

## Research Article

# Multiple Description Coding Using Data Hiding and Regions of Interest for Broadcasting Applications

M. Verza,<sup>1</sup> N. Conci,<sup>2</sup> G. Boato,<sup>1</sup> and F. G. B. De Natale<sup>1</sup>

<sup>1</sup>Department of Information Engineering and Computer Science(DISI), University of Trento, Via Sommarive 14, 38100 Trento, Italy

<sup>2</sup>Multimedia and Vision Research Group, School of Electronic Engineering and Computer Science, Queen Mary, University of London, London E1 4NS, UK

Correspondence should be addressed to N. Conci, nicola.conci@elec.qmul.ac.uk

Received 27 February 2009; Accepted 1 June 2009

Recommended by Maurizio Murrioni

We propose an innovative scheme for multiple description coding (MDC) with regions of interest (ROI) support to be adopted in high-quality television. The scheme proposes to split the stream into two separate descriptors and to preserve the quality of the region of interest, even in case one descriptor is completely lost. The residual part of the frame (the background) is instead modeled through a checkerboard pattern, alternating the strength of the quantization. The decoder is provided with the necessary side-information to reconstruct the frame properly, namely, the ROI parameters and location, via a suitable data hiding procedure. Using data hiding, reconstruction parameters are embedded in the transform coefficients, thus allowing an improvement in PSNR of the single descriptions at the cost of a negligible overhead. To demonstrate its effectiveness, the algorithm has been implemented in two different scenarios, using the reference H.264/AVC codec and an MJPEG framework to evaluate the performance in absence of motion-compensated frames on 720p video sequences.

Copyright © 2009 M. Verza et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

## 1. Introduction

The widespread of digital TV and high-quality broadcasting requires the definition of innovative schemes to guarantee improved robustness and error resilience to video streams. The robustness requirement is particularly necessary in the areas where the digital signal is not strong enough and frequent errors that damage slices or even entire frames are common. In fact, one of the most significant drawbacks of current technologies for terrestrial video broadcasting is that if the signal-to-noise ratio drops under a certain threshold, the program is obscured until the quality of the signal returns back to the standard operational range. This is actually one of the main concerns that arose with Digital Video Broadcasting (DVB), and it is one of the pending issues that need to be addressed. With this respect, scalable coding can be regarded as a potential solution to the problem, allowing the split of the stream into a number of substreams. The substreams can provide reduced quality at lower bit rate, and the full quality can be achieved by suitably combining all layers. In such a framework, either layered coding or MDC could be adopted as a viable approach [1].

MDC [2] has emerged as a promising approach and a valid alternative to more common layered coding standards, to improve error resilience in a video. Starting from a video source, the core idea consists of splitting the original stream in a predefined number of substreams, called descriptors. Descriptors are typically balanced, meaning that each of them has comparable bit-rates and provides similar quality when individually decoded. This results in a big advantage if compared to layered coding (from MPEG-2 to H.264/SVC), where the hierarchical organization of the stream strongly relies on the correct decoding of the base layer. In MDC, instead, each descriptor can be independently decoded at low but acceptable quality. The higher the number of received descriptors, the better the quality. If all substreams are received and properly merged, the original quality can be guaranteed.

In our work, we propose to spatially differentiate the quality of each description, defining a suitable region of interest (ROI) where the quantization strength is reduced in order to improve the perceived quality within the desired area. Since this information (location, dimension, quality factor, etc.) must be shared with the decoder to correctly

display the visual content, we exploit a watermarking algorithm, which allows improving the PSNR of each single description without significantly affecting the resulting overhead. The algorithm is tested on both MJPEG and H.264/AVC. The choice of performing the validation on two different coding schemes is motivated by the recent developments in HDTV broadcasting research. In this area, bandwidth is not a critical issue anymore as it used to be, and, therefore, the transmission of static images instead of motion compensated videos could provide better perceptual quality, and higher robustness to error propagation. The validation phase has been carried out taking into account standard metrics like PSNR, together with VQM, in order to determine the perceptual impact of the ROI and the watermarking scheme.

The paper is structured as follows. Section 2 reports a review of the most relevant works in the field of MDC. In the same section, also the adoption of data hiding tools for nonconventional applications is discussed. The proposed MDC scheme is described in detail in Section 3, while ROI embedding is presented in Section 4. Experimental results are reported in Section 5, and Section 6 presents some concluding remarks.

## 2. Related Work

*2.1. Multiple Description Coding.* The generation of multiple streams imposes the introduction of a variable amount of overhead (also referred to as redundancy  $\rho \in [0, 1]$ ), which is related to the number of descriptors and to the requirements of robustness of the specific application. Given  $N$  descriptors, the overhead is calculated in general as in (1), where  $\text{size}_{\text{SDC}}$  refers to the video encoding in single descriptor mode, traditionally:

$$\rho = \frac{\sum_{n=1}^N \text{size}(n) - \text{size}_{\text{SDC}}}{\text{size}_{\text{SDC}}}. \quad (1)$$

The overhead is caused by the need of storing in each descriptor some auxiliary information like

- (i) headers, trailers, synchronization markers, and presentation data that must be replicated in order to make each single stream visible;
- (ii) side information to facilitate the reconstruction in presence of losses.

One of the first effective applications of MDC to image coding that is still a reference method was proposed by Vaishampayan in [3]. His idea of a Multiple Description Scalar Quantizer was demonstrated, for the sake of generality, with application to a memoryless Gaussian source; however, it turns out to be effective also in image and video coding applications.

In literature we find a large amount of algorithms based on MDC, and one of the expected results has always been the ability to deploy effective implementations capable of maximizing quality in the single descriptor reconstruction while minimizing overhead. This can be achieved through

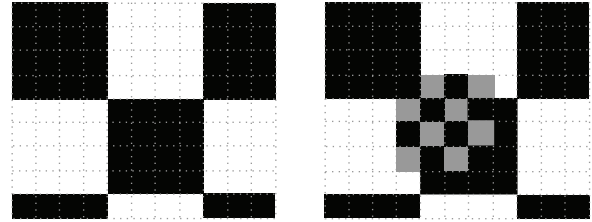


FIGURE 1: Macroblock grid without (left) and with the ROI (right).

different techniques. The most common approaches are on the one hand related to the improvement of the single descriptor quality, by sharing relevant data among descriptors, and on the other hand focused on the maximization of the number of substreams to provide diversity. In fact, the intrinsic nature of MDC does not impose any limitation in the number of descriptors to be considered and it has been demonstrated that this goal can be achieved by applying, for example, wavelet-based approaches. There are several implementations dealing with these problems as it can be seen in [4–6] where the different strategies have been tested considering up to sixteen descriptors. The methods proposed in these articles mostly deal with static images, since wavelet-based tools for video coding have not received great attention yet. However, the implementation of a high number of descriptors introduces often a non-negligible redundancy.

If not so, the achieved quality of the single descriptor reconstruction is very poor and the streaming provider usually prefers other alternatives such as FEC as discussed in [7]. In addition, a large number of descriptors would introduce a significant increase in the complexity both at the encoder and at the decoder, making it almost impossible (with the current technologies) to perform live video transmission. Therefore, the most likely number of descriptors to deal with can be limited to a few units, according to the channel capacity and the network setup, even though most of the works in literature adopt two descriptors only.

Even though several MDC techniques have been proposed in the video framework in both the form of MJPEG contents and motion-compensated coding standards, still little research has been conducted on the combination of MDC and H.264. In [8] the authors focus on the implementation of a coding scheme to be implemented in a MIMO architecture. In [9] an innovative standard compliant approach is presented: descriptors are composed by alternating primary and redundant slices in a complementary manner, so that losses can be mitigated by replacing each missing portion of data with its redundant slice from the complementary descriptor. More recently, another MDC system was proposed in [10]. Here, the smoothness and edge features of DCT blocks are modeled in such a way that their perceptual tolerance against visual distortion is measured; the key components identified via this method are duplicated, while the remaining ones are effectively split among descriptors.

*2.2. Data Hiding.* As far as data hiding is concerned, we have applied the concept in a slightly different manner with

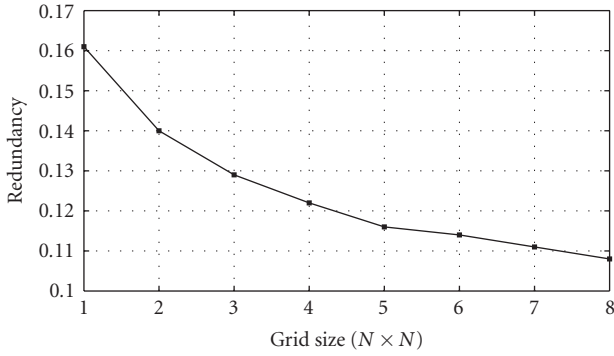


FIGURE 2: Resulting overhead by varying the gridsize in the H.264/AVC configuration.

respect to the standard applications. Data hiding schemes are usually applied to digital rights management [11]. Such methods invisibly embed a watermark into a cover data (a portion of the image in which the mark is inserted), allowing the identification of copyright violation, presence of illegal copies or illegal distributors, by recovering the embedded information. In particular, classical watermarking applications consist of copy control, broadcast monitoring, fingerprinting, authentication, copyright protection, and access control [12]. Since early 1990s, digital watermarking plays a key role in security of multimedia communication and digital right management. In the last few years, the watermarking schemes have evolved significantly and the classical Spread Spectrum techniques have been outperformed by the so-called informed watermarking, the Quantization Index Modulation (QIM) algorithms [13, 14].

As far as video watermarking techniques are concerned, H.264 requires the definition of new appropriate tools for both copyright protection and authentication. Few works have addressed this problem so far. A hybrid approach is presented in [15] that faces both features by embedding a robust mark in the transform coefficients and a fragile mark in the motion vectors. Authentication is instead achieved in [16] via a fragile watermark insertion using skipped macroblocks of H.264 compressed sequences. Recently, the work in [17] proposed an H.264 video watermarking method which exploits a perceptual model to select the coefficients for the mark insertion. The goal consists, as usual, in increasing the size of the payload, guaranteeing a reduced impact on the visual distortion, and therefore focusing on the standard problem of robustness to common signal processing attacks. The same issue is faced in [18].

In the last few years, data hiding has received considerable attention also in different research areas and has been adopted also to nonsecurity oriented applications. As far as error concealment is concerned, there are some works exploiting watermarking, where embedded information is extracted to detect and conceal errors. In [19] important data for each macroblock (MB) are extracted and inserted into the next frame with suitable MB-interleaving slice-based data hiding techniques for I and P frames, while in the decoding phase, they can be exploited to conceal

the corrupted MBs. Hidden information about the original (high-quality) data can be also used to estimate the quality degradation of the received data. Indeed, the extraction of the embedded information can support the reduced-reference methods for quality assessment [20]. Recently, a simple data hiding technique has been adopted also in video surveillance applications to enhance the visual quality of faces [21].

A first proof of concept that demonstrates the suitability of data hiding tools also in the MDC has been proposed in our previous work [22]. Here we underline the advantages of data hiding in improving MDC schemes while reducing the overhead. The approach embeds the DC coefficient of each block of the JPEG frame into the AC coefficients, therefore, avoiding its transmission.

### 3. The Proposed MDC Coding Scheme

According to the typical constraints of MDC, each descriptor should be

- (i) independent,
- (ii) nonhierarchical,
- (iii) complementary,
- (iv) balanced.

Following these principles, we propose to modify the quality of each MB on the basis of a predefined checkerboard pattern of  $N \times M$  MBs, in which high- and low-quality MBs are alternated. In case only one descriptor out of two is correctly received and decoded, the achievable perceptual quality is reduced, due to the checkerboard pattern. In Figure 1 (left) an example of a  $4 \times 4$  checkerboard pattern is shown. White MBs are coarsely quantized being the least expensive in terms of bit allocation, while a finer quantization is applied to black MBs.

Being the amount of overhead and, therefore, the final bitrate strongly related to the checkerboard structure and the quantization factors of the MBs, we show in Figure 2 the relationship between the redundancy and the size of the grid when adopting H.264/AVC as the coding algorithm. In H.264/AVC, the quality factor is given by the quantization parameter (QP), which values span in the range [1, 51], being 1 the best achievable quality. In the presented configuration, we have chosen  $QP_H = 20$  for high-quality macroblocks, while the coarsely quantized MBs have  $QP_L = 40$ . The same evaluation could be carried out using the MJPEG encoder. However, since only H.264 has predictive features also in the intra frames, the MJPEG curve would be almost flat, since the number of MBs with high and low quality is constant regardless of the grid size and in particular, each block is independently coded and has no relationship with the adjacent ones. For the MJPEG configuration, we have instead set the quality factor (QF) to  $QF_H = 92$  and  $QF_L = 22$ .

The choice of a simple checkerboard pattern and QP (QF) alternation would be reasonable in presence of limited packet losses, and it may generate annoying artefacts in presence of significant losses. However, according to the human visual system, we can say that users typically focus

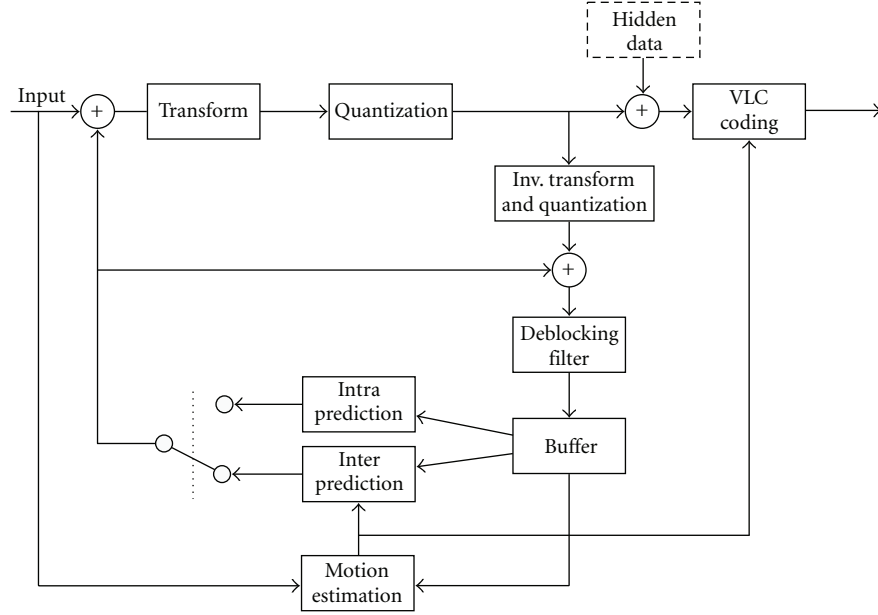


FIGURE 3: The proposed ROI data hiding scheme.

TABLE 1: Set of parameters for ROI coding.

Param	Description	Range	n.bits
select <sub>D</sub>	Identifies D1 or D2	BOOL	1
N	MBs in the horizontal grid	[1–8]	3
M	MBs in the vertical grid	[1–8]	3
ROI	Activate ROI	BOOL	1
QP <sub>L</sub> or QF <sub>L</sub>	Set low quality quantizer	[1–max]	6
X <sub>st</sub>	Start of ROI in X (%)	[1–100]	7
Y <sub>st</sub>	Start of ROI in Y (%)	[1–100]	7
X <sub>end</sub>	End of ROI in X (%)	[1–100]	7
Y <sub>end</sub>	End of ROI in Y (%)	[1–100]	7
ΔQP (ΔQF)	QP <sub>R</sub> – QP <sub>L</sub> (QF <sub>R</sub> – QF <sub>L</sub> )	[0–max]	6
Total required bits			48

the attention on a specific area of the image when watching a video [23]; therefore, we have introduced a mechanism to provide better perceptual quality of the single descriptions reconstruction in a ROI, by (i) reducing the gap between the two quantization levels in the ROI and (ii) decreasing the size of the grid in that area to the minimum size of  $1 \times 1$  MBs. Figure 1 (right) illustrates an example where the grid of the background is set to  $4 \times 4$  MBs and the ROI grid is set to  $1 \times 1$  MBs. A new parameter QP<sub>R</sub> (QP<sub>L</sub> < QP<sub>R</sub> < QP<sub>H</sub>) for the H.264/AVC configuration and QF<sub>R</sub> (QF<sub>L</sub> < QF<sub>R</sub> < QF<sub>H</sub>) for the MJPEG configuration is introduced to replace QP<sub>L</sub> and QF<sub>L</sub> within the ROI and it is represented by the gray MBs in the figure. This parameter allows reducing the gap between the two category of blocks, smoothing the quality differences. In order to make the scheme more flexible, we keep all parameters (Table 1) configurable every intra-frame, so to adapt the ROI to the visual content. Furthermore,

the size of the checkerboard can also be chosen a priori and changed every GOP (in H.264) or every frame (in MJPEG), making it possible to adjust the quality and the corresponding overhead to the video source.

In order to allow the correct decoding of the video sequence, the ROI parameters necessarily require to be transmitted to the receiver/decoder. This is usually achieved by either delivering the data out-of-band or by introducing new syntax elements in the stream. In our approach we have preferred to exploit data hiding techniques due to (i) the negligible impact on the resulting bitrate and (ii) the intrinsic nature of data hiding in terms of data protection, allowing only authorized users to recover the embedded information. The procedure of data embedding and recovery is explicitly described in the next section.

#### 4. ROI Parameters Embedding

The choice of a suitable data hiding scheme requires the estimation of the payload to be transmitted together with the video. For the sake of demonstration, we have roughly calculated a reasonable number of bits that need to be embedded as watermark. According to our design we have considered that 48 bits are sufficient to encode the ROI information, including the size of the grid, the quantization parameters for the background and the ROI, and its coordinates in the image plane (see Table 1). In order to allow a one-to-one comparison between the two coding schemes, ΔQP and ΔQF values are represented with 6 bits, meaning that in the JPEG case, only every second value can be chosen, since  $1 < QF < 100$ , and this would require instead 7 bits. This approximation is reasonable and does not affect the achieved results significantly, since the visual appearance variation is negligible.

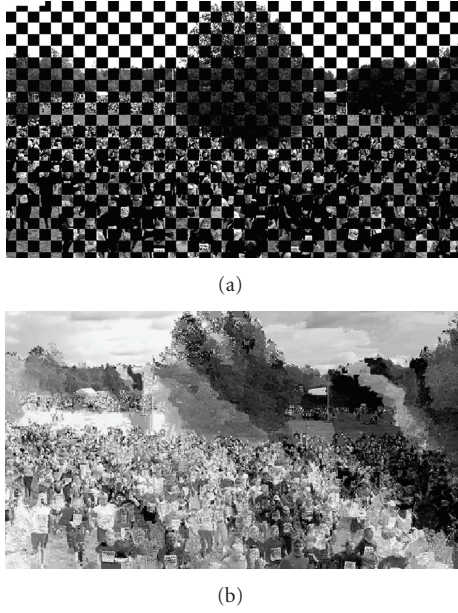


FIGURE 4: Decoding without knowing the watermark extraction key on the *crowd* sequence in case of MJPEG (a) and H.264 (b).

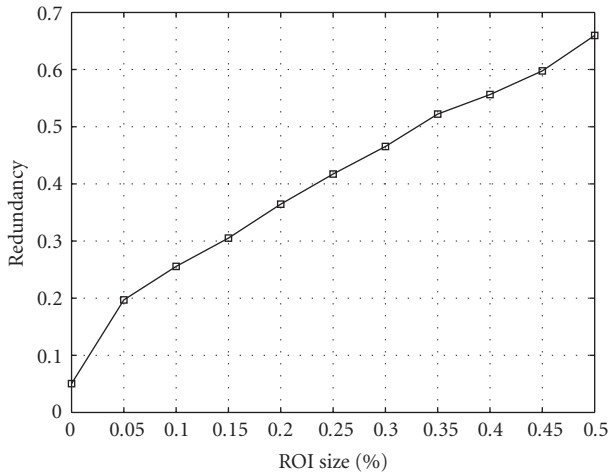


FIGURE 5: Relationship between ROI size and resulting overhead.

In order to limit the additional computational burden of the algorithm, the complexity of the adopted data hiding technique must be kept as low as possible. Therefore, we have chosen a simple yet flexible watermarking scheme in the class of quantization index modulation (QIM) techniques [13]. The adopted embedding technique is the Binary Scalar Costa Scheme (BSCS) [24], and the formula is as follows:

$$q_n = Q_\Delta \left\{ x_n - \Delta \left( \frac{d_n}{D} + k_n \right) \right\} - \left( x_n - \Delta \left( \frac{d_n}{D} + k_n \right) \right), \quad (2)$$

where  $d_n$  is the watermark symbol,  $x_n$  is the coefficient to be marked,  $D$  is the cardinality of the alphabet of the watermark (2 in our case),  $k_n$  is the watermarking key, and  $Q_\Delta(\cdot)$  represents scalar uniform quantization with step  $\Delta$ . We selected  $\Delta = 2$ ,  $k_n = 1$  to minimize the impact of the

inserted information thus leading to perfect mark invisibility and very low complexity. Since the embedding of the ROI parameters has a negligible impact on the redundancy, it provides a powerful tool to improve the perceptual quality of single description, as experimentally verified in Section 5.

In our embedding scheme, we have decided to replicate the mark in three different locations to improve the robustness in case a portion of the image is damaged and we have chosen the first seven MBs of the first, middle, and last row as a cover for the watermark insertion. In each MB, we embed 8 payload bits by modifying the first two AC coefficients of each  $8 \times 8$  block composing the MB after quantization. At the decoder side, the knowledge of the watermark extraction key (i.e., its location) allows the extraction of the embedded mark (i.e., ROI parameters). Notice that the first parameters of Table 1 (select<sub>D</sub>, N, M, ROI, and QP<sub>H</sub>—or QF<sub>H</sub>) are fundamental to correctly start the decoding process. Therefore, to ensure their reconstruction, we set QP (QF) of MB<sub>0</sub> and MB<sub>1</sub> to QP<sub>H</sub>(QF<sub>H</sub>) in both descriptions  $D_1$  and  $D_2$ . The proposed data hiding exploitation adds a significant level of protection to the MDC scheme. As a matter of fact, it is impossible to reconstruct an acceptable version of the video without knowing the watermark extraction key. If the embedded ROI parameters cannot be extracted (in case for example of nonauthorized users) neither descriptor can be correctly decoded. Details about this concept are provided in Section 5. The overall encoding scheme for H.264/AVC is shown in Figure 3. As it can be noticed, the proposed architecture does not alter the coding process, returning a completely standard-compliant stream, resulting in another advantage of this architecture with respect to other solutions. Indeed, the information related to the ROI is included in the transform coefficients and no additional syntax elements need to be defined. The corresponding scheme for the MJPEG implementation is similar to the one in Figure 3, but clearly without the motion compensation loop.

## 5. Experimental Results

This section refers to the validation of the system using three different test sequences retrieved from the European Broadcasting Union (EBU) website [25] in the 720 p format, namely,  $1280 \times 720$ . We will first review the results obtained in the MJPEG framework and successively using H.264. In both cases, the encoders have been configured in order to achieve a PSNR higher than 40 dB on the luminance component in case both descriptors have been correctly received.

As a common achievement in both implementations, we demonstrate that the data hiding procedure serves also as a protection tool, which does not allow enjoying the visual content properly; if the user does not know how to interpret the embedded key, the exact quantization values are known only for the black MBs in Figure 1, while QP<sub>L</sub>/QF<sub>L</sub> are hidden in the cover. The effects on the resulting image depend on the adopted codec, but in general the visual content is not preserved. An example is presented in Figure 4

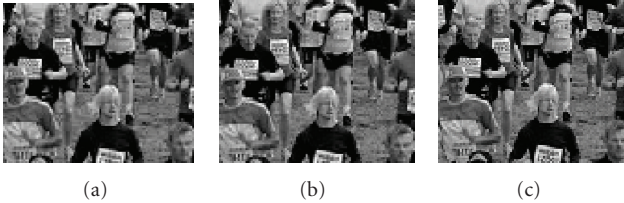


(a)



(b)

FIGURE 6: Full reconstruction (a) and single descriptor reconstruction (b) for *Crowd Run* sequence using MJPEG.



(a)

(b)

(c)

FIGURE 7: ROI disabled (a), ROI enabled with  $\Delta QF = 35$  (b), and ROI enabled with  $\Delta QF = 70$  (c) for *Crowd Run* sequence using MJPEG.

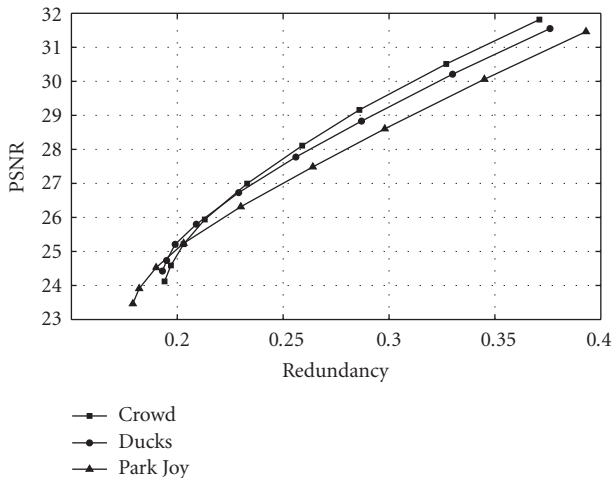


FIGURE 8: Redundancy versus PSNR in case of one descriptor is completely lost for different values of  $QP_H$  and for the *Park Joy*, *Ducks take off*, and *Crowd Run* video sequences.

TABLE 2: Location of ROIs in the test sequences.

Test sequence	$X_{start}$	$Y_{start}$	$X_{end}$	$Y_{end}$
Park Joy	40	40	80	70
Crowd Run	10	60	50	90
Ducks take off	30	40	70	70

TABLE 3: Overhead and average PSNR using MJPEG on the *Crowd Run* sequence.

$\Delta QF$	$\rho$	PSNR (dB)
0 (no ROI)	0.23	29.8
35	0.25	30.69
70 (high quality ROI)	0.34	30.91

without the ROI, for MJPEG and H.264, respectively. The MJPEG implementation shows an evident grid. In fact, during the process of inverse quantization, values are not reported at the original magnitude, and the details of the picture are lost. In case of H.264, instead, the content appears to be more *scrambled*. This is due to the intra prediction algorithm of the codec that tends to compensate possible errors that may occur in intrablocks as well. For the tests, we have set a grid of  $2 \times 2$  MBs, in order to visually reduce the blocking artifacts due to the high- and low-quality block alternation. In the ROI, the grid is set to  $1 \times 1$  MB. The location of the ROI for each test sequence is set in terms of percentage as shown in Table 2.

It is clear that the ROI size has an impact on the total overhead. In our system, the increase of the redundancy is proportional to the ROI size, as demonstrated in Figure 5. Here we have chosen the central portion of the *Ducks take off* video sequence and we have progressively increased the area occupied by the ROI at regular intervals.

**5.1. MJPEG Experiments.** This first set of results aims at demonstrating the viability of the proposed approach in absence of motion compensation videos. It is worth noting that although we have chosen a simple MJPEG codec, there is no practical constraint in terms of compressing standard, since the data hiding procedure can be applied to any transform-based coding scheme. Focusing on the target quality of 40 dB in case of full reconstruction, we have set  $QP_H = 92$  and  $QP_L = 22$  in order to achieve a reasonable compromise between perceived quality and overhead. This configuration returns an average PSNR of the video sequence around 30 dB. In Figure 6, a comparison between the full reconstruction and the single descriptor decoding is shown for the *Crowd Run* sequence. In Figure 7 a detail of the picture is shown, in which the improvements given by the ROI are highlighted at different values for  $\Delta QF$ . The corresponding numerical values are reported in Table 3.

**5.2. H.264 Experiments.** Also in this case, the encoder has been configured in order to ensure a PSNR in the range of 40dB in case both descriptors are received. In our experiments we have set  $QP_H = 20$  and  $QP_L = 40$ . All

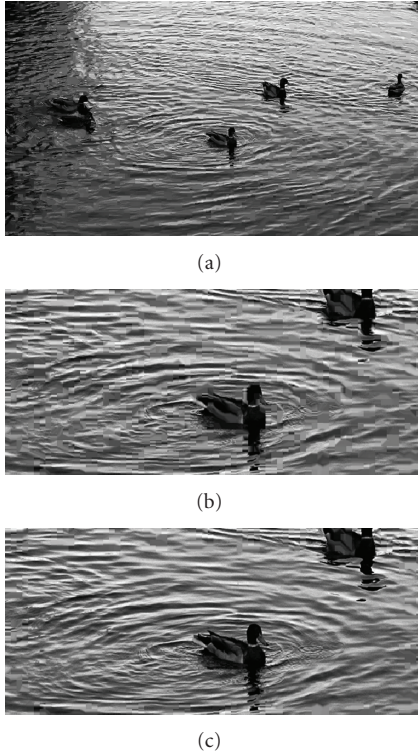


FIGURE 9: *Ducks take off* sequence: entire frame and ROI disabled (a); detail without (b), and with ROI (c).



FIGURE 10: *Crowd Run* sequence detail with disabled (a) and enabled (b) ROI.

the tests presented in this section have been performed using the JM reference software (v12.4) [26]. In Figure 8, the relationship between redundancy and distortion for different values of  $QP_L$  is reported, which demonstrates the versatility of the proposed method in modulating the overhead. We can notice that, on one hand, very low redundancy values can be achieved at the cost of low single description quality, while, on the other hand, it is possible to obtain higher quality

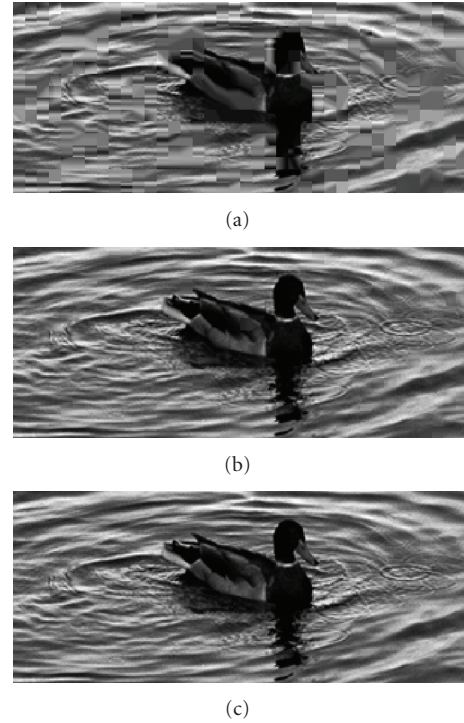


FIGURE 11: Detail of the ROI area for the *Ducks take off* sequence with different  $\Delta QP$ : 0 (a), 15 (b), and 30 (c).

reconstruction by increasing the bitrate. In any case, even increasing  $QP_L$  significantly, the overhead is still constrained below 40% with an average distortion of the whole frame around 30 dB.

Figure 9(a) shows a single description reconstruction ( $QP_L = 50$ ) with disabled ROI, with Figure 9(b) being its detail, while in Figure 9(c), the ROI is inserted. The same experiment is presented in Figure 10 for the *Crowd Run* sequence. In Figure 11, we illustrate the details of the ROI by applying different values of  $\Delta QP$ . As it can be noticed from the picture, the quality of the ROI can be significantly improved by filling the gap in terms of quantization parameter.

Figures 12 and 13 highlight the performance improvement due to the ROI introduction: although this results in an unavoidable bitrate increase, the PSNR in the ROI can be significantly improved by varying the  $QP_R$ , letting the user perceive a better quality in the significant area of the video. Both Figures 12 and 13 refer to a configuration in which  $QP_H$  is set to 20,  $QP_L$  is set to 50, and  $QP_R$  is between 45 and 25.

To assess the perceived quality improvement introduced by the ROI we employed the NTIA General Video Quality Metric (VQM) [27], which returns values in the 0–1 range, with 1 representing worst quality. This metric has been proved to be highly correlated to subjective ratings of processed HD video sequences [28]. The results are reported in Table 4.

As a final comment on the achieved results, we underline that the proposed method embeds the ROI data within a reduced number of MBs, altering a maximum of only

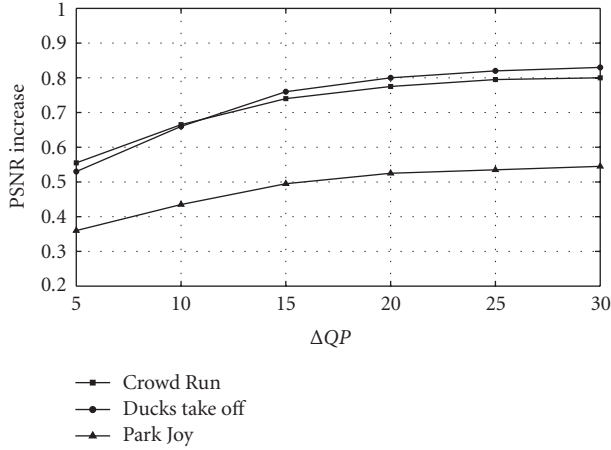


FIGURE 12: Quality improvement due to ROI introduction for the *Park Joy*, *Ducks take off*, and *Crowd Run* video sequences.

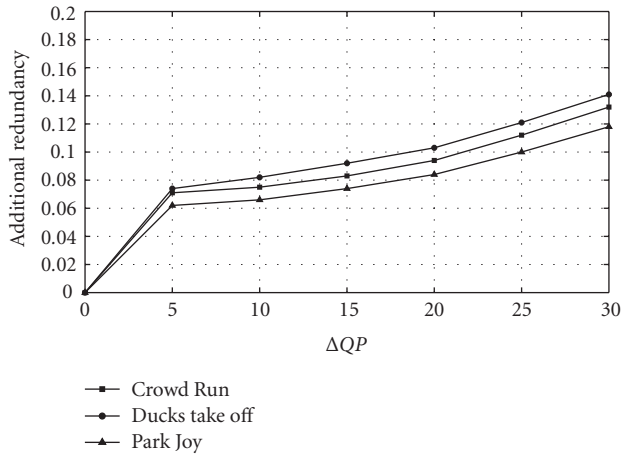


FIGURE 13: Overhead increase due to ROI introduction for the *Park Joy*, *Ducks take off*, and *Crowd Run* video sequences.

TABLE 4: Comparison of VQM results between enabled and disabled ROI for the *Park Joy*, *Ducks take off*, and *Crowd Run* video sequences.

Sequence	VQM (ROI Off)	VQM (ROI On)
Park Joy	0.5955	0.5689
Crowd Run	0.6011	0.5773
Ducks take off	0.7061	0.6848

two transform coefficients per block. Therefore, the quality degradation due to data hiding is completely negligible. Simulation results refer that the average loss in a single watermarked macroblock can be assessed around 0.45 dB in both MJPEG and H.264/AVC, therefore, minimally affecting the visual appearance of the image. We have replicated the watermark three times in different areas of the frame, thus reducing the probability of ROI details being lost due to channel errors. Even in this case, the impact on stream quality and bitrate is minimal: our experiments refer negligible additional distortion ( $<0.005$  dB) and increased

redundancy (0.004%) when applying three fully redundant watermarks on the *Ducks take off* sequence.

## 6. Conclusions

In this paper, we presented a novel MDC technique for HDTV video sequences, using MJPEG and H.264/AVC. The proposed method exploits data hiding to improve the perceptual quality of each single description within specific regions of interest. ROIs are defined a priori and coded with better quantization factors, and the corresponding information required at the decoder side is transmitted via a suitable watermarking algorithm. We reported extensive experimental results demonstrating the effectiveness of the proposed method, which allow increasing the PSNR of single descriptions at the cost of a slightly higher redundancy.

## Acknowledgment

This work is partially funded by the Province of Trento (Italy) in the framework of the TRITON Project.

## References

- [1] Y.-C. Lee, J. Kim, Y. Altunbasak, and R. M. Mersereau, "Layered coded vs. multiple description coded video over error-prone networks," *Signal Processing: Image Communication*, vol. 18, no. 5, pp. 337–356, 2003.
- [2] Y. Wang, A. R. Reibman, and S. Lin, "Multiple description coding for video delivery," *Proceedings of the IEEE*, vol. 93, no. 1, pp. 57–70, 2005.
- [3] V. A. Vaishampayan, "Design of multiple description scalar quantizers," *IEEE Transactions on Information Theory*, vol. 39, no. 3, pp. 821–834, 1993.
- [4] W. Jiang and A. Ortega, "Multiple description coding via polyphase transform and selective quantization," in *Proceedings of the Visual Communications and Image Processing*, vol. 3653 of *Proceedings of SPIE*, pp. 998–1008, San Jose, Calif, USA, January 1999.
- [5] A. C. Miguel, A. E. Mohr, and E. A. Riskin, "SPIHT for generalized multiple description coding," in *Proceedings of IEEE International Conference on Image Processing*, vol. 3, pp. 842–846, Kobe, Japan, October 1999.
- [6] S. D. Servetto, K. Ramchandran, V. A. Vaishampayan, and K. Nahrstedt, "Multiple description wavelet based image coding," *IEEE Transactions on Image Processing*, vol. 9, no. 5, pp. 813–826, 2000.
- [7] C. Soldani, G. Leduc, F. Verdichio, and A. Munteanu, "Multiple description coding versus transport layer FEC for resilient video transmission," in *Proceedings of the International Conference on Digital Telecommunications (ICDT '06)*, 2006.
- [8] M. Tesanovic, D. R. Bull, A. Doufexi, and A. R. Nix, "H.264-based multiple description coding for robust video transmission over MIMO systems," *Electronics Letters*, vol. 42, no. 18, pp. 1028–1029, 2006.
- [9] T. Tillo, M. Grangetto, and G. Olmo, "Redundant slice optimal allocation for H.264 multiple description coding," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 18, no. 1, pp. 59–70, 2008.



- [10] S. Moradi, S. Gazor, and T. Linder, "A multiple description video coding motivated by human visual perception," in *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP '08)*, pp. 689–692, Las Vegas, Nev, USA, March 2008.
- [11] E. T. Lin, A. M. Eskicioglu, R. L. Lagendijk, and E. J. Delp, "Advances in digital video content protection," *Proceedings of the IEEE*, vol. 93, no. 1, pp. 171–182, 2005.
- [12] G. Doerr and J.-L. Dugelay, "A guide tour of video watermarking," *Signal Processing: Image Communication*, vol. 18, no. 4, pp. 263–282, 2003.
- [13] B. Chen and G. W. Wornell, "Quantization index modulation methods for digital watermarking and information embedding of multimedia," *Journal of VLSI Signal Processing Systems for Signal, Image, and Video Technology*, vol. 27, no. 1-2, pp. 7–33, 2001.
- [14] L. Pérez-Freire, F. Pérez-González, and S. Voloshynovskiy, "Revealing the true achievable rates of Scalar Costa scheme," in *Proceedings of the 6th IEEE Workshop on Multimedia Signal Processing*, pp. 203–206, September–October 2004.
- [15] J. Zhang, A. T. S. Ho, G. Qiu, and P. Marziliano, "Robust video watermarking of H.264/AVC," *IEEE Transactions on Circuits and Systems II*, vol. 54, no. 2, pp. 205–209, 2007.
- [16] D. Profrock, H. Richter, M. Schlauweg, and E. Müller, "H.264/AVC video authentication using slapped macroblocks for an erasable watermark," in *Visual Communications and Image Processing*, vol. 5960 of *Proceedings of SPIE*, pp. 1480–1489, Beijing, China, July 2005.
- [17] M. Noorkami and R. M. Mersereau, "A framework for robust watermarking of H.264-encoded video with controllable detection performance," *IEEE Transactions on Information Forensics and Security*, vol. 2, no. 1, pp. 14–23, 2007.
- [18] D. Profrock, M. Schlauweg, and E. Müller, "A new uncompressed-domain video watermarking approach robust to H.264/AVC compression," in *Proceedings of the 3rd IASTED International Conference on Signal Processing, Pattern Recognition, and Applications*, pp. 99–104, 2006.
- [19] L.-W. Kang and J.-J. Leou, "An error resilient coding scheme for H.264/AVC video transmission based on data embedding," *Journal of Visual Communication and Image Representation*, vol. 16, no. 1, pp. 93–114, 2005.
- [20] Z. Wang, G. Wu, H. R. Sheikh, E. P. Simoncelli, E.-H. Yang, and A. C. Bovik, "Quality-aware images," *IEEE Transactions on Image Processing*, vol. 15, no. 6, pp. 1680–1689, 2006.
- [21] P. Meuel, M. Chaumont, and W. Puech, "Data hiding in H.264 video for lossless reconstruction of region of interest," in *Proceedings of European Signal Processing Conference*, 2007.
- [22] N. Conci, M. Carli, and G. Boato, "Data hiding tools for multiple description video coding," *Electronics Letters*, vol. 45, no. 2, pp. 98–99, 2009.
- [23] Z. Wang, L. Lu, and A. C. Bovik, "Video quality assessment based on structural distortion measurement," *Signal Processing: Image Communication*, vol. 19, no. 2, pp. 121–132, 2004.
- [24] J. J. Eggers, R. Bauml, R. Tzschoppe, and B. Girod, "Scalar Costa scheme for information embedding," *IEEE Transactions on Signal Processing*, vol. 51, no. 4, pp. 1003–1019, 2003.
- [25] European Broadcasting Union, <http://www.ebu.ch>.
- [26] I. 14496-10 and I. R. H.264, Advanced Video Coding 2003.
- [27] VQEG, "Final report from the Video Quality Experts Group on the validation of objective models of video quality assessment, Phase II," August 2003.
- [28] S. Wolf and M. H. Pinson, "Application of the NTIA general video quality metric (VQM) to HDTV quality monitoring," in *Proceedings of the 3rd International Workshop on Video Processing and Quality Metrics for Consumer Electronics (VPQM '07)*, Scottsdale, Ariz, USA, January 2007.

## Special Issue on Network Coding for Wireless Networks

### Call for Papers

The main idea in network coding was introduced in 2000 by Ahlswede et al. With network coding, an intermediate node in a network cannot only forward its incoming packets but also encode them. It has been shown that the use of network coding can enhance the performance of wired networks significantly. Recent work indicates that network coding can also offer significant benefits for wireless networks.

Communications over wireless channels are error-prone and unpredictable due to fading, mobility, and intermittent connectivity. Moreover, in wireless networks, transmissions are broadcast and can be overheard by neighbors, which is treated in current systems as interference. Finally, security poses new challenges in wireless networks, where both passive and active attacks have quite different premises than in wired networks. Ideas in network coding promise to help toward all these issues, allowing to gracefully add redundancy to combat errors, take advantage of the broadcast nature of the wireless medium and achieve opportunistic diversity, exploit interference rather than be limited by it, and provide secure network communication against adversarial attacks.

In this special issue, we are interested in original research articles which can carry the momentum further and take the wireless network coding research to the next level. The areas of interest include novel network code designs and algorithms, new applications of wireless network coding, network coding capacity, and performance analysis. In addition to original research articles, we are open to review articles. The following list indicates topics of interest which is by no means exhaustive:

- Network codes and algorithms for wireless networks
- Physical layer network coding
- Joint source coding and network coding
- Graph codes and network coding
- Reduced complexity decoding for network coding
- Secure network coding
- Capacity and fundamental bounds on network coding performance
- Cross-layer optimization and network coding
- Energy-efficient network coding
- TCP, routing, MAC, or scheduling algorithms for network codes

- Wireless network coding for multimedia application
- Wireless network coding for bio-medical application

Before submission authors should carefully read over the journal's Author Guidelines, which are located at <http://www.hindawi.com/journals/wcn/guidelines.html>. Prospective authors should submit an electronic copy of their complete manuscript through the journal Manuscript Tracking System at <http://mts.hindawi.com/> according to the following timetable:

Manuscript Due	October 1, 2009
First Round of Reviews	January 1, 2010
Publication Date	April 1, 2010

### Lead Guest Editor

**Heung-No Lee**, Department of Information and Communications, Gwangju Institute of Science and Technology (GIST), Gwangju, South Korea; [heungno@gist.ac.kr](mailto:heungno@gist.ac.kr)

### Guest Editors

**Sae-Young Chung**, School of EECS, KAIST, Daejeon, South Korea; [sychung@ee.kaist.ac.kr](mailto:sychung@ee.kaist.ac.kr)

**Christina Fragouli**, School of Computer and Communication Sciences, EPFL, Lausanne, Switzerland; [christina.fragouli@epfl.ch](mailto:christina.fragouli@epfl.ch)

**Zhi-Hong Mao**, ECE/Bio-Medical Dept., the University of Pittsburgh, Pittsburgh, PA, USA; [zhm4@pitt.edu](mailto:zhm4@pitt.edu)

## Special Issue on Video Analysis, Abstraction, and Retrieval: Techniques and Applications

### Call for Papers

The proliferation of TV broadcast channels and programs has led to an explosion of digital video content, which results in large personal and public video databases. However, the rapidly increasing availability of video data has not yet been accompanied by an increase in its accessibility. This is due to the situation that video data are naturally different to traditional forms of data, which can be easily accessed and searched based on text. Therefore, how to efficiently organize broadcast video, such as TV news and sports, into more compact forms and extract semantically meaningful information becomes more and more important. In the past ten years, the majority of research has gradually converged to three fundamental areas, namely, video analysis, video abstraction, and video retrieval. Video analysis is utilized to extract both general and domain-specific visual features, such as color, texture, shape, human faces, and human motion. Video abstraction is to generate a representation of visual information, which is similar to the extraction of keywords or summaries in text document processing. Basically, video abstraction is associated with key-frame detection, shot clustering, and the extraction of domain knowledge of the targeted video source. The content attributes found in video analysis and abstraction processes are often referred to as metadata. In many information systems, we need fast schemes and tools to use content metadata to query, search, and browse large video databases. Although a lot of efforts have been devoted into this area, both computational cost and accuracy of the existing systems are still far from satisfactory.

This special issue aims at capturing the latest advances of the research community working in video analysis, abstraction, and retrieval for broadcasting applications. The objectives of this special issue are twofold: (1) publishing novel fundamental techniques, and (2) showcasing robust systems to treat popular broadcast videos, such as TV news and sports video. Topics of interest include, but are not limited to:

- Feature extraction and description from broadcast video
- Object detection, tracking, and recognition in broadcast video
- Shot boundary detection and scene segmentation

- Key frame extraction and video summarization
- Efficient methods for video indexing and concepts modeling
- Semantic content understanding and recognition
- Video browsing/visualization tools for the broadcast video
- Semantic annotations of video content
- Metadata Standards for Video Analysis, Abstraction and Retrieval
- Multimodal data generation and fusion
- User interface for media browsing and search
- General framework for video retrieval
- Evaluation techniques and methodologies for video abstraction and retrieval
- Robust systems: TV news, sports, and so forth.

Before submission authors should carefully read over the journal's Author Guidelines, which are located at <http://www.hindawi.com/journals/ijdmb/guidelines.html>. Prospective authors should submit an electronic copy of their complete manuscript through the journal Manuscript Tracking System at <http://mts.hindawi.com/>, according to the following timetable:

Manuscript Due	September 1, 2009
First Round of Reviews	December 1, 2009
Publication Date	March 1, 2010

### Guest Editors

**Jungong Han**, Eindhoven University of Technology, 5600 MB Eindhoven, The Netherlands; [jg.han@tue.nl](mailto:jg.han@tue.nl)

**Ling Shao**, Philips Research Laboratories, 5656 AA Eindhoven, The Netherlands; [l.shao@philips.com](mailto:l.shao@philips.com)

**Peter H. N. de With**, CycloMedia/Eindhoven University of Technology, 4180 BB Waardenburg, The Netherlands; [p.h.n.de.with@tue.nl](mailto:p.h.n.de.with@tue.nl)

**Ling Guan**, Ryerson University, Toronto, ON, Canada M5B 2K3; [lguan@ee.ryerson.ca](mailto:lguan@ee.ryerson.ca)

## Special Issue on Microphone Array Speech Processing

### Call for Papers

Significant knowledge about microphone arrays has been gained from years of intense research and product development. There have been numerous applications suggested, for example, from large arrays (on the order of  $>100$  elements) for use in auditoriums to small arrays with only 2 or 3 elements for hearing aids and mobile telephones. Apart from that, array technology has been widely applied in the areas of speech recognition and more recently surveillance. Traditional techniques that have been used for microphone arrays include the fixed spatial filter as well as optimal and adaptive beamforming. These techniques model input or calibration signals as well as localization information for their design. Today contemporary techniques using blind signal separation (BSS) and time frequency masking techniques have attracted significant attraction. Those techniques are less reliant on array modeling and localization, but more on the statistical properties of speech signals such as sparseness, non-Gaussianity, nonstationarity, and so forth. The main advantage that multiple microphones add from a theoretical perspective is the spatial diversity, which is an effective tool to combat interference, reverberation, and noise when used according to the theoretical assumption. Combining spatial information with time-frequency information and perceptual cues will lead to innovative techniques and new methods, which will provide improved communication capabilities in challenging acoustic environments.

To further enhance current research and to promote new applications, this special issue aims to collect and present the latest research efforts in signal processing methods and algorithms for microphone arrays.

Topics of interest include, but are not limited to:

- Optimal and adaptive beamforming
- Blind signal extraction methods
- Multichannel dereverberation techniques
- Microphone array-assisted multichannel acoustic echo cancellation
- Spatial filtering techniques
- Sound source localization and tracking
- Psychoacoustically motivated procedures and algorithms such as perceptual cues, hearing thresholds, and spatial masking effects

- Distributed microphone networks
- Spherical array of microphones and Eigen/Modal beamforming

Before submission authors should carefully read over the journal's Author Guidelines, which are located at <http://www.hindawi.com/journals/asp/guidelines.html>. Prospective authors should submit an electronic copy of their complete manuscript through the journal Manuscript Tracking System at <http://mts.hindawi.com/>, according to the following timetable:

Manuscript Due	August 1, 2009
First Round of Reviews	November 1, 2009
Publication Date	February 1, 2010

### Guest Editors

**Sven Nordholm**, Western Australian Telecommunications Research Institute (WATRI), Curtin University of Technology, Perth, WA 6009, Australia; [sven@watri.org.au](mailto:sven@watri.org.au)

**Thushara Abhayapala**, Department of Information Engineering, Research School of Information Sciences and Engineering, ANU College of Engineering and Computer Sciences, The Australian National University, Canberra, ACT 0200, Australia; [thushara.abhayapala@anu.edu.au](mailto:thushara.abhayapala@anu.edu.au)

**Simon Doclo**, NXP Semiconductors, Corporate I&T-Research, Interleuvenlaan 80, 3001 Leuven, Belgium; [simon.doclo@nxp.com](mailto:simon.doclo@nxp.com)

**Sharon Gannot**, School of Engineering, Bar-Ilan University, 52900 Ramat-Gan, Israel; [gannot@macs.biu.ac.il](mailto:gannot@macs.biu.ac.il)

**Patrick Naylor**, Department of Electrical and Electronic Engineering, Imperial College, London SW7 2AZ, UK; [p.naylor@imperial.ac.uk](mailto:p.naylor@imperial.ac.uk)

**Ivan Tashev**, Microsoft Research Redmond, One Microsoft Way, Redmond WA, 98052-6399, USA; [ivantash@microsoft.com](mailto:ivantash@microsoft.com)