Efficient Motion Estimation and Predictive Coding Methods for Compression of Spatio-temporal Sequences

A Thesis submitted by
Tushar Shankar Shinde

in partial fulfillment of the requirements for the award of the degree of
Doctor of Philosophy

॥ त्वं ज्ञानमयो विज्ञानमयोऽसि ॥

Indian Institute of Technology Jodhpur
Department of Electrical Engineering
June 2020

6
Conclusions and Future works

6.1 CONCLUSIONS

This Thesis work developed the novel mechanisms for efficient motion search and compression of Spatio-temporal skeleton sequences. The following conclusions can be drawn:

The fact that the motion estimation process consumes a significant proportion of time in video compression has posed a stiff computational challenge. An efficient solution to this problem is presented in Chapter 3. The proposed algorithm uses the dynamic switch between the search region based on the SAD minimum location. Moreover, the proposed adaptive search range selection ensured the fast convergence of the block matching process without compromising on performance. The proposed direction-oriented search patterns are found to be useful for varying motion content sequences as compared to FS, DS, HS, CDHS, ARPS, and TZS. Moreover, these patterns are less susceptible to becoming trapped in the local minimum. Computational cost is further reduced by employing partial SAD computations for each candidate block. The experimental results show that the proposed EDOS achieved speed-up of 2.69, 1.28, 1.05, and 1.12 times over ARPS for slow, medium, fast, and directional motion sequences, respectively, without any significant degradation in block matching quality. For directional motion video sequences, this technique even achieved improvement up to 1.58 dB and 8.8% in PSNR and SSIM, respectively. This improvement in matching quality over FS along with reduced computational cost mainly comes from the proposed dynamic switch between SR based on the location of DISC, adaptive SR dimension selection, adaptability of direction-oriented search patterns, and optimal threshold value selection for partial SAD computations. Improvement in performance as compared to FS and reduction in computational cost to state-of-the-art fast methods is the major contribution of the proposed work. Thus, EDOS is very suitable for a wide range of applications, such as high-speed and high-quality video conferencing.

Although efficient motion search algorithms provided improved performance but these algorithms needed to be tailored for specific applications such as surveillance video coding. In Chapter 4, a fast background-foreground-boundary aware block matching algorithm is proposed for surveillance videos. Our method firstly used a three-level saliency detection mechanism to classify blocks into three classes: BG, BD, and FG blocks. This classification captured different characteristics of surveillance videos. Secondly, the proposed method used different search strategies for different classes. Our search patterns are found to be useful for directional motion content blocks in the surveillance sequences. Thirdly, we have also proposed BD block partitioning criteria for efficient block matching. The proposed approach provided better matching quality as compared to other existing algorithms. At the same time, it has also been able to provide four times speed-up for BG blocks and more than two times speed-up for BD and FG blocks as compared to the BFDS method. Experimental results confirm the efficacy of the proposed solution in different scenarios, such as surveillance sequences containing varying motion types and motion areas. Thus our method is suitable for real-time surveillance video coding applications.

The importance of the skeleton information in surveillance videos pose a significant storage challenge. In Chapter 5, a novel lossless compression scheme for encoding skeleton sequences is presented. For this, we introduce a multimodal prediction scheme that switches between a set

of prediction modes to effectively exploit spatial and temporal correlations present in the skeleton sequences. Next, we introduce two entropy coding schemes to exploit the coding redundancy available in the prediction residuals further. Our method is lossless, and both encoder and decoder work symmetrically. Experimental results show that our adaptive method significantly outperforms the direct coding scheme. The proposed scheme provided about 70% improved coding performance for challenging skeleton sequences chosen from Posetrack and Surveillance datasets. Thus our method is suitable for real-time skeleton sequence coding. The skeleton sequences could play a vital role in security applications.

In a nutshell, this thesis provided efficient and effective solutions for motion search in the traditional video compression process. The motion search process is further tailored for surveillance video coding applications. Lastly, the skeleton sequence information present in the surveillance videos is significantly compressed.

6.2 RECOMMENDATIONS FOR FUTURE WORK

With this thesis, following future directions could be explored:

- In future work, the direction oriented search patterns could be made adaptive for high definition video sequences containing varying motion content. The proposed algorithm could be tailored for critical applications such as real-time sports video coding scenarios. More advanced search patterns exploring advantages of both state-of-the-art TZS patterns and our direction-oriented search pattern could be investigated. Moreover, the partial distortion mechanism could be extended for various applications where sub-sampling in the pixel domain plays a vital role. We also aim to include the proposed scheme in the state-of-the-art HEVC framework.

- Our background foreground boundary aware block classification scheme can be employed in real-time video conferencing applications. The proposed surveillance video coding scheme could also be implemented in the HEVC framework.

- Further, we would like to extend the skeleton coding framework for the lossy compression by introducing some quantization parameters and investigate the performance for various action recognition tasks. Our skeleton coding scheme could further be incorporated in state-of-the-art video compression standards as an extension.

...