The Control Theory of Progressive Transmission

Computer Vision Group[1*]

[1]Department of Computer Science and Artificial Intelligence, University of Granada,
CITIC-UGR, 18071 Granada, Spain

*To whom correspondence should be addressed; E-mail: jags@decsai.ugr.es

The studies based on visual perception indicate that human observers will mostly focus on just a small but important part of the video sequence. Because of the need of resources for learning more about the optimal use of visual information in the transmission problem, here we propose a novel theory that deals with monitoring and controlling processes in progressive transmission. For example, we treat with the emergence of a region-based approach, the paradigm of prioritization and the dynamics of low-cost transmission and exploratory effort, the optimal bit saving path and the equilibrium of a transmission. We also study the relevance of knowledge from bit-saving in progressive transmission.

Introduction

Very low bit rate coding methods are basically from two different classes. The first includes block-based techniques such as H.263 and MPEG-2 (*1, 2*) which are relatively easy to implement and give fair image quality at low bit rates. But artifacts become visible and the visual quality is bad at very low bit rates. The second class includes the second-generation video-

1

coding methods (*3*), such as segmentation-based coding (*4*) as well as object-based coding (*5,6*). They use object regions instead of rectangular blocks for efficient compression, but the quality and bit rate vary dramatically depending on the approach and heavily depend on the quality of object segmentation and the representation of a region. In most of the object-based coding techniques, segmentation and motion estimation techniques are computationally very expensive and the accurate representation of the shape of moving objects is the necessary condition to achieve the goal of good compression result (*7*). In contrast, block-based video coding such as H.263 (*8*) does not suffer from these problems but it does not support any of the content-based functionalities which are required by the MPEG-4 (*9*).

The studies based on visual perception indicate that human observers will mostly focus on just a small but important part of the video sequence. It is so proposed as the attention model (*10*). Thus, visual information with more attention-attraction potential will be more sensitive to coding error, and should be presented at a higher quality to get a better overall quality. But human visual system is extremely complex, and many of its properties are not well understood even today. Humans tend to use high-level concepts in everyday life, however what computer techniques often extract are mostly low-level features.

Because of the need of resources for learning more about the optimal use of visual information in the transmission problem, we propose a novel theory that deals with monitoring and controlling processes in progressive transmission. In the development of this control theory of progressive transmission, we made repeated use of Pontryagin's Maximum Principle, (*11*). To this aim, optimization problems need to be put into the form of a Bolzano problem which involves a number of state variables which can change over time. Following this approach, we study the emergence of a region-based approach, the paradigm of prioritization and the dynamics of low-cost transmission and exploratory effort, the optimal bit saving path, and finally the relevance of knowledge from bit-saving in progressive transmission.

2

The Optimal Use of Visual Information

In the absence of a priori knowledge about regions of interest, does dysfunctional behavior in the subband pyramid lead to the emergence of a region-based approach to image transmission as complexity grows? A model of image transmission which links simplification of information and time to produce a transmission plan to the bit-rate cost of the transmission, and where the decoded output is affected negatively by those factors that reduce the quality of the transmission plan and delay its preparation, will answer the question in two steps (*12*).

First we need to characterize the variables (degrees of simplification and subordination) through which we expect to control, at any truncation time, the optimum transmission system. In this case we may observe that transform-domain coefficients exhibit self-seeking behavior if errors which result from the transmission plan increase as the distance of the coefficient from the top of the hierarchy grows.

Second we use the degree of detail which arrives at the central control in any system of progressive transmission to derive conditions under which the likelihood of emergence of a region-based approach to progressive transmission grows with the increase of complexity of the picture. We find that the presence of self-seeking coefficients leads to a higher likelihood of emergence of a region-based approach as complexity rises.

Another interesting problem is the prioritization of several deposits of visual information which are equally accessible within the image transform but differ in relevance for humans. The paradigm of prioritization is to select the highest-importance deposit of visual information to be transmitted first. From the study of the optimal prioritization profile of a progressive image transmission with several deposits of visual information, it follows that there exist circumstances in which it is optimal to transmit from relatively low-importance deposits of visual information before turning to relatively high-importance deposits of image information. That is,

we will provide a set of sufficient conditions under which optimality does not imply the transmission of deposits in strict sequence of importance, beginning with the highest-importance deposit.

If we assume that the rent component of profit on the optimal transmission path (i.e., profit minus transmission cost) would steadily decrease, the system should prefer to transmit a unit of visual information at this time over conserving it for the future. But if the rent component shifts from declining to rising over time, heavy losses may be incurred in the form of foregone transmission opportunities in the future. In particular, (*13*) shows the difficulty of predicting the rent component of profit, using a model of bit depletion from control theory. The transmission system must simultaneously determine transmission rates that balance total profit from bit allocation with transmission costs. The objective is maximize the present value of the satisfaction derived from progressive transmission over the future, subject to constraints imposed by the initial conditions and the prioritization and transmission technologies. The analytical results demonstrate an oscillatory behaviour over time of the rent component of profit depending on the incremental cost of cumulative transmission and not the current marginal transmission cost.

The oscillatory behaviour of rent over time necessitates devising a stabilizing scheme for its optimal use in the transmission, since heavy losses may be incurred in the form of foregone transmission opportunities in the future if the transmission system prefers to transmit data at this time over conserving (at least a part of them) for the future. It calls for a saving rate and an accumulation of low-cost bits at the present time than would be unnecessary in the case of a steadily declining rent component of profit. The system would then modify the transmission condition accordingly to render it dynamically efficient over time.

But the exhaustibility of bit streams with low transmission cost (using some coding strategy) could also restrict the efficiency at low bit rates. Hence there is a concern for this problem and understanding its implications is essential to the attainment of efficiency in intertemporal mod-

els of progressive transmission. In fact the issue of depletion in low-cost bit reserves forces us to think about bit allocation over time: The more we allocate low-cost bit reserves for transmission at this time, the less will be available at higher bit rates.

Using Control Theory it follows that there are at least three forces offsetting the limitations imposed by exhaustible (low-cost) bit reserves in progressive transmission, (*14*): Optimal exploratory activity to build the knowledge base up to a level that reduces encoding costs; optimal bit saving path to render the transmission condition dynamically efficient over time; and sustained improvement in transmission technology.

Bit streams can be prioritized in response to the use of a prioritization protocol, but only a part of the prioritized low-cost bitstreams is to be transmitted to the decoder at this time. The level of transmission technology at any time may also be augmented by investing part of the gross income in transmission improvement. The rest of prioritized bitstreams is to be added to a knowledge base from bit-saving, which increases the knowledge of the transmission system about the significance of this particular information which was not sent to the decoder at the time of its prioritization.

Thus, the knowledge base from bit-saving allows the transmission system to distinguish between the prioritized bitstreams not sent to the decoder at the time of their prioritization, and the rest of image information which was not prioritized, yet. It is a bit saving just reducing the amount of transmitted bits at this present time, but additions to the knowledge base may be transmitted later following an improved transmission technology at higher bit rates.

Another control problem is to choose bit-allocation, bit-reserves use and investment in transmission improvement over time so as to maximize the discounted utility of low-cost transmission. It uncovers the exact role and relevance of knowledge from bit-saving in the optimal transmission path (*15*). For the postulated model we have found that along the optimal transmission path the rate of change of the rate of return on bit-reserves use equals the rate of return

on knowledge from bit-saving.

Also, we have found that the discounted instantaneous marginal utility of bit-allocation changes at a rate equal to the rate of the return on knowledge from bit-saving minus the discount rate (which will be introduced in the analysis to be able to compute the discounted utility of transmission at higher bit rates).

We have also determined the rule for optimal investment in transmission technology (over time): Along the optimal transmission path, investment in technology should be carried out up to the point where the growth rate in the marginal rate of accumulation of technology equals the difference between the marginal product due to an extra unit of investment and the marginal product of knowledge from bit-saving.

Experimental Results

In order to test the practical relevance of this control theory several experiments are performed as follows.

The H-REVIC video coder, (*16*), is an example of application of the control theory for progressive transmission of video sequences. The software of the video codec may be accessed in `http://decsai.ugr.es/cvg/revic`.

The 3D spatio-temporal orientation trees coupled with powerful SPIHT sorting and refinement, (*17*), renders 3D-SPIHT video coder, (*18*), so efficient that it provides performance superior to that of MPEG-2 and comparable to that of H.263 (which are generally used nowadays) with minimal system complexity. That is, 3D-SPIHT is one wavelet-based coder that exhibits state-of-the-art compression performance.

The comparative performance of the 3D-SPIHT and the proposed coder H-REVIC without motion filtering, is here tested on a set of real-world videos. In both cases the embedded coding was performed without motion compensated filtering, since the transmission of parameters is

6

critical at very low bit rates (if motion compensated temporal filtering would be applied we will need to code motion vectors).

The compound gain (CG) between a test image I and decoded outcome O is a generalization of the Kullback-Leibler joint information gain of various random variables such that, it satisfies a number of natural properties as given in (*19*). The main result of the study in Reference (*20*) is the finding that a generalization of the Kullback-Leibler joint information gain of various random variables, called "compound gain", appears to relate to visual distinctness as perceived by human observers. Given any coding scheme the CG may then be applied to quantify the visual distinctness by means of the difference between the original image I and decoded images at various bit rates. It allows us to analyze the behavior of coders from the viewpoint of the visual distinctness of their decoded outputs, taking into account that an optimal coder in this sense tends to produce the lowest value of the CG. The software and documentation of the compound gain may be accessed in the Internet site with URL of `http://decsai.ugr.es/cvg/CG` or by anonymous ftp to `decsai.ugr.es` with the path `pub/cvg/software` in the compressed tar file `cg.tar.gz`.

Figs. 1 and 2, are given to illustrate the objective quality of H-REVIC and SPIHT. These figures give the comparative performance of the two methods using the Compound Gain (CG) as distortion measure. As can be seen from these figures, the CG distortion values predict a better visual fidelity using H-REVIC than with the SPIHT reconstructions.

We have developed an algorithm, called as the SAVING coder, (*15*), which permits us to implement a bit saving path following the control theory of bit saving. It is based on the utility-per-coding-bit optimization (*21*). The software and documentation may be accessed in the Internet site with URL of `http://decsai.ugr.es/cvg/REWIC`.

Here we perform a comparison in objective performance of the state of the art codec in progressive transmission, i.e., the set partitioning in hierarchical trees (SPIHT), (*17*), against
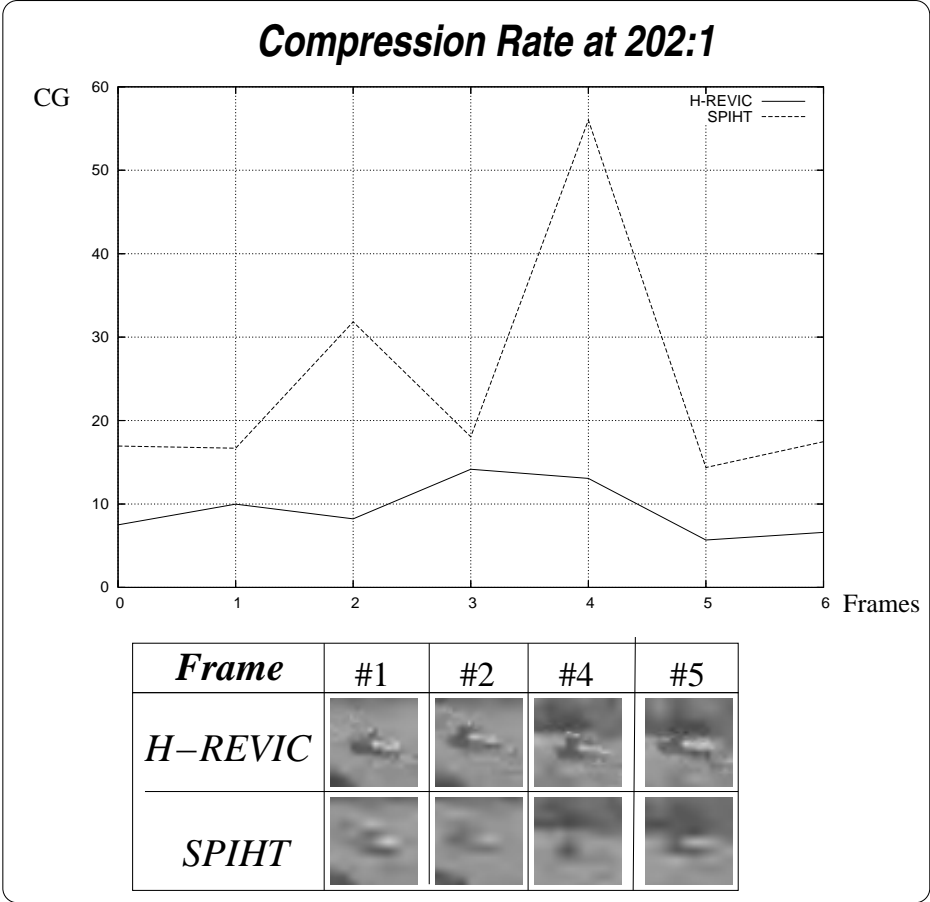
7

Figure 1: Objective quality of H-REVIC and SPIHT using the Compound Gain (CG) as distortion measure, for a sequence of moving targets.
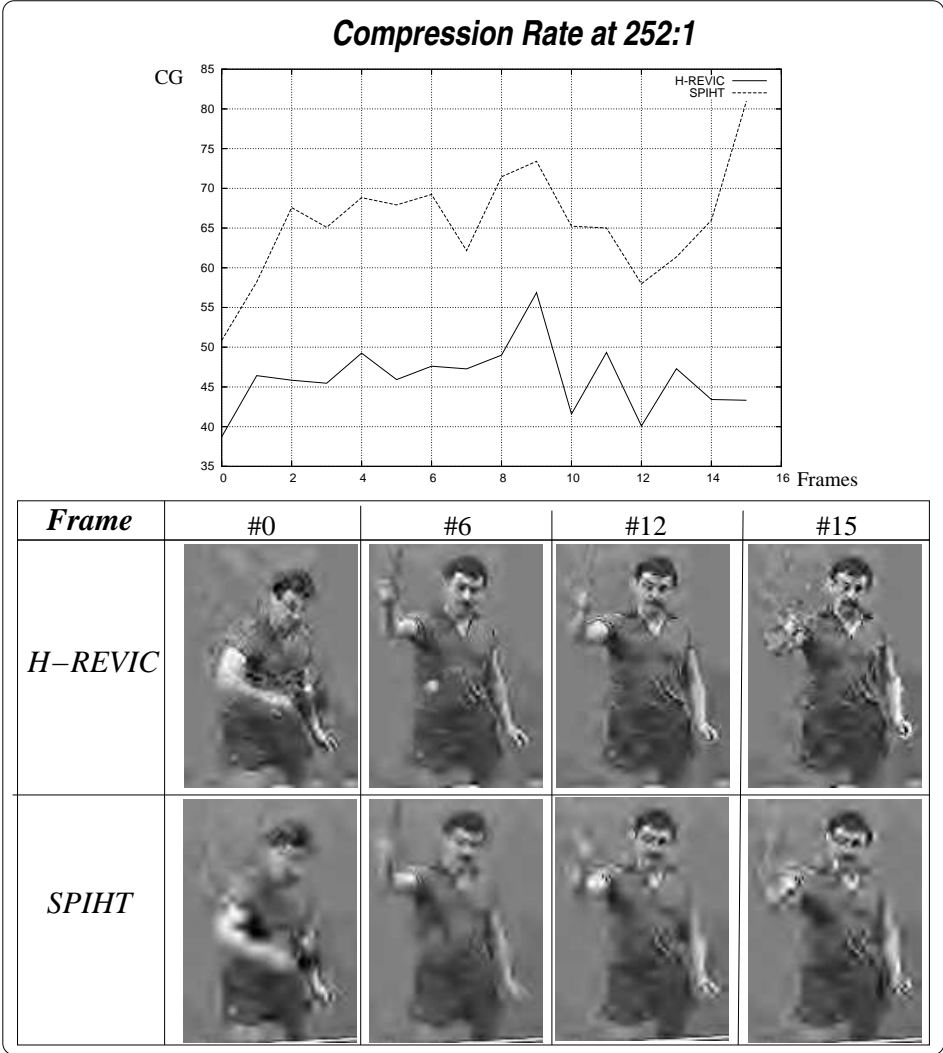
Figure 2: Objective quality of H-REVIC and SPIHT for Tennis sequence.

9

the SAVING coder using an objective procedure for coder selection, given in (*20*). The test dataset is composed of twenty 512×512 grayscale images (Fig. 3).

Given a test image I, let $\{I^{spiht}_{q(1)}, \cdots, I^{spiht}_{q(K)}\}$ be the set of SPIHT decoded outputs at bit rates $q(1), \cdots, q(K)$; $\{I^{SAVING}_{q(1)}, \cdots, I^{SAVING}_{q(K)}\}$ be the set of SAVING reconstructions at the same bit rates $q(1), \cdots, q(K)$. The compound gain CG may then be applied to quantify the visual distinctness by means of the difference between the original image I and reconstructions at bit rates $q(i)$:

$$f(\text{SPIHT}, i) = CG(I, I^{spiht}_{q(i)}). \tag{1}$$

and similarly, $f(\text{SAVING}, i)$. It allows us to analyze the behavior of both coders across bit rates $q(i)$. The optimal coder achieves the lowest value of $f(\dagger, i)$, for all $i = 1, 2, \cdots, K$, over lossy coders in the class being compared (e.g., SAVING and SPIHT).

Once distortion functions $f(\dagger, i)$ have been calculated following equation (1), we make use of an objective criterion for coder selection based on the overall difference between the two functions $f(\text{SPIHT}, i)$ and $f(\text{SAVING}, i)$, which can be measured by a Kolmogorov-Smirnov (K-S) test to a certain required level of significance as follows:

Coder Selection Procedure: SAVING vs. SPIHT. *In the language of statistical hypothesis testing, the coding scheme SAVING is significantly better than SPIHT for test image I if the following two conditions are true:*

(1) $f(SAVING, i) \leq f(SPIHT, i)$, for the 90 per cent of the bit rates $q(i)$, with $i = 1, 2, \cdots, K$; and

(2) we disprove, to a certain required level of significance, the null hypothesis of a Kolmogorov-Smirnov test that two data sets $\{f(SAVING, i) \mid i = 1, 2, \cdots, K\}$ and $\{f(SPIHT, i) \mid i = 1, 2, \cdots, K\}$ are drawn from the same population distribution function.

10

Figure 3: Dataset of 512×512 images

Condition 1 takes into account that an optimal coder tends to produce the lowest value of $f(\dagger, i)$ across bit rates, and disproving the null hypothesis in condition 2 in effect proves data sets $\{f(\text{SAVING}, i) \mid i = 1, 2, \cdots, K\}$ and $\{f(\text{SPIHT}, i) \mid i = 1, 2, \cdots, K\}$ are from different distributions.

For the dataset of 512×512 images (Fig. 3), eight out of twenty test images (40 %) have passed conditions (1) and (2) in the coder selection procedure (SAVING versus SPIHT): Test images #7, #9, #10, #11, #12, #16, #19, and #20 (see Fig. 1). Hence, SAVING is significantly better than SPIHT with high confidence level for forty per cent of the dataset of 512×512 test images.

By the contrary SPIHT is significantly better than SAVING with high confidence level for five per cent of the test dataset. One out of twenty test images has passed the two respective

11

conditions in the coder selection procedure (SPIHT versus SAVING): Test image #17 (Fig. 3).

Fig. 4 illustrates the subjective quality of the SAVING coder and the SPIHT coder. This figure shows that SPIHT reconstruction exhibits bad visual fidelity at 80:1 compression ratio since noise data was incorrectly prioritized before significant visual information from the mammographic area. But the SAVING output is fair at 80:1 since granular microcalcifications are clearly detected in the mammographic section.

References and Notes

1. R.J. Clarks, *Digital Compression of Still Images and Video,* San Diego, CA: Academic, 1995.

2. *MPEG - Video Simulation Model Three,* ISO/IECJTC/SC2/WG11, MPEG90/041. (1990)

3. M. Kunt, A. Ikonomopoulus, and M. Kochner, "Second-generation image coding techniques", Proc. IEEE, Vol. 73, pp. 459-574, (1985).

4. T. Ebrahimi, E. Reusens, and W. Li, "New Trends in Very Low Bitrate Video Coding", Proc. IEEE, Vol. 83, pp. 877-891, (1995).

5. B. Talluri, K. Oehler, T. Bannon, J.D. Curtney, A. Das, and J. Liao, "A robust, scalable, object-based video compression technique for very low bit-rate coding", IEEE Trans. Circuits Syst. Video Technol., Vol. 7, No. 1, pp. 221-233, (1997).

6. D.S. Cho, and R.H. Park, "Object-based Very Low Bitrate Using Motion Parameter Estimation Based on Multiple Frame Prediction", J. Visual commun. Image Repres., Vol. 10, pp. 291-305, (1999).
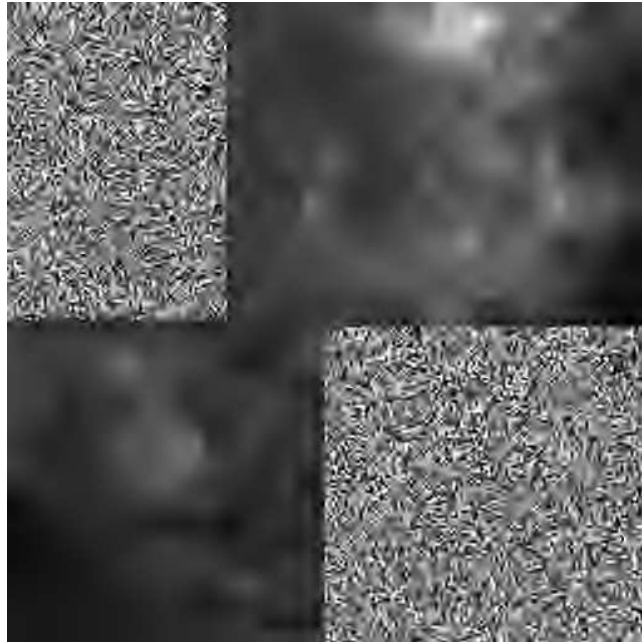
SPIHT at 80:1
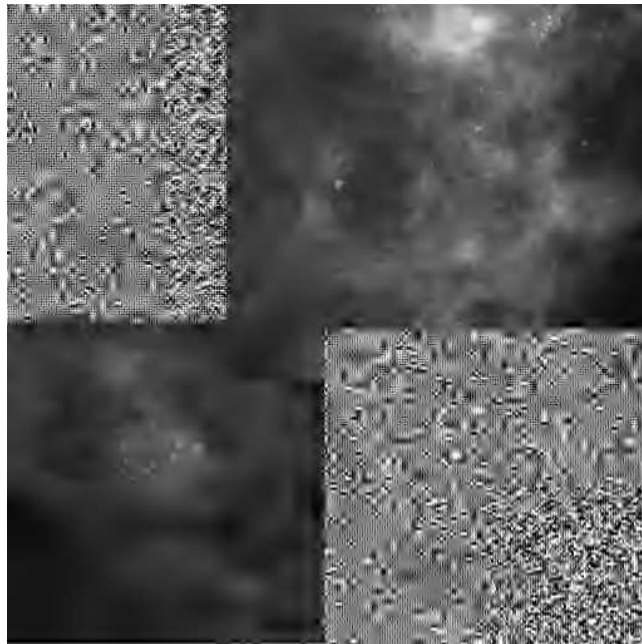
SAVING at 80:1

Figure 4: Subjective comparisons of decoded outcomes at $80 : 1$ using the state of the art codec in progressive transmission, i.e., the set partitioning in hierarchical trees (SPIHT), against the SAVING coder

7. Shyi-Chyi Cheng, "Visual Pattern Matching in Motion Estimation for Object-Based Very Low Bitrate Coding Using Moment Preserving Edge Detection", IEEE Trans. on Multimedia, Vol. 7, No. 2, (2005).

8. I. Kompatsiaris, and M. G. Strintzis, "Video Coding for Low Bitrate Communication Draft Recommendation H.263", ITU-T, Study Group 15, (1995).

9. *MPEG-4 Video Verification Model Version 16.0,* Video Group ISO/IECJTCI/SC29/WG11/N3312, (2000).

10. L. Itti, and C. Koch, "A Comparison of Feature Combination Strategies for Saliency-Based Visual Attention Systems", SPIE Human Vision and Electronic Imageing IV, San Jose, CA, Vol. 3644, pp. 373-382, (1999).

11. L.S. Pontryagin, V.G. Boltayanskii, R.V. Gamkrelidze, E.F. Mishchenko, "The Mathematical Theory of Optimal Processes", Wiley (1962) (Translated from Russian).

12. J.A. Garcia, Rosa Rodriguez-Sanchez, J. Fdez-Valdivia, "Emergence of a region-based approach to image transmission", *Optical Engineering*, Vol. 44, 067004, (2005).

13. J.A. Garcia, Rosa Rodriguez-Sanchez, J. Fdez-Valdivia, "Dynamics of low-cost transmission on the optimal path", *Optical Engineering*, Vol. 46 (3), DOI: 10.1117/1.2717132, (2007).

14. J.A. Garcia, Rosa Rodriguez-Sanchez, J. Fdez-Valdivia, "Optimal Exploratory Effort to Build Knowledge for Video Transmission", *Optical Engineering*, Vol. 46 (4), 047401, (2007).

15. J.A. Garcia, Rosa Rodriguez-Sanchez, J. Fdez-Valdivia, "Relevance of knowledge from bit-saving in progressive transmission", *Journal of Visual Communication and Image Representation*, Vol. 21(7), pp: 741-750, (2010).

16. Rosa Rodriguez-Sanchez, J.A. Garcia, J. Fdez-Valdivia, Antonio Garrido, "Automatic and optimal hierarchical quantizer decomposition to build knowledge for video transmission", *Optical Engineering*, Vol. 46 (10), doi:10.1117/1.2799092, (2007).

17. A. Said, and W.A. Pearlman, "A new, fast and efficient image codec based on set partitioning in hierarchical trees," *IEEE Trans. on Circuit and System for Video Technology*, **6**(3), 243-250, (1996).

18. B.-J., Kim, Z. Xiong, and W.A. Pearlman, "Low bit-rate scalable video coding with 3D set partitioning in hierarchical trees (3D-SPIHT)", *IEEE Trans. Circuits and Systems for Video Technology*, Vol. 10, pp. 1374-1387, December 2000.

19. J.A. Garcia, J. Fdez-Valdivia, Xose R. Fdez-Vidal, and Rosa Rodriguez-Sanchez, "Information Theoretic Measure for Visual Target Distinctness," *IEEE Trans. on Pattern Analysis and Machine Intelligence,* **23**(4), 362-383, (2001).

20. J.A. Garcia, J. Fdez-Valdivia, X.R. Fdez-Vidal, and R. Rodriguez-Sanchez, *Computational models for predicting visual target distinctness*, SPIE Optical Engineering Press, PM-95, Bellingham, Washington USA, (2001).

21. J.A. Garcia, R. Rodriguez-Sanchez, and J. Fdez-Valdivia, *Progressive image transmission: The role of rationality, cooperation and justice*, SPIE Optical Engineering Press, PM-140, Bellingham, Washington USA, (2004).