation. *2015 IEEE International Conference on Data Mining Workshop (ICDMW)*, 1196–1203. https://doi.org/10.1109/ICDMW.2015.203

Barbara, S., Yang, X., & Gibson, J. (2013). *Distinctive and Efficient Local Features for Real-Time Mobile Applications*.

Bay, H., Ess, A., Tuytelaars, T., & Van Gool, L. (2008). Speeded-Up Robust Features (SURF). *Computer Vision and Image Understanding*, *110*(3), 346–359. https://doi.org/10.1016/j.cviu.2007.09.014

Bay, H., Tuytelaars, T., & Van Gool, L. (2006). SURF: Speeded up robust features. *Lecture Notes in Computer Science (Including Subseries Lecture*

Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), *3951 LNCS*, 404–417. https://doi.org/10.1007/11744023_32

Bosch, M., Zhu, F., Khanna, N., Boushey, C. J., & Delp, E. J. (2011). Combining Global and Local Features for Food Identification in Dietary Assessment. In *IEEE International Conference on Image Processing* (pp. 1789–1792).

Bucak, S. S., Jin, R., & Jain, A. (2014). Multiple Kernel Learning for Visual Object Recognition: A Review. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *36*(7), 1354–1369. https://doi.org/10.1109/TPAMI.2013.212

Buoncompagni, S., Maio, D., Maltoni, D., & Papi, S. (2015). Saliency-based keypoint selection for fast object detection and matching. *Pattern Recognition Letters*, *62*, 32–40. https://doi.org/10.1016/j.patrec.2015.04.019

Chatfield, K., Lempitsky, V., Vedaldi, A., & Zisserman, A. (2011). The devil is in the details : an evaluation of recent feature encoding methods. In *Proceedings of the British Machine Vision Conference* (pp. 1–12).

Chen, Q., Zhang, C., Zhao, J., & Ouyang, Q. (2013). Recent advances in emerging imaging techniques for non-destructive detection of food quality and safety. *TrAC Trends in Analytical Chemistry*, *52*, 261–274. https://doi.org/10.1016/j.trac.2013.09.007

Chen, Shuangshuang, Liu, H., Zeng, X., Qian, S., Wei, W., Wu, G., & Duan, B. (2018). Local patch vectors encoded by fisher vectors for image classification. *Information (Switzerland)*, *9*(2), 1–12. https://doi.org/10.3390/info9020038

Chen, Shuhan, Shi, W., & Lv, X. (2015). Feature coding for image classification combining global saliency and local difference. *Pattern Recognition Letters*, *51*, 44–49. https://doi.org/10.1016/j.patrec.2014.08.008

Chien, H.-J., Chuang, C.-C., Chen, C.-Y., & Klette, R. (2016). When to use what feature? SIFT, SURF, ORB, or A-KAZE features for monocular visual odometry. *2016 International Conference on Image and Vision Computing New Zealand (IVCNZ)*, (1), 1–6. https://doi.org/10.1109/IVCNZ.2016.7804434

Chougrad, H., Zouaki, H., & Alheyane, O. (2015). Soft assignment vs hard assignment coding for bag of visual words. *2015 10th International Conference on Intelligent Systems: Theories and Applications, SITA 2015*. https://doi.org/10.1109/SITA.2015.7358422

Ciocca, G., B, P. N., & Schettini, R. (2015). Food Recognition and Leftover Estimation for Daily Diet Monitoring. In *International Conference on Image Analysis and Processing* (Vol. 9281, pp. 334–341). https://doi.org/10.1007/978-3-319-23222-5

Coulston, A. M., Boushey, C. j., & Ferruzzi, G., M. (2013). Nutrition in the

Prevention and Treatment of Disease. In *Nutrition in the Prevention and Treatment of Disease* (3rd ed., pp. 5–30). Academic Press.

Csurka, G., Dance, C., Fan, L., Willamowski, J., & Cedric Bray. (2004). Visual categorization with bag of keypoints. *International Workshop on Statistical Learning in Computer Vision*, 1–22. https://doi.org/10.1234/12345678

Cui, J., Cui, M., Xiao, B., & Li, G. (2015). Compact and discriminative representation of Bag-of-Features. *Neurocomputing*, *169*, 55–67. https://doi.org/10.1016/j.neucom.2014.09.104

De Choudhury, M., Sharma, S., & Kiciman, E. (2016). Characterizing Dietary Choices, Nutrition, and Language in Food Deserts via Social Media. In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing* (pp. 1157–1170). https://doi.org/10.1145/2818048.2819956

Dell'Agnello, D., Carneiro, G., Chin, T.-J., Castellano, G., & Fanelli, A. M. (2013). Fuzzy clustering based encoding for visual object classification. *Proc. of the 2013 IFSA World Congress - NAFIPS Annual Meeting*, 1439–1444. https://doi.org/10.1109/IFSA-NAFIPS.2013.6608613

Deng, L., & Yu, D. (2014). *Deep Learning: Methods and Application*. NOW Publisher.

Donoser, M., Riemenschneider, H., & Bischof, H. (2010). Shape Guided Maximally Stable Extremal Region (MSER) Tracking. In *International Conference on Pattern Recognition* (pp. 1800–1803). https://doi.org/10.1109/ICPR.2010.444

Du, C., & Sun, D. (2006). Learning techniques used in computer vision for food quality evaluation : a review. *Journal of Food Engineering*, *72*(1), 39–55. https://doi.org/10.1016/j.jfoodeng.2004.11.017

Dubey, A. K., & Jaffery, Z. A. (2016). Maximally Stable Extremal Region Marking-Based Railway Track Surface Defect Sensing. *IEEE Sensors Journal*, *16*(24), 9047–9052. https://doi.org/10.1109/JSEN.2016.2615333

Ege, T., & Yanai, K. (2018). Image-Based Food Calorie Estimation Using Recipe Information. *Special Section on Machine Vision and Its Applications*, (5), 1333–1341.

Escalante, H. J., Ponce-López, V., Escalera, S., Baró, X., Morales-Reyes, A., & Martínez-Carranza, J. (2016). Evolving weighting schemes for the Bag of Visual Words. *Neural Computing and Applications*, 1–15. https://doi.org/10.1007/s00521-016-2223-x

Fan, R., Wang, X., & Lin, C. (2015). LIBLINEAR : A Library for Large Linear Classification. *Journal of Machine Learning Research*, *9*(2008), 1871–1874.

Fanyu Kong, J. T. (2011). DietCam: Regular Shape Food Recognition with a Camera Phone. In *International Conference on Body Sensor Networks*. IEEE. https://doi.org/10.1109/BSN.2011.19

Faria, F. A., Alex, J., Rocha, A., & Torres, R. S. (2012). Automatic Classifier Fusion for Produce Recognition. In *25th SIBGRAPI Conference on Graphics, Patterns and Images*. https://doi.org/10.1109/SIBGRAPI.2012.42

Farinella, G. M., Allegra, D., Moltisanti, M., Stanco, F., & Battiato, S. (2016). Retrieval and classification of food images. *Computers in Biology and Medicine*, *77*, 23–39. https://doi.org/10.1016/j.compbiomed.2016.07.006

Farinella, G. M., Moltisanti, M., & Battiato, S. (2015). Food recognition using consensus vocabularies. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *9281*, 384–392. https://doi.org/10.1007/978-3-319-23222-5_47

Fengqing, Z., Bosch, M., Woo, I., Kim, S., J. Boushey, C., S. Ebert, D., & J. Delp, E. (2010). The Use of Mobile Devices in Aiding Dietary Assessment and Evaluation. *IEEE Journal of Selected Topics in Signal Processing*, *4*(4), 756–766.

Ferraz, C. T., & Gonzaga, A. (2017). *Object classification using a local texture descriptor and a support vector machine*. *Multimedia Tools and Applications* (Vol. 76). Multimedia Tools and Applications. https://doi.org/10.1007/s11042-016-4003-0

Fried, D., Surdeanu, M., Kobourov, S., Hingle, M., & Bell, D. (2015). Analyzing the language of food on social media. *Proceedings - 2014 IEEE International Conference on Big Data, IEEE Big Data 2014*, (Section II), 778–783. https://doi.org/10.1109/BigData.2014.7004305

Gao, H., & Yang, Z. (2011). Integrated Visual Saliency Based Local Feature Selection for Image Retrieval. *Intelligence Information Processing and Trusted Computing (IPTC), 2011 2nd International Symposium On*, 47–50. https://doi.org/10.1109/IPTC.2011.19

Garg, V., Vempati, S., & Jawahar, C. V. (2011). Bag of visual words: A soft clustering based exposition. *Proceedings - 2011 3rd National Conference on Computer Vision, Pattern Recognition, Image Processing and Graphics, NCVPRIPG 2011*, 37–40. https://doi.org/10.1109/NCVPRIPG.2011.15

Ge, T., Ke, Q., & Sun, J. (2013). Sparse-Coded Features for Image Retrieval. *Procedings of the British Machine Vision Conference 2013*, 132.1-132.11. https://doi.org/10.5244/C.27.132

Ghosh, Soumi, & Dubey, S. K. S. (2013). Comparative analysis of k-means and fuzzy c-means algorithms. *International Journal of Advanced Computer Science and Applications*, *4*(4), 35–38.

https://doi.org/10.14569/IJACSA.2013.040406

Ghosh, Soumyadeep, Dhamecha, T. I., Keshari, R., Singh, R., & Vatsa, M. (2015). Feature and keypoint selection for visible to near-infrared face matching. *2015 IEEE 7th International Conference on Biometrics Theory, Applications and Systems, BTAS 2015*. https://doi.org/10.1109/BTAS.2015.7358760

Gil, A., & Hern, M. (2007). Evaluation of interest point detectors for Visual SLAM. In *Current Topics in Artificial Intelligence*. Springer-Verlag Berlin. https://doi.org/10.1007/978-3-540-75271-4_18

Giovany, S., Putra, A., Hariawan, A. S., & Wulandhari, L. A. (2017). Machine Learning and SIFT Approach for Indonesian Food Image. *Procedia Computer Science*, *116*, 612–620. https://doi.org/10.1016/j.procs.2017.10.020

Gosain, A., & Dahiya, S. (2016). Performance Analysis of Various Fuzzy Clustering Algorithms: A Review. *Procedia Computer Science*, *79*, 100–111. https://doi.org/10.1016/j.procs.2016.03.014

Harzallah, H., Jurie, F., & Schmid, C. (2009). Combining efficient object localization and image classification. *Proceedings of the IEEE International Conference on Computer Vision*, 237–244. https://doi.org/10.1109/ICCV.2009.5459257

Hassannejad, H., Matrella, G., Ciampolini, P., De Munari, I., Mordonini, M., & Cagnoni, S. (2016). Food Image Recognition Using Very Deep Convolutional Networks. *Proceedings of the 2nd International Workshop on Multimedia Assisted Dietary Management - MADiMa '16*, 41–49. https://doi.org/10.1145/2986035.2986042

Hassannejad, H., Matrella, G., Ciampolini, P., De Munari, I., Mordonini, M., & Cagnoni, S. (2017). A new approach to image-based estimation of food volume. *Algorithms*, *10*(2). https://doi.org/10.3390/a10020066

He, H., Kong, F., & Tan, J. (2016). DietCam : Multiview Food Recognition Using a Multikernel SVM. *IEEE Journal OF Biomedical and Health Informatics*, *20*(3), 848–855.

Hiba, C., Hamid, Z., & Omar, A. (2016). Bag of Features Model Using the New Approaches: A Comprehensive Study. *International Journal of Advanced Computer Science & Applications*, *1*(7), 226–234.

Hoashi, H., Joutou, T., & Yanai, K. (2010). Image Recognition of 85 Food Categories by Feature Fusion. In *IEEE International Symposium on Multimedia*. https://doi.org/10.1109/ISM.2010.51

Hosang, J., Benenson, R., Dollár, P., & Schiele, B. (2015). What makes for effective detection proposals? *Computer Vision and Pattern Recognition*, 2014. https://doi.org/10.1109/TPAMI.2015.2465908

Hu, Y.-T., Lin, Y.-Y., Chen, H.-Y., Hsu, K.-J., & Chen, B.-Y. (2014). Descriptor Ensemble: An Unsupervised Approach to Descriptor Fusion in the Homography Space. *Computer Vision and Pattern Recognition*.

Huang, Y., Huang, K., Yu, Y., & Tan, T. (2011). Salient coding for image classification. *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference On*, 1753–1760. https://doi.org/10.1109/CVPR.2011.5995682

Huang, Y., & Tan, T. (2014). *Feature Coding for Image Representation and Recognition*. Springer Heidelberg. https://doi.org/10.1007/978-3-662-45000-0

Huang, Y., Wu, Z., Wang, L., & Tan, T. (2014). Feature coding in image classification: A comprehensive study. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *36*(3), 493–506. https://doi.org/10.1109/TPAMI.2013.113

Ionescu, B., Benois-Pineau, J., Piatrik, T., & Quenot, G. (2014). Fusion in Computer Vision. *Advances in Computer Vision and Pattern Recognition*, 272. https://doi.org/10.1007/978-3-319-05696-8

Jabeen, S., Mehmood, Z., Mahmood, T., Saba, T., Rehman, A., & Mahmood, M. T. (2018). An effective content-based image retrieval technique for image visuals representation based on the bag-of-visual-words model. *PLoS ONE*, 1–24.

Jiang, Y., Yang, J., Ngo, C., & Hauptmann, A. G. (2010). Representations of Keypoint-Based Semantic Concept Detection : A Comprehensive Study Representations of Keypoint-Based Semantic Concept Detection : A Comprehensive Study. *IEEE Transaction on Multimedia*, *12*(1), 42–53.

Joutou, T., & Yanai, K. (2009). A food image recognition system with Multiple Kernel Learning. In *2009 16th IEEE International Conference on Image Processing (ICIP)* (pp. 285–288). https://doi.org/10.1109/ICIP.2009.5413400

Kagaya, H. (2014). Food Detection and Recognition Using Convolutional Neural Network. In *Proceedings of the 22nd ACM international conference on Multimedia* (pp. 1085–1088).

Kagaya, H., & Aizawa, K. (2015). Highly Accurate Food/Non-Food Image Classification Based on a Deep Convolutional Neural Network. In *International Conference on Image Analysis and Processing* (Vol. 9281, pp. 350–357). https://doi.org/10.1007/978-3-319-23222-5

Kawano, Y., & Yanai, K. (2014). FoodCam : A Real-Time Mobile Food Recognition System Employing Fisher Vector. In *International Conference on Multimedia Modeling* (pp. 369–373).

Kawano, Y., & Yanai, K. (2015). FoodCam: A real-time food recognition system on a smartphone. *Multimedia Tools and Applications*, *74*(14), 5263–5287. https://doi.org/10.1007/s11042-014-2000-8

Keypoints, S., & Lowe, D. G. (2004). Distinctive Image Features from. *International Journal of Computer Vision*, *60*(2), 91–110. https://doi.org/10.1023/B:VISI.0000029664.99615.94

Kibria, S. B. (2017). An analysis of Feature extraction and Classification Algorithms for Dangerous Object Detection. *2017 2nd International Conference on Electrical & Electronic Engineering (ICEEE)*, 1–4. https://doi.org/10.1109/CEEE.2017.8412846

Knez, S., & Sajn, L. (2015). Food Object Recognition Using a Mobile Device: State of the Art. In *International Conference on Image Analysis and Processing* (Vol. 9281, pp. 366–374). https://doi.org/10.1007/978-3-319-23222-5

Kong, F. (2012). *Automatic Food Intake Assessment Using Camera Phones*.

Kong, F., He, H., A. Raynor, H., & Tan, J. (2015). DietCam: Multi-view regular shape food recognition with a camera phone. *Pervasive and Mobile Computing*, *19*(C), 108–121.

Kong, F., & Tan, J. (2012). DietCam: Automatic dietary assessment with mobile camera phones. *Pervasive and Mobile Computing*, *8*(1), 147–163.

Krig, S. (2014a). Interest Point Detector and Feature Descriptor Survey. In *Computer Vision Metrics* (pp. 217–282). Apress, Berkeley, CA.

Krig, S. (2014b). Local Feature Design Concepts, Classification, and Learning. *Computer Vision Metrics*, 131–189. https://doi.org/10.1007/978-1-4302-5930-5_4

Krizhevsky, A., Sulskever, Ii., & Hinton, G. E. (2012). ImageNet Classification with Deep Convolutional Neural Networks. *Advances in Neural Information and Processing Systems (NIPS)*, 1–9.

Lee, M. H., & Park, I. K. (2017). Performance evaluation of local descriptors for maximally stable extremal regions. *Journal of Visual Communication and Image Representation*, *47*, 62–72. https://doi.org/10.1016/j.jvcir.2017.05.008

Li, K., Wang, F., & Zhang, L. (2016). A new algorithm for image recognition and classification based on improved Bag of Features algorithm. *Optik - International Journal for Light and Electron Optics*, *127*(11), 4736–4740. https://doi.org/http://dx.doi.org/10.1016/j.ijleo.2015.08.219

Li, W., Dong, P., Xiao, B., & Zhou, L. (2016). Object recognition based on the Region of Interest and optimal Bag of Words model. *Neurocomputing*, *172*, 271–280. https://doi.org/10.1016/j.neucom.2015.01.083

Li, Yali, Wang, S., Tian, Q., & Ding, X. (2015a). A survey of recent advances in visual feature detection. *Neurocomputing*, *149*(PB), 736–751. https://doi.org/10.1016/j.neucom.2014.08.003

Li, Yali, Wang, S., Tian, Q., & Ding, X. (2015b). Feature representation for statistical-learning-based object detection: A review. *Pattern Recognition*, *48*(11), 3542–3559. https://doi.org/10.1016/j.patcog.2015.04.018

Li, Yanshan, Liu, W., Huang, Q., & Li, X. (2016). Fuzzy bag of words for social image description. *Multimedia Tools and Applications*, *75*(3), 1371–1390. https://doi.org/10.1007/s11042-014-2138-4

Lin, G., Fan, C., Zhu, H., Miu, Y., & Kang, X. (2017). Visual feature coding based on heterogeneous structure fusion for image classification. *Information Fusion*, *36*, 275–283. https://doi.org/10.1016/j.inffus.2016.12.010

Lin, W., Tsai, C., Chen, Z., & Ke, S. (2016). Keypoint selection for efficient bag-of-words feature generation and effective image classification. *Information Sciences*, *329*, 33–51. https://doi.org/10.1016/j.ins.2015.08.021

Lisin, D. A., Mattar, M. A., Blaschko, M. B., Learned-Miller, E. G., & Benfield, M. C. (2005). Combining Local and Global Image Features for Object Class Recognition. *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05) - Workshops*, *3*, 47–47. https://doi.org/10.1109/CVPR.2005.433

Liu, J., Huang, Y., Wang, L., & Wu, S. (2014). Hierarchical feature coding for image classification. *Neurocomputing*, *144*, 509–515. https://doi.org/10.1016/j.neucom.2014.04.022

Liu, L., Wang, L., & Liu, X. (2011). In defense of soft-assignment coding. *Proceedings of the IEEE International Conference on Computer Vision*, 2486–2493. https://doi.org/10.1109/ICCV.2011.6126534

Liu, W. X., Hou, J., & Karimi, H. R. (2014). Research on vocabulary sizes and codebook universality. *Abstract and Applied Analysis*, *2014*. https://doi.org/10.1155/2014/697245

Liu, Z., Ling-Yu Duan, Jie Chen, & Huang, T. (2016). Depth-based Local Feature Selection for Mobile Visual Search. In *International Conference on Image Processing (ICIP)*. IEEE.

Loncomilla, P., Ruiz-del-Solar, J., & Martínez, L. (2016). Object recognition using local invariant features for robotic applications: A survey. *Pattern Recognition*, *60*, 499–514. https://doi.org/10.1016/j.patcog.2016.05.021

Loussaief, S., & Abdelkrim, A. (2018). Deep Learning vs . Bag of Features in Machine Learning for Image Classification. *2018 International Conference on Advanced Systems and Electric Technologies (IC_ASET)*, 6–10. https://doi.org/10.1109/ASET.2018.8379825

M. Anthimopoulos, M., Gianola, L., Scarnato, L., Diem, P., & G.Mougiakakou, S. (2014). A Food Recognition System for Diabetic Patients Based on an Optimized Bag-of-Features Model. *IEEE Journal OF Biomedical and Health Informatics*, *18*(4), 1261–1271.

Ma, P., Seeland, M., Rzanny, M., Alaqraa, N., & Wa, J. (2017). Plant species classification using flower images — A comparative study of local feature representations. *PLoS ONE*, 1–30. https://doi.org/10.7910/DVN/QDHYST

Mangai, U., Samanta, S., Das, S., & Chowdhury, P. (2010). A Survey of Decision Fusion and Feature Fusion Strategies for Pattern Classification. *IETE Technical Review*, *27*(4), 293. https://doi.org/10.4103/0256-4602.64604

Mansourian, L., Abdullah, M. T., Abdullah, L. N., Azman, A., & Mustaffa, M. R. (2018). An effective fusion model for image retrieval. *Multimedia Tools and Applications*, *77*(13), 16131–16154. https://doi.org/10.1007/s11042-017-5192-x

Maria Farinella, G., Allegra, D., Stanco, F., Battiato, S., Hisham, R., & Nor, M. (2015). On the Exploitation of One Class Classification to Distinguish Food Vs Non-Food Images. In *International Conference on Image Analysis and Processing* (Vol. 9281, pp. 375–383). https://doi.org/10.1007/978-3-319-23222-5

Martinel, N., Foresti, G. L., & Micheloni, C. (2016). Wide-Slice Residual Networks for Food Recognition. In *Computer Vision and Pattern Recognition*.

Martinel, N., Piciarelli, C., & Micheloni, C. (2016). A supervised extreme learning committee for food recognition. *Computer Vision and Image Understanding*, *148*, 67–86. https://doi.org/10.1016/j.cviu.2016.01.012

Martinel, N., Piciarelli, C., Micheloni, C., & Foresti, G. L. (2016). A Structured Committee for Food Recognition. *Proceedings of the IEEE International Conference on Computer Vision*, *2016-Febru*, 484–492. https://doi.org/10.1109/ICCVW.2015.70

Matas, J., Chum, O., Urban, M., & Pajdla, T. (2002). Robust Wide Baseline Stereo from. *In British Machine Vision Conference*, 384–393. https://doi.org/10.5244/C.16.36

Matsuda, Y., Hoashi, H., & Yanai, K. (2012). Recognition of multiple-food images by detecting candidate regions. In *Proceedings - IEEE International Conference on Multimedia and Expo* (pp. 25–30). https://doi.org/10.1109/ICME.2012.157

Min, W., Jiang, S., Liu, L., Rui, Y., & Jain, R. (2018). A Survey on Food Computing. *Computers and Society*.

Mokhtar, S. A. E. H., & Elnemr, H. A. (2018). A comparative study of data mining approaches for Bag of Visual Words based image classification. *Journal of Computer Science*, *14*(1). https://doi.org/10.3844/jcssp.2018.53.66

Mukherjee, P., Srivastava, S., & Lall, B. (2016). Salient keypoint selection for object representation. *2016 22nd National Conference on Communication, NCC 2016*. https://doi.org/10.1109/NCC.2016.7561176

Murino, V., Puppo, E., Sona, D., Cristani, M., & Sansone, C. (2015). Dish Detection and Segmentation for Dietary Assessment on Smartphones. In *International Conference on Image Analysis and Processing* (Vol. 9281, pp. 433–440). https://doi.org/10.1007/978-3-319-23222-5

Nistér, D., & Henrik, S. (2008). Linear time maximally stable extremal region. In *European Conference on Computer Vision* (pp. 183–196).

Nguyen, D., Zong, Z., O.Ogunbona, P., Probst, Y., & Li, W. (2014). Food image classification using local appearance and global structural information. *Neurocomputing*, *140*, 242–251.

Nowak, E., Jurie, F., & Triggs, B. (2006). Sampling strategies for bag-of-features image classification. In *9th European Conference on Computer Vision* (Vol. 3954 LNCS, pp. 490–503). Springer-Verlag Berlin Heidelberg. https://doi.org/10.1007/11744085_38

Okawa, M. (2018). From BoVW to VLAD with KAZE features: Offline signature verification considering cognitive processes of forensic experts. *Pattern Recognition Letters*, *0*, 1–8. https://doi.org/10.1016/j.patrec.2018.05.019

Oliveira, L., Costa, V., Neves, G., Oliveira, T., Jorge, E., & Lizarraga, M. (2014). A mobile, lightweight, poll-based food identification system. *Pattern Recognition*, *47*, 1941–1952.

Parisa Pouladzadeh Aslan Bakirov, Ahmet Bulut, Abdulsalam Yassine, S. S. (2015). Cloud-based SVM for food categorization. *Multimedia Tools Application*, *74*, 5243–5260.

Park, D. (2013). Image Data Classification Using Fuzzy c-Means Algorithm with Different Distance Measures. In *International Symposium on Neural Networks* (pp. 489–496).

Peng, X., Wang, L., Wang, X., & Qiao, Y. (2015). Bag of visual words and fusion methods for action recognition: Comprehensive study and good practice. *Computer Vision and Image Understanding*, *150*, 109–125. https://doi.org/10.1016/j.cviu.2016.03.013

Perronnin, F., & Dance, C. (2006). Fisher Kernels on Visual Vocabularies for Image Categorizaton. In *Conference on Computer Vision and Pattern Recognition*. IEEE.

Perronnin, F., Sanchez, J., & Mensink, T. (2010). Improving the Fisher Kernel for Large-Scale Image Classificatio. In *European Conference on Computer Vision* (pp. 143–156).

Pooja, H., & Madival, P. S. A. (2016). Food Recognition and Calorie Extraction using Bag-of- SURF and Spatial Pyramid Matching Methods. *International Journal of Computer Science and Mobile Computing*, *5*(5), 387–393.

Pouladzadeh, P., Shirmohammadi, S., Bakirov, A., Bulut, A., & Yassine, A. (2015). Cloud-based SVM for food categorization. *Multimedia Tools and Applications*, 5243–5260. https://doi.org/10.1007/s11042-014-2116-x

Pouladzadeh, P., Shirmohammadi, S., Member, S., & Al-maghrabi, R. (2014). Measuring Calorie and Nutrition From Food Image. *IEEE Transactiions on Instrumentation and Measurement Society*, *63*(8), 1947–1956.

Pouladzadeh, P., Yassine, A., & Shirmohammadi, S. (2015). New Trends in Image Analysis and Processing -- ICIAP 2015 Workshops. In *International Conference on Image Analysis and Processing* (Vol. 9281, pp. 441–448). https://doi.org/10.1007/978-3-319-23222-5

Qiu, Q., Cao, Q., & Adachi, M. (2014). Filtering out background features from BoF representation by generating fuzzy signatures. In *International Conference on Fuzzy Theory and Its Applications (iFUZZY2014)* (pp. 14–18).

Ragusa, F., Tomaselli, V., Furnari, A., Battiato, S., & Farinella, G. M. (2016). Food vs Non-Food Classification. In *2nd International Workshop on Multimedia Assisted Dietary Management* (pp. 77–81).

Ramesh, B., Xiang, C., & Lee, T. H. (2017). Multiple object cues for high performance vector quantization. *Pattern Recognition*, *67*, 380–395. https://doi.org/10.1016/j.patcog.2017.02.024

Ravi, D., Lo, B., & Yang, G. Z. (2015). Real-time food intake classification and energy expenditure estimation on a mobile device. *2015 IEEE 12th International Conference on Wearable and Implantable Body Sensor Networks, BSN 2015*. https://doi.org/10.1109/BSN.2015.7299410

Ren, T., Qiu, Z., Liu, Y., Yu, T., & Bei, J. (2014). Soft-assigned bag of features for object tracking. *Multimedia Systems*, *21*(2), 189–205. https://doi.org/10.1007/s00530-014-0384-y

Rich, J., Haddadi, H., & Hospedales, T. M. (2016). Towards Bottom-Up Analysis of Social Food in Social Media. In *6th International Conference on Digital Health* (pp. 111–120).

Riemenschneider, H. (2008). *Online Object Recognition using MSER Tracking*. Master Thesis, Graz University of Technology.

Rocha, A., Hauagge, D. C., Wainer, J., & Goldenstein, S. (2010). Automatic fruit and vegetable classification from images. *Computers and Electronics in Agriculture*, *70*, 96–104. https://doi.org/10.1016/j.compag.2009.09.002

Rosten, E., & Drummond, T. (2006). Machine learning for high-speed corner detection. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *3951 LNCS*, 430–443. https://doi.org/10.1007/11744023_34

Rudinac, M., Lenseigne, B., & Jonker, P. (2009). Keypoint extraction and selection for object recognition. *MVA2009 IAPR Conference on Machine Vision Applications*, 191–194.

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

Sasano, S., Han, X.-H., & Chen, Y. (2016). Food Recognition by Combined Bags of Color Features and Texture Features. In *International Congress on Image and Signal Processing, BioMedical Engineering and Informatics* (pp. 815–819).

Satpathy, A., Jiang, X., & Eng, H. L. (2014). LBP-based edge-texture features for object recognition. *IEEE Transactions on Image Processing*, *23*(5), 1953–1964. https://doi.org/10.1109/TIP.2014.2310123

Shahed, M., Khan, A. U., & Chowdhury, S. A. (2017). Detection and Recognition of Bangladeshi Road Sign Based on Maximally Stable Extremal Region. In *International Conference on Electrical Information and Communication Technology* (pp. 7–9).

Shinomiya, Y., & Hoshino, Y. (2016). A Feature Encoding Based on Fuzzy Codebook for Large-Scale Image Recognition. *Proceedings - 2015 IEEE International Conference on Systems, Man, and Cybernetics, SMC 2015*, 2908–2913. https://doi.org/10.1109/SMC.2015.506

Smith, S. (1992). A New Class of Corner Finder. In *British Machine Vision Conference*. https://doi.org/10.5244/C.6.15

Song, X., Jiang, S., Xu, R., & Herranz, L. (2015). Semantic features for food image recognition with geo-constraints. *IEEE International Conference on Data Mining Workshops, ICDMW*, *2015-Janua*(January), 1020–1025. https://doi.org/10.1109/ICDMW.2014.144

Su, F., Ding, W., Wang, L., Shan, S., & Xu, H. (2017). Text Proposals Based on Windowed Maximally Stable Extremal Region for Scene Text Detection. In *International Conference on Document Analysis and Recognition Text* (pp. 1–6). https://doi.org/10.1109/ICDAR.2017.69

Sujatha, K. S., Keerthana, P., Suga Priya, S., Kaavya, E., & Vinod, B. (2012). Fuzzy based multiple dictionary bag of words for image classification. *Procedia Engineering*, *38*, 2196–2206. https://doi.org/10.1016/j.proeng.2012.06.264

Szegedy, C., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., & Rabinovich, A. (2015). Going deeper with convolutions. *2015 IEEE*

Conference on Computer Vision and Pattern Recognition (CVPR), 1–9. https://doi.org/10.1109/CVPR.2015.7298594

Takeishi, N., Tanimoto, A., Yairi, T., Tsuda, Y., Terui, F., Ogawa, N., & Mimasu, Y. (2015). Evaluation of Interest-region Detectors and Descriptors for Automatic Landmark Tracking on Asteroids. *Transactions of the Japan Society for Aeronautical and Space Sciences*, *58*(1), 45–53. https://doi.org/10.2322/tjsass.58.45

Tanno, R., Okamoto, K., & Yanai, K. (2018). DeepFoodCam : A DCNN-based Real-time Mobile Food Recognition System. In *Proceedings of the 2nd International Workshop on Multimedia Assisted Dietary Management* (pp. 89–89). https://doi.org/10.1145/2986035.2986044

Tsai, C.-F. (2012). Bag-of-Words Representation in Image Annotation: A Review. *ISRN Artificial Intelligence*, *2012*, 1–19. https://doi.org/10.5402/2012/376804

Turcot, P., & Lowe, D. G. (2009). Better matching with fewer features: The selection of useful features in large database recognition problems. *2009 IEEE 12th International Conference on Computer Vision Workshops, ICCV Workshops 2009*, 2109–2116. https://doi.org/10.1109/ICCVW.2009.5457541

Tuytelaars, T., & Mikolajczyk, K. (2007). Local Invariant Feature Detectors: A Survey. *Foundations and Trends® in Computer Graphics and Vision*, *3*(3), 177–280. https://doi.org/10.1561/0600000017

Van Gemert, J. C., Veenman, C. J., Smeulders, A. W. M., & Geusebroek, J. M. (2010). Visual word ambiguity. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *32*(7), 1271–1283. https://doi.org/10.1109/TPAMI.2009.132

Verschae, R., & Ruiz-del-solar, J. (2012). à lireee LNCS 7431 - TCAS: A Multiclass Object Detector for Robot and Computer Vision Applications, 632–641.

Verschae, R., & Ruiz-del-Solar, J. (2015). Object Detection: Current and Future Directions. *Frontiers in Robotics and AI*, *2*(November), 1–7. https://doi.org/10.3389/frobt.2015.00029

Wang, H., & Deng, W. (2015). Face Recognition via Compact Fisher Vector. In *Chinese Conference on Biometric Recognition* (pp. 68–77). https://doi.org/10.1007/978-3-319-25417-3

Wang, J., Yang, J., Yu, K., Lv, F., Huang, T., & Gong, Y. (2010). Locality-constrained linear coding for image classification. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 3360–3367. https://doi.org/10.1109/CVPR.2010.5540018

Wang, X., Wang, L., & Qiao, Y. (2013). A comparative study of encoding, pooling and normalization methods for action recognition. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *7726 LNCS*(PART 3), 572–585. https://doi.org/10.1007/978-3-642-37431-9_44

Wang, Y., Member, S., Li, S., & Kot, A. C. (2015). DeepBag : Recognizing Handbag Models. *IEEE Transactions on Multimedia*, *17*(11), 2072–2083. https://doi.org/10.1109/TMM.2015.2480228

Warfield, N. I. W., K, S., Weisenfeld, N. I., & Warfield, S. K. (2008). Kernel Codebooks for Scene Categorization. *European Conference on Computer Vision*, 696–709. https://doi.org/10.1007/978-3-642-23626-6_40

Wazumi, M., Han, X.-H., & Chen, Y.-W. (2013). Food recognition using Codebook-based model with sparse-coding. *Proceedings of the 2013 IEEE/SICE International Symposium on System Integration*, 482–485. https://doi.org/10.1109/SII.2013.6776730

WHO. (2018). Obesity and Overweight. Retrieved from http://www.who.int/en/news-room/fact-sheets/detail/obesity-and-overweight

Wu, J., Cui, Z., Sheng, V. S., Zhao, P., Su, D., & Gong, S. (2013). A Comparative Study of SIFT and its Variants. *Measurement Science Review*, *13*(3), 122–131. https://doi.org/10.2478/msr-2013-0021

Xie, L., Tian, Q., & Zhang, B. (2016). Simple Techniques Make Sense: Feature Pooling and Normalization for Image Classification. *IEEE Transactions on Circuits and Systems for Video Technology*, *26*(7), 1251–1264. https://doi.org/10.1109/TCSVT.2015.2461978

Xu, R., Herranz, L., Jiang, S., Wang, S., Song, X., & Jain, R. (2015). Geolocalized Modeling for Dish Recognition. *IEEE Transactions on Multimedia*, *17*(8), 1187–1199. https://doi.org/10.1109/TMM.2015.2438717

Xu, X., Mu, N., & Zhang, H. (2015). Inferring Visual Perceptual Object by Adaptive Fusion of Image Salient Features. *Mathematical Problems in Engineering*, *2015*. https://doi.org/10.1155/2015/973241

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

Yang, J., Yu, K., Gong, Y., & Beckman, T. H. (2009). Linear Spatial Pyramid Matching Using Sparse Coding for Image Classification. *IEEE Compmuter Society Conference on Computer Vision and Pattern Recognition (CVPR)*, 1794–1801. https://doi.org/10.1109/CVPR.2009.5206757

Yang, L., Hu, Q., Zhao, L., & Li, Y. (2015). Salience based hierarchical fuzzy representation for object recognition. *Proceedings - International Conference on Image Processing, ICIP*, *2015-Decem*, 4873–4877. https://doi.org/10.1109/ICIP.2015.7351733

Yang, S., Chen, M., Pomerleau, D., & Sukthankar, R. (2010). Food Recognition Using Statistics of Pairwise Local Features. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. https://doi.org/10.1109/CVPR.2010.5539907

Yoo, S., & Kim, C. (2013). Background subtraction using hybrid feature coding in the bag-of-features framework. *Pattern Recognition Letters*, *34*(16), 2086–2093. https://doi.org/10.1016/j.patrec.2013.07.008

Yoshiyuki Kawano, K. Y. (2015). FoodCam: A real-time food recognition system on a smartphone. *Multimedia Tools Applications*, *74*(14), 5263–5287.

Yu, J., Qin, Z., Wan, T., & Zhang, X. (2013). Feature integration analysis of bag-of-features model for image retrieval. *Neurocomputing*, *120*, 355–364. https://doi.org/10.1016/j.neucom.2012.08.061

Yu, K., America, N. E. C. L., & Road, N. W. (2010). Improved Local Coordinate Coding using Local Tangents. In *Improved Local Coordinate Coding using Local Tangents* (pp. 1215–1222).

Yu, K., & Zhang, T. (2009). High Dimensional Nonlinear Learning using Local Coordinate Coding, 1–9. https://doi.org/citeulike-article-id:11997015

Zadeh, L.A. (1965). Fuzzy sets. *Information and Control*, *8*(3), 338–353. https://doi.org/10.1016/S0019-9958(65)90241-X

Zamani, M., & Kremer, S. C. (2013). *Handbook on Neural Information Processing* (Vol. 49). https://doi.org/10.1007/978-3-642-36657-4

Zhang, C., Wen, G., Lin, Z., Yao, N., Shang, Z., & Zhong, C. (2016). An Effective Bag-of-visual-word Scheme for Object Recognition. In *International Congress on Image and Signal Processing, BioMedical Engineering and Informatics* (pp. 417–421). https://doi.org/10.1109/CISP-BMEI.2016.7852747

Zhang, F., Song, Y., Cai, W., Hauptmann, A. G., Liu, S., Liu, S., … Chen, M. (2015). Ranking-Based Vocabulary Pruning in Bag-of-Features for Image Retrieval. *Artificial Life and Computational Intelligence*, *8955*, 436–445.

Zhang, S., & Xie, M. (2013). Beyond sliding windows: Object detection based on hierarchical segmentation model. *2013 International Conference on Communications, Circuits and Systems, ICCCAS 2013*, *1*(1), 263–266. https://doi.org/10.1109/ICCCAS.2013.6765229

Zhang, X., Yang, Y.-H., Han, Z., Wang, H., & Gao, C. (2013). Object Class Detection: A Survey. *ACM Computing Surveys*, *46*(1), 1–46.

Zheng, J., Wang, Z. J., & Member, S. (2016). Superpixel-based Image Recognition for Food Images. In *IEEE Canadian Conference on Electrical and Computer Engineering (CCECE)*.

Zheng, J., Wang, Z. J., & Zhu, C. (2017). Food Image Recognition via Superpixel Based Low-Level and Mid-Level Distance Coding for Smart Home Applications. *Sustainability*, *9*(5). https://doi.org/10.3390/su9050856

Zhou, X., Yu, K., Zhang, T., & Huang, T. S. (2010). Image Classification Using Super-Vector Coding of Local Image Descriptors. In *European Conference on Computer Vision* (pp. 141–154). https://doi.org/10.1007/978-3-642-15555-0_11

Zhu, F., Bosch, M., Khanna, N., J. Boushey, C., & J. Delp, E. (2015). Multiple Hypotheses Image Segmentation and ClassificationWith Application to Dietary Assessment. *IEEE Journal Of Biomedical and Health Informatics*, *19*(1), 377–388.

Ziomek, A., & Oszust, M. (2016). Evaluation of Interest Point Detectors in Presence of Noise. *International Journal of Intelligent Systems and Applications*, *8*(3), 26–33. https://doi.org/10.5815/ijisa.2016.03.03

Zong, Z., Nguyen, D. T., Ogunbona, P., & Li, W. (2010). On the combination of local texture and global structure for food classification. *Proceedings - 2010 IEEE International Symposium on Multimedia, ISM 2010*, 204–211. https://doi.org/10.1109/ISM.2010.37

Zuchun, D. (2013). An Effective Keypoint Selection Algorithm in SIFT. In *International Journal of Signal Processing, Image Processing and Pattern Recognition* (Vol. 6, pp. 155–164). https://doi.org/10.1.1.641.1181