# REUSE OF MOTION PROCESSING FOR CAMERA STABILIZATION AND VIDEO CODING

*Bao Lei[1], Rene Klein Gunnewiek[2], Peter H. N. de With[3]*

Philips Research Laboratories, Eindhoven, 5656AA, The Netherlands[12]
University of Technology Eindhoven, VCA Dept, 5600MB, The Netherlands[3]
rene.klein.gunnewiek@philips.com[2], p.h.n.de.with@tue.nl[3]

## ABSTRACT

The low bit rate of existing video encoders relies heavily on the accuracy of estimating actual motion in the input video sequence. In this paper, we propose a Video Stabilization and Encoding (ViSE) system to achieve a higher coding efficiency through a preceding motion processing stage (to the compression), of which the stabilization part should compensate for vibrating camera motion. The improved motion prediction is obtained by differentiating between the temporal coherent motion and a more noisy motion component which is orthogonal to the coherent one. The system compensates the latter undesirable motion, so that it is eliminated prior to video encoding. To reduce the computational complexity of integrating a digital stabilization algorithm with video encoding, we propose a system that reuses the already evaluated motion vector from the stabilization stage in the compression. As compared to H.264, our system shows a 14% reduction in bit rate yet obtaining an increase of about 0.5 dB in SNR.

## 1. INTRODUCTION

The efficient compression of existing video encoders is to a large extent based on accurately estimating the actual motion in the input video sequence. Generations of video encoding standards, from MPEG-1 in 1993 to H.264 in 2003, have become remarkably efficient, mainly due to an increasingly advanced motion modeling and a correspondingly accurate prediction of the incoming video signal. However, this progress was achieved at the expense of a high computational complexity. Among the modules in a video encoder, motion estimation has always been a critical component that significantly contributes to efficiency and complexity simultaneously.

Starting with MPEG-4, modern video encoders perform increasingly accurate motion estimation by modeling the motion in the input sequence. H.264 provides enhanced motion-estimation capabilities with seven block-partition modes for inter prediction, quarter-pixel accuracy, multiple reference frames and intraframe prediction. For approximately 50% bit-rate savings at an equivalent perceptual quality [1], H.264 was reported to be significantly more complex than H.263 [2].

Rather than focusing exclusively on estimating and encoding the actual motion in a video sequence, we concentrate in this paper on achieving a higher bit-rate reduction by a preceding motion processing stage. The increased compression ratio is obtained by differentiating between the temporal coherent motion and a more noisy motion component which is orthogonal to the coherent one. The latter motion component occurs particularly in mobile devices like mobile phones. In this application, camera vibration is the most common type of undesirable motion in video sequences. Therefore, our approach is to first accurately compensate this motion, so that it is eliminated from the succeeding coding process. As a result, this undesirable motion component is left out of the encoded bit stream entirely.

The above approach calls for integrating a camera stabilization system and a video encoder. A so-called Digital Image Stabilization (DIS) system has been proposed as the digital domain solution for compensating camera vibration. The DIS system traces back to 1986 when it was first patented in the US by a group of researchers from MEI, Japan [3]. More recent DIS systems rely on global motion-estimation algorithms to obtain the camera-motion vectors needed for stabilization.

Despite the fact that both the DIS system and the video encoder co-existed for a long time, integration of the two has not been exploited because of the involved computational complexity. This is especially true for mobile devices, which generally need stabilization, but have tight constraints on the available processing resources.

Our proposed system is a synergetic integration of a Digital Image Stabilization system and a video encoder, called Video Stabilization and Encoding (ViSE) system. We address the complexity issue through reusing estimated candidate motion vectors from the DIS system in the video encoder. The ViSE system combines low-complexity and high-efficiency, so that a better complexity-efficiency trade-off is made.

The sequel of the paper is as follows. Section 2 presents the analysis of the relationship between camera stabilization and encoding efficiency. Section 3 presents the design rationale and architecture of the ViSE system, which is followed by Section 4 where it is compared with H.264 reference encoder. Finally, conclusion is given in Section 5.

## 2. CAMERA STABILITY AND VIDEO ENCODING EFFICIENCY

It is expected that video sequences containing undesirable camera vibrations require more bits to encode than the sequences without those vibrations. However, to effectively reduce the impact of vibration on the encoding efficiency, it is necessary to study how many and where extra bits are caused by camera vibration, and how they can be saved most efficiently in terms of computation complexity.

In our research, we simulated vibrations with different amplitudes, frequencies and directions. H.264 was used to encode the sequences with several alternative encoder settings. Based on the encoding results, we are able to provide an indicative description on the relationship between camera vibration and the video encoding efficiency.

### 2.1. Vibration–Efficiency Simulation

Camera vibration can occur along six directions, which are commonly defined as (horizontal and vertical) translation, rotation, zoom, tilt and pan as shown in Figure 1. Several assumptions are made under common usage scenarios to efficiently represent vibration: (1) Camera is shooting at a normal frame rate of 25 fps; (2) Speed of tilt and pan is slow; (3) Object distance[1] can be assumed to be reasonably long. Under these assumptions, translation vibration can properly approximate tilt and pan vibrations. Hence, we shall constrain ourselves to considering translation, rotation and zoom vibrations only.

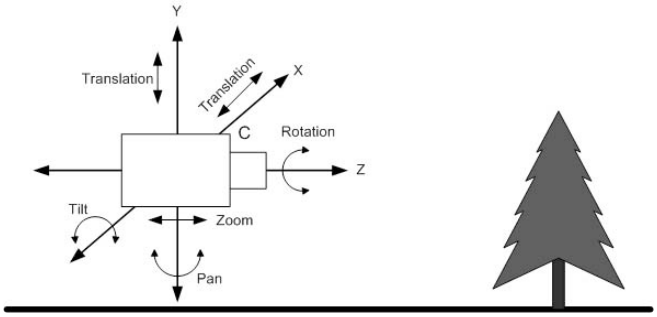

Figure 1 Possible directions of camera vibration.

[1] The distance between the object and the camera

Artificial video sequences with predefined and reproducible vibrations were created through manipulation of the *lenna* image to facilitate quantitative studies in the simulation. The setup of the vibration–efficiency simulation is shown in Figure 2.

An H.264 encoder consecutively compresses each of the 32 artificial video sequences with each of the four encoder configurations. The encoded stream and meta-data were then collected at the output of each encoding process.

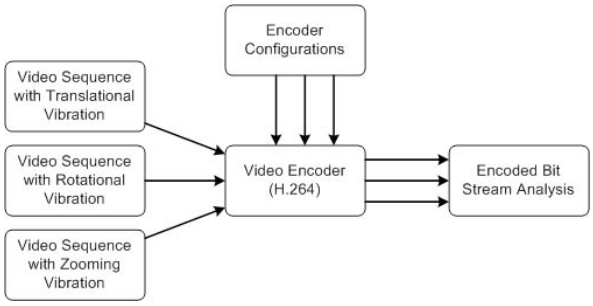

Figure 2 Set-up of vibration–efficiency simulation.

### 2.2. Camera Stability – Encoding Efficiency Relationship

The relationship between camera stability and encoding efficiency is summarized in Figure 3. The curves represent the relationship between encoding efficiencies and vibration amplitudes. The four sub-figures represent the impact of selected encoder configurations on the previous relationship.

As shown in Figure 3, camera vibration has a significant negative impact on the encoding efficiency. From 31 kbps at zero vibration, bit-rate increases are observed in all vibration scenarios and encoder configurations. In other words, bit-rate reduction is feasible through camera stabilization.

Furthermore, rotational vibration has the most serious impact on coding efficiency, as compared to the translation and zooming. The total encoded bit rate increases from 30 kbps to 284 kbps for rotational vibration from still status to 1° in amplitude and 15 Hz in frequency. Considering the fact that rotation is seldom intentional during video shooting, such vibration is highly undesirable.

The current video encoders perform motion estimation by searching for the minimum Sum-of-Absolute-Difference (SAD) between the current motion-estimation block and the motion-candidate blocks. Because the search is only based on translations, rotation and zooming motion usually result in large SAD values. Blocks that cannot be sufficiently well matched are intra-encoded, which results in an increased bit rate. The H.264 encoder may try smaller estimation blocks for a better match. However, this will introduce extra bit rate in motion-information encoding. As a result, current video encoders handle rotational and zooming vibrations poorly.

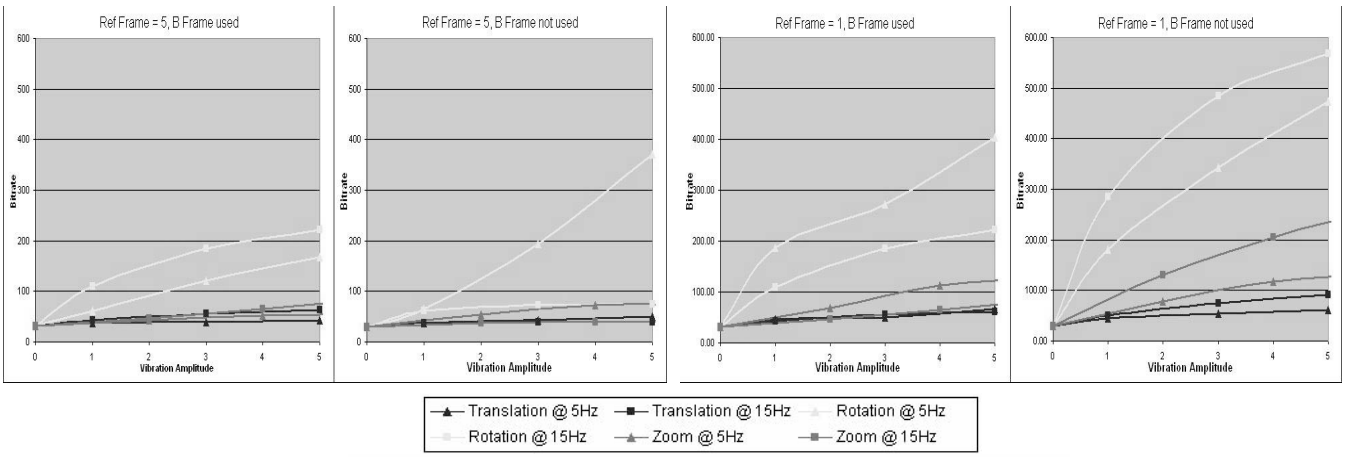


Figure 3 Relationships between camera stability and encoding efficiency.

Finally, the encoder configurations play an important role in the vibration–efficiency simulation. While using B-frames generally reduces the output bit rate, the effect is not essential in the figure. Instead, the vibration–efficiency curve is more sensitive to the number of reference frames used, because extra reference frames provide more possible candidate blocks for motion estimation.

## 3. VIDEO STABILIZATION AND ENCODING (ViSE) SYSTEM

The Video Stabilization and Encoding (ViSE) system stabilizes input video prior to encoding. Consequently, the ViSE system reduces the bit rate caused by camera vibrations. As a result, the integrated system produces stable video sequences at low bit rates, as compared to a conventional video encoder.

The ViSE system forms a synergetic integration of a Digital Image Stabilization system and a video encoder. Despite the complexity of the two systems, we will show that the integration can be computationally efficient by reusing the candidate motion vectors in both sub-systems.

The structure of a typical DIS system is portrayed by Figure 4. Input video frames are fed into two paths. One path involves the control of the global motion-estimation module and the motion-smoothing module. This path generates motion-compensation vectors and motion vectors of the stabilized video sequence. The second path involves data manipulations in the frame memory. Here, video frames are transformed using compensation vectors to generate stable video signals. The structure of a commonly known block-based motion-compensated hybrid video encoder, like H.264, is shown in Figure 5.

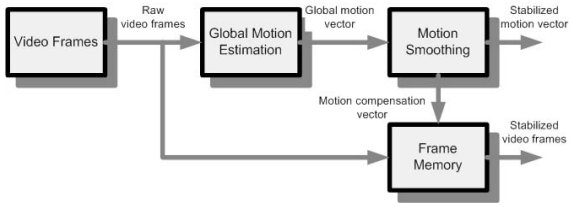


Figure 4 Architecture of the DIS system.

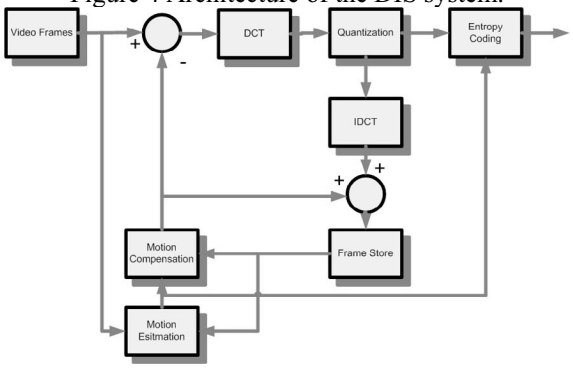


Figure 5 Architecture of a MC-DCT video encoder.

The new architecture of the ViSE system that results from integrating the two sub-systems is shown in Figure 6. The motion-estimation modules in both systems, which are computationally intensive, are elegantly combined. The original motion-estimation module in the video encoder is replaced by the DIS system, as indicated by the dotted box at the lower left corner. Global motion estimation contributes to roughly 60% of the computation in DIS, whereas motion estimation in a typical MPEG video encoder contributes for 60-80%. After some analysis, we have estimated a 30-40% reduction in computational complexity for merging the two systems, though the actual complexity of the proposed system highly depends on the level of algorithmic optimization and the target processing platform. Furthermore, both motion-estimation modules employ frame buffers. These memories can be reduced by exploring reuse as well.

The new motion-estimation module receives two inputs and generates two outputs. An input video frame is first stabilized with reference to its preceding frame. Subsequently, the remaining global-motion vector (without vibration) is used to warp the previous frame for motion compensation. Therefore, the actual input to the DCT transform can be viewed as the input video sequence after removal of the vibration motion and compensation for the remaining motion. Independent motion blocks are intraframe coded with zero motion vectors from the previous frame. Alternatively, regular motion-estimation algorithms can be selectively applied to these blocks. Finally, the residual codes and global motion vector are sent after entropy coding. The remainder of the video encoding process is unchanged.

| | |
|---|---|
| **Input** | Current video frame from capturing device |
| **Input** | Previous video frame from local decoding |
| **Output** | Stabilized current video frame to be subtracted with motion-compensated frame |
| **Output** | Motion vector of current video frame after compensation for vibrations |

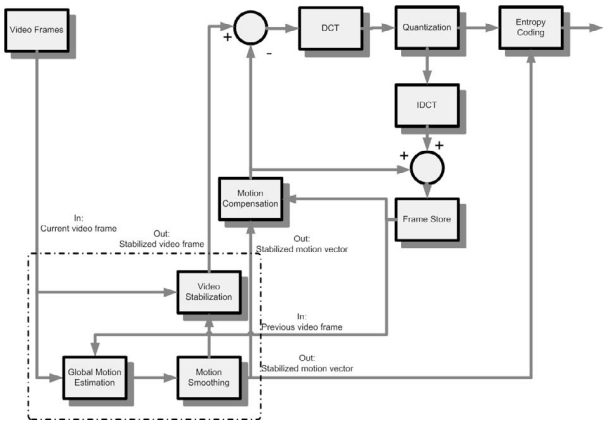

Figure 6 New architecture of ViSE system.

## 4. BENCHMARKING AND ANALYSIS

We benchmarked the performance of our ViSE system with an H.264 reference encoder (JM 9.3). The eight sample sequences used for benchmarking represented a wide range of real-life camera vibrations, including video shooting when standing still, walking, running, riding on transportation, etc. Bit-rate reduction is observed in all sequences with an average amount of 14.03% and 0.68 dB gain in SNR (Figure 7). Observation of the encoder meta-data shows that the bits used for coding motion vectors are reduced, indicating less vibration motion in the encoded sequence. As the same time, the number of blocks coded using mode 0 is increased, which suggests that more blocks were simply copied from its motion reference-frame due to improved stability. Both phenomena contribute to the overall bit-rate reduction. The individual bit-rate savings depend on the camera motion and video content.

## 5. CONCLUSIONS

It has been shown that removal of the undesirable vibration motion in a compression-preceding stage improves the efficiency of the video encoding for low-rate mobile applications. The Video Stabilization and Encoding (ViSE) System described in this paper is a feasible proposal for integrating the DIS system with motion processing for compression. As a result, undesirable camera vibration is smoothed prior to video encoding by the DIS system. The bit rate is reduced by 14.03% combined with a 0.68 dB gain in SNR. The reuse of already evaluated motion-vector candidates from the stabilization stage in the video compression relieves the computational complexity. In the new proposal where we combine motion processing, the computational complexity is estimated to be about 30-40% lower than the original case based on separate processing. Evidently, this number strongly depends on the chosen computing platform and the level of algorithmic optimization.

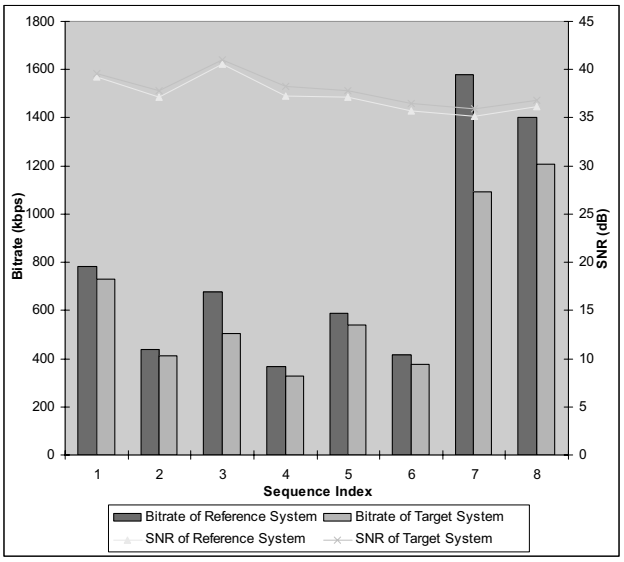

Figure 7 Benchmarking results for ViSE system.

## 6. REFERENCES

[1] Thomas Wiegand, Gary J. Sullivan, Gisle Bjøntegaard, Ajay Luthra, "Overview of the H264/AVC Video Coding Standard", *IEEE Transactions on Circuits and Systems for Video Technology*, Vol. 13, No. 7, July 2003

[2] Ville Lappalainen, Antti Hallapuro and Timo D. Hamalainen, "Performance of H.26L Video Encoder on General-Purpose Processor", IEEE, 2001

[3] M. Oshima, T. Hayashi, S. Fujioka, T. Lnaji, H. Mitani, J. Kajino, K. Ikeda, K. Komoda, "VHS Camcorder with Electronic Image Stabilizer", *IEEE Transactions on Consumer Electronics,* Vol. 35, No. 4, November 1989