Segmentation-Based Image Coding by Morphological Local Monotonicity

Joseph Bosworth and Scott T. Acton
University of Virginia
Department of Electrical Engineering
Charlottesville, VA 22904

Abstract

Segmentation-based image and video coding is desirable for many multimedia applications due to the additional functionality provided by object-based representation. Methods of object-based coding have generally treated the segmentation and encoding processes as separate problems. Here, we present an integrated segmentation and coding method unified by the theoretical structure of *morphological local monotonicity.* This unified segmentation/coding scheme utilizes morphological operators within a nonlinear scale-space to generate a segmentation. The segmented regions are independently coded and reconstructed using a morphological generalization of Laplace's equation in a multiresolution framework. The coding procedure is appropriate for non-textured imagery and avoids arbitrarily chosen constants. Examples are given for two-dimensional grayscale imagery.

1. Introduction

Object-based coding requires both a segmentation method and a means of compression. The segmentation step typically employs an edge-detection or measure of discontinuity, while the compression step typically uses an assumption of continuity or correlation between neighboring pixels. Though both tasks rely on some measure of intra-object homogeneity, they are treated as separate procedures, each with its own arbitrary constants and methods for measuring homogeneity.

We wish to unify the processes of segmentation and object-based coding by relating the parameters used in each process and employing a coding method that is naturally object-based. Here, we restrict our initial study to the case of non-textured imagery. In this special case, intra-object homogeneity can be equated with the smoothness of graylevel values. The more general case of textured imagery may be explored in future research where homogeneity is evaluated by texture features rather than simple graylevel values.

For non-textured imagery, edge detection should measure a lack of smoothness of a magnitude exceeding some perceptually significant threshold. Likewise, coding should consist of generating a minimal description of a smooth surface for efficient transmission and reconstruction at the decoder. Our goal is to employ the same mathematical operators in both the detection of smooth surfaces (or lack thereof) and the reconstruction of smooth surfaces from minimal descriptors.

Toward this goal, we exploit a measure of signal smoothness – local monotonicity. This signal characteristic can be expressed in terms of self-dual morphological filters. Here, the operators systematically replace classic image processing tools, such as Gaussian filters, gradient and edge detectors, and Laplacian operators, with their self-dual morphological counterparts while avoiding the use of arbitrarily chosen parameters. Advantages of the morphological approach also include efficient implementation, edge localization, and specific scale properties by the use of fixed-sized structuring elements. We begin with a review of local monotonicity and its relationship to morphology.

2. Morphological Local Monotonicity

In one dimension (1-D), a signal is locally monotonic of degree or scale n ("lomo-n") if and only if the signal is monotonic within every interval of length n. This concept has been generalized to higher dimensions by the use of self-dual morphology [1], and we refer to this generalization as morphological local monotonicity to distinguish it from other proposed definitions [2]. Such multidimensional lomo signals possess specific scale properties. Here we restrict our discussion to the discrete 2-D case and note two of these properties [1]:

Property 1. A lomo-n signal is a root signal of the morphological lomo filter:

$$f(x,y) \leftarrow \frac{f(x,y)\circ k(x,y) + f(x,y)\bullet k(x,y)}{2}, \qquad (1)$$

0-7803-6514-3/00/$10.00©2000 IEEE

where the signal f is iteratively filtered using a constant-valued circular structuring element k until convergence. The structuring element consists of all pixels with Euclidean distance from the center pixel is less than or equal to radius r, where r is related to the scale n by

$$r(n) = (n-2)/2. \quad (2)$$

Property 2. Each local extremum belongs to a constant-valued plateau enclosing the structuring element.

The self-dual property and use of circular structuring elements make the lomo filter appropriate for scale-space generation by avoiding graylevel and orientation biases. Also, as is shown in the following section, the well-defined plateau property facilitates sampling that occurs in a pyramidal coding technique.

3. Locally Monotonic Coding

Here, we outline the use of multiscale locally monotonic image representations for object-based coding. It is assumed that a segmentation has been previously generated by an arbitrary method. In section 3.4, however, we suggest employing a segmentation method that is coupled to the coding method through the concept of local monotonicity. Inter-object segmentation boundaries are transmitted separately from intra-object graylevel content, and we do not make a contribution to such contour coding techniques.

3.1 The Lomo Pyramid

Progressive encoding is accomplished by the use of multi-resolution pyramids. At the encoder, the original image is used for the generation of a *lomo pyramid*, formed by lomo filtering and uniform sub-sampling. This pyramid is then coded in a multi-resolution coarse to fine order.

Pyramidal representations computed by filtering and sub-sampling have been used successfully for image coding [5]. Here, we show the generation of the lomo pyramid, which uses a self-dual lomo filter. In order to ensure proper sampling on a rectangular image grid, a 2 x 2 structuring element is employed during the filtering stage. Thus, the lomo plateau size (Property 2) is a 2 x 2 square. With this condition met at the local signal extrema, one-of-two horizontal and vertical sampling is guaranteed to include all extrema, or equivalently all level-set connected components. This pyramidal structure is efficient in deriving successive approximations to the original image.

For a given segment at a given pyramidal resolution, single-pixel graylevel values within the segment serve as boundary conditions for solving the morphological analogy to Laplace's equation. Here, the classic Laplace's equation,

$$\nabla^2 f(x,y) = 0, \quad (3)$$

is replaced by

$$\nabla^2_{morph} f(x,y) = s(x,y) - f(x,y) = 0, \quad (4)$$

with $s(x,y)$ defined as the mean between the dilation and erosion with respect to structuring element $k(x,y)$:

$$s(x,y) = \frac{f(x,y) \oplus k(x,y) + f(x,y) \ominus k(x,y)}{2}. \quad (5)$$

For 1-D lomo-n signals, this morphological analogy proportional to the common second derivative approximation

$$\frac{\partial^2}{\partial x^2} f(x) \approx \frac{f(x+\Delta) - 2f(x) + f(x-\Delta)}{2\Delta} \quad (6)$$

for lomo-n signals, where $\Delta = r(n)$.

This solution smoothly interpolates between the known boundary condition points. This definition of smoothness is again used in the segmentation process of the next section. The boundary condition data are transmitted to the decoder side, and the corresponding solution is used as an initial estimate of the segment. Additional points are chosen one by one at the locations of the greatest error between the decoded and original pyramid level.

The overall coding procedure then consists of determining the set of boundary condition points for each segment that serve to reconstruct the graylevels at the decoder. Because the coding is progressive, at each pyramid level there exists an estimate of the decoded image. Errors from this estimate are transmitted, rather than absolute graylevel values. These errors are quantized in accordance with an adjustable error tolerance *threshold* (a parameter which we relate to the edge detection threshold in our segmentation). Visually displeasing quantization artifacts such as staircasing are avoided, because the quantized values are superimposed upon the smoothly varying solution to Laplace's equation. Fig. 1 diagrams the encoding process.

3.2 The Morphological Laplace's Equation

A crucial step in the encoding process is the solution of the morphological Laplace's equation for interpolation between encoded points. The numerical solution is similar to its classical counterpart. It should be noted that the geodesic propagation and smoothing of [3], though not explicitly stated, can be viewed as an algorithm for the solution of the morphological Laplace's equation. In [3], the boundary conditions are the graylevel values from alongside edges, akin to the classical Dirichlet problem in partial differential equations. This reliance upon graylevels from the object boundaries makes the method very sensitive to edge localization errors. Even step-like edges which are properly localized but contain transition regions wider than a single pixel (*e.g.* due to camera focus) cause significant errors to be propagated into the interpolated regions.

In contrast, our method imposes isolated point values within a segment as boundary conditions rather than complete reliance upon precise edge localization or sharp step edges. This method imposes essential boundary conditions at this finite number of points, rather than along the entire segment boundary. The interpolation from these points extends to infinity, so that graylevel content of an object may be calculated independently of the shape of its contour. For example, the first value generates a constant-valued surface, three values determine a plane, etc. Upon reconstruction, this surface is truncated to match the contour shape of the corresponding segment. In this way, each segment is independently interpolated. An advantage of this approach for video sequences is that is automatically estimates occluded portions of objects.

The stopping condition for the encoding of individual boundary condition points is chosen to be the iteration in which no error values exceed a "perceptually significant graylevel difference" T as defined by the user. This parameter is also used in the segmentation process itself as an edge detection threshold, thus providing a natural coupling between the segmentation and coding method. In fact, the concepts of smoothness by local monotonicity and the morphological Laplacian that define the coding process are also the basis for the segmentation process, as explained in section 3.4.

For the computational solution of the morphological Laplace's equation, multiple techniques are possible. The method of [3] is one such possibility. Another, more conventional approach is similar to one used in the numerical solution of the classic Laplace's equation [6]. In the classic (2-D) solution, the mean of the values in a concentric circle about a given point replaces the value at that center point. With the boundary condition points serving as seeds, the iteration continues until convergence. (The circles must be small enough not to enclose boundary condition points within their interior.) When the solution is reached, all points equal the mean of their immediate surroundings, which, in the differential limit, means that the Laplacian is zero.

For the morphological case, we desire a solution in which the morphological Laplacian is zero. That is, the mean of the (circular) dilation and erosion equals the center point, rather than the mean of the entire circle. Thus, at each iteration the mean of the circular dilation and erosion replaces the center pixel value.

3.3 Multiresolution Error Encoding

The pyramidal coding process provides both compression and functionality. Compression is achieved by removing detail below a given scale or resolution and capitalizing on the underlying smoothness between neighboring pixels. Such smooth regions are efficiently represented by a small number of boundary condition pixels, which serve as seeds for the morphological smooth interpolation. The multi-resolution successive approximation within a given resolution allows the increased functionality of variable bit-rate transmission and progressive image display at the receiver.

The image estimation procedure begins at the highest pyramid level (1 pixel x 1 pixel), with an arbitrary estimated value (*e.g.* 128 for a 256-graylevel image). The error is calculated at the transmitter as the difference between the estimate and the corresponding level of the lomo pyramid. The most significant single error is then chosen for use as a boundary condition point. While this error value determines the point to be selected as a boundary condition, its error value is not directly transmitted. Instead, the quantized difference between the lomo pyramid and a reference image is used.

The reference image is chosen such that the order of points chosen does not need to be transmitted to the receiver. We choose the reference image to be the previous reconstructed pyramid level (upsampled to the current resolution). This way, the entire batch of boundary condition values corresponding to a given pyramid level can be sent to the transmitter together, rather than point-by-point, which would necessitate transmission of each point's spatial location.

Graylevel difference values are first uniformly quantized using a step size T, then entropy coded before transmission. For smooth surfaces, most of these difference values are zero and the resulting compression approaches one bit per pixel at full resolution. (Overall compression at less than one bit per pixel is possible if the selection of new boundary condition points is terminated before the full resolution is reached.) Additional compression may be possible if longer run-lengths of zero values can be correlated with small difference values. This is a matter for further investigation. For our examples, no run-length coding is used, only a single Huffman coding of all quantized difference values.

Note that while the quantization in difference values significantly improves entropy compression, it does not translate into quantization in the possible values of the reconstructed image. The smoothly interpolated solutions to Laplace's equation may take on the full range of graylevel intensities, because the isolated quantized values are superimposed upon smoothly interpolated surfaces with a continuous range of values. Thus, the adjustment of these intensities by quantized amounts before re-interpolation does not result in perceptually undesirable quantization effects. This is observed in the examples shown in the next section.

Coding continues at each scale until all errors are below the specified threshold, or until a user-defined bit limit is reached. The estimate at a completed level is then upsampled and Laplace's equation re-solved for use as the initial estimate at the next finer level. The upsampling (by a factor of two in each dimension) simply takes the coarse resolution values of the previous level and smoothly interpolates between them (by a 2x2 pixel lomo filter). Consult the summary of the coding method in Fig. 1.

3.4 Lomo Segmentation

Here we outline the multiscale lomo segmentation procedure. First, a scale-space of increasing degree of local monotonicity is generated. Each scale is generated by applying (1) with a structuring element of a given radius to the previous scale until convergence. Faster convergence may be achieved by applying the alternative lomo filter:

$$f \leftarrow \frac{(f \circ k) \bullet k + (f \bullet k) \circ k}{2} \tag{7}$$

as discussed in [1]. Because this scale-space is generated solely for accurate boundary detection and segmentation (it is not transmitted), sampling is not employed here, though could be considered for computational efficiency.

Within the lomo scale-space, the morphological Laplacian operator (4) performs edge detection at each scale. Edges are detected as zero-crossings of this Laplacian, with a user-defined threshold on the morphological gradient magnitude given by:

$$\left|\nabla\right|_{morph} f(x,y) = \frac{f(x,y) \oplus k(x,y) - f(x,y) \ominus k(x,y)}{2}. \tag{8}$$

The threshold value represents the perceptually significant graylevel difference as is used in the encoding.

This edge detection may be viewed as a morphological analogy to the linear Laplacian of Gaussian method, where a morphological locally monotonic scale-space replaces the linear Gaussian scale-space. As described in [4], gaps in the detected zero-crossing contours are closed by scale-dependent dilation (structuring element radius of $r/2$) and thinning. Between scales, the thinning process is biased towards the retention of finer-scale edges. Thus, wherever edges overlap between scales, the segment boundaries are refined as scale decreases.

The two parameters, segmentation scale and gradient threshold, are selected by the user in order to extract semantically meaningful objects. Their values, therefore, depend upon the application, but should be image independent.

4. Results and Conclusions

We present in Fig. 2 sample results for a synthetic image and for a portion of the well-known "Lena" test image, to allow easy comparison to other methods. The synthetic image contains simple geometric shapes with sharp edges and smoothly varying surfaces. The selected region of the natural image also contains non-textured objects. Both images are 128x128 pixels, with 256 graylevels. For the segmentation of the synthetic image, an initial scale corresponding to a structuring element of radius of 3 pixels is used, along with the selection of a gradient threshold of 8 graylevels. For the natural image, these parameters are 5 and 10, respectively. We show decoded results for the lomo method as well as progressive JPEG encoding at comparable compression rates. The costs of coding the segment contours themselves are not included in the comparison.

In the case of the ideal synthetic non-textured imagery, the proposed method outperforms the block-based and frequency-based JPEG method due to the precise boundaries and lack of blocking artifacts. However, the current lomo method is appropriate only for non-textured imagery, an assumption that is violated for most natural scenes. The areas that contain texture or edges not included in the segmentation, *e.g.* the lips in the Lena image, are not well reconstructed. Therefore, further compression is required in order to make the proposed method competitive with more established techniques, even for non-textured natural images. Future work will include an attempt to compress further the sparse boundary condition images, and to generalize the lomo method to include texture coding.

In summary, a novel segmentation-based image coding technique has been introduced for non-textured imagery. The object-based coding approach is unified with the theoretical basis of the segmentation itself through morphological local monotonicity. Using the morphological interpretation of Laplace's equation within a lomo pyramid, an interpolation method for smooth segments from sparse points is introduced and applied to multiresolution coding.

5. References

[1] J. Bosworth and S. T. Acton, "The Morphological Lomo Filter for Multiscale Image Processing, " *Proc. IEEE Int. Conf. on Image Processing, ICIP '99*, Kobe, Japan, Oct. 25-29, 1999.

[2] S. T. Acton and A. Restrepo (Palacios), "Locally Monotonic Models for Image and Video Processing," *Proc. IEEE Int. Conf. on Image Processing, ICIP '99*, Kobe, Japan, Oct. 25-29, 1999.

[3] J. R. Casas, P. S. Salembier, and L. Torres, "Morphological Interpolation for Texture Coding," *Proc. IEEE Int. Conf. on Image Processing, ICIP '95*, Washington, D.C., Oct., 1995.

[4] J. Bosworth and S. T. Acton "Morphological Image Segmentation by Local Monotonicity," *Proc. of the Asilomar Conference on Signals, Systems, and Computers*, Pacific Grove, California, October 24-27, 1999.

[5] P. J. Burt and E. H. Adelson, "The Laplacian Pyramid as a Compact Image Code," *IEEE Trans. on Communications*, vol. COM-31, no.4, pp. 532-540, 1983.

[6] L. M. Kells, *Elementary Differential Equations*. McGraw-Hill, New York, 1954.

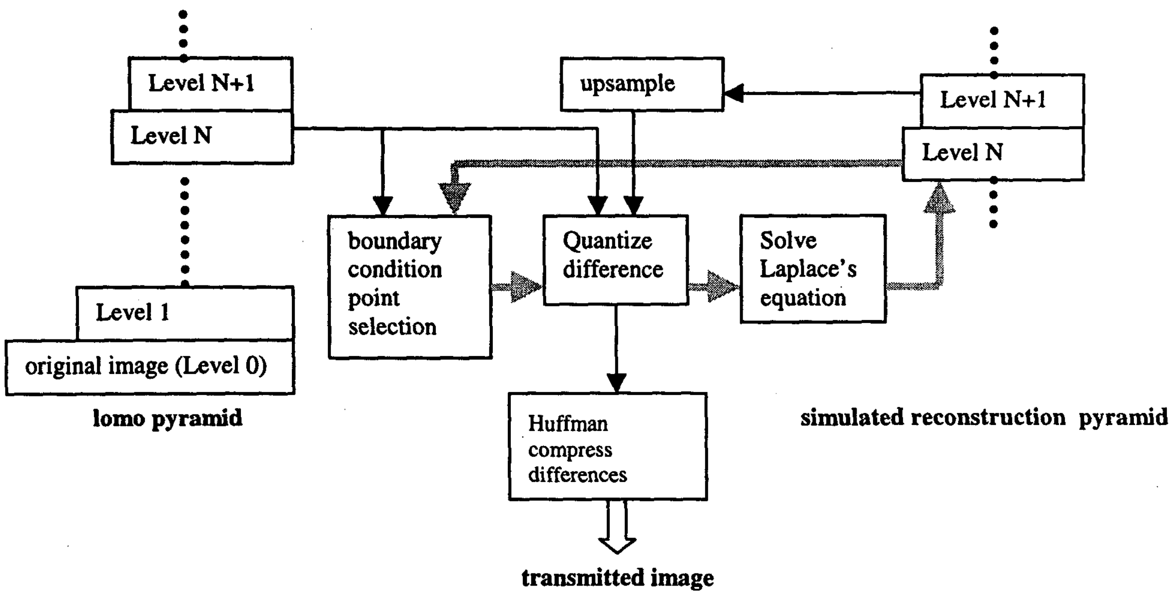

Figure 1. The operation of the encoder at a given pyramid level, consisting of the multiresolution lomo pyramid, and a copy of the reconstruction pyramid of the receiver. Though boundary condition points are determined recursively point-by-point (see the gray arrows), a given pyramid level is not transmitted to the receiver until it is complete.

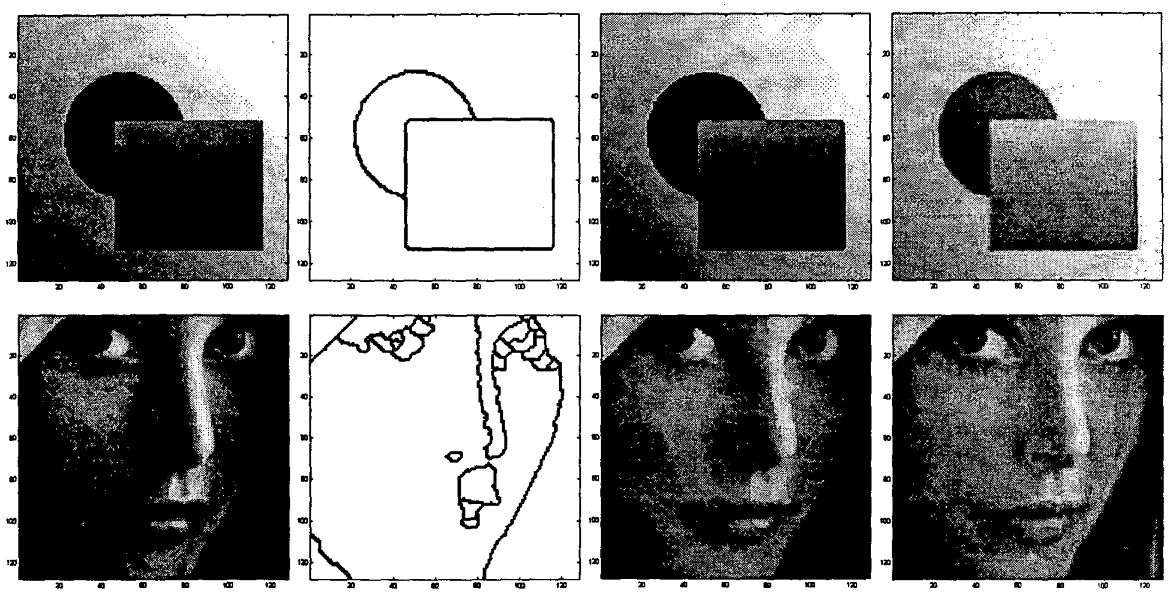

Figure 2. Results of lomo encoding compared to progressive JPEG. All images are 128x128 pixel, 256 graylevels. Top row: synthetic image, lomo segmentation, lomo encoded at 0.25 bpp (not including contour code), PSNR 37.9 dB, progressive JPEG encoded at 0.38 bpp, PSNR 33.8 dB. Bottom row: original image, lomo segmentation lomo encoded image at 0.71 bpp (not including contour code), PSNR 32.1 dB, JPEG at 0.71 bpp, PSNR 35.7 dB.