International Journal of Recent Technology and Engineering (IJRTE)
ISSN: 2277-3878 (Online), Volume-8, Issue-1, May 2019

Bag-of-Surf and Spatial Pyramid Matching for Food Recognition and Calorie Extraction

M. Aravind Kumar, D. Naresh, Srinivas Bachu

Abstract: ***Now a days the diet control become habitat to persons all over the world. Hence, in this paper the estimation of calories from recognized food is explained using various algorithms. Basically, the main techniques for extraction of features are bag-of surf method and SPM method and also classification is based on support vector machine classification. The classification rate is about 86% and the results are examined based on the PFID dataset. This technique can also be experimented by using Android.***

Index Terms**: *SURF, PFID, food recognition, spatial pyramid matching, calorie estimation.***

I. INTRODUCTON

Dietary meals vitamins and minerals are the fundamental supply for maintain existence. Food being the essential wellspring of force it's expended in an assortment of style with different introductions [1, 5]. Prior nourishment was once devoured customarily as entire vegetables, natural products, grain, dairy item and nectar. Instruction of any sustenance with the top notch supplement esteems is of vital test both in home and modern condition.

Sustenance classification acknowledgment offers with the mechanized awareness of dinners objects/types with capacities relating to customized dietary appraisal, nourishment recommendation contributions and web-based social networking (suppers) photographs examination. This zone is turning into a thriving report territory with the expansion of reasonable phones and online life administrations. Picture preparing and AI are at the coronary heart of such procedures with different innovations equal to 3D recreation moreover being connected when divide/amount estimation is required to inexact caloric amounts. We suggest two methods to respect food objects: Speeded up robust features (SURF) descriptor established by Bag-of features model and Spatial Pyramid Matching (SPM) [5, 6]. In bag-of-SURF approach, we first build a dictionary of code words, then generate a histogram of code words for all trained pixels and use linear kernel classification scheme. Spatial pyramid matching procedure tries to account for the spatial information via dividing and subdividing the given food image and developing the histogram of code words of individual areas. We then teach a classifier with spatial pyramid kernel utilizing LIBSVM bundle [7, 8].

Revised Manuscript Received on 30 May 2019.

***** Correspondence Author

Dr. M. Aravind Kumar*, Department of ECE, GVIT College of Engineering / Bhimavaram, A.P, India.

D. Naresh, Department of ECE, Guru Nanak Institutions Technical Campus (Autonomous) / Hyderabad, Telangana, India.

Dr. Srinivas Bachu, Department of ECE, Marri Laxman Reddy Institute of Technology and Management/Dundigal, Hyderabad, Telangana, India.

II. RELATED WORKS

Herbert Bay et al., [1], proposed novel scale-and pivot invariant scale factor identifier and descriptor, authored as SURF (quickened solid highlights). It approximates and even beats earlier proposed plans with appreciate to repeatability, zone of interests, and power, yet will likewise be processed and in correlation so much quick and adaptable.

That is finished by methods for depending on different pictures for picture convolutions; by methods for developing on the qualities of the dominating surf identifiers and descriptors and with the guide of rearranging these techniques to the prevalent. This outcome in a blend of novel location, portrayal, and coordinating advances.

The paper offers experimental final result on a traditional analysis set, as nice as on imagery got throughout the context of an actual-existence object attention utility. Both methods shows SURF method has strong efficiency.

S Lazebnik, C Schmid, J Ponce [2], presented a technique for perceiving scene classes headquartered on rough world geometric correspondence. This procedure works by methods for using apportioning the picture into increasingly more fine sub-locales and processing histograms of close-by angles found within each sub-provincial. The resulting spatial pyramid is a simple and computationally ground-breaking augmentation of a request less pack of-surf picture portrayal, and it indicates massively increased effectivity on convoluted scene arrangement obligations. Fundamentally, our proposed methodology surpasses the best in class on the Caltech-one hundred and one database and accomplishes extreme precision on a tremendous database of fifteen normal scene classes.

Chih-Chung Chang and Chih-Jen Lin [3], Proposed LIBSVM library for support Vector Machines (SVMs). We now have been actively opening this bundle in view that the 12 months 2000. The intent is to support persons to effortlessly conform to SVM to their purposes. LIBSVM has got big status in computer imaginative and prescient services. Listed here, we reward all implementation small print of LIBSVM.

Disorders an identical to fixing SVM optimization disorders, theoretical convergence, multi-type classification, possibility estimates, and parameter decisions are mentioned in features. Mei Chen et al., [4], presented the essential visual dataset of speedy nourishments with an entire of four,545 still pix, 606 stereo sets, 303 360° recordings for structure from development, and 27 protection keeping up motion pictures of expending interests of volunteers. This work was once impacted by utilizing research on quick supper's awareness for dietary correlation. The information used to be gathered by methods for acquiring three events of one hundred and one sustenance's from eleven favored quick dinner's chains, and catching pictures and recordings in every eatery stipulations and a controlled lab environment. We benchmark the dataset utilizing two general strategies, shading histogram and sack of Filter includes together with a discriminative classifier. Our dataset and the benchmarks are intended to animate research in this control and will be propelled uninhibitedly to the examination gathering.

Various methods are proposed for recognition of food and calorie extraction. The methods are below: Object recognition process plays one of the key roles in present generations. This recognition process involves feature extraction based on local features such as SIFT features and global features such as GIST features; images are typically represented as bags of these features without explicit spatial information [9, 10]. There are other methods in which the spatial information is taken as feature extraction. Other method involves the formation of pixels by considering n-1 vectors from n vectors using the range of pixels. These features are extracted based on the deformable shapes such as polygons. A straightway forward feature is also calculated in order to acquire the class using interplay classifier. Based on the thin plate spline parameterization the mean shape of the object in the image is considered as another feature.

These methods show valid results for recognition of object but the recognition of food is not valid. These features are mainly edges, counters. But food doesn't possess any edges or counters [11]. Detection of food with these features is very hard to detect because the food that we intake has randomness in shape. The total accuracy is very less.

III. METHODOLOGY

This section provides the detailed explanation of the proposed methods. We first describe SURF, local feature that we use in our methods and then propose two methods to classify food items that are: bag-of-SURF and SPM approach.

A. SURF:

In this process, the local features are calculated based on the surf descriptors. By learning the SIFT features, SURF features are developed for various applications such as object recognition, speech recognition etc. Similar, to SIFT features SURF features are also rotation invariant. Basically, the length of the SURF features are 64 or 128 in length, But the 64 dimension SURF features provide less memory overhead when compared to 128 dimension SURF feature descriptor.

B. BAG-of-SURF Method:

A Bag of SURF method is one that represents images as order less collections of local features. This section provides an explanation of the Bag of Features image representation, focusing on the high-level process independent of the application.

The term vector that speaks to the record is a sparse vector where every component is a term in the lexicon and the estimation of that component is the occasions the term shows up in the archive separated by the complete number of word reference words in the report. The term vector is the Pack of Words record portrayal – called a "bag" since all requesting of the words in the archive have been lost.

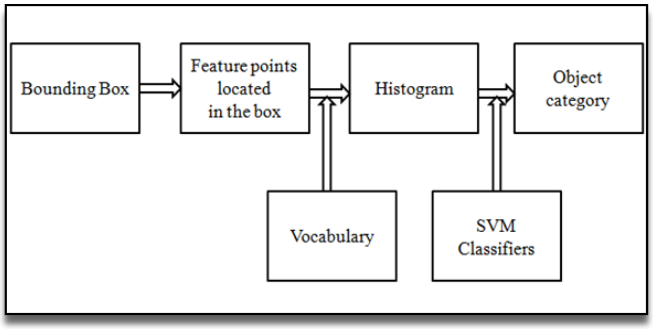

Figure 1. The process of object identification for a test image

As the bounding box has been created in the test image is shown in Figure 1, we can obtain all SIFT feature points extracted from the test image which are in the region of the bounding box. It is based on feature point's positions in the test image. With the help of these feature point descriptors, it is able to create the 'word' with the help of 'vocabularies' from the training set of images. The 'word' is then the representative of the detected object. With constructed classifiers from training process, the detected object can be identified to its belonging class.

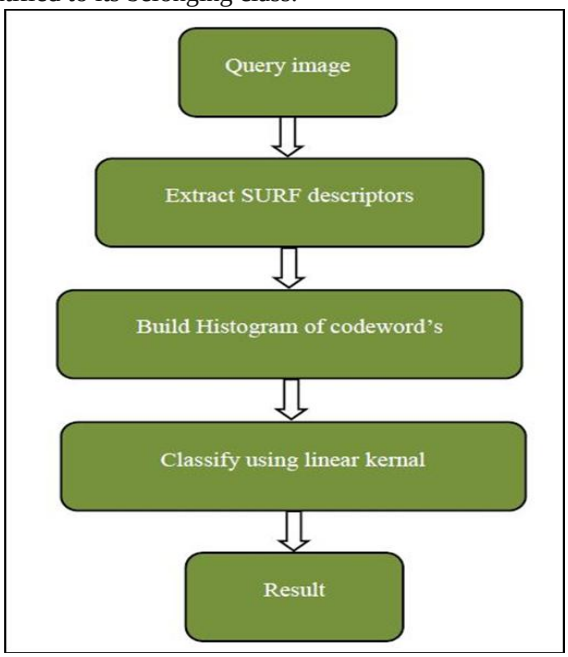

Figure 2. Bag-of-SURF model

The bag of SURF method block diagram is shown in Figure 2. In the bag of SURF method the first step involves the building of code words. To build this code dictionary the group of images is selected. As said earlier, for this process the 64 dimension SURF descriptor is used for all these subsets.

The descriptors from these subsets of images are plotted on high dimensional subspace and divided into 200 clusters based on k-means clustering algorithm, which resembles the dictionary size. The number of clusters are also increased to 300 or decreased to 100, but the descriptors with 200 sub clusters provide better performance. The process next for this is to train images using the classifier. For this, we extract 64 dimensional SURF descriptors from each of the training images. Then after by calculating the Euclidian distance, the nearest neighbor pixel values are calculated and histograms of the code words are generated for all the 200 clusters. Finally, we train a classifier using LIBSVM package using linear kernel. We also experimented using chi-square kernel method.

The process of testing is also similar as for training process. In this testing process, the code words are generated and predict based on the linear SVM classifier.

C. *Spatial Pyramid Matching*

Spatial Pyramid Matching (SPM) strategy works by method for dividing the image into more and littler sub-zones and registering histograms of spatial features found within each sub neighborhood is shown in Figure 3.

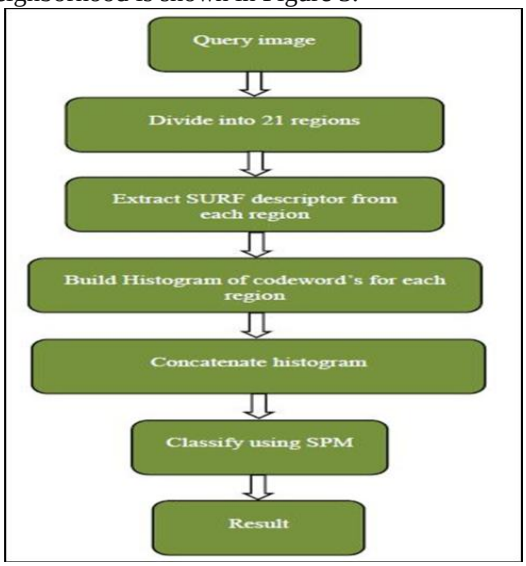

Figure 3. Spatial pyramid matching method

Through doing this, we fix the spatial data of scene in our histogram. Like pack of-SURF process, we begin by methods for structure lexicon of code words by methods for separating the SURF feature vectors from a subset of preparing previews, and afterward we plot the descriptors in 64 dimensional regions and bunch them into 200 groups making utilization of k-implies calculation.

In our investigation, we utilized 3 levels SPM: level 0, level 1 and dimension 2. Level 0 speaks to the entire picture; we get level 1 by apportioning the picture into 4 sub areas lastly level 2 is developed by parceling into 16 sub districts. In absolute we will have 21 locales. We at that point separate SURF descriptors for every one of the areas and afterward a histogram of codeword's. Histograms of all sub-areas are connected together to frame a long element vector of measurement 21*200 = 4200. Finally this histogram is normalized so that it is probability density.

In SPM, the SVM classifier is used for classification. The code words definition for X and Y values are defined as:

$$K^L(X,Y) = \frac{1}{2^L} * I^o + \sum_{t=1}^{L} \frac{1}{2^{(L-t+1)}} I^1 \qquad (1)$$

Whereas L is higher level of the SVM kernel and the intersection of this level to code words is defined as

$$I(h_1,h_2) = \sum_{t=0}^{N} \min\left(h_1(i), h_2(i)\right) \qquad (2)$$

In our experiments, L=2. Similarly, as in the bag of surf method, in this method also the test image is subdivided into 21 regions and code words are calculated and the class is predict using LIBSVM classifier.

IV. RESULTS AND DISCUSSIONS

In order to create a smaller dataset, considered one of the input images from the food items, that is chapathi, guava, kitkat, carrot, drink are shown in Figure 4, Figure 6, Figure 8, Figure 10 and Figure 12. The histogram of the corresponding images is shown in the Figure 5, Figure 7, Figure 9, Figure 11 and Figure 13. Where X-axis represents visual word index and the Y-axis represents frequency of occurrence.

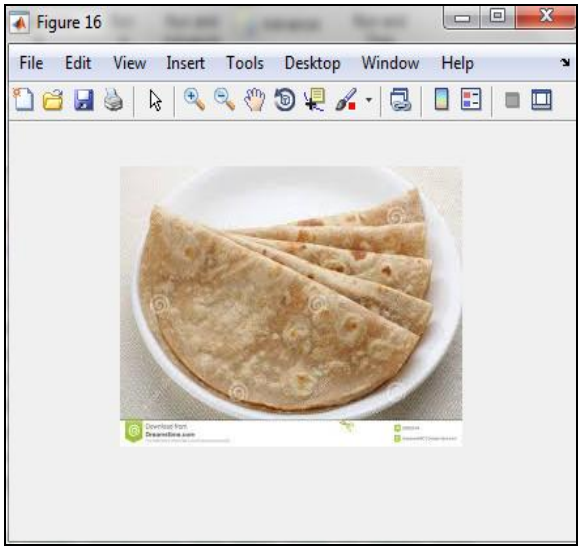

Figure 4. Sample image 1 (Chapathi)

The input and output observations like, creating bag-of features, extracting SURF features, balancing the count, visual vocabulary, SVM, iteration process and clustering are shown in MATLAB command window.

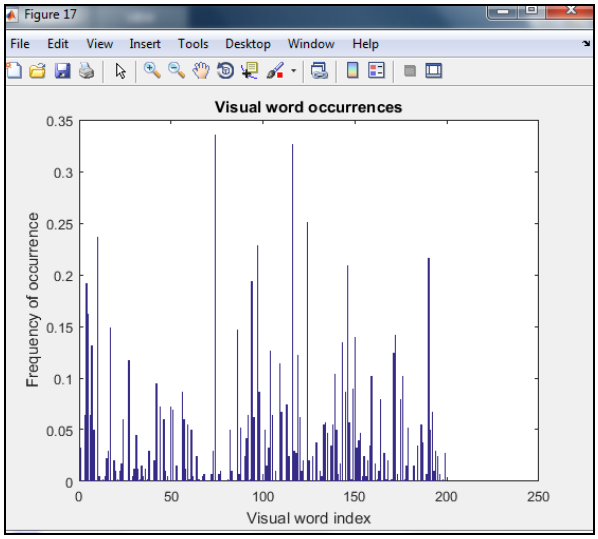

Figure 5.Histogram of sample image 1(Chapathi)

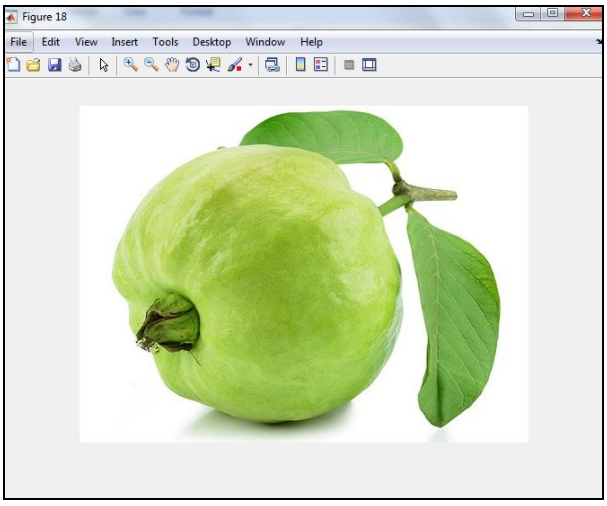

Figure 6. Sample image 2 (Guava)

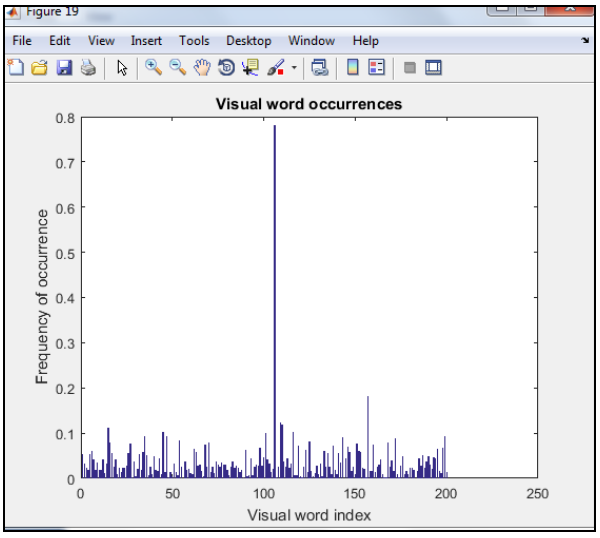

Figure 7. Histogram of sample image 2(Guava)

Figure 8. Sample image 3 (Kitkat)

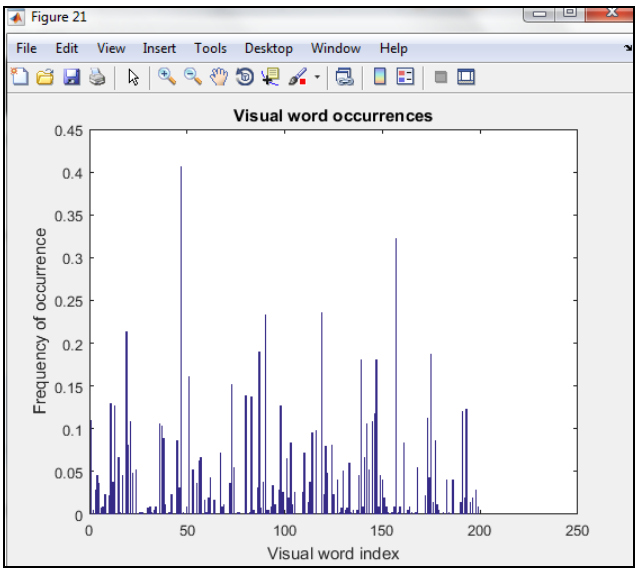

Figure 9. Histogram of sample image 3(Kitkat)

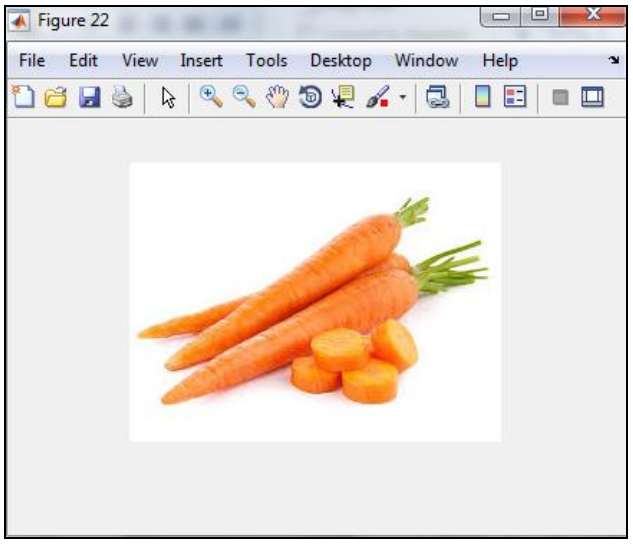

Figure 10. Sample image 4 (Carrot)

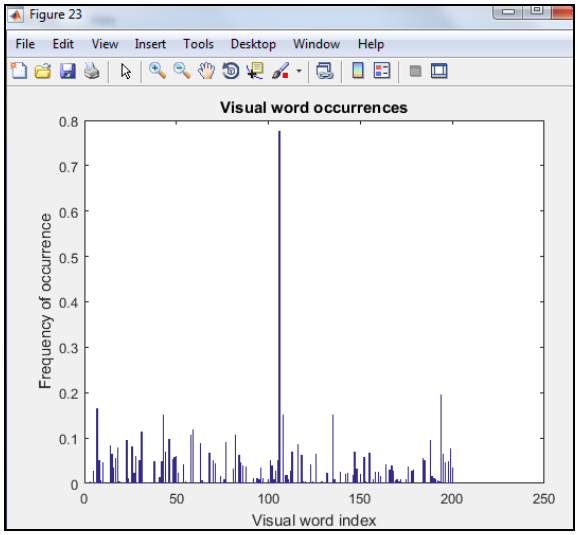

Figure 11. Histogram of sample image 4(Carrot)

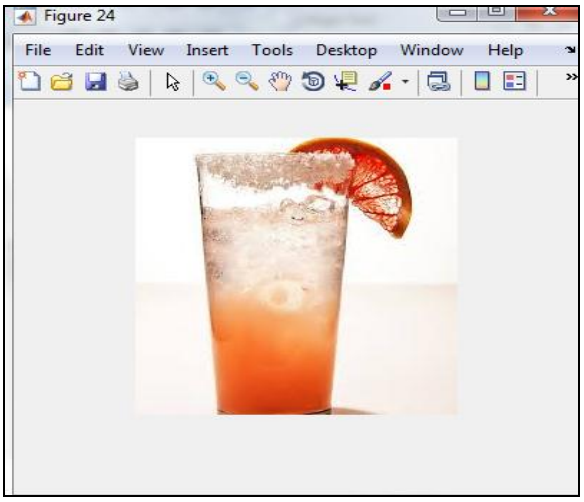

Figure 12. Sample image 5 (Drink)

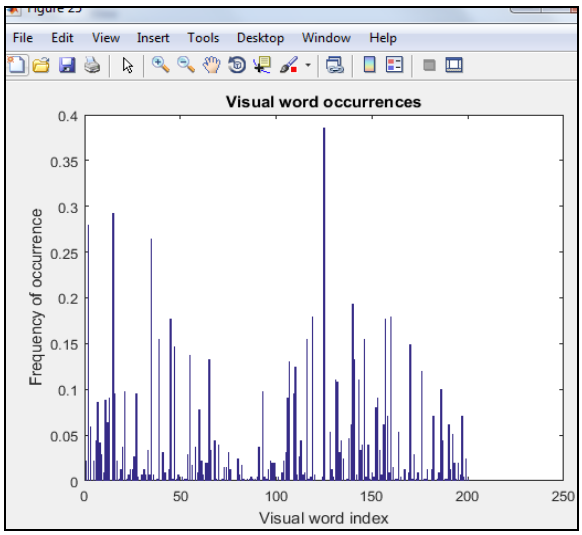

Figure 13. Histogram of sample image 5 (Drink)

Choose the test image (kitkat) from one of the food items for classification and evaluation. The classified image is shown in Figure 15. Average accuracy and test accuracy are shown in below.

>>* Average Accuracy is 1.00.

The test accuracy is 1.

Select a test image 3

calorie values

726.4009

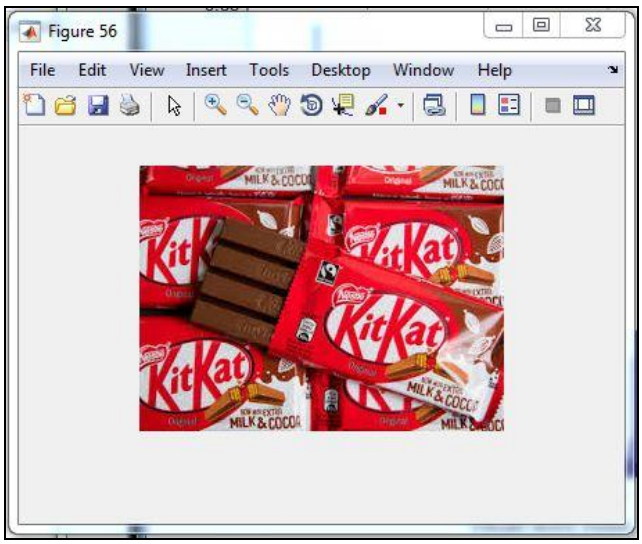

Figure 14. Test image for evaluation

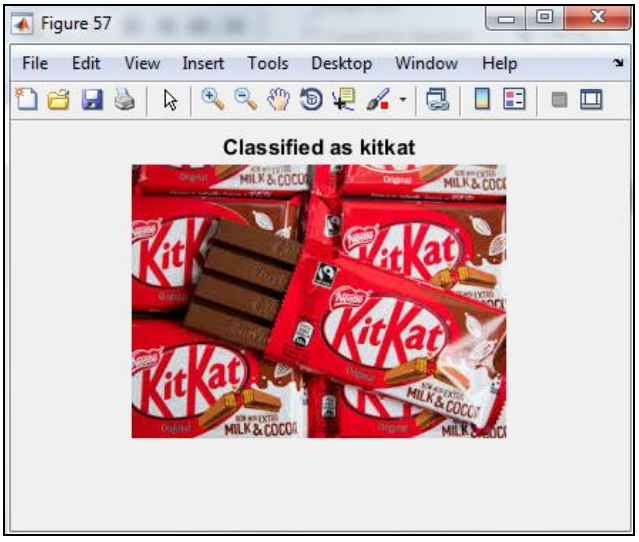

Figure 15. Classified image

Advantages

- Classification results are high.
- Recognition rate is high.
- All type food is recognized without misclassification results.
- The calorie can also be extracted using calorie table.

V. CONCLUSION

In this paper the calorific value can be estimated for each food item. The methods bag of surf and spatial pyramid matching provides valid results with high classification rate and recognition rate by using the SVM classifier.

Bag-of-SURF method with linear kernel resulted in classification accuracy of 80%. Both Bag-of-SURF and SPM based methods resulted in classification accuracy of 86%. We experimented with dictionary size of 100, 200 and 300, dictionary size of 200 performs the best in our experiment on this dataset. This method can be applied in android applications for diet control.

REFERENCES

1. Herbert Bay, Tinne Tuytelaars, and Luc Van Gool. ”SURF: Speeded Up Robust Features”. In ECCV, 2006.
2. S Lazebnik, C Schmid, J Ponce. ”Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories”. Proceedings of CVPR, 2006.
3. Chih-Chung Chang and Chih-Jen Lin. ”LIBSVM: A Library for Support Vector Machines”. ACM Transactions on Intelligent Systems and Technology, Volume 2 Issue 3, April 2011.
4. Mei Chen, Kapil Dhingra , Wen Wu, Lei Yang, Rahul Sukthankar, Jie Yang. ”PFID: Pittsburgh Fast-food Image Dataset”. Proceedings of International Conference on Image Processing, 2009.
5. S. Yang, M. Chen, D. Pomerleau, and R. Sukthankar. ”Food recognition using statistics of pairwise local features”. In Proc. of IEEE Computer Vision and Pattern Recognition, 2010.
6. Yoshiyuki Kawano and Keiji Yanai. ”Real-time Mobile Food Recognition System”. In CPVR 2013.
7. Y. Matsuda, H. Hoashi, and K. Yanai. ”Recognition of multiple-food images by detecting candidate regions”. In Proc. of IEEE International Conference on Multimedia and Expo, pages 15541564, 2012.
8. G. Shroff, A. Smailagic, and D. Siewiorek. ”Wearable context-aware food recognition for calorie monitoring”. In Proceedings of International Symposium on Wear- able Computing, 2008.
9. L.J. Li, H. Su, E. Xing, and L. Fei-Fei, Object bank: A high-level image representation for scene classication and semantic feature sparsication Advances in Neural Information Processing Systems, 2010.
10. D. G. Lowe, Distinctive image features from scale- invariant keypoints, Int. J. Computer Vision, vol. 60, no.2, pp. 91110, 2004.
11. Cynthia L. Ogden, Margaret D. Carroll, Brian K. Kit and Katherine M. Flegal. ”Prevalence of Obesity in the United States”.