21st International Conference on Pattern Recognition (ICPR 2012)
November 11-15, 2012. Tsukuba, Japan

Multiple-Food Recognition Considering Co-occurrence Employing Manifold Ranking

Yuji Matsuda and Keiji Yanai
Graduate School of Informatics, The University of Electro-Communications
Email: matsuda-y@mm.cs.uec.ac.jp, yanai@cs.uec.ac.jp

Abstract

In this paper, we propose a method to recognize food images which include multiple food items considering co-occurrence statistics of food items. The proposed method employs a manifold ranking method which has been applied to image retrieval successfully in the literature. In the experiments, we prepared co-occurrence matrices of 100 food items using various kinds of data sources including Web texts, Web food blogs and our own food database, and evaluated the final results obtained by applying manifold ranking. As results, it has been proved that co-occurrence statistics obtained from a food photo database is very helpful to improve the classification rate within the top ten candidates.

1 Introduction

Recently, personal services to recode people's food habits by taking meal photos with mobile phones have become popular. Currently, labeling food names to meal photos requires human labor, which is a quite troublesome task. Therefore, it is desired to make recording of food items more easier and quickly. To this end, several methods to recognize food images have been proposed so far [1, 2, 3, 4].

Most of the existing works assumed that one meal image contained only one food item. They cannot handle a meal photo which contains two or more food items such as a hamburger-and-french-fries image. Then, in [4], we proposed a new method for recognizing meal photos which contain two or more food items as shown in Figure 1. In our proposed method, firstly, we detect candidate regions with several methods including Felzenszwalb's deformable part model (DPM) [5], a circle detector and the JSEG region segmentation [6]. Then, we extract various kinds of image features from each candidate region. After applying the classification models trained by multiple kernel learning [7], we obtain the names of the top N food item candidates over the given image.

In the proposed method [4], each food item is recognized independently. Meanwhile, in the research of scene recognition the targets of which usually contain multiple objects, relations between objects were considered as important cue for scene recognition in some works [8, 9]. Inspired by these works, we introduce relation information between food items for recognizing multiple-food meal photos. As relations which have been used in object recognition research so far, co-occurrence [8] and relative location [9] are common. In case of meal photos, we think co-occurrence relation is more important than relative locations, because some combinations of foods such as "hamburger and french fries" are very common, while the way to place food items on the table is not strictly restricted in general.

Figure 1. Examples of multiple-food photos.

In this paper, we propose a method to recognize multiple-food meal photos considering co-occurrence statistics by extending our previous work [4]. To do that, we use manifold ranking [10] which have been used as a method for relevance feedback of image retrieval [11]. Manifold ranking is a ranking method to consider similarities between items. In the image retrieval research, manifold ranking was used to rank images considering both user preference and visual similarities between images. In this paper, we use the manifold ranking method to rank food candidates considering both recognition results by our previous method [4] and co-occurrence statistics between food items. That is, the method proposed in this paper re-ranks the original candidate ranking obtained by the multiple-food recognition method, which does not consider co-occurrence statistics, by taking into account co-occurrence statistics. To our best knowledge, this is the first work to recognize multiple-food items in one meal image taking into account co-occurrence statistics.

The rest of this paper is organized as follows: Section 2 explains the proposed method, and Sec-

978-4-9906441-1-6 ©2012 IAPR

tion 3 describes the experimental results. Finally in Section 4 we conclude this paper.

2 Proposed Method

The proposed method refine the results obtained by our previous method [4] with manifold ranking [10]. Before explaining the propose method, we describe the previous method to detect food items for multiple-food photos as shown in Figure 1.

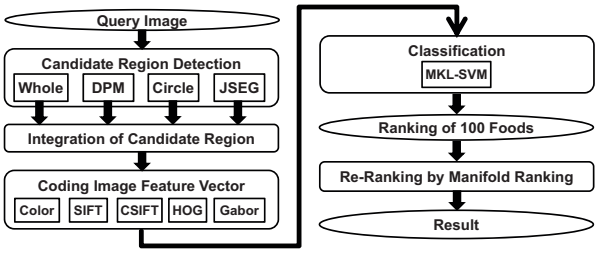

Figure 2. Recognition Flow

2.1 Previous Method

In [4], we proposed a food image recognition system which outputs the names of food items that are expected to be shown in a given meal photo. We show the overview of the processing flow of the proposed system including the co-occurrence extension proposed in this paper in Figure 2. Given an input image, first, the system detects candidate regions of dishes. We use four types of detectors including the deformable part model (DPM) [5], a circle detector, the JSEG region segmentation [6], and whole image. Next, we integrate bounding boxes of the candidate regions detected by the four methods. Then, we check the aspect ratio of width and height of the bounding boxes, and exclude irrelevant bounding boxes regarding their shapes from the candidate set.

Next, the system extracts various kinds of image features including Bag-of-Features (BoF) of SIFT and color SIFT with spatial pyramid, Histogram of Oriented Gradients (HoG), and Gabor texture features from the selected regions, and calculate SVM scores by multiple-kernel learning (MKL) [7] with chi-square RBF kernels for each candidate region in the one-vs-rest manner. MKL can estimate the optimal weights for linear combination of kernels each of which corresponds to one kind of a visual feature.

After that, we obtain SVM output scores for all the candidate regions and all the food categories. Note that our objective is not associating extracted regions with possible names of food items directly, but listing all the names of the food items which are estimated to be shown in the given meal photo. Then, we select the maximum SVM output score over all the candidate region regarding each food category. Finally, we obtain the evaluation scores which express likelihood that the corresponding food items appear in the given photo. As a system output, we obtain the names of the top N food items over the given image regarding the descending order of the evaluation score.

2.2 Manifold Ranking with Co-occurrence Statistics

Some common combinations of food items exist such as "hamburger and french-fries" and "rice and miso-soup", while unlikely combinations exist such as "sushi and hamburger" or "sashimi and french-fries". From these observation, co-occurrence statistics is expected to be able to enhance the performance of multiple-food image recognition. By considering co-occurrence statistics, we can reduce unlikely combinations and boost possible combinations which are included in the higher ranked candidates. As results, obtained ranking of food item candidates become more precise.

For multi-food recognition with co-occurrence statistics, we use manifold ranking [10], which is a re-ranking method to consider similarities between items. In this paper, we propose to re-rank the food candidate ranking obtained by our previous work [4] with the manifold ranking method.

The equation of the manifold ranking is as follows:

$$\boldsymbol{r}^* = (I - \alpha S)^{-1}\boldsymbol{r}, \tag{1}$$

where $\boldsymbol{r}$ and $\boldsymbol{r}^*$ represent the initial ranking vector and the manifold ranking vector, and I and S represent an identity matrix and a similarity matrix, respectively. Note that α is a constant varying from 0 to 1, which adjusts the effect of S for the initial vector $\boldsymbol{r}$.

The initial ranking vector $\boldsymbol{r}$ is calculated by applying a standard sigmoid function to the SVM output value $\boldsymbol{v}_i$ of each category and L1-normalized as shown in the following equation:

$$\boldsymbol{r}_i = \frac{(1+\exp(-\boldsymbol{v}_i))^{-1}}{\sum_j (1+\exp(-\boldsymbol{v}_j))^{-1}}, \tag{2}$$

As a similarity matrix to re-rank the initial ranking, we use a co-occurrence probability matrix, which can be calculated by counting the number of co-occurrence of two food item pairs in a training dataset. Each element of the co-occurrence matrix $S_{i,j}$ can be obtained with

$$S_{i,j} = \frac{c_{i,j}}{\sum_{k \in F \wedge k \neq j} c_{k,j}}, \tag{3}$$

Figure 3. 100 food categories in our dataset. Please see this figure on a PDF viewer with magnification.

where $c_{i,j}$ represents the number of co-occurrence pairs of food category i and j over whole the training dataset, and F represents a set of all the predefined food categories. Note that all the diagonal elements of S are defined as 0.

In addition to a food image database, we also use World Wide Web as knowledge source to calculate co-occurrence matrix S. In [8], in addition to co-occurrence statistics extracted from image database, they proposed constructing co-occurrence matrix with Google Web Search. Following that, we construct co-occurrence matrix using Web data. Note that Rabinovich et al. used conditional random field (CRF) [8], while we use the manifold ranking method to re-rank the initial output.

As a method to estimate co-occurrence of two words from the Web, we use Normalized Google Distance (NGD), which is calculated based on the hit number of Web search results. NGD between word x and word y is given by

$$NGD(x,y) = \frac{\max(\log f(x), \log f(y)) - \log f(x,y)}{\log M - \min(\log f(x), \log f(y))}, \tag{4}$$

where M is the total number of Web pages over whole the Web, and $f(x)$ and $f(y)$ represent the number of hit pages for the search word x and y, respectively. Then, a co-occurrence matrix can be computed as follows:

$$S'_{i,j} = \frac{\exp(-NGD(\mathrm{Name}_i, \mathrm{Name}_j)}{\sum_{k \in F \wedge k \neq j} \exp(-NGD(\mathrm{Name}_k, \mathrm{Name}_j))}, \tag{5}$$

where "Name$_i$" and "Name$_j$" means the name text of food category i and j, respectively.

As a system output, we obtain the names of the top N food items over the given image regarding the descending order of the elements of the manifold ranking vector $\boldsymbol{r}^*$.

3 Experiments

In the experiments, we used our own food image dataset built for [4] as shown in Figure 3 which includes 100 Japanese food categories with bounding boxes on each food item. It contains about one hundred images for each category and 9132 images totally. For the experiments, we selected 500 multiple food-item images from them which contain 1178 food items for test, and used the rest of all the images for training.

To evaluate the performance, we use a classification rate within the top N candidates CR@N regarding food items, which is defined in the following equation:

$$\mathrm{CR@}N = \frac{\text{num. of correctly-detected food items in top N}}{\text{num. of all the food items in all the test image}}$$

If the top N candidates include the names of the food items appearing in the given food image, we count them as the correctly-detected food items.

Firstly, we compare the results before and after applying manifold ranking using the co-occurrence matrix built from the food image dataset. Figure 4 shows the classification rates within the top N candidates. In this figure, "Baseline" represents the initial results obtained by the previous method [4], while "DB" represents the refined results by manifold ranking. Regarding classification rate within top 10 (CR@10), "DB" outperformed "Baseline" by 7.67 points, which shows the effectiveness of the proposed co-occurrence-based refinement method. Figure 4 shows co-occurrence frequency among the top 15 frequent foods in our database. "Red boxes" means more high frequent co-occurrence pairs.

In the previous experiment, we set the constant value α in Eq.1 as 0.1. This value was decided by the experimental results varying α from 0.0 to 0.9 as shown in Figure 6. α is a constant adjusting how extent co-occurrence is taken into account in the manifold ranking computation. When α is 0.1, CR@10 achieved the maximum value, 63.50%. That is why we set α as 0.1.

Next, we carried out manifold ranking with co-occurrence matrix obtained by the knowledge on the Web as well. As explained in the previous section, we used Normalized Google Distance which can be calculated based on the hit numbers of given words returned by Google Search. We built three kinds of NGD-based co-occurrence matrices: the first one is computed using Google Search, the second one is obtained using Google Image Search instead of normal Google Search, and the third one is calculated using Google Search the search target sites of which were restricted to blogs mainly related to foods. Figure 7 shows the results by three kinds of matrices as well as baseline, and

Table 1. Classification rate within the top ten candidate (CR@10).

Baseline	DB	NGD(text)	NGD(img)	NGD(blog)
55.84	63.50	56.27	56.71	56.10
gain ⇒	+7.67	+0.43	+0.87	+0.26

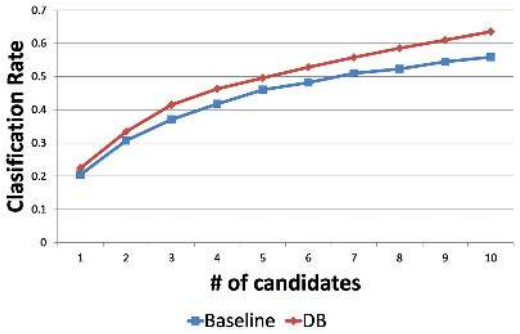

Figure 4. Classification rate before and after applying manifold ranking

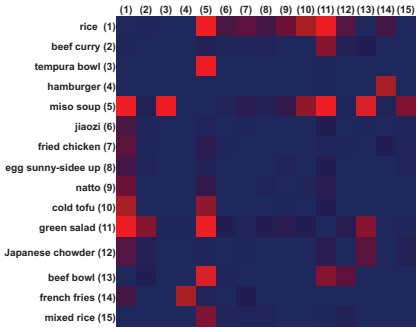

Figure 5. Co-occurrence matrix extracted from our food photo database

Table 1 shows the value of CR@10 of them. Although Google Search with food blogs achieved the best results among the Web-based methods, its improvement was not so much. Overall, Web-based co-occurrence was not effective in this experiment.

4 Conclusions

In this paper, we proposed a method to detect multiple food items from one food image considering co-occurrence statistics with the manifold ranking method. In the experiments, we could improved the result from 55.85% to 65.62% regarding the classification rate within top ten using co-occurrence statistics estimated from a food image dataset. For these results, it has been shown that the proposed method is very effective. On the other hand, Web-based co-occurrence does not improve the initial results so much.

For future work, we need to examine how to use external knowledge such as Web texts more deeply, and compare other methods than manifold ranking to consider co-occurrence statistics for multiple object recognition such as conditional random fields (CRF).

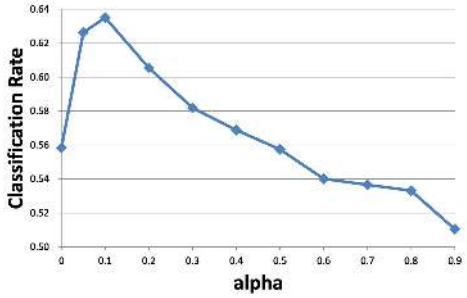

Figure 6. Classification rate varying the value of α

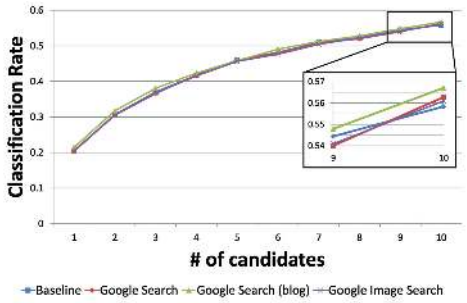

Figure 7. Classification rate with Web resources

References

[1] T. Joutou and K. Yanai, "A food image recognition system with multiple kernel learning," in *ICIP*, pp. 285–288, 2009.

[2] S. Yang, M. Chen, D. Pomerleau, and R. Sukthankar, "Food recognition using statistics of pairwise local features," in *CVPR*, 2010.

[3] Z. Zong, D.T. Nguyen, P. Ogunbona, and W. Li, "On the combination of local texture and global structure for food classification," in *ISM*, pp. 204–211, 2010.

[4] Y. Matsuda and K. Yanai, "Recognition of multiple-food images by detecting candidate regions," in *ICME*, 2012.

[5] P. F. Felzenszwalb, R. B. Girshick, D. McAllester, and D. Ramanan, "Object detection with discriminatively trained part-based models," *PAMI*, vol. 32, no. 9, pp. 1627–1645, 2010.

[6] Y. Deng and B. S. Manjunath, "Unsupervised segmentation of color-texture regions in images and video," *PAMI*, vol. 23, no. 8, pp. 800–810, 2001.

[7] S. Sonnenburg, G. Rätsch, C. Schäfer, and B. Schölkopf, "Large Scale Multiple Kernel Learning," *JMLR*, vol. 7, pp. 1531–1565, 2006.

[8] A. Rabinovich, A. Vedaldi, C. Galleguillos, E. Wiewiora, and S. Belongie, "Objects in context," in *ICCV*, 2007.

[9] E. B. Sudderth, A. Torralba, W. T. Freeman, and A. S. Willsky, "Learning hierarchical models of scenes, objects, and parts," in *ICCV*, pp. 1331–1338, 2005.

[10] D. Zhou, J. Weston, A. Gretton, O. Bousquet, and B. Schölkopf, "Ranking on data manifolds," *NIPS*, vol. 16, pp. 169–176, 2004.

[11] J. He, M. Li, H.J. Zhang, H. Tong, and C. Zhang, "Manifold-ranking based image retrieval," in *ACM MM*, pp. 9–16, 2004.