A Resolution-Independent Image Representation for Digital Cinema

Sylvain Brugnot Xiangyang Ju Paul Cockshott Paul Siebert
Computing Science Department, University of Glasgow, G12 8QQ, Glasgow, UK
{sylvain, xju, wpc, psiebert} @dcs.gla.ac.uk

ABSTRACT

We present a generic architecture for a novel type of image and movie representation designed to allow resolution-independent manipulation of movie sequences. By expressing the input signal as a sum of analytically-defined basis functions across space, time and scale, it becomes possible to compute image values at any location in the sequence, thereby providing a basis for both spatial and temporal resolution-independence. Distortion-constrained encoding of the basis function coefficients is furthermore built into the architecture to achieve data compression whilst minimizing signal degradation.

Keywords

resolution-independence, image coding, video encoding, post-production.

1. INTRODUCTION

Filmmakers today are increasingly turning towards an all-digital solution, from image capture to post-production and projection. Due to its fairly recent appearance, the digital cinema chain still suffers from limitations which can hamper the productivity and creativity of cinematographers and production companies.

Firstly, the variety of means by which digital material can be obtained (e.g., 35mm film scans, high-definition film cameras, video cameras) and the various display methods (TV, HDTV, projection) have led to the coexistence of multiple resolution standards, such as: PAL (720x576 pixels), 1080i (1920x1080 pixels), 2K (2048x1080 pixels) and more recently 4K (4096x2160 pixels). Similarly, this also occurs in the temporal domain, where values of 24, 30 or 60 frames/second coexist. The combination of these two factors results in over 25 different video formats. This absence of a unified standard of spatial and temporal resolution implies that digital film material from different sources is more costly to combine and composite at the production stage.

Secondly, the current standard for the storage and transfer of film data is the CINEON or DPX format (Digital Picture eXchange). Each DPX file encodes an individual frame as an array of uncompressed pixel values. Such files are extremely large, at typically over 10MB per frame, which corresponds to over a Terabyte per hour of film at 24 frames/second. Despite the ongoing technological advances in disk capacity and network speed, this makes the storage and transfer of movie data between collaborators along the digital cinema chain impractical and costly.

2. AN ARCHITECTURE FOR THE ENCODING OF MOVING IMAGES

In order to address the two issues mentioned above, we present an architecture for the resolution-independent encoding of digital cine material, which offers compression whilst minimizing resulting artifacts. This generic architecture allows a variety of specific implementations, each using different encoding schemes. The following concepts are the key to the architecture:

1. The cine sequence is treated as a three dimensional array of samples handled by voxel compression techniques.
2. For each block, a scale-space approach analogous to image pyramids is used to capture correlations across voxels in time and space.
3. At each level of the pyramid, the signal is expressed as a sum of analytically defined basis functions across space and time.

Posters proceedings ISBN 80-86943-04-6
WSCG'2006, January 30-February 3, 2006
Plzen, Czech Republic.

4. A closed loop feedback system is used to compensate for artifacts introduced by quantization at higher levels.
5. The quantization is constrained so that the resulting artifacts always remain within the statistical boundaries of in-camera shot noise.

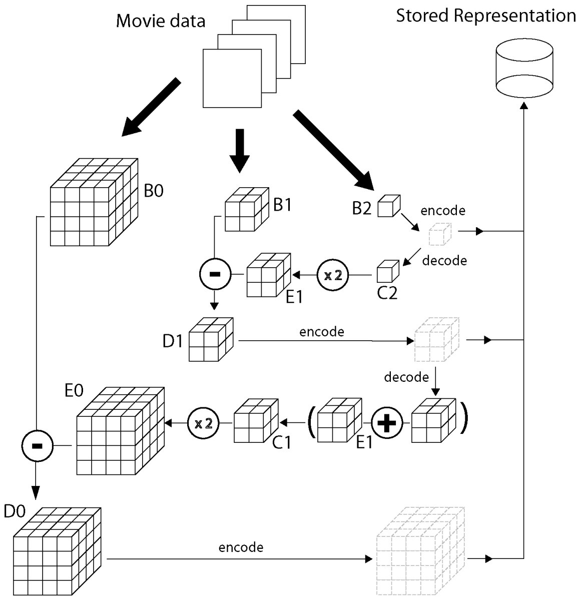

Figure 1. Encoding Architecture

The first phase of the algorithm, illustrated in Figure 1, gathers individual digital frames into three-dimensional blocks of pixels. Each block is then spatially and temporally filtered and downsampled repeatedly to form a structure where each successive block is smaller in each dimension by a scale factor *r* (typically 2 or 3). Blocks B0, B1 and B2 in Figure 1 form such a structure. We refer to this sequence of blocks as a pyramid by analogy with the commonly used image pyramids of [Bur83a]. We refer to the largest block as being at the base of the pyramid and the smallest as being at the top.

Consider the top block, B2 in the case of Figure 1. We apply to it some encoder (which, in the case of the top block, may be null). A copy of the encoded representation is sent to the output stream after which the encoded representation is internally decoded by the compressor and stored in block C2.

C2:= decode(encode(B2))

An interpolation algorithm, which may exploit features of the encoding previously applied, is then used to expand C2 by a scale factor of *r* to obtain block E1 which is the same size as block B1 but lacks high frequency information.

E1:=expandby(r,C2)

E1 may include some artifacts due to quantization or the interpolation technique used. We then form

D1:= B1-E1

so that D1 is a differential block containing the information from B1 that was not captured by the encoding of E1.

D1 is then itself encoded, perhaps using a different encoder than that used at the top level. The encoded output is appended to that produced for the top level of the pyramid. Again the encoded form of the current level is internally decoded and added to E1 to produce C1.

C1:= E1 + decode(encode(D1))

The block C1 should now be a close approximation to the original downsampled block B1. It can be expanded again by a factor of *r* to form a block E0, as large as the original source block B0. Subtracting them we again obtain a difference block:

E0:=expandby(r,C1)

D0:=B0-E0

which we encode, appending the ecoding to the output stream.

3. PRELIMINARY RESULTS

For evaluation purposes, the architecture presented here was implemented using trivariate polynomials as basis functions, resulting in an approach similar to the one found in [Cyg96a]. When no quantization scheme is applied, it was shown that the original signal can be reproduced with 100% accuracy, as well as rendered at varying resolutions with a result visually similar to that of well-established interpolation techniques such as bicubic resampling.

This architecture lends itself well to vector quantization methods, such as described in [Lin80a] and [Tao05a]. The performance of such a system is currently being investigated.

4. REFERENCES

[Bur83a] "The Laplacian Pyramid as a Compact Image Code", P.J. Burt, E.H. Adelson, IEEE Transactions on Communications, vol. 31, n. 4, 1983, pp 532-540

[Cyg96a] “3-D Block Decomposition Approach for Video Coding”, Boguslaw Cyganek, Image Processing & Communications, An International Journal, 1996, vol. 2, n. 4

[Lin80a] “An Algorithm for Vector Quantizer Design” Y. Linde, A. Buzo, et al., IEE Transactions on Communication, 1980, vol. 28, pp. 84-95

[Tao05a] “Distortion-Constrained Compression of Three-Dimensional CLSM Images Using Image Pyramid and Vector Quantization”, Y. Tao, PhD Thesis, University of Glasgow, 2005.