

# Image Sequence Coding by Octrees

*Riccardo Leonardi*

Visual Communications Research Department

AT&T Bell Laboratories / rm 4C-507

Crawfords Corner Road

Holmdel, NJ 07733 (U.S.A.)

## ABSTRACT

This work addresses the problem of representing an image sequence as a set of octrees. The purpose is to generate a flexible data structure to model video signals, for applications such as motion estimation, video coding and/or analysis.

An image sequence can be represented as a 3-dimensional causal signal, which becomes a 3 dimensional array of data when the signal has been digitized. If it is desirable to track long-term spatio-temporal correlation, a series of octree structures may be embedded on this 3D array. Each octree looks at a subset of data in the spatio-temporal space. At the lowest level (leaves of the octree), adjacent pixels of neighboring frames are captured. A combination of these is represented at the parent level of each group of 8 children. This combination may result in a more compact representation of the information of these pixels (coding application) or in a local estimate of some feature of interest (e.g., velocity, classification, object boundary). This combination can be iterated bottom-up to get a hierarchical description of the image sequence characteristics.

A coding strategy using such data structure involves the description of the octree shape using one bit per node except for leaves of the tree located at the lowest level, and the value (or parametric model) assigned to each one of these leaves. Experiments have been performed to represent Common Image Format (CIF) sequences.

## 1. INTRODUCTION

Image sequences are used in a variety of applications ranging from computer vision for robotics applications, surveillance systems, 3-dimensional object recognition, to video transmission. Most systems exploit the correlation that exists between 2 successive frames of the image sequence over time. Recently, research efforts have tried to combine information present in several frames to reduce or better use redundancy of information over time [1-3]. Decomposition of images into regions, i.e. segmentation, has been efficiently applied for image processing applications, ranging from classification to compression [4]. This leads naturally to the idea of extending such concept to the time dimension and generating regions defining 3D volumes resulting from the partition the 3-dimensional array defining a digital image sequence.

In two dimensions, segmentation is a difficult task, especially when it is desired to obtain regions corresponding to objects of the natural scene of which the image is a projection. Adding the time dimension might ease the segmentation task by combining luminance/chrominance information to motion information. Once a good segmentation is achieved, a description of the image sequence may be performed by describing how the different objects evolve with respect to time (motion, deformation, occlusion, ...) Intuitively, it is possible to describe the 3-dimensional shape of the various volumes that result from the clustering of the same object over several frames. Unfortunately, the exact boundary representation of such volumes is costly and complex.

On the other hand, adequate 3-dimensional data structures may ease the representation and analysis of image sequences, without requiring the complex description of 3-dimensional volumes. Variable block-size schemes using quadtrees have been suggested to this extent in two dimensions [5-7]. This paper extends this concept by the use of octrees to model image sequences. Section 2 describes how to build such data structure for video sequences. Section 3 focuses on the use of octrees for video compression applications.

## 2. OCTREE REPRESENTATION OF VIDEO SEQUENCES

To track long-term spatio-temporal correlation, a series of octree structures is embedded on this 3D array. Each octree looks at a subset of data in the spatio-temporal space. At the lowest level (leaves of the octree), adjacent pixels of neighboring frames are captured. A combination of these is represented at the parent level of each group of 8 children. This combination may result in a more compact representation of the information of these pixels (coding application) or in a local estimate of some feature of interest (e.g., velocity, classification, object boundary). Such process can be iterated bottom-up to get a hierarchical description of the image sequence characteristics. Conversely, a top-down approach can be derived, by iteratively dividing a cube of data into 8 identical sub-cubes, as long as a global description of the initial cube can be derived at the highest level and subsequently to all lower levels. If an intermediary level node is retained to be accurate to describe the information of interest, lower level children can be omitted, resulting in a partitioning of the original 3-dimensional array into variable size cubes. Figure 1 describes such a decomposition process.

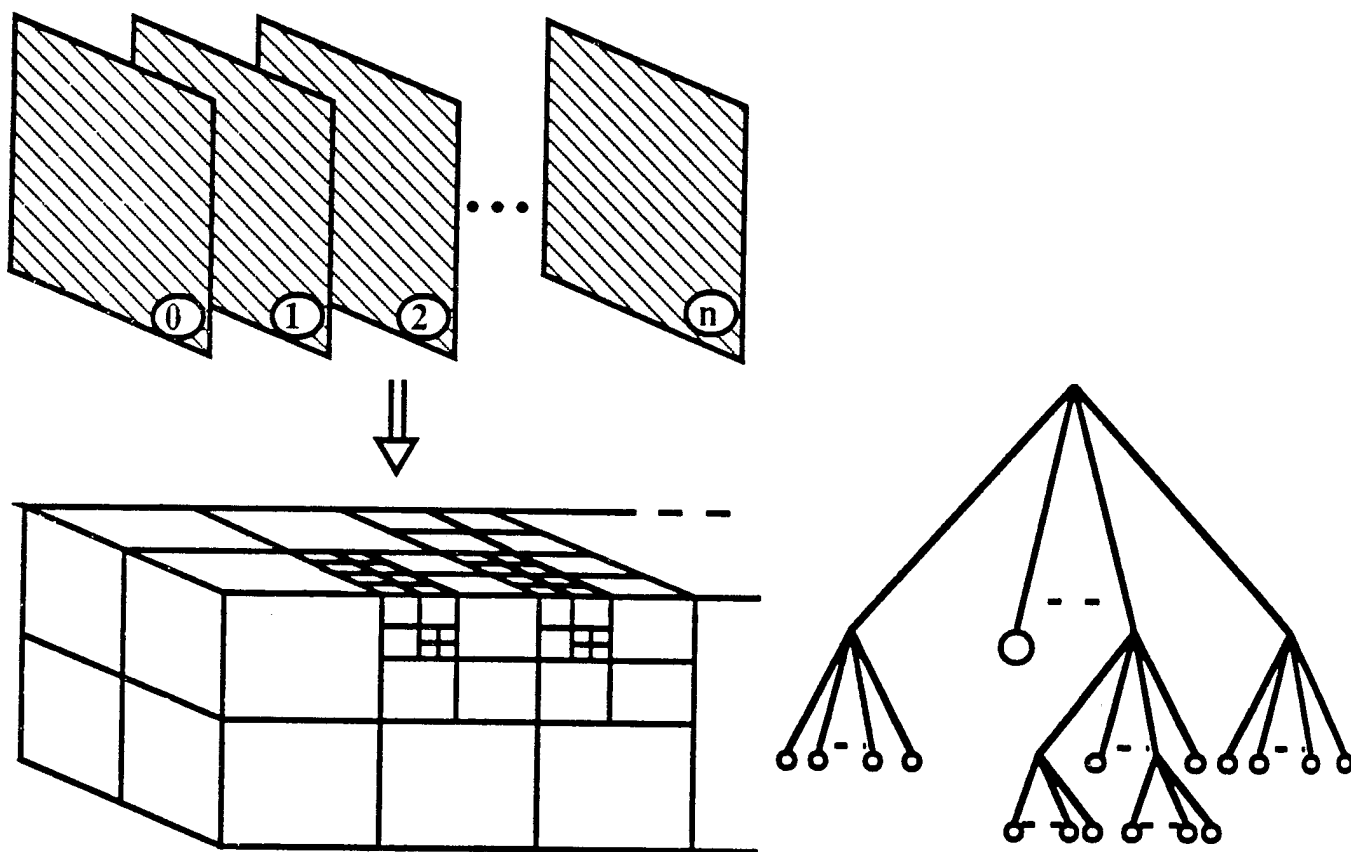


Figure 1: Octree decomposition of an image sequence.

### 3. OCTREE FOR VIDEO COMPRESSION

We focus here on coding applications. By choosing an appropriate distortion measure, it is possible to reconstruct the signal from high-level nodes (combination up in the tree) for certain locations in the spatio-temporal domain, whereas lower nodes are used for other locations (areas with new motion or contours). This decomposition will be sensitive both in respect to the way of combining the various children of a node in the octree, as to the distortion measure chosen to estimate the quality of the reconstruction. To further improve the performance of a coding algorithm, we suggest to use the prediction of the octree decomposition between successive octrees along the time axis as only areas with motion will change the decomposition. Prediction can be accomplished as well within one octree between nodes corresponding to neighboring areas in the spatio-temporal domain.

Delay constraints may limit the depth of the octree decomposition along the time axis. On the other hand, reconstruction may be rapidly achieved by scanning a reconstructed octree at the decoder.

An implementation of this coding strategy has been tried out using a very simple combination of the various children of each parent node, i.e. their mean. The processed data use the Common Image Format (288x360 progressive 4:2:2 frames). A set of 9 adjacent octrees was built upon each group of 8 successive frames, each one containing at most  $128 \times 128 \times 8$  pixels. Within each octree, both decompositions with respect to luminance Y and chrominance difference signals (Cr, Cb) have been considered. The decomposition of the original sequence into the octree is obtained by using one bit per node except for leaves of the tree located at the lowest level. The value of each node is being DPCM coded with respect to neighboring nodes.

If the distortion measure is defined by the variance of each cubic block, the threshold over this variance measure has to be kept relatively low to guarantee an accurate reconstruction of the original signal. It is desirable to extract contour-motion masks to facilitate the decomposition of the 3-dimensional spatio-temporal array into a reduced size octree. Current effort is being placed in the design of spatio-temporal filters to obtain these control masks.

#### References

- [1] A. Puri, "Conditional Motion-Compensated Interpolation and Coding", *Second International Workshop on 64 kbit/s coding of Moving Video*, paper No. 1-5, Hannover, Sept. 4-6, 1989.
- [2] P. Willemin, T. Reed and M. Kunt, "Image Sequence Coding at Very Low Bit Rates with a 3-D Split and Merge Algorithm", *Second International Workshop on 64 kbit/s coding of Moving Video*, paper No. 3-4, Hannover, Sept. 4-6, 1989.
- [3] M. Orchard and R. Leonardi, "Motion Segmentation Prediction for Hybrid Video Coding", Submitted to Picture Coding Symposium, March 1990.
- [4] R. Leonardi, "Adaptive Segmentation for Image Coding", *Ph.D. thesis No. 691*, Dept. of Electrical Eng., Swiss Fed. Inst. of Tech., Lausanne, Switzerland, June 1987.
- [5] R. Wilson, "Quadtree Predictive Coding - A New Class of Image Data Compressions Algorithms", in *Proceedings of the Intern. Conf. on ASSP*, paper 29.3.1, San Diego, March 1984.
- [6] R. Leonardi and M. Kunt, "Adaptive Split for Image Coding", in *Proceedings IASTED Intern. Symp. for Applied Signal Proc. and Digital Filtering*, Paris, June 85.
- [7] D.J. Vaisey and A. Gersho, "Variable Blocksize Image Coding", in *Proceedings Intern. Conf. on ASSP*, pp. 1051-1054, Dallas, TX, April 1987.