International Journal of Database Theory and Application
Vol.8, No.1 (2015), pp.95-104
http://dx.doi.org/10.14257/ijdta.2015.8.1.11

Contextual Co-occurrence Information for Object Representation and Categorization

[1]Soheila Sheikhbahaei and [2]Zahra Sadeghi

[1]*Department of Computer Engineering, Kharazmi University, Tehran, Iran*
[2]*School of Electrical and Computer Engineering, University of Tehran, Tehran, Iran*
Zahra.sadeghi@ut.ac.ir

Abstract

Object categorization based on hierarchical context modeling has shown to be useful in large database of object categories, especially, when a large number of object classes needs to be recognized from a range of different scene categories. However, average precision of categorization is still low compared to other existing methods. This may reflect that the contribution of underlying relations between objects has not been fully considered. In this paper, we improve average precision of contextual object recognition by taking advantage of objects co-occurrence information. Our method consists of two main phases. In the first phase, object representation is derived by considering the frequency of objects appeared in each image. The second phase is focused on classification of objects by applying a decision tree algorithm. We use SUN09 database to evaluate our proposed method. This database consists of images spanning from different scene categories and object instances. Our experimental results demonstrate that our proposed method achieves a higher average precision in comparison to a recent similar method by encoding contextual information in an efficient way.

Keywords: *Object categorization; contextual information; semantic representation*

1. Introduction

Object recognition is one of the most important tasks in computer vision. The main goal is to predict the presence of the desired object in a scene. It plays a critical role in applications such as optical character recognition [1], face detection [2] and video surveillance [3]. However, object recognition is considered a very difficult task. Part of this difficulty is caused by vast number of objects, their diversities in shape and color and small-sized objects. To address such difficulties, context-based models have been extensively studied and widely been used with promising results. Any information that is not directly produced by appearance of an object can be regarded as contextual information. There are three types of contextual information: Semantic context information, which is extracted by exploring each object in relationship with other objects and their occurrence in a scene; Spatial context information, which is related to likelihood of an object according to its position to other objects in a scene; And scale context information, which is about the size of the objects with respect to other existing objects in the scene [4]. In this work we use semantic contextual information to predict the presence of each object. Several different methods have previously exploited contextual relations to improve object recognition tasks and to reduce processing time (*ex.*, [4-6]). One drawback of such approaches is that they are tested on datasets with only a few object categories and most images contain only one or two object classes [7]. Therefore, the potential benefit of contextual information has not been considered in these methods. To address this problem, Choi *et al.*, proposed a

ISSN: 2005-4270 IJDTA

model based on hierarchical contextual model. This algorithm has gained attention due to introducing a new dataset named SUN dataset with images that contain many instances of different object categories. Recently, object co-occurrence statistics has been used for investigation of conceptual relationships in a subset of SUN2012 database [8]. In the current study, we create a feature vector for each object based on the co-occurrence information of object categories available in SUN09 database. Object are then represented by encoding the frequency information collected from other objects that co-occur in similar scenes. Finally, object categories are decided by a decision tree classifier. The rest of the paper is organized as follows: in section 2, we briefly review decision tree reconstruction method we then describe our proposed method in Section 3. Experimental results and conclusions are presented in Sections 4 and 5 respectively.

2. Decision Tree

Decision tree classifiers break down a complex decision-making process into a collection of simpler problems, thus providing a solution which is often easier to interpret. By rank ordering attributes based on their performance, decision trees can explore a large space quickly and efficiently. Decision tree idea is very simple and intuitive. In this method there are questions which specify the path of classification. Traveling from one node to another is based on the answer of the related node question. In our proposed method we use CART algorithm, one of the most important tools in modern data mining. In this section we introduce briefly how this algorithm works. The Notations used are shown in Table 1.

In each node t, the aim is to maximize splitting criterion $\Delta i(s,t)$. Estimation of p(j, t), p(t), and p(j|t) are as follows [9]:

$$P(j, t) = \pi(j)\, N_{wj}(t) / N_{w,j} . \quad (1)$$

$$P(t) = \Sigma_j\, p(j, t). \quad (2)$$

$$P(j|\, t) = P(j, t)/P(t) = P(j, t)/\Sigma_j\, P(j, t). \quad (3)$$

$$N_{w,j} = \Sigma_{n \in h}\, w_n\, f_n\, I(y_n = j). \quad (4)$$

$$N_{w,j}(t) = \Sigma_{n \in h(t)}\, w_n\, f_n\, I(y_n = j). \quad (5)$$

I (a=b) is indicator function and has the value 1 when a=b, 0 otherwise. If Y is continuous, the following splitting criterion is used with the Least Squares Deviation (LSD) impurity measure.

$$\Delta i(s, t) = i(t) - p_L\, i(t_L) - P_R\, i(t_R). \quad (6)$$

$$i(t) = (\Sigma_{n \in h(t)}\, w_n\, f_n\, (y_n - y(t))^2)/ (\Sigma_{n \in h(t)}\, w_n\, f_{n)} . \quad (7)$$

$$p_L = N_w(t_L) / N_w(t),\ P_R = N_w(t_R)/N_w(t),\ N_w(t) = \Sigma_{n \in h(t)}\, w_n\, f_n . \quad (8)$$

$$y(t) = \Sigma_{n \in h(t)} w_n f_n y_n / N_w(t). \quad (9)$$

Stopping rules used in this algorithm are as follows:

- If all cases in a node have identical values of the dependent variable.
- If tree depth reaches the user specified maximum tree depth value.
- If the size of a node is less than the user specified minimum node size value.
- The improvement is smaller than the user specified minimum improvement.

Table 1. Notations used in Cart Algorithm

Variable name	Definition
Y	Dependent target variable. If Y is categorical with *J* classes, its class takes values in C = {1, …, *J*}.
X_m , m=1,..,M	Set of all predictor variables.
h= {X_n, Y_n} n=1,..,N	Learning sample.
h(t)	Learning samples that fall in node *t*.
w_n	Case weight associated with case *n*.
f_n	Frequency weight associated with case *n*.
π(j) , j=1,…,J	Prior probability of *Y* = *j*, *j* = 1, …, *J*.
p(j, t)	Probability of a case in class *j* and node *t*.
p(t)	Probability of a case in node *t*.
P(j |t)	Probability of a case in class *j* given that it falls into node *t*.
C(i |j)	Cost of miss-classifying a class *j* case as a class *i* case. *C*(*j* | *j*) = 0 .

3. Proposed Model based on Semantic Contextual Information

In this section, a contextual model is proposed which is efficient for large-scale object recognition. This algorithm consists of three steps: 1- construction of the contextual frequency matrix according to training images. 2- Building the classification decision tree according to this matrix. 3- Using contextual vector of each test image to determine the label of the object in test images.

3.1. Benchmark Data

The experiments and performance evaluation were carried out on SUN 09. It consists of 4367 training and 4317 test images. These images contain more than 500 objects from more than 800 different scenes. We perform object categorization task on 107 object categories of this dataset. These 107 categories are the same as collected by Choi et al. PASCAL dataset is another standard dataset which has been widely used for the task of object recognition [10]. This dataset is not beneficial for contextual object recognition as it contains few objects in each image. Samples of this dataset are shown in Figure 1.

Figure 1. Two Sample Images of PASCAL Dataset. Left Image is Labeled with Two Objects (Airplane and Person). Right Image is Labeled with One Object (Boat)

3.2. Contextual Frequency Matrix Construction

As stated earlier, the contextual frequency matrix is an image by object matrix which is created by leveraging object co-occurrence relationships. This matrix is illustrated in Figure 2. Each element a_{ij} corresponds to the number of object j in image i. We then collect objects of interest for categorization task and transform it into an object representation matrix. For the sake of comparison, we used the same set of objects as were used in the hierarchical contextual model [11]. Hence, each row of object representation matrix contains information for one sample for one of the category of objects. In this way, we obtain another matrix that stores high-order contextual representation for all objects (Figure 3). This matrix is then utilized for training a decision tree classifier.

3.3. Decision Tree Construction

Each row of the object representation matrix is treated as a training sample for building a decision tree. As explained in Section 2 we use CART method for creation of a decision tree. Each branch of the constructed tree specifies a decision rule and each terminal leaf are labeled with the predicted value for that node. In our method, decision rules are obtained by attending to contextual information derived from the frequency of presence of all objects in each image and the label is the name of the target object we want to detect. After the training phase, we achieve a tree which is constructed over all objects. This tree provides us with a specific path for categorization of each individual object.

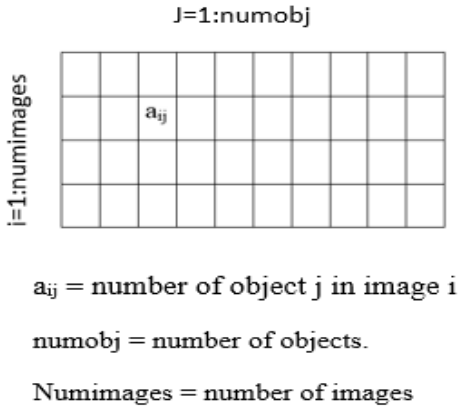

Figure 2. Contextual Frequency Matrix

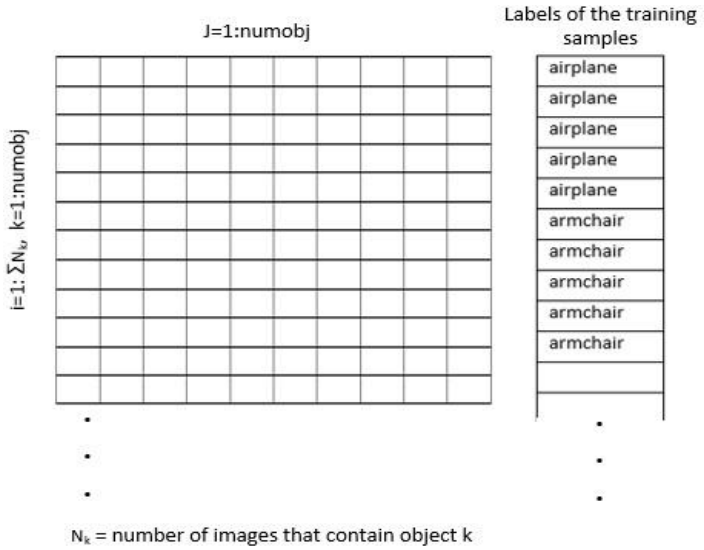

Figure 2. Contextual Representation Matrix

4. Evaluation

In this section, we compare our proposed method with a similar algorithm based on hierarchical contextual information [11]. We performed object classification task on 107 object categories. These categories span from regions (*e.g.*, road, sky, and building) of well-defined objects (*e.g.*, river, towel, and curtain). First, we show that relationship of all objects is not considered in the hierarchical context model. Then we indicate that our proposed method improves recognition by bringing the interaction between objects by using all other objects and their frequencies to predict the presence of a desired object. Hierarchical context model uses a hierarchical structure to represent object relationships. In this model a tree is learned from SUN 09 according to Chow-Liu algorithm [12] that maximizes the likelihood of the data. This tree is shown in Figure 4. For example, most of the objects that commonly can be found in a kitchen are appeared as descendants of the node sink, and all the vehicles are descendants of road. In this model, prediction of each object is based on the information available from the parent of a node until the root (Figure 4.) and other objects are not considered. Presence of each object i in an image is indicated by node b_i in the tree. The joint probability of all binary variables is computed according to the tree structure:

$$p(b) = p(b_{root}) \prod p(b_i \mid b_{pa(i)}). \quad (9)$$

Where $pa(i)$ is the parent of node i and $b \equiv \{b_i\}$. When two objects appear together in a scene like *floor* and *wall*, their relationship is positive in the tree and otherwise it's negative. For example, probability of object *dish* is computed using equation (10).

$$p(b_{dish}) = p(b_{sky})\, p(b_{floor}|\, b_{sky})\, p(b_{wall}|b_{floor})\, p(b_{sink}|b_{wall})\, p(b_{countertop}|b_{sink})\, p(b_{cabinet}|b_{countertop})\, p(b_{bowl}|b_{cabinet})\, p(b_{plate}|b_{bowl})\, p(b_{dish}|b_{plate}). \quad (10)$$

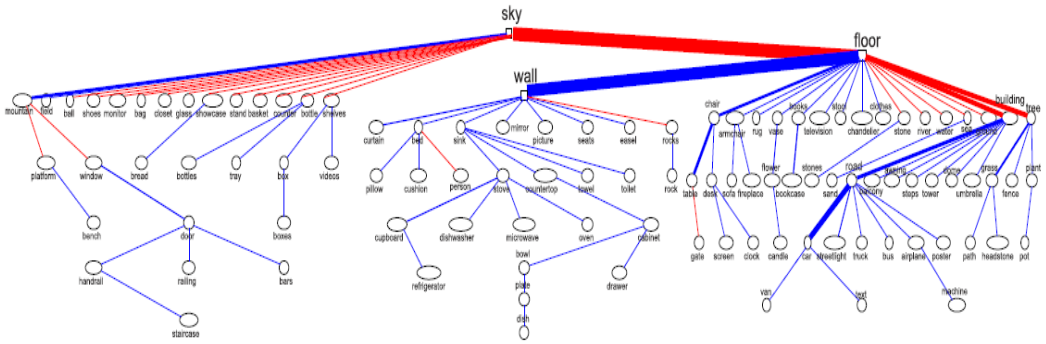

Figure 3. Tree Structure Learned from SUN 09 [1]. Negative Relations between Categories are shown by Red Edges and the Strength of the Link is Represented by Edge Thickness

As it can be inferred, to predict the presence of class *dish*, only 8 object categories are considered. Prediction of other classes is performed by considering even less number of categories and in most of the cases, it's limited to two or three categories. We show that by using information of all object categories we can increase the precision. In the proposed model object prediction is based on the presence of all other objects and their frequencies. For example, in Figure 5 an image is shown which contains 10 objects (sky, airplane, car, unmatched, building, fence, finger, person, window, rand road). Using the hierarchical model, objects with the highest probability (p=1) are box, headstone, and grass. As we see these objects are not present in this image. The probability of presence of the airplane in this image has been predicted 0.05 which is close to zero and this algorithm has predicted that airplane is not present.

We applied our algorithm to the same set of object categories as used in the hierarchical model. Quantitative comparisons are performed according to mean average precision and are shown in Fig 6. Object categorization precision is calculated by eq. 11,

$$\text{Precision of object i detection} = \text{correct predictions of object i} / \text{all predictions of object i} . \quad (11)$$

Our results are based on the average precision obtained by 10 independent runs for all objects with a different randomized order of training inputs. In addition, mean average precision (MAP) of the proposed algorithm is compared with hierarchical model in Table 2. Normalized confusion matrix for 10 objects with the highest precision is shown in Figure 7. As evident, diagonal numbers are almost one and most of the non-diagonal numbers are zero which demonstrates that our proposed algorithm achieved a high accuracy in object classification. The overall improvement percentage obtained by our proposed algorithm in comparison to the hierarchical contextual approach is up to 23.93%.

Figure 4. An Image of SUN 09 Test Dataset. This Image Contains 10 Object Categories. This Image Contains 10 Object Categories. Probability of Presence of Airplane Object in this Image According to Hierarchical Model is 0.05, while this Object is Predicted in the Proposed Method

5. Conclusion and Discussion

In this paper, we proposed an object categorization method based on semantic contextual information for images with large number of objects spanning from different scenes. To exploit effectiveness of contextual information, there need to be a large number of images containing different object categories with various sizes. Hence, we evaluated the performance of our proposed algorithm on a large number of objects available in SUN 09 database which contains more than 200 object categories in a wide range of scene categories and obtained superior performance than that of a similar contextual object recognition strategy. In our method, first we use contextual information by taking advantage of presence of other objects and their frequencies to compute contextual frequency matrix. This matrix is then subject to a tree decision classifier. By using objects presented in the image and frequency of their instances, we improved object recognition accuracy in comparison to a similar hierarchical contextual model. It is worth to note that in the hierarchical contextual model, authors optimized an effective pathway for recognition of each individual object. In contrast, in our proposed model, we considered the interaction between all objects to find a rich representation for each category. Therefore, it might appear that our method tolerates redundancy in representation. While the prevailing structure of connections between concepts is still an open investigation, it seems intuitive that human lives in a world intermingled with huge categories. This view may lead to the point that although object recognition procedure executes fast in the brain, there's a semantic link even between apparently unrelated concepts. Moreover, a combination of different aspects of information obtained from seemingly different categories may influence mental object representation. Our model is consistent with the recent finding that challenges the view of object-category selective regions in human visual cortex for object representation [13] and is a possible avenue for future investigation about the network of semantic connections between different concepts.

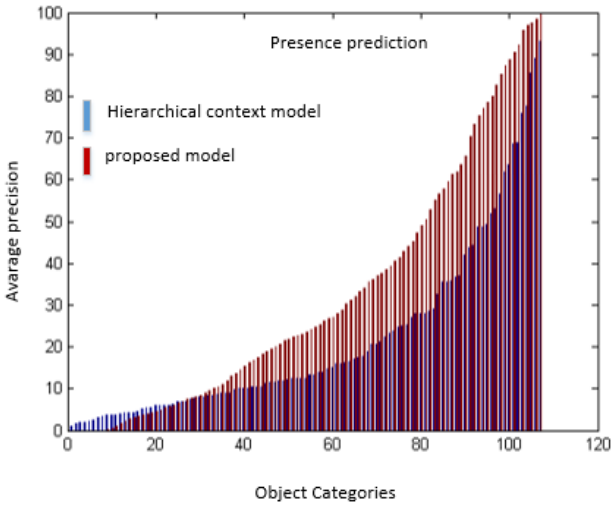

Figure 5. Average Precision of Proposed Method and Hierarchical Context Model. Object Categories are Sorted by AP

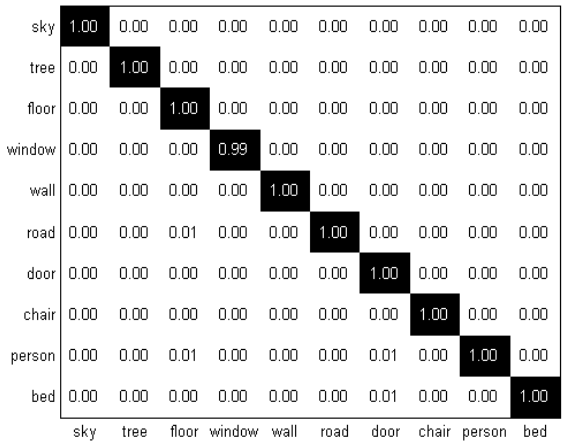

Figure 7. Normalized Confusion Matrix for 10 Objects with the Highest Precision on SUN 09 Dataset

Table 2. Mean AP (Averaged Across all Object Categories) for Presence Prediction on SUN 09

	Hierarchical context model	*Proposed method*
Mean average precision	*26.08*	***32.32***

References

[1] Permaloff and C. Grafton, "Optical Character Recognition", Political Science and Politics, vol. 25, no. 3, **(1992)**, pp. 523-531.
[2] T. Serre, J. Louie, M. Riesenhuber and T. Poggio, "On the Role of Object-Specific features for Real World Object Recognition in Biological Vision", Workshop on Biologically Motivated Computer Vision, **(2002)**.
[3] M. J. Hossain and M. A. A. Dewan, "Moving object detection for real time video surveillance: An edge based approach", IEICE transactions on communications, no. 12, **(2007)**, pp. 3654-3664.
[4] C. Galleguillos and S. Belongie, "Context based object categorization: A critical survey", Computer Vision and Image Understanding, vol. 114, **(2010)**, pp. 712-722.

[5] A. Torralba, K. Murphy and W. Freeman, "Contextual models for object detection using boosted random fields", Advances in Neural Info. Proc. Systems, **(2004)**.
[6] Z. Tu, "Auto-context and Its Application to High-level Vision Tasks", IEEE Computer society conference on computer vision and pattern recognition, **(2008)**.
[7] M. Jin, C. A. Torralba and A. S. Willsky, "A Tree-Based Context Model for Object Recognition", Pattern Analysis and Machine Intelligence, IEEE Transactions on, vol. 34, **(2012)**, pp. 240-252.
[8] Z. Sadeghi, J. L. Mcclelland and P. Hoffman, "You shall know an object by the company it keeps: An investigation of semantic representations derived from object co-occurrence in visual scenes", Neuropsychologia, **(2014)**.
[9] L. Breiman, J. H. Friedman, R. A. Olshen and C. J. Stone, "Classification and Regression Trees Wadsworth Statistics/Probability", **(1984)**.
[10] M. Everingham, L. V. Gool, C. Williams, J. Winn and A. Zisserman, "The Pascal Visual Object Classes (VOC) Challenge", International Journal of Computer Vision, vol. 88, **(2010)**, pp. 303-338.
[11] M. J. Choi, J. Lim, A. Torralba and A. S. Willsky, "Exploiting hierarchical context on a large database of object categories", Computer Vision and Pattern Recognition (CVPR), **(2010)**.
[12] C. Chow and C. Liu, "Approximating discrete probability distributions with dependence trees", Information Theory, IEEE Transactions on Information Theory, vol. 14, **(1986)**.
[13] A. G. Huth, S. Nishimoto, A. T. Vu and J. L. Gallant, "A Continuous Semantic Space Describes the Representation of Thousands of Object and Action Categories across the Human Brain", Neuron, vol. 76, no. 6, **(2014)**.

