# Approaches for Automated Object Recognition and Extraction from Images – a Study

S. Margret Anouncia and J. Godwin Joseph

School of Computing Sciences, VIT University, Vellore, Tamil Nadu, India

Digital Image Analysis is one of the most challenging and important tasks in many scientific and engineering applications. The two vital subtasks in image analysis are recognition and extraction of object(s) of interest (OOI) from an image. When these tasks are manually performed, it calls for human experts, making them more time consuming, more expensive and highly constrained. These negative factors led to the development of various computer systems performing an automatic recognition and extraction of visual information to bring consistency, efficiency and accuracy in image analysis. This paper focuses on the study of various existing automated approaches for recognition and extraction of OOI from an image in various scientific and engineering applications. In this study, a categorization is made based on the four principle factors (Input, Object, Feature, Attention) with which each approach is driven. All the approaches discussed in this paper are proved to work efficiently in real environment.

*Keywords:* image analysis, segmentation, feature extraction, computer vision, image retrieval, ontology

## 1. Introduction

Automatic image analysis is a vital process which is indispensable today in numerous applications of different domains. When such tasks are hand-engineered, it demands for human experts, making the system more constrained, difficult to maintain, less portable and economically weak. In response to the aforementioned problems, many researchers have focused their work to contribute solutions for these problems by developing methodologies which automate the entire image analysis process with little or no human interventions. Generally, image analysis incorporates two key tasks namely object recognition and extraction of the Object(s) of Interest (OOI) from a digital image. The vital tasks involved are segmenting the image into meaningful parts, extracting the desired features from the image and finally carrying out the extraction procedure. This work focuses on a tentative tassonomy of pre-determined systems in the domain of computer vision for extracting OOI, so that a conceptual framework for discussion, analysis and evolution of novel techniques in the domain can be achieved in the near future. All the works studied in this paper are selected based on the complexity involved, its advantage and the adaptable nature of the technique. This study will enable the researchers in this domain to have a quick review of these works and find which technique fits best to their application. This study also provides a way to bring novel ideas in this domain, by making an in-depth analysis of pitfalls in the surveyed works.

The paper is organized into six sections. Input-driven approaches are discussed in Section 2, Section 3 deals with the Feature-driven approaches. In Section 4, a study on Object-driven approaches is made. Section 5 deals with the Attention-driven approaches. Section 6 draws a summary of all studied approaches and finally Section 7 concludes this study and points out a new approach which may perform better in analyzing digital images.

## 2. Input-driven Approach

The methodologies categorized under this approach heavily depend on the images stored in memory and the image under analysis. The

study made states that when this kind of approach is adopted, it should be noted that preprocessing activities such as enhancing the image under analysis become mandatory. The images used for referencing must also be of the desired quality. The image compression tasks must also be considered to a certain extent so that an efficient storage structure is established for storing images. In this section, three methodologies under this approach were studied. The first relies heavily on the camera inputs, followed by a method based on affinely invariant regions to bring in automation in the extraction process, finally a method which completely depends on the images stored in the database, and which is highly domain specific.

## 2.1. Object Extraction Using Multiple Camera Inputs

A novel algorithm was developed by Jiro Katto and Mutsumi Ohta wherein they have extracted the OOI using multiple camera inputs [2]. The framework developed applies segmentation techniques to multiple camera inputs. The entire methodology is clearly depicted in Figure 1.

*Methodology*

The methodology developed by Jiro Katto et al uses four cameras where each of the two pairs of cameras shares optical axes. One pair of cameras focuses on the front objects of the image, while the other pair focuses on the rear objects of the image, following which a fusion (by K-means clustering) is done between the pair of images obtained out of different focuses. The

resulting pair of images is compared for detecting disparity, deriving a one-dimensional disparity vector. This disparity vector reduces prediction errors and contributes to the compression function which derives efficient disparity maps. This prevents encoding of both obtained stereo pairs, so only one image is appended with the disparity vector. Image fusion and disparity detection make the segmentation task efficient. From the focus map and the disparity map, an extraction map is derived. When this extraction map is masked with one of the fused images, an initial object extraction is done, but it may include a loss of certain correct clusters. A parallel region synthesis is carried out by gathering clusters which have similar statistical properties. The number of masked pixels is then counted in each region, determined by the region synthesis. Finally, regions containing more masked pixels than that of the predetermined ratio are selected and integrated to retrieve the OOI.

## 2.2. Object Extraction Using Invarient Regions

The system under study is developed by T. Tuytelaars, A. Zaatri, Luc Van Gool and H. Van Brussel in which a generic algorithm based on local, affinely invariant regions is introduced for bringing in automation in object recognition [3]. The entire methodology is clearly depicted in Figure 2.

*Methodology*

The methodology developed by T. Tuytelaars et al, captures the image under study in such a way that the distance between the camera and
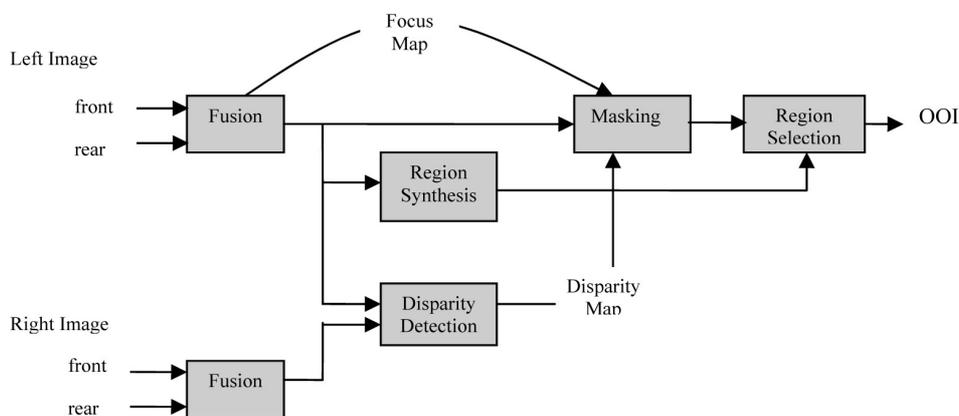


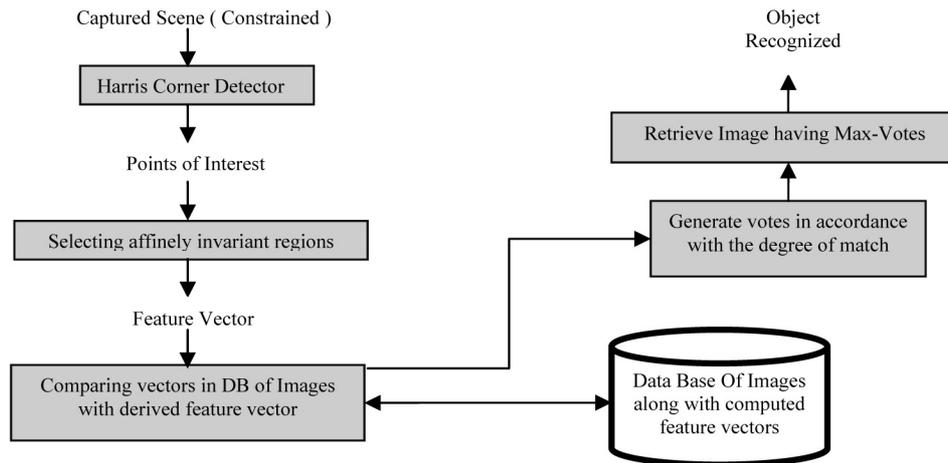*Figure 1.* Object extraction using multiple camera inputs.

*Figure 2.* Automated object recognition system.

the scene is sufficiently large and the scene is locally planar. The methodology is appearance-based an image object is selected from a database of images to derive a relevant object that matches the captured image. This selection process purely depends on affine moment invariants computed over the affinely invariant regions. Initially some points are selected within the captured image which delivers high degree of information. These are corners detected with Harris corner detector. Around each of the chosen points, affinely invariant regions are selected, and a feature vector for affine moment invariants (i.e. grey value invariants on local neighborhood) is computed. Finally, these vectors are used to retrieve similar regions from the database of images using Mahalanobis-distance (i.e. distances are measured in units of standard deviation). The most similar image is then se-

lected using a voting mechanism. So the image in the database with the highest number of votes is automatically recognized as the desired object.

## 2.3. Flag Identification Using Object Extraction Technique

Researchers Eduardo Hart, Sung-Hyuk Cha, and Charles Tappert proposed an interactive image retrieval technique for identifying flags [4]. Figure 3 depicts the entire methodology.

*Methodology*

The proposal made by Eduardo Hart et al is an interactive object extraction approach. Users interact with the system twice, firstly, in cropping the Region of Interest (ROI) from the
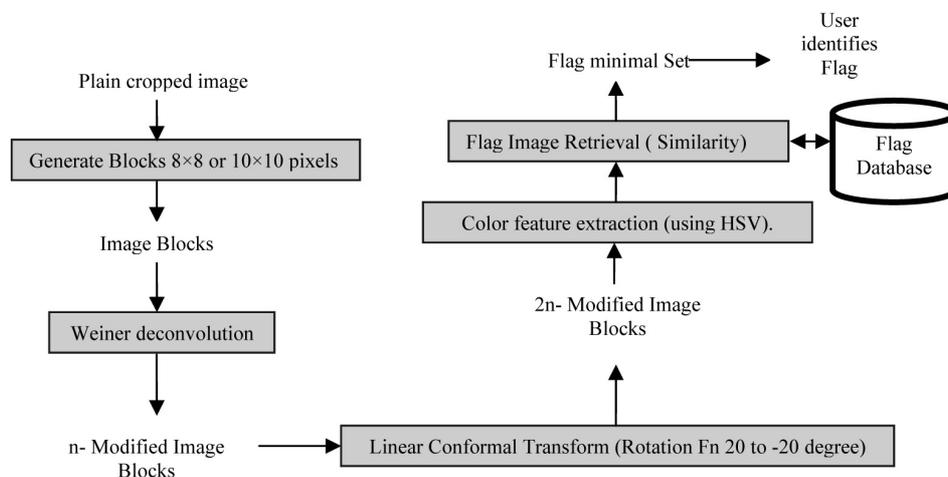


*Figure 3.* Flag identification using image retrieval technique.

photo (i.e. Flag part), secondly, to choose the desired flag from a minimal choice set. The technique adopted is a content-based image retrieval scheme where initially the cropping of the flag is made and the image is divided into $8 \times 8$ or $10 \times 10$ pixels block. The color and texture feature extraction process is carried out for each block. The target features are the color and their corresponding percentage contribution to the segmented portion of the flag. The features extracted are compared with a flag database which is built in the following way. Apart from the original image in the database, each flag image is divided into five sections where the first four contain $1/4^{th}$ of the cropped flag image and the fifth one is brought in by considering 30% of the central area of the total image. So there are totally 6 images on which Wiener Deconvolution technique is applied so that the effects due to the camera movements and 'out of scope' problems are eliminated. Finally, for structuring the database, with the set of 12 samples, 24 more samples are created by using the Linear Conformal Transform over those samples. Now the color feature extraction is done by clustering the colors into 9 colors using the HSV color value. The final task to be carried out in this methodology retrieves based upon the country and synthesis. This methodology showed that its performance lies between 82% and 93%.

## 3. Feature-driven Approach

Following the input driven approaches, feature-driven approach focuses entirely on high dimensionality of the segmented region for object recognition. Suppose an image segmentation technique divides an image into n-regions, each region can be determined with various attributes and properties. These features are distinctly categorized into two types namely, geometric features (i.e. angle, shape etc.) and statistical features (i.e. distance, mean etc.) [10, 8], so that when a best-input subset from a given image space is selected, the recognition and extraction of the objects can be done with more accuracy and in a cost effective manner compared to that of the Input-driven approach. In this section, four methodologies are surveyed. The first methodology analyzed under this category uses Artificial Intelligence for classifying outdoor images. The second technique is a research

work, where an edge-based segmentation technique is proposed for object recognition. The third methodology makes the entire feature extraction process in an automated way by using a novel framework and the final methodology discussed under this section is a robust real time object detection technique.

### 3.1. Artificial Intelligence Technique for Outdoor Image Classification

Researchers Neill W. Campbell, William P.J. Mackeown, Barry T. Thomas and Tom Troscianko developed a methodology for classification of outdoor images in an automated fashion [5].

*Methodology*

The methodology proposed by Neill W. Campbell et al uses a Bristol image database containing a large set of high quality color images. The image under consideration is segmented using K-means clustering of the grey level histogram. The value K is taken to be 4 so that there remains an optimal balance between under and over segmentation. From these segmented closed regions, 28 features are extracted, among which some of the important features are average color, position, size, rotation, texture (using Gabor filter), and shape (using principle components). The feature measure, in the context of the segmented region, is also derived. Finally, a neural approach is brought in for the classification purpose, where the system is trained with 7000 sample regions and tested for 3000 test regions. Using the conjugate gradient descent optimization, the neural net is found to have 24 nodes in the hidden layer so the encoding will be 1-in-n with 11 possible output labels. Misclassifications do occur in this type of approach (e.g. road and pavement) due to under or over segmentation. This approach classifies 80% of the regions correctly, corresponding to 91.1% of the image area. Classification accuracy by multilayer perception classifier is 91.1% and by learning vector quantization 76.7%.

### 3.2. Scale Map-based Image Segmentation for Object Recognition

A research work done by S. Kondra focused on usage of low level visual features, for edge detection, segmentation and recognition of the object [6].

*Methodology*

In the research made by S. Kondra, a scale map is introduced which is very similar to that of Lindeberg Scale Formula, but uses 8 directions and takes the sum of all orientations in order to determine the scale for each point in the image. The work also introduced rotational scale map where, instead of using differential filters, rotational filters are employed. In the resulting image, black indicates a lower scale while white indicates a higher scale. These two scales play a vital role in the segmentation. Initially, the image is segmented using watershed segmentation. A novel LINK method is introduced in which more than two segments are merged in one step. For comparing the segments, histogram distribution of the scale map, rotational map and CIELAB color are used. Each segment is compared with the neighboring segments and in the next step they are aggregated. Some small segments remain which cannot be merged based only on texture, so for these segments the difference of CIELAB distribution is considered only at the boundary of the segment, if the boundary is not salient then the technique proceeds with merging the segments. This technique has a human global consistency error measure of 0.184 which is comparable to the human GCE which is 0.08.

## 3.3. Adaptive Object Recognition Framework for Image Analysis

A research made by Ilya Levner, Vandim Bulitko, Lihong Li, Grreg Lee and Russel Greiner for automating the image feature extraction process by developing a novel Multi-Resolution Adaptive Object Recognition framework (MR ADORE) [7].

*Methodology*

MR ADORE framework developed by Ilya Levner et al uses a readily available off-the-shelf Image Processing Library (IPL). The process involves a significant amount of complexity as the libraries are domain independent. It calls for an intelligent control policy for applying various library operators, since a series of operators are required to make the entire image interpretation process more efficient and reliable. The decided IPL operators are termed as *MDP actions* and after applying these operators the results

are termed as *MDP states.* In order to recognize the object, they employ off-line and on-line machine learning techniques. Each training datum consists of two images, the input image and its user annotated counterpart allowing the output of the system to be compared to the desired image label. In the off-line training scheme, the initial image is applied with several sequences of IPL operators up to a certain user-controlled length, the resulting in MDP states. This is followed by computing a Q-function for each of the sampled states defined as $Q(s,a)$ where 's' is the state and 'a' is the actions. This Q-value is evaluated against a scoring metric rewarding scheme and a desired label for the image is learned by the system. When features are also considered, it makes supervised machine learning methods practically feasible. In this case the $Q-$function can be modified as $Q(f(s),a)$ where 'f' is the feature based on which the actions from IPL are adopted. The only purpose of employing this off-line learning phase is to construct an on-line control policy. This prevents the decisions based on frequently incomplete information provided by imperfect features, which in turn leads to having a significant increase in the interpretation quality. As the control policy calls for automated extraction of high quality features, a K-nearest neighbor (KNN) algorithm (basically a case-based approach) which approximates the Q-value of the input token based on distance(s) to the nearest training example(s) is adopted. This work implements and tests five policies of the aforementioned techniques and it has found that 1-NN and ANN (Artificial Neural Network) together with hand crafted feature, provide the best interpretation in which the accuracy is about 83%. The only pitfall is that huge training set is needed for enhancing the performance of 1-NN algorithm. A higher degree K-NN algorithm can be adopted to overcome this pitfall.

## 3.4. Rapid & Robust Object Detection Framework for Face Recognition

This technique was developed by Paul Viola and Michael Jones for robust real time object detection [12].

*Methodology*

The technique proposed by Viola-Jones aimes at developing an object detection framework that is capable of processing image rapidly and

achieving high detection rates. Three key contributions enable this technique to achieve the desired framework. Firstly, a new image representation called integral image, where in pixel value at (x,y) is the sum of pixel values of the original image above and to the left of (x,y). So this integral image can be computed by one pass through the image. This allows quick computation of features which are to be used by the detector. Second contribution is the AdaBoost learning algorithm which selects a small number of critical visual features by yielding an efficient classifier. The classifier is learned from labeled data, about 4916 hand labeled faces are used and all these faces are normalized. In addition to it, the training data also considers the variation factors like illumination, pose etc. The final contribution is a method of combining classifiers in a cascaded way so that background regions can be quickly discarded while spending more time on object like regions. This technique, when implemented on a conversional desktop, executes face detection at speed of 15 frames per second.

## 4. Object-driven Approach

Following the feature-based techniques, methodologies adopting, a kind of approach where in analyzing digital images, heavily depends upon the attributes contained in the OOI. Adopting this approach for addressing the image interpretation problem is widespread due to its intrinsic flexibility and the effectiveness of the result. Four methodologies are surveyed under this category. First methodology studied under this category is used for classifying CAD models. The second comes from an ontological approach for object recognition, followed by a novel technique for an enhanced segmentation-based object recognition and, finally, a method for detecting objects based on foreground image extraction is surveyed.

### 4.1. A Supervised Learning Technique for Classification of CAD Models

The methodology developed by researchers Cheuk Yiu Ip and William C. Regli focused on Content-based classification of CAD (Compute Aided Design) models [9].

*Methodology*

The method proposed by Cheuk Yiu Ip et al is a posteriory technique wherein an algorithm runs on a model and it finally produces a category or label for it. In comparing 3D CAD models, basically two basic techniques exist. They are the feature-based and the shape-based techniques. In the first technique, a graph based data structure is used for process plan generation, while integrating the ideas in the data base with the generated plan indexing and clustering of CAD models is achieved. The second technique is based on the computer geometry, computer vision and computer graphics.

As most of the engineering models have multiple classifications based on the manufacturing cost and the manufacturing process, a machine learning mechanism is applied for the classification purpose. Initially, a selection of invariant features is made from the models to form a set of n comparable vectors represented as $<a_1, a_2, \ldots, a_n>$. For certain samples say $S_1$, $S_2$, $S_3$, if we need to check whether these samples belong to the same category say C and if the a distance between two samples is given by $D(S_1, S_2)$ which is derived from the feature vectors, and also if s1 and s2 belongs to same category, then $D(S_1, S_2)$ is always lesser than $D(S_1, S_3)$. A technical approach is adopted wherein for each CAD model a shape distribution histogram is derived. Matching two shape distribution models calls for the following steps. Initially, a shape function is selected followed by the sampling of random points after which a point-pair distance is calculated from which a shape distribution histogram is derived. While comparing the shape distributions, Murkowski distance is employed. The classification process is done by using the nearest neighborhood technique. Experiments were conducted to test the classification correctness. The proposed method has 76% of highest classified correctness and 63% of average classified correctness. The correctness rates are not very attractive. Still, it can be considered, since there is no automated system for classifying 3D model, the correctness rate can be significantly increased by improving the quality of the training sets.

## 4.2. Recognition Model for Semantic Image Indexing

The methodology developed by Nicolas Maillot and Monique Thonnat is an object recognition model for semantic image indexing [11] where the main focus lies in finding out the semantic class of a physical object observed in the image using the concept of ontology.

*Methodology*

The research work done by Nicolas Mailloit et al focused on finding out the semantic class of a physical object observed in the image using the concept of ontology. Ontology, in general, is a set of concepts and relations used to describe a domain. The proposed methodology for performing image indexing, calls for two phases, a knowledge formulization phase which is done by hierarchy of object classes described by visual concepts and a machine learning phase in which the numerical features are mapped to semantic concepts. The learning is made by adopting a multi-layer neural network in which both positive and negative samples of the visual concept are considered, followed by feature extraction for each sample is which the required features are selected (color, shape and textural) for training the machine which enables the task of visual concept detection. When an input image is given for analysis, an automatic segmentation is made, followed by the feature extraction process so that all the classes from the root to the leaf in the ontological structure are scanned to get a confidence level, for the given image. After getting the confidence level, a semantic query is generated implicitly and the image is retrieved in accordance with the query. Experimental results determine that the precision of the proposed approach has 73% precision value and 19% recall value. Today the concept of ontology is implemented to address various problems in image interpretation. Medical image analysis, defect detections and real-time objects monitoring systems are some core areas wherein many researchers focus their work adopting this ontological approach for image analysis.

## 4.3. Reinforced Segmentation Mechanism for Object Recognition

The two approaches surveyed earlier have their performance determined by a rich training set and expert knowledge. In those domains where there is a lack of a priori knowledge and less samples for training, a new learning mechanism called Reinforced Learning (RL) can be adopted. An attempt was made by Farhang Sahba, Hamid R. Tizhoosh, Magdy M.A. Salama to increase the object recognition rate using the reinforced segmentation mechanism [13].

*Methodology*

The methodology proposed by Farhang Sahba et al focuses on making the Object Recognition (OR) system more adaptable and reacts cleverly to the changes in the environment. The concept of reinforced learning is mainly based on rewards or punishments to be assigned to a behavior under a specific context, which enables the system to perform better. The method adopted for RL is the Q-learning method proposed by Watkins[1]. In this technique, the construction of RL relies heavily on 3 defined components called state, action and reward. A Q-matrix is built depending upon the definitions of states and actions, followed by a matching criterion wherein, if there exist more than one object which has its matching features, the object with the closest match with respect to that of a threshold is considered. For defining states, they have considered an object descriptor (a shape descriptor representing the pixel arrangement), area (area of the extracted object) and finally the number of objects (considered with respect to the threshold value). In order to define the actions, the agent must adjust the threshold value and the size of the structuring element for morphological operations. The assignment of reward or punishment is based on a similarity measure. The system is trained using some reference objects, where a Q-matrix is developed using a Boltzman policy. Agents must find appropriate threshold and post processing parameters so that a reference object can be verified. Incorporating more states and more data sets enables this approach to be compared with the well established recognition techniques.

## 4.4. Object Extraction Technique for Foreground Images

Final methodology discussed under this category is a work done by researchers Yea-Shuan Huang and Fang-Hsuan Cheng, resulting in a

technique for foreground image extraction and object detection [14]. The proposed methodology is clearly depicted in Figure 4.

*Methodology*

In the methodology proposed by Yea-Shuan Huang et al, objects projected over the foreground of the image are considered essential for the detection and extraction purpose. Initially a Frame-based Distinct Pixel Extraction (FDPE) is performed, which completely extracts pixels having quite different features from their background, after which an object detection task is initiated that checks for instances of the needed object. Finally, a Region-based Foreground Object extraction (RFOE) is made wherein the focus is on extracting the foreground pixels of the Object Of Interest (OOI). As this method basically focuses on foreground pixel extraction, the background modeling (as the foreground pixels are both object and application dependant) plays a vital role. The object-oriented approach employed here combines both the conventional background subtraction and object detection techniques. In this method $I(x,y)$ represents a pixel in the image, $B(x,y)$ is a pixel in the background and $D(x,y)$ is the distinctness index, $D(x,y) = 1$ when $Prob(I(x,y),B(x,y)) < T$ ($T$ – predefined probability threshold value) and $D(x,y) = 0$, otherwise. So, when the distinctness index is nearing 1, it means that it's a foreground pixel. Following this task, the region scan is initiated where the shape of the region is considered, and those regions having enough number of distinct pixels are used for object detection. An AdaBoosting algorithm is adopted, which uses a re-weighting rule to construct a strong classifier. Finally for extracting the objects, a criteria denoted by $L(x,y)$ which is 1 when $Prob(I(x,y),B(x,y)) < T(x,y)$ where $T(x,y)$ is the threshold of pixel $(x,y)$ which is the function of position x and y. This determines whether a pixel is in the foreground or in the background. When the $L(x,y)$ and $D(x,y)$ are determined, an integration process defined as $L(x,y)$ *OR* $D(x,y)$ is evaluated and if this *OR* operation results in 1, it is a foreground pixel. Background updating is also done in order to have more satisfactory results. An experimental result of the proposed approach clearly states that it performs a speedy and accurate extraction tasks.

## 5. Attention-driven Approach

Visual attention plays a vital role in analyzing images, by focusing on objects that attract the attention of human observers. The general open problem is the gap between the low-feature content and the semantic image interpretation. Most working models based on the attention-driven technique are considered potentially well suited for addressing content-based image retrieval (CBIR) problems. Most of the early CBIR systems rely completely on region-based extraction while attention-driven approach identifies those regions which are more informative and feature rich, in order to perform an effective image retrieval task.
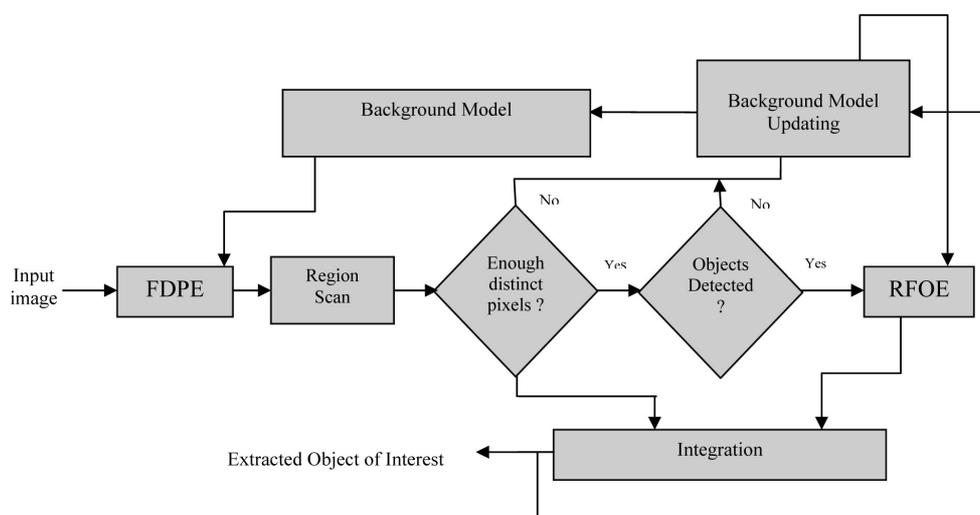


*Figure 4.* Object-oriented foreground image extraction.

## 5.1. Attention-driven Model for Content Base Image Retrieval

Researchers Ying-Hua Lu, Xiao-Hua Zhang, Jun Kong and Xue-Feng Wang proposed an attention-driven model for addressing the CBIR problem [15]. The entire methodology is clearly depicted in Figure 5.

*Methodology*

The Methodology proposed by Ying-Hua Lu et al carries out the segmentation task by a novel clustering algorithm called Expectation Maximum (EM) which automatically segments the image into several homogeneous regions, based only on the objects. This EM algorithm incorporates two key activities. Initially, feature extraction is done depending upon a three dimensional color descriptor.

Following the segmentation, a modified Itti-Koch model of visual attention is employed to play a role in attention selection, from a bottom-up perspective. Salient points are extracted by this model. Initially, the feature extraction process is carried out using a wavelet which could decrease the interference of the background information. Wavelets are used to create 9 different spatial scales and each level is decomposed into red, green, blue, yellow, intensity and local orientation. If the red, green and blue channels are normalized by image intensity, a local orientation information is obtained by applying wavelets to the intensity. From these channels, center-surrounded feature map is constructed. In their work, 36 feature maps are computed: 6 for intensity, 12 for color and 18 for orientation. Over the obtained feature maps, a combination scheme is made to normalize each feature map to a fixed dynamic range. Next, the summation of these maps is done for deriving the desired saliency map.

Following the saliency map construction, the technique concentrates on extraction of OOI, using the segmented image and the saliency map derived. The salient points produced by Itti-Koch algorithm are used to initiate a controlled region growth to the boundaries established by EM (Expectation Maximum) algorithm. Finally, the resulting image contains the extracted object of interest. This model results in an image including the prominent object in the scene. So from the extracted objects, feature vectors from features like Color (HSV) and Texture (mean, standard deviation, smoothness, third moment, uniformity and energy) are extracted. Model proposed by Ying-Hua Lu et al completely relies on the result of the segmentation. Objects cannot be extracted form poorly segmented regions.

A similarity measure is carried out by matching regions, using a novel distance method called Region-based quadratic distance to extract the desired OOI. This R-quadratic distance measure is achieved by calculating the distance between any two regions. A distance measure used in this method is the Euclidean distance. Experiments were conducted on the 5000 images from the COREL database. It shows that the proposed model ranks best with an overall retrieval accuracy. To qualitatively evaluate the retrieval effectiveness of the proposed model, over the entire image database 75 images of 5 categories are randomly selected as query images. A comparison with 7 peer retrieval methods is done whose result states the proposed model to have high overall effectiveness.
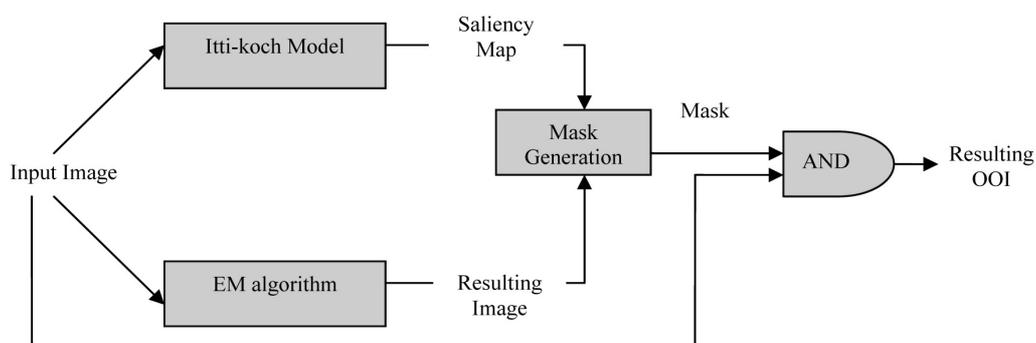


*Figure 5.* Attention-driven model for CBIR.

The significance of the proposed attention-driven model is its effectiveness in conquering some of the main limitations prevailing in several region-based image retrieval models. As it employs the result of a biologically-inspired bottom-up model to detect most salient peaks in an image, these peaks are often used to extract attentive objects, which in many real time cases corresponds to semantically meaningful objects.

## 6. Summary

Table 1 summarizes all the discussed methodologies. There are many methodologies available today to address the image analysis task, but those methodologies surveyed in this paper showed attractive figures for rate of accuracy in recognition and extraction of OOI. A study over the advantages and limitations of every method is made. In this summary, the domains where each method can be applied are also analyzed.

| APPROACH | METHODOLOGY | ADVANTAGE | LIMITATION | APPLICATION |
|---|---|---|---|---|
| **Input-driven Approach** | **1.** Uses Multiple camera inputs, fuses the inputs and then detects disparity for segmentation followed by region synthesis and masking, to extract the OOI. | Improved extraction performance. | Performance directly depends on the number of differently focused images. | 1. MPEG4 – Region Standardization. 2. Virtual-Studio – extract objects from uniform background. |
| | **2.** Harris Corner detector gives the points from which affinely invariant regions are selected. Feature vectors are derived for each region and a appropriate match is derived from the database by a voting mechanism. | Works efficiently even when there are large changes in the view point. | Sensitive to environmental changes (i.e. illumination, shadow etc.). | In supervisory control systems to recognize the objects. |
| | **3.** The Object of interest is identified in an interactive way, the cropped flag image is divided into image blocks for which Weiner deconvolution is applied, followed by a linear conformal transformation. Finally, the HSV color feature extraction is done to get a minimal choice set from which the user selects a flag. | Attractive flag identification results ($82\% - 93\%$). | Demands for more effective segmentation task and feature extraction process. | In flag identification systems. |
| **Feature-driven Approach** | **1.** A Bristol database is used for storing high quality images. K-means clustering is employed for segmentation, task followed by feature extraction which extracts about 28 features. Finally, a neural network is trained with 7000 sample regions in order to enhance the accuracy of the classification process. | Increased classification accuracy 91.1%. | More training data is required. | In monitoring systems which recognize the captured objects. |
| | **2.** Two scale maps namely scale map and rotational scale map are introduced. A novel LINK method is introduced which combines more than one region in one step by comparing the histogram distribution of the scale map, the rotational map and the CIELAB color. Finally, the aggregated region results in the extraction of the OOI. | Is easily adaptable to grey scale images using intensity. | On linking smaller segments illumination problems occur. | This novel approach can be deployed widely in extracting OOI from colored images and grey scale images. |

| APPROACH | METHODOLOGY | ADVANTAGE | LIMITATION | APPLICATION |
|---|---|---|---|---|
| | **3.** A novel MR ADORE framework which uses an IPL is employed to extract the OOI. Both off-line and on-line learning mechanisms are adopted and finally a KNN algorithm is used for recognizing the objects. | Completely automated feature extraction process. Accuracy is about 83%. | Large training set is required to enhance the performance. | Can be deployed widely in automatic recognition systems. |
| | **4.** A robust real time object detection technique enables rapid image processing with high detection rates. It uses a new integral image representation, an AdaBoost learning algorithm and a cascaded classifier for the desired face detection. | High speed detection, experimentally face detection proceeds at 15 frames per second. | Large constrained training data like under variations of illumination etc. | In digital cameras for face detection. |
| **Object-driven Approach** | **1.** It is a posteriory technique wherein a invariant feature vector for a model under consideration is derived and tested against the pre-determined feature set to find the appropriate match in order to decide a label for the model. | Proposed technique is fully automated, customizable. | 1. Classification rate heavily depends on the quality of training set. 2. Less adaptable to other domains. | In automated systems which perform part classification schemes for 3D CAD data. |
| | **2.** This technique is focused on finding out the semantic class of a physical object observed in the image using the concept of ontology. It is a purely object-oriented approach. | Highly adaptable in nature. | Depends on training data and ontology structuring. | In diverse domains where knowledge can be easily structured in terms of concepts and relations. |
| | **3.** In this methodology an effective reinforced learning mechanism is introduced by adopting Q-learning technique. A threshold technique for recognizing and extracting the objects from the image. | Highly adaptable as it relies less on expert knowledge. | Depends on large training data sets. | Being generic, a method which can be employed in systems which identify and analyze the captured objects. |
| | **4.** In this technique a FDPE is done followed by a region scan to decide whether the objects are detected from extracted pixels, if there are enough pixels. Otherwise, RFOE is made to recognize and extract the required OOI. | Speedy and more accurat in classification. | Demands for constrained environments. | In visual studios to extract objects from uniform background. |
| **Attention-driven Approach** | **1.** In this technique only those regions which are feature rich are considered. An EM algorithm results in clustered groups and the Itti-Kochi model derives a saliency map from which a mask is generated. When this mask is intersected with the image under study, the OOI is extracted. | Best overall retrieval accuracy (comparison made with 7 peer retrieval models). | Demands refinement in extracted OOI. | In CBIR systems such as medical image analysis, traffic monitoring systems etc. |

*Table 1.* Comparison of various methodologies.

## 7. Conclusion

The paper focuses on a survey of various approaches evolved in the last decade which facilitated the image analysis task, by bringing in automation in object recognition and extraction. Those approaches discussed in this survey are categorized based on the defined principles (Input, Feature, Object and Attention) with which each approach is driven. In spite of the fact that

most of the discussed approaches perform well, there is a possibility to bring in more objectivity in the interpretation task in those systems which have adopted these approaches. From the tentative tassonomy of pre-determined systems in a real environment, as most of the image interpretation systems are highly domain specific and rely much on expert knowledge, a memory-based problem solving principle may be adopted in the near future to enhance the analyzing performance of the system. In spite of relying completely on the expert knowledge, an adaptation can be incorporated within the system, which can make them derive better results. One of the novel methodologies under this principle is Case-based Reasoning (CBR) which can definitely facilitate the domain specific systems to perform with more accuracy, consistency and efficiency.

## References

[1] C. J. C. H. WATKINS, P. DAYAN, Q-Learning Machine Learning. Vol. 8, pp. 279–292, 1992.

[2] KATTO, J. OHTA, M., Novel Algorithm for Object Extraction using Multiple Camera Inputs. *Proc. of International Conference on Image Processing*, Vol. 2, pp. 863–866, 1996.

[3] TUYTELAARS, T. ZAATRI, A. VAN GOOL, L. VAN BRUSSEL, H., Automatic Object Recognition as a part of an Integrated Supervisory Control System. *Proc. of International Conference on Robotics & Automation*, Vol. 4, pp. 3707–3712, 2000.

[4] E. HART, S.-H. CHA, C. C. TAPPERT, Interactive Flag Identification using Image Retrieval Techniques. In: *CISST 2004*, pp. 441–445, 2004.

[5] N. W. CAMPBELL, W. P. J. MACKEOWN, B. T. THOMAS, T. TROSCIANKO, The Automatic Classification of Outdoor Images. *Proc. of International Conference on Engineering Applications of Neural Networks*, pp. 339–342, 1996.

[6] S. KONDRA, Image segmentation for object recognition. In: *Sixth Frame Work Programme, Marie Curie Actions*, Human Resource and Mobility Marie Curie Research Trainings Networks (RTN), 2004.

[7] I. LEVNER, V. BULITKO, L. LI, G. LEE, R. GREINER, Automated feature extraction for object recognition. In: *Image and Vision Computing'03 New Zealand*, Palmerston North, New Zealand, 2003.

[8] S. MANSOURI ALGHALANDIS, GH. NOZAD ALAMDARI, Welding Defect Pattern Recognition in Radiographic Images of Gas Pipelines using Adaptive Feature Extraction Method and Neural network Classifier. In: *23$^{rd}$ World Gas Conference, Amsterdam*, 2006.

[9] C. Y. IP, W. C. REGLI, Content-based Classification of CAD models with Supervise Learning. In: *Computer-aided Design and Applications*, Vol. 2, no. 5, pp. 609–617, 2005.

[10] ANOUNCIA S. M., SARAVANAN R., Non-destructive Testing using Radiographic Images – a Survey. *British Journal of Non-Destructive Testing, Insight*, Vol. 48, no. 10, pp. 592–596, 2006.

[11] N. MAILLOT, M. THONNAT, Object Recognition for Semantic Image Indexing. In: *Orion team INRIA Sophia Antipolis FRANCE*.

[12] P. VIOLA, M. JONES, Robust Real Time Object Detection. In: *Second International workshop on statistical and computational theories of vision – modeling, learning, computing, and sampling*, 2001.

[13] SAHBA, F. TIZHOOSH, H. R. SALAMA, M. M. A., Increasing Object Recognition Rate Using Reinforced Segmentation *Proc. of IEEE International Conference on Image Processing*, pp. 781–784, 2006.

[14] Y.-S. HUANG, F.-H. CHENG, Object-orientation Foreground Image Extraction. *Proc. of 2$^{nd}$ International Conference on Innovative Computing, Information and Control*, pp. 407, 2007.

[15] Y.-H. LU, X.-H. ZHANG, J. KONG, X.-F. WANG, A Novel Content-based Image Retrieval Approach Based on Attention-driven Model. *Proc. of International Conference on Wavelet Analysis and Pattern Recognition*, pp. 510–515, 2007.

*Contact address:*

S. Margret Anouncia
J. Godwin Joseph
School of Computing Sciences
VIT University, Vellore
Tamil Nadu, India
e-mail: anouncia@rediffmail.com

DR. S. MARGRET ANOUNCIA is a Professor at the School of Computing Science and Engineering at VIT University, Vellore, India. She received a Bachelor's degree in Computer Science and Engineering from Bharathidasan University in 1993 and a Masters in Software Engineering from Anna University in 2000. She has been awarded doctorate in Computer Science and Engineering at VIT University in 2008. She has teaching and research experience of about fifteen years. Her area of specialization includes Digital Image Processing, Software Engineering and Knowledge Engineering. She is a life member of the Computer Society of India and of the Indian Society for Technical Education. She has published five papers in international journals and presented around fifteen papers at various national and international conferences.

MR. J. GODWIN JOSEPH is currently a M.S. scholar at VIT University.