17.3: Color Motion Video Coded by Perceptual Components

A. B. Watson
NASA Ames Research Center, Moffett Field, CA
C. L. M. Tiana
Sterling Software, Moffett, Field, CA

Abstract

We describe an implementation of an architecture for coding and compression of color motion video that is based upon the partition of the visual signal by the early human visual system. The architecture consists of a pyramid in space and time, with separate bands for static and moving picture elements. Experiments on a highly dynamic 256x256 color image sequence suggest acceptable quality at 1 bit/pixel.

Introduction

A wide range of emerging applications, including digital movies, HDTV, and transmission and visualization of scientific imagery(Jaworski, 1990), will require efficient methods of coding color motion video. For much of this imagery, the ultimate consumer is the human eye, and image codes should be designed to match the visual capacities of the human observer. Elsewhere we have proposed a general Perceptual Components Architecture (PCA) for digital video based upon these ideas (Watson, 1990a,b, 1991). In this report we briefly describe an implementation of PCA coding of color motion video, and provide some preliminary results on the effectiveness of the scheme.

Perceptual Components Architecture

The Perceptual Components Architecture is based on our current understanding of how imagery is decomposed in early human vision, and consists of transformations and partitions of color, spatial, and temporal dimensions. Beginning with a digital image sequence whose pixels are indexed by row, column, frame, and color (R, G, B), the color dimension is first transformed from RGB into an opponent color space, nominally white /black (WB), red/green (RG), and blue /yellow (BY). The spatial dimension is partitioned into a number of bands of spatial frequency and orientation . Finally the temporal dimension is partitioned into four bands: low, left, right, and high . The spatiotemporal bands are grouped in such a way that the signal is partitioned into components moving in particular directions. In the three-dimensional spatiotemporal frequency domain, these moving components correspond to paired regions on either side of the origin.

In general form, the PCA is a type of analysis/synthesis filter bank (Woods, 1991). The signal is first decomposed by a bank of analysis filters, downsampled appropriately, quantized, upsampled, and reconstructed by means of a synthesis filter. In the present case, analysis and synthesis filters are equivalent.

Color Transform

We transform from RGB monitor primaries to a perceptually based opponent color space using the following matrix:

$$\begin{bmatrix} WB \\ RG \\ BY \end{bmatrix} = \begin{bmatrix} 0.4523 & 0.8724 & 0.1853 \\ 0.7976 & -0.5499 & -0.2477 \\ -0.2946 & -0.5132 & 0.8062 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix}$$

This matrix is based on the so-called "cardinal directions" of color space (Derrington, Krauskopf, & Lennie, 1984; Krauskopf, Williams, & Heeley, 1982; Mulligan & Ahumada, 1992) . We have not at this time properly tuned this matrix to the chromaticity coordinates of our display monitor.

Spatial Partition

The spatial partition is implemented by means of the Cortex Transform (Watson 1987a,b). This is an invertible pyramid transform which divides the frequency space using concentric rings and radial wedges (Fig. 1). The rings are a constant logarithmic distance apart. In the present work, we used radial bandwidths of one octave and orientation bandwidths of 45 degrees.

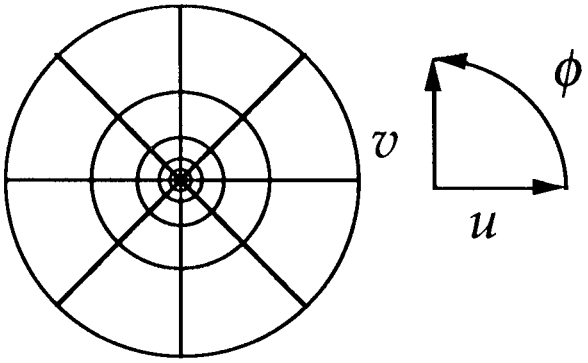

Fig. 1. Partition of the spatial frequency domain. Spatial frequencies and orientation are indicated by (u,v) and ϕ.

The borders of each region are "softened" by convolution with a Gaussian whose width is proportional to the frequency of the filter. Each sector of the resulting partition is equivalent to all others after scaling and rotation, so that this is an example of a "wavelet" transform (Grossman & Morlet, 1984).

These filters are self-inverting, in the sense that sending an image through the set of filters twice reproduces the image. This means that the sum of the squares of the filters equals 1.

Because the input sequence is real, its Fourier transform has conjugate symmetry, and we consequently only need to encode half the frequency plane. We do this by using only four of the possible eight sectors within each ring, covering the upper half of the spatial frequency plane.

This spatial partition is applied separately to each of the three color channels (WB, RG, and BY).

Temporal Partition

The temporal partition is accomplished by means of four filters. Each filter consists of two opposed cumulative Gaussians, as defined by the following Mathematica (Wolfram, 1991) function,

```
filter[length_,scale_,corner_] := Module[ {tmp,x},
 tmp=Table[ Sqrt[.5 * Erfc[(x-corner*length/2-1)
  / scale ] ] ,{x, length/2 + 1}];
 Join[tmp,Reverse[ Take[tmp,{2,length/2}]]]]
```

where Erfc is the complementary error function (cumulative Gaussian), length is the number of frames in a coded segment, scale is a scale factor, and corner defines the 50% cutoff of each flank, expressed in terms of the Nyquist frequency. We used length=8, scale=0.5, corner=0.25. The four filters are four copies of this filter, shifted by increments of length/4, as pictured in Fig. 2. As with the spatial filters, the temporal filters are self-inverting.

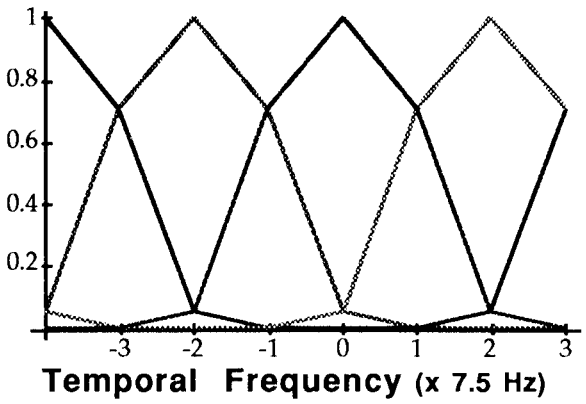

Fig. 2. Temporal filters.

Motion Components

As noted above, we exploit the conjugate symmetry of the real input sequence by using only four of the eight possible orientation filters at each spatial frequency. When the four temporal filters are applied separably to one of these four spatial filters, four spatio-temporal filters result, two of which are selective for motion components. For example, considering the spatial filter encompassing 0-45 degrees of orientation, the four resulting spatio-temporal filters are: stationary low temporal frequency, rightward motion, stationary high temporal frequency, and leftward motion.

The result of applying one of these spatiotemporal filters to the image sequence twice, once in the analysis stage and once in the synthesis stage, is a complex image sequence. Retaining the real part yields the appropriate motion component.

Although there have been a few recent experiments with three-dimensional subband coding of video (Karlsson, & Vetterli, 1988; Kovacevic, 1991), this is to our knowledge the first use of explicit motion components.

Sampling

After spatial and temporal filtering, each band is subsampled in space and time. The spatial subsampling is via a sampling matrix k S, where

$$S = \begin{bmatrix} 1 & 3 \\ 3 & 1 \end{bmatrix}$$

and where k = image width/ filter width (Watson, 1987b). For an image of width 256, the highest spatial frequency filter has a width of 256 (in the frequency domain). Temporal subsampling was by 2. The resulting collection of complex samples is overcomplete by a factor of 8/3.

Quantization

Each band was quantized uniformly with a particular divisor (bin width). Various different schemes were tried, but in general divisors increased with spatial frequency, with temporal frequency, and with color (BY>RG>WB). First order entropy was computed for each band and accumulated.

Test Material

To test the implementation, we have worked with a short (8 frame) segment from an MPEG test sequence (football). The original material was cropped from 256 by 192 to 242 by 192 to remove black borders, and expanded to 256 by 256 by bicubic interpolation. The sequence contains saturated colors, high contrast luminance and color borders, and a very large amount of motion, including panning and object motion. A single frame is shown in Fig. 3.

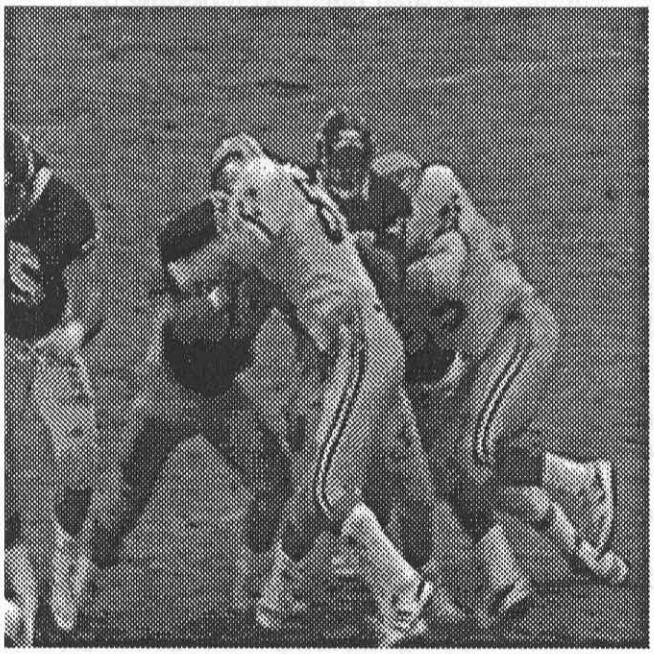

Fig. 3. A single frame from the test sequence. The actual test sequence was in color.

For subjective evaluation, reconstructed sequences were compared to the original at a viewing distance of approximately three picture heights.

Results

Application of the method to the test sequence suggests acceptable quality rendition at rates around one bit/pixel. Optimal compression requires appropriate setting of the quantization factor for each band, which is a difficult multidimensional problem.

Our search strategy was to first establish the range of sample values for each component (combination of spatial frequency, orientation, temporal filter, and color channel). The initial quantization bin width (w_0) for each component was set to the range for that component divided by 256, to produce 256 possible bins. We then compressed all four orientations of one component, using bin widths of $2^k w_0$, k={0,...8}. We then examined each reconstructed sequence to determine the bin width yielding a perceptually lossless result. This was repeated for all spatial frequencies, temporal frequencies, and colors. For the WB channel, k was typically around 4 or 5, and nearly independent of spatial frequency. For the color channels, k= 8 for the upper 2 spatial frequencies, and around 6 for the lower resolutions.

Conclusions and Discussion

We have implemented a prototype Perceptual Components Architecture for digital color image sequence coding. Preliminary results on a brightly colored, rapidly moving , 256 by 256 test sequence suggest acceptable quality at around 1 bit/pixel. Higher resolution sequences will generally require lower bit rates (in bits/pixel) for an equivalent viewing distance (in picture heights), since the added resolution will be at relatively less visible high spatial frequencies.

The current scheme is overcomplete by a factor of 8/3. This is largely due to down sampling in time by only a factor of two, in spite of the use of four time filters. We are currently examining the use of four-channel perfect reconstruction filter banks, with downsampling by 4 in time (Vaidyanathan, 1990), to reduce the redundancy to a factor of 4/3.

Acknowledgments

We thank Eero Simoncelli for useful discussions. This work was supported by NASA RTOP 506-71-51.

References

Derrington, A. M., Krauskopf, J., & Lennie, P. (1984). Chromatic mechanisms in lateral geniculate nucleus of macaque. J. Physiol, London 357, 241-265.

Grossman, A., & Morlet, J. (1984). Decomposition of Hardy functions into square integrable wavelets of constant shape. SIAM J. Math. 15, 723-736.

Jaworski, A. (1990). Earth Observing System (EOS) Data and Information System (DIS) software interface standards. Pasadena, CA: American Institute of Aeronautics and Astronautics.

Karlsson, G., & Vetterli, M. (1988). Three dimensional subband coding of video. New York: 1100-1103.

Kovacevic, J. (1991). Filter banks and wavelets: Extensions and applications Columbia University, Center for Telecommunications Research Technical Report CU/CTR/TR 257-91-38.

Krauskopf, J., Williams, D. R., & Heeley, D. W. (1982). Cardinal directions of color space. Vision Research 22, 1123-1131.

Mulligan, J. B., & Ahumada, A. J., Jr. (1992). Principled methods for color dithering based on models of the human visual system. Society for Information Display Digest of Technical Papers 23.

Vaidyanathan, P. P. (1990). Multirate digital filters, filter banks, polyphase networks, and applications: a tutorial. Proceedings of the IEEE 78(1), 56-93.

Watson, A. B. (1987a). The cortex transform: Rapid computation of simulated neural images. Computer Vision, Graphics, and Image Processing 39(3), 311-327.

Watson, A. B. (1987b). Efficiency of an image code based on human vision. Journal of the Optical Society of America A 4(12), 2401-2417.

Watson, A. B. (1990a). Digital visual communications using a perceptual components architecture. Pasadena, CA: American Institute of Aeronautics and Astronautics.

Watson, A. B. (1990b). Perceptual-components architecture for digital video. J. opt. Soc. Amer. A 7(10), 1943-1954.

Watson, A. B. (1991). Multidimensional pyramids in vision and video. In A. Gorea (Ed.), Representations of vision: trends and tacit assumptions in vision research (pp. 17-26). Cambridge: Cambridge University Press.

Wolfram, S. (1991). Mathematica: A system for doing mathematics by computer (Second Edition ed.). New York: Addison-Wesley.

Woods, J. W. (1991). Subband image coding. Norwell, MA: Kluwer Academic Publishers.