World Academy of Science, Engineering and Technology
International Journal of Computer and Information Engineering
Vol:13, No:6, 2019

Object Recognition Approach Based on Generalized Hough Transform and Color Distribution Serving in Generating Arabic Sentences

Nada Farhani, Naim Terbeh, Mounir Zrigui

Open Science Index, Computer and Information Engineering Vol:13, No:6, 2019 publications.waset.org/10010509/pdf

***Abstract*—**The recognition of the objects contained in images has always presented a challenge in the field of research because of several difficulties that the researcher can envisage because of the variability of shape, position, contrast of objects, etc. In this paper, we will be interested in the recognition of objects. The classical Hough Transform (HT) presented a tool for detecting straight line segments in images. The technique of HT has been generalized (GHT) for the detection of arbitrary forms. With GHT, the forms sought are not necessarily defined analytically but rather by a particular silhouette. For more precision, we proposed to combine the results from the GHT with the results from a calculation of similarity between the histograms and the spatiograms of the images. The main purpose of our work is to use the concepts from recognition to generate sentences in Arabic that summarize the content of the image.

***Keywords*—**Recognition of shape, generalized hough transformation, histogram, Spatiogram, learning.

I. Introduction

SHAPE recognition is a key in the development of computer vision based on image analysis. Computers recognize shapes by analyzing the various aspects of an object that needs to be simplified in a representation called "shape descriptor" and containing the most relevant information to identify an object, then a comparison between these characteristics and the characteristics of the models stored in a database will take place. According to Neisser [1], recognition of an object takes place in three steps: a selection step, which is used to extract the most relevant information by simplifying the object in a representation, will be performed. A second step is the detection of objects which designates the location of the objects of a given category. And finally, the last step is to segment classes of objects and this by determining the pixels of the image belonging to an object of one of the classes.

When we talk about locating or detecting objects, we are talking about two different problems. The first is the location of object instances which consists in the detection and the localization of the same object in different images. The second is the location of categories of objects used for the detection of different objects but belonging to the same category.

Object recognition consists of the prediction, using an algorithm, of the recognition of object(s) that is in an image. In other words the input of the algorithm is an image and the output will be the class of the object in the case of object category recognition or the reference of a well-defined object in the case of recognition of the object instances. In case the image contains various objects, we can have at the exit of the algorithm a list of objects. In our work, we are interested in the detection of object instances.

This paper is organized as follows. An overview about the main methods used in the state of the art on the recognition of object instances is presented in the second section. In the third section, the related works are presented. The fourth section presents our contribution with the different steps to achieve our goal (preprocessing, contour detection, image feature extraction, application of the Generalized HT and phase of recognition). Some experimental results as well as the discussion will be presented in Section V. Finally, Section VI concludes this paper.

II. Methods for Object Instance Recognition

We present successively the methods of the state of the art concerning the object instance recognition which are geometric, global and local methods [2].

A. Geometric Methods

The first category of object instances recognition methods are the geometric methods which consist in the representation of the reference objects by their contours. These methods are divided into two types, hash-based methods and model-object alignment-based methods. The second category is that of geometric alignment methods that have a 3D model of the object sought. The primitives (such as lines) composing this model are aligned with the primitives detected in the image and the quality of the alignment determines the presence or not of the object. An example of these methods the interpretation trees; nevertheless, these methods suffer from two defects. The first is that in certain types of categories of objects, such as trees, they cannot be defined only by their contours.

The second lies in the sensitivity of these methods against imprecise detection of the contours generally due to image acquisition conditions.

B. Global Methods

This type of recognition methods consists of calculating a signature of the image that can be as an example, the distribution of colors in the image as color histograms and

Nada Farhani is with the LaTICE Laboratory, University of Monastir, Tunisia, ISITCOM Hammam Sousse, University of Sousse, Tunisia (e-mail: farhaninada@yahoo.fr)

Naim Terbeh and Mounir Zrigui are with the LaTICE Laboratory, University of Monastir, Tunisia (e-mail: naim.terbeh@gmail.com, mounir.zrigui@fsm.rnu.tn).

Open Science Index, Computer and Information Engineering Vol:13, No:6, 2019 publications.waset.org/10010509/pdf

textures. This signature is therefore taken in the entirety of the image and is calculated for a set of images representing an object instance. Generally, these methods are very simple as they are robust to changes in contrast and illumination under the condition of having a large image base at different illumination conditions. This is why global methods require a large amount of data. As they give bad results in the case of presence of charged background and occultation.

C. Local Methods

Local methods come to overcome the main problems posed by global methods. They do not treat the images as a whole but as a collection of local regions, which are generally parts of images or rectangular or square and correspond to flat surfaces, so their scale and orientation can be reduced to standard values. When these regions are segmented within the objects, they become independent and do not influence the charged background. When these regions are small, they are most often totally obscured or completely present. The principle of local methods is to represent the reference images through local regions and to store the descriptors of these regions in a database.

Local methods are the most used today for the visual recognition of object instances and other domains also [3]-[5] thanks to their good management of occultation, charged background and changes of point of view as well as their speed.

III. Related Work

To simplify the formation of deeper networks than previously used networks, in [6], the authors presented a residual learning framework. In an explicit way, Zhang et al. did not choose to learn unreferenced functions but they reformulated the layers as residual learning functions with reference to the layer inputs.

In [7], the authors propose a model generating natural language descriptions of images and their regions. To learn more about the correspondence between images and language, the approach proposed by Karpathy et al. exploits the dataset of text descriptions of images. The base of the alignment model is a new combination of two-way recursive neural networks on sentences, convolutional neural networks on image regions, and a structured objective that serves to align the two modalities via multimodal integration.

In [8], the proposed system is divided into two stages. In fact, this system is used to transmit the input image via the Deep MANTA network which produces the visibility properties of the parts, the 2D bounding boxes and the associated vehicle geometry. The Deep MANTA network is created in order to have other attributes (similarity of models, visibility of parts and location of vehicle parts), as well as vehicle detection through a coarse-to-fine bounding box proposal.

The approach proposed in [9], the focal loss applying a modulating term to the loss of cross entropy in order to weight the many easy negatives and to focus learning on concrete examples. Vaillant et al. applied in their work [10] convolutional neural networks to the recognition of handwritten figures.

In order to detect faces, in their works [11], the authors exploited detectors of boosted objects. The use of integral channels [12] and HOG functions [13] has led to the emergence of effective methods for pedestrian detection. In the classical computer vision, the sliding window approach was the main model of detection, with the appearance of deep learning [14].

With regard to work based on two-stage detectors, in [15], the first phase is the generation of an isolated set of candidate propositions in the obligation to include all the objects as well as the filtering of the majority negative locations. The second phase classifies proposals into foreground/background. For the purpose of improving the second-stage classifier, R-CNN [16] has made modifications to this level to have a convolutional network that offers significant gains in accuracy. In turn, R-CNN has also been improved over the years, at the same time at the speed level [17] and by exploiting proposals for learned objects [18].

To improve timeliness and to have a faster classifier than that used by the RCNN, Region Proposal Networks (RPPs) incorporated proposal generation with the second-stage classifier into a single convolutional network [18]. As well, several extensions have been proposed in this framework [19]-[22].

An example of the first modern object detectors, is OverFeat [23] which is a one-stage detector based on deep networks. One of the newest detectors is SSD (Single Shot MultiBox Detector) [24], where its approach is based on a convolutional feed-forward network generating scores for the presence of object class instances in fixed size bounding boxes that were generated before, then a non-maximal deletion step for the production of the final detections.

YOLO9000 [25] also a real-time framework is designed to detect more than 9000 categories of objects while optimizing the classification and detection.

IV. Methodology

Pattern recognition is one of the most important challenges today [26].

The work of [27] followed some common steps such as pre-processing, which is an important step in the image processing domain and is used to improve the quality and content of an image to extract information.

The next step is edge detection, which significantly reduces the amount of data and also eliminates information that may be considered less relevant, while preserving the important structural properties of the image.

After edge detection, segmentation is the next step that allows the extraction of structural information from an image that cannot be seen with the naked eye. So, the aim is the cutting of the image into various regions, in which the pixels must verify a certain criterion of homogeneity and finally, the features extraction since in the fields of pattern recognition, image processing and machine learning, features extraction plays an important role. The extracted features must contain

Open Science Index, Computer and Information Engineering Vol:13, No:6, 2019 publications.waset.org/10010509/pdf

relevant data from the input data. A number of features for each image are extracted, describing its high level content information. Then, according to the similarity of these vectors, a comparison between these two images is made. In our contribution, we will use the color information with the Generalized HT to recognize objects present in an image to use them later in the production of sentences in Arabic.

A. Global View of the Proposed System

To achieve our goal of this work, we must go through the object recognition phase where the input image passes through a succession of steps:

Pretreatment is followed by the extraction of the color distribution in the image, and then a contour detection step to extract the characteristics of the image at the gray level using the Generalized HT, to obtain the recognized objects in the end. Fig. 1 gives an overview of the proposed system.

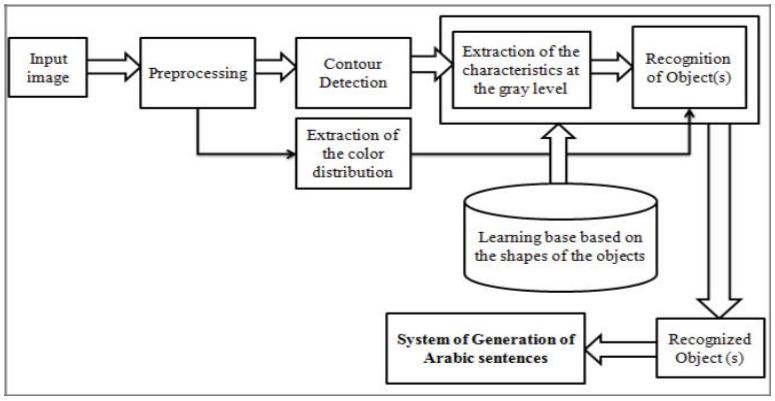


Fig. 1 System overview of the main steps to recognize object(s) in an image

B. Detailed System Description

1. Pretreatment

The pretreatment of the images is a preliminary step before starting the application of digital processing on the image to be processed, and mainly involves pre-treatments such as contrast enhancement which includes the increase in dynamic range, the isolation and the improvement of the perception of attributes. Also, it consists of the morphological operations such as insulation of regions, plugging of holes, and elimination of small noisy structures. We have also the filtering operation in order to attenuate noise but filtering while preserving the edges. The pretreatment also touches the color throw the conversion between color spaces: RGB, XYZ, HSV and YUV.

2. Contour Detection

Contour detection is primarily an important tool in the field of image processing for the detection and extraction of features. According to [28], studies were made for comparison between the different contour detection filters and they showed that the canny filter gave better results. In fact, in 1986, Canny proposed a study on the detection of the contour and he formalized three criteria that must validate a detector of the contour.

- Detection: noise robustness
- Location: precision of the location of the contour point
- Uniqueness: only one answer per contour

3. Extraction of the Color Distribution

As color is important for an image, we proposed to store these features to help in decision making afterwards.

We use:

- Histograms that present a tool for counting the number of occurrences of each value in an image. For a color image, we will have a histogram on each component of the system of representation of the color that is RGB, XYZ, HSV or YUV.

The lack of spatial information, in standard color histograms, has led us to the use of Spatiogram(s).

- Spatiogram(s) are generalized histograms describing more than the occurrence of the pixels in each color box, the average and the covariance of the coordinates of the pixels, which makes it possible to capture the spatial distributions of the different image colors.

4. Object Recognition

- **Application of the Generalized Hough Transform (GHT)**

The Hough transformation was first developed for the detection of analytically defined forms (such as lines, circles, ellipses, etc.). GHT can be used for the detection of arbitrary forms (forms that do not have a simple analytical form).

It requires the complete specification of the exact form of the target object. In the original formulation called classical GHT, a characteristic space (composed of contour points of the model and their vectors towards a reference point) is transformed into a Hough space (rotation, scale, and displacement of the model in picture). In this transformation, the rotated and scaled versions of the vector of each point of model outline are superimposed on each edge point found in the image to vote in the Hough space.

The maximum value in this Hough space corresponds to the rotation, scaling, and moving parameters of the model in the image. This type of transformation is robust to partial or slightly deformed shapes as well as the presence of additional

Open Science Index, Computer and Information Engineering Vol:13, No:6, 2019 publications.waset.org/10010509/pdf

structures in the image. It is also noise tolerant and he can find several occurrences of a form during the same treatment. The instructions of the GHT algorithm are presented below.

➢ **First Steps:**

For a form type (reference form):

(1) Choose a reference point O "center of gravity", with coordinates (x_o,y_o).

For each point P of the contour:

(2) Draw: PT the tangent to the contour on P, $\vec{G}$ the direction of the gradient (normal to PT), $\bar{r}$= PO (distance between P and O) and γ: Angle under which is seen the center of gravity.

(3) Calculate Θ: Gradient angle.

(4) Save the reference point (xo, yo) as a function of Θ

(5) Finally the model is presented of the form; Model = list of triplets {Θ, γ, r}.

We group the triplets of the model in an array R (R-table) where we associate to each value of Θ a pair (γ, r) such that these pairs correspond to the same value Θ. The table is filled in the following way with $m \neq n \neq k$ but we can also have equal values of these parameters. So we will have an independent R-table for each different form (different object). These tables allow us to use contour points and gradient angle Θ to recalculate the location of the reference point in the detection phase. Fig. 2 shows the GHT principle.

TABLE I
THE GHT

Value of Θi	*Pair(s) (γ,r)*
Θ_1	$(\gamma_1^1,r_1^1), (\gamma_2^1,r_2^1),...(\gamma_m^1,r_m^1)$
Θ_2	$(\gamma_1^2,r_1^2), (\gamma_2^2,r_2^2),...(\gamma_n^2,r_n^2)$
⋮	⋮
Θ_i	$(\gamma_1^i,r_1^i), (\gamma_2^i,r_2^i),...(\gamma_k^i,r_k^i)$

➢ **Detection:**

(1) For each edge point P (x, y)

(1.1) Using the gradient angle Θ, retrieve from the table R all the values (γ, r) indexed under Θ.

(1.2) We can say that the portion of the contour centered on P votes for a center of gravity whose coordinates are given by (γ, r):

$$x_o = x + r\cos\gamma.$$
$$y_o = y + r\sin\gamma. \quad (1)$$

(1.3) Increase the counters (vote)

(2) If many points vote for the same point, the form is recognized.

Fig. 3 outlines the steps to be followed for object recognition.

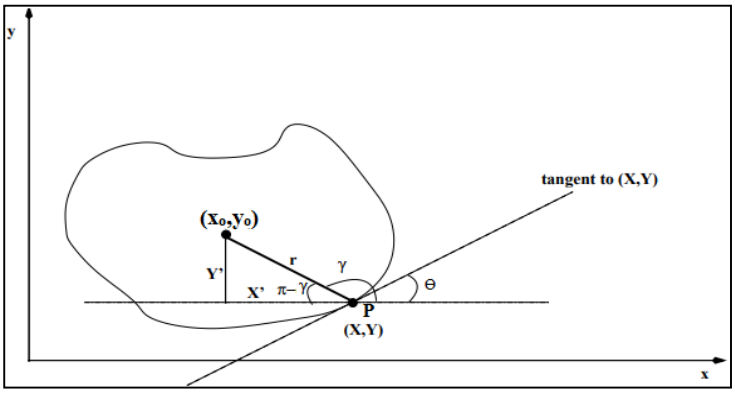


Fig. 2 Steps (1), (2) and (3) of the GHT

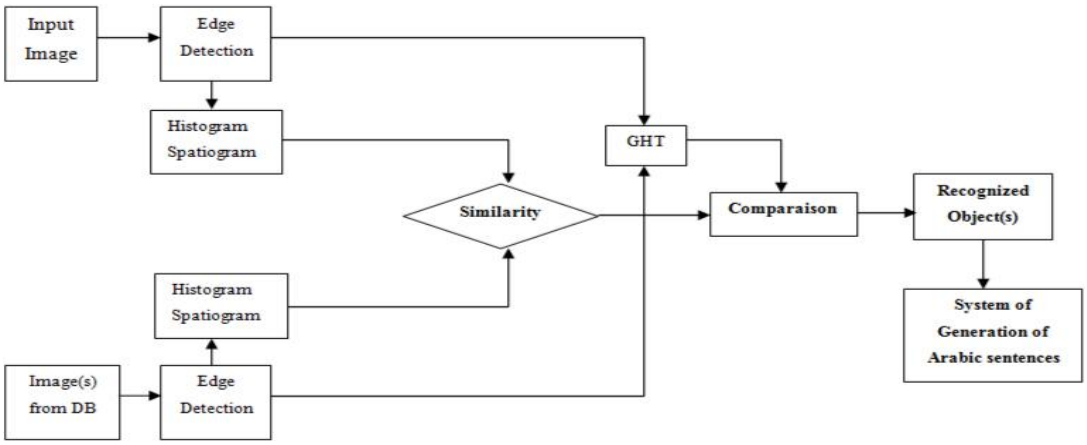


Fig. 3 Overview of the object reorganization steps for an image

Open Science Index, Computer and Information Engineering Vol:13, No:6, 2019 publications.waset.org/10010509/pdf

V. Tests and Results

This section describes some experimental results to evaluate the ability of histogram(s), spatiogram(s) and GHT to locate and recognize objects in an image.

A. Image Database

For the database we used images from ImageNet and Pascal VOC and also we have collected images from the Internet. All these images are divided according to their categories into test images and others into images forming the learning base. We tried to collect a large number of images according to their category so that each one can cover the maximum of images in different positions and different contrasts.

TABLE II
Number of Images for Each Category.

Categories	*Number of images*
Person	*432*
Dog	*189*
Elephant	*202*
Giraffe	*182*
Apple	*35*
Flours	*103*
Guitar	*186*
:	*:*
House	*280*

B. Experimental Results

From the image database collected. We will test our algorithms on some categories of objects. The objects in the learning database are tagged to use the tags after. Subsequently, we will associate the object to recognize its histogram and its spatiogram to compare them later to those of the objects of the learning base by calculating a similarity factor.

➢ **Practical case**

To give some practical results of the algorithm, we will take the case of an apple as an object to recognize. We will make the comparison with objects of circular shapes and the object having a similarity factor close to 1; it is considered that the object is recognized.



Fig. 4 The test image of an apple

Table III shows the comparison results with different objects.

From the results, we find that the values close to 1 are those associated with the images which are the most similar to the test image, despite having several objects of the same circular form.

C. Discussion

This approach has a part that serves to preserve the color information and that is consists on the use of histograms and spatiograms to help the recognition and this by a comparison between the histograms and the spatiograms of the template (form from the image database, compared one by one) with those of the unknown object. And for more precision, by averaging between the two values from the comparison, the closest object is chosen. The results obtained show that the use of the average between the two gives a better rate than the use of each alone. The comparison between the approaches gave a precision rate of 96.6667%. Compared with another works, [29] where the precision rate was 93% and also with [30], where the precision rate for histograms was 95% and spatiograms was also 95%, we have obtained good results that allow us to use it to generate Arabic sentences from the recognized objects.

TABLE III
Results of a Comparison Between the Test Image and Some Images from Database with Circular Form

Image	*Hist_similarity*	*Spatio_similarity*	*Average*
	***0.947066***	***0.923508***	***0.9353***
	***0.977446***	***0.937913***	***0.9577***
	*0.229233*	*0.156628*	*0.1929*
	*0.058742*	*0.035666*	*0.0472*
	*0.027292*	*0.019593*	*0.0234*
	*0.034039*	*0.021104*	*0.0276*
	*0.069248*	*0.051686*	*0.0605*

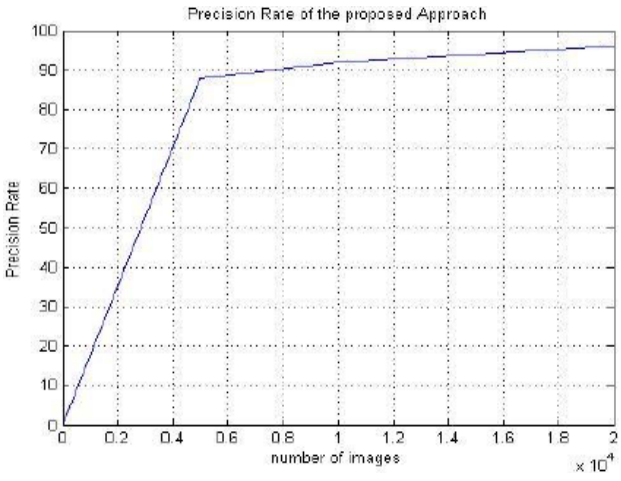


Fig. 5 Precision rate of the proposed approach

VI. Conclusion

This paper presents an approach to object recognition in an image which is divided into two parts; the use of histograms and spatiograms and also the use of GHT. In this approach, we

try to have advantages from the color image and when it is converted in grayscale image (on which we apply GHT) in order to detect and recognize arbitrary shape with different positions and contrast. So we have combined the use of histograms, spatiograms and GHT to have better rate of recognition. To improve this approach, it is intended to replace GHT by Invariant Generalized Hough Transform (IGHT) which further takes into consideration rotations and scaling translations to properly cover position-point objects. Also we can use the HOG (Histogram of Oriented Gradients) to improve the results. We can therefore conclude the importance of the results obtained to use the concepts extracted from the input images to generate Arabic sentences [31], [32] that summarize the content of the image [33].

REFERENCES

[1] Ulric Neisser. Cognitive Psychology. Appleton-Century-Crofts, New York (1967).
[2] Eric Nowak. Recognize object categories and object instances using local representations. Informatique (cs). Institut National Polytechnique de Grenoble – INPG. (2008).
[3] N Terbeh, M Labidi, M Zrigui, Automatic speech correction: A step to speech recognition for people with disabilities.Information and Communication Technology and Accessibility (ICTA). (2013).
[4] A Zouaghi, L Merhbene, M Zrigui. A hybrid approach for arabic word sense disambiguation. International Journal of Computer Processing Of Languages 24 (02), 133-151 (2012).
[5] S Mansouri, M Charhad, M Zrigui. A Heuristic Approach to Detect and Localize Text on Arabic News Video. Computación y Sistemas 22 (1). (2018).
[6] He, K., Zhang, X., Ren, S., & Sun, J. Deep residual learning for image recognition. pp. 770-778. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (2016).
[7] Karpathy, A., & Fei-Fei, L. Deep visual-semantic alignments for generating image descriptions. pp. 3128-3137. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (2015).
[8] Chabot, F., Chaouch, M., Rabarisoa, J., Teulière, C., & Chateau, T. Deep MANTA: A Coarse-to-fine Many-Task Network for joint 2D and 3D vehicle analysis from monocular image (2017).
[9] Lin, T. Y., Goyal, P., Girshick, R., He, K., & Dollár, P. Focal loss for dense object detection (2017).
[10] R. Vaillant, C. Monrocq, and Y. LeCun. Original approach for the localisation of objects in images. IEE Proc. on Vision, Image, and Signal Processing (1994).
[11] P. Viola and M. Jones. Rapid object detection using a boosted cascade of simple features. In CVPR (2001).
[12] P. Dollar, Z. Tu, P. Perona, and S. Belongie. Integral channel ´ features (2009).
[13] N. Dalal and B. Triggs. Histograms of oriented gradients for human detection. In CVPR (2005).
[14] Krizhevsky, I. Sutskever, and G. Hinton. ImageNet classification with deep convolutional neural networks. In NIPS (2012).
[15] R. Uijlings, K. E. van de Sande, T. Gevers, and A. W. Smeulders. Selective search for object recognition. IJCV (2013).
[16] He, G. Gkioxari, P. Dollar, and R. Girshick. Mask R- ´ CNN. In ICCV (2017).
[17] He, X. Zhang, S. Ren, and J. Sun. Spatial pyramid pooling in deep convolutional networks for visual recognition. In ECCV (2014).
[18] S. Ren, K. He, R. Girshick, and J. Sun. Faster R-CNN: Towards real-time object detection with region proposal networks. In NIPS (2015).
[19] T.-Y. Lin, P. Dollar, R. Girshick, K. He, B. Hariharan, and ´ S. Belongie. Feature pyramid networks for object detection. In CVPR (2017).
[20] Shrivastava, A. Gupta, and R. Girshick. Training regionbased object detectors with online hard example mining. In CVPR (2016).
[21] Shrivastava, R. Sukthankar, J. Malik, and A. Gupta. Beyond skip connections: Top-down modulation for object detection (2016).
[22] R. Girshick, J. Donahue, T. Darrell, and J. Malik. Rich feature hierarchies for accurate object detection and semantic segmentation. In CVPR (2014).
[23] P. Sermanet, D. Eigen, X. Zhang, M. Mathieu, R. Fergus, and Y. LeCun. Overfeat: Integrated recognition, localization and detection using convolutional networks. In ICLR (2014).
[24] Liu, D. Anguelov, D. Erhan, C. Szegedy, and S. Reed. SSD: Single shot multibox detector. In ECCV (2016).
[25] J. Redmon and A. Farhadi. YOLO9000: Better, faster, stronger. In CVPR (2017).
[26] Awad, Dounia. Vers un système perceptuel de reconnaissance d'objets. Diss. Université de La Rochelle (2014).
[27] Yogesh N. Shinde1, Mrunmayee Patil."Translating Images into Text Descriptions and Speech Synthesis for Learning Purpose", International Journal for Research in Applied Science & Engineering Technology (IJRASET), Volume 4 Issue VI. (2016).
[28] ACHARJYA, Pinaki Pratim, DAS, Ritaban, et GHOSHAL, Dibyendu. Study and comparison of different edge detectors for image segmentation. Global Journal of Computer Science and Technology (2012).
[29] Wu, J., & Xiao, Z. Video surveillance object recognition based on shape and color features. (Vol. 1, pp. 451-454). In *Image and Signal Processing (CISP), 2010 3rd International Congress on* IEEE. (2010).
[30] Auguste, R., Aissaoui, A., Martinet, J., & Djeraba, C. Spatio-temporal histograms for the re-identification of people in television news. *Compression et Représentation des Signaux Audiovisuels (CORESA)*, article30. (2012).
[31] Hkiri, E., Mallat, S., & Zrigui, M. Constructing a Lexicon of Arabic-English Named Entity using SMT and Semantic Linked Data. *(2017).*
[32] Lhioui, C., Zouaghi, A., & Zrigui, M. A Rule-based Semantic Frame Annotation of Arabic Speech Turns for Automatic (2017).
[33] Farhani, N., Terbeh, N., & Zrigui, M. Image to Text Conversion: State of the Art and Extended Work. 2017 IEEE/ACS 14th International Conference on Computer Systems and Applications (AICCSA), 937-943. (2017).

Open Science Index, Computer and Information Engineering Vol:13, No:6, 2019 publications.waset.org/10010509/pdf