CLUSTERING MEAL IMAGES IN A WEB-BASED DIETARY MANAGEMENT SYSTEM

Anonymous ICME submission

Workshop on Multimedia Services and Technologies for E-health
Paper ID 45

ABSTRACT

We investigate the possibility of improving interactions in a multi-user, image-based dietary assessment system. A segmentation algorithm detects food regions in each meal image. Pairwise matching of images based on color and texture information of these regions forms a similarity matrix of the image collection. We cluster the nodes of the graph constructed using this matrix, to identify natural groupings of meal images according to their content. Representative images from these clusters form summaries of the large image collection. We conduct a user study to evaluate the effectiveness of the proposed algorithms in summarizing meal image collections, and report the results.

Index Terms— Meal summary, FoodLog, segmentation, clustering, meal image analysis

1. INTRODUCTION

With increased attention to dietary control in the field of healthcare, there has been a rapid growth in research and development of innovative tools for supporting dietary assessment [1]. There are several such tools that can be used on computers and mobile phones. Some of these also use photos of meals, taken with camera phones or digital cameras, to make recording dietary data easier [2, 3].

Most of these tools function as personal information systems. They allow a user to store data regarding his/her dietary behavior, for his/her own use or to be shown to a health expert during consultation. Some others are Internet based systems that have a large number of users. Such systems also allow users to interact with each other by forming groups, sharing information, encouraging each other, and sometimes competing with each other to achieve goals such as weight loss.

A Multi-user online dietary assessment tool can accumulate, with time, a large amount of information regarding people's dietary behavior. Such information can be used in a number of ways such as making meal recommendations to users, identifying dietary trends and patterns, etc. However, for such data to be useful, there should be techniques for automatically analyzing and classifying the data. This is a challenging task because data submitted by users to Internet-based services can be incomplete, disorganized, and even inaccurate. Data analysis is particularly difficult when users submit images, instead of text or menu items, as inputs.

This paper presents our work on analyzing food images from a web based dietary information system called *FoodLog* [4]. A FoodLog user records his dietary information by submitting photos of his/her daily meals, taken with a digital camera and/or a camera phone. The dietary balance of the meals in these images is calculated according to the "Food Balance Guide" by the Ministry of Agriculture, Forestry and Fisheries of Japan [5]. It categorizes food into five dietary components; grains, vegetables, meat/beans, fruit, and dairy products. The quantities of these components are measured in "servings (SV)." The results are stored together with other image metadata such as date, time and location (if available). A user can access FoodLog via a web browser and view the recorded information in different formats such as graphs, calendars and maps. Such visualizations help a user to keep track of the balance and variety of his dietary intake and correct any off-balance trends.

While FoodLog has approximately 1500 registered users and more than 60,000 images, there is little interaction among different users. A user can see meal images submitted by others, organized on a map or a grid, according to the time and location of image capture. However, more meaningful visualizations such as "other users who had similar meals" or "a summary of what *FoodLoggers* ate today" will create more interest and interaction. De Silva et al. [6] demonstrated that communities of users who take similar meals can be formed by clustering the dietary balance data. However, they also reported that errors in the dietary balance estimation can drastically affect clustering.

In this paper, we propose to use image features directly to find similarities between meal images and use the results for improved visualizations and interactions on FoodLog. We first detect image regions corresponding to food, and extract features that represent the regions. A pairwise matching algorithm compares the meal images and creates a similarity matrix for the image collection. A clustering algorithm extracts natural groupings in this collection. We demonstate how the results of clustering can be applied to summarize a collection of meal images, for improved visualization. We also evaluate the proposed techniques with a user study, and report the results.

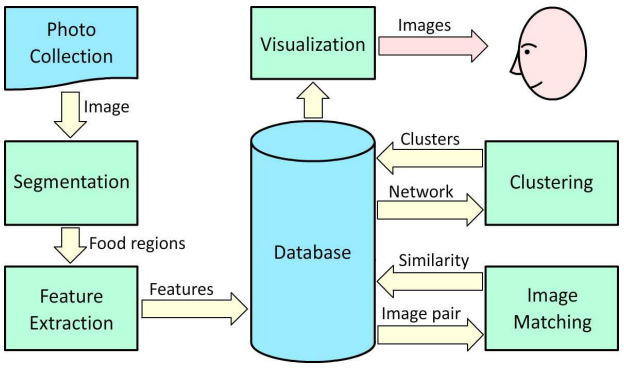

Fig. 1. System overview.

2. SYSTEM DESCRIPTION

Figure 1 shows the functional outline of the proposed system. Images uploaded to FoodLog are segmented to identify regions corresponding to different types of food. A set of features extracted from these regions are stored in a database. A pairwise matching algorithm compares these features to generate a similarity graph for the image collection. A clustering algorithm forms groups of images from which images are selected for summarization and visualization. The following subsections describe these functions in detail.

2.1. Food region detection

A typical meal image consists of regions corresponding to defferent types of food, crockery and cutlery, glassware, and background regions (such as trays, table surfaces etc). We analyzed a collection of approximately 32,000 meal images to identify characteristics that distinguish food regions from the rest. The following were the dominant characteristics of most types of food regions:

- A set of colors that are quite similar to each other
- A shape that is close to a circle (due to arrangment for serving)
- Presence of a large amount of irregular edges
- location at the center of a non-food region (e.g.: plate)

Based on the above characteristics, we designed the following algorithm for segmenting food regions. The images were first segmented using pyramidal segmentation based on similarity of pixels in RGB color space. This partitions the image into a set of non-overlapping regions. The arrangement and the presence of irregular edges are represented by the number and size of the circles detected by applying Hough transform on the gray-scaled original image. The results of these processes are combined to determine the regions that correspond to food. The parameters of pyramidal segmentation and Hough transform were emprically determined using the image collection mentioned above.

Figure 2 demonstrates the result of image segmentation on three different images. It is evident that some regions can be split to more than one segment, and some parts of food regions might not get segmented. However, we consider the results sufficiently good for feature extraction.

It should be noted that we purposely avoided supervised learning approaches based on low level features, for food region detection. While 30,000 meal images is a fairly large collection, it comes from a smaller number of users most of whom are from Japan. Therefore, supervised learning can lead to the risk of overfitting classifiers to the dietary behavior of these users. Edge irregularity and similarity of colors, on the other hand, are more general and less likely to cause overfitting.

2.2. Feature extraction

We select size, color and texture as the features to represent each food region. The size of a food region is determined by dividing the number of pixels in the corresponding region by the total image area in pixels. Color is represented using a three dimensiobal RGB color histogram of the region. A total of 512 bins ($8\times8\times8$) with 8 levels for each color is used. Texture is represented using a 256-bin Local Binary Pattern (LBP) histogram [7]. We selected LBP histogram due to its relative robustness to rotation and scaling.

Sometimes, one type of food can be partitioned to multiple regions due to the presence of some other food item on it. This results in multiple food regions with nearly identical features, and can cause errors in finding similar images. Therefore, food regions are matched against each other and merged if they have a more than 90% overlap in both texture and color histograms. The collection of the features for all food regions in an image forms the feature vector for the image. This vector is stored in a database for similarity matching, which is described in the next step.

2.3. Similarity matching

Based on the features we extracted, we establish rules for determining the similarity between two meals. Two meals are considered "similar" if the number of food regions in the corresponding images are similar, and features of the regions in one image have a 1:1 mapping with those of the other image. We also want the similarity to be a real number between 0 and 1, with 0 denoting completely different meals and 1 denoting identical meals. With the above rules in mind, we define the similarity between two meals as follows.

Consider a collection of images R. For two images I_p and I_q where $p, q \ni R$, we define similarity between I_p and I_q as

$$S_{pq} = \frac{2N_{pq}}{N_p + N_q}$$

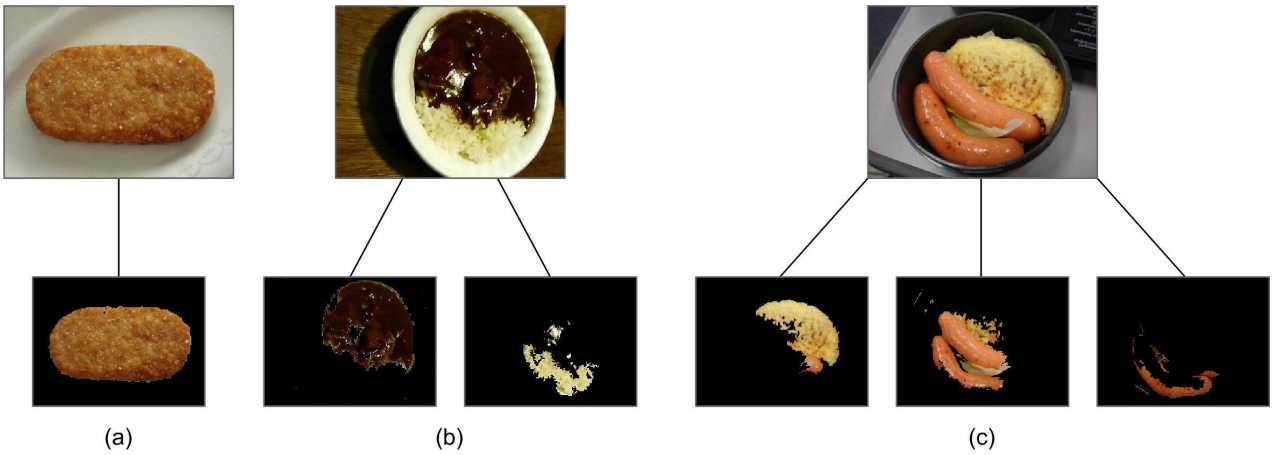

Fig. 2. Example results of food region detection.

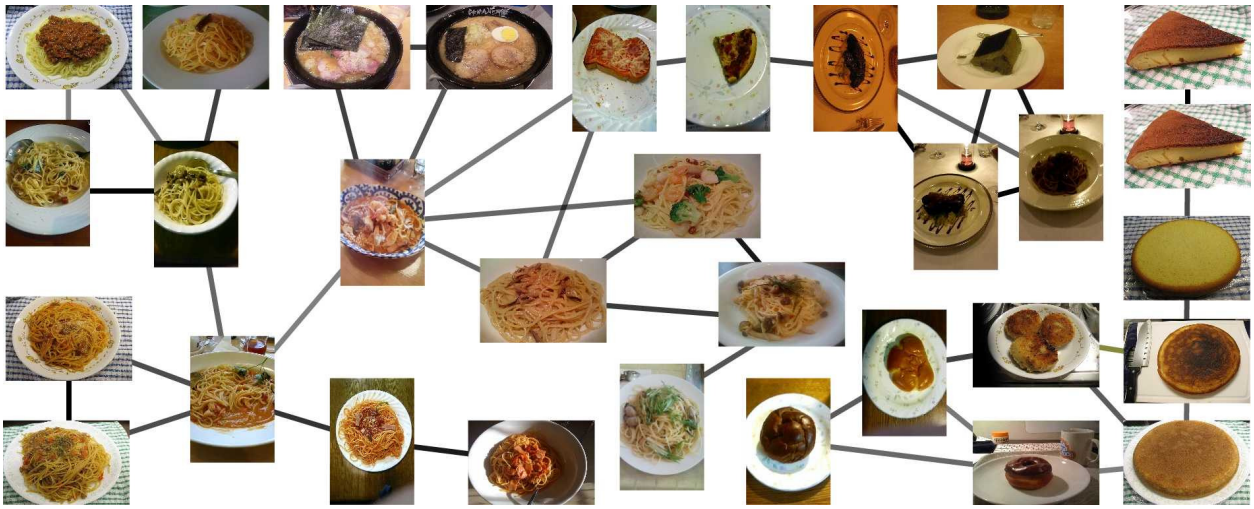

Fig. 3. An example similarity graph.

where N_p and N_q are the numbers of food regions in I_p and I_q respectively and N_{pq} is the number of matching food segments between I_p and I_q.

Two food regions R_i and R_j are considered similar if

$$S^C(i,j) > 0.4$$

$$S^T(i,j) > 0.4$$

and

$$wS^C(i,j) + (1-w)S^T(i,j) > 0.7$$

where $S^C(i,j)$ and $S^T(i,j)$ are the proportions of overlap in color and texture histograms between R_i and R_j. The threshold values were empirically determined using aproximately 500 images. We set w to 0.5 so that the relative influence of texture and color on similarity is equal.

The algorithm for calculating is somewhat similar to that proposed by Jing et al. [8]. SIFT features are used there, instead of color and texture features. While using SIFT features is common in image matching algorithms, we observed that they do not perform well in matching food regions.

2.4. Clustering

The result of similarity matching is a symmetric square matrix $S = [S_{ij}]_{N \times N}$ where N is the number of images used for matching. The non-diagonal elements S_{ij} of the matrix correspond to the similarity between images i and j. This matrix can also be visualized as a bi-directional graph where the N images are nodes and the non-diagonal elements are weighted edges between nodes. Figure 3 shows an example graph created by matching 31 meal images with each other. Darker edges correspond to higher similarity. Edges corresponding to similarity less than 0.3 have been removed for clarity.

We intend to form clusters of similar meal images, so that we can find natural groupings among them. While there are several methods for clustering, CNM algorithm [9] is the most

Fig. 4. Sample results of clustering.

widely used. We use the modified CNM algorithm proposed by Wakita and Tsurumi [10] due to its faster performance with large datasets. The colored areas containing the images in Figure 4 indicate the clusters formed by applying the above algorithm to a collection of images. It is evident that the images have a natural grouping, according to the types of food. However, there are some misclassifications due to errors in segmentation and images with a high degree of compression. Further, the results are hard to observe once the technique is applied to a large collection of images.

2.5. Summarization

Clustering meal images according to their similarity can be used as a starting point for many applications in a dietary assessment system. Clustering makes it possible to summarize a large collection of images. It is possible to recommend meals to a user by selecting meal images from different clusters that are more likely to match his/her dietary patterns. Since each meal image is owned by a user, clustering meal images provides a means of connecting users; for instance, to suggest friends and/or groups.

In this work, we apply the results of clustering to summarize the collection of images uploaded by FoodLog users. At the current state, the users can see the most recently uploaded meal images on a grid or a map (if location data are available). This leads to less interesting results, such as a large number of similar images, especially just after meal times. We intend to use the results of clustering to create a more diverse and interesting summary of meal images.

We propose to create a summary by selecting a number of *representatives* from each cluster. The number of representatives can vary according to the number of clusters created and the size of the summary that is required. An image is considered to represent its cluster well if it is similar to many members of the same cluster. We use the degree centrality of the node represented by the image in the subgraph corresponding to the cluster, as the measure of representativenes. If the size of the summary N_s is smaller than the number of images, the best representatives from the largest N_s clusters are selected. If N_s is larger than the number of clusters, the best representatives from each cluster are selected, in proportion to the cluster sizes. Figure 5 shows the representatives selected from the clusters in Figure 4. In this example, one representative from each cluster has been selected.

Fig. 5. The summary of images displayed in Figure 4.

3. EVALUATION

We conducted a user study on selecting representative images from a collection of meal images, with the following objectives:

- Investigate whether ordinary Internet users have a common agreement in selecting representative images from clusters of photos of similar meals
- Evaluate the effectiveness of the proposed technique for representative selection
- Evaluate the effect of using the clusters for summarizing meal photo collections for visualization
- Evaluate the effectiveness of food image segmentation
- Identify possible improvements and future directions

The study consisted of four sections, each containing three repetitive tasks. The following is a description of the tasks in different sections:

- **Section 1:** In each task, the participant looks at a set of meal images, and selects three representatives from each set. Each set consists of three categories of meals. The responses from multiple participants will indicate whether there is common agreement among different persons when it comes to selecting representatives from meal images.
- **Section 2:** In each task, a participants looks at a set of meal images and selects up to three representatives from each set. Each set consists of one category of food (for instance, spaghetti). They can rank the three according to the order of preference (i.e., 1, 2, and 3). The responses from multiple participants will indicate whether there is common agreement among different users when it comes to selecting and ranking representatives from FoodLog images.
- **Section 3:** In each task, the participant looks at a set of meal images, and selects two representatives for each set from 6 candidates. They can rank the representatives according to the order of preference (e.g.: 1, 2). The candidates are selected as
  - the first two representatives selected using the proposed method,
  - the first two representatives selected by matching visual features without segmentation of food regions, and
  - two randomly selected images.

  The responses will indicate the effectiveness of food region detection and the proposed method for representative selection.
- **Section 4:** In each task, the participant looks at a set of meal images, and four candidate summaries for that set of images. Thereafter he/she selects the best summary for each set, according to his/her opinion. The four summaries are created by
  1. collecting one representative from each cluster using the proposed method,
  2. collecting one representative from each cluster using non-segmented images,
  3. randomly selecting the same number of images as in summary 1, and
  4. randomly selecting the same number of images as in summary 2

  The responses will indicate the effectiveness of the proposed methods for food region detection and summarization.

Instructions on how to complete each section were given in both verbal and written forms, at the start of the section. One of the authors was present during the entire experiment, to provide additional explanations if needed.

The sets of images contained 24 to 96 images, all extracted from actual data uploaded by FoodLog users. Each set was arranged in tabular format. The positions of images weres changed for different participants, to prevent biases towards relative positioning of the images.

Sixteen persons in the age range of 22 to 30 years participated in the user study. The participants were given book vouchers as tokens of appreciation for taking part. The participants took 17 to 40 minutes to complete the study. The average time for completing the experiment was appproximately 22 minutes. This time included breaks in between sections and explanations on how to complete each section.

4. RESULTS

The results of Sections 1 and 2 showed that the participants found a common agreement in selecting representative images. In each task in Section 1, 25% of the images were selected by 64% of the participants. The results did not show any bias towards the placement of images, or the photographic quality of images. However, most of the images were of medium photographic quality and therefore it cannot be assumed whether the selection is independent of the same.

In Section 2, 25% of the images were selected by 53% of the particpiants. However, the participants mostly disagreed on ranks assigned to images. The result showed that it is hard for a user to rank representative images in a meal image collection by observation.

In Section 3, representatives selected using the proposed method received 59% of the votes. Representatives selected

by clustering meal images without food region detection received 23% of the total vote. This result demonstrates that food region detection before feature extraction leads to better selection of representatives.

In Section 4, the summaries created using the proposed method received 52% of the total votes. The summaries created by clustering images without food region detection received 25% of the votes. It is evident that better summaries can be created using the proposed method.

However, there was another interesting finding in the results in Sections 3 and 4. In each section, there was one task for which there were a slightly higher number of votes for images selected without food region detection. Close examination of the results found the reason to be the inclusion of candidate images with fairly low photographic quality was the reason for this result. Therefore, it is evident that photographic quality should also be taken into account when summarizing meal image collections.

5. CONCLUSION

We proposed a set of techniques for clustering and summarizing a collection of meal images from a multi-user dietary assessment system. The proposed segmentation algorithm was able to detect food regions in a meal image with reasonable accuracy. Pairwise similarity matching using color and texture features followed by clustering showed a natural grouping of images. The results of the user study indicated that the proposed algorithm for summarizing image collections was able to select appropriate representative images, and create good summaries, with an average approval rating of approximately 56% competing with two other methods.

At the current state, the similarity matrix of meals is symmetric. However, depending on how the results of clustering are used, it may be useful to have an asymmetric similarity matrix for the image collection. We are working on incorporating context data such as meal time and location to improve the clustering results. We also intend to connect users with each other based on clustering results, by identifying users with simiar dietary behavior.

6. ACKNOWLEDGMENTS

Acknowledgments have been removed for anonimity.

7. REFERENCES

[1] Fengqing Zhu, Anand Mariappan, Carol J. Boushey, Deb Kerr, Kyle D. Lutes, David S. Ebert, and Edward J. Delp, "Technology-assisted dietary assessment.," in *Computational Imaging*, Charles A. Bouman, Eric L. Miller, and Ilya Pollak, Eds. 2008, vol. 6814 of *SPIE Proceedings*, p. 681411, SPIE.

[2] Keigo Kitamura, Toshihiko Yamasaki, and Kiyoharu Aizawa, "Foodlog: capture, analysis and retrieval of personal food images via web," in *Proceedings of the ACM multimedia 2009 workshop on Multimedia for cooking and eating activities*, New York, NY, USA, 2009, CEA '09, pp. 23–30, ACM.

[3] SungYe Kim, TusaRebecca Schap, Marc Bosch, Ross Maciejewski, Edward J. Delp, David S. Ebert, and Carol J. Boushey, "Development of a mobile user interface for image-based dietary assessment," in *Proceedings of the 9th International Conference on Mobile and Ubiquitous Multimedia*, New York, NY, USA, 2010, MUM '10, pp. 13:1–13:7, ACM.

[4] Foo.log Inc., "Foodlog, http://www.foodlog.jp," May 2010.

[5] Forestry Ministry of Agriculture and Fisheries of Japan., "Food balance guide, http://www.maff.go.jp/j/balance guide/b_about/index.html.," July 2007.

[6] Gamhewage C. de Silva and Kiyoharu Aizawa, "Image-based dietary information mining for community creation in a social network," in *Proceedings of second ACM SIGMM workshop on Social media*, New York, NY, USA, 2010, WSM '10, pp. 53–58, ACM.

[7] Timo Ojala, Matti Pietikinen, and Topi Menp, "A generalized local binary pattern operator for multiresolution gray scale and rotation invariant texture classification," in *in Proc. of Second Inter. Conf. on Advances in Pattern Recognition, Rio de Janeiro*, 2001, pp. 397–406.

[8] Yushi Jing and Shumeet Baluja, "Visualrank: Applying pagerank to large-scale image search," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 30, pp. 1877–1890, 2008.

[9] A. Clauset, M. E. J. Newman, and C. Moore, "Finding community structure in very large networks," *Physical Review E*, vol. 69:066111, 2004.

[10] K. Wakita and T. Tsurumi, "Finding community structure in mega-scale socialnetworks," *cs.CY072048*, 2 2004.