Carbohydrate contents detection in food Using Bag of Feature

Prof. R. S. Parte, Sanas Supriya, Gaikwad Ashwini, Gupta Pooja, Kavitake Kajal

Department of Computer Engineering, Jaywantrao Sawant College of Engineering,
Savitribai Phule Pune University, Hadapsar, Pune, Maharashtra, India
gupta.pooja865@gmail.com, sanas.supriya25@gmail.com

---***---

Abstract - *Now days, due to fast food addiction, at very small age people are suffering from diabetes and blood pressure problems. Computer vision based proposed system identifies the contents of carbohydrates into a dish using BOF (Bag-Of-Feature) Model. It uses dense local feature that gives the information of speeded up robust features (SURF) over the HSV (hue, saturation, value) color space. The system builds a visual dictionary of 10000 visual words by implementation of hierarchical K-means clustering on visual based Dataset. Food images dataset is classified with high diversity using Linear support vector machine classifier which improves the accuracy over the previous system. The system includes a food segmentation stage before applying the proposed identification module, so that images with multiple food types can also be addressed. Finally food volume will be identified by using multi-view reconstruction and carbohydrates content will be calculated based on the computer vision results and nutritional tables.*

Key Words: **Bag-Of-Features, Diabetes, food identification, image classification, feature extraction, SURF, Computer Vision, Naive Bayes**

1. INTRODUCTION

Now a day's many people are addicted towards fast food and hotelling. Because of that food they face the problem of diabetes. Diabetes occurs when the level of sugar (glucose) in the blood becomes higher than normal. There are two main types of diabetes - type 1 diabetes and type 2diabetes.Type 2 diabetes occurs mainly in people aged over 40. Type 1 diabetes is treated with insulin injections and diet .It is necessary to detect whether the food is healthy or not, this system is designed. Users have access to capture food image using the mobile and these images are sent to the server for processing. The server side contain the large dataset of images classified using BOF. Server is processing on the image using the feature of image like size, color, shape, pixel value etc. It detects the carbohydrate content in food from

This paper covers many sections. Section II covers the literature survey carried out for this paper. Section III Discussion about comparative study of all papers. Section IV depicts the proposed approach. Section V is conclusion and future work.

2. LITERATURE SURVEY

In our day to day life diet is very important because people are suffering from diseases like obesity, diabetes, cancer etc. They don't know about diet so for measuring the accurate diet they are using novel mobile telephone food record which gives the accurate account of daily food and nutrient intake .Using that mobile devices they are capturing images before and after eating to estimate the amount and type of food consume. For classification they used SVM for identifying the food item they have used statistical food item techniques. They are using concept of volume estimation which consist of camera calibration and 3-D volume reconstruction. For input they are using two images one of that food image taken by user and other is segmented image [1].

The accuracy of carbohydrate content for Adolescents with Type1 diabetes. It includes the 48 Adolescents ages between 12-18 year with Type1 diabetes of >1 year who used ratio of insulin: carbohydrate for at least one meal per day. T tests were used for assess the importance of over or underestimation of carbohydrate content. Each meals accuracy was categorized as accurate if it is within 10 grams or estimated if it is greater than 10 grams or underestimated if it less than 10 grams. The mean difference between carbohydrate content estimated by patient and Actual content. System studies and looked only for 1 time measurement of carbohydrate and its counting accuracy which changes time to time. This is self managing system which made for the patient and provides the data to make more informed decisions to patient [2].

Presently the first visual dataset for fast foods with total of 4545 still images, 606 stereo pairs, 303 360 for structure from nation. They are using two standard approaches color histogram & bag of SIFT features with a discriminative classifier. They have provided dataset of 101 fast foods. They have provided freely available dataset for computer vision research on the food classification. Their goal link PFID to the food and netutrient database for Dietary studies database which help to simulate other research to work on the food recognition problem [3].

Accuracy of food is improved using the computer vision technique. Drawback of tradition system is inaccurate assessment or complex lab measurement so they have implemented system which use mobile phone for capturing images of food and identify the food type, estimate their respective volume and return quantitative nutrition information.

The System has a combination multiple vision techniques to achieve the quantitative food intake estimation. Food intake visual and voice recognizer system is used to measure nutritional content of a user meal. The mobiles are ubiquitous devices and mostly come equipped with cameras. The FIVR system works a calibrated cell phone as a capturing device [4].

Dish extraction method by neural network firstly the system extracts the food image before and after food intake. Secondly the system compares the amount of food intake between initial food images and remained one. Finally amount of food intake is determined by measuring system input data of NN are dish image shapes, diameter, width, height. This method cannot specify the dish position accurately by itself [5].

System is implemented using SURF: Speeded - Up Robust Features algorithm for cutting edge image feature scheme. Because the dense algorithm can be used only on the high resolution images to achieve efficient performance, they generally have very slow processing speed, so the SURF algorithm is used. SURF implies the feature detector using the Gaussian Second Derivative mask, the Local Haar Wavelet is used for getting the feature descriptor. This concept is much conceptually similar to the most widely used feature detector in computer vision community i.e. the Scale Invariant Feature Transform (SIFT). The research has proved that SURF gives more output both in terms of Speed and accuracy [6].

The SURF detector algorithm has following steps:

1. Form the scale - space response by convolving the source image using DoH filters with different σ

2. Search for local maxima across neighbouring pixels and adjacent scales within different octaves

3. Interpolate the location of each local maxima found

4. For each point of interest, return x, y, σ, the DoH magnitude, and the Laplacian's sign

Bag-of-words [BOW], contain image by the histogram of local patches on the basis of a visual vocabulary, because of good performance and flexibility having good attention in the visual categorization this paper implement a novel Contextual bag-of-words [CBOW] representation that display relations between local patches, i.e., a semantic conceptual relation and a spatial neighboring relation. Using multiple semantic levels according to the similarity of class distribution. Visual words are combined and images are represented according to local patches are encoded [7].

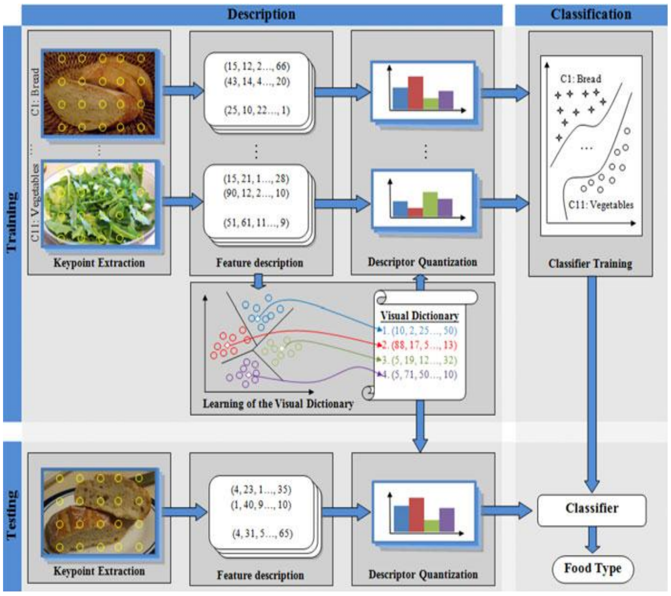

Fig -1: System Architecutre

2. PROPOSED SYSTEM

System makes several contributions to the field of food recognition. A visual dataset with nearly 5000 homemade food images was created, reflecting the nutritional habits in central Europe. The foods appearing in the images have been organized into 11 classes of high intravariability. Based on the aforementioned dataset, we conducted an extensive investigation for the optimal components and parameters within the BoF architecture. Three key point extraction techniques, fourteen lo-cal image descriptors, two clustering methods for the creation of the visual dictionary, and six classifiers were tested and comparatively assessed. Descriptors' fusion and feature selection were also tested. Moreover, the effects of various parameters like the number of extracted key points, the descriptor size(s), and the number of visual words are illustrated after conducting extensive experiments. Finally, a system for the recognition of generic food is proposed based on an optimized BoF model.

3. INTRODUCTION OF SPEEDED-UP ROBUST FEATURE (SURF)

3.1 SURF detector

The SURF detector focuses its attention on blob-like struct ures in the image. These structures can be found at corner s ofobjects, but also at locations where the reflection of lig ht on specular surfaces is maximal (i.e. light speckles).

In 1998, Lindeberg noticed that the Monge-Ampère operator and operator and Gaussian derivative filters could be used to locate features. Specifically, he detected blobs by convolving the source image with the determinant of the Hessian (DoH) matrix, which contains different 2-D Gaussian second order derivatives. This metric is then divided by the Gaussian's variance, σ2, to normalize its response:

$$DoH(x,y,\sigma) = \frac{G_{xx}(x,y,\sigma) \cdot G_{yy}(x,y,\sigma) - G_{xy}(x,y,\sigma)^2}{\sigma^2}$$

$$where\ G_{ij}(x,y,\sigma) = \frac{\partial N(0,\sigma)^2}{\partial i \cdot \partial j} * image(x,y)$$

The local maxima of this filter response occur in regions where both G_{xx} &$G_{yy}$are strongly positive,and where G_{xy} is strongly negative. Therefore, these extrema occur in regions in the image with large intensity gradient variations in multiple directions, as well as at saddle points. Visually, this means that blob-like structures refer to corners and speckles.

The other reason why many feature detection schemes rely on Gaussian filters is to get rid of noisy data by blurring the image. As a side-effect, Gaussian blurring highlights image details at or near a single unique scale.

As Marr wrote in the Theory of edge Detection, "no single filter can be optimal simultaneously at all scales, so it follows that one should seek a way of dealing separately with the changes occurring atdifferent scales." He goes on to propose to extract features using the combined information of multiple responses, generated using the same family of filters but at different scales. This resulting stack of convolutions is typically referred to as the scale space.

3.2 SURF Descriptor

To describe each feature, SURF summarizes the pixel information within a local neighbourhood. The first step is determining an orientation for each feature, by convolving pixels in its neighbourhood with the horizontal and the vertical Haar wavelet filters. Shown in Fig 7, these filters can be thought of as block-based methods to compute directional derivatives of the image's intensity. By using intensity changes to characterize orientation, this descriptor is able to describe features in the same manner regardless of the specific orientation of objects or of the camera. This rotational invariance property allows SURF features to accurately identify objects within images taken from different perspectives

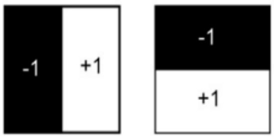

Fig 7: Horizontal and vertical Haar wavelet filters

In fact, the intensity gradient information can also reliably characterize these pixel regions. By looking at the normalized gradient responses, features in images taken in a dark room versus a light room, and those taken using different exposure settings will all have identical descriptor values. Therefore, by using Haar wavelet responses to generate a unit vector representing each feature and its neighbourhood, the SURF feature framework inherits two desirable properties lighting invariance and contrast invariance.

Because the neighbourhood can be broken down into smaller windows in arbitrary manners and because the individual Haar wavelet responses can be summarized differently, multiple variants of SURF have been proposed using different combinations of settings. Additionally, in applications such as mobile robotic vision, images are captured using statically-positioned cameras, so these photos are all oriented in similar directions. To save computation speed , a variant of the algorithm called U-SURF (where U stands for up-right) foregoes the orientation step, and computes the horizontal and vertical Haar wavelet responses using the cardinal axes of the image directly. The original authors have experimentally demonstrated that U-SURF features are orientation-invariant up to 15°.

3.3 SURF Comparator

We begin this section by formally defining the setup for our object recognition application: we are given an image library of labelled objects, where each image contains a single object in full view without obstruction, over a plain dark background. Each object is represented by multiple images, which are taken from slightly different viewpoints. They are generated using a fixed camera, taking pictures of the object as it sits on top of a controlled turntable. The object's identity and orientation are provided to the user.

The approach to object recognition is to compute a database of features for each image in the library. When given a query image at runtime, we generate the set of query features and attempt to match it to other sets within the database. Once we find the database object with the best feature matching based on some comparison metric, we conclude that this object is present in the query image. This type of approach is very similar to appearance-based identification algorithms, using eigenspace and Principal Component Analysis (PCA). With an identical database setup, pixel-based PCA methods compute a high-dimensional vector per image based on the Eigen space of the pixel covariance matrix.

Therefore, each object can be represented by a manifold in the high dimensional space of the principal component weights. To match objects using PCA, the query image is projected onto the principal components of each database object's eigenspace, and the associated weight vector is compared to existing manifolds. The object identity of the closest manifold is assigned as the recognition match. By using features, each image is now represented by a set of feature vectors rather of a single weight vector. Because individual feature descriptor values are not related numerically, each object is no longer represented by a manifold, but instead corresponds to a cloud in feature space containing the features belonging to all of its images. Similarly, the query data can also be seen as a cloud of features. We use the Euclidean distance metric of the near est neighbouring feature to assess the distance between poin t clods.

4. CONCLUSION AND FUTURE WORK

In this proposed paper we are implementing a system which detects the CHO contents using the Bag-Of-Features algorithm. In the literature survey it was seen that SFIT algorithm did not provide much accuracy in extracting the key points from the image so the proposed system implements SURF algorithm that provides more accuracy and overcomes the drawbacks of SIFT. The System uses BOF for storing the descriptor, this descriptor are of 64 bits long. Using the Naive Bayes classifier the food images dataset classification is done.

Enhanced features system to be designed in which the CHO contains is to be detected but depending on the amount of food consumed or the quantity of food which is consumed. More precise results to be generated using computer vision. Moreover the enhancement of visual dataset with more images will improve the classification rate especially for the classes with high diversity.

ACKNOWLEDGEMENT

We express true sense of gratitude towards our project guide Prof. R. S. Parte, Associate Professor Computer Department for her invaluable co-operation and guidance that she gave us throughout our Project. We specially thank our project coordinator Prof. M. V. Pawar for inspiring us and for providing us all the lab facilities. We would also like to express our appreciation and thanks to HOD Prof. H. A. Hingoliwala and Principal Dr. M. G. Jadhav and all our friends who have assisted us throughout our hard work.

REFERENCES

[1] The Use of Mobile Devices in Aiding Dietary Assessment and Evaluation Fengqing Zhu, Student Member, IEEE, Marc Bosch, Student Member, IEEE, Insoo Woo, SungYe Kim,Carol J. Boushey, David S. Ebert, Fellow, IEEE, and Edward J. Delp, Fellow, IEEE M. Young, The Technical Writer's Handbook. Mill Valley, CA: University Science, 1989.

[2] The Carbohydrate Counting in Adolescents With Type 1 Diabetes (CCAT) Study, Franziska K. Bishop, MS, David M. Maahs, MD, Gail Spiegel, MS, RD, CDE, Darcy Owen, MS, RD, CDE, Georgeanna J. Klingensmith, MD, Andrey Bortsov, MD, Joan Thomas, MS, RD, and Elizabeth J. Mayer-Davis, PhD, RD.

[3] PFID: PITTSBURGH FAST-FOOD IMAGE DATASET Mei Chen1, Kapil Dhingra3, Wen Wu2, Lei Yang2, Rahul Sukthankar1, Jie Yang2 1Intel Labs Pittsburgh, 2Carnegie Mellon University, 3Columbia University http://pfid.intel-research.net

[4] Recognition and Volume Estimation of Food Intake using a Mobile Device Manika Puri Zhiwei Zhu Qian Yu Ajay Divakaran Harpreet Sawhney Sarnoff Corporation 201 Washington Rd,Princeton, NJ, 08540

[5] Dish Extraction Method with Neural Network for Food Intake Measuring System on Medical UseFumiaki Takeda, Kanako Kumada and Motoko Takara Development of Information Systems Engineering, Kochi University of Technology, Kochi, 782-8502, Japan Tel: +81-887-57-2300 Fax:+81-887-57-2220

Email: takeda.fumiaki@kochGtech.ac.jp

[6] SURF: Speeded-Up Robust Features COMP 558 – Project Report Presented By: Anqi Xu (260148014) & Gaurav Namit (260307292)

[7] Contextual Bag-of-Words for Visual CategorizationTeng Li, Tao Mei, In-So Kweon, Member, IEEE, and Xian-Sheng Hua, Member,IEEE