# THE USE OF SPATIAL RELATIONSHIPS AND OBJECT IDENTIFICATION IN IMAGE UNDERSTANDING

**Hamid Sadeq Mahdi Alsultani, Shaymaa Taha Ahmed and Ban Jawad Khadhim**
Department of Computer Science, College of Basic Education,
University of Diyala, Diyala, Iraq

**Qusay Kanaan Kadhim**
Faculty of Information and Communication Technology (FTMK), Universiti Technology
Malaysia Melaka (UTeM), Malaysia, Hang Tuah Jaya, 76100 Durian Tunggal,
Malacca, Malaysia

## ABSTRACT

*Image understanding includes mathematical and geometrical abilities. It requires analyzing, classifying, labeling to identify requirements, involving difference comparing or appreciated gaps in an image or object analysis, and these investigation cases use different methods. Image understanding supports many knowledge fields as Image Processing, Artificial Intelligence, Computer Graphics, Psychology, Object Recognition and many other fields. From another side, the information of spatial relationships holds enormous and vast inputs for the study of image understanding. This paper concentrated on the techniques of image understanding by the use of spatial relationships and object identification, whereas staying nearer to the related issues.*

**Key words:** Understanding, Object Identification, Spatial Relationships, Context, Categorization, Computations Modeling

**Cite this Article:** Hamid Sadeq Mahdi Alsultani, Shaymaa Taha Ahmed, Ban Jawad Khadhim and Qusay Kanaan Kadhim, The Use of Spatial Relationships and Object Identification in Image Understanding, International Journal of Civil Engineering and Technology, 9(5), 2018, pp. 487–496.
http://www.iaeme.com/IJCIET/issues.asp?JType=IJCIET&VType=9&IType=5

## 1. INTRODUCTION

The definition of image understanding is the method of an image interpreting into different objects or areas. This method helps to find out alteration happens in a specific image. The next steps are learning objects meaning containing their specific spatial relationships with the other objects or the entire image. The previous definition express the necessity of a suitable process, or technique for image studying and understanding. So, firstly the image needs to be divided into parts, and different objects to interpret the image (George, 2016).

The understanding of image needs perfect interpreting for its own spatial info too. As an example, the contextual model allows to explore these details via the quantification of area spatial relationship. Moreover, this process assists in resolving the suspicions of low level features included to classify an image and to detect an object (Kamavisdar et al., 2013). Actually, different approaches are used in image understanding to model spatial relationships and interrelationships (Blaschke, 2003; Kamavisdar et al., 2013).

## 2. BRIEF BACKGROUND

At any scene, the objects appear concordantly, and that is ralated to the objects' spatial order and probability of occurrence (Blaschke, 2003; Pinz, 2005; Kamavisdar et al., 2013).

According to (Pinz, 2005), two major groups of image understanding computation models were improved for assisting in analyzing of object configuration at various task requests. The first group can deals with the ambiguity of objects in a specific scene (Torralba and Sinha, 2001; Pinz, 2005; Shotton et al., 2006; Athanasiadis et al., 2007; Liu et al., 2016), while the second group deals more with the observation based on analysis (Kumar and Hebert, 2005; Pinz, 2005; Shotton et al., 2006; Casaca et al., 2013; Liu et al., 2016).

Clearly: The first models group deals with the exploration of object from bottom level to up level. It represents an image at low level and it can be considered as an "Essence" to draw up contextual concepts before real recognition of the object (Pinz, 2005). However, there is little divergence in this group. Thus, several models take into account the low-level features correlation through images that surrounding the object or through the group (Pinz, 2005; Shotton et al., 2006; Athanasiadis et al., 2007; Liu et al., 2016). However, with the progression of researches, the most recent methods include an occurrence of high-level features that offer a huge supporting to the contextual limitations (Pinz, 2005; Zhong et al., 2017).

The second models group refers nearly to the techniques used in the recognition of object, because they are using further complicated methods for analyzing. This group contains techniqes including spatial context and relationships (Pinz, 2005; Shotton et al., 2006; Liu et al., 2016). This group contains also pair-wise relationships (analyzing of images inter-regions) and semantic context (enhancement the accuracy of recognition) beneficial to understand occurrence of the objects' order in the scene (Kumar and Hebert, 2005). Examples under this group are *top-down* methods, that are called also *model-based* (George, 2016).

## 3. IMAGE UNDERSTANDING APPLICATIONS

For long time ago image understanding has been a significant topic in psychology and computer-vision, and this topic has effects on localization, visual-analysis and the performance of recognition (Simons and Wang, 1998; Pinz, 2005; Oliva and Torralba, 2007; Lalaoui and Mohamadi, 2013).

Image retrieving or retrieval is an area that relates to search and browse the digital images through a set of database (Wang, 2013). This area very associated with image understanding through its requirements of retrieval actions. According to using or consuming image in many actions, the function of image retrival has a great effect on image understanding at the level of applications. So, searching and browsing digital images have widely interest in various areas such as multimedia, digital image processing, digital library, databases and other areas (Ogiela and Tadeusiewicz, 2002; Tadeusiewicz and Ogiela, 2004; Wang, 2013).

# 4. IMAGE UNDERSTANDING THROUGH OBJECT CATEGORIZATION

Object categorization allows to identify, position and verify the important features of object categorization in the image. The purpose of designing different object categorization models is to simplify minimizing many faults (i.e. noise, low quality), that represent an obstruction in the recognition of an object (Galleguillos and Belongie, 2010; Kamavisdar et al., 2013). Actually, to solve such an obstruction, object's appearance information and context information are used to optimize the recognition accuracy in those models (Galleguillos and Belongie, 2010). Figure 1 shows appearance features (i.e. responses of edge, colour and shape) that considered as essential details to identify object categories because these details arer sensitive to the changes in object categories (Pinz, 2005; Galleguillos and Belongie, 2010; Kamavisdar et al., 2013; George, 2016). In figure 1, the object's appearance in (b) can not give sufficient info about identity of the object. The whole scene (a) makes the object in (b) identifiable as a kettle by adding contextual info.

An appropriate process to identify objects that are based on context info needs perfect understanding to know their positions in the environment of the real world. The classes of relations to analyze of objects in real world scenes context are familiar size, support, probability, position and interposition. In the area of computer vision, context features (e.g. interactions information among the objects exist in a specific scene), are used for enhancing the recognition of objects, and reducing the time of process (Galleguillos and Belongie, 2010).

There are at most four stages that can be used to explore an object in a specific image with the assist of the developed principles of object categorization through several researchs works. Next sections present these stages.
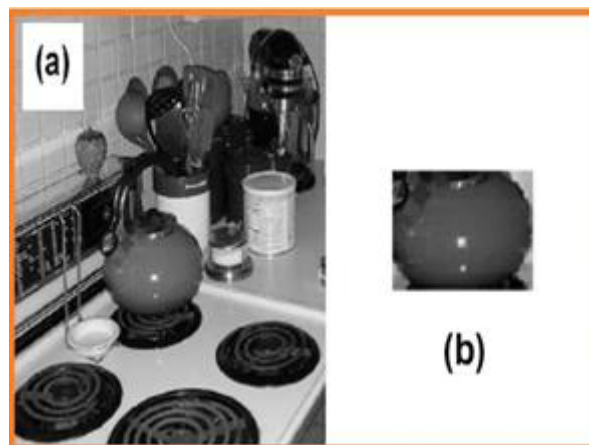


**Figure 1** Appearance features. (a) the complete scene, (b) the  object's  appearance

## Object Categorization Context Types

These types are essential and known as *contextual features*. They consists of three classes, which are *semantic-context*, *spatial-context* and *scale-context*, that participated in object categorization obtaining from the real world scene (Galleguillos and Belongie, 2010).

Often, in the real world scene context, the main classes that can be used to analyze the word images appearance are *semantic-context* and *spatial-context*. (Bar and Aminoff, 2003; Pinz, 2005; Galleguillos and Belongie, 2010; George, 2016;).

*Semantic-context* deals with an object's presence probability in several and unique scenes only. Every object semantic-context implies its occurrence with another objects comparatively to the objects' presence in the scene. *Semantic-context* may be gotten through several labels/pixel to show a pixel presence at a specific object, containing context of the external knowledge (Bar and Aminoff, 2003; Pinz, 2005; Galleguillos and Belongie, 2010).

*Spatial-context* is a derivation of *Biederman's* position class. It deals with the probability of an object's presence only in several positions compared with another objects' occurrence around it (Galleguillos and Belongie, 2010).

*Scale-context* deals with the size of the object; from bottom level (pixels) to up level (real object shape). This class includes identification of another object, then the depth and spatial relationships between this object and the target object in the scene.

## Context Level Methods

These methods are *global-context* and *local-context* (Figure 2). *Global-context* concerns with the information taken from the entire image including the object, while *local-context* concerns with the details at the object neighborhood's areas (Galleguillos and Belongie, 2010; George, 2016). Psychology studies related object recognition are used to learn these essentials (Galleguillos and Belongie, 2010). In (figure 2), straight window shows *global-context*, while dotted window shows *local-context*.



**Figure 2** Global-context and  local-context

## Contextual Interactions

Contextual interactions may be analyzed within either *global-context* or *local-context*. If the analysis happens at the global level, then they are known as interactions among objects and scenes. On the other side, if the analysis happens at the local level, (i.e. at pixel, region or object level) then they are known as pixel, region or object interactions respectively (Galleguillos and Belongie, 2010).

## Integrating Context

In spite of the complexity of tasks, several learning methods are used to obtain the advantage of their useful probability algorithms. The main groups associating many methods to the integrating context are the *classifiers* and *graphical models*. Classifiers assist graphical models to integrate both context and appearance features. The essential role of *classifiers* is to combine contextual features with the local appearance detectors' outputs. There are two stages to construct the context feature. In the first stage, the low level and semantic info are

calculated when processing of the image. In the second stage, the samples of previously calculated features are collected to calculate the context feature at all points (Galleguillos and Belongie, 2010).

## 5. OBJECT IDENTIFICATION AND SPATIAL RELATIONSHIP IN IMAGE UNDERSTANDING

Object identification is usual performed incorrectly either by usual or little experience persons. Generally, humans tend to concentrate on clear fact that may be accessed almost everyone. The details that would be produced by a lot of obsevers are colour, size, nature and shape of object. In general, spatial-context and the contextual identification of object are missed from such report. However, with expert persons the concept of object identification take a deep and wide perception. When humans are looking at specific image, then they can see the world beyond that image. On the other hand, most observers may talk on the objects in the image. So, the observers may disregard the fact that all things surrounding specific object are important to identify that object (Kim et al., 2013).

The perception of humans requires several mathematical models for confirming their points of view and assumptions against results of the computer (Tadeusiewicz and Ogiela, 2004; Szczepaniak and Tadeusiewicz, 2010). These models are related to the popular methods used in the analysis of image understanding. Also, these models are required to understand the scene from a single image. Scene understanding needs powerful assumptions on the real world, and these asuumptions contain several complicated probability formulas to form the algorithms of image processing (Ogiela and Tadeusiewicz, 2002; Tadeusiewicz and Ogiela, 2004; Lalaoui and Mohamadi, 2013).

Actually, spatial information are applied by contextual models through the region spatial relationships quantification for solving doubts in low-level features that are used to detect object and classify image. (Aksoy et al., 2005) discussed axiomatic, adaptable and effective methods to model spatial relationships, and pairwise directional. a *fuzzy landscape* was defined by these methods. In this landscape, a value is assigned to each point in the image to quantify its position relatively to the reference object(s) and the relationship type.

(Rosman and Ramamoorthy, 2011) presented an algorithm that used a layered representation to redescribe a scene by labeling objects' point clouds in the scene. This layered representation gives useful symbolical meaning to the inter-object relationships for the future common sense logic and decision making.

An essential context in categorization of object to identify and understand the image is the low-level (bottom-up) method (Klinker et al., 1990; Pinz, 2005; Galleguillos and Belongie, 2010; Kim et al., 2013; Lalaoui and Mohamadi, 2013). Nevertheless, low-level features are not guarantee to describe high-level conceptions in minds of users. Thus, with extra development in researches of image retrieving, this conception is moved from a word to low-level then to semantic features to get minimal time consuming and objective with simplicity in the process (Wang, 2013).

## 6. IMAGE UNDERSTANDING PROCESSING METHODS

Image understanding is a multi-disciplinary domain of numerous areas. A lot of areas require adequate knowledges, methods, processes, and theories from math, computer sciences, engineering, etc. So, image understanding may be clearly a complicated mission. However, selecting the correct method to analysis and study image understanding considered a significant step to start (Tadeusiewicz and Ogiela, 2004).

Usually, when persons are requested for describing specific image, then, the objects in the image with their related locations are given, and this description causes neglecting many details (Klinker et al., 1990). Furthermore, image understanding generally consists of two classical stages: extracting features stage, and reasoning stage. The second stage is in charge of the exploration of the features of the image comparatively to the features of the described object (Klinker et al., 1990; Lalaoui and Mohamadi, 2013).

The process of object exploration, helps in preventing the confusion of the object appearance under many physical changes such as illumination, influence of noise, heat, etc. (Klinker et al., 1990; Pinz, 2005; Oliva and Torralba, 2007; Galleguillos and Belongie, 2010; Szczepaniak and Tadeusiewicz, 2010; Rosman and Ramamoorthy, 2011; Kamavisdar et al., 2013; George, 2016).

Actually, the difficulty of image understanding is making it ambiguous to list down the methods taking part in image understanding works. However, the next paragraphs would review several of image understanding methods.

The first image understanding method is *bottom-up* method. It segments the image to parts to form several shapes of object, then it uses representations to draw up the objects. The second image understanding method is *top-down* method. Firstly, it designs several suppositions, then it applies the image data to the suppositions testing, and finally it gets the conclusions (Klinker et al., 1990; Pinz, 2005; Kamavisdar et al., 2013; George, 2016).

Another famous image understanding method is object-oriented classification. Instead of classifying individual pixels, this method permits users to take advantage of structural info to implement the classification based on regions or areas (Aksoy et al., 2005).

The last discussed image understanding method is *a method to color*. This method segments and analyzes the surfaces according to the changes of color, that come from highlight and shadow. It depends on the *Dichromatic Reflection Model*. Using this model enables separating the colored image into two essential images: the highlighted image, and the main unhighlighted image. This model may be used to contain segmentation of colored image in analyzing image (Klinker et al., 1990).

Spatial relationship understanding among objects has a great role in analyzing image understanding. Practically, objects' qualitative structure in an environment and the relationships among them can be used to define an environment composition, and to allow efficient plans construction for enabling the completion of different accurate tasks (Athanasiadis et al., 2007).

## 7. COMPUTATIONS MODELING OF IMAGE UNDERSTANDING

The standard relationships studied in the literature during image understanding works, concerned with structures, semantics, statistics, topology, and geometric relationships. Actually, many methods of computation modeling together with the above specified relationships are used to specify manner of modeling for areas or objects. The usually methods contain centroids, bounding box and grid-based spatial representation (Blaschke, 2003).

The following are image mathematical modeling examples out of object categorization as in (Galleguillos and Belongie, 2010.

## Spatial Context Method

Some years ago, spatial context was acknowledged by numerous methods as appropriate in image understanding recognition accuracy improving. It is a combination of both the statistics of inter-pixel and relations of pairwise among areas in images. However, depending on (Galleguillos and Belongie, 2010), (Shotton et al., 2006) contributed in enhancing the process by introducing the statistics of inter-pixel to categorize object. According to their own framework, the spatial interaction among class labels of neighbouring pixel is captured by the unary classifier ($\lambda_i$). A lookup table, an index of pixel ($i$) and class entry ($C_i$) are used to represent spatial context with normalized index ($\hat{i}$) of ($i$). The parameters of the model are represented by ($\Theta_\lambda$), such as in equation (1).

$$\lambda_i (C_i, i; \Theta_\lambda) = \log \Theta_\lambda (C_i, \hat{i}) \text{ (1)}$$

### Computational Process of Local Context

What is meant by **local-context info** is the information obtained from every area around a specific object. There are numerous models of object categorization, that have used for applying **local-context** from all objects, or pixel info around the selected object. Schiele and Kruppa contributed in their own algorithms for detecting face. The local arrangements of quantized coefficients are captured by the detector features, depending on the classifier of Naive Bayes, such as in equation (2).

$$\prod_{k=1}^{n} \prod_{x, y \in region} \frac{p_k\big(pattern_k(x, y), i(x), j(y)\big|object\big)}{p_k\big(pattern_k(x, y), i(x), j(y)\big|nonobject\big)} > \theta \text{ (2)}$$

($\Theta$) refers to the acceptance threshold; and ($P_k$) refers to the coarse quantization probability functions ($i(x)$ and $j(y)$) of the feature location respectively (Galleguillos and Belongie, 2010).

## Contextual Interactions Method

Essentially, there are two types of contextual interactions: *local interactions* and *global interactions*.

### Local Interactions

In extremely disarranged scene, bottom-up attentional frameworks may be applied to improve the performance of object recognition (i.e. local-context analyzing). The associated features of frameworks assure the quality of outcomes. In fact, bottom-up processing analyzes the interactions of pixel, to apply a similarity concept in neighbouring pixels. At pixel level, numerous works made a contribution to object categorization frameworks (Galleguillos and Belongie, 2010). Particularly, the case of (Galleguillos and Belongie, 2010) has made a solution to the obtaining contextual features problem by the use of interactions of pixel.

### Global Interactions

In this type of interaction, the representation of scene centered may be used to recognize global context. Actually, the interactions of object-scene are modeled by the use of image clusters training, that gives hints about the depicted objects in the image with their own probable locations. The relationships among the categories of the object (o), the spatial location of the categories in the image (x), with their appearance (g) may be modeled by equation (3).

$$p(o, x, g|\theta, \phi, \eta) = \prod_{i=1}^{N} \prod_{j=1}^{M_i} x \sum_{h_{ij}=0}^{1} p\big(o_{ij}|h_{ij}, \theta\big) p\big(x_{ij}|h_{ij}, \phi\big) p\big(g_{ij}|o_{ij}, h_{ij}, \eta\big) \text{ (3)}$$

($N$) is the number of images, such that each image with ($M_i$) proposals of object over the categories of the object, and ($P(oij|hij,\Theta)$) is the probability of the categories of the object that exist in the image. This represents the interactions of object-scene.

## Graphical Models

*Graphical Models* provide easy methods for visualizing a probabilistic model structure. These models offer a strong framework to implement the distributions of global probability determined by relatively local restrictions. The distributions of global probability are determined on directed graphs to express the causative relationships among random variables. Equation (4) may be used to compute the distribution of the joint probability for the directed graphical models.

$$P(x) = \prod_i P(x_i|pa_i) \quad (4)$$

Where: ($pa_i$) represents the potential function on the maximal graph's cliques. These models suppose using of objects that are conditionally independent for giving the scene (Galleguillos and Belongie, 2010).

## 8. CONCLUSIONS

Image understanding articles that are reviewed here show two common methods to learn and explore any object in a specific image. The first method explores the object from bottom level to up level. The contextual ideas are achieved before the actual recognition of object. While in the second method, techniques like semantic context, spatial relationships and spatial context pairwise relation are applied.

Depending on the reviewed articles, contextual models exploit spatial information through region spatial relationships quantification, and this exploitation assists in resolving the uncertainty of low-level features used to classify images and detect of objects.

According to the information of modeling and quantifying, spatial information and efficiency of the relationship models participate in developing the accuracy of image understanding applications.

Studies illustrated that the optimizing accuracy in image understanding exploration is accomplished via the combination of co-occurrence with spatial context instead of co-occurrence only.

The suitable perception of object categorization may assist in analyzing image and modeling the computations.

Several particular perspectives may be used to discuss object categorization comprehensively. These perspectives contain object categorization essential contexts (semantic, spatial and scale contexts), the contextual interactions (global and local contexts), the levels of context (bottom-up and top-down levels), and finally, the integrating context (graphical models and classifiers). Particularly, two essential contexts used to analyze image appearance that are semantic and spatial contexts.

## REFERENCES

[1]     Aksoy, S., R.G. Cinbiş and H.G. Akçay, 2005. Image classification and object detection using spatial contextual constraints. Bilkent University, Turkey.

[2]     Athanasiadis, T., P. Mylonas, Y. Avrithis and S. Kollias, 2007. Semantic image segmentation and object labeling. IEEE Transactions on Circuits and Systems for Video Technology, 17(3): 298-312. DOI: 10.1109/TCSVT.2007.890636

[3]     Bar, M. and E. Aminoff, 2003. Cortical analysis of visual context. Neuron, 38(2): 347-358.

[4]     Blaschke, T., 2003. Object-Based Contextual Image Classification Built on Image Segmentation. IEEE Workshop on Advances in Techniques for Analysis of Remotely Sensed Data, Oct. 27-28, IEEE Xplore, USA. DOI: 10.1109/WARSD.2003.1295182

[5]     Casaca, W., A. Paiva, E.G. Nieto, P. Joia and L.G. Nonato, 2013. Spectral image segmentation using image decomposition and inner product-based metric. Journal of Mathematical Imaging and Vision, 45(3): 227-238. DOI: 10.1007/s10851-012-0359-6

[6]     Galleguillos, C. and S. Belongie, 2010. Context based object categorization: a critical survey. Computer Vision and Image Understanding, 114(6): 712-722. DOI: 10.1016/j.cviu.2010.02.004

[7]     George, M.N.A., 2016. Objects in relation for scene understanding. PhD Thesis, Zurich University, Switzerland.

[8]     Kamavisdar, P., S. Saluja and S. Agrawal, 2013. A survey on image classification approaches and   techniques. International Journal of Advanced Research in Computer and Communication Engineering, 2(1): 1005-1009.

[9]     Kim, B.S., P. Kohli and S. Savarese, 2013. 3D Scene Understanding by Voxel-CRF. Proceedings of the IEEE International Conference on Computer Vision, (ICCV' 13), Dec. 1-8, IEEE Xplore, Australia. DOI: 10.1109/ICCV.2013.180

[10]    Klinker, G.J., S.A. Shafer and T. Kanade, 1990. A physical approach to color image understanding. International Journal of Computer Vision, 4(1): 7-38. DOI: 10.1007/BF00137441

[11]    Kumar, S. and M. Hebert, 2005. A Hierarchical Field Framework for Unified Context-Based Classification. Proceedings of the IEEE International Conference on Computer Vision, (ICCV' 05), Oct. 17-21, IEEE Xplore, China. DOI: 10.1109/ICCV.2005.9

[12]    Lalaoui, L. and T. Mohamadi, 2013. A comparative study of image region-based segmentation algorithms. International Journal of Advanced Computer Science and Applications, 4(6): 198-206.

[13]    Liu, T., X. Huang and J. Ma, 2016. Conditional random fields for image labeling. Mathematical Problems in Engineering, 2016: 1-15. DOI: 10.1155/2016/3846125

[14]    Ogiela, M.R. and R. Tadeusiewicz, 2002. Image understanding methods in biomedical informatics and digital imaging. Journal of Biomedical Informatics, 34(6): 377-386. DOI: 10.1006/jbin.2002.1034

[15]    Oliva, A. and A. Torralba, 2007. The role of context in object recognition. Trends in Cognitive Sciences, 11(12): 520-527. DOI: 10.1016/j.tics.2007.09.009

[16]    Pinz, A., 2005. Object categorization. Foundations and Trends® in Computer Graphics and Vision, 1(4): 255-353. DOI: 10.1561/0600000003

[17]    Rosman, B. and S. Ramamoorthy, 2011. Learning spatial relationships between objects. The International Journal of Robotics Research, 30(11): 1328- 1342. DOI: 10.1177/0278364911408155

[18]    Shotton, J., J. Winn, C. Rother and A. Criminisi, 2006. TextonBoost: Joint Appearance, Shape and Context Modeling for Multi-class Object Recognition and Segmentation. In: Lecture Notes in Computer Science, Leonardis A., H. Bischof and A. Pinz, (Eds.), Springer, Berlin, Print ISBN: 978-3-540-33832-1, pp: 1-15. DOI: 10.1007/11744023_1

[19] Simons, D.J. and R.F. Wang, 1998. Perceiving Real-World Viewpoint Changes. Psychological Science, 9(4): 315-320. DOI: 10.1111/1467-9280.00062

[20] Szczepaniak, P.S. and R. Tadeusiewicz, 2010. The role of artificial intelligence, knowledge and wisdom in automatic image understanding. Journal of Applied Computer Science, 18(1): 75-85.

[21] Tadeusiewicz, R. and M.R. Ogiela, 2004. Structural approach to medical image understanding. Technical Sciences, 52(2): 131-139.

[22] Torralba, A. and P. Sinha, 2001. Statistical Context Priming for Object Detection. Proceedings of the IEEE International Conference on Computer Vision, (ICCV' 01), Jul. 7-14, IEEE Xplore, Canada. DOI: 10.1109/ICCV.2001.937604

[23] Wang,, H.H., 2013. Performance and analysis of the automated semantic object and spatial relationships extraction in traffic images. Advances in Computer Science: an International Journal, 2(4(5)): 37-42.

[24] Zhong, Y., S. Wu and B. Zhao, 2017. Scene semantic understanding based on the spatial context relations of multiple objects. Remote Sensing, 9(10): 1-19. DOI:10.3390/rs9101030