Neurocomputing 226 (2017) 35–45

Contents lists available at ScienceDirect

Neurocomputing

journal homepage: www.elsevier.com/locate/neucom

Adaptive weighted non-parametric background model for efficient video coding

Subrata Chakraborty[a,*], Manoranjan Paul[b], Manzur Murshed[c], Mortuza Ali[c]

[a] School of Management & Enterprise, University of Southern Queensland, Sinnathamby Boulevard, PO Box 4196, Springfield Central, Queensland 4300, Australia
[b] School of Computing & Mathematics, Charles Sturt University, Panorama Avenue, New South Wales 2795, Australia
[c] School of Information Technology, Federation University, Northways Road, Churchill, PO Box 3191, Gippsland Mail Centre, VIC 3841, Australia

ARTICLE INFO

Keywords:
Background model
Coding efficiency
Coding performance
HEVC
Non-parametric model
Video Coding

ABSTRACT

Dynamic background frame based video coding using *mixture of Gaussian* (MoG) based background modelling has achieved better rate distortion performance compared to the H.264 standard. However, they suffer from high computation time, low coding efficiency for dynamic videos, and prior knowledge requirement of video content. In this paper, we introduce the application of the *non-parametric* (NP) background modelling approach for video coding domain. We present a novel background modelling technique, called *weighted non-parametric* (WNP) which balances the historical trend and the recent value of the pixel intensities adaptively based on the content and characteristics of any particular video. WNP is successfully embedded into the latest HEVC video coding standard for better rate-distortion performance. Moreover, a novel *scene adaptive non-parametric* (SANP) technique is also developed to handle video sequences with high dynamic background. Being non-parametric, the proposed techniques naturally exhibit superior performance in dynamic background modelling without *a priori* knowledge of video data distribution.

1. Introduction

The latest HEVC video coding standard [1] has improved the coding performance by applying a number of innovative tools compared to its predecessor H.264/AVC [2,3] including a wider range of variable block size motion estimation (ME), motion compensation (MC), prediction, and transformation units. The use of multiple reference frames (MRFs) with variable block sizes typically provides better coding performance than the single reference frame approach [4] for video with repetitive motion, uncovered background, non-integer pixel displacement, lighting change, etc. However, MRFs-based schemes require index codes to identify a particular reference frame and the computational time increases almost linearly with each additional reference frames due to ME and MC. The decision on appropriate number of reference frames is dependent on the video content and the computational time constraint which may not always allow large number of reference frames [5,6].

Some fast coding techniques [7–10] have achieved significant time saving compared to H.264 but failed to outperform it in coding performance for challenging video sequences [5]. Dual reference frames based schemes [11,12] try to solve the challenges in MRFs by using only two reference frames where the immediate previous frame is used as the short term reference (STR) frame and a frame from previously coded frames is used as long term reference (LTR). The rationality of dual reference frames is to use STR frame for local motion and LTR frame for background or global motion. Video segmentation based coding techniques exploit the stable parts in a frame by treating them as background [13–15]; however they are highly computationally intensive. Object segmentation based sprite coding techniques [16,17] were also introduced but they suffer from high computation burden and their performance degrades at high bit rates [18]. A newer method called sparse coding has been successfully used in image deniosing [19,20]. A recent study showed good performance using sparse coding technique where significant efficiency is achieved against the HEVC [21]. Although the new technique shows good performance study showing performances in HD videos or videos with background motion is not currently available.

A video coding scheme named McFIS (most common frame in a scene) [5,6,22,23] was introduced to utilise the highly accepted Mixture of Gaussian (MoG) dynamic background modelling (DBM) technique. The McFIS scheme further instilled the fact that using a good quality background frame as a reference frame improves coding

* Corresponding author.
E-mail addresses: subrata.chakraborty@usq.edu.au (S. Chakraborty), mpaul@csu.edu.au (M. Paul), manzur.murshed@federation.edu.au (M. Murshed), mortuza.ali@federation.edu.au (M. Ali).

http://dx.doi.org/10.1016/j.neucom.2016.11.016
Received 14 January 2015; Received in revised form 13 September 2016; Accepted 14 November 2016
Available online 19 November 2016
0925-2312/

performance and efficiency compared to the MRFs using a number of previously decoded frames. It also established the need for using a good DBM technique for practical usage. A number of studies have highlighted the improved performance and efficiency of the DBM based McFIS scheme [5,6,22–24]. DBM is also applied in recent studies in the area of transcoding technique for video surveillance [25].

MoG based DBM works at the pixel level where each pixel of a scene is modeled independently using a mixture of Gaussian distributions (generally 3–5) [24–29]. Although the MoG based DBM has proved successful and is widely used by researchers and practitioners, it requires the user to assume the data distribution in advance and relevant parameters must be set based on this underlying assumption. It also performs poorly for fast changing background environments [28,30]. A relatively new non-parametric (NP) technique [28,30,31] has gained the attention of many researchers due to its ability to perform well in highly dynamic scenarios and provides a set of stable parameters with no requirement for an initial assumption about the underlying data distribution [28,30,32].

Although the high sensitivity to dynamic background makes the NP technique very attractive for applications such as object detection and tracking, this poses a serious challenge for video coding. If the background is updated frequently then the insertion of more reference frames would be necessary, leading to a higher bit rate requirement. Also, in the existing NP technique [28,30] the background is generated using the pixel intensity values of the last frame only and the historical pixel intensity values are only used for the purpose of probability estimation. Hence, the background is heavily biased towards the last frame and loses the historical trend value. To resolve these issues we propose a new *weighted non-parametric* (WNP) technique where we generate more stable background using historical pixel values and pixel values from the latest frame. The new technique uses the weighted average of (i) a probabilistic pixel intensity value (calculated based on the median of historical values and randomly scaled standard deviation) and (ii) the latest pixel intensity value. The weight ratio between the historical and the recent pixel values is decided in an adaptive manner considering specific video contents. An intuitive weight ratio (α) selection procedure is presented which selects the most appropriate α based on the quality of backgrounds produced by a set of potential weight ratios. The proposed WNP technique inherits the advantages of the NP technique such as the capacity to better detect dynamic backgrounds and the ability to perform probability estimation with dynamic data distribution. The additional ability of WNP to provide a more stable background makes it more suitable for video purposes as it reduces the computational time significantly and provides better video coding performance.

The WNP technique is further modified to develop the *scene adaptive non-parametric* (SANP) technique which eliminates the need for coding of the background reference frame by utilizing the coded frames to generate the background reference frame. The key advantage of SANP is the ability to adapt quickly to subtle background changes which are not generally considered as scene changes. This adaptability increases the usability and improves the coding performance for videos with high dynamic backgrounds.

Note that the main objective of video coding applications using background modelling is to reduce the residual error for improving compression performance, while the main objective of object detection applications is to find the object in its original shape. In this study we developed novel schemes to exploit the parameter stability and superior change adaptation capabilities of the NP technique and make it applicable for video coding applications.

We present the motivations and contributions of this study in section II, followed by the proposed WNP background modelling technique and the adaptive αselection procedure along with the SANP technique in Section III. Extensive experimental results are presented in Section IV followed by conclusive remarks in Section V.

2. Motivations and contributions

Traditional DBM is performed at pixel level, where each pixel of a frame is modeled independently by a mixture of L (normally 3–5 models are used) Gaussian distributions [26,27,29]. Each Gaussian model represents the intensity distribution of one of the different environment components e.g., moving objects, static background, shadow, illumination, cloud changes, etc. observed by the pixel in frames. We assume that the l-th Gaussian at time t representing a pixel intensity is $\eta_{l,t}$ with mean$\mu_{l,t}$, variance$\sigma_{l,t}^2$, and weight $w_{l,t}$ such that $\sum w_{l,t} = 1$ for all l. A learning parameter $\Omega = 0.1$ [30] is used to balance the contribution of the current and past values of parameters such as weight, variance, mean, etc. Then $1/\Omega$ defines the time constant which determines the speed at which the distribution's parameters change. The system starts with an empty set of models and then for every new observation Z_t at the current time t, it is first matched against the existing models in order to find one (say the l-th) such that $|Z_t - \mu_{l,t}| <= 2.5\sigma_{l,t}$. If such a model exists, its associated parameters are updated. Otherwise, a new Gaussian is introduced with $\mu_{l,t} = Z_t$, arbitrarily high σ=30, and arbitrarily low w=0.001 by evicting $\eta_{l,t}$ if it exists [30].

Although the MoG based DBM such as the McFIS-DBM showed better performance compared to H.264, the key challenges with the MoG based DBM is the appropriate parameter value selection. The number of distribution model L if selected high may improve the background stability with more computation as a drawback. On the other hand using small number of L will lead to background being changed frequently thus loosing stability from the coding perspective. Similarly the learning parameter Ω has proportional impact on convergence. A value of $\Omega = 0.1$ will need 10 frames to update a background. Higher Ω value will provide less stable background as it uses less number of frames while smaller Ω value will provide more stable background by using large number of frames. It is very difficult to accurately identify the appropriate parameter values as they are highly dependent on the video sequence and require precise setting to get the best results.

These challenges led us to our first contribution "***Contribution 1****: Integrating traditional NP technique into the HEVC video coding scheme*". The NP technique [25,27] is able to work well without the explicit parameter settings required by MoG. It is also found to be performing better than MoG in object detection applications. We develop a DBM based coding scheme where the NP technique is used for generating the background frame to be used as an LTR frame during coding.

Although the NP technique performs well for object detection applications, the NP based coding scheme is not the best for video coding applications. We have found that the NP based scheme performs better than HEVC (with 2 reference frames) and almost as good as the MoG based scheme. Through further investigation we have identified that the background frame development process of the NP scheme is responsible for its ordinary coding performance. In the traditional NP technique parameter estimation is conducted based on the historical pixel values, however the background is generated based on a pixel's recent value only. Using the recent values of the pixels may be fine for object detection but it is not appropriate for coding purpose. For improving the coding performance we require a stable background frame comprising static and uncovered background areas so that the motion estimation can be reduced. The traditional NP technique is unable to provide a stable background frame as the previously uncovered background area is lost from the frame due to the recent pixel value usage.

The quest for a more stable background using the NP technique led us to our second contribution "***Contribution 2:*** *Developing the weighted non-parametric (WNP) technique*". We have developed the novel WNP technique where the background is generated by incorporating both the historical and the recent values of the pixels. This