Multiple Description Coding for Video Transmission over Wireless Channels

Nikola Bozinovic

09-01-2001

Boston University

Department of Electrical and Computer Engineering

Technical Report No. ECE-2001-01

BOSTON UNIVERSITY

Multiple Description Coding for Video Transmission over Wireless Channels

Nikola Bozinovic

Boston University
Department of Electrical and Computer Engineering
8 Saint Mary's Street
Boston, MA 02215
`www.bu.edu/ece`

09-01-2001

Technical Report No. ECE-2001-01

Summary

The objective of multiple description coding (MDC) is to encode a source into multiple bitstreams such that a decoder which receives an arbitrary subset of the bitstreams may produce a useful reconstruction. Because of these properties, the technique is of special interest in video communications over error-prone channels, like wireless or Internet. Most of the recent MDC proposals make use of spatial redundancy in a video sequence to generate multiple bitstreams.In this report we investigate possibilities for the use of multiple descriptions generated from time-separated group of frames (GoF). We argue that the loss of frames due to the loss of one description could be compensated for by using other descriptions. The idea is to reconstruct missing frames from the neighboring GoFs by using the motion smoothness property. We perform characterization of videoconferencing sequences and investigate possibilities for the application of this technique to video transmissions over wireless channels.

Contents

List of Figures

1 Introduction

In this report, we present an idea for a new multiple description coding strategy for robust video communication over unreliable channels. Examples of such channels are wireless channel and the best-effort packet networks, such as today's Internet. Other typical scenario might require data to move from a fiber link to a wireless link, which necessitates dropping packets to accomodate the lower capacity of the latter. In our work we will concentrate on the wireless transmission although principles and techniques described here can be easily deployed in other networks with similar channels properties.

Main source of errors in the transmission over a wireless channel are fading and multipath. Since such errors are generated in bursts, large segments of the transmitted data may be lost or useless. This severely degrades the received signal quality if the missing data is not recovered.

Traditional video compression algorithms are not robust to transmission errors. The sole objective of compression is to maximize coding gain, assuming error-free channels. Most video coding schemes rely on temporal-difference coding to achieve coding efficiency, thereby introducing a pervasive dependency structure into a bit stream. Hence, losses due to dropped packets or late packet arrivals result in the loss of subsequent dependent frames, leading to visual artifacts that can be long lasting and annoying.

Existing techniques to recover the lost data or mitigate the loss impact include ARQ retransmission [1], FEC using error-correcting codes [2] and receiver reconstructions using only the received data by exploiting the residual correlation in the encoded data [3, 4].

For delay-constrained applications like real time video or multicast applications (e.g., multiparty teleconference), ARQ would obviously not be an appropriate choice. For large bursts of bit erasures, error-correcting codes, such as block codes and convolutional codes, cannot provide sufficient protection without excessive delay and computation.

Recently, multiple description coding techniques have been shown to be effective in protection against channel failures. Three examples are: Multiple Description Scalar Quantizer (MDSQ) [5, 6, 7], Multiple Description Transform Coding [8, 9] and Multiple Description via Polyphase Transform (MDTP) [10]. Robustness to channel errors in these systems is achieved at the expense of relatively large system complexity. For example, MDSQ requires careful index assignments while MDTC necessitates another correlating transform besides the conventional decorrelating transform.

In this report, we propose a scheme where multiple bitstreams are generated from groups of frames (GoF) separated in time. This coding scheme is based on time division multiplexing of the original stream where each description is generated from certain number of consecutive frames. At the decoder, different streams are used alternately to generate final video sequence. We argue that the loss of frames due to the loss of one description could be compensated for based on neighboring descriptions by using the motion smoothness property.

This report is organized as follows: in the next section we give an overview of current state-of-the-art wireless networks, describe types of transmission standards and characterize error patterns typical for wireless channels. The overview of different multiple description coding schemes proposed recently is given in Section 3. In Section 4, we propose our method for generating multiple bitstreams that make use of motion smoothness property of the video signal. In Section 5 we present initial results in characterization of videoconferencing sequence for use in such a scheme. We draw conclusions in Section 6.

2 Overview of Current Wireless Technology

The ability of a handheld device to send and receive voice has been used since the late 70's. The wireless communication was restricted to point-to-point speech communication and primarily used in specialized services like army or police. Since the introduction of a commercial mobile users network in late 80's, the number and popularity of cellular phones was growing with unprecedented speed, when compared to any other electronics device. In the last couple of years, the attempts of wireless data transfer started to draw attention of the users, although their success was less than expected, primarily due to low data-rates. Today, wireless networks achieve transmission speeds between 14.4 Kb/s and 19.2 Kb/s, depending on the underlying network technology.

New standards that are proposed for multipurpose wireless transmission promise to change the way we are communicating by introducing new multimedia content to mobile users. The third-generation wireless technology (3G or IMT-2000) is expected to support data rates from 384Kb/s to 2 Mb/s, which is sufficient for video applications. On the other hand, new video coding schemes promise to offer better compression, higher resistance to errors and interference, and more flexibility. All this should greatly improve current experience with video over wireless channels, and provide whole new world of communications.

Current standards for wireless transmission networks include GSM and IS-95 ($\sim 10kb/s$) as two most popular standards in widespread cellular networks, and WCDMA and cdma-2000 as emerging standards for 3G personal communication networks ($\sim 100kb/s$).

Wireless channel is known to be error prone, which seriously limits the data rates that can be achieved. Probability of error in the personal communication systems is typically between 10^{-3} to 10^{-5}, which is couple of orders of magnitude worse then performance of the wired systems. In a few cases, a radio that will be guaranteed line of sight and that will always be indoors with static conditions and shielding from external EM sources will not need special protection, but in most wireless environments, it becomes necessary to have some form of error correction coding if reliability is required.

In some typical cases for a channel with a high capacity and certain retransmission scheme, low bit error rate (10^{-5} and lower) is needed to get any data across the channel if there is a need for data to be received intact. For example, for high-rate channels (1 Mb/s or more) and bit error rates on the order of 10^{-5}, the throughput of TCP/IP begins dropping dramatically and becomes almost zero at 10^{-4}. A successful coding scheme for Video over Wireless (VoW) transmission should be able to keep the visual quality of the transmitted stream at a satisfactory level even in the presence of high number of successive errors, which are typical of wireless transmission. Here we present a list of typical problems that can occur in a wireless channel, all of which produce similar error patterns in the signal:

- Raleigh fading, also known as "shadow fading", occurs when the receiver is starting to move behind a large object. Essentially, the receiver is moving into a shadow from the EM waves coming from the transmitter. Ordinarily Raleigh fading only occurs in a mobile environment.
- Ricean effects are those due to scattering and multipath. Even if the receiver and transmitter are immobile, the position of bodies around them, such as rain and cars, can significantly affect the channel.
- Co-channel interference occurs when two transmitters in close proximity are transmitting, and their signals are interfering with one another.

Typical channel models used in the analysis of wireless transmission include memoryless channel, symmetric channel, additive white Gaussian noise (AWGN) channel and bursty channel. Here we give a more detailed description of each of these models:

- Memoryless channel is also known as a random bit error channel and is characterize by error that is independent from one symbol to the next.
- Symmetric channel - the probability of a transmitted symbol value i being received as a value j is the same as that of a transmitted symbol j being received as i, for all values of i and j. A commonly encountered example is the binary symmetric channel (BSC) with a probability p of bit error. If the channel is BSC, this allows certain assumptions to be made that allow codes to be more efficient.
- Additive white Gaussian noise (AWGN) channel is a memoryless channel in which the transmitted signal suffers from the addition of wide-band noise whose amplitude is a normally distributed random variable. AWGN is the most common form of a memoryless channel.
- Bursty channel is the channel where errors are characterized by periods of relatively high symbol error rate separated by periods of relatively low, or zero, error rate. A burst error means that the probability of error is dependent from one symbol to the next.

It is the bursty nature of the wireless channel that makes it hard for standard "layered" video coding schemes (e.g., MPEG-2) to be easily and efficiently deployed. Our effort in generating multiple descriptions from successive GoF is directly motivated by this typical error pattern, and represents a way to minimize the impact of successive errors in the channel.

3 Basic Principles of Multiple Description Coding

From the reasons we mentioned in the previous section, we can conclude that for proper transmission of video sequence over an unreliable channel, strategies that add error resilience to coding algorithms have to be used. Coding algorithms are almost always used in video transmission because video data has ample redundancies and is compressed before it is sent. There are two types of non-redundant robust coding algorithms that are resilient to errors in transmission:

1. In layered coding, data is partitioned into a base layer and a few enhancement layers. The base layer contains visually important video data that can be used to produce video output of acceptable quality, whereas the enhancement layers contain complementary information that allows higher-quality video data to be generated. In networks with priority support, the base layer is normally assigned a higher priority so that it has a larger chance to be delivered error free when network conditions worsen. Layered coding has been popular with ATM networks but may not be suitable for wireless transmission for two reasons. First, it is impossible to provide priority deliveries for different layers when using wireless transmission. Second, when the packet-loss rate is high and part of the base layer is lost, it is hard to reconstruct the lost data since no redundancy is present.

2. Multiple description coding divides video data into "equally" important streams such that the decoding quality using any subset is acceptable, and that better quality is obtained by more descriptions . It is assumed in MDC that the probability of losing all the descriptions is very low. As we mentioned earlier, to date MDC has been implemented in several ways. Here we describe three proposed MDC schemes:

- A scalar-quantizer (MDSQ) [5, 6, 7] applies two side scalar quantizers in order to produce two descriptions. Multiple description quantizers can conceptually be seen as the use of a set of independent scalar (or vector) quantizers to obtain a number of descriptions of a source sample (vector). Each description is then transmitted (in packets) to the receiver using as many channels as descriptions. The channel is assumed to only introduce packet loss errors. At the receiver the source sample (vector) is reconstructed by combining the descriptions arrived. The more descriptions that arrive at the receiver the lower the distortion between the original and reconstructed source sample (vector) becomes. In order to minimize reconstruction errors when both descriptions are received, system then maps a proper subset of index pairs formed from side quantizers to central-quantizer intervals. The difficulties with this approach are that optimal index assignments are hard to achieve in real time, and that suboptimal approaches, such as A2 index assignment [5], introduce a large overhead in bit rate [11].

- Multiple description transform coding (MDTC) proposed in [9] produces statistically correlated streams such that lost streams can be estimated from the received data. Authors propose two techniques: Square Correlating Transforms and Overcomplete Frame Expansions.

  In the method of Square Correlating Transforms, a block of n independent, zero mean variables with different variances are transformed to a block of n transform coefficients in order to create a known statistical correlation between transform coefficients. The transform coefficients from one block are distributed to different packets so in the case of a packet loss, the lost coefficients can be estimated from the received coefficients. The redundacy comes from the relative inefficiency of scalar entropy coding on correlated variables.

  The coding of a source vector x proceeds as follows:

  1. x is quantized with a uniform scalar quantizer with step size $\Delta : x_{q_i} = [x_i]_\Delta$, where $[\cdot]_\Delta$ denotes rounding to the nearest multiple of Δ.

  2. The vector $x_q = [x_{q_1}, x_{q_2}, \cdots, x_{q_n}]^T$is transformed with an invertible, discrete transform $\widehat{T} : \Delta Z^n \rightarrow \Delta Z^n, y = \widehat{T}(x_q)$.

  3. The components of y are independently entropy coded.

  Descrete transform $\widehat{T}$ and continuous tranform T are related through "lifting":

$$\widehat{T}(x_q) = \left[T_1 \left[T_2 ... \left[T_k x_q\right]_\Delta\right]_\Delta\right]_\Delta$$

  The lifting structure ensures that the inverse of $\widehat{T}$ can be implemented by reversing the calculations:

$$\widehat{T}^{-1}(y) = \left[T_k^{-1} \cdots \left[T_2^{-1} \left[T_1^{-1} y\right]_\Delta\right]_\Delta\right]_\Delta$$

  When all the components of y are received, the reconstruction process is to (exactly) invert the transform $\widehat{T}$ to get $\widehat{x} = x_q$. The distortion is precisely the quantization error from Step 1. If some components of y are lost, they are estimated from the received components using the statistical correlation introduced by the transform $\widehat{T}$. The estimate $\widehat{x}$ is then generated by inverting T.

  Authors also give the optimal design of the transform $\widehat{T}$ for Gaussian sources, where arbitrary packet loss probabilities are allowed.

In contrast to statistical redundancy in the described method, in the technique of Overcomplete Frame Expansions a deterministic redundancy is used between descriptions. A linear transform from $\Re^k$ to $\Re^n$, followed by scalar quantization is used to generate n descriptions of a k-dimensional source. These n descriptions are such that a good reconstruction can be computed from any k descriptions, but also descriptions beyond of the kth are useful and reconstructions from less than k descriptions are easy to compute (see [9] and references therein for algebraic details).

- Multiple Description via Polyphase Transform (MDTP) proposed in [10], explicitly separates description generation and redundancy addition which reduces the implementation complexity, especially for systems with more than two descriptions. The system proposed realizes a Balanced Multiple Description Coding (BMDC) framework that can generate descriptions of statistically equal rate and importance. For a given total coding rate, the problem of optimal bit allocation between source coding and redundancy coding is solved in order to achieve the minimum average distortion for different channel failure rates.

3.1 Multiple Description Coding Terminologies

Here we introduce MDC definitions and then contrast MDC with two other coding techniques: Simulcast Coding (SC) and Layered Coding (LC).

Given an information source χ and distortion measure $d(\chi; \overline{\chi})$:

Two Description Coding involves finding two rate distortion codes $\{C_i, i = 1, 2\}$ such that C_1 achieves the rate distortion pair (R_1, D_1), C_2 achieves the rate distortion pair (R_2, D_2) and (C_1, C_2) achieves rate distortion pair $(R_1 + R_2, D_0)$ with $D_0 < D_1$ and $D_0 < D_2$.

Multiple Description Coding (MDC) involves finding multiple rate distortion codes $\{C_i, i = 1, 2, \ldots, M\}$ such that C_i achieves the rate distortion pair (R_i, D_i), any combinations of more than one code (total number $\sum_{k=2}^{M} \left(\begin{array}{c} M \\ k \end{array} \right)$) achieves smaller distortion (smaller than $\min\{D_i\}, i = 1, 2, \ldots, M$), and $(C_1, C_2, \ \ldots, C_M)$ achieves the global minimum distortion D_0 (rate distortion code $(\sum_{k=1}^{M} R_k, D_0)$).

Whereas the notation is somewhat cumbersome here, two important observations can be made:

1. Each code C_i is independently decodable since it has its own pair of encoding-decoding functions (f_i, g_i).

2. Each code C_i carries new information about the original source that indicates that the more codes are used for reconstruction, the smaller the overall distortion one can achieve. These are two features that distinguish MDC from SC and LC.

In an LC system, the source is encoded into multiple bitstream layers $\{L_0, L_1, \ldots, L_M\}$, which correspond to the multiple descriptions $(C_0, C_1, \ldots, C_M)$ in an MDC system. However, layers are usually not independent of each other. Higher layers can only be encoded/decoded after lower layers have been encoded/decoded. For example, L_2 is encoded and decoded with the help of L_0 and L_1. Obviously, for perfect channels, LC systems can achieve higher rate-distortion gain compared to MDC systems and are thus more appropriate for communication networks which provide delivery with different priority levels. However, LC systems are more susceptible to channel errors due to this inter-layer dependency on non-priority networks, such as wireless network or today's Internet.

In an SC system, the source is actually encoded into multiple bitstreams $C_0, C_1, \ldots, C_M$ that can be independently decoded to yield different reconstruction qualities. Normally, each code is specifically designed for a specific class of users. For example, C_0 is for users with the smallest bandwidth while C_M is for users with the largest bandwidth. However, codes do not usually complement each other but rather serve as independent bitsources so that a useful reconstruction is possible from any of the received bitstreams. In the case of reception of two or more bitstreams of a different quality level, final reconstruction quality is limited by the quality of "better" stream, and additional information from other bitstream won't result in improvement in overall reconstruction.

On the other hand, in MDC systems each C_i carries new information compared to codes upstream $C_0, C_1, \ldots, C_{i-1}$. This shows that, for example, (C_0, C_1) will only give reconstruction quality equal to that can be achieved by C_1 in a SC system, yet better reconstruction than that using only C_0 or C_1 in an MDC system. Clearly, this shows that MDC has a better rate-distortion gain compared to that of SC techniques. In summary, MDC is better than SC in the rate-distortion sense while it outperforms LC in the channel robustness sense at least in the case when no priorities exist.

4 The Proposed MDC System

As we have seen in the previous section, most of the proposed MDC schemes make use of spatial redundancy in a video sequence to generate multiple bitstreams. Here we describe a new system that exploits the motion smoothness property of a video sequence.

Consider a video sequence that consists of N frames, shown on Fig. 1. Basic coding scheme that uses only two descriptions and GoF of equal length looks like this: frames 1 to M are used to generate the beginning of the first description A, then frames $M+1$ to $2M$ are used to generate the beginning of the second description B, frames $2M+1$ to $3M$ are used to generate the next part of the description A and so on. Descriptions A and B then can be transmitted separately, over the same or different paths. If there is an error in the transmission, and part of one description is lost, the coding scheme should be designed in a such way that the receiver can recover particular GoF from the other description that carries information about adjacent GoFs.

There is one interesting question that arises here: Is the coding system defined here a real MDC system in terms of the definition given in Section 3? Obviously, multiple descriptions are used, and distortion achieved with all descriptions properly

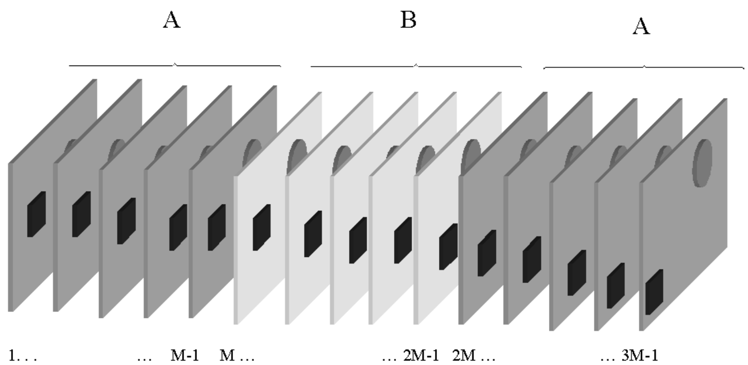

Figure 1: Group of Frames for MDC with two descriptions.

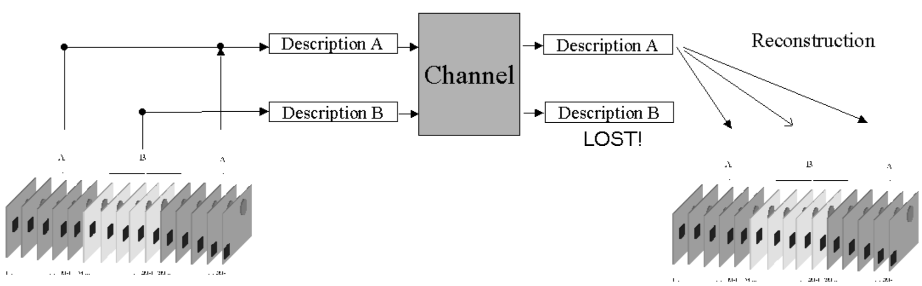

Figure 2: Block diagram of the proposed system for k=2: when one description is lost (B), the other description (A) is used to restore GoF corresponding to the description B.

received is certainly smaller then the distortion achieved when some of the descriptions are missing. However, there is a subtle difference with respect to the conventional "spatial" MDC case. One can argue that *if* a description is received, then the corresponding GoF is decoded with minimal distortion *by using that description only*. In other words, there can be no improvement in the reconstruction of the particular GoF in the decoded sequence if we introduce "adjacent" descriptions. This makes sense, since each description is trying to reconstruct complete frames, and the joint use of different descriptions in the reconstruction of a single GoF doesn't result in an improvement when compared to the reconstruction from a single successfully-received description.

The basic assumption made in our design is that motion estimated over multiple frames is continuous and "smooth" when going from one frame to another. In other words, we expect that an object will keep its trajectory over the next GoF, making it possible to find a reconstruction based on the knowledge of the object position and motion in the previous GoF. Extension of the basic model with two descriptions and GoFs of the same length would be a system with k descriptions ($k > 2$). Although we first consider same-length GoFs, ultimately adaptive-length GoFs could further improve performance by exploiting the characteristics of the video sequence.

Main advantage of the scheme proposed here is that it performs well in the case when errors are present in bursts. Also, since we are estimating motion over multiple fields, we can have reasonable gains in compression, because each GoF is described by its first frame and the motion field for particular GoF only. Low achievable data rate is always of great importance when dealing with narrow bandwidth of typical wireless channel.

There are couple of disadvantages however, one of which is large computational complexity. In order to have well estimated motion over each group of frames, expensive methods for motion estimation (in our case regularized block matching) have to be implemented. The questions that are of great importance for proper evaluation of the proposed scheme and that require further investigation are:

1. What is the most efficient method for motion estimation over multiple frames?
2. Is there a way to improve the reconstruction of GoF if all the descriptions are received over the case when only its own description is received?
3. What is the optimal way to combine more than 2 descriptions?
4. What is the criterion/optimal length of GoF over which the motion field is estimated?
5. How can we deal with sub-pixel positions derived from the trajectory estimated over GoF?
6. What's the optimal number of parameters for description of such a trajectory?

5 Experimental Results

In this section we present our initial results on the characterization of video sequences for the use in transmission over wireless channels. In order to estimate motion over multiple frames, we first carry out pixel-precision motion estimation over each pair of frames from a GoF, and we combine the results together. As test sequences we used the QCIF grayscale *coastguard* sequence. For motion estimation we used regularized block matching technique. Standard (non regularized) block matching was first applied to 8×8 blocks to obtain initial estimate. Then, each block was divided into 4 blocks of size 4×4, and block-matching search was applied around the positions obtained from the initial motion vector estimation. Finally, the same procedure was repeated for 2×2 blocks. The search range was $\pm 10, \pm 6$ and ± 6 pixels for 8×8, 4×4 and 2×2 blocks, respectively. In this way, the maximum possible value for each of the motion vector coordinates was $\pm(10+6+6) = \pm 22$. The histogram of estimated motion vectors coordinates shown in Fig. 3. supports the choice of such a search range, since most of the estimated vectors were relatively small. On the other hand, we can intuitively expect motion vectors to be small when looking for motion field between two adjacent frames.

The only regularization applied in this first pass (8×8 blocks) was introduced through a constraint on the motion vector length. The cost function used was

$$J = cost_0 \times (t + length)^{\alpha}$$

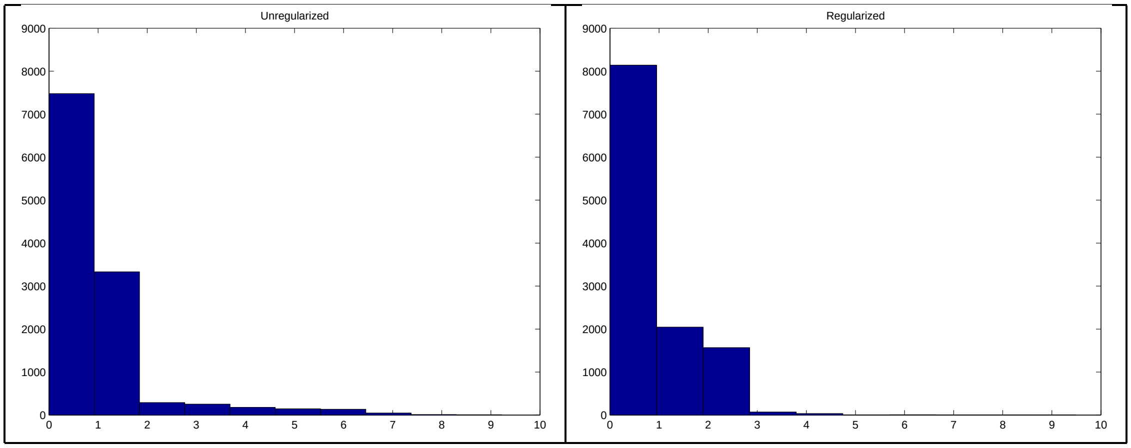

Figure 3: Histograms of length of unregularized and regularized motion vectors

where $cost_0$ is the sum of absolute values of pixel differences in the current block and the corresponding block in the reference frame. Parameters t and α define the length impact on the cost function; the larger the t, the smaller the impact of the length penalty. On the other hand, a large α will increase the total cost if length of the estimated motion vector is large. In other words, large α's and small t's will give us more regularization over small α's and large t's. In our work, we used $t = 5$ and $\alpha = 1$. The motivation for the use of this approach, when compared with the classical cost function of the form: $J = J_0 + \alpha J_1$ (where J_0 represents data-fitting term and J_1 represents additional constraint term) was the fact that in the latter one the two cost terms were of different dimensions (one representing pixel difference values and other lengths of motion vectors) and that the parameter choice was a difficult task.

In the next step, we repeated the block matching 2×2 blocks using motion field estimated in the first pass as a starting point. This time the cost function is:

$J' = cost_0 \times (t + length)^\alpha \times (t_s + s)^\beta$

where the last term in the product forces global smoothness of the motion field by penalizing large differences between adjacent motion vectors. In this expression, s is the sum of length differences between the current motion vector and its 4 nearest neighbor vectors (from 2×2 blocks up, left, down and right relative to the current block). The values of parameters chosen for the second pass were $t = 2$, $a = b = t_s = 1$. In this pass, the search range was limited to ± 6 pixels with the center at the position estimated in the previous pass. Finally, we repeated the regularization iteration once again, using the same procedure. The only difference from the previous pass was that we were using motion field estimated in the second pass as the initial field. From the Fig. 3. we can see that regularization greatly reduced the number of "huge" vectors (more than an order of magnitude).

The results of motion field estimation for all passes and for both cases of no temporal subsampling and temporal subsampling by two are given on Fig. 4 through 7. We can see that for our choice of parameters, most of regularization is done in the first pass, and that there is not much change in the estimated MF in the second iteration. Moreover, there is a small number of motion vectors that remain inconsistent with their neighbors. Picking better regularization parameters and/or running more iterations more could probably solve this problem. Another important regularization issue is to introduce "edge detection" term into the cost function, which would give us better motion estimation for the blocks that are on the object boundary in a video sequence.

For the given *coastguard* sequence, we see that our motion estimation algorithm did a good job in finding motion in the sequence. We can conclude that there is a global camera panning to the left (since motion vectors for the coast are almost uniformly pointed to the right), that large boat is moving to the right (motion vectors for this boat are larger then for the coast) and that the camera pan closely tracks the small boat. The problem of non-uniform motion field in the water, due to speckles, is greatly reduced by regularization.

In order to find the maximum length of GoF, we investigate how well our motion estimation is working over multiple frames. There are two ways to do this: to play the sequence from the first frame using only motion fields estimated earlier and to visually inspect the quality of the sequence (motion projection), or to do backward motion compensation and see how well we can compensate the real motion (motion compensation). In other words, we want to use estimated MF's to "return" all frames in the sequence to the first frame position. Then, we can use visual inspection to judge the total motion in the "frozen sequence" or plot the total variation for each motion-compensated frame when compared to the first frame in the sequence.

Below, we present results obtained using the second approach. We plot the first frame together with the motion-compensated last frame for different estimates of motion fields (obtained with and without regularization). We also plot the values of the total variation per pixel for each of the motion-compensated frames in the sequence. These plots are given in Figs. 8 and 9.

We can see that all the motion-compensated frames keep resembling to first frame in the sequence, even after 35 frames in the case of no temporal subsampling or 50 frames - when temporal subsampling is used. An interesting observation is that the unregularized MF gave the best results both visually and numerically, which is not unexpected since in this method only the data matching term is used. We can see that the regularized MFs perform especially poorly near the edges of objects in the sequence. This is probably because of the overregularization in these areas, since we didn't use a term that would relieve the smoothness constraint at the edges. We think that a better choice of regularization parameters and the introduction of "edge aware" term would help a lot. There is another detail worth noticing: although the unregularized MF performed better in our motion compensation test, we can see that some details and structures are preserved better by using regularized MF's (e.g., figure in the small boat is almost completely lost after 35 frames when using unregularized MF, but well preserved in the case of regularized MF). This gives us additional expectation that regularization can produce results comparable to unregularized case, simultaneously providing opportunities for compression and error resilience.

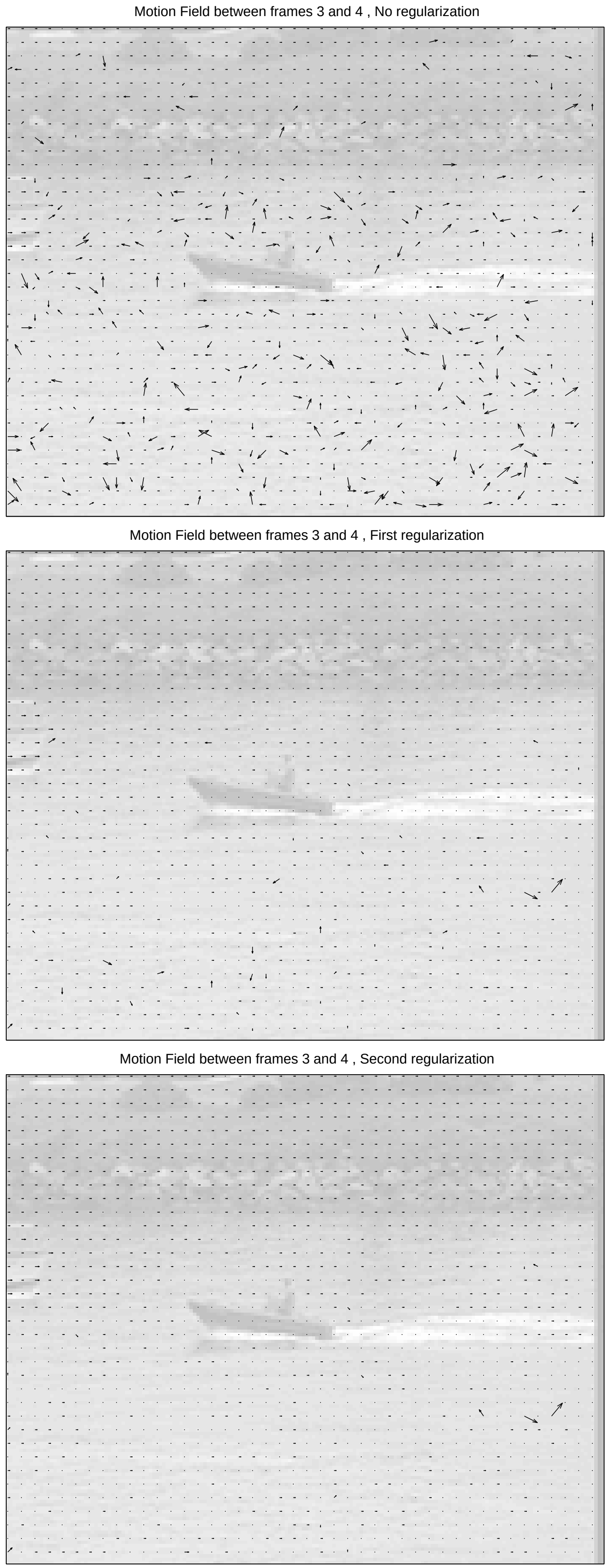

Figure 4: Motion field between frames 3 and 4

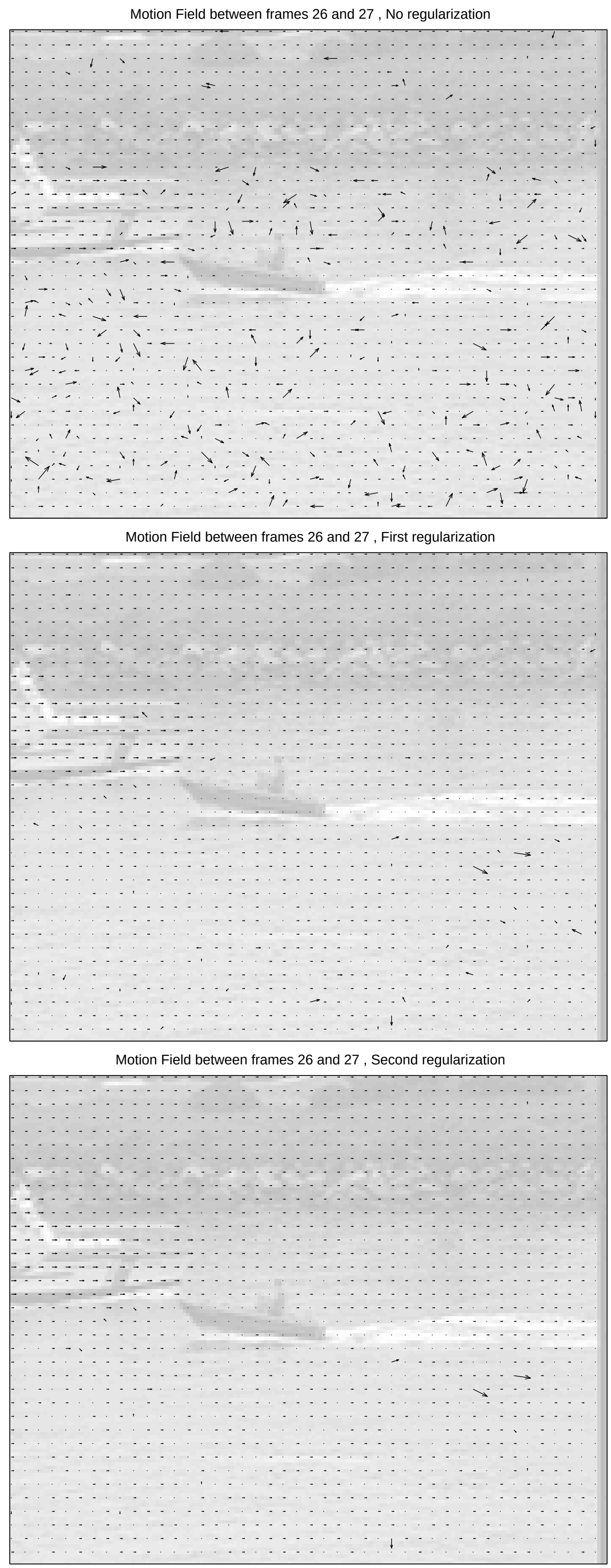

Figure 5: Motion field between frames 26 and 27

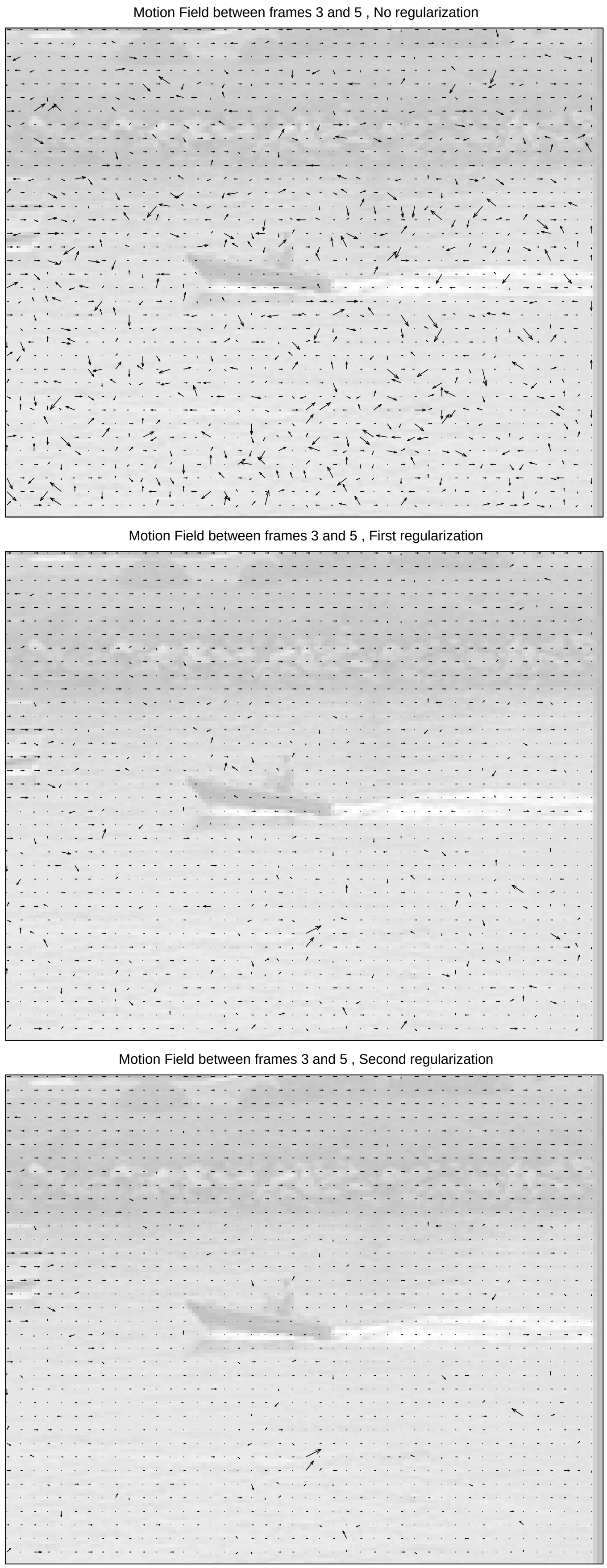

Figure 6: Motion field between frames 3 and 5

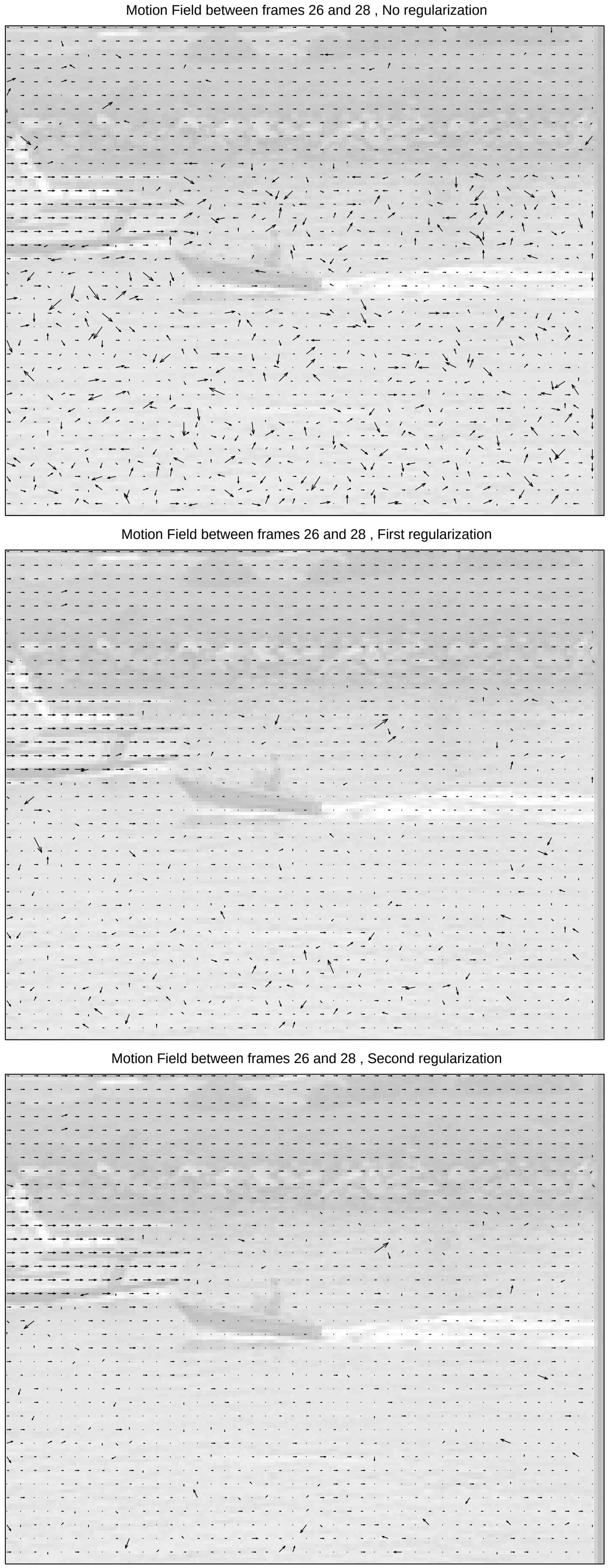

Figure 7: Motion field between frames 26 and 27

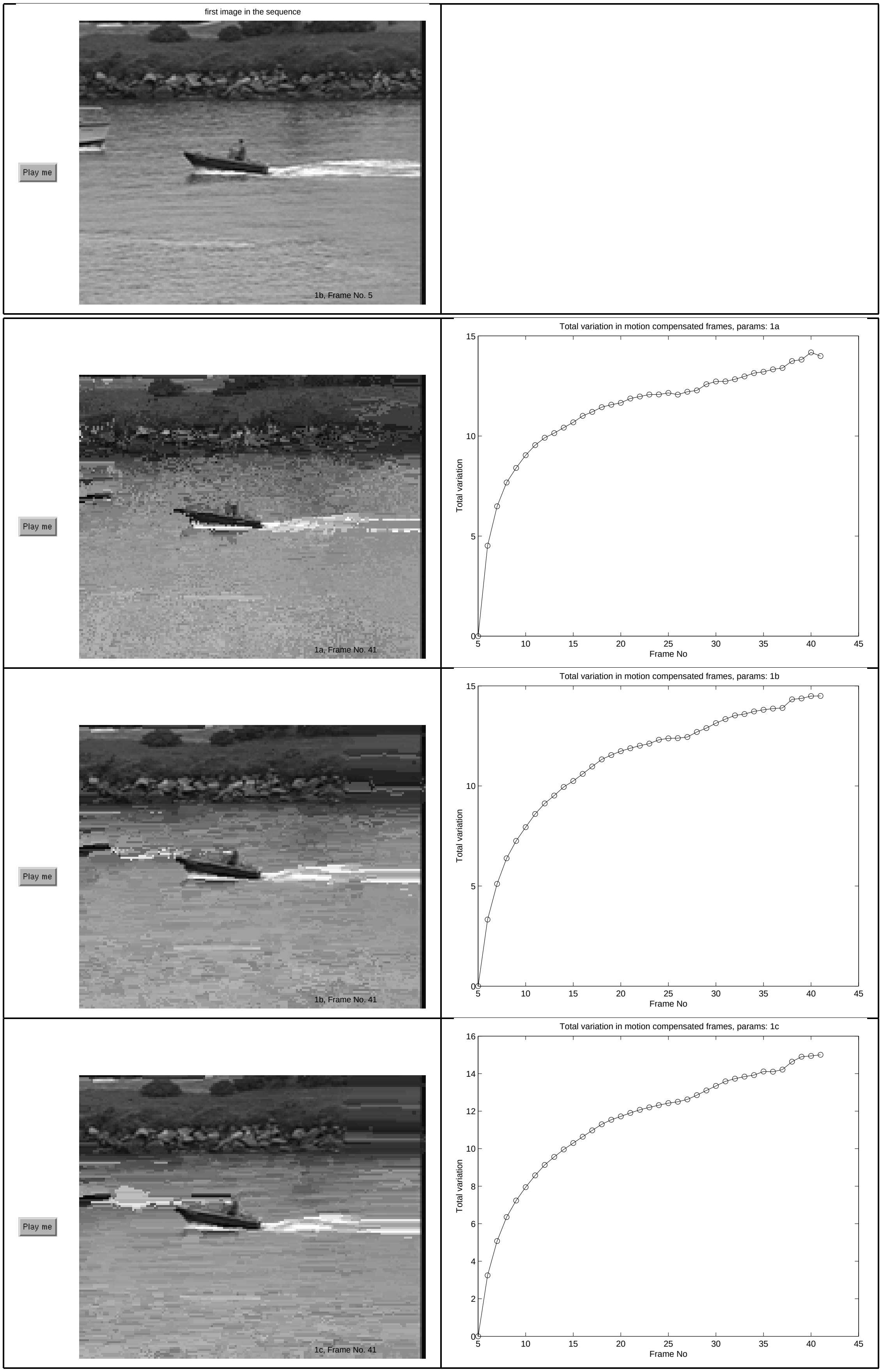

Figure 8: Last motion-compensated frames, no temporal subsampling

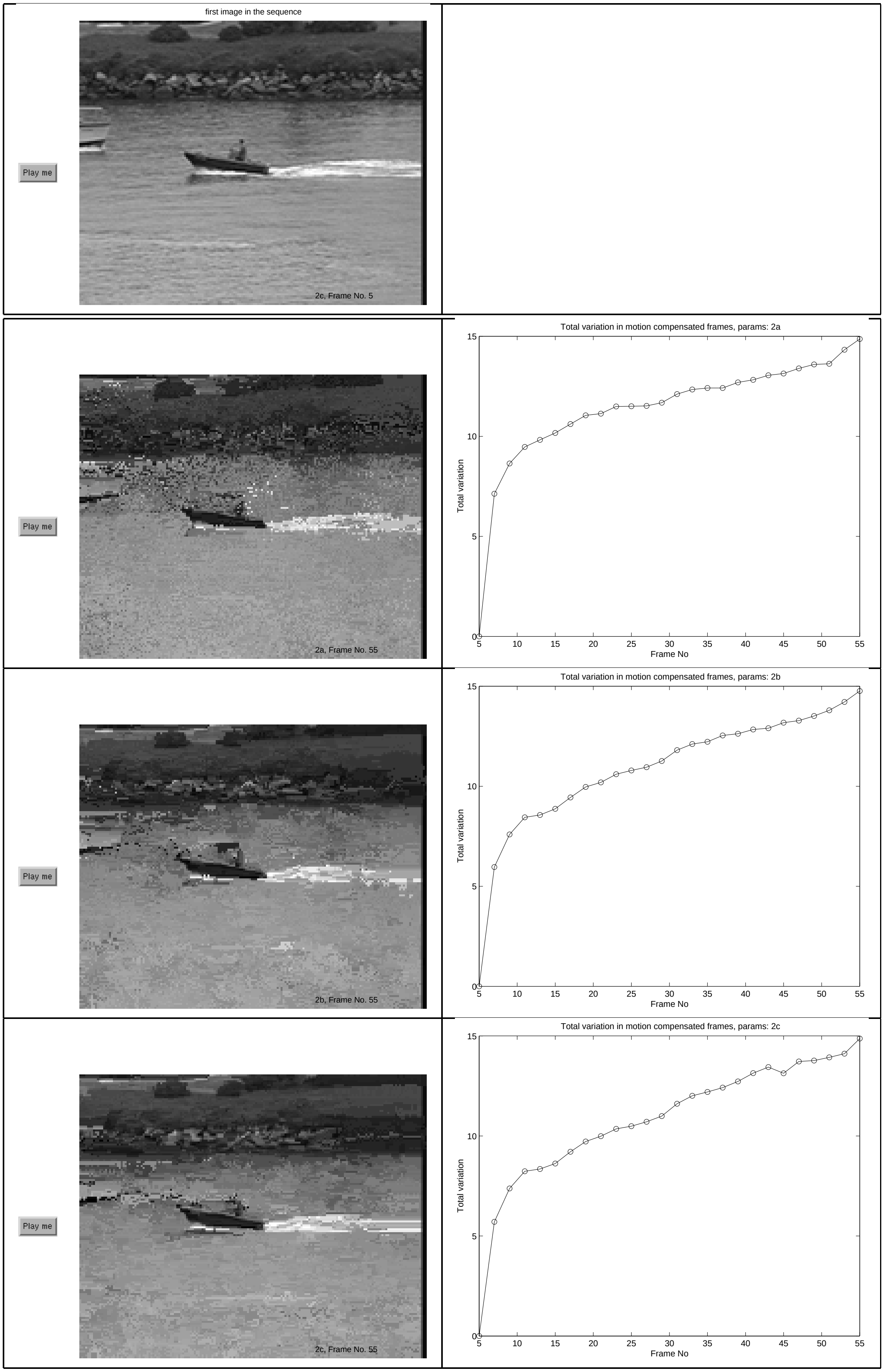

Figure 9: Last motion-compensated frames, temporal subsampling equal to 2

6 Conclusions

Since this work is in early stages, it is hard to give definite conclusions whether the proposed method could be implemented for real-time video transmission over wireless channel. Of special interest for future work would be to conduct a research/survey on error patters in wireless channels, continue with investigation of possibilities for estimation of motion trajectories over multiple frames and to further examine questions posed in Section 4. First results are encouraging and we will continue to investigate this issue with the hope that significant gains can be achieved.

References

[1] C. Y. Hsu, A. Ortega, and M. Khansari, "Rate control for robust video transmission over burst-error wireless channels" IEEE Journal on Selected Areas in Communications, Special Issue on Multimedia Network Radios, 1998.

[2] P. G. Sherwood and K. Zeger, "Joint source and channel coding for internet image transmission," in Proc. of DCC'97, 1997.

[3] N. S. Jayant and S. W. Christensen, "Effects of packet losses in waveform coded speech and improvements due to an odd-even sample-interpolation procedure," IEEE Trans. Communications, vol. COM-29, no. 2, pp. 101-109, Feb. 1981.

[4] G. Karlsson and M. Vetterli, "Subband coding of video for packet networks," Optical Engineering, vol. 27, no. 574-586, 1988.

[5] V. A. Vaishampayan, "Design of multiple description scalar quantizers," IEEE Trans. Information Theory, vol. 39, no. 3, pp. 821-834, 1993.

[6] V. A. Vaishampayan and J.-C. Batllo, "Asymptotic analysis of multiple description quantizers," IEEE Trans. on Information Theory, vol. 44, no. 1, pp. 278-284, Jan. 1998.

[7] D. Servetto, K. Ramchandran, V. Vaishampayan, and K. Nahrstedt, "Multiple description wavelet based image coding," in ICIP'98, 1998.

[8] Y. Wang, M. Orchard, and A. R. Reibman, "Multiple description image coding for noisy channels by paring transform coeffcients," in Proc. IEEE 1997 First Workshop on Multimedia Signal Processing, 1997.

[9] V. K. Goyal, J. Kovacevic, R. Arean, and M. Vetterli, "Multiple description transform coding of images," in ICIP'98, 1998.

[10] Wenqing Jiang and Antonio Ortega "Multiple Description Coding via Polyphase Transform and Selective Quantization", In Proc. of Visual Communic. and Image Proc., VCIP '99, San Jose, USA, Jan 1999.

[11] Y. Wang and Q. Zhu, "Error control and concealment for video commu-nications: A review," Proc. IEEE, vol. 86, pp. 974-997, May 1998.

[12] A. A. El-Gamal and T. M. Cover, "Achievable rates for multiple descriptions," IEEE Trans. Information Theory, vol. IT-28, no. 6, pp. 851-857, Nov. 1982.