Surveillance Image Coding for Interpretability Using Importance Maps

A. Nguyen†, V. Chandran†, S. Sridharan† and R. Prandolini‡

†Research Concentration in Speech, Audio and Video Technology
Queensland University of Technology, GPO Box 2434, Brisbane QLD 4001, Australia
an.nguyen@qut.edu.au, v.chandran@qut.edu.au, s.sridharan@qut.edu.au

‡DSTO C3 Research Centre
Fern Hill Park, Department of Defence, Canberra ACT 2600, Australia
robert.prandolini@dsto.defence.gov.au

Abstract

An image coding technique called importance coding is developed to improve the interpretability versus bit-rate performance for an image compression system. The term interpretability is a subjective image quality measure of the content recognition performance by human observers. Importance coding aims to improve the image interpretability by optimising the amount of useful visual information encoded for a given bit rate. Traditional PSNR scalable compression schemes such as the EZW and SPIHT rely on descending wavelet coefficient magnitudes to prioritise the encoded bit-stream for rate-distortion optimisation. However, for large surveillance imagery, prioritisation of the wavelet coefficients based on the order of importance of contents in an image would be more desirable. An importance map, which provides a systematic approach for the assignment of relative importance to coefficients, is used to aid the prioritisation of the encoded bit-stream according to coefficient importance for interpretability. The importance coding framework discussed in this paper can be incorporated into the JPEG2000 international standard for image coding. Subjective evaluations indicate that importance coding is better than traditional PSNR scalable coders for very low bit-rate content recognition.

1. Introduction

The interpretability of an image to achieve maximum image content recognition is important in surveillance applications, where trained image analysts need to utilise the imagery to support decision processes in strategic, operational and tactical tasks.

Importance coding aims to optimise the interpretability of an image for a given bit rate by selecting and distributing the bit allocation to contents of importance. The quality of an importance prioritised encoded image may be severely restricted at very low bit-rates, but will preserve interpretable content that are important for human interpretation and examination.

Although, surveillance imagery are typically complex natural scenes and can be hand-marked for regions of interests based on expert analysis, the automation of such a task with reliability is beyond current technology and the direct use of image content or image interpretation is presently not possible for importance analysis. What is required is an analysis tool that generates an image map that assigns relative importance to regions of interests in an image. This image map is called an *importance map* [3].

Prandolini [10] and Nguyen et al [6, 7] have applied importance maps derived from estimates of the local fractal dimension in image regions and quad-tree dynamics to importance coding. Results have shown that importance prioritisation is preferred over traditional PSNR scalable compression schemas in subjective evaluations.

The paper will first review the notion of importance maps and how it is used for importance analysis and then discuss the framework for importance coding using a quad-tree dynamics importance map algorithm [5]. The relationship between importance coding and JPEG2000 is also outlined.

2. Image interpretability analysis

Visual examination and interpretation of surveillance imagery is based on the recognition of objects. This recognition often takes place without any conscious effort by humans. Nonetheless, there are several basic factors that can aid the examination and interpretation of surveillance imagery.

Studies in eye movements and visual perception provide insight into higher image understanding techniques that can influence the work in importance coding. Research on visual attention and eye movements [8, 14] has shown that humans generally only fixate or attend to a few regions in an image. These regions are highly correlated amongst several subjects and are determined in part by the information content of regions, with more fixations being directed to more informative regions.

Models of visual attention can be classed as multi-resolution and/or region based. Multi-resolution based models decompose a scene into spatial-frequency and/or orientation channels, whereas region based models have an image segmented into several regions or objects. These visual attention models are driven and influenced by a number of factors, which are typically classified as either top-down (task driven) or bottom-up (stimulus driven). For example, in surveillance applications, bottom-up factors may include shape, size, pattern, shadows, tone and texture [2]. Bottom-up models comprise of individual features that make up the scene. Whereas, top-down higher level cognitive factors may include context. Context associates different low-level features or objects and is important for aiding interpretation and providing an overall impression of the entire image. This is often a matter of intuition for image interpretation by humans, but not as much so for machines and computers. For a more detailed discussion on the usefulness of each factor, see [5].

A systematic approach to realise where informative regions are likely to be located is to use an importance map [3]. This map denotes the relative weights of importance to each location in an image. These importance weights are normalised in the range between 0 and 1, such that an importance score of 0 represents low importance while an importance score of 1 represents high importance. Importance maps have previously been used for visually lossless compression where improved compression has been reported with high-perceived image quality [9]. The concept of importance maps can be extended to importance coding to identify regions in an image that are of most importance and encode these first while discarding the rest of the image. The definition of what constitutes importance is thus crucial in developing such a system.

3. Quad-tree dynamics importance map

The importance map algorithm that is presented in this section is based on a quad-tree computational platform that accesses the dynamics of image regions according to a number of visual attention properties. It is conjectured that partitioning an image into regions for several visual attention factors would extract relevant information from an image that would identify regions in an image that would be of particular importance for an image recognition task.

Quad-tree partitioning employs the recursive splitting of image regions using a selected predicate such that the resulting partition can be represented by a tree structure. A simplified illustration of such a quad-tree structure is shown in Figure 1. This partitioning has been used to determine appropriate sized blocks for processing in many image processing techniques. Here the technique is generalised and an importance map is generated from the depth of the quad-tree for the various different predicates. The predicate used comprises of a rule based on some image property and are chosen based on the human visual attention processes and known importance criteria for the class of imagery. The range of values is chosen heuristically in this paper but may be obtained through a training process.

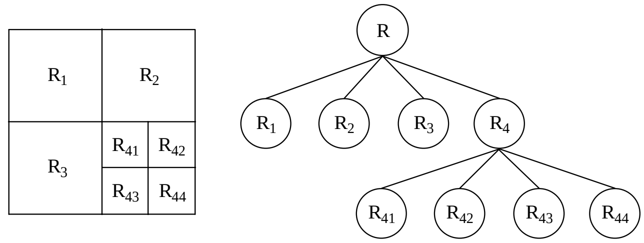


Figure 1. Partitioned image (left) and corresponding quad-tree (right). Quad-tree partitioning is determined by “feature continuity” within the region tested by a predicate. Importance of a region is proportional to the depth of the tree for the given predicate.

The importance of a given region is then assigned proportionally to the height of the quad-tree for each predicate such that a node terminated higher up in the tree would be assigned lesser importance than a node that terminated further down the tree. The overall importance map for the image is generated from a number of such predicates. The key idea is to capture feature continuity across the “field of view” as additional information to generate an importance map. A more detailed explanation of each stage follows.

Let R represent the original normalised [0 1] image region, which is to be decomposed, and P be a criterion or predicate that will be used for the quad-tree decomposition. The approach subdivides R successively into smaller and smaller quadrant regions such that for any region R_i, $P(R_i)$ = TRUE. That is, if $P(R)$ = FALSE, divide the image into quadrants. Then if P is FALSE for any quadrant, subdivide that quadrant into sub-quadrants, and so on.

The image properties incorporated that determine region importance are listed below. More details with regards to each predicate can be found in [5].

- *Contrast.* The contrast feature has been known to be a fundamental characteristic in the human visual system and affects the detection of many kinds of image features such as edges, textures and hence regions. This feature is used to assess the local contrast in each region R_i, which corresponds to the local adaptation of the human eye. The contrast predicate, P_c, is determined as follows:

$$P_c : \frac{g_{\max}(R_i) - g_{\min}(R_i)}{g_{\max}(R_i)} < C \qquad (1)$$

where $g(R_i)$ is the grey-level in region R_i, and C is the contrast threshold. Other contrast definitions such as that of Weber may be used but such definitions may not be scalable with region sizes and are more computationally complex.

- *Relative Brightness.* The brightness or luminance of regions compared to that of the background is another key visual factor that has been found to be informative and influential in visual attention studies. The brightness predicate, P_b, is given by:

$$P_b : \frac{g_{\max}(R_i)}{g_{\max}(R)} < B \qquad (2)$$

where B is a brightness threshold. Other formulations can be derived based on the application and type of scenery that is being processed.

- *Variance*: The variance enables the detection of regions with considerable energy. The predicate for variance, P_v, is calculated as follows:

$$P_v : \sigma^2(R_i) < V \qquad (3)$$

where $\sigma^2(R_i)$ is the variance of region R_i, and V is the variance threshold.

- *Edge density*: Edges are perhaps the most important of all features used by humans and play a fundamental role in image recognition and interpretation. Edges that might otherwise be hardly noticeable could be emphasised by this predicate, while the prominence of uniformly shaded areas could be decreased. The predicate for edge density, P_e, is determined as:

$$P_e : \varepsilon(R_i) < E \qquad (4)$$

where $\varepsilon(R_i)$ is the number of edge pixel in region R_i resulting from a Canny edge detection of R.

Using the four predicates formulated above, quad-tree partitioning can be performed to generate quad-tree feature maps for each of the four factors. For each quad-tree feature map, an importance weighting is then assigned to each region R_i according to the level of information or detail that each region contributed in the resultant quad-tree maps. Since quad-tree partitioning recursively splits regions further and further into narrower fields of view, the dynamics and importance for that feature grows. The importance value, I, of a region R_i, can be assigned as a function of the level of decomposition of the region:

$$I_{R_i} = \frac{1}{\log_2\left(D_{R_i}\right) + 1} \qquad (5)$$

where D_{Ri} is the dimension of region R_i. The formulation assigns pixel level regions with the highest importance value of 1, with regions of higher dimensions assigned lesser importance. This importance assignment formulation is unique, effective and computationally simple.

The overall importance map, IM, for the image can be obtained through a weighted-combination of the importance-weighted quad-tree maps.

$$IM = \sum_k w_k * (I_k)^q \quad where \quad \sum_p w_p = 1 \qquad (6)$$

where k sums through all the importance-weighted feature maps, w is the scaling factor for each feature and q is a power index to give non-linear weighting of importance values for $q>1$. Although edges may be considered the most important of all features used by humans for interpreting images, the relative importance of the different factors is unknown and may change from one image to the next and may require a training set to determine. The relative importance for each factor is treated here as being of equal importance with w set to 0.25.

The next section presents the coding aspect of the paper and will discuss how importance maps can be incorporated into an importance coding system.

4. Importance progressive coding

Progressive or embedded coding has an attractive feature that an encoded bit stream can be truncated at any point and still decode a perceptible image. Progressive coding prioritises the bit-stream according to importance, with the most “important” bits coded first.

In the default implementations of the JPEG2000 [1] and its predecessors, the embedded block coding with optimised truncation (EBCOT) [13], embedded zero-tree wavelet (EZW) [12] and set partitioning in hierarchical trees (SPIHT) [11], uses an importance prioritisation schema that is implemented for rate-distortion optimisation. The term importance referred to coefficients that yielded the least distortion according to the Mean Squared Error (MSE) distortion measure. These coders encode coefficient bit-planes from the most significant bit down relying only on wavelet coefficient magnitudes at multiple scales to determine their importance for prioritisation. Thus coefficients with larger magnitudes would be encoded first since they would contribute to the most improvement in distortion. But this strategy may not be desirable, since at very low bit-rates, the bit allocation is spent on minimising distortion and not on encoding features that are crucial for image interpretability. These objective quality measures treat all impairments equally important.

Importance coding aims to improve the interpretability of an image for a given bit rate by selecting and distributing the bit allocation to spatial-frequency transform coefficients by optimising the amount of useful visual information encoded for a given bit rate.

One highly desirable property of the JPEG2000 is that the algorithm has the functionality to implement a modified schema for coefficient prioritisation. Consequently, the most important features required for interpretability can be transmitted and reconstructed earlier.

4.1 Importance map scale-space pyramid

Given that the Mallat [4] wavelet decomposition produces horizontal, $h_l(i,j)$, vertical, $v_l(i,j)$, and diagonal, $d_l(i,j)$, coefficients at several scales, l, and the approximation coefficients, $a(i,j)$, a technique must be adopted to determine the relative importance of these coefficients across scales. A scale-space pyramid for importance maps can be achieved in three ways, namely:

1. Produce a dyadic pyramid of the input image and analyse the image at each scale to produce the importance map pyramid.
2. Analyse the image to produce the bottom-level importance map, and then produce the dyadic pyramid of importance maps by down-sampling the bottom-level map.
3. Decompose the input image using the Wavelet transform into the Mallat structure and produce importance maps for each and every sub-band. Other wavelet decomposition structures may be used.

Clearly, the first and second methods of constructing importance map scale-space pyramids require additional formulation to define the relative importance between coefficients in each directional sub-band of a Mallat wavelet decomposition. This can be formulated in several ways. A simple and efficient method can use linear weighting of coefficient values given by:

$$I_{x_l}(i,j) = I_l(i,j)\left.\frac{x_l(i,j)}{h_l(i,j)+v_l(i,j)+d_l(i,j)}\right|_{x_l=\langle h_l,v_l,d_l\rangle} \quad (7)$$

where $I(i,j)$ is the importance at each coefficient location (i,j). The importance of the approximation image, $a(i,j)$, at the lowest scale can be obtained directly from the importance map at the lowest scale (i.e. top-level scale-space importance map).

The third approach, however, would by default, have evaluated the importance for each and every coefficient in all the sub-bands of the Mallat wavelet decomposition. So equation (7) is not required.

Another approach that is a hybrid of the first and third approaches is discussed here. The method uses the direct results of a Mallat wavelet decomposition to construct the importance map scale-space pyramid. Importance maps can be obtained for each scale by applying the importance map algorithm to each $a(i,j)$ of the wavelet decomposition. This allows the adaptation of the importance map to the details present at each scale. Since an approximation image at a given scale would decompose to produce details at the next lower scale, the importance map at that scale would be used for the sets of coefficients at the next lower scale. The approach produces a more accurate representation of image content across the scales.

4.2 Prioritisation of coefficients

In EZW and SPIHT, the two compression schemes are essentially bit-plane coders with a two-pass execution through each bit-plane. In the dominant (sorting) pass, coefficients become significant if their values are greater or equal to the bit-plane thresholds. The most significant bit is encoded for the new significant coefficient. A subordinate (refinement) pass then follows after each dominant pass for the refinement of coefficients that were significant in the previous dominant passes. These coders are termed progressive PSNR optimal; although we note that a zero bit may be encoded once the coefficient has become significant, which does not improve the PSNR.

For the importance coding schemes adopted in [7, 10], the prioritisation of coefficients were similar to that used in EZW and SPIHT, but with the exception that all coefficients that were significant within a bit-plane were prioritised according to its importance as defined in equation (7).

This paper presents a similar prioritisation schema, but with the added degree of freedom of importance being the highest priority for prioritisation. The importance coder implemented prioritises the ordering of the coefficient bits according to its importance as defined in the scale-space importance map. A two-pass approach was utilised such that the dominant pass encoded the coefficient's most significant bit if the coefficient's importance fell within the highest importance band. The subordinate pass then refines coefficients according to its importance within the highest considered bit-plane as in [7, 10] and then decrements the importance band for the next dominant pass.

Not only is there flexibility for ordering coefficients in this scheme, but it can also be implemented by the JPEG2000. Blocks of wavelet coefficients are independently bit-plane coded using a three pass arithmetic coder (i.e. three coding passes per bit-plane). The coder generates a layered bit-stream organisation from which one can put any (very flexible) number of coding passes for any blocks into each layer such that each layer will incrementally improve the overall image quality for the entire image at full resolution. So to prioritise a region of an image (bounded by the location of the blocks) and the particular sub-band of frequencies (given by the sub-band the blocks are in), one can put more coding passes from that block into earlier coding layers. This makes it possible to experiment with the prioritisation coding and still implement a practical image coder. Previously, the EZW and SPIHT coders were very constrained.

5. Results

The notion and usefulness of importance coding can be illustrated by simulating results for the importance and PSNR progressive coders similar to the bit-plane coders used in EZW and SPIHT. Overhead information such as coefficient addressing and arithmetic or entropy coding has been ignored for both types of coders to illustrate the concept. It should be noted that the importance prioritisation schema described, does not need to encode the importance map for the bit-stream to be properly decoded. The flexible ordering of the bit-stream as described in the JPEG2000 standard [1] is implemented such that the overhead coefficient addressing information for both an importance prioritised and a progressive PSNR bit-plane coded bit-stream remains approximately the same.

The quad-tree dynamic importance maps were generated and examined for a class of aerial surveillance imagery and showed promising results. The technique assigns higher importance to buildings, roads, and vehicles, which are of importance in surveillance applications. A typical 1024-by-1024 grey-level surveillance image that was extracted from a much larger image is shown in Figure 2(a) with its importance map at the highest resolution given in Figure 2(b).

Figure 2(c) and 2(d) shows an importance and progressive PSNR bit-plane encoded image at 0.002 bits per pixel. Edge and outline information of man-made structures can be seen to be beginning to form in the importance prioritised image. Since the bit budget of the PSNR progressive coder is spent on minimising the distortion of the reconstructed image, salient and interpretable features that are important for interpretation are not encoded. Figure 2(e) and (f) further illustrates the improvement in interpretability for an importance prioritised coder over traditional PSNR progressive coders. The images were decoded at 0.008 bits per pixel.

The PSNR and accumulated importance versus bit-rate performance score plots for the aerial image are shown in Figure 3. The PSNR quality metric is defined as

$$PSNR = 10\log\frac{(2^n-1)^2}{MSE} \tag{8}$$

where

$$MSE = \frac{1}{MN}\sum_{i=0}^{M-1}\sum_{j=0}^{N-1}\left(x(i,j)-\hat{x}(i,j)\right)^2 \tag{9}$$

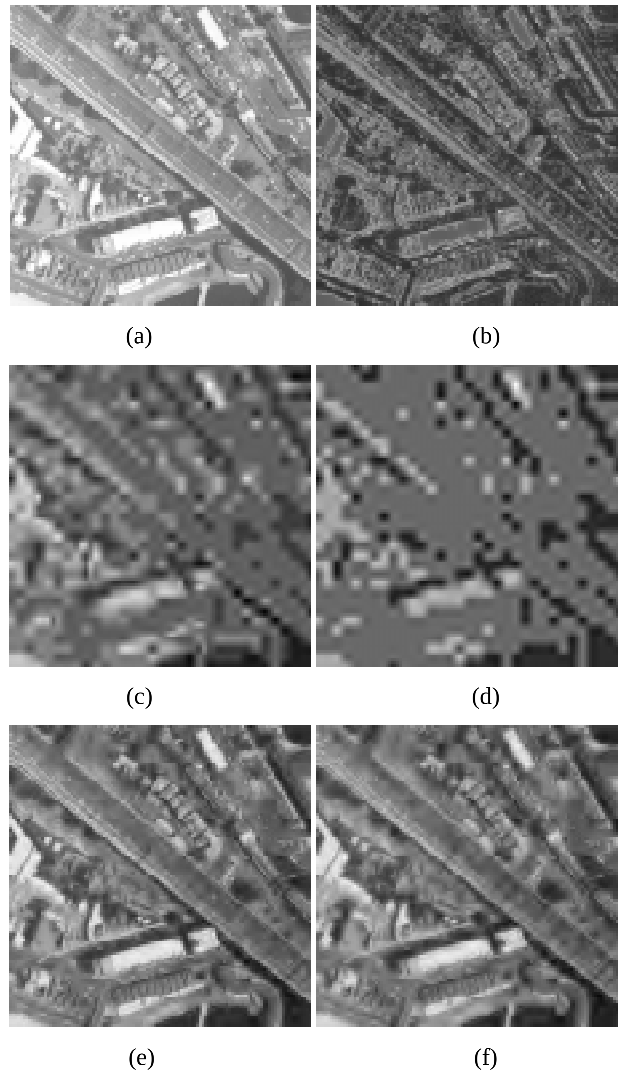
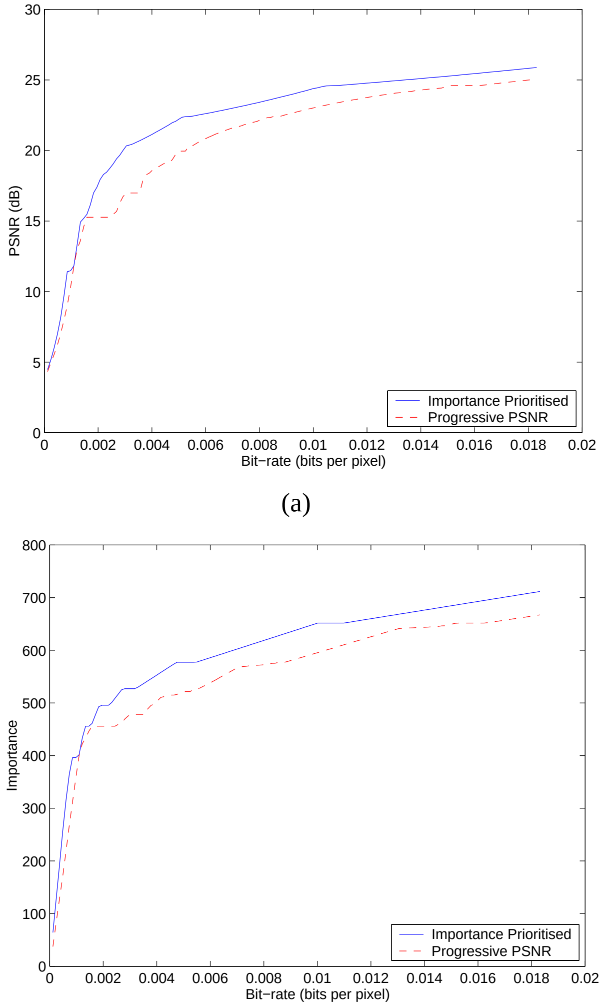


Figure 2. (a) Original 1024-by-1024 aerial image, (b) Importance map at the highest resolution (brighter regions represent higher importance), (c) Importance prioritised image (0.002 bits/pixel, PSNR = 17.94, importance = 495.67), (d) PSNR progressive image (0.002 bits/pixel, PSNR = 15.27, importance = 455.94), (e) Importance prioritised image (0.008 bits/pixel, PSNR = 23.39, importance = 617.61), and (f) PSNR progressive image (0.008 bits/pixel, PSNR = 22.08, importance = 572.13).

M, N are the height and width of the image in pixels, $x(i,j)$ is the value of the original pixel at location (i,j), $\hat{x}(i,j)$ is the value of the reconstructed pixel at location (i,j), and n is the number of bits per pixel used to represent the original image.

The accumulated importance scores were calculated from the bit-plane contributions of coefficient importance values as defined in equation (7). As each coefficient bit is encoded, the importance associated with that bit is added to the overall importance for the image, with the



(a)

(b)

Figure 3. Performance scores for the importance prioritised and PSNR progressive bit-plane coders for a 1024-by-1024 aerial surveillance image: (a) PSNR versus bit-rate, and (b) accumulated importance versus bit-rate.

residual importance corresponding to the remaining bits. The accumulated importance score correlates with what would subjectively be considered better for visual interpretation and quality, which is not the same as PSNR.

Figure 5 shows a significant improvement for the importance-prioritised coder over the progressive PSNR bit-plane coder for both the PSNR and importance scores. The two results show that we are not trading between importance and PSNR to achieve improved interpretability. It should also be noted that typical surveillance imagery are often much larger than 1024-by-1024, so allocating and prioritising a few bits at the important areas such as roads and buildings would provide even greater compression.

6. Conclusion

This paper has presented a framework for importance coding using importance maps. A quad-tree computational platform for the assignment of importance to regions in surveillance imagery using several bottom-up factors is presented. The importance map is used to select and distribute the bit allocation to spatial-frequency transform coefficients and the importance prioritised coder is shown to outperform progressive PSNR bit-plane coders. The concept of importance coding may also benefit future low bit rate video coding for wireless low bandwidth applications

7. References

[1] M. Boliek, C. Christopoulos, and E. Majani, "JPEG 2000 Part I Final Draft International Standard,", ISO/IEC JTC 1/SC 29/WG 1 N1890R, 25 September 2000.

[2] E. Falkner, *Aerial mapping: methods and applications*. Boca Raton, Lewis, 1995.

[3] A. J. Maeder, "Importance maps for adaptive information reduction in visual scenes," *ANZIIS*, Perth, WA, Australia, pp. 24-29, 1995.

[4] S. G. Mallat, "A theory for multiresolution signal decomposition: the wavelet representation," *IEEE Trans. on PAMI*, vol. 11, no. 7, pp. 674-693, 1989.

[5] A. Nguyen, V. Chandran, S. Sridharan, and R. Prandolini, "Importance assignment to regions in surveillance imagery to aid visual examination and interpretation of compressed images," *ISIMP*, Hong Kong, pp. 385-388, 2-4 May, 2001.

[6] A. Nguyen, V. Chandran, S. Sridharan, and R. Prandolini, "Importance coding of still imagery based on importance maps of visually interpretable regions," *ICIP*, Thessaloniki, Greece, 7-10 October, 2001.

[7] A. Nguyen, V. Chandran, S. Sridharan, and R. Prandolini, "Importance coding of surveillance imagery for interpretability using quadtree dynamic importance maps," *ISSPA*, Kuala Lumpur, Malaysia, 13-16 August, 2001.

[8] D. Noton and L. Stark, "Eye Movements and Visual Perception," *Scientific America*, vol. 224, pp. 34-43, 1971.

[9] W. Osberger, A. J. Maeder, and N. Bergmann, "Perceptually based quantization technique for MPEG encoding," *SPIE*, vol. 3299, pp. 148-159, 1998.

[10] R. Prandolini, "Coding of surveillance imagery for interpretability using local dimension estimates," *VCIP*, Perth, Australia, pp. 516-526, 2000.

[11] A. Said and W. A. Pearlman, "A new, fast, and efficient image codec based on set partitioning in hierarchical trees," *IEEE Trans. on CSVT*, vol. 6, no. 3, pp. 243-250, 1996.

[12] J. M. Shapiro, "Embedded image coding using zerotrees of wavelet coefficients," *IEEE Trans. on SP*, vol. 41, no. 12, pp. 3445-3462, 1993.

[13] D. Taubman, "High performance scalable image compression with EBCOT," *IEEE Trans. on IP*, vol. 9, no. 7, pp. 1158-1170, 2000.

[14] A. L. Yarbus, *Eye movements and vision*. New York, Plenum, 1967.