Intelligent Scan Image Progressive Transmission

Marina Sherman and Ran Ginosar
Electrical Engineering Department
Technion - Israel Institute of Technology
Haifa 32000, Israel

Abstract

Progressive transmission of high resolution still images over low speed channels enables the receiver to view an interpolated version of the image even before the complete image data have arrived. Previously proposed methods are based on sub-sampling on regular grids, icons, and linear encoding based on Gaussian/Laplacian pyramids [1, 2]. A novel progressive image transmission method is proposed, employing *Intelligent Scan* (I/S) encoding. It comprises an image encoding and fast image interpolation steps. Pixels are transmitted in the order in which they are most useful to human image recognition, rather than in the order in which they are arranged in the image file.

1. Introduction

Transmission of large digital still image files over low and medium speed global communication channels, such as the phone and ISDN networks and the internet, often takes many seconds or even minutes per image. Still image transmission is required for applications such as communications, education, video conferencing, medical, news media and publicity, real estate, and image archiving. One of the techniques for addressing this problem is progressive transmission, which is also applicable to image retrieval from data storage media such as disks.

Two common progressive transmission methods are based Raster Scan and on Icons (Directories). Whereas raster scan images are transmitted in full, icon based methods transmit small image versions (icons) and allow the user to select which images are transmitted in full. In the raster scan method, about 3/4 of the image must be transmitted before the image can be recognized in most cases. While icons allow image selection, they are often misinterpreted due to their small size, which still leads to unnecessary transmissions and wasted viewer time. Transmission of the full image, even after an icon has been selected, still requires raster scan transmission.

The *Intelligent Scan Image Progressive Transmission* is designed to overcome those difficulties by rearranging the order of pixel transmission for faster image understanding.

Raster scan successive transmission of pixels is replaced by other scan methods. The system is adaptive to natural visual properties, and applies a search method based on peripheral altering and foveal analysis [3]. People perform a vision task by continually moving their eyes to examine objects of interest, relying on peripheral vision at low resolution to detect regions of attention. Subsequently, attention is concentrated on selected regions at high resolution to examine these regions in detail. Therefore the system must provide dense image sampling in the centers of attention and sparse sampling in the periphery (similar to the organization of the retinal).

We employ a multiresolution pyramid [1], by which the original image is decomposed into a hierarchy of images. The pyramidal structure may be regarded as a model of early processing in natural vision because it provides a direct means for controlling the resolution at which data is represented for analysis and the domain over which that analysis is performed. At the same time, the nature of pyramidal structure is ideal for progressive transmission because the reduced image on the top of the pyramid can be used as Icon for initial transmission and it can be expanded progressively by adding selected information from lower layers of the pyramid.

The multiresolution pyramid proposed by Burt and Adelson [1] often employs some lossy image decomposition techniques for data compression. It can also provide lossless transmission, but at the cost of transmitting 1.33 more pixels than there are in the original image. The method proposed in this paper offers lossless progressive transmission without increasing the number of pixels that must be sent.

The paper describes the encoder, decoder and interpolation algorithms in Sections 2, 3 and 4, respectively. The proposed real-time VLSI architecture is described in Section 5, and simulation results are discussed in Section 6.

2. Image Encoder

Encoding is based on a unique pyramidal structure, as follows (see Figure 1):

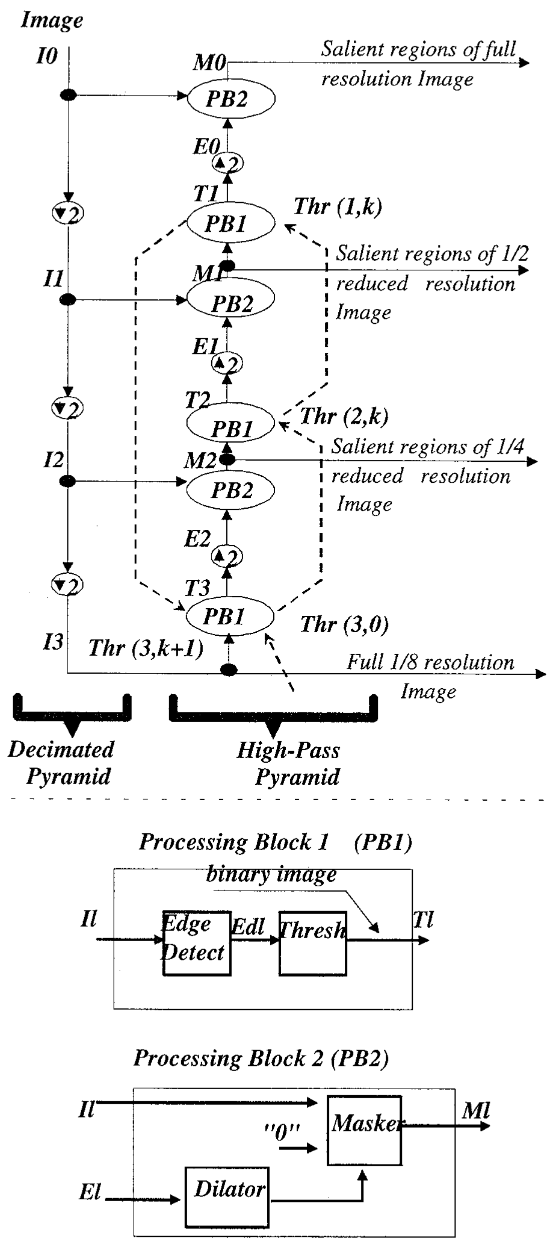

Fig.1. Intelligent Scan Encoder Algorithm.

1. A multiresolution pyramid is created, where each level of the hierarchy is a decimated version of the previous level. Note that this is not a Gaussian pyramid, as no LPF is employed (see below). The highest (lowest resolution) level is selected for initial transmission to provide the user a general understanding of the image. Typically, the initial image contains every eighth pixel of the original (namely, following three successive steps of binary decimation).
2. A second, non-linear high pass pyramid is constructed. Processing begins with the lowest resolution level. Using edge detection and threshold, regions of salient information are selected according to the a specific threshold value. An intermediate binary image is created, where selected pixels are marked by '1'. This image content-related subsampling has inspired the name Intelligent Scan for this method.
3. The intermediate binary image is expanded, using zero-order interpolation, to the image size at the previous level of the pyramid.
4. The salient regions of this binary image, which are defined by the '1's, are morphologically dilated by a tier one pixel wide. The dilated binary image is used to select the pixels of the corresponding level of the decimated image pyramid. The selected pixels are transmitted.
5. Steps 2-4 are repeated up to the highest resolution level, completing one iteration. The result of each iteration (the sub-sampled, transmitted image) is used as the data for the next iteration, thus providing successive refinement of the salient features.
6. Steps 2-5 are iterated multiple times, using successively decreasing threshold levels, until the complete image data has been transmitted.

The I/S encoding algorithm produces the transmitted pixels in sorted order of importance. In the example above, the importance is based on edge contents; other criteria are also possible. In addition an external computer vision algorithm can supply overriding control which assigns priorities to image segments (e.g. human faces vs. office background), thus modifying the sorting order and affecting even faster recognition by the receiver of the critical features of the image.

A decimated pyramid is employed, rather than a Gaussian one, for the following reasons. First, progressive transmission must yield the original pixels (i.e. lossless transmission) rather than their low-pass filtered version. Second, it must do so while transmitting only the original number of pixels, whereas most methods using Gaussian (and Laplacian) pyramids result in encoding containing more pixels than the original image. Third, aliasing leads to only transient errors in the proposed encoding: when transmission is complete, a full lossless copy of the original image has been received.

The multiple threshold values are critical to proper operation of the algorithm, affecting the contents of the various salient regions during transmission. An adaptive method has been developed, which simplifies threshold specification and optimizes the resulting image quality. Adaptive thresholding provides mechanism for control of pixel selection according to a `coarse-to-fine' principle. Threshold value variations depend on the resolution level and on the iteration steps:

$$Thr_{l,k} = (Thr_{l_{max},0} - B * k) * 2^{(4-l)}$$

where:

l	resolution level (l=3,2,1,0);
k	number of iterations;
$l_{max}=3$	lowest resolution level;
Thr	threshold value;
B	constant.

Thus, the first threshold value must be pre-defined; all successive values are computed.

3. Image Decoder

The Image Decoder (see Figure 2) consists of two parts: Coordinate Generator and Interpolator.

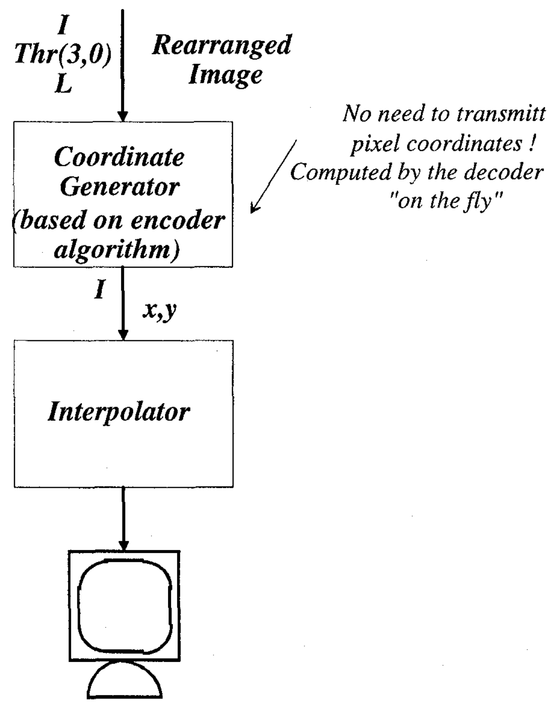

Fig.2 . Intelligent Scan Decoder Algorithm.

The encoder trathe image pixels in an irregular order. The pixels transmitted without any identifying address, so their coordinates must be generated by the decoder. Note that the unique structure of the pyramid encoding algorithm enables such coordinate generation, where the location of pixel i can be uniquely determined by processing pixels 1...i-1.

Fast image interpolation takes advantage of I/S encoding. Image interpolation from scattered, non-uniformly distributed data is more complicated than from conventional regularly distributed data (raster scan). Therefore, image interpolation and reconstruction is a critical issue of the Intelligent Scan Image Progressive Transmission

4. Interpolator

The interpolator must fulfill the following requirements:

1. *I/S interpolation* from non-uniformly distributed data points.
2. *Efficient interpolation* based on small amount of data.
3. *Adaptive Interpolation* updating only the changed regions.
4. *Efficient implementation* in hardware.

Zero-order interpolation, based on Voronoi-Dirixle tessellation [4], attaches to unknown points the value of their nearest known neighbor, as follows. Given a set of m known points ($P_1, P_2, .. Pm$), the tessellation associates the region T_i with each point P_i. The region contains the set of unknown points $\{x\}$,

$$T_n = \{x: \; d(x, P_n) \le d(x, P_k) \text{ for all } k \ne n\}$$

where d is distance between two points.

Zero-Order interpolation is fast and requires relatively simple and low cost computations, but results in ``blocky" image and jagged edges. Consequently, we employ a more complicated method developed by McLain [5]. The method includes plane triangulation and interpolation as the weighted average of three functions corresponding to the values at the triangle vertices (the known points):

$$F = w_1 * f_1 + w_2 * f_2 + w_3 * f_3$$

where:

f_1, f_2, f_3	values at the vertices of the triangle;
F	value at the interpolated point;
w_1, w_2, w_3	weights.

The weight w_i is proportional to the n'th power of the distance from the edge opposite vertex i to the

interpolated point. This ensures smooth transition from one triangle to the next, without blockiness.
Given that the distance from the interpolated point at coordinates *(x,y)* to the edge opposite vertex i is d_i

$$d_i = l_i * x + m_i + y + n_i$$

where l_i, m_i, n_i are coefficients of the triangle edges.
Then the weight is:

$$w_i = \frac{d_i^n}{(d_1^n + d_2^n + d_3^n)}$$

where:

d_i distance from interpolated point to one of the triangle sides (I=1,2,3);

n order of smoothing between triangles, n=2,3;

The challenge of this method is plane triangulation over an irregular grid. A triangulation is regarded as optimal for the purpose of interpolation if the triangles are nearly equiangular [6]. The proposed triangulation is based on non-optimal triangulation for the following reasons. First, the cost of computing optimal triangulation is prohibitive: One new point can cause, through a recursive process, triangulation changes in a very wide region. Second, the interpolation algorithm must be suitable for efficient hardware implementation. Third, comparisons of SNR results of optimal and non-optimal triangulations show only insignificant differences. Our sub-optimal triangulation is based on the fact that transmission begins with every eighth pixel, and thus triangulation is always bounded to the insides of 8×8 blocks.

A newly received pixel changes only its parent triangle (the triangle inside which it is presently placed). First triangulation (in the lowest resolution image) is always regular and well defined, since the grid at that low level is regular. In irregular triangulation, one new point always produces two new triangles: If the new point falls inside the parent triangle, it produces three child triangles (adding two triangles to the global list of triangles). If the new point falls on the boundary between two triangles, it produces two child triangles for every parent (also adding two triangles to the global list of triangles). Thus, triangulation consists of the following steps:

1. At the lowest level of resolution, perform triangulation and interpolation over a regular grid.
2. New point received: Scan 8×8 block, find parent triangle of the new point.
3. If the new point falls inside parent triangle, change triangle data list according to rules for single parent triangle.

   If the new point falls on the triangle side, find second parent triangle and change triangle data list according to rules for dual parent triangles.

4. Find all points of parent triangles, identify to which child triangle each one belongs, and re-interpolate it.

A combination of triangulation and zero-order interpolation methods is employed, where the former is used at the early transmission stages, and the latter is used at later stages. While triangulation yields better results at the early stages, the higher cost associated with it becomes unjustified at later transmission stages.

5. Architecture

VLSI architectures for both the encoder and the interpolator have been designed. The encoder and decoder coordinate generator are implemented as a simple pipeline, based on the algorithm structure.
The interpolator architecture (Fig. 3) includes a database for the progressively changing plane triangulation points, and pipeline computational elements to compute the interpolated pixels based on the triangulation, as well as a zero-order interpolator, as follows. RAM1 and RAM2 are off-chip memories. RAM1 has as many words as there are pixels in the image (2^{18} in the example). Its 12-bit word stores the intensity of every pixel (8 bits, received or interpolated) and distances (4 bits) to the nearest pixel which was actually received (for use by zero-order interpolation).

RAM2 stores the database for the triangulation. RAM2 contains 2^{15} words in this example, enabling 2^{15} triangles; experimentally we have found that after ½×2^{15} received points (2^{15} triangles out of 2^{19} possible for a 512×512 image), the difference between triangulation and zero-order interpolation is insignificant. Every RAM2 word contains intra-block 4-bit coordinates x and y of the three triangle vertices (totaling 6×4=24 bits) and a 15-bit pointer to the next triangle in the present block.

The Block Coordinate Generator contains two units: A scanner examines the 64 points inside the 8×8 block to which the new point belongs, and a controller checks a number of boundary conditions and exceptions. The on-chip RAM3 stores eight 18-bit words, each containing the l_i, m_i, and n_i coefficients (5, 5, and 7 bits respectively) for the d_i computation of all possible edges of the new triangles around a point, and a bit identifying whether the new point is inside the triangle.

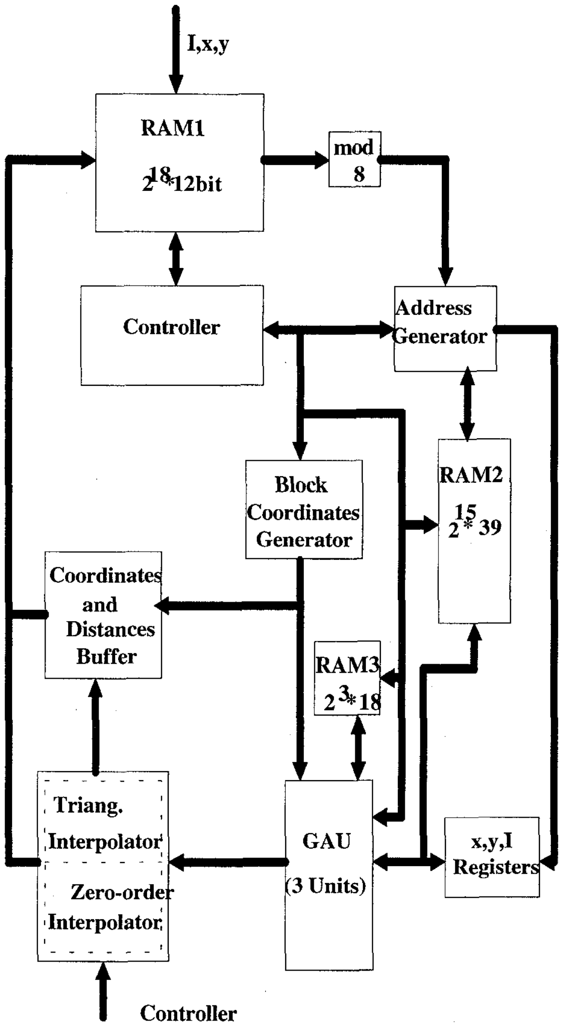

Fig.3 . Interpolator Global Architecture.

The General Arithmetic Unit (GAU) includes three arithmetic units for computing l_i, m_i, and n_i of the newly defined triangles, and for computing d_i for each interpolated point. The Triangulation Interpolator receives d_i $(i=1,2,3)$ and, having w_i stored in ROM, it computes F for each interpolated point. The Zero-Order Interpolator begins working at the later transmission stages, as described above. The Address Generator controls RAM2 operations. The Controller controls the GAU, RAM3, and RAM1. The interpolation architecture assures real-time operation at the receiver.

6. Results

The proposed architecture has been designed and simulated in VHDL. Simulations show that the triangulation algorithm requires up to 128 cycles/pixel, of which coordinate generation takes 64 cycles and 64 cycles or less are required for the remaining computations. Clock frequency is assumed at conservative 20 MHz. Thus, the proposed architecture achieves real time image interpolation for transmission channels operating at data rates as low as telephony and as high as 1.2 Mbit/second (typical LANs).

7. Conclusion

One application of the proposed method is for distribution of still video images over public telephone networks to home computers, to wireless terminals, and other targets where no high bandwidth channels are availab. It is also applicable to WAN image transmission, such as over WWW (World Wide Web) links. Conventional line-by-line (ra) transmission does not permit rapid recognition of picture contents. The proposed Intelligent Scan Image Progressive Transmission Method will alleviate this problem. Other applications include fast image retrieval from slow (low cost) mass storage.

References

[1] P. J. Burt, E.H. Adelson, ``The Laplasian Pyramid as a Compact Image Code," IEEE Transactions on Communications , vol. Com-31, no.4, April 1983, pp. 532-541.

[2] Kon-Hu Tzou, ``Progressive Image Transmission: Review and Comparison of Techniques," Society of Photo-Optical Engeneering, vol. 26(7), pp. 529-537, 1987.

[3] H. Zabrodsky, S. Peleg, ``Attentive Transmission," Journal of Visual Communications and Image Representation, Nov. 1990, vol. 1, pp. 189-198.

[4] P.J. Green, R. Sibson, ``Computing Dirichlet Tessellation in the Plane," The Computer Journal, vol. 21, no. 2, 1978, pp. 168-173.

[5] D.H. McLain, ``Two Dimensional Interpolation from Random Data," The Computer Journal, vol. 19, no. 2, 1976, pp. 178-181.

[6] R.Sibson, ``Locally Equiangular Triangulation," The Computer Journal, vol. 21, no. 3, 1978, pp. 243-245.