Asian Journal of Applied Science and Technology (AJAST)
Volume 1, Issue 5, Pages 104-107, June 2017

Video Stabilization and GOP - Codec Estimation

Prerana Sawant[1] and N.Aswini[2]

[1]UG Student, Department of ECE, MVJ College of Engineering, Whitefield, Bengaluru, Karnataka, India.
[2]Assistant Professor, Department of ECE, MVJ College of Engineering, Whitefield, Bengaluru, Karnataka, India.

Article Received: 16 May 2017 *Article Accepted: 01 June 2017* *Article Published: 04 June 2017*

ABSTRACT

Video stabilization and codec identification is an important task to prove the authenticity of its content. Video content is acquired, compressed and then distributed in a digital format. This paper gives various algorithms used to identify codec and GOP (Group of Pictures). It also gives a method for stabilizing the video. It compares the complexity and accuracy of various algorithms used. This analysis is very useful in forensics, since an analyst might be able to identify the characteristics of the previous coding steps (like its codec, GOP size) and he can also stabilize the video. These methods are also used other than in forensics like video quality assessment.

Keywords: GOP Estimation, Codec, Stabilization, Forensic video analysis and Video quality.

1. INTRODUCTION

The popularization of digital devices with video cameras has provided us to get fast and sustained videos whenever required. However, visual quality of captured videos is varied and extremely stricken by uncertainties associated with environmental setting, photographic skills, etc. Typically, it's difficult for amateur user to hold a camera steadily within the whole capture method, therefore generating noticeable shaking impact within the output video sequence. Thus, video stabilization becomes a crucial and exigent drawback in video process that strives for eliminating or reducing the visual shake for a smoother display. Some hardware-based solutions are applied to tackle the annoying shake, like tripods, dollies, steady-cams, that are heavy-load and inflexible in use for casual capture. But in camera's which have in-built optical or electronic stabilizers solely resist high-frequency noise. They could bog down within the prevalence of low-frequency disturbance like video shot by a walking person. On the contrary, recent competition strategies treat video stabilization as a task of image post-processing, that is freed from any hardware and able to influence a large variety of styles of shaky motion. Such strategies don't essentially pursue physically correct motion set up, however obtain to rework the first video frames for steady look as an entire sequence.

When the video is recorded by a video camera or camcorder it is originally encoded by the acquisition device, and then re-encoded, either after an editing operation, or when it is uploaded to a sharing website. The analysis of the bit stream gives us the details about the last compression step (i.e., the codec that was adopted and its corresponding encoding parameters), while masking the previous compression history. This paper gives a method in order to identify the codec used in the first coding step and also its GOP size.

A GOP (Group of Pictures) contains different picture types, they are, I-Frame (intracoded picture), B-Frame (Bipredictive coded picture), P-Frame (Predictive coded picture) and D-Frame (DC direct coded picture). Each GOP begins with I-Frame. I-Frame is a picture that is coded independently of all other picture. A GOP structure is referred by 2 numbers M and N. M represents the distance between two anchor frames (I-frame or P-frame). N represents the distance between two images (I-frames), this is called the GOP Size.

A Video Codec is an electronic software (or a circuit) that compresses or decompresses a digital video. The Codec either converts a raw digital video to a compressed format or vice versa. The compression of the video can be either lossy or lossless compression. The two types of compression are (i) Spatial compression (intraframe coding) (ii) Temporal compression (interframe coding). The various codecs are MPEG2 (Moving picture Experts group), MPEG4, H.264/AVC (Advanced video coding), DIRAC, etc.

Section II discusses about the Video stabilization techniques. Section III deals about the Codec identification. Section IV deals with the GOP estimation. Section V concludes the paper, providing information about future challenges.

2. VIDEO STABILIZATION METHODS

Global approach - Fast Video Stabilization

This approach presents a completely unique formulation of video stabilization, by directly providing a solution for optimal image warps towards stable sequence. With the calculable shaky motion via long or short feature trajectories, this approach encodes another 2 steps, motion compensation and image warp, into one global optimization method, instead of operating as 2 individual steps. This method is done with solely positions of embedded mesh vertices as common variables. Spatial and temporal coherence is then re-formulated with similarity invariant illustration of motion trajectories and intra-(and inter-)frame consistency of comparable transformations with regard to mesh vertices. Such a one-shot formulation is used to convert video stabilization into a quadratic energy minimization drawback defined for image warps, so is efficiently resolved by exploitation sturdy convergent thinker for sparse linear system

[1].

This method directly finds a set of image warps with appropriate parameterization, which are shake fee sequences in the entire video sequence. Therefore the two steps of motion compensation and image warping are brought together and combined to form a single global video stabilization. First the shaky motion is estimated and then the global approach is performed. The main aim of stabilization is to provide a smooth display of video sequence, so that both high and low frequency vibration can be suppressed in order to produce shake-free appearance.

This approach gains fast stabilization computation that is comparable with the spatial- temporal optimization method. It is faster than the other methods. Besides, this method can deal with shaky videos that cannot be processed by previous method like subspace video stabilization, and well preserve the original appearance with less visual distortion and frame cropping. The primary time consuming is based on constructing and solving the sparse linear system. Generally for 200 frames with 1280 × 720 resolution, the feature tracking for trajectories takes about 3.3 seconds, then the construction and solving the sparse linear system takes about 4 seconds; we finally obtain the stabilized frames after cropping outranged regions and rendering them with texture mapping technique, which consumes about 3.8 seconds for 200 frames. In short this method can achieve an 18fps processing toward fast video stabilization.

Limitation: This method is based on robust trajectories detection, therefore it might possibly crack itself in case feature tracking fails. A solution to this problem is to adopt local parametric similar transformations between adjacent frames. This method might generate artifacts for videos with apparent line structure. A solution to this problem is by adding line-preserving constraints.

Adaptive Local Subspace of Feature Point Classification

This approach first classify the feature point into inliers and outliers based on global motion estimation to remove the feature points on moving objects, this stabilizes the camera movements without the interference of outliers. Second, it assembles the trajectory matrix with inliers across adaptive frames such that it guarantees sufficient complete trajectories for factorization. After this every frame is smoothed in separate local subspace. This method [2] is more flexible than the global subspace.

The subspace video stabilization method cannot be used when scenes are varying fast in moving camera. Hence, this method [2] came into existence. First, the feature points are classified into two sets based on global motion estimation to exclude outliers. Therefore the effect of outliers on motion vector estimation is reduced.

A Geodesic Video Stabilization

This provides a unique formulation of video stabilization within the space of geometric transformations. With the setting of the Riemannian metric, the optimized smooth path is cast because the geodesics on the Lie group embedded in transformation space. Whereas determining the geodesics includes a closed-form expression in a certain space, path smoothing will be simply enforced by geometric interpolation, instead of optimizing any space time energy function. Specially, by the use of geodesic solution within the space of rigid transformations, this approach even gains quickening 10× quicker than state of the arts methods for path smoothing and motion compensation, and guarantees no further distortion drawn into the stable frames.

Limitation: The motion estimation of this approach is based on the feature points detection and matching between frames. Therefore the existence of large homogeneous or reduplicate regions impairs feature matching, which might degenerate stabilization results that are provided by this method.

3. CODEC IDENTIFICATION

There exists a difficulty in dealing with video sequences rather than still images. First, while considering the video, there is no single coding standard that is universally used. Conversely, different standards are being used based on the scenario, so that sequences might be encoded with legacy MPEG-2 and MPEG-4 encoders, as well as with the more recent AVC and HEVC standards. Second, the number of degrees of freedom while configuring the video encoder is significantly larger than in JPEG. Third, complex statistical dependencies are created between the quantized coefficient values among the different frames, due to the adoption of motion-compensated prediction.

Here in [4] a processing chain of two coding steps is considered and a method is proposed that aims at identifying the type of codec used in the first step, with the help of coding-based footprints. The method is based on the fact that lossy coding is an idempotent operation, that tell that re-encoding the reconstructed sequence with the same codec and the same coding parameters, will produce a sequence that is highly correlated with the input one. This method finds several applications in the field of multi-media forensics, e.g. to identify the device that generated the original video stream or detect collages of different sequences.

In Fig1. The output of codec 2, is codec with c3 belonging to set of codecs C (e.g. MPEG2, AVC etc.). When c3 does not match c1, then $\hat{x}_3 \neq \hat{x}_1$. . If c3 matches c1, the PSNR vs. QP function presents a local maximum which corresponds to the QP value originally used by c1 to encode the frame Refer Fig.2.

Asian Journal of Applied Science and Technology (AJAST)
Volume 1, Issue 5, Pages 104-107, June 2017

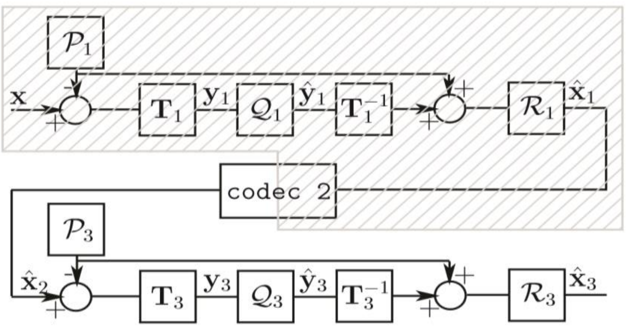

Figure 1: A series of three coding steps

Limitation: It deals with a closed-group setting where the different codec implementations are known and can be enumerated.

In [5] three motion estimation algorithms for H.264 standard have been implemented and some of its features have been tested. These algorithms are full search and two fast search method. In Full search method in this method, all possible modes are checked, with performing the motion estimation for every blocks and calculating the R-D criterion for all of them. This is time consuming. In Fast search method the fast motion estimation with modified diamond search for variable block sizes is performed. The Motion vector field adaptive search technique (MVFAST) uses a different initial search point and it finds patterns with selective application of large diamond search (LDS) and small diamond search (SSD) based on the characteristics of motion activity assessed by the similarity of motion vector field among contiguous blocks.

In [6] a modified method is formulated for the codec identification algorithm to increase its robustness to the masking introduced by the second coding step. In this new algorithm it does not require a linear behavior between the Quantization Parameter (QP) and its Peak Signal to noise Ratio (PSNR), which was already assumed in [4] and [8]. In this method it receive the input after the second encoder C_2, it decodes it to $\hat{x}_2$ and then re-encodes this sequence using a third codec C_3, obtaining $\hat{x}_3$. Repeating the encoding $|C|\times|QP|$ times and for each pair, the PSNR between $\hat{x}_2$ and $\hat{x}_3$ on a frame by frame basis is calculated. In this method the complexity can be reduced by reducing the temporal resolution that is by increasing ωT. For an example it is possible to reduce the computational complexity by a factor 15 by using $\omega QP = 3$ and $\omega T = 5$, by decreasing the average accuracy by only 4%.

4. GOP ESTIMATION

A GOP (Group of Pictures) size is the distance between two I-frames. Estimating the period of the Group of Pictures (GOP) in the first compression is a clue to detect the double compression of the video, as it gives a way to trace the processing history of the video. In [9,10] a new forensic footprint was proposed and based on it a method for detecting whether a video has been encoded twice or not was found out and the size of the Group of Pictures (GOP) was employed during the first encoding. The footprints proves to be very robust even in realistic settings that are rarely addressed by existing techniques. Here they propose to use a robust and a distinctive footprint based on the variation of the macro block prediction types in the re- encoded P-frames.

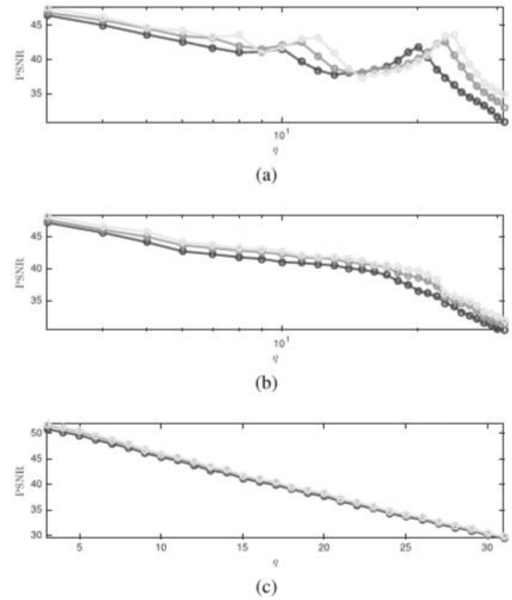

Figure 2: PSNR vs. q curves for different I-Frames of the first GOP (a) C3-MPEG2 matches C1 (b) C3-MPEG4 does not match C1 (c) C3-AVC does not match C1

This Variation of Prediction Footprint (VPF) has an advantage that it is present in the twice encoded video without the need of re-compression. Furthermore, the VPF becomes apparent only in P-frames that were intra- coded in the first encoding step, a method to estimate the size of the GOP used in the first compression was proposed. Estimation of the GOP size is not only important to assess the processing history of a digital video, but can also act as a catalyst for further forensic analysis, e.g., tampering detection.

In [10] the algorithm is proposed for blind GOP period detection from the decoded frames of a video and it is based on the Mean Absolute Motion Error (MAME) computation of each frame with reference to the previous frame. GOP Period estimation is done using motion error in frames.

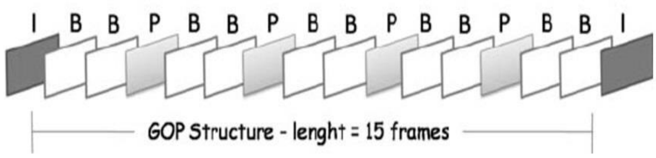

Figure 3: GOP Structure

In [6] a simple and effective method is proposed for periodicity estimation that enables GOP detection at reduced complexity. The GOP is estimated as follows

i) For every $c_3 \in C$

First compute the absolute value of Fourier transform of every row of P''_{c3} and then average the result along every column to obtain g_{c3}.

$$g_{c3} = \frac{1}{|QP|} \sum_{q \in QP} |F\{ P''_{c3}(q, .)\}| \qquad (1)$$

Where P''_{c3} is the matrix containing the PSNR values of $\hat{x}_2$ and $\hat{x}_3$.

Next use the filter bank to find out the periodicity of the peaks in g_{c3}, which will give an estimate value of the GOP size.

It is shown that the time required to analyze a ten seconds sequence with this simple implementation using available laptop with Intel Core i7 processor is approximately 100 seconds. Complexity can be reduced by reducing the number of tested QP because it does not affect the accuracy.

5. CONCLUSION

This paper gives a brief information about the various methods used to determine the video codec and its GOP size. It also gives a method to stabilize the video, in case of blur videos. It gives information about the drawbacks of some methods used. The complexity of the methods can be reduced by skipping one step, but it will affect the accuracy. Moreover, in forensics time is not a constraint, and accuracy is more important. The further development in this paper would be to add new technologies that will provide efficient and accurate results faster, with less complexity. The other features of video like the date and the time when the video was taken will be the topic of our future research.

REFERENCES

[1] Lei Zhang, Qian-Kun Xu, Hua Huang, "A Global Approach for Fast Video Stabilization" *IEEE Transactions on Circuits System and Video Technology 2015.*

[2] Xiaohong Ma, Zhongyin Cao, "Video Stabilization Based on Adaptive Local Subspace of Feature Point Classification" *IEEE 2016*.

[3] Lei Zhang, Xiao-Quan Chen, Xin-Yi Kong, Hua Huang, "Geodesic Video Stabilization in Transformation Space" *IEEE Transactions On Image Processing, 2017.*

[4] P. Bestagini, A. Allam, S. Milani, M. Tagliasacchi, and S. Tubaro, "Video codec identification" *in Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP),* Mar. 2012, pp. 2257–2260.

[5] A.Ahmadi, M.M.Azadfar, "Implementation of fast motion estimation algorithms and comparison with full search method in H.264" *IJCSNS International Journal of Computer Science and Network Security, VOL.8 No.3, March 2008*.

[6] Paolo Bestagini, Simone Milani, Marco Tagliasacchi, Stefano Tubaro, "Codec and GOP Identification in Double Compressed Videos", *IEEE Transactions On Image Processing, Vol. 25, No. 5, May 2016*.

[7] P. Bestagini, S. Milani, M. Tagliasacchi, and S. Tubaro, "Video codec identification extending the idempotency property," *in Proc. 4th Eur. Workshop Vis. Inf. Process. (EUVIP), Jun. 2013, pp. 220–225.*

[8] D. V´azquez-Pad´, M. Fontani, T. Bianchi, P. Comesa˜na, A. Piva, M. Barni, "Detection of video double encoding with GOP size estimation", *in National Inter-University Consotium for telecommunications Marta 3, 50139 Firenze, Italy. 2012 IEEE.*

[9] Perla Ramakrishna, Aniruddha Mazumdar and Prabin K. Bora, "Blind Forensics Method for GOP Period Detection in Motion Compensated Video", *2016 IEEE.*

[10] Mahdi, O.A., Mohammed, M.A. and Mohamed, A.J., 2012. Implementing a novel approach and convert audio compression to text coding via hybrid technique. *International Journal of Computer Science Issues,* 9(6), pp.53-59.