# ONTOLOGY BASED SEMANTIC IMAGE SEGMENTATION USING MIXTURE MODELS AND MULTIPLE CRFS

1. SHAIK.KARIMUNNISA, 2. N.SUBBARAO

1.M.Tech Student, Dept. of ECE, QUBA COLLEGE OF ENGINEERING& TECHNOLOGY, Venkatachalam, SPSR Nellore

2. Guide, Associate Professor, Dept. of ECE, QUBA COLLEGE OF ENGINEERING& TECHNOLOGY, Venkatachalam, SPSR Nellore

**Abstract:** Images / Videos are major source of content on the internet and the content is increasing rapidly due to the advancement in this area. Image analysis and retrieval is one of the active research field and researchers from the last decade have proposed many efficient approaches for the same. Semantic technologies like ontology offers promising approach to image retrieval as it tries to map the low level image features to high level ontology concepts. This paper proposes ontology-based semantic image segmentation (OBSIS) approach that jointly models image segmentation and object detection. In particular, a Dirichlet process mixture model transforms the low-level visual space into an intermediate semantic space, which drastically reduces the feature dimensionality. These features are then individually weighed and independently learned within the context, using multiple CRFs. The segmentation of images into object parts is hence reduced to a classification task, where object inference is passed to an ontology model. The main objective of this project is to detect objects in image by segmentation using ontology-based semantic image segmentation (OBSIS) approach.

**Keywords –** ontology, semantic image segmentation, web ontology language, image segmentation, object detection .

## I. INTRODUCTION

The word ontology refers to the science of metaphysics which defines the nature with its properties and relations. In Computer Science, ontology is a systematic arrangement of concepts, their properties and relations which exist in domain. Common components of ontology includes Individuals, Classes, Attributes, Relations, Function terms, Restrictions, Rules, and Axioms. Ontology can be domain-specific or generic; the former means ontology concepts are defined with reference to the specific domain whereas the later means the concepts are defined in general (i.e. the meaning / relationship of these concepts are already defined by English language). The implementation of ontology is generally a hierarchal representation defining concepts and their relationships. Three kind of relationships namely is-a, instance-of and part-of are generally used in the ontology. An ontology is defined as a formal and explicit representation of a shared conceptualization. It defines a set of representational primitives used to model a domain of interest and describes the semantics concepts and their interrelationships. Ontologies are based on a descriptive logic that axiomatizes the semantic classes and their correlations with relationships such as symmetry, transitivity, and equivalences as reliable and consensual knowledge.

Ontologies are specified in a language that allows rich structures by formalizing the information and knowledge about a domain of interest.

Most approaches use DL-based languages for describing knowledge domains and the relationships. The Web Ontology Language (OWL) is one of most popular languages to represent ontologies. It allows a domain to be specified in terms of concepts (or classes), rules (or properties), individuals and axioms with a uniform syntax and unambiguous semantics. Accordingly, the terms involved in the ontology include:

• classes, which represent types of objects, kinds of things, or concepts in the domain. They are usually organized in taxonomies;

• rules, which specify binary relations between classes and individuals. They include conditions and consequences that define the logical inferences derived from an assertion. Rules are denoted as a set of tuples, which are subsets of the cross products of the objects in the domain of discourse; • individuals, which refer to the actual instances in the domain; and

• axioms, which are used to model assertions. They include logical rules. In this work, a sublanguage of OWL namely OWL 2 DL is used as the ontology modeling language since it is more expressive and consists of more axioms than other sublanguages.

Ontology are usually develop to share common understanding of information among entities or softwares where each node in the ontology is a concept containing set of attributes and relationships. In the last decade, Ontologies have been widely used for knowledge representation and sharing. Ontology-based systems have been used in diverse areas such as software maintenance, Business Process Management, Biomedical Informatics, Knowledge Sharing, Knowledge Integration, Semantic Web, Fuzzy Systems, Supply chain management, Healthcare, Text Classification, Medical Domain, Robotics, Autonomic Computing, System Modelling, etc. The idea of using the ontologies in Image processing for content used retrieval is not new; in the last decade, researchers have proposed many efficient solutions using Ontologies for content based Image processing and retrieval. The existing approaches can broadly be categorized into three types namely
1) Color based techniques
2) Shape based technique and
3) Texture based technique.

## II. BACKGROUND WORK

A lot of research has been conducted on Image Retrieval (IR) on the basis of content similarity.
Many techniques have been used to enhance the results of image search. These approaches include hierarchical knowledge-based systems for Image Retrieval as researched by Kurtz, Camille, et al in 2014.

The semantic gap between the low-level image features and their high level semantics has always ruined the retrieval quality. So to cope up with this problem, Fernandez Miriam et al. used an ontology based approach for the enhancements of the image semantics. This research aimed to solve the restriction of the keyword based searching to support the semantic based Image Retrieval. The concept of semantic indexing has also been studied in the field of ontology based retrieval systems.

The literature review on Image Retrieval based on semantic concepts by Riad Alaa et al. had a great impact on the Image Retrieval field as it was very helpful for improving the semantic image retrieval systems accuracy. In this research various image search techniques are described for reduction of semantic gap. Furthermore, based on existing methods and application requirements author have suggested few future assessments. Another important survey was conducted by Liu Ying et al. in 2007 about the recent technical achievements on semantic based Image Retrieval; majority of the recent publications were included as the test data for the survey covering diverse amount of aspects in this area. Similar work has also been conducted on medical images by Xu J et al. , the authors focused on the key features of the image (e.g., shape, texture) in this research. The authors concluded that the performance of most CBIR systems is forced by these features because they cannot efficiently model the expectations of the user. All of existing studies helped in improving the results of content based images retrieval and lowering down the semantic gap between the user requirements and the search results.

Image segmentation is a challenging task in computer vision and has extensively been approached using the bottom-up paradigm based on low-level image features. Semantic image segmentation is even more challenging as it involves assigning semantic labels to the segmented regions. Primary studies on semantic image segmentation has dealt with this issue through pixel-wise labeling. For instance, Shotton et al. classified each pixel in an image according to joint appearance, shape, and context models. Although pixel-wise labeling is simple and straightforward, image pixels often contain limited information, which does not allow effective discrimination between classes. In contrast, region-level cues can be potentially more informative as well as more robust to noise, illumination, clutter, and variation.

This has led researchers to shift from the pixel-based paradigm to region-based. Region-based labeling normally begins with an initial unsupervised segmentation step as pre-processing.

Some prominent studies determine superpixels as basic units in the images. In a superpixel-based classification model was proposed using a unified energy function over image appearances and scene geometry structures.

Li et al. used multi-layered superpixels to capture diverse and multi-scale visual patterns of images based on bipartite graph partitioning.

These methods demonstrate that region-level segmentation is able to outperform their pixel-level counterparts. Despite the improvements of region-level segmentation methods, one pressing issue still exists where the segmented regions have little or no inherent semantic meaning. This makes effective labeling a seemingly prohibitive task. With the hope of obtaining more discriminative representations, the low-level bottom-up and global top-down features were adopted to be exploited and integrated into a segmentation model.

## III. PROPOSED METHODOLOGY

The main advantage of the proposed ontology inference is that it can handle different types of relationships at differing neighborhood and abstraction levels, which makes it applicable to other applications. The overall diagram of OBSIS is illustrated in Fig. 1 where semi-semantic segmentation and semantic labeling are incorporated to build semantic segmentation. In semi-

semantic segmentation, the visual features from segmented super pixels are exploited and individually learned in the new feature space using Dirichlet mixture models and CRFs. Therefore, object parts can be associated with the intermediate labels, which are then used in the semantic learning step. In this step, the obtained object parts along with their intermediate labels, image objects, and relationships at different levels are incorporated into the semantic ontology. Semantic segmentation is finally achieved by induced labels of the semantic ontology..
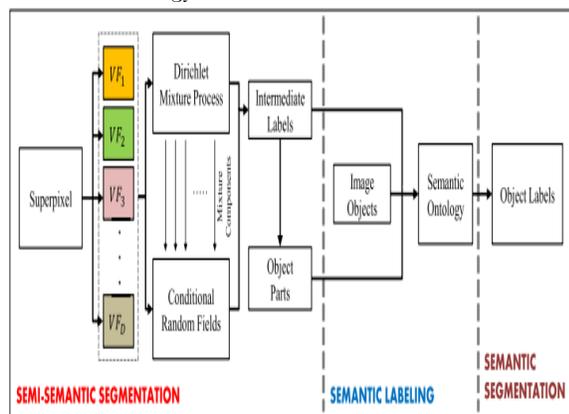


**Fig. 1** Proposed OBSIS

In probability theory, Dirichlet processes (after Peter Gustav Lejeune Dirichlet) are a family of stochastic processes whose realizations are probability distributions. In other words, a Dirichlet process is a probability distribution whose range is itself a set of probability distributions.

### 1. Dirichlet Process Mixture

The Dirichlet process is specified by a base distribution {\displaystyle H} H and a positive real number {\displaystyle \alpha } \alpha called the concentration parameter (also known as scaling parameter). The base distribution is the expected value of the process, i.e., the Dirichlet process draws distributions "around" the base distribution the way a normal distribution draws real numbers around its mean.

However, even if the base distribution is continuous, the distributions drawn from the Dirichlet process are almost surely discrete. The scaling parameter specifies how strong this discretization is: in the limit of 0, the realizations are all concentrated at a single value, while in the limit of {\displaystyle \alpha \rightarrow \infty } \alpha \rightarrow \infty the realizations become continuous.

Between the two extremes the realizations are discrete distributions with less and less concentration as increases.

The Dirichlet process can also be seen as the infinite-dimensional generalization of the Dirichlet distribution. In the same way as the Dirichlet distribution is the conjugate prior for the categorical distribution, the Dirichlet process is the conjugate prior for infinite, nonparametric discrete distributions. A particularly important application of Dirichlet processes is as a prior probability distribution in infinite mixture models. For example, we might be interested in how people will vote on a number of questions in an upcoming election. A reasonable model for this situation might be to classify each voter as a liberal, a conservative or a moderate and then model the event that a voter says "Yes" to any particular question as a Bernoulli random variable with probability dependent on which political cluster they belong to. By looking at how votes were cast in previous years on similar pieces of legislation one could fit a predictive model using a simple clustering algorithm such as k-means. That algorithm, however, requires knowing in advance the number of clusters that generated the data. In many situations it is not possible to determine this ahead of time, and even when we can reasonably assume a number of clusters we would still like to be able to check this assumption. For example, in the voting example above the division into liberal, conservative and moderate might not be finely tuned enough; attributes such as a religion, class or race could also be critical for modeling voter behavior.

### 2 Conditional Random Fields

CRF is simply a probabilistic framework with an arbitrary graphical structure for the labeling and classification of structured data. In CRFs, many statistically correlated attributes of the inputs are adopted, and their dependencies do not need to be explicitly declared due to the conditionality of the model. Moreover, in contrast to Hidden Markov Models (HMM), CRFs can model linear-sequence structures as well as arbitrary structures. Unlike sequential classifiers that are myopic about the impact of the current decisions on later decisions, CRFs can be trained discriminatively and are able to trade off decisions at different positions to achieve a globally optimum decision.

In OBSIS, CRFs are trained based on the samples of the visual features mixture components obtained from the Dirichlet process mixtures.

This is in order to learn the inter-mediate semantic space in the contextual space of the labels. Specifically, three CRFs corresponding to color, texture, and shape are independently trained to associate image object parts with intermediate labels.

These labels and their relationships along with concepts and their interactions are further utilized in an ontology for the final inference in the proposed OBSIS.

### 3 Semi-Semantic Image Segmentation

Although several segmentation algorithms are available. The problem of image segmentation is still a great challenge in computer vision. Despite the advances made in recent years, the perceptual grouping of image

regions is still unreliable. Image segmentation into perceptual regions is crucial since many applications rely on it such as object detection, object recognition, scene understanding, as well as semantic image segmentation. With the aim of robust semantic segmentation, this paper proposes a semi-semantic segmentation algorithm which efficiently group super pixels to generate semantic object parts. The generated segments are uniform in a sense where they are meaningful enough to be labeled in the inference stage. This method operates on over-segmented regions, such as that produced by SLIC super pixel segmentation, and tries to group these homogenous regions into a set of semantic object parts based on their low-level features. Particularly, the segmentation problem is treated as a classification task since extracted features from the super pixels are mapped to a set of low-level semantic labels. These labels correspond to color, texture and shape features in a higher level feature space. For each feature, the exploited attribute values from the training samples are grouped to an unknown number of classes using the aforementioned Dirichlet process mixture that does not require the number of classes to be known beforehand.

The proposed semi-semantic segmentation method is in a sense similar to traditional methods that assume image features are related to the statistical structure of the perceived environment, and consequently, they utilize clustering algorithms such as K-means, fuzzy c-means (FCM) or Gaussian mixture models (GMM) to exploit the representative vectors as centroids of the clusters. The extracted features of an image are then mapped onto the corresponding representative bins, and finally, pixels in the same cluster constitute a region. These methods determine the number of clusters beforehand, or empirically by testing several values. However, the consistency within and across different clusters cannot be guaranteed. Moreover, the number of representatives may vary due to the different parameter settings. Our intuition is that the aforementioned problems can be overcome in the proposed method where a new feature space is formed by automatically analyzing the data and using the Dirichlet process mixture models and CRFs.

Specifically, the intermediate labels that represent the visual features in the new feature space are learned through the training of the generated mixture components of the Dirichlet process mixture models using multiple CRFs.

Accordingly, the global context of the visual features can be incorporated in the feature learning because in the CRF training, all the samples with the same posterior distribution for a specific feature are investigated in their contexts.

In addition, the potential quantization error produced during the centroid computation is inherently avoided since the centroids of the mixture components are not explicitly computed.

We construct the Dirichlet process mixture model using a stick-breaking process, which represents it as a mixture model containing an infinite number of components with random mixing weights. The shortcomings of CRFs caused by the large graphical structures can be avoided since they are used in a higher order feature space with lower dimensionality. Note that context plays an important role for objects recognition in human vision. Therefore, it is considered not only in the feature level where visual features are specified in their contexts, but also in the concept level where ontology's are further utilized to characterize the concept relationships.

The training samples used to learn the intermediate color and texture-based labels are extracted from the irregular shaped super pixels while the semantic object parts are used for learning shape feature labels. This is because the shape descriptors mostly exploit the shape information from the edges and contours which are not accurate for the irregular shaped regions. If both color and texture labels of the neighboring regions are the same, they will be grouped into a larger region. The region labels are then identically assigned to the larger region while its exploited shape descriptor specifies the shape label. This region can be viewed as a semantic subpart of an object with some degrees of homogeneity, which for example, can be a car's body, tire or head-lamp. Otherwise, the shape feature is used to decide on whether regions are combined or separated. For this purpose, the shape descriptors are exploited for each region separately and also for their combination. If the likelihood of the most probable label $\hat{y}$ computed from the combination of regions is greater than that of individual regions, the regions are grouped into a larger region, otherwise, they remain separated.

At the end of the segmentation step, semantic object parts are labeled with their color, texture and shape labels, which will be then used for semantic labeling. Hence, by generating the object parts, image segmentation can be tailored for semantic segmentation in the inference stage.

## 4 Semantic Labeling

The importance of multi-level descriptions for structuring heterogeneous information is explored for semantic labeling in this work. The intuition is to generate an ontology to facilitate machine image visualization in the similar manner humans visualize images. Ontologies enable the recognition of the object labels by connecting the image space, the feature space, and the ontology space. The low-level visual features are exploited from images in the image space.

In this work, these visual features are transferred to a higher level intermediate semantic feature space. The

ontology space includes the ontology and a reasoning engine for the object labeling.

The intermediate labels in the feature space can boost the reasoning from the ontology's reasoning engine since they connect the high-dimensional and low-level features to usable terms in the ontology. For instance, the human recognition of a color as light blue is associated to a discretized numerical value in the feature space, which is seen as more efficient compared to existing approaches that use textual strings. It is noted that there are no appropriate names known to humans for all distinguishable values of all the features.

A three-layered ontology-based approach is proposed to allow for the most comprehensive semantic description of the image content. As shown by the semantic graph in Fig.3.3(a) the first layer consists of the ontology that captures the contextual relationships between image objects. Object decomposition into object subparts is modelled in the second layer. The third layer integrates the visual content into the ontology by describing the visual features of the object parts in terms of intermediate semantic labels. The semantic hierarchy indicates that the first layer extends the semantics of the underlying layers.

The overall goal is to explore the use of this ontology to previously unseen images to reason on the consistency of the labels of segmented objects. The problem of semantic labeling can be formulated as the reasoning in an ontology that contains the explicit knowledge about concepts, subsumption relationships, and contextual dependencies among image objects.

In particular, three kinds of relationships can be specified in the proposed three-layered ontological representation of the domain knowledge. The contextual relationships are exploited through connections between image objects in the first layer. The image objects are connected to the object parts in the second layer via subsumption relations such as 'is a part of '. In the third layer, the visual features are linked to the object parts to represent discrimination relatedness of each feature class for each object part. The intermediate labels in this layer are weighed from the training data. To model these relationships in the ontology, binary relations are usually utilized. However, binary predicates do not efficiently capture the useful contextual knowledge in the connections between image objects (in the first layer), and also, they cannot represent the discrimination weights of the intermediate labels (in the third layer). Therefore, these relationships are formulated using fuzzy description logics, which can signify a degree of relatedness to the connections. This is demonstrated in Fig.3.3(b). where fuzzy descriptions are used to characterize the co-occurrence of the two objects 'car' and 'road', and the degree of which each object part is represented by an intermediate label. Note that several

intermediate labels with different relatedness degrees can be used to represent a specific visual feature for an object part.

In order to deduce a relevant decision through the ontology, description logic (DL) is used. DL includes the required logics for representing structured knowledge.

In the semantic graph, the nodes are objects, object parts, feature class, or intermediate labels, which are modeled as OWL classes. The language OWL 2 DL assumes all classes are overlapped unless it is explicitly indicated that they are disjoint using a disjoint axiom. For each OWL class, a rule class is defined to include the sufficient conditions to characterize the relationships related to the corresponding class. It also presents the necessary conditions to bind the rule class to the OWL class.

Semantic role labeling, sometimes also called shallow semantic parsing, is a task in natural language processing consisting of the detection of the semantic arguments associated with the predicate or verb of a sentence and their classification into their specific roles. For example, given a sentence like "Mary sold the book to John", the task would be to recognize the verb "to sell" as representing the predicate, "Mary" as representing the seller (agent), "the book" as representing the goods (theme), and "John" as representing the recipient. This is an important step towards making sense of the meaning of a sentence. A semantic representation of this sort is at a higher-level of abstraction than a syntax tree. For instance, the sentence "The book was sold by Mary to John" has a different syntactic form, but the same semantic roles.
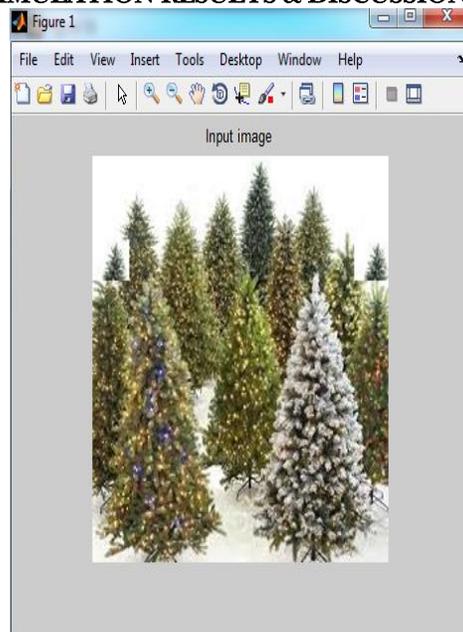
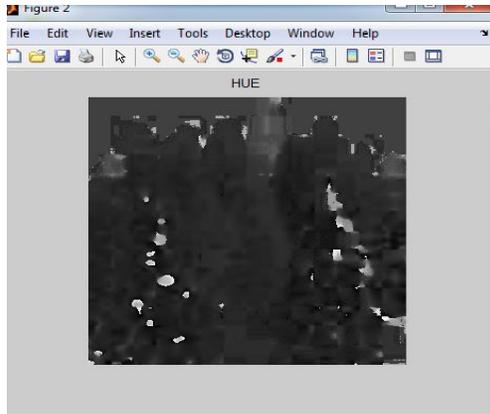## IV. SIMULATION RESULTS & DISCUSSIONS



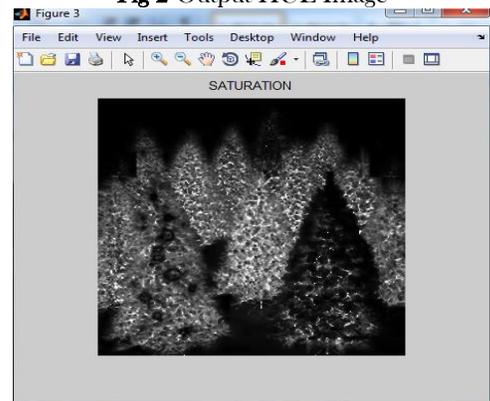**Fig 1** Input Image

**Fig 2** Output HUE Image



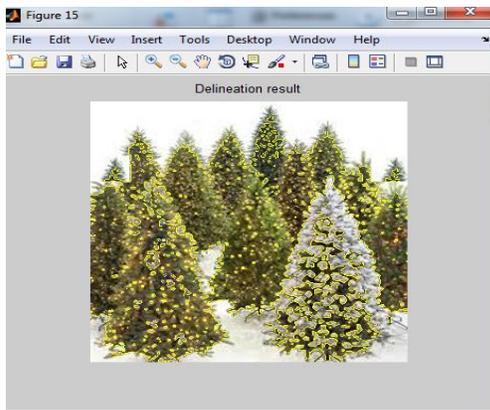**Fig 3** Output Saturation Image



**Fig 4** Output delineation Image

## V. CONCLUSION AND FUTURE WORK

This project proposes an ontology-based semantic image segmentation method (OBSIS) which efficiently employs different types of information at the proper levels. It bridges low- and high-level features by incorporating semantic knowledge in a gradual process from the very beginning. The low-level visual space is transformed into an intermediate semantic space of reduced dimensionality using the Dirichlet process mixture models and multiple CRFs. The visual features in this higher level space include intermediate labels that are used for the region labeling. A semantic ontology is constructed from the higher level features in the intermediate space, semantic concepts, and their relationships, where the final inference performed by this ontology model. Clustering the visual space using the Dirichlet process, and learning the cluster representations using CRFs offerconsiderable advantages over existing methods.

However, same with the other compared approaches, the proposed method fails to segment of touching objects with similar features. Another weakness is in the extracted relation-ships between features and concepts in the ontology, which are not effective enough due to the limited number of concepts in the benchmark databases. Finally, we observe that semantic segmentation must make trade-offs between the complexity of the model, complexity of the inference, and performance. This work forms a basis of our future work towards designing a more well-defined ontology for image retrieval applications. We also plan to use more comprehensive image representations such as prototype-based features, and conduct experiments with larger databases.

## REFERENCES

[1] Guang-Hai Liu, Lei Zhang, Ying-Kun Hou, Zuo-Yong Li, Jing-Yu Yang, Image retrieval based on multi-texton histogram, Pattern Recognition (2010), Volume 43, Pages 2380–2389.

[2] Alexander Toshev, Ben Taskar and Kostas Daniilidis, Shape-Based Object Detection via Boundary Structure Segmentation, International Journal of Computer Vision (2012), Volume 99, Number 2, Pages 123- 146.

[3] Umar Manzoor, Samia Nefti, Yacine Rezgui, ―Categorization of malicious behaviors using ontology-based cognitive agents‖, Data & Knowledge Engineering (2013), Volume 85, May 2013, Pages 40–56.

[4] Y. Liu,D. Zhang,G. Lu,W.-Y. Ma,A survey of content-based image retrieval with high-level semantics, Pattern Recognition (2007), Volume 40, Issue 11, Pages 262–282.

[s5] Umar Manzoor, Samia Nefti, ―iDetect: Content Based Monitoring of Complex Networks using Mobile Agents‖, Applied Soft Computing, Volume 12, Issue 5, May 2012, Pages 1607-1619.

[6] Shotton J, Winn J, Rother C, Criminisi A (2009) Texton boost for image understanding: Multi-class object recognition and segmentation by jointly modeling texture, layout, and context. International Journal of Computer Vision, 81(1), 2009.

[7] Shotton J, Blake A, Chipolla R, Contour-based learning for object detection. In Proceeding of International Conference on Computer Vision (2005).

[8] Awatef Al Azemi, Samia Nefti, Umar Manzoor, Yacine Rezgui ―Building a Bilingual Bio-Ontology

Platform for Knowledge Discovery‖, International Journal of Innovative Computing, Information and Control, Volume 7, Number 12, Dec 2011, Pages 7067-7075.

[9] T. Quack, U. Monich,L. Thiele, B.S. Manjunath, Cortina: a system for large-scale, content-based web image retrieval, in:Proceedings of the 12th annual ACM international conference on Multimedia, 2004.

[10] Naveed Ejaz, Umar Manzoor, Samia Nefti, Sung Wook Baik —A Collaborative Multi-Agent Framework for Abnormal Activity Detection in Crowded Areas‖, International Journal of Innovative Computing, Information and Control, Volume 8, Number 6, June 2012, Pages 4219- 4234.