

## Automatic Semantic Content Extraction in Videos Using a Fuzzy Ontology and Rule-Based Model

**Satish D Mali**

ME,

Department of Computer Science and Engineering,  
T.P.C.T.'s College of Engineering,  
Osmanabad.

**Pravin P Kalyankar**

Professor,

Department of Computer Science and Engineering,  
T.P.C.T.'s College of Engineering,  
Osmanabad.

### ABSTRACT

*Recent increase in the use of video-based applications has revealed the need for extracting the content in videos. Raw data and low-level features alone are not sufficient to fulfill the user's needs; that is, a deeper understanding of the content at the semantic level is required. Currently, manual techniques, which are inefficient, subjective and costly in time and limit the querying capabilities, are being used to bridge the gap between low-level representative features and high-level semantic content. Here, we propose a semantic content extraction system that allows the user to query and retrieve objects, events, and concepts that are extracted automatically. We introduce an ontology-based fuzzy video semantic content model that uses spatial/temporal relations in event and concept definitions. This metaontology definition provides a wide-domain applicable rule construction standard that allows the user to construct ontology for a given domain. In addition to domain ontologies, we use additional rule definitions (without using ontology) to lower spatial relation computation cost and to be able to define some complex situations more effectively. The proposed framework has been fully implemented and tested on three different domains. We have obtained satisfactory precision and recall rates for object, event and concept extraction.*

### INTRODUCTION

The rapid increase in the available amount of video data has caused an urgent need to develop intelligent methods to model and extract the video content. Typical applications in which modeling and extracting

video content are crucial include surveillance, video-on-demand systems, intrusion detection, border monitoring, sport events, criminal investigation systems, and many others. The ultimate goal is to enable users to retrieve some desired content from massive amounts of video data in an efficient and semantically meaningful manner.

There are basically three levels of video content which are raw video data, low-level features and semantic content. First, raw video data consist of elementary physical video units together with some general video attributes such as format, length, and frame rate. Second, low-level features are characterized by audio, text, and visual features such as texture, color distribution, shape, motion, etc. Third, semantic content contains high-level concepts such as objects and events. The first two levels on which content modeling and extraction approaches are based use automatically extracted data, which represent the low-level content of a video, but they hardly provide semantics which is much more appropriate for users. Users are mostly interested in querying and retrieving the video in terms of what the video contains. Therefore, raw video data and low-level features alone are not sufficient to fulfill the user's need; that is, a deeper understanding of the information at the semantic level is required in many video-based applications.

However, it is very difficult to extract semantic content directly from raw video data. This is because video is a temporal sequence of frames without a direct relation to its semantic content. Therefore, many different

representations using different sets of data such as audio, visual features, objects, events, time, motion, and spatial relations are partially or fully used to model and extract the semantic content. No matter which type of data set is used, the process of extracting semantic content is complex and requires domain knowledge or user interaction.

There are many research works in this area. Most of them use manual semantic content extraction methods. Manual extraction approaches are tedious, subjective, and time consuming, which limit querying capabilities. Besides, the studies that perform automatic or semiautomatic extraction do not provide a satisfying solution. Although there are several studies employing different methodologies such as object detection and tracking, multimodality and spatiotemporal derivatives, the most of these studies propose techniques for specific event type extraction or work for specific cases and assumptions. Simple periodic events are recognized where the success of event extraction is highly dependent on robustness of tracking. The event recognition methods described are based on a heuristic method that could not handle multiple-actor events. Event definitions are made through predefined object motions and their temporal behavior. The shortcoming of this study is its dependence on motion detection. Scenario events are modeled from shape and trajectory features using a hierarchical activity representation extended from. Hakeem and Shah propose a method to detect events in terms of a temporally related chain of directly measurable and highly correlated lowlevel actions (subevents) by using only temporal relations.

**Image:**

An image is a two-dimensional picture, which has a similar appearance to some subject usually a physical object or a person.

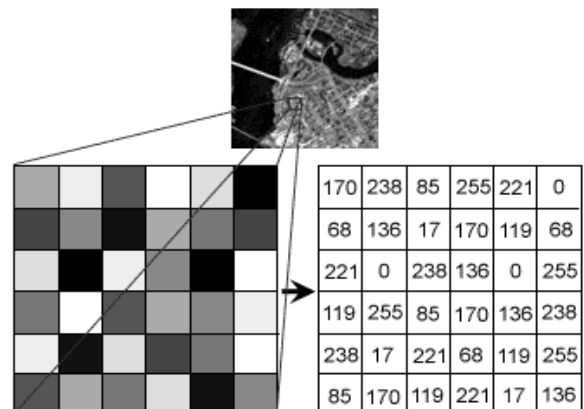
Image is a two-dimensional, such as a photograph, screen display, and as well as a three-dimensional, such as a statue. They may be captured by optical devices—such as cameras, mirrors, lenses, telescopes,

microscopes, etc. and natural objects and phenomena, such as the human eye or water surfaces.

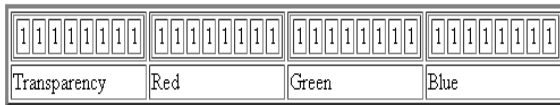
The word image is also used in the broader sense of any two-dimensional figure such as a map, a graph, a pie chart, or an abstract painting. In this wider sense, images can also be rendered manually, such as by drawing, painting, carving, rendered automatically by printing or computer graphics technology, or developed by a combination of methods, especially in a pseudo-photograph.



An image is a rectangular grid of pixels. It has a definite height and a definite width counted in pixels. Each pixel is square and has a fixed size on a given display. However different computer monitors may use different sized pixels. The pixels that constitute an image are ordered as a grid (columns and rows); each pixel consists of numbers representing magnitudes of brightness and color.



Each pixel has a color. The color is a 32-bit integer. The first eight bits determine the redness of the pixel, the next eight bits the greenness, the next eight bits the blueness, and the remaining eight bits the transparency of the pixel.



### Image processing:

Digital image processing, the manipulation of images by computer, is relatively recent development in terms of man's ancient fascination with visual stimuli. In its short history, it has been applied to practically every type of images with varying degree of success. The inherent subjective appeal of pictorial displays attracts perhaps a disproportionate amount of attention from the scientists and also from the layman. Digital image processing like other glamour fields, suffers from myths, mis-connections, mis-understandings and mis-information. It is vast umbrella under which fall diverse aspect of optics, electronics, mathematics, photography graphics and computer technology. It is truly multidisciplinary endeavor ploughed with imprecise jargon.

Several factor combine to indicate a lively future for digital image processing. A major factor is the declining cost of computer equipment. Several new technological trends promise to further promote digital image processing. These include parallel processing mode practical by low cost microprocessors, and the use of charge coupled devices (CCDs) for digitizing, storage during processing and display and large low cost of image storage arrays.

### Image Enhancement:

Image enhancement is among the simplest and most appealing areas of digital image processing. Basically, the idea behind enhancement techniques is to bring out detail that is obscured, or simply to highlight certain features of interesting an image. A familiar example of enhancement is when we increase the contrast of an

image because "it looks better." It is important to keep in mind that enhancement is a very subjective area of image processing.



### Image Restoration:

Image restoration is an area that also deals with improving the appearance of an image. However, unlike enhancement, which is subjective, image restoration is objective, in the sense that restoration techniques tend to be based on mathematical or probabilistic models of image degradation.



Enhancement, on the other hand, is based on human subjective preferences regarding what constitutes a "good" enhancement result. For example, contrast stretching is considered an enhancement technique because it is based primarily on the pleasing aspects it might present to the viewer, where as removal of image blur by applying a deblurring function is considered a restoration technique.

### Representation and Description:

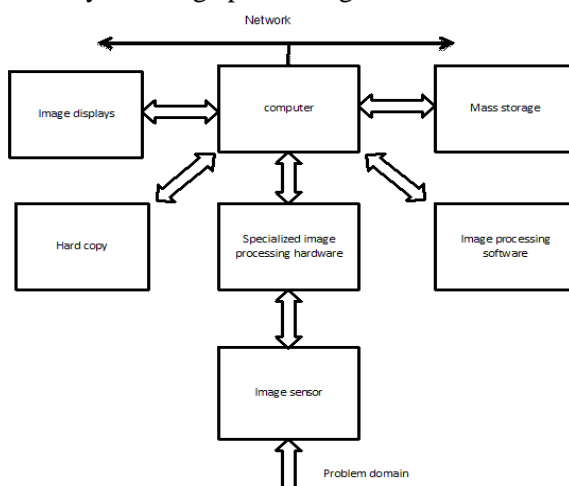
Representation and description almost always follow the output of a segmentation stage, which usually is raw pixel data, constituting either the boundary of a region (i.e., the set of pixels separating one image region from another) or all the points in the region itself. In either case, converting the data to a form suitable for computer processing is necessary. The first decision that must be made is whether the data should be represented as a boundary or as a complete region.

Boundary representation is appropriate when the focus is on external shape characteristics, such as corners and inflections.

Regional representation is appropriate when the focus is on internal properties, such as texture or skeletal shape. In some applications, these representations complement each other. Choosing a representation is only part of the solution for transforming raw data into a form suitable for subsequent computer processing. A method must also be specified for describing the data so that features of interest are highlighted. Description, also called feature selection, deals with extracting attributes that result in some quantitative information of interest or are basic for differentiating one class of objects from another.

### Components of an Image Processing System

As recently as the mid-1980s, numerous models of image processing systems being sold throughout the world were rather substantial peripheral devices that attached to equally substantial host computers. Late in the 1980s and early in the 1990s, the market shifted to image processing hardware in the form of single boards designed to be compatible with industry standard buses and to fit into engineering workstation cabinets and personal computers. In addition to lowering costs, this market shift also served as a catalyst for a significant number of new companies whose specialty is the development of software written specifically for image processing.



Although large-scale image processing systems still are being sold for massive imaging applications, such as processing of satellite images, the trend continues toward miniaturizing and blending of general-purpose small computers with specialized image processing hardware. Figure 1.24 shows the basic components comprising a typical general-purpose system used for digital image processing. The function of each component is discussed in the following paragraphs, starting with image sensing.

### CONCLUSION

The primary aim of this research is to develop a framework for an automatic semantic content extraction system for videos which can be utilized in various areas, such as surveillance, sport events, and news video applications. The novel idea here is to utilize domain ontologies generated with a domain-independent ontology-based semantic content metaontology model and a set of special rule definitions.

Automatic Semantic Content Extraction Framework contributes in several ways to semantic video modeling and semantic content extraction research areas. First of all, the semantic content extraction process is done automatically. In addition, a generic ontology-based semantic metaontology model for videos (VISCOM) is proposed. Moreover, the semantic content representation capability and extraction success are improved by adding fuzziness in class, relation, and rule definitions. An automatic Genetic Algorithm-based object extraction method is integrated to the proposed system to capture semantic content.

In every component of the framework, ontology-based modeling and extraction capabilities are used. The test results clearly show the success of the developed system. As a further study, one can improve the model and the extraction capabilities of the framework for spatial relation extraction by considering the viewing angle of camera and the motions in the depth dimension.

## REFERENCES

- [1] M. Petkovic and W. Jonker, "An Overview of Data Models and Query Languages for Content-Based Video Retrieval," Proc. Int'l Conf. Advances in Infrastructure for E-Business, Science, and Education on the Internet, Aug. 2000.
- [2] M. Petkovic and W. Jonker, "Content-Based Video Retrieval by Integrating Spatio-Temporal and Stochastic Recognition of Events," Proc. IEEE Int'l Workshop Detection and Recognition of Events in Video, pp. 75-82, 2001.
- [3] L.S. Davis, S. Fejes, D. Harwood, Y. Yacoob, I. Haratoglu, and M.J. Black, "Visual Surveillance of Human Activity," Proc. Third Asian Conf. Computer Vision (ACCV), vol. 2, pp. 267-274, 1998.
- [4] G.G. Medioni, I. Cohen, F. Bre´mond, S. Hongeng, and R. Nevatia, "Event Detection and Analysis from Video Streams," IEEE Trans. Pattern Analysis Machine Intelligence, vol. 23, no. 8, pp. 873-889, Aug. 2001.
- [5] S. Hongeng, R. Nevatia, and F. Bre´mond, "Video-Based Event Recognition: Activity Representation and Probabilistic Recognition Methods," Computer Vision and Image Understanding, vol. 96, no. 2, pp. 129-162, 2004.
- [6] A. Hakeem and M. Shah, "Multiple Agent Event Detection and Representation in Videos," Proc. 20th Nat'l Conf. Artificial Intelligence (AAAI), pp. 89-94, 2005.
- [7] M.E. Do¨nderler, E. Saykol, U. Arslan, O¨ . Ulusoy, and U. Gu¨du¨kbay, "Bilvideo: Design and Implementation of a Video Database Management System," Multimedia Tools Applications, vol. 27, no. 1, pp. 79-104, 2005.
- [8] T. Sevilmiş, M. Bastan, U. Gu¨du¨kbay, and O¨ . Ulusoy, "Automatic Detection of Salient Objects and Spatial Relations in Videos for a Video Database System," Image Vision Computing, vol. 26, no. 10, pp. 1384-1396, 2008.
- [9] M. Ko¨pru¨lu¨, N.K. Cicekli, and A. Yazici, "Spatio-Temporal Querying in Video Databases," Information Sciences, vol. 160, nos. 1-4, pp. 131-152, 2004.
- [10] J. Fan, W. Aref, A. Elmagarmid, M. Hacid, M. Marzouk, and X. Zhu, "Multiview: Multilevel Video Content Representation and Retrieval," J. Electronic Imaging, vol. 10, no. 4, pp. 895-908, 2001.
- [11] J. Fan, A.K. Elmagarmid, X. Zhu, W.G. Aref, and L. Wu, "Classview: Hierarchical Video Shot Classification, Indexing, and Accessing," IEEE Trans. Multimedia, vol. 6, no. 1, pp. 70-86, Feb. 2004.
- [12] L. Bai, S.Y. Lao, G. Jones, and A.F. Smeaton, "Video Semantic Content Analysis Based on Ontology," IMVIP '07: Proc. 11th Int'l Machine Vision and Image Processing Conf., pp. 117-124, 2007.
- [13] R. Nevatia and P. Natarajan, "EDF: A Framework for Semantic Annotation of Video," Proc. 10th IEEE Int'l Conf. Computer Vision Workshops (ICCVW '05), p. 1876, 2005.