

## Object Based Video Analysis, Interpretation and Tracking

<sup>1</sup>A. Umamakeswari and <sup>1</sup>A. Rajaraman

<sup>1</sup>Department of Computer Science and Engineering, School of Computing,  
SASTRA University, Thanjavur, Tamil Nadu, South India-613402

<sup>2</sup>Indian Institute of Technology, Chennai, Tamil Nadu, South India-600 036

**Abstract:** The role of computers in different facets of human life is increasing everyday, from one of supplementing his needs to one of integrating in the different activities, he is involved. This has become more predominant with the prolific developments in communication and internet. This brings in a need to raise the level of computers to the level of human beings and a paradigm shift from hard computing to soft computing towards the turn of this century has reinforced this. The present focus of the study is on implementing visual capabilities in computers so that involvement and interaction with humans are easier. The paper presents the details of object based video analysis for conventional engineering applications.

**Key words:** Dynamic scenes, attributes, reconstruction, feature extraction, recognition, tracking

### INTRODUCTION TO VIDEO PROCESSING

Complex and dynamic scenes in video and television broadcasting have been a challenge for computerization of both in terms of storage and interaction. Scenes can be complex wherein there are many cluttered objects with different sizes, shapes, colors and can be dynamic with multiple interacting moving objects in a constantly changing background. Examples of such videos abound in applications like sports, air traffic, car traffic and cloud transformations etc<sup>[1]</sup>. Due to the advancement in technologies and also because of increasing demand for managing the vast amount of visual data in video, reliable and efficient systems that are capable of understanding and analyzing scenes need to be developed.

Video sequences provide more information about how objects and scenarios change over time, as compared to still images. They can provide more information than text, graphics and static images can do. This can relate to the position, distance, temporal and spatial relationships that are included in the video data implicitly. However, video needs more space for storage and wider bandwidth for transmission<sup>[2]</sup>. The advantage of using video instead of a single image for recognition is that, different pose of the object can be visible and also smoothness in pose transitions can be used to the greatest advantage for recognition<sup>[3]</sup>.

Thus, the ability to process digital video to automatically identify shots and scenes has led to a wide range of multimedia applications. Research on video processing is still at an early stage but is now attracting attention. Motion detection, object tracking and object recognition are very useful in many applications. Automatic understanding of objects from images has become a central focus of research and tracking, reconstruction and recognition from video and still remains as challenging problems to computer scientists<sup>[4]</sup>. The methods which are used to describe the contents of video are selected keywords which are time-based and attributes generated by looking at the content which is done almost manually. In the context of computer integration this is done by analyzing and understanding the video through techniques for object identification and recognition, using AI and Soft computing methods<sup>[5]</sup>.

### EXTRACTION/RETRIEVAL OF VISUAL INFORMATION FROM VIDEO IMAGES

Object extraction from images and video streams uses the ideas and inspirations from computer graphics and computer vision areas. Visual content can be modeled as a hierarchy of abstractions. At the first level are the raw pixels with color or brightness information. Further processing yields features such as edges,

**Corresponding Author:** A.Umamakeswari, Department of Computer Science and Engineering, School of Computing, Sastra University, Thanjavur, India, 613 402 Tel: +9194434 51898

corners, lines, curves, and color regions. A higher abstraction layer may combine and interpret these features as objects and their attributes. At the highest level are the human level concepts involving one or more objects and relationships among them. Most of the work is done using low level features which are not used by human to interpret video sequences (human does by means of semantics). The main challenge in the retrieval of images or video is bridging the gap between the high level query from the human and the low level features that can be easily measured and computed. This gap arises mainly due to the lack of understanding of the “meanings” of the video, “meaning” of a query and also by the way in which the result incorporates knowledge<sup>[6]</sup>.

The visual information retrieval in retrieving images or image sequences from a database is essentially based on query. This arose due to emerging multimedia applications, the availability of large image and video archives and also due to sharing and distributing image video data over large bandwidth computer networks. Early visual information retrieval systems used querying by strings or text and were found difficult to work with and later systems used shape or visual features such as the outline of an object. But in the new generation visual information retrieval systems, image processing and computer vision techniques are applied for automatic extraction of visual features like color, shape and texture from the image data which is similar to human perception.

MPEG-7, the recent member of the MPEG family, specifies visual descriptors for color, shape, texture, and motion and description schemes for specifying structure and semantic relationships among descriptors to provide efficient and fast content-based access and search<sup>[7]</sup>. Essential tools in the form of fully automatic algorithms for multimedia analysis and interpretation are needed to integrate MPEG-7 description schemes into multimedia applications. Nowadays, development of visual retrieval systems is gaining importance and at the same time becoming a very important constituent of the vision systems.

### **VIDEO FEATURE EXTRACTION PROCEDURES**

The two main procedures in video feature extraction are frame-based and object based. In frame-based approaches, to analyze the video sequences, low-level features such as color, histogram, texture and motion of frames (in uncompressed or compressed domains) are extracted and are used. The advantage of this approach is that they are relatively easy to compute

and the drawback is it is relatively difficult to use them to address the semantics of the video contents, especially when we are interested in the behavior of objects in the video sequences. All the features are in frame level and they contain all of this information. However, by extracting features in the frame level, the semantic information is implicit, and requires great effort to recover later. Successful results have been reported but they are generally very difficult to use for fine-grained semantic interpretation of video sequences.

In recent years, object-based approaches have been gaining popularity. The basic unit of analysis is low-level features from individual objects, instead of from frames. These are video objects (VOs), a concept introduced by MPEG-4 which are first segmented from the video sequences and then associated with semantic meaning. The advantage of this approach is that the semantic information can be explicitly expressed via the object features. Finer-grained semantic interpretation of video sequences can be obtained more easily by the features directly derived from objects and it gives us more flexibility in choosing the mapping models that are closer to human understanding of video sequences. The major drawback of this approach is that it will introduce higher computation complexity since generic VO segmentation is an extremely difficult task<sup>[8]</sup>.

### **VISUAL TRACKING OF VIDEO IMAGES**

Object tracking in video aims at detecting the appearances of an object and also the position of a moving object in a video sequence. It has got many applications and the major drawback of this approach is in the time needed for processing and the large size of the files that are used for processing. The first application of motion detection was done in video compression algorithms as they reduced the size of video data by storing motion vectors instead of the value of each pixel. The first step in object tracking algorithms is locating areas of motion. Estimating the motion manifested in a set of images or in an image sequence is a fundamental problem in both image/video processing and computer vision. The computer vision system is to interpret the world using visual information sensed by a video imaging system. The interpreted information in any scene provides the necessary information for estimating the motion.

Recognition and tracking are implemented by combining the analysis of single image frame with the analysis of consecutive image frames. Object tracking and recognition differ only in the set of objects used by them: recognition uses a predefined set of objects and

tracking identifies objects in a set of previously identified objects<sup>[5]</sup>. So object tracking represents the spatio-temporal relationship of objects in video sequences and object recognition is a direct derivative of object matching.

In object tracking, an object's spatial and temporal changes during a video sequence are monitored, including its presence, position, size, shape, etc. by solving the problem of matching the target region in successive frames of a sequence of images taken at closely spaced time intervals. The detection of objects in videos involves verifying the presence of an object in image sequences and possibly locating it precisely for recognition. These two processes are closely related because tracking usually starts with detecting objects, while detecting an object repeatedly in subsequent image sequence is often necessary to help and verify tracking.

Object tracking requires identifying the specified object in each frame of the video sequence. Thus it faithfully locates the targets in successive video frames by using the shape features in the current/previous frames and the object information is used in the detection of objects in the next frame. From the tracked features, the type of moving objects and their interaction are then analyzed. Thus the predictions on the motion at any instant can be made on the basis of their previous trajectories<sup>[9]</sup>.

The complexity of the problem also increases if multiple moving objects need to be tracked, as is the case in many applications including surveillance, sports reporting, video annotation, supervisory management of shopping outlets, production lines and traffic control.

### **METHOD OF TRACKING IN VIDEO IMAGES**

To successfully track moving objects in a simple video is a challenging problem. The reason is that the objects may change shape or may be subject to occlusions. Most of the tracking algorithms can be broadly classified into the following four categories<sup>[10]</sup>:

- Gradient-based methods locate target objects in the subsequent frame by minimizing a cost function
- Feature-based approaches use features extracted from image attributes such as intensity, color, edges and contours for tracking target objects
- Knowledge-based tracking algorithms use a priori knowledge of target objects such as shape, object skeleton, skin color models and silhouette

- Learning-based approaches use pattern recognition algorithms to learn the target objects in order to search them in an image sequence

The normal method is to identify the object on a single frame and then track its evolution across subsequent frames. Each frame is treated as an image. The essential information conveyed by the video is analyzed by the boundary of the object as it changes with time. The time variation provides cues about the identity of the object, activity performed by the object and also the nature of interaction between different objects in the same scene<sup>[3]</sup>. The algorithm is capable of tracking objects between two consecutive frames and given any two frames it is able to track the object in the video sequence also. Even one can attempt to use information from the past and future frames using which the object shape as well as the position may be predicted. Also the transformations undergone by the object are computed<sup>[11]</sup>. Sometimes tracking may get lost at some points due to noise, occlusions etc and hence methods of reconstruction need to be attempted to track and also identify the object in the manner similar to human beings.

### **ABORT, A PLATFORM FOR VIDEO IMAGES**

With the background provided for both object recognition and image tracking in video sequences a platform, ABORT-Attribute Based Object Recognition and Tracking-was developed with added features to take care of deficiencies in the video images. This System recognizes the objects based on attributes and also tracks them by plotting the trajectory. Many times the individual video images of a video clip may be unfocussed, incomplete or even imprecise<sup>[12]</sup>. Further a video clip may contain missing frames resulting in problems both in terms of identification and reconstruction. As the platform ABORT contains modules to take care of individual frames as well a sequence of frames, one will be able to reconstruct and regenerate using computer methods so that for engineering applications it will be easier.

### **RESULTS**

Figure 1a and b, the inaugural snapshots of ABORT system are shown. In Fig. 2, typical acquired images with deficiencies along with the reconstructed image by ABORT system are shown. Better applications of this

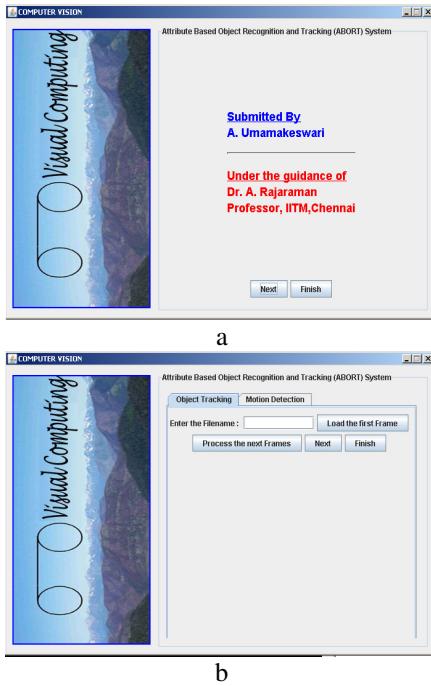


Fig. 1: a and b ABORT System

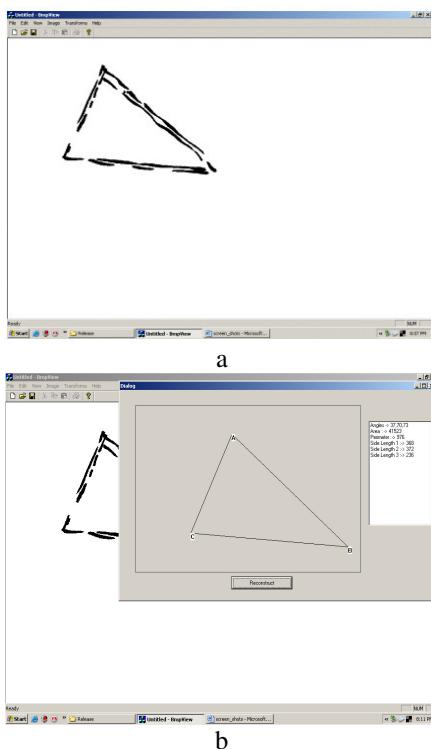


Fig. 2: a and b Acquired Image and its Reconstruction by ABORT system

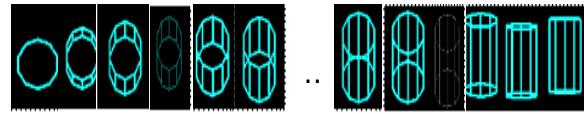


Fig. 3: Two frames are incomplete

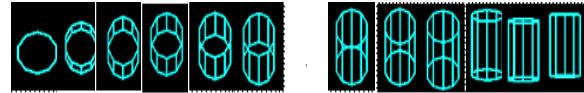


Fig. 4: All frames are complete (reconstructed above two incomplete frames)

platform ABORT can be seen in dynamic images and this is shown in Fig.3 and 4. Fig. 3 gives a video clip which is incomplete with some frames missing. Further some of the images in the clip have deficiencies. Fig.4 shows a complete reconstruction of the video clip with inclusion of missing frames and regeneration of deficient images.

## REFERENCES

- Kenji Okuma James J. Little David Lowe, Automatic Acquisition of Motion Trajectories: Tracking Hockey Players, Internet Imaging V San Jose, California, 2003.
- Yiwei Wang, John F. Doherty, Robert E. Van Dyck, Moving Object Tracking in Video, AIPR, 29th Applied Imagery Pattern Recognition Workshop (AIPR'00), 2000.
- Omar Javed, Mubarak Shah and Dorin Comaniciu, A Probabilistic Framework For Object Recognition In Video, International Conference on Image Processing, 2004.
- A. Hilton, P. Fua and R. Ronfard, Modeling people: Vision-based understanding of a person's shape, appearance, movement, and behaviour, Computer Vision and Image Understanding 104(2006): 87-89.
- Gaetan Noel, Identification of moving objects in color digital video sequences, Ph.D Thesis, University of Ottawa (Canada), 2002.
- Yan Liu and John R. Kender, Fast video segment retrieval by Sort-Merge feature selection, boundary refinement, and lazy evaluation, Computer Vision and Image Understanding 92 (2003) 147-175.
- Gozde Bozkurt, Curve and Polygon Evolution Techniques for Image Processing, Ph.D Thesis, North Carolina State University, 2002.

8. Ying Luo, Tzong-Der Wu and Jenq-Neng Hwang, Object-based analysis and interpretation of human motion in sports video sequences by dynamic bayesian networks, CVIU, Computer Vision and Image Understanding 92(2003): 196-216.
9. Xianghong, (Henry) Liu, Development of a vision-based object detection and recognition system for intelligent vehicle, The University of Wisconsin-Madison, 2000.
10. R.Venkatesh Babu, Patrick Perez, Patrick Bouthemy, Robust tracking with motion estimation and local Kernel-based color modeling, ICIP, 2005.
11. A. Umamakeswari, Dr. A. Rajaraman, 2004, Role of Object Attributes of Dynamic Images in Visual Computing for Engineering Applications, International Conference on Computational Intelligence, (ICCI 2004), Istanbul, Turkey, ISBN 975-98458-1-4, pp: 145-148.
11. A. Umamakeswari, Dr. A.Rajaraman, 2004, Visual Computing: A new Methodology for Engineering Applications, Sixth International Conference on Cognitive Systems, (ICCS 2004), CRCS, IIT, New Delhi.