International Journal on Recent and Innovation Trends in Computing and Communication
Volume: 3 Issue: 3
ISSN: 2321-8169
973 - 978

Large Scale Learning for Food Image Classification

Abbirami R S[1], Abhinaya A[2], Kavivarthini P[3], Rupika T[4]
Department of Information Technology
Panimalar Engineering College
Chennai.
abbi.nikki@gmail.com[1], abhi.abhinaya1993@gmail.com[2], kavivarthini@gmail.com[3], rupikagnanam@gmail.com[4]

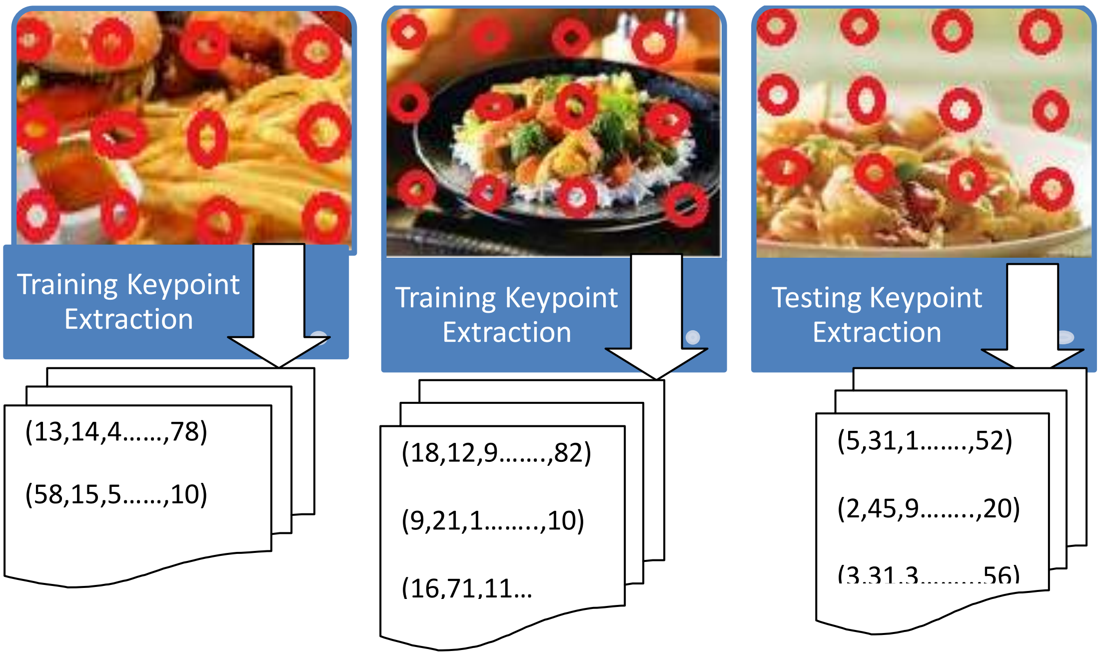
***Abstract*:-** Since health care on foods is drawing people's attention recently, in this paper we propose a computer vision based food recognition system could be used to estimate food for diabetes patients. This study proposes a methodology for automatic food recognition, based on the Bag of Features (BoF) model. We present an approach to find out the group and location of objects in images. The system computes dense local features using scale invariant features. It performs very fast classification of each pixel in an image. For the design and valuation of the proposed system, a image dataset with nearly 5010 food images was created and organized into 11 classes. This system has achieved the accuracy of 78%.

*****

I. INTRODUCTION:

The increased number of people affected by diabetics worldwide, together with their proven inability to evaluate their diet correctly raised the need to develop systems that will support T1D patients during intake of food. To extreme, a wide range of mobile phone applications have been projected in the literature, ranging from interactive diaries [7] to dietary monitoring based on on-body sensors [8].Object detection is one of the core problems in computer vision, and it is a very extensively examine topic. Due to appearance inconsistency caused for example by gentility, surroundings disorder, variations in viewpoint, directions, it is a hard problem. A food recognition application was introduced by Shroff et al. [9] for the classification of fast-food images into various classes. For each food item, a quantity of color (normalized RGB values), size, texture, form and background features is computed and fed to a feed-forward Artificial Neural Network (ANN).Color descriptors have been used to improve illumination invariance and discriminative power. The color invariant descriptors in the context of image category recognition are required though many descriptors exist.

The pixel intensities, the color components and the Gabor filter responses features was used by Zhu et al. [10], together with a Support Vector Machine classifier, for the recognition of many food classes, leading to a recognition rate of the order of 94% for food replicas and 58% for real food items. Kong et al. [11] proposed the use of Scale-Invariant Feature Transform (SIFT) features clustered into visual words and fed to a simple Bayesian probabilistic classifier that matches the food items to a food database containing images of junk food, and various types of fruits and vegetable.



ISSN: 2321-8169
973 - 978

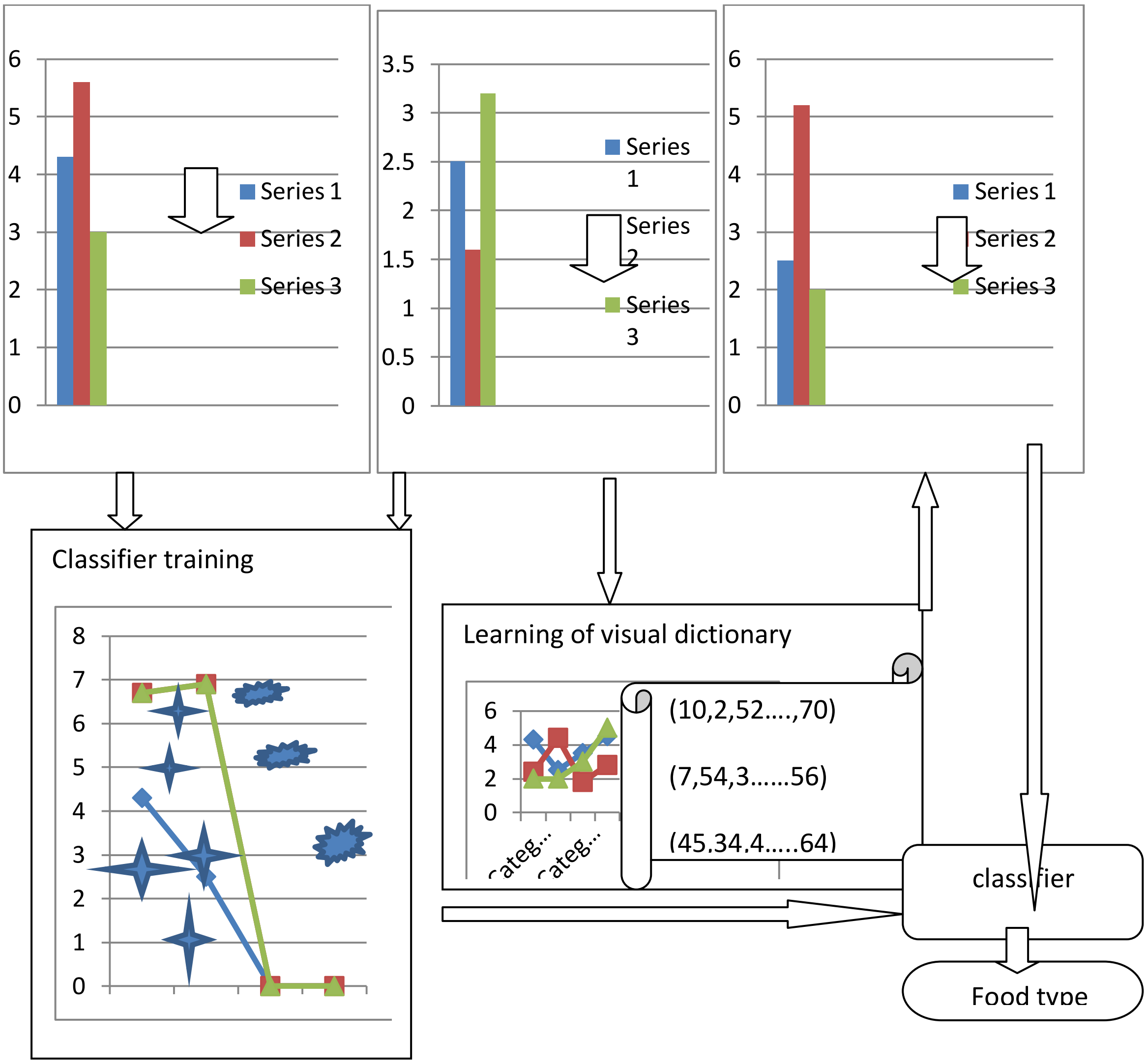


Fig.1.Architecture of the proposed large scale learning for food image classification

.

II. RELATED WORK

Sampling of food image: They evolved from texton methods in surface analysis. The essential idea is to handle images as loose collections of separate patches, sampling a envoy set of patches from the images, evaluating a visual descriptor vector for each patch separately, and using the resultant allocation of samples in descriptor space as a characterization of the image. Dense sampling is a sampling of food image in a grid by grid manner. Random sampling is based on the random selection of point coordinates. Dense sampling has the best performance among all the strategies but with high spatial and process of computing complexity, random sampling gives better results than other sparse sampling methods.

Fig2.sampling technique(a)Dense sampling (b)Random sampling.

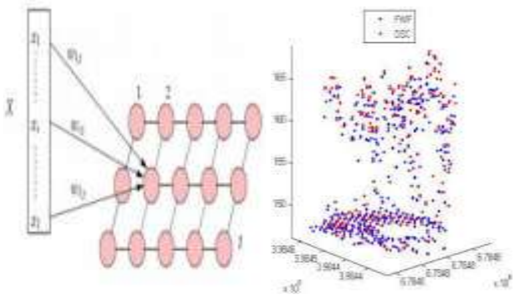


Fig3.Graph (a) Dense sampling (b) Random sampling.

Color Descriptors and Invariant Properties: In this part, color descriptors are on hand and their invariance properties are summarized. First, color descriptors based on histograms are discussed. After that, color moments and the color invariants are available. In conclusion, the descriptors were based on SIFT.

RGB histogram: It is a combination of three 1-D histograms based on the R, G and B channels of the RGB color space and it doesn't possesses any invariance properties.

Opponent histogram: It is a combination of three 1-D histograms based on the channels of the opponent color space:

$$\begin{pmatrix} O1 \\ O2 \\ O3 \end{pmatrix} = \begin{pmatrix} \frac{R-G}{\sqrt{2}} \\ \frac{R+G-2B}{\sqrt{6}} \\ \frac{R+G+B}{\sqrt{3}} \end{pmatrix}.$$

The intensity is represented in channel O3 and the color information is in channels O1 and O2. Due to the subtraction in O1 and O2, the balance will be cancelled out if they are equal for all channels (e.g. a white light source). Therefore, these color models are shift-invariant with respect to light intensity.
The channel O3 has no invariance properties. The intervals between the histograms for the opponent color space have ranges different from the RGB model. In the HSV color space, the hue histogram is recognized that the hue becomes unbalanced around the grey axis. To this end, Van de Weijer et al. [21] applied an error analysis to the hue. The analysis shows that the certainty of the hue is inversely proportional to the saturation. Therefore, the hue histogram is made more robust by weighing each sample of the hue by its saturation. The H and the S color models are scale invariant and shift-invariant with respect to light intensity.
rghistogram: In the normalized RGB color model, the chromaticity components r and g describe the color information in the image (b is redundant as r + g + b = 1):

$$\begin{pmatrix} r \\ g \\ b \end{pmatrix} = \begin{pmatrix} \frac{R}{R+G+B} \\ \frac{G}{R+G+B} \\ \frac{B}{R+G+B} \end{pmatrix}.$$

Because of the normalization, r and g are scale-invariant and thereby invariant to changes in light intensity, and shading [9].The distorted colors distribution of an RGB histogram is not invariant to changes in light conditions. Though, by normalizing the pixel value distributions, the scale-invariance and the shift-invariance is achieved. Because each channel is independently normalized and the descriptor is also normalized against changes in color and arbitrary offsets:

$$\begin{pmatrix} R' \\ G' \\ B' \end{pmatrix} = \begin{pmatrix} \frac{R-\mu_R}{\sigma_R} \\ \frac{G-\mu_G}{\sigma_G} \\ \frac{B-\mu_B}{\sigma_B} \end{pmatrix},$$

with _C the mean and _C the standard deviation of the distribution in the channel. This yields for every channel a distribution where _ = 0 and _ = 1.

SIFT descriptor:

Now key point descriptor is created. A 16x16 neighborhood around the key point is taken. It is divided into 16 sub-blocks of 4x4 sizes. For each sub-block 8 bin orientation histogram is created. So a full of 128 bin standards are existing. It is represented as a vector to form key point descriptor. It also used to identify the intensity gradient orientation.

Color SIFT descriptor:

HueSIFT:

Van de Weijer et al. [21] introduce a concatenation of the hue histogram with the SIFT descriptor. When compared to HSV-SIFT, the practice of the weighed hue histogram addresses the unsteadiness of the hue around the grey axis. Because the bins of the hue histogram are independent, there are no problems with the periodicity of the hue strait for HueSIFT when comparable to the hue histogram. However, only the SIFT component of this descriptor is invariant to clarification color changes or shifts the hue.

HSV-SIFT:

Bosch et al. [2] compute SIFT descriptors over all three channels of the HSV color model, in its place of over the strength channel only. This gives 3x128 dimensions per descriptor, 128 per channel. Drawback of this approach is that the unsteadiness of the hue for low diffusion is ignored. The properties of both channel H and S also applied to this descriptor: it is scale invariant and shift-invariant. However, the H and the S SIFT descriptors are not invariant to brightness color changes; only the strength SIFT descriptor is invariant to this. As a result, the descriptor is only somewhat invariant to light color change.
rgSIFT: Intended for the rgSIFT descriptor, descriptors be added for the r and g chromaticity workings of the normalize RGB color model from eq. (11), which is already scale invariant. since the SIFT descriptor use derivative of the input channels, the rgSIFT descriptor becomes shift invariant as well. However, the color part of the descriptor is not invariant to change in lighting color.

Clustering Technique:

Clustering means the division of a dataset into a number of groups such that similar items falls or belong to same groups. In direct to come together the database, K-means algorithm employs an iterative approach.

K-means clustering algorithm:

K-means clustering is a well recognized partition technique. In this objects are classified as belonging to solitary of K-groups. The consequences of partition method are a set of K clusters, everything of data set belong to solitary cluster. In every cluster there might be a centroid or a cluster envoy. In case where we think real-valued data, the arithmetic mean of the quality vectors for all substance within a cluster provides an appropriate envoy.

4. Hierarchical Clustering [3]

Hierarchical methods are well known clustering technique that can be potentially very useful for a range of data mining tasks. A hierarchical clustering plan produces a sequence of clusterings in which each clustering is nested into the next clustering in the succession. Since hierarchical cluster is a greedy search algorithm base on a local search.

Classifiers:

SVM:

Support vector machines are worn to analyze data and recognize patterns for classification.

In order to recognize the appropriate classifier for the specific problem, several experiments with three classification methods were conducted: SVM, ANN and Random Forests (RF).

SVM [27] with linear or non-linear kernels constitutes the most common classifier with the BoF approach. The next three kernels were worn in the experiments of the in progress study:

Linear:

$linear(\mathbf{x1},\mathbf{x2}) = \mathbf{x1}^T * \mathbf{x2}$ (7)

RBF:

$k_{RBF}(\mathbf{x1},\mathbf{x2}) = \exp(-\gamma * \|\mathbf{x1}-\mathbf{x2}\|^2)$ (8)

Exponential X2:

$k_{x2}(\mathbf{x1},\mathbf{x2}) = \exp - \gamma 2 * \Sigma (x1i - x2i)2 x1i + x2$ (9)

Where **x1** and **x2** are feature vectors and γ is a scaling factor that requires being tune. Thus, the following SVM are tested

ANNs also include popular machine data models for solving versatile computer vision problems [28]. Two different feed-forward ANN model were used through the current study: a linear *ANN*nh and a non-linear with one hidden layer - *ANN*wh. *ANN*nh be knowledgeable using the easy gradient-descend algorithm, while *ANN*wh used the scale conjugate gradient back-propagation algorithm [28]. Frequently, the conjugate gradient back-propagation algorithm leads to faster junction to better minima than pattern steepest descent methods [29]. Both ANNs are fully associated, with early weights randomly particular in the range [-1.0, 1.0]. As launch function, the saturated linear was used for the output layer and the hyperbolic tangent sigmoid for the hidden layer of the second system. The *ANN*wh topology and the internal parameters be resolute using a trial-and-error process.

RFs have happened to popular because of their too much efficiency, the ease of training and their capability to give approximation of the variables significance. They are a set of decision trees such that every tree depends on the values of a random vector sampled independently with the same allocation for all trees in the forest [30]. The forest chooses the categorization having the popular of votes over all the trees in the forest. In this study, one RF was used for the experiment consisting of 31 trees, through each split randomly choosing a figure of features equal to the squared root of the total number of features.

III. EXPERIMENTAL SETUP

A. Dataset Preprocessing:

Intended for the tentative wants of the system developed a dataset of 5000 color images be created by collecting images from the system.

Input: click to choose image files from dataset

Fig.4.model images of the foodstuff dataset.

Output: view selected image.

 ISSN: 2321-8169
973 - 978

B. Key point mining:

Keypoints are chosen points on an image that describe the centers of local patches where descriptors will be pertain.

Input: choose key points from images.

Fig.5. (a) SIFT Sampling, (b) Dense sampling ,(c) Random sampling.

Output: selected key points.

C.Color evaluation:

A 16x16 pixel region is divided into 4x4 sub-regions as well as for each one of them an 8-bin histogram of the intensity gradient orientation is computed, leading to a 128-dimensional feature vector.

Input: choose image.

Output: classify RGB values.

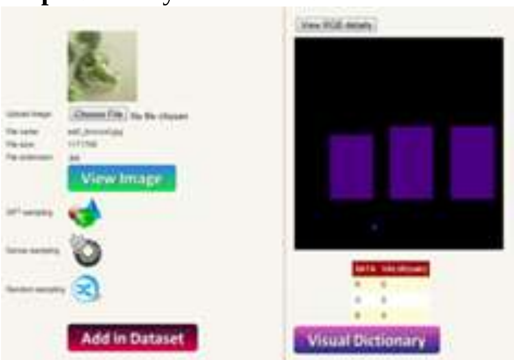


Fig.6.represents feature description

D.Classification of food images:

Food image classification level undergoes training and testing stage. In this classification, the methods which are used are SVM, ANN, and RF.

Input: analyze descriptor values to food type.

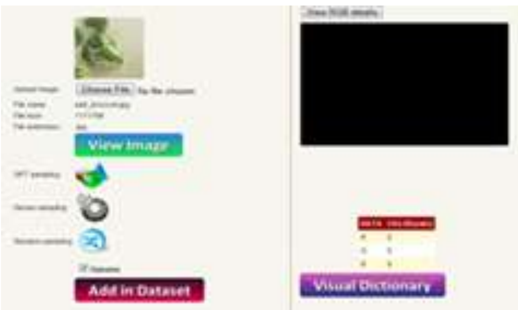


Fig.7. shows the design for identifying food image content which will match to predefined food list contents

Output: cluster images & stored.

IV. PERFORMANCE MEASURES

In the bag of words technique, it is found that k-means produces more delegate small dictionaries though, as the number of visual words considered increases, while *hk*-means provides the same results while at the same instance it decreases the computational cost in both training and testing. This significant improvement is due to the tree structure of the *hk*-means dictionary which consequences in more efficient vector quantization

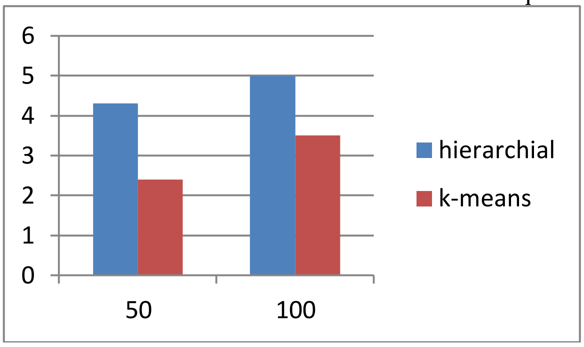


Fig.8. Represents consumption of time by hierarchical k-means and k-means clustering.

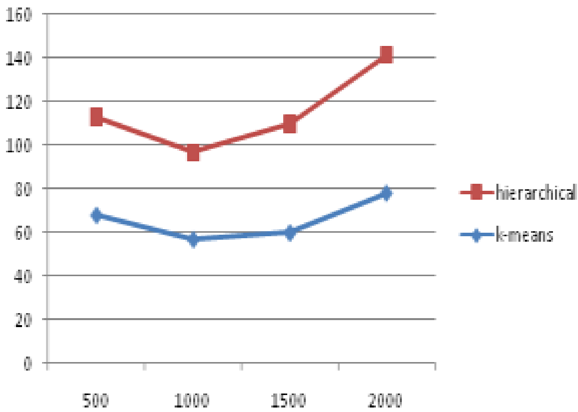


Fig.9. represents the accuracy of clustering images

V. SUMMARY AND CONCLUSION

In this paper, we intend an automatic food recognition system that could be used to calculate approximately food for diabetes patients. The Bag of features technique is used which are more appropriate and accurate for the result. The experiments were conducted on the foodstuff image dataset with 5000 images of food belonging to 11 dissimilar food classes. The experiment undergoes a sequence of five major experiments for choosing and optimizing the concerned components and parameters of the system.

While earlier concepts require a detailed manual annotation of the images in the training database, the proposed model can learn characteristic intermediate "themes" of scenes with no supervision, nor human intervention and achieves comparable performance.

In the first experiment, the key points are extracted from the food images using SVM, ANN and RF methods. The second experiment investigated the result of the descriptor's size on the final performance. The best results were obtained by the descriptor combination with sizes of 16, 24 and 32. By using various sizes of descriptors, the BoF system gained multi-

resolution properties that improved the final performance, since the food scale may vary among the food images.
Followed by, the hsvSIFT was selected amongst fourteen different color and texture descriptors which gives the best results. hsvSIFT constitutes a differential descriptor which gives the local texture in all the color channels of the HSV color space. This fact enables it to include other information, apart from texture, and we should keep some invariances in intensity and color changes. The Bag of words was determined to be around 10,000, because smaller number of words resulted in visibly poorer results and additional words did not improve the performance.
The optimized system achieved overall identification accuracy in the order of 78%, proving the feasibility of a BoF-based system for the food recognition problem. The enrichment of the dataset with additional images will improve the categorization rates, especially for the classes with high variety. The system will additionally include a food segmentation phase before applying the proposed recognition module, so that images with multiple food types can be identified. For the future work, apply rating for the foodstuff which will be more effective while consuming food.

VI. ACKNOWLEDGEMENT

The authors would like to thank our guide Mrs. A.Vinodhine, our project coordinator Mrs.Dr.K.Karunkuzali and our Head of the department Mrs.Dr.M.Helda Mercy for their support in the creation of the visual dataset and the definition of the various food types.

REFERENCES:

[1] Navjot Kaur, Jaspreet Kaur Sahiwal, Navneet Kaur" *Efficent K-means Clustring Algorithm Using Ranking Method In Data Mining*"ISSN: 2278 – 1323 International Journal of Advanced Research in Computer Engineering & Technology Volume 1, Issue 3, May2012.

[2] Kehar Singh , Dimple Malik and Naveen Sharma " Evolving limitations in K-means algorithm in data mining and" IJCEM International Journal of Computational Engineering & Management, Vol. 12, April 2011.

[3] C. E. Smart, K. Ross, J. A. Edge, C. E. Collins, K. Colyvas, and B. R. King, "Children and adolescents on intensive insulin therapy maintain postprandial glycaemic control without precise carbohydrate counting," *Diabetic Medicine*.

[4] C. E. Smart, B. R. King, P. McElduff, and C. E. Collins, "In children using intensive insulin therapy, a 20-g variation in carbohydrate amount significantly impacts on postprandial glycaemia," *Diabetic Medicine*,

[5] M. Graff, T. Gross, S. Juth, J. Charlson, "How well are individuals on intensive insulin therapy counting carbohydrates?," *Diabetes Research and Clinical Practice*,

[6] F. K. Bishop, D. M. Maahs, G. Spiegel, D. Owen, G. J. Klingensmith, A. Bortsov, J. Thomas, and E. J. Mayer-Davis, "The carbohydrate counting in adolescents with type 1 diabetes (CCAT) study," *Diabetes Spectrum*.

[7] C.E. Smart, K. Ross, J.A. Edge, B.R. King, P. McElduff, C.E. Collins, "Can children with type 1 diabetes and their caregivers estimate the carbohydrate content of meals and snacks?" *Diabetic Medicine*.

[8] O. Amft, G. Tröster, "Recognition of dietary activity events using on-body sensors," *Artificial Intelligence in Medicine*,

[9] G. Shroff, A. Smailagic, and D. P. Siewiorek, "Wearable context-aware food recognition for calorie monitoring," in *12th IEEE International Symposium on Wearable Computers*,

[10] F. Zhu, M. Bosch, I. Woo, S. Y. Kim, C. J. Boushey, D. S. Ebert, and E. J. Delp, "The use of mobile devices in aiding dietary assessment and evaluation," *IEEE Journal of selected Topics in Signal Processing*

[11] F. Kong and J. Tan, "DietCam: Automatic dietary assessment with mobile camera phones," *Pervasive and Mobile Computing*.

[12] J.N. Matthews, W. Hu, M. Hapuarachchi, T. Deshane, D. Dimatos, G. Hamilton, M. McCabe, and J. Owens, "Quantifying the Performance Isolation Properties of Virtualization Systems," Proc. Workshop Experimental Computer Science (ExpCS '07), 2007.

[13] Amazon Elastic Compute Cloud,2012.

[14] D. Milojicic, I.M. Llorente, and R.S. Montero, "Opennebula: A Cloud Management Tool," IEEE Internet Computing, vol. 15, no. 2, pp. 11-14, Mar./Apr. 2011.

[15] S. Boyd and L. Vandenberghe, Convex Optimization. Cambridge Univ. Press, 2009.

[16] E. Imamagic, B. Radic, and D. Dobrenic, "An Approach to Grid Scheduling by Using Condor-G Matchmaking Mechanism," Proc. 28th Int'l Conf. Information Technology Interfaces.