Hindawi Publishing Corporation
International Journal of Distributed Sensor Networks
Volume 2013, Article ID 209318, 10 pages
http://dx.doi.org/10.1155/2013/209318

Research Article

A Coding and Postprocessing Framework of Multiview Distributed Video for Wireless Video Sensor Networks

Chong Han,[1] Lijuan Sun,[1,2,3] and Jian Guo[1,2,3]

[1] *College of Computer, Nanjing University of Posts and Telecommunications, Nanjing 210003, China*
[2] *Jiangsu High Technology Research Key Laboratory for Wireless Sensor Networks, Nanjing 210003, China*
[3] *Key Lab of Broadband Wireless Communication and Sensor Network Technology, Ministry of Education Jiangsu Province, Nanjing 210003, China*

Correspondence should be addressed to Lijuan Sun; sunlj@njupt.edu.cn

Received 14 July 2013; Accepted 22 August 2013

Academic Editor: Zhijie Han

In many surveillance application scenarios of wireless video sensor networks (WVSNs), a number of video sensors are deployed, and multidimension monitored the visual information in a region of interest, forming multiview videos. Since the power, computing capability, and bandwidth are very limited in WVSNs, the conventional multiview video coding method is no longer applicable. So multiview distributed video coding (MDVC) emerged and developed rapidly. In this paper, we propose a new multiview video coding and postprocessing framework for multiview videos. First, in coding scheme, motion intense regions (MIRs) and nonmotion intense regions (NMIRs) based on sum of absolute difference (SAD) criteria are distinguished. For the MIR, the side information (SI) is generated by fusion temporal SI and interview spatial SI at the pixel level. But for the NMIR, the temporal SI is directly use as the ultimate SI. Then, to further improve the quality of the decoded image, an image postprocessing scheme is designed by using deblocking and deringing artifact filters on decoded image. Finally, a set of experimental results show that the proposed fusion SI approach can bring improvements up to 0.2–0.5 dB when compares with only temporal SI used. The subsequent decoded videos postprocessing simulation proves that the proposed postprocessing scheme can provide an additional improvement of about 0.1 dB to the decoded video sequences.

1. Introduction

Nowadays, with the rapid development of wireless communication and microelectronics technologies, a series of devices and systems relevant with wireless network and video techniques, such as wireless video sensor networks (WVSNs), wireless IP camera, mobile video phone, dense camera arrays, satellite communication systems, and television systems for multiview virtual meeting, have wide application prospects. These kind of video equipment or systems have a common feature, that is, the computing, data storage, and power consumption capacity are all limited, but all abilities in receiving ends are very powerful. Meanwhile, along with the demand for high-quality multimedia/video surveillance, environmental monitoring, and industrial process control, and more and more video sensors are used in WVSNs. In WVSNs, a number of video sensor nodes are deployed and multidimension shot visual information in a region of interest. As a result, the visual information retrieved from adjacent video senor nodes is called multiview video, which usually exhibit high levels of correlation and give rise to considerable data redundancy in the network.

In fact, multiview video is a kind of video, which has interactive manipulation functions and stereoscopic impression. It is the video record acquired from video sensors placed under a particular array in a same scene and provides the capability of scene roaming and viewpoint selection for users. Since all video sensors shoot the same scene from multiple views, the video network contains a large number of interview statistical dependencies, high-quality images/videos could collect by combining intraview (motion estimation), and interview prediction (disparity estimation). Although the multiview video technology provides a richer viewing experience than traditional video technology, it needs to deal with a huge amount of data, and this brings new challenges to the data compression. Traditional video coding standards, such as MPEG-x and H.26x [1], mainly rely on the hybrid architecture, and encoder uses motion estimation to fully exploit the video sequences of time and spatial correlation information. Since the heavy computing burden of the motion estimation and compensation task in these video compression standards, the encoder is overwhelmingly more complex than the decoder; for example, the H.264/AVC encoder/decoder complexity ratio is in the order of 10 for basic configurations and can grow up to 2 orders of magnitude for complex ones [2, 3]. So traditional standardized video coding technologies are difficult to meet the low-complexity coding of the new video application in WVSNs.

To address this above requirement, distributed video coding (DVC) which first practical schemes proposed in [4, 5] is a solution because it is based on Slepian-Wolf [6] and Wyner-Ziv theories [7] and relies on a new statistical framework, instead of the deterministic approach of conventional coding techniques. DVC allows shifting the complexity from the encoder to the decoder and encoding with very low complexity then gives the decoder the task to exploit the source statistics to achieve efficient compression. DVC is also well suited for camera/video sensor networks, where the correlation across multiple views can be exploited at the decoder, without communications between the cameras/video sensor nodes [8]. This kind of coding scheme for multiview distributed videos is called multiview distributed video coding (MDVC).

No matter the conventional DVC or MDVC is, the side information (SI) generation is the key link in coding framework. But different from traditional monoview DVC, the SI generation of MDVC can be computed not only from previous and next decoded frames in the same view but also from frames in other spatially proximal views. Meanwhile, DVC and MDVC as source compression and coding scheme transform coding have over the years emerged as the dominating compression strategy. Transformations are decomposition and representation of the image information. Energy of transformed image focused on the transform domain determines the object of quantization coding, which is the core part of image and video coding. However, the transformation itself will not bring about distortion and losses of image information. The distortion segment of image and video coding is quantization process. In traditional image and video coding, standards such as JPEG and JPEG 2000 and video compression standards such as H.26x, the discrete cosine transformation (DCT), and quantization based on divided image blocks all lead to coding effect. So in this paper, we study the MDVC coding and propose a new MDVC framework, which is constituted by video coding scheme and image postprocessing scheme in WVSNs. Our main contributions include the following.

(1) Based on the video sensor characteristic of WVSNs, we present a new MDVC scheme, which not only contains the temporal SI fusion but also includes the spatial SI fusion method.

(2) For distortion issue caused by quantization processing of image and video coding, to further improve the quality of decoded video, image postprocessing based on spatial, using deblock effect and dering effect on the decoded video is designed.

The remainder of this paper is organized as follows. Section 2 discusses some previous works on DVC and MDVC which motivated our work. Section 3 introduces the proposed multiview distributed video coding scheme. Section 4 describes the image postprocessing scheme in the decoder. The performance evaluations of the proposed framework are presented in Section 5. Finally, conclusions and future work are derived in Section 6.

2. Related Works

DVC is the important application of distributed source coding (DSC) for video coding. The theoretical basis of DSC is Slepian-Wolf theorem [6] and Wyner-Ziv theorem [7]. Theory of Slepian-Wolf shows that even if correlated sources are encoded without getting information from each other, coding performance can be as good as dependent encoding if the compressed signals can be jointly decoded. Wyner-Ziv theory extends this conclusion to the lossy source coding with side information. The Slepian-Wolf and Wyner-Ziv theorems suggest that it is possible to compress two statistically dependent signals in a distributed way (separate encoding, joint decoding), approaching the coding efficiency of conventional predictive coding schemes (joint encoding and decoding). Based on these theorems, DVC has emerged and became a hot research topic rapidly [4, 5, 9–12]. The typical DVC solutions are Berkeley WZ video codec [4, 11] and Stanford WZ video coding architecture [5, 10]. The Berkeley WZ video coding solution is mainly characterized by block-based coding with decoder motion estimation, works at block level, and does not require a feedback channel; the Stanford architecture is mainly characterized by frame-based Slepian-Wolf coding, typically using turbo codes, and a feedback channel to perform rate control at the decoder.

Along with the rise of distributed camera/video network and the development of Multiview video coding, the architecture of DVC is considered using in multiview video coding [13–17]. In [13], DVC strategy is first extended to multiview video coding, and a more flexible side information generation algorithm considering both temporal and view-directional correlations is proposed to achieve high prediction accuracy. In [14], video sensors are arranged in an array to monitor the same scene from different view points. The impact of disparity fields at the central decoder and how to estimate the centralized disparity compensation at the decoder to improve the efficiency of the video sensor networks are discussed. In [15], based on multiview videos and DVC, a scheme for coding video surveillance camera networks is introduced. Then a new fusion technique between temporal side information and homography-based side information is proposed to improve the rated-distortion performance. In [16], based on WVSNs, a low-complexity video compression algorithm that uses the edges of objects in the frames to estimate and compensate for motion is put forward, and two schemes that balance energy consumption among nodes in a cluster on a WVSN are proposed. In [17], based on wireless multimedia sensor networks (WMSNs), a power-rate-distortion (PRD) optimized resource-scalable low-complexity multiview video encoding scheme is proposed, and resource allocation achieved at the encoder while optimizing the reconstructed video quality is discussed.

From the above existing works in DVC and MDVC, lots of outstanding accomplishments have achieved, but there are some shortcomings still existing; for example, some coding architectures need feedback channel to perform rate control at the decoder, which would result in a large amount feedback loops. Obviously, this is unrealistic when the sensor node scale of WVSNs is very tremendous. Another example, in the encoder, the type of the more than half of the encoded frames is traditional intracoded in current most MDVC schemes. The coding of these frames is complexity and inefficiency. Moreover, using the image postprocessing technique could effectively improve the quality of decoded video in decoder, which is rarely involved in existing MDVC scheme. Meanwhile, based on turbo or LDPC, the Wyner-Ziv frame of MDVC in all regions without distinction, the side information is all fused by the temporal SI and interview spatial SI.

For the above problems, in this paper, we propose an improved MDVC coding and postprocessing framework. First, in the encoder of main perspective, according to the Sum of absolute difference (SAD) criteria, we differentiate the motion intense regions (MIRs) and the nonmotion intense regions (NMIRs) of a raw video frame and encode severally. In the decoder, for the MIR, the side information (SI) is generated by fusion temporal SI and interview SI. But for the NMIR, we directly use the temporal SI, the scheme utilizes motion compensated temporal interpolation (MCTI) to generate temporal SI. After the above steps in frame SI generation, we process the image postprocessing of the single decoded frame. Then mix the decoded frames in order to produce the decoded video.

3. Proposed Multiview Distributed Video Coding Scheme

In this section, we present a new MDVC scheme for WVSNs. First, we introduce the principles of DVC, then the frame and coding structure, the temporal, and spatial SI calculation techniques based on SAD criteria, and SI mask fusion method of this proposed MDVC scheme is described.

3.1. Principles of DVC. As we know, the principles of DVC are Slepian-Wolf [6] and Wyner-Ziv [7] theorems. The rate boundaries defined by the Slepian-Wolf theorem for the independent encoding and joint decoding of two statistically dependent discrete random independent and identically distributed (i.i.d.) sources are illustrated in Figure 1.

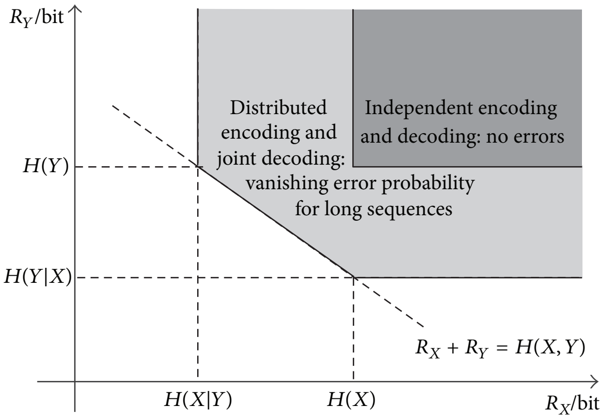


Figure 1: Rate boundaries defined by the Slepian-Wolf theorem.

In Figure 1, X and Y are two i.i.d. random variables/sequences; that is, the raw signals, $H(X)$ and $H(Y)$ are the entropies of X and Y. R_X and R_Y are the bit rates of lossless coding of X and Y, respectively. In traditional entropy encoding, we can get $R_X \geqslant H(X)$ and $R_Y \geqslant H(Y)$. According to the information theory, X and Y can be encoded using joint coding at the bit rate of conditional entropy $R_X \geqslant H(X \mid Y)$ and $R_Y \geqslant H(Y \mid X)$, respectively, and the total bit rate is their joint entropy $R_X + R_Y \geqslant H(X,Y)$. Conversely, with distributed coding, these two signals are independently encoded but jointly decoded. In this case, the Slepian-Wolf theorem proves that the minimum rate is still $H(X,Y)$ with a residual error probability which tends towards 0 for long sequences. In other words, Slepian-Wolf coding allows the same coding efficiency to be asymptotically attained.

Subsequently, Wyner-Ziv theorem extended the Slepian-Wolf theorem by characterizing the achievable rate-distortion region for lossy coding with SI. Wyner-Ziv theorem studied a particular case of distributed source coding, asymmetric coding, that deals with lossy compression of source X associated with the availability of the Y source at the decoder but not at the encoder, and Y (or a derivation of Y) is known as side information. A conclusion is derived that, typically, there is a rate loss incurred when the side information is not available at the encoder. But when performing independent encoding with side information under certain conditions, that is, when X and Y are jointly Gaussian, memoryless sequences and a mean-squared error distortion measure are considered. There is no coding efficiency loss with respect to the case when joint encoding is performed, even if the coding process is lossy. The structure of Wyner-Ziv codec is shown in Figure 2. Together, the Slepian-Wolf and Wyner-Ziv theorems suggest that it is possible to compress two statistically dependent signals in a distributed way, namely, separate encoding, joint decoding, approaching the coding efficiency of conventional joint encoding, and decoding predictive coding schemes. And in general, DVC uses Wyner-Ziv coding scheme as its lossy particular case [18, 19].

3.2. Frame and Coding Structure of the Proposed MDVC. In this section, based on the sensor characteristic of WVSNs, we propose a feasible MDVC scheme, which could be regarded as an extension of conventional DVC. In a WVSN, video sensors are separated at certain distances and angles by each

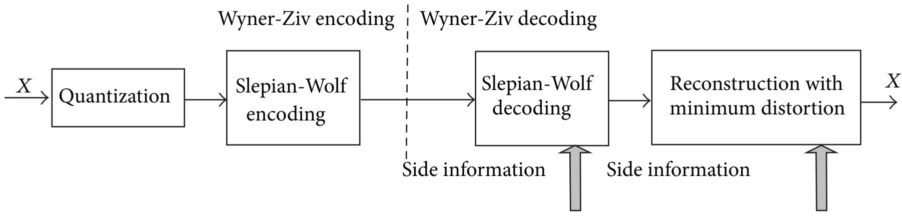


Figure 2: The structure of Wyner-Ziv codec.

other. All nodes synchronized shoot the same scenario in an area of interest of WVSN, and the relevant video sequences are produced. Then the encoder, that is, the video sensor node, encodes the captured sequences in order independently. In encoder, MDVC does not use the complexity encoder for encoding, and interview does not perform data communication. It exploits the source statistics of intraview and interview to obtain high quality decoded video, so it is superior to traditional multiview coding. The frame and coding structure of the proposed MDVC scheme is shown in Figure 3. Multiview video frames are classified into two categories: key frames and Wyner-Ziv frames, noted by K and WZ, respectively. The $C_v, C_{v-1}, C_{v+1}, \ldots$ denote the video sensors in adjacent views, and the $F_t, F_{t-1}, F_{t+1}, \ldots$ represent the successive frames collected by the video sensors with time order. In encoder, the key frames are encoded by the intraframe codec with the traditional DCT based intracoding method, and the Wyner-Ziv frames are encoded by Wyner-Ziv codec. In decoder, the key frames are individually decoded by conventional intraframe decoder and the Wyner-Ziv frames are joint decoded by fusion information of the temporal SI from previous and next key frames in same view and spatial SI from the key frames in adjacent two views. Since Wyner-Ziv frames can be intraencoded and inter decoded, the whole scheme consists of independent encoder and joint decoder. Figure 4 shows a sample of WZ frame coding scheme in MDVC. All key frames are separately intraframe encoding and decoding. For the Wyner-Ziv frame, we use intraframe encoding and interviews decoding with temporal SI and spatial SI.

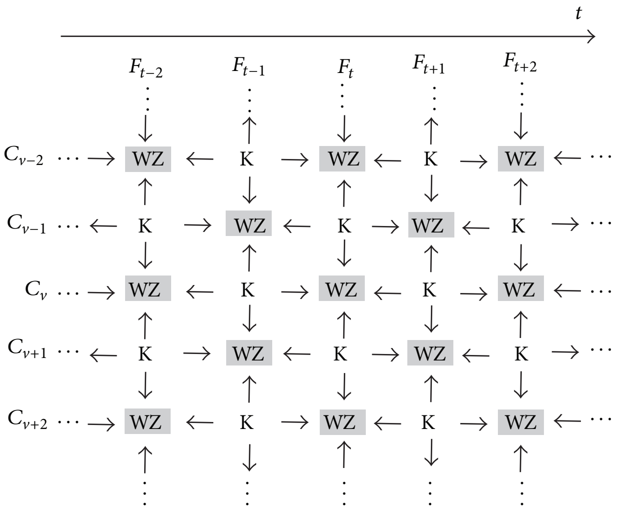


Figure 3: Successive timestamp frame structure for MDVC scheme in a WVSN.

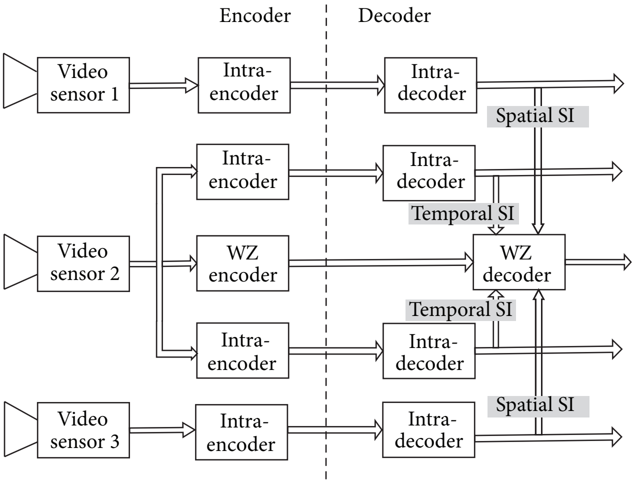


Figure 4: A sample of Wyner-Ziv frame coding scheme in MDVC.

3.3. *MDVC Scheme Based on the SAD Criteria.*

The key technique of the MDVC is about exploiting both temporal and interview correlations in an efficient way. Most of the existing works on DVC/MDVC are based on on turbo or LDPC, the Wyner-Ziv frame coding in all regions without distinction, motion estimation techniques cannot accurately predict the area which are more intense exercise. The decoder needs to request more feedback information, thus not only the rate increases, but the decoded image is still not accurate enough. For this problem, we propose an improved MDVC scheme based on SAD criteria.

In some video sequences, motion vector of many macroblocks is equal to zero or very small, so only a small part of macroblocks has large moving. For the macroblocks with motion vector same as being zero or very small, motion compensated temporal interpolation (MCTI) can make a good predictive coding performance. While for other intense motion macroblocks, decoder needs to use information from other cameras. In the encoder of Wyner-Ziv frame, according to the SAD criteria, we can get the motion intense regions (MIRs) and the nonmotion intense regions (NMIRs). For the MIR, the side information is generated by fusion temporal side information and inter-camera side information at the pixel level. Inversely, for the NMIR, we directly use

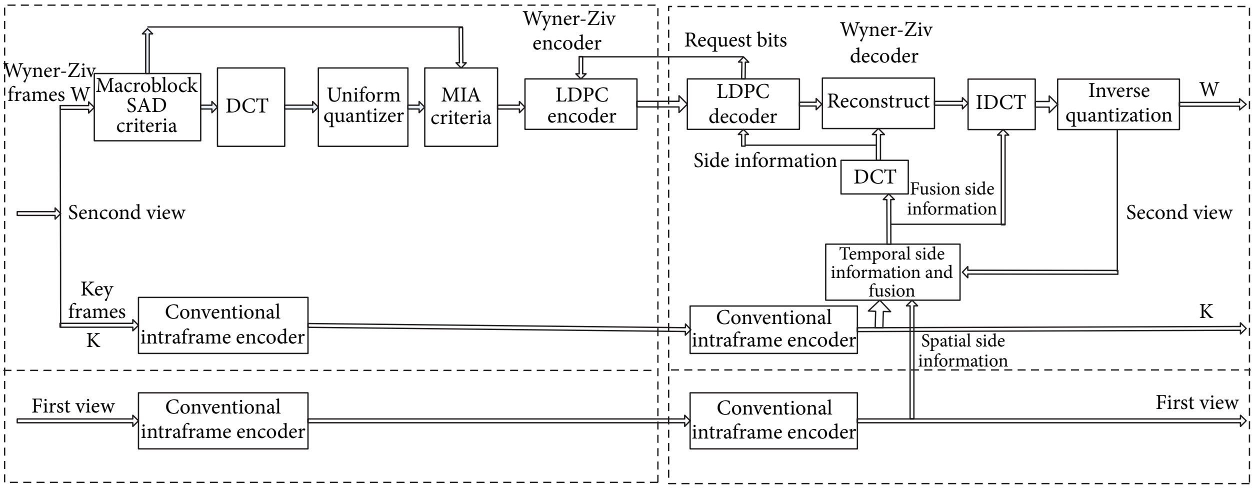


Figure 5: Block diagram of the encoder/decoder of the proposed MDVC Scheme based on SAD.

the temporal side information. Encoder uses SAD criteria to get MIR and NMIR macroblock and mark the MIR macroblock. Suppose that M_w is the current frame and M_p is the previous reference frame, so SAD of a macroblock is computed as (1). Consider

$$\mathrm{SAD} = \sum_{(x,y)\in B_i} \left| M_w(x,y) - M_p(x,y) \right|. \tag{1}$$

B_i represents an 8 × 8 macroblock of the image, x, y is, respectively, horizontal and vertical coordinates of a pixel. If $\mathrm{SAD} \geqslant T$, we regard the macroblock as MIR. An adequate threshold T has been found experimentally.

The detailed block diagram of the encoder/decoder of the proposed MDVC Scheme based on SAD criteria is illustrated in Figure 5. We use two video sensor views to present the encoding and decoding process of Wyner-Ziv frames for brevity. The K frames of the second view are encoded and decoded by conventional intraframe scheme, while for the WZ frames, according to the SAD criteria, we get MIR (8 × 8 macroblock) and NMIR (8×8 macroblock). In encoder, mark the MIR and NMIR macroblock, respectively. The K frame in first view is used conventional intracoding mode, and the decoded video stream is transformed into the interview SI for WZ frame of the second view by homography compensated interview interpolation (HCII). Side information is used in the reconstruction, to obtain the decoded DCT coefficients; then, IDCT and IQ (inverse quantization) are applied to generate the decoded frame W', as the structure of Wyner-Ziv codec shown in Figure 2.

3.4. Side Information Generation by Fusing. Unlike DVC, the generation of SI in MDVC must synthetically consider the temporal SI and interviews spatial SI. In this paper, for the temporal SI, we choose to use motion compensated temporal interpolation (MCTI) technique estimating temporal motion vector from the previous frame towards the forward frame. Then, the motion vectors are interpolated at midpoint to generate the SI [13, 15].

For the interviews spatial SI, in the scene of WVSN, all nodes synchronized monitor the same scenario in an area of interest, and the video sensor network contains a large number of interview statistical dependencies. The position and view of the cameras have been fixed; homography matrix [20] could be used to generate the interview SI from other cameras. The homography is a 3 × 3 matrix that relates video sensor 1 to video sensor 2 in the homogenous coordinates system. There are 8 parameters in matrix, we use a gradient descent method to get the parameters [21]. The pixel point of video sensor 1 is mapped to a pixel point (x_2, y_2) of video sensor 2 up to a scale μ such that

$$\mu \begin{bmatrix} x_2 \\ y_2 \\ 1 \end{bmatrix} = \begin{bmatrix} b_0 & b_1 & b_2 \\ b_3 & b_4 & b_5 \\ b_6 & b_7 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ y_1 \\ 1 \end{bmatrix},$$

$$x_2 = \frac{b_2 + b_0 x_1 + b_1 y_1}{b_6 x_1 + b_7 y_1 + 1}, \tag{2}$$

$$y_2 = \frac{b_5 + b_3 x_1 + b_4 y_1}{b_6 x_1 + b_7 y_1 + 1}.$$

Harris corner and edge detector [22] also could be used in iteratively calculating and adjusting the parameters in homography of two video sensor nodes. Then, we can use homography to generate the spatial SI of the Wyner-Ziv frame in adjacent sensor view.

When the temporal SI and spatial SI of a preparing Wyner-Ziv frame have generated, we present a binary masks fusion method to generate SI according the MIR or NMIR type of macroblocks transmitted from encoder. Figure 6 shows how to fuse the binary masks from two interview sensors, which is done by a simple logical OR operation. We look for the pixel that predicts Wyner-Ziv frame better from both side information. In binary mask, 0 represents that the pixel from the same position of interview SI is reliable and 1 represents that the pixel from the same position of temporal SI is reliable. In the fusion process of the MIR, we take a simple difference operation between the current

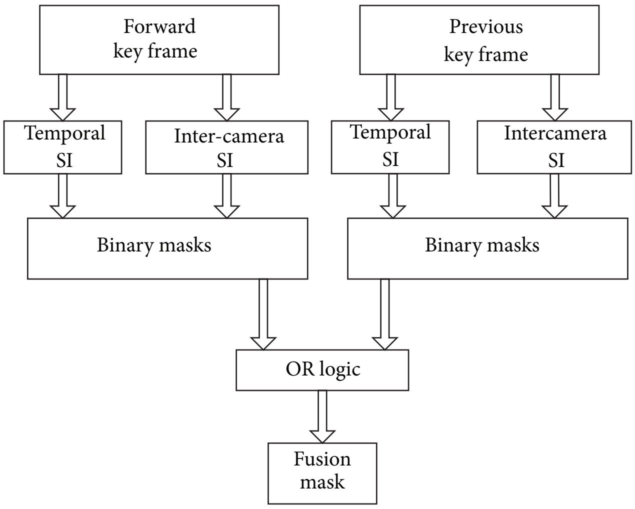


Figure 6: Side information fusion of Wyner-Ziv frame in decoder.

pixel from the previous key frame and the pixel an the same position from interview SI to obtain a value of A and take a simple difference operation between the current pixel from the previous key frame and the pixel at the same position from Temporal side information to obtain a value of B. If $|A| \geqslant |B|$; we set binary mask to 1; otherwise, we set binary mask to 0. For the NMIR, we set binary mask to 0, and it save a lot of computing time. We perform the same process with forward key frame. Thus, we obtain a second binary mask. Finally, we perform an OR logic operation between both binary masks of MIR to obtain the binary fusion mask.

4. Decoded Image Postprocessing Based on Spatial Domain

In Wyner-Ziv DVC and MDVC, quantification is a significant step. But the quantization matrix or quantization coefficient adopted in quantization process would affect the quality of image decoding. In image processing, the most common divided blocks processing technique, discrete cosine transforms (DCT), and quantization could lead to coding artifacts, such as blocking and ringing artifacts. In this section, we try to conduct the image postprocessing for the video decoded from the proposed MDVC by using filtering techniques.

4.1. Deblocking Filter. The quantization process for DCT coefficients causes the blocking effect during image and video coding. Because the quantization error is different in each block, the coding based on blocks, for example, 4 × 4, 8 × 8 (which is adopted in this paper), or 16 × 16, brings to the discontinuity in the border of adjacent blocks. In video coding, motion compensation would spread the blocking effect to the prediction coding frames. This is because the location of blocking effect is unfixed, and it would move with the motion compensation. The key technique of blocking effect eliminating is how to eliminate the blocking effect with protecting the real edges in the border of blocks. In general, the blocking effect eliminating algorithm should just handle the pixels near the border of block.

The algorithm of deblocking filter along the border of the 8 × 8 macroblock executes one-dimension filtering in the decoder. First, using horizontal filter eliminate the horizontal blocking effect; then, using vertical filter eliminates the vertical blocking effect. If a pixel had changed in the last filtering process, the modified pixel would be used in the next filtering. The filtering process has three modules: mode selection, DC offset mode for smooth area, and default mode for complex area.

To eliminate the blocking effect more powerful, in this paper, according the satisfied condition of near pixels at the border of 8 × 8 blocks, different filtering patterns are adopted in filtering process: DC filter of strong filtering for smooth area and default mode for pixels in the border of blocks for complex area.

Suppose that $p_0, p_1, \ldots, p_9$ are ten consistent pixels in horizontal direction, and the value of eq_cnt denotes the smooth level of the near image of blocks border; and T_1 is a small threshold, we have

$$\text{eq\_cnt} = \sum_{i=0}^{8} \phi\left(p_i - p_{i+1}\right), \qquad (3)$$

where

$$\phi(\sigma) = \begin{cases} 1 & \text{if } |\sigma| \leqslant T_1, \\ 0 & \text{otherwise.} \end{cases} \qquad (4)$$

If present pixels $p_0, p_1, \ldots, p_9$ are in smooth area, the value of eq_cnt would be relatively great. On the contrary, if the present pixels are in complex area, the value of eq_cnt would be very small. We employ (5) to determine the desired filtering mode, DC offset mode or default mode, denoted by filter_mode. Consider

$$\text{filter\_mode} = \begin{cases} \text{DC\_offset} & \text{if eq\_cnt } T_2, \\ \text{default} & \text{otherwise.} \end{cases} \qquad (5)$$

Suppose that T_2 is a threshold larger than T_1. Examining the blocking effect areas cause by little DC offset. If eq_cnt $\geqslant T_2$, the blocking area is smooth area and DC filtering mode is adopted. Inversely, the blocking area is complex area and default filtering mode is selected.

4.2. Deringing Effect Filter. The coarse quantification of the high-frequency components of image results in ringing effect. If high-frequency components, which corresponding to the strong edges of image, such as high contrast, occur quantization error, the nearby region of the strong edges would appear to be fake edge. The core of the ringing effect eliminating algorithm is how to distinguish the ringing and real edge. The most common method is examining the edge of image first, then using low-pass filter screening the no-edge pixels to achieve the objective of eliminating ringing effect.

The deringing effect filter includes threshold determination, identifier evaluation and adaptive smoothing. Every pixel in 8 × 8 macroblock all needs to filter. In fact, 10 × 10 pixels are used in the process of every 8 × 8 macroblock filtering.

In threshold determination module, first, we divide the 8×8 macroblock to 4 smaller blocks, then find the maximum and minimum grey value max[i] and min[i] of the divided ith block. Calculate the threshold thr[i] and the dynamic range of grey range[i] by (6) and (7), respectively. Consider

$$\text{thr}\,[i] = \frac{\max\,[i] + \min\,[i] + 1}{2}, \qquad (6)$$

$$\text{range}\,[i] = \max\,[i] - \min\,[i]\,. \qquad (7)$$

Calculate the dynamic range of 4 luminance blocks, and number the block, which has maximum dynamic range; as $i_{\max}$, we have

$$\max\_\text{range} = \text{range}\left[i_{\max}\right]. \qquad (8)$$

Revise the threshold of 4 smaller blocks by (9). Consider

$$\underset{i\in[0,3]}{\text{thr}[i]} = \begin{cases} \text{thr}\left[i_{\max}\right] & \text{if } \left(\text{range}\,[i] < 32\&\&\max\_\text{range} \geqslant 64\right), \\ 0 & \text{if } \max\_\text{range} < 64. \end{cases} \qquad (9)$$

In identifier evaluation module; after threshold determined, the following operations are all in original 8 × 8 macroblock. The grey value and binary identifier of the pixel (h, v) are denoted by rec(h, v) and bin(h, v), respectively. The determination criterion of binary identifier is shown in (10). Consider

$$\underset{h,v\in[0,7]}{\text{bin}(h,v)} = \begin{cases} 1 & \text{if rec}\,(h,v) \geqslant \text{thr}, \\ 0 & \text{otherwise}, \end{cases} \qquad (10)$$

where thr is the threshold of the current block i. The purpose of binary identifier processing is to distinguish the grey value with the range with thr in block.

Adaptive smoothing module is constituted by adaptive filtering and numerical pruning two parts. Calculate the binary identifiers of all the pixels in the 3×3 window, whose center is current pixel. If all binary identifiers are "1" or "0", this region is smoothed area, so to filter this current pixel by smooth filtering. The reconstructed value after filtering flt$'(h, v)$ is computed by (11). Consider

$$\text{flt}'\,(h,v) = \frac{1}{16}\left[\sum_{i=-1}^{1}\sum_{j=-1}^{1}\text{coef}\left(i,j\right)\cdot\text{rec}\left(h+i,v+j\right)+8\right], \qquad (11)$$

where coef(i, j) denotes the coefficient, $i, j = -1, 0, 1$, (i, j) represents the location of pixel in 3 × 3 image window.

In order to prevent excessive handling of pixel, the gray level difference dif(h, v) between the reconstructed values with smooth filtering and the original pixel value must change limiting in the range from the change of quantization parameters. fit$''(h, v)$ and fit$'(h, v)$ indicate the reconstructed value with smooth filtering and the without limited reconstructed value pixel value. We have

$$\text{flt}''\,(h,v) = \begin{cases} \text{rec}\,(h,v) + \max\_\text{dif} & \text{if } \left(\text{dif}\,(h,v) > \max\_\text{dif}\right), \\ \text{rec}\,(h,v) - \max\_\text{dif} & \text{if } \left(\text{dif}\,(h,v) < \max\_\text{dif}\right), \\ \text{flt}'\,(h,v) & \text{otherwise}, \end{cases} \qquad (12)$$

where OP denotes quantization parameter and max_dif is OP/2 no matter the macroblock in internal coding or non-internal coding.

5. Performance Evaluation

In this section, we design some simulations to evaluate the effectiveness of the proposed coding and postprocessing framework by using public and representative multiview video sequences.

We use the two multiview sequences: "Exit" and "Vassar" which are made public available by Mitsubishi Electric Research Laboratories (MERL) [23]. For reasons of computation complexity, the spatial resolution was halved from VGA (640×480, YUV 4 : 2 : 0) to QCIF (176×144, YUV 4 : 2 : 0). For both, the time resolution is 25 fps and we used the 2 cameras with 100 frames per view, but only luminance component is evaluated for per frame. Our simulated environment is Microsoft Visual Studio 2005 on Windows XP SP3 system with Intel Core 2 CPU 2.40 GHz and 2.00 GB memory. We extend the exiting Wyner-Ziv video codec [24] to MDVC, and all codes are written in ANSI C++. In the experiments, LDPC codes are generated by PEG algorithm [25], and the rate of LDPC code is 7/8. After several experimental analyses and comparisons, 64 is the ideal threshold of SAD criteria. According to the frame structure shown in Figure 3, the video streams of Camera 0 and Camera 1 all use distributed video coding, but the key frames and Wyner-Ziv frames coding sequence for Camera 0 and Camera 1 are "W-K-W-K" and "K-W-K-W", respectively. The key frames K and Wyner-Ziv frames W are alternative coding, and the key frame of Camera 1 is used to assist the Wyner-Ziv frame decoded in Camera 0, which is main perspective in our experiments.

To evaluate the rate-distortion performance of our proposed scheme, we compare the following five schemes:

(1) H.263+ "I-I-I-I": H.263 intraframe coding (I-I-I-I). H.263+ codec uses TMN8 [26].

(2) H.263+ "I-P-I-P": H.263 interframe coding (I-P-I-P). Like H.263+ "I-I-I-I" scheme, the codec used also is TMN8, but in different coding options.

(3) Only Temporal SI: MDVC with only temporal SI used in side generation.

(4) Only Temporal SI: MDVC with only spatial SI used in side generation.

(5) Fusion SI: MDVC with fusion temporal and spatial SI used in side generation.

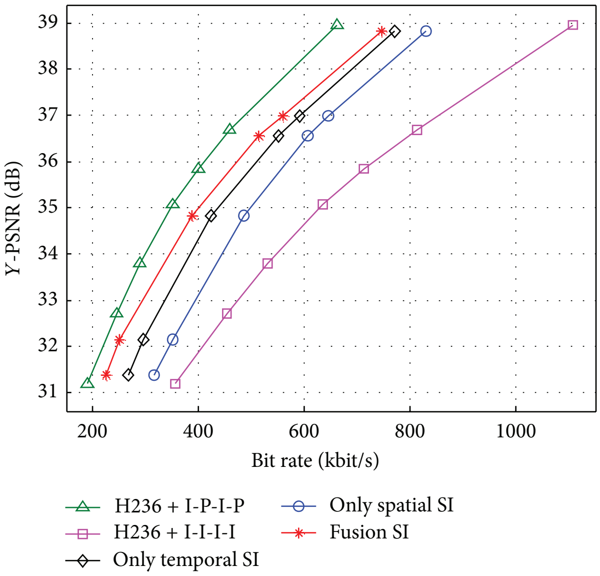


FIGURE 7: Luminance PSNR versus average bit rate for different coding schemes of "Exit" multiview sequences.

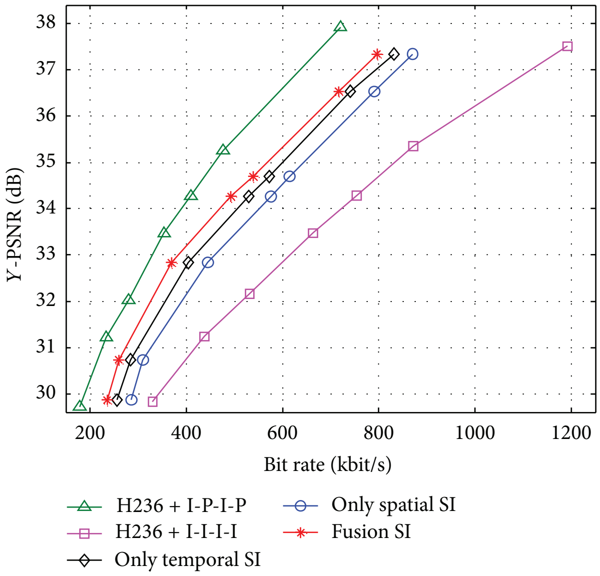
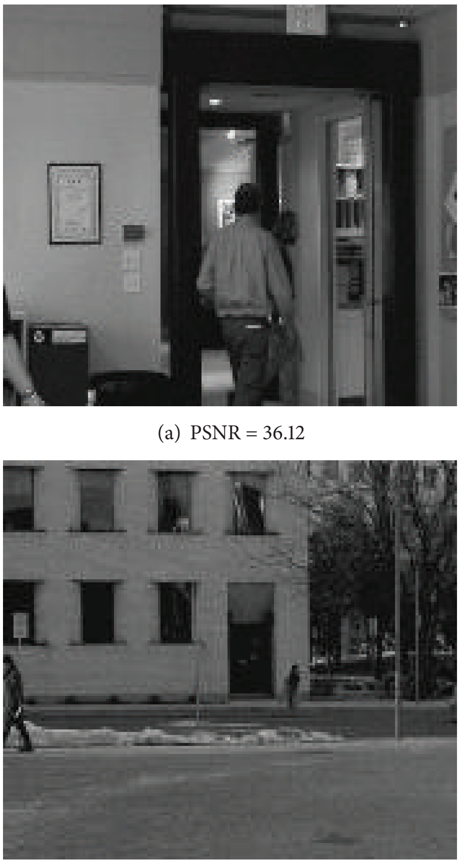


FIGURE 8: Luminance PSNR versus average bit rate for different coding schemes of "Vassar" multiview sequences.

The luminance PSNR performance as a function of the different average bit rate for the two multiview sequences "Exit" and "Vassar" is shown in Figures 7 and 8, respectively. We can find that the proposed MDVC scheme has significantly better performance 2-3 dB than that of H.263 intraframe coding, and MDVC system is less than the overall complexity of the H.263+ coding. Although there is still a performance gap between H.263 interframe coding and our proposed MDVC scheme, our scheme is easier at encoder which is fit for WVSNs. The proposed fusion SI approach can bring improvements up to 0.2–0.5 dB when compared to the MDVC with only temporal SI used in the same average bit rate. Our proposed MDVC scheme can gain more accurate motion estimation in the intense motion region, to save a lot of computing time. Figures 9(a) and 9(b), respectively, is 15th decoded frame (WZ frame) of "Exit" sequence and "Vassar" sequence using our proposed MDVC scheme; Figures 10(a) and 10(b), respectively, is 15th decoded frame (WZ frame) to "Exit" sequence and "Vassar" sequence using temporal SI. From the comparisons, it can be seen that the decoded frame in fusion SI in MDVC is significantly better than nonfusion scheme, for instance, only temporal SI scheme.

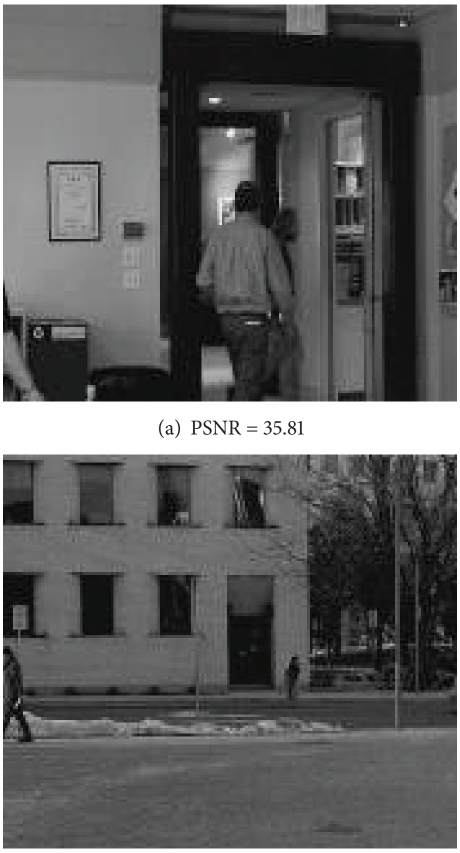


(a) PSNR = 36.12

(b) PSNR = 34.82

FIGURE 9: The decoded 15th frame in proposed fusion SI scheme. (a) "Exit". (b) "Vassar".

For evaluating the performance of postprocessing framework proposed in this paper, we set up a comparison experiments with the postprocessing and without postprocessing for the decoded multiview sequences. The comparison results of average YUV-PSNR versus QP with decoded "Exit" and "Vassar" sequences are shown in Figure 11. We can find that the postprocessing method could provide an additional improvement of about 0.1 dB to the decoded video sequences from Figure 11. The decoded video after the postprocessing filter can be reference data in the subsequently image processing and is useful to enhance the compression image quality in objective and subjective.



(a) PSNR = 35.81

(b) PSNR = 34.49

Figure 10: The decoded 15th frame in only Temporal SI scheme. (a) “Exit”. (b) “Vassar”.

6. Conclusions and Future Work

In this paper, we proposes an improved coding and postprocessing framework for multiview distributed video in WVSNs. In coding scheme, through distinguishing the motion intense regions and the nonmotion intense regions based on sum of absolute difference criteria, the coding scheme encodes macroblocks adaptively, and a fusion temporal and spatial side information is adopted to improve the quality of side generation in the decoder. To further enhance the quality of decoded video sequences, a postprocessing scheme is designed to get more additional compression gain. Experiments demonstrate the validity of the proposed framework.

In the postprocessing scheme, we enforce the image postprocessing based on spatial domain without considering time relevance of the previous and next frames, so the image quality gain we got in this scheme is limited. In our future work, we would consider talking about the postprocessing based on temporal filtering to guarantee the temporal continuity in the reconstructed video sequences to get higher video quality and more compression gain.

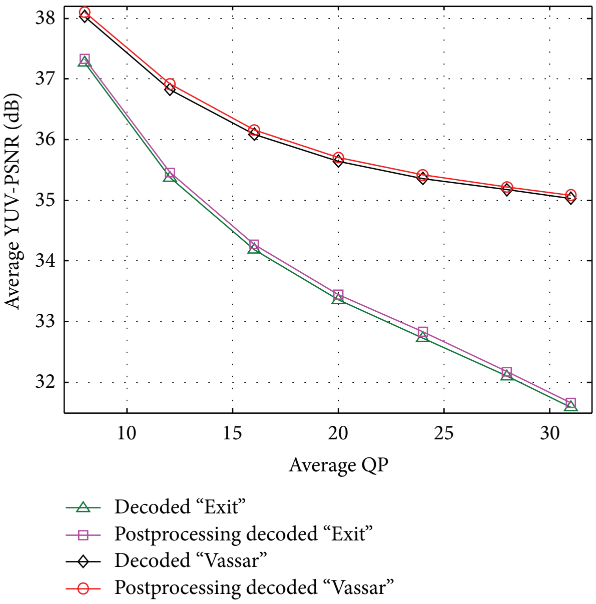


Figure 11: The decoded 15th frame in only Temporal SI scheme. (a) “Exit”. (b) “Vassar”.

Acknowledgments

This work is supported in part by the National Natural Science Foundation of China under Grant nos. 61171053, 61300239, and 61373137; the Doctoral Fund of Ministry of Education of China under Grant no. 20113223110002; the Science & Technology Innovation Fund for Higher Education Institutions of Jiangsu Province under Grant no. CXZZ12_0480.

References

[1] T. Wiegand, G. J. Sullivan, G. Bjøntegaard, and A. Luthra, “Overview of the H.264/AVC video coding standard,” *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 13, no. 7, pp. 560–576, 2003.

[2] S. Saponara, C. Blanch, K. Denolf, and J. Bormans, “The JVT advanced video coding standard: complexity and performance analysis on a tool-bytool basis,” in *Proceedings of the International Packet Video Workshop (PV ’03)*, Nantes, France, April 2003.

[3] J. Ostermann, J. Bormans, P. List et al., “Video coding with H.264/AVC: tools, performance, and complexity,” *IEEE Circuits and Systems Magazine*, vol. 4, no. 1, pp. 7–28, 2004.

[4] R. Puri and K. Ramchandran, “PRISM: a new robust video coding architecture based on distributed compression principles,” in *Proceedings of the Allerton Conference on Communication, Control and Computing*, Allerton, Ill, USA, October 2002.

[5] A. Aaron, R. Zhang, and B. Girod, “Wyner-Ziv coding of motion video,” in *Proceedings of the 36th Asilomar Conference on Signals Systems and Computers*, pp. 240–244, Pacific Grove, Calif, USA, November 2002.

[6] D. Slepian and J. K. Wolf, “Noiseless coding of correlated information sources,” *IEEE Transactions on Information Theory*, vol. 19, no. 4, pp. 471–480, 1973.

[7] A. D. Wyner and J. Ziv, "The rate-distortion functions for source coding with side information at the decoder," *IEEE Transactions on Information Theory*, vol. 22, no. 1, pp. 1–10, 1976.

[8] F. Dufaux, W. Gao, S. Tubaro, and A. Vetro, "Distributed video coding: trends and perspectives," *EURASIP Journal on Image and Video Processing*, vol. 2009, Article ID 508167, 2009.

[9] D. Varodayan, A. Aaron, and B. Girod, "Exploiting spatial correlation in pixel-domain distributed image compression," in *Proceedings of the International Picture Coding Symposium (PCS '06)*, Beijing, China, April 2006.

[10] A. Aaron, D. Varodayan, and B. Girod, "Wyner-Ziv residual coding of video," in *Proceedings of the International Picture Coding Symposium (PCS '06)*, Beijing, China, April 2006.

[11] R. Puri, A. Majumdar, and K. Ramchandran, "PRISM: a video coding paradigm with motion estimation at the decoder," *IEEE Transactions on Image Processing*, vol. 16, no. 10, pp. 2436–2448, 2007.

[12] Q. Xu and Z. Xion, "Layered Wyner-Ziv video coding," *IEEE Transactions on Image Processing*, vol. 15, no. 12, pp. 3791–3803, 2006.

[13] X. Guo, Y. Lu, F. Wu, W. Gao, and S. Li, "Distributed multiview video coding," in *Visual Communications and Image Processing (VCIP)*, vol. 6077 of *Proceedings of SPIE*, San Jose, Calif, USA, January 2006.

[14] M. Flierl and B. Girod, "Coding of multi-view image sequences with video sensors," in *Proceedings of the IEEE International Conference on Image Processing (ICIP '06)*, pp. 609–612, Atlanta, Ga, USA, October 2006.

[15] M. Ouaret, F. Dufaux, and T. Ebrahimi, "Multiview distributed video coding with encoder driven fusion," in *Proceedings of the 15th European Signal Processing Conference (EUSIPCO '07)*, Poznan, Poland, September 2007.

[16] P. N. Huu, V. Tran-Quang, and T. Miyoshi, "Video compression schemes using edge feature on wireless video sensor networks," *Journal of Electrical and Computer Engineering*, vol. 2012, Article ID 421307, 20 pages, 2012.

[17] L.-W. Kang, C.-S. Lu, and C.-Y. Lin, "Low-complexity video coding via power-rate-distortion optimization," *Journal of Visual Communication and Image Representation*, vol. 23, no. 3, pp. 569–585, 2012.

[18] P. L. Dragotti and M. Gastpar, *Distributed Source Coding: Theory, Algorithms and Applications*, Academic Press, New York, NY, USA, 2009.

[19] H. Bai, A. Wang, Y. Zhao, J.-S. Pan, and A. Abraham, *Distributed Multiple Description Coding: Principles, Algorithms and Systems*, Springer, London, UK, 2011.

[20] D. A. Forsyth and J. Ponce, *Computer Vision: A Modern Approach*, Prentice Hall, Upper Saddle River, NJ, USA, 2nd edition, 2011.

[21] F. Dufaux and J. Konrad, "Efficient, robust, and fast global motion estimation for video coding," *IEEE Transactions on Image Processing*, vol. 9, no. 3, pp. 497–501, 2000.

[22] C. Harris and M. Stephens, "A combined corner and edge detector," in *Proceedings of the 4th Alvey Vision Conference*, pp. 147–151, Manchester, UK, August 1988.

[23] Mitsubishi Electric Research Laboratories, "MERL multiview video sequences," ftp://ftp.merl.com/pub/avetro/mvc-testseq/.

[24] D. Chen, D. Varodayan, M. Flierl, and B. Girod, "Wyner-Ziv video codec with unsupervised motion learning," http://msw3.stanford.edu/~dchen/DVC/.

[25] X.-Y. Hu, E. Eleftheriou, and D. M. Arnold, "Regular and irregular progressive edge-growth tanner graphs," *IEEE Transactions on Information Theory*, vol. 51, no. 1, pp. 386–398, 2005.

[26] Video codec test model, ITU-T/SG15, TMN8, Portland, Ore, USA, June 1997.

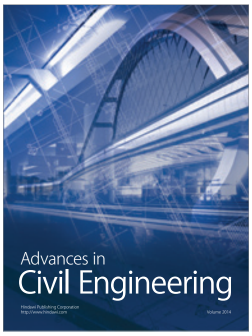
Journal of
Engineering

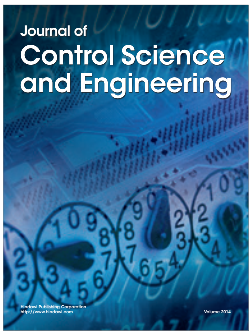
The Scientific World Journal

International Journal of
Rotating Machinery

Journal of
Sensors

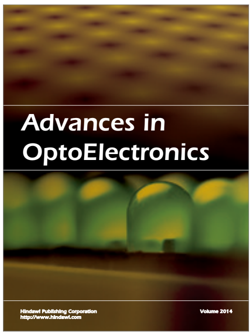
International Journal of
Distributed Sensor Networks


Advances in
Civil Engineering


Journal of
Control Science and Engineering

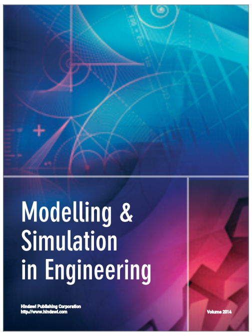
Journal of
Robotics

Journal of
Electrical and Computer Engineering

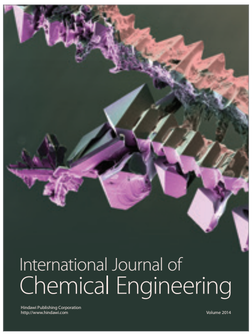

Advances in
OptoElectronics

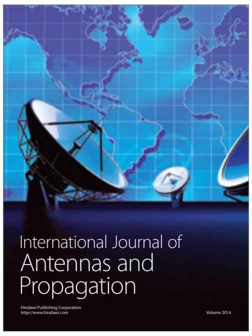
VLSI Design

International Journal of
Navigation and Observation

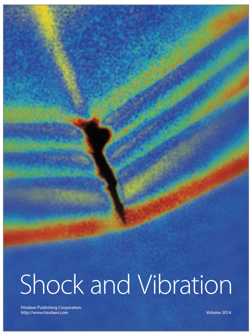

Modelling & Simulation in Engineering

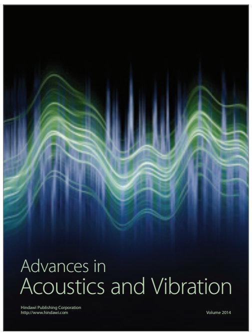
International Journal of
Aerospace Engineering


International Journal of
Chemical Engineering


International Journal of
Antennas and Propagation

Active and Passive Electronic Components


Shock and Vibration


Advances in
Acoustics and Vibration