Image Understanding Using Object Identification and Spatial Relationship

HAMID SADEQ MAHDI
University of Diyala / Basic Education College / Computer Science Department, Iraq

ABSTRACT

IU involves psycho-thinking, math and geometrical experiences/ability; it may require analyzing, sorting, classifying / categorizing/ for identification needs, including comparing for difference or gaps appreciating in an image/photo/picture or object (shape, color, appearance…, etc.) diagnosis. Such an investigations instance makes uses of various methods. IU is in background support to multiple fields of knowledge (e.g. AI, Robotics, Computer graphics and multimedia, Psychology, Image processing and objects recognition; etc.). In other hands, spatial relationship information holds huge and great inputs for high-level image understanding study. This articles review has focused on IU processes using object identification and object Spatial relationship, while remaining closer to related topics.

Keywords: Image understanding, object, image, spatial, contextual, pairwise, computer-vision, scene, perception.

1 INTRODUCTION

1.1 Defining IU

Image Understanding (IU) can be defined as the process of systematically interpreting an image broken down into regions/objects. This process aims at [finding out] figuring out what change occurs in the image. And advanced steps would be learning the meaning of formed objects including their individual spatial relationship with others or to the whole **[4].** This definition indirectly tells about the need for a proper procedure, practically working method if to study and understand an image. Thus, the first step requires dissecting (or fragmenting) the image into pieces, various object shapes /forms for imagery interpretation. Therefore, IU can be undertaken from different perspectives and approaches [depending] through either of processing methods for the image better understanding.

Advanced image understanding requires good interpretation of its spatial information as well. For instance / Such as/ For, contextual models enable deeply exploring such details through the region spatial relationships quantification; Furthermore, this procedure helps resolve the uncertainties in low-level features often involved in image classification and object detection. In fact, various methods are applied in process of IU-- intuitive, flexible and efficient… are examples of methods used in modeling spatial relationships and interrelationship.

1.2 ***IU Brief Background***

The environment scene shows concordant appearance of objects, which is linked to their spatial arrangement and chance of co-occurrence [2][1][31]. However, advanced image understanding requires good interpretation of its spatial information as well. For, contextual models (i.e. Co-occurrence and relative location [2]) enable deeply exploring such details through the region spatial relationships quantification [1]. Furthermore, this procedure helps resolve the uncertainties in low-level features often involved in image classification and object detection [1]. In fact, various methods are applied in process of IU -- **intuitive, flexible and efficient; etc., are examples of method**s used in modeling spatial relationships and interrelationship [1][31].

According to [2], there are two broad categories of IU computational models for which each has been developed to help under different task requirements in object configuration analysis. One of the two --say [a] enables identifying ambiguous objects in a scene [2][16][11] [13][14][19][20]; whereas those in [b] deal much more with observations based analysis [2][11][14][18][20][10]12]. Explicitly:

The first methods group [a] handles the object's exploration from bottom level upward; it is described as low level representation of an image and it is further labelled "Gist" for drawing up contextual ideas before the object actual recognition [2]. However, few diverge on their approach. Thus, some of them consider a correlation of low level features across images surrounding the object or across the category [2][11]13][14][19][20]; while with research progresses, the latest techniques contain co-occurrence of high level features offering instead great support to contextual constraints [2][15].

The second group –i.e. [b] refers almost to the methods applied in advanced of object recognition, since using more complex interpretations and analysis techniques. Among others, this group includes methods involving spatial relationships, spatial context (supporting inter-pixel statistics) [2][11][14][18][20]; pairwise relation (for images inter regions analysis) [12] and semantic context (Enabling recognition accuracy enhancement) useful for co-occurrence understanding label agreement of objects in scene.

1.3 IU Applications & Processing Methods

1.3.1 Applications

For long time ago IU or (scientific) configuration of objects has been an important subject in psychology studies and then in computer vision; all in search for its effects in visual analysis, localization and recognition performance [2][5][6][7][8][9][29]. And afterward, other fields of study (e.g. AI/robotics, engineering fields; etc. ;) got involved.

Image retrieval (IR) is a study field that deals with searching and browsing digital images from database collection [32]. It is closely related to IU from its functional operations— i.e. retrieval action requirements. And due to many involvements of image use or consumption in various [human] works activities, IR function has a high impact in IU at applications' level. Hence, both are hugely interesting in different fields/application areas such as in the fields of image processing, multimedia, digital libraries, remote sensing, astronomy, database applications and related area [32][17][27]

1.3. 2 IU Processing Methods

IU is a multidisciplinary field with many applications. And most of research areas need sufficient knowledge, theories, methods, and techniques from computer science, engineering, mathematics, and even from general and specialized domains; etc., [27][25]. In other words, referring to such background's knowledge coverage, IU process can be obviously a complex task. Nevertheless, choosing a right method /approach for its study and analysis is an important step to get started.

Usually, "when humans are asked to describe a picture, they generally give a list of objects and their relative positions in the scene. (And yet) a closer look at the image reveals that they have omitted a lot of detail in their description" [26]. Moreover, in vast majority of related research works, IU problem is usually broken down into two traditional phases: a low-level segmentation or feature extraction phase, and a higher-level reasoning phase. This latter explores the image features relatively to the object features described in object models of the scene [26][28][29].

Computations modelling are needed in IU advanced study level. The environment scene shows concordant appearance of objects, which is linked to their spatial arrangement and the probability of co-occurrence [2][1][24][26]. However, these details that can be generally grab by systematic observations, requires some computations and even statistical analysis in order to well appreciate the object's image observations accuracy in a given context [24][25][26]27][28][29]. According to [2] there are two broad categories of IU computational models, have been developed to help under different task requirements in object configuration analysis. One of the two --say [a] enables identifying ambiguous objects in a scene [2] [20][16][14] [19] [13][11]; whereas those in [b] deal much more with observations based analysis [2] [11] [14] [18] [20] [10] [12].

The first methods group handles the object's exploration from bottom level upward; it is described as low level representation of an image and it is further labelled "Gist" for drawing up contextual ideas before the object actual recognition [2][16]. However, few diverge on their approach. Thus, some of them consider a correlation of low level features across images surrounding the object or across the category [2] [19] [11] [14] [20] [13]; while with research progresses, the latest techniques [~~which~~] contain co-occurrence of high level features offering instead great support to contextual constraints [2][15].

The second group [i.e. 'b'] refers almost to methods applied in advanced of object recognition, since using more complex interpretations and analysis techniques. This group includes methods involving spatial relationships, spatial context (supporting inter-pixel statistics) [2][11][14][18][20]. The second group includes also pairwise relation (for images inter regions analysis) [2][12] and semantic context (for recognition accuracy enhancement). Examples under category are: "top-down techniques", which are also known as "model-based" and "deformable models" [4]. Deformable models imply some possibility of making change in the model to fit the data in a desired way.

1.4 – **IU through Objects Categorization**

Object categorization (OC) enables positioning, identifying and verifying inside an image all necessary attributes of an object category. Different OC models are designed to facilitate minimizing various defects (e.g. poor quality, noise/background clutter), which stand as barrier in object recognition within an image [21] [1]. In fact, as solution to such a barrier, the recognition accuracy in those models is optimized by using object's appearance information (i.e. Facts for recognizing object classes using visual cues) and context information (i.e. Facts from interaction among objects in context) [21]. In practical instances (Figure 1.4.1 appearance features (e.g. Edge responses, shape and color) are fundamental details considered for object classes identification in real world images; that because of their sensitivity to change in objects classes [1][2][4][21].

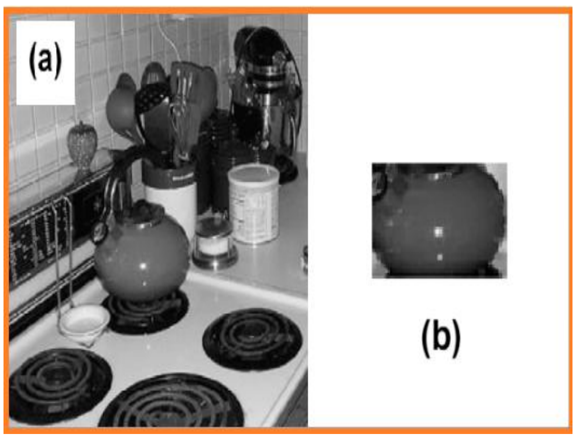

Figure 1.4.1 -- an object appearance (b) does not tell enough about its identity. The scene (a) adds contextual info [i.e. complete scene (a)] about (b) object full details, identifiable now as a kettle.

A suitable process of identifying objects based on contextual information requires good understanding of their positioning in real-world environment. Interposition, support, probability, position and familiar size are the classes of relations developed for analysis of objects in real world scenes context [21][22]. In computer vision, semantic relations (i.e. info about interactions among objects in the scene) also known as context features have made much of their use/application to enhance objects recognition and at the meantime to reduce process time.

There are almost four levels or stages where an object can be explored on an image with the help of object categorization principles developed out of different research work. The four cases are introduced in the following paragraphs.

1.4.1 OC Context Types

They are fundamental and are also called **contextual Features** in IU analysis. And, they are of three groups. That is, semantic context, spatial context (~~position~~) and scale contexts (~~size~~), which are contributed in obtaining objects categorization (locating and identifying objects ~~instances~~) in real-world appearance/view [21][22]. In fact, they have to do with outside object's details produced by other objects surrounding its location plan.

❖ **OC Contexts at Glance:**

With respect to object categorization in real-world scenes context, "semantic" and "spatial" contexts are two principal types in which the word images appearance can be analyzed [2][4][21][22][23].

- Semantic context method has to do with the probability for an object existence only in some and unique scenes. Therefore, objects semantic context imply their co-occurrence with other objects relatively to their appearance or existence in the scenes [21][23][2]. It can be obtained through some labels per pixel showing a pixel existence in a given object, including external knowledge context.

- Spatial context method is derived from Biederman's position class [21]. Contrary to semantic context, it refers to the likelihood or possible chances for an object to exist only in some positions with respect to the occurrence of other objects around it.
- Scale context method has to do with object's size – from pixel level to actual object shape; involves: another object's identification and then the details about specific spatial and depth relations between the target and this object in the setting.

1.4.2 Context Levels Methods

There are two cases: local and global contexts (Figure 1.4.2). The first is about only the details at the object vicinity's areas; whereas the second is about the information taken from the whole image setting containing the object [4][21].These principles are learnt from psychology studies related OR (object recognition) [21].

Figure 1.4.2 Dotted window indicates local context and the straight window indicates the region of interest for the appearance features.

1.4.3 Contextual Interactions

They can then be analyzed within local context (–i.e. contextual interactions at pixel, region and object level) and referring to as pixel, region and object interactions; or at global context level and known as interactions between objects and scenes [21].

1.4.4 Integrating Context

Despite the tasks complexity, some learning techniques have been borrowed in order to benefit from their powerful probabilistic algorithms. However, based [21] classifiers and graphical models are the main groups associating various methods for integrating context. Classifiers help models integrate their context with their appearance features; they are known as more efficient in combining outputs of local appearance detectors with any of contextual features. And **classifiers** main role is to combining the outputs of local appearance detectors with contextual features as obtained from either local or global statistics. The construction of the context feature is done in two stages. In the first stage, the image is processed to calculate the low level and semantic information. An in the second stage, the context feature is calculated at each point by collecting samples of the previously computed features.

2 OBJECT IDENTIFICATION & SPATIAL RELATIONSHIP IN IU

Object identification through image in scene is ordinary wrongly or insufficiently executed whether by an ordinary or say by little specialized person on such a task. In general, people tend to assert, consider, and focus on the most obvious fact accessible /apprehensible by almost everybody. Object's shape, size, color and nature are the common details many observers would produce. The spatial context and the object's contextual identification are generally missing from such report.

However, in the case of learnt researchers and or experts, the concept and perceptions go deeper and wider for an image identification task. According to [30], "When humans look at an image, they see not just a pattern of color and texture, but the world behind the image". Otherwise, the majority of observers would speak/tell only about the object's image. Thus, such above scientist's statement means that the image's observers tend to ignore the role that everything surrounding the given object plays for its identification. Similarly ~~/Likewise~~, [26] stated also stated that "when humans are asked to describe a picture, they generally give a list of objects and their relative positions in the scene. (*Whereas*) a closer look at the image reveals that they have omitted a lot of details in their description."

The above declarations and many related ones from other researchers show altogether the importance of object identification and the spatial relationship in IU. The requirements for this purpose were introduced in previous section as found in reviewed articles. But, humans' perception (though superior) needs some mathematical models to confirm their viewpoint and theoretical analysis versus the computer's results version [24][25[27]. Therefore, a set of mathematical modeling is presented in a coming section. They are directly relevant to common IU analysis methods. Moreover, these computations are necessary also because scene understanding from a single image requires strong assumptions about the world [25][27][28][17][29]. And these assumptions often involve some complex probabilistic formulations into modelled the image processing algorithms. Hence, mathematics/statistics results facilitate drawing up reasonable conclusions when needed.

Again, spatial information holds a very important role in high-level of IU. In fact, Contextual models apply spatial information through the quantification of region spatial relationships to solve about uncertainties in low-level features used for image classification and object detection [31]. And this reference's work discussed intuitive, flexible and efficient methods for modelling pairwise directional, spatial relationships and the ternary between relationship using fuzzy mathematical morphology. For these methods, [31]'s authors have defined **a fuzzy landscape** where directional mathematical dilation with fuzzy structuring elements is used to compute this landscape. And then, they proved how using spatial constraints derived from shadow regions can improve building detection accuracy [31].

The [33]'s outcome provides a layered representation of a scene, which gives some symbolic meaning to the inter-object relationships. Such inter-object relationships are useful for subsequent commonsense reasoning and decision making

Low-level or bottom-up method is known as the fundamental context in object categorization toward image identification and image understanding [26][2][21][29][30]. However, low level features is not warranty for describing high level concepts in the users' mind [32]. Hence, with more evolution in image retrieval researches the concept is moving from keyword to low level and to semantic features (less objective and time consuming) with more easiness in process. Concretely, "all objects in the scene can be identified based on low level features extraction integrated with a proposed line detection techniques". With these considerations [32] proposed a novel technique for objects spatial relationships semantics extraction and representation among objects exists in images. It is about 8 spatial relationship concepts namely: "Front", "Back", "Right", "Left", "Right-Front", "Left-Front", "Right-Back", "Left-Back" concept. Complete concept's details available in reference [32].

Another interesting type of image for objects and spatial relationship is that between fixed and animated object –e.g. robots and object around as case example. With respect to such category of images, [33] presented an algorithm able to describes one more time a scene in a layered representation manner, from labeled point clouds of the objects in the scene using qualitative description of the structure of the objects, and symbolic relationships. This is achieved by constructing contact point networks of the objects considered as their topological representations in the scene, and the regions of contact between those objects.

In every accessible article included into this papers review section, the emphasis was found to be around the high position given to image's objects description/identification. And much more it has been about the role [that] played by the spatial relationship to improve objects' IU, including enhancing the outcome accuracy. Some other important analysis details are left in reviewed and related articles, but available upon your curiosity when would go through them,

3 IU METHODS -- PROCESSING METHODS

IU is not only a multidisciplinary field with many applications. But also, most of research areas need sufficient knowledge, theories, methods, and techniques from computer science, engineering, mathematics, and even from general and specialized domains [27]. In other words, referring to such background's knowledge coverage, IU process can be obviously a complex task. Nevertheless, choosing a right method /approach for its study and analysis is an important step to get started.

Usually, "when humans are asked to describe a picture, they generally give a list of objects and their relative positions in the scene. But, taking) a closer look at the image reveals that they have omitted [missed out] a lot of detail in their description" [26]. Moreover, in vast majority of related research works, IU problem is usually broken down into two traditional phases: a low-level segmentation or feature extraction phase, and a higher-level reasoning phase. This latter explores the image features relatively to the object features described in object models of the scene [26][28][29]}. And with the advance in technology, the applications of computer visions knowledge contributed enough in such task initially done under human control alone. It has much help IU in finding usually missing details with reference to human heuristic analysis.

Naturally the environment scene shows concordant appearance of objects, which is linked to their spatial arrangement and chance of co-occurrence [2][1][Db-1ext: IU="33"] . However, in object learning / exploration process, understanding this detail is strongly helpful since it prevents confusing the object look [when] under light or noise's influence for instance; this includes other physical variations due to environment changes (e.g. shadow, heat, illumination…) [2][1][4][3] [8] [21] [24][26].

In fact, the complexity of IU subject (Background & applications) makes it just vague to attempt listing down the available methods participating to IU works. However, this section reviewed some of common IU methods encountered in the articles used into this paper.

The first among the most obviously used IU methods are "bottom-up and Top-down method discussed in [4] lecture material. The first requires segmenting/fragmenting the image into regions forming some object shapes; then drawing up the obtained objects using representations. This method corresponds to IU low-level processing – goes from raw image data**,** to bring about the object shape representation, and ends up with structured analysis or decision [1][2][4][26]. The other method is obviously the advance way. It involves designing some hypotheses on choices made; then applying the image data in these hypotheses testing "accept/reject"; and then draw up final conclusions.

In general, spatial details and relationship models are fundamental facts/inputs in IU. In fact, many great improvements in accuracy in IU applications show the importance of spatial information and the effectiveness of the relationship models in modeling and quantifying this information [31]. However, according to [1][2], intuitive, flexible and efficient methods are the most used in practice for pairwise and directional spatial relationships and the **ternary** between relationship when applying fuzzy mathematical modeling morphology as shortly explained next based on [31].

Object-oriented classification is another great IU method [31]. It allows users exploiting structural information to perform region-based classification instead of classifying individual pixels. In examples, various works have succeeded performing classification using the spatial context of each pixel according to a hierarchical multi-level representation of the scene. And **[31].**

According to [25] a sound theoretical study of image understanding can be done out of the following study point and analysis perspectives:

a) General understanding and representation (geometry and perspective projections and Euclidean) of an object and its image
b) Image characteristics and coordinates
c) Irreducible representations and 3D rotation; 3D rotation representation;
d) Algebraic invariance of image characteristics
e) Scene and images characterizations
f) Shape from motion, angle, texture and surface.

The material presents a description of various processes of object imaging and geometrical representations. These are also supported by some computation models and well-known mathematical theorems. This study material differs from [26] (whose analysis is more mathematical oriented) from not involving color image; but closely similar in theory objectives.

3.2 Other Processing Methods / Techniques

Under IU, **approach to color** can be used to segment and analyze surfaces with color variations, which come from highlights and shading. This method has been use into [26] work, which was based on a theory-the Dichromatic Reflection Model. The use of dichromatic theory enable separating a color image into two intrinsic reflection images: an image with the highlights, and the original image without highlights. Relatively to IU using object identification and spatial relationship, this approach is an interesting contribution, especially with comparison to the works done by [21] [4][23] and many related ones. For, a reflection model can be applied to include color image segmentation into an image analysis. **Overall** the expected result is the (non-color and) color image understanding system that is able to produce physical descriptions of the reflection processes occurring in the scene [26].

Understanding spatial relationship between objects has always played a great role in IU analysis/study. And practically "***it is the qualitative structure of the objects in an environment and the relationships between them which define the composition of that environment, and allow for the construction of efficient plans to enable the completion of various elaborate tasks***" [19]. By the way, according to [31], a structural way of modelling context in images is through the quantification of spatial relationships. All these show that in additional to the object characteristics (color, texture, size, shape; etc.), the objects' observer/explorers need or take advantage of the ***environment and the relationships*** as the features offered in their image in order to run easily the study required task.

4 IU PROCESSING COMPUTATIONS MODELING

It is interesting to remember that throughout IU works, typical relationships studied in the literature include geometric (size, position, shape, and orientation), topological (set relationship and neighborhood structure), semantic (similarity and causality), statistical (frequency and co-occurrence), and structural (spatial configuration and arrangement patterns) relationships. In fact, the methods applied for computations modelling along with above stated are subject to which way the objects/regions are modelled. Therefore, commonly used approaches include grid-based representations, centroids and minimum bounding rectangles. But, fixed sized grids are also not generally applicable, as part of processes limitation [31]

Here are the examples of image mathematical modelling out of OC as in [21]; and some related details were introduced earlier in this paper.

4.1 Spatial context (SC) Computational Method

Several year ago, many approaches acknowledged spatial context as suitable in IU improving recognition accuracy. Practically it is incorporated from inter-pixel statistics and from pairwise relations between regions in images . However, in recent years based on [21], a contribution from Shotton et al., had enhanced the process by introducing inter-pixel statistics for OC (object categorization).. *In their framework,* a unary classifier λ_i captures spatial interactions between class labels of neighboring pixel. And SC (Spatial context) is represented by a look-up table with an entry for each class c_i and pixel index (i) such as $\lambda_i(c_i, i; \theta_\lambda) = \log \theta_\lambda(c_i, \hat{i})$; with index $\hat{i}$ a normalized version of the pixel index for the image. Here, θ_λ represents the model's parameters..

- Local context (LC) Computational Process

Local context information is that from areas around the object. In literatures, there are many OC models, which have applied LC from objects, patches and pixel information around targeted object. And the procedure allowed them achieving successfully OC task [21]. A particular contributors are Kruppa and Schiele with their face detection algorithms [21]. The features of their detector capture local arrangements of quantized wavelet coefficients, based on Naive Bayes classifier:

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

Further investigations showed the role of local context for face detection algorithms. At the core of the detector there is Naive Bayes classifier:

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

. θ Corresponds to acceptance threshold; and P_k the likelihood functions on the coarse quantization i(x) and j(y) of the feature position; etc.

4.2 Contextual Interactions (CI) Computational Method

Basically, here are summarized some important information under this title.

a) **Local interactions**

With the objects in highly cluttered scenes, recognition performance [can be improved dramatically] enhancement is fairly possible by applying bottom-up (i.e. LC analysis) attentional frameworks. The frameworks' associate features ensure the result/outcome quality. Actually, bottom-up processing goes about analyzing pixel interactions, which applies a concept of similarity in neighboring pixels. Many works contributed to OC frameworks at pixel level; and the case of He et al., [21] has particularly solved the problem of obtaining contextual features by using pixel level interactions.

b) Global interactions

Global context can be recognized by means of a scene-centered representation. And the image modelling basic computations are as follow. In fact, the Object-scene interactions are modeled using training image clusters, which give hints about what objects are in the query image. The relationship between object categories O, their spatial location x within an image, and their appearance g can be modeled using the following joint distribution computation:

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

With: N, the number of images each with M_i object proposals over L object categories; and $p(o_{i,j}|h_{i,j}=m,\theta_m)$ the likelihood of the object categories that appear in the image – these representing the object-scene interactions.

Object-scene interactions are modeled using training image clusters, which give hints as to what objects are depicted in the query image and their likely location.

c) Integrating Context (CI) Computational Method

Basically, here are summarized some important information under this title.

4.3 *Graphical models*

Graphical Models offer simple ways to visualize the structure of a probabilistic model. They provide a powerful and flexible framework for implementing global probability distributions defined by relatively local constraints. And global probability distributions are defined on directed graphs for expressing causal relationships between random variables. Thus a joint probability distribution for the directed graphical models can be computed by:

$$P(x) = \prod_i P(x_i|pa_i)$$; with pa_i the potential function over the maximal cliques C of the graph. Such graphical models assume that objects are conditionally independent given the scene [21].

There are also special cases of undirected graphical models for modeling context that include Markov random fields (MRFs) [6] and conditional random fields (CRFs). However, further computations modeling can be performed for "conditional random fields", which will allow learning about an image's conditional distribution over the class labeling; and those details and others can be found in Ref. [21].

5 REVIEWED ARTICLES BASED CONCLUSIONS

IU reviewed articles in this paper showed that there are two general methods /approaches of learning or exploring an image's object. One of the two is object's exploration from bottom level upward or image low level representation. This brings about the contextual ideas before the object actual recognition. And the second approach applies methods such as spatial relationships, spatial context pairwise relation and semantic context.

Based reviewed article, contextual models help resolve for uncertainties in low-level features used for image classification and object detection by exploiting spatial information through the quantification of region spatial relationships (e.g. Co-Occurrence and Relative Location [2]).

In modeling and quantifying ~~the~~ information, spatial information and the effectiveness of the relationship models contribute significantly to the improving accuracy in IU applications.

Studies showed that [greater] accuracy optimization in IU exploration can be achieved by combining co-occurrence and spatial context rather than using co-occurrence alone.

A good understanding of OC (Object categorization) can better help in image analysis and computations modelling.

OC can be comprehensively discussed under some specific perspectives. These include OC fundamental contexts (i.e. Semantic, spatial & Scale contexts); Context levels; Contextual interaction (i.e. Local vs. Global); and Integrating context (Classifiers, Graphical model). However, overall these cases "semantic" and "spatial" contexts are two principal types in which the word images appearance can be better analyzed, hence the OC fundamental context as the whole.

Etc.

REFERENCES

- [1] Image Classification and Object Detection Using Spatial Contextual Constraints, Selim Aksoy, R. G¨okberk Cinbi¸s, H. G¨okhan Ak¸cay, 2005; Bilkent University Department of Computer Engineering, Bilkent, 06800, Ankara, Turkey
- [2] Carolina Galleguillos, Andrew Rabinovich and Serge Belongie, (2008) Object Categorization using Co-Occurrence, Location and Appearance;
- [3] --- Hui Hui Wang, Dzulkifli Mohamad & N.A. Ismail, Semantic Gap in CBIR: Automatic Objects Spatial Relationships --Semantic Extraction and Representation, International Journal Of Image Processing (IJIP), Volume (4): Issue (3).
- [4] Lecture 21 Image Understanding; by Bryan S. Morse, Brigham Young University, 1998–2000; Last modified on March 20, 2000 at 6:00 PM,
- [5] M. Bar and S. Ullman. Spatial context in recognition. Perception. 25:343-352., 1993.
- [6] I. Biederman. Perceiving real-world scenes. Science, 177(7):77–80, 1972.
- [7] Biederman, R. J. Mezzanotte, and J. C. Rabinowitz. Scene perception: Detecting and judging objects undergoing relational violations. Cognitive Psychology, 14(2):143–177, April 1982. [11] P. Lipson, E. Grimson, and P. Sinha. Configuration based scene classification and image indexing. In CVPR, 1997.
- [8] S. E. Palmer. The effects of contextual scenes on the identification of objects. Memory and Cognition, 1975.
- [9] L. Wixson and D. Ballard. Using intermediate objects to improve the efficiency of visual search. IJCV, 12(2):209–230, 1994.
- [10] T. Cour, F. Benezit, and J. Shi. Spectral segmentation with multi-scale graph decomposition. In CVPR, 2005.
  →~~Replaceable~~…
- [11] X. H [5] T. Cour, F. Benezit, and J. Shi. Spectral segmentation with multi-scale graph decomposition. In CVPR, 2005.e, R. S. Zemel, and M. ´ A. Carreira-Perpi ˜ n´ an. Multiscale conditional random fields for image labeling. In CVPR, 2004.
- [12] S. Kumar and M. Hebert. A hierarchical field framework for unified context-based classification. ICCV, 2005.
- [13] K. Murphy, A. Torralba, and W. Freeman. Using the forest to see the tree: a graphical model relating features, objects and the scenes. NIPS, 2003.
- [14] J. Shotton, J. Winn, C. Rother, and A. Criminisi. Textonboost for image understanding: Multi-class object recognition and segmentation by jointly modeling appearance, shape and context. IJCV, pages 1–22, 2007.
- [15] A. Singhal, J. Luo, and W. Zhu. Probabilistic spatial context models for scene content understanding. In CVPR, 2003.
- [16] A. Torralba. Contextual priming for object detection. IJCV, 53(2):169–191, 2003.
- [17] Image Understanding Methods in Biomedical Informatics and Digital Imaging; by: Marek R. Ogiela1 and Ryszard Tadeusiewicz; Journal of Biomedical

Informatics 34, 377–386 (2001), Elsevier /Direct science; Volume 34, Issue 6, December 2001, Journal of Biomedical Informatics; DOI:10.1006/jbin.2002.1034.

- [18] A. Torralba, K. Murphy, W. Freeman, and M. Rubin. Context-based vision system for place and object recognition. In CVPR, 2003.
- [19] J. Verbeek and B. Triggs. Scene segmentation with CRFs learned from partially labeled images. In NIPS, 2007.
- [20] L.Wolf and S. Bileschi. A critical view of context. IJCV 69(2):251– 261, 2006.
- [21] Carolina Galleguillos & Serge Belongie (2010) Context based object categorization: A critical survey , Computer Vision and Image Understanding, Computer Vision and Image Understanding xxx (2010) xxx–xxx, 1077-3142/$ - see front matter 2010 Elsevier Inc. All rights reserved. DOI:10.1016/j.cviu.2010.02.004
- [22] I. Biederman, Perceiving real-world scenes, Science 177 (7) (1972) 77–80.
- [23] M. Bar, Visual objects in context, Nature Reviews Neuroscience 5 (8) (2004) 617–629.
- [24] Artificial Intelligence and the Science of Image Understanding; by: B. K. P. HORN; Massachusetts Institute of Technology, Cambridge, Massachusetts; References: Computer vision and Sensor-based Robots, Edited by Geoges G.D. dd and Lotha Ressel (Planum Publishing Corporation, 1979); pp.75-77
- [25] Group-Theoretical Methods in Image Understanding --With 138 Figures; by: Kenichi Kanatani; Springer-Verlag Berlin Heidelberg, New York, London, Paris, Tokyo, Hong Kong
- [26] A Physical Approach to Color Image Understanding; by Gudrun J. Klinker, Steven A. Shafer, and Takeo Kanade; International Journal of Computer Vision. 4, 7-38 (1990) © 1990 Kluwer Academic Publishers. Manufactured in The Netherlands.
- [27] Overview of Biomedical Image Understanding Methods; by: Joo-Hwee Lim, Sim-Heng Ong, Wei Xiong;, Jierong Cheng, Ying Gu and Shimiao Li; DOI: 10.1002/9781118715321.ch1; Copyright © 2015 John Wiley & Sons, Inc.
- [28] A.R. Hanson and E.M. Riseman, "VISIONS: A computer system for interpreting scenes." In A.R. Hanson and E.M. Riseman (4s.). Computer Vision Systems, New York: Academic Press, pp. 303-333, 1978.
- [29] T. Kanade, "Region segmentation: Signal vs. semantics." Proc. 4th Intern. Joinf Conf Pattern Recog. pp. 95-105. IEEE, Kyoto, Japan, November 1978.
- [30] Seeing the World Behind the Image -- Spatial Layout for 3D Scene Understanding; by: Derek Hoiem, 2007; Thesis Doctor of Philosophy Robotics Institute; Carnegie Mellon University Pittsburgh, William T. Freeman, Massachusetts Institute of Technology; Copyright ©2007 by D H. All rights reserved.
- [31] Image Classification and Object Detection Using Spatial Contextual Constraints; by: Selim Aksoy, R. Gokberk Cinbis, and H. Gokhan Akça; Bilkent University Department of Computer Engineering Bilkent, 06800, Ankara, Turkey
- [32] Semantic Gap in CBIR: Automatic Objects Spatial Relationships --Semantic Extraction and Representation ; by: Hui Hui Wang, Dzulkifli Mohamad and N. A. Ismail; International Journal Of Image Processing (IJIP), Volume (4): Issue (3)
- [33] Learning Spatial Relationships between Objects; Benjamin Rosman_ and Subramanian Ramamoorthy; March 31, 2011.