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

A R T I C L E I N F O

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

A B S T R A C T

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$\sigma^2_{l,t}$, 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 process helps retaining the previously uncovered background areas and also incorporating newly found background areas. The generated background is much more stable and suitable for coding applications.

WNP showed much improved coding performances in terms of rate-distortion performance; however, the key challenge with WNP is to determine the appropriate ratio between the historical values and the recent value of a pixel while generating the pixel value for the background frame. The ratio is found to be dependent on the video sequences. The challenge to selecting the appropriate ratio for every video sequence leads us to devise an adaptive background pixel value selection process. We develop an adaptive process which has the capability to test various ratios (define as α) for best suitability for any particular video sequence. The process selects the best ratio based on the amount of background detection.

The WNP technique shows superior results, however, it requires better adaptation to highly dynamic scenarios. This leads us to our 3rd contribution "***Contribution 3:*** *Developing the scene adaptive non-parametric (SANP) technique*". Unlike the WNP where original frame is used, the SANP technique uses coded frames to generate the background reference frames and has the capabilities to update the background frame with each newly coded frame. This recursive updating process makes SANP more adaptable to subtle changes in the background. In section III we describe the contributions in detail followed by extensive experimental results in section IV.

3. Proposed techniques

First we incorporated the NP technique (traditionally used for object detection) [28,30] for video coding purpose. We applied the NP used for object detection to identify foreground and background pixels and retain the background pixel values in the background frame.

We then develop WNP and SANP techniques based on the NP coding in this study which use a dual reference frames technique where the LTR is a high quality background frame (generated by the proposed WNP background modelling) and the STR is the immediate previous frame of the current frame. The background frame is modeled with a small number of original frames of a scene and all frames in the scene are encoded using the LTR and STR frames. The Lagrangian multiplier [33,34] is finally used to select the reference frame for a block.

In the proposed techniques we also incorporate a *scene change detection* (SCD) strategy to trigger the background frame reset. The SCD is determined applying a simple metric computed utilising the McFIS and the current frame. The *sum of absolute difference* (SAD) between the McFIS and the current frame is computed. If the SAD for the current frame is 70% greater than that of the previous frame of a scene, we consider the SCD occurred [6]. When a scene change occurs we reset the background modelling and generate a new background frame for the new scene. To reduce the computational time, we only use a small search range for ME and MC when the LTR frame is used as a reference. As the LTR frame is referenced mainly for static and uncovered areas, we do not need to use a large search range. Thus, we can reduce the computational time significantly with respect to HEVC.

The proposed WNP technique is primarily based on the well-known NP technique [28,30,31]. The WNP aims to generate a stable background frame from a set of initial training (input) frames. The generated background frame integrates both historical pixel intensity value and recent pixel intensity value from the last training frame for each pixel. This retains the past pixel intensity trend as well as the recent pixel value which in turn will provide a more stable background for video coding purposes. The use of historical pixel values also eliminates a sudden pixel intensity change (i.e., noise or dynamic background impulse) of the latest frame compared to the historical trend. For object detection applications, considering the pixel intensity of the latest frame to generate a background frame as done in the NP technique provides a better object in dynamic background situations. However, considering a combination of the pixel intensity of the latest frame and the historical pixel value provides better rate distortion performance in video coding applications.

The NP, WNP and SANP techniques uses background modelling and background generation method as described by following steps:

3.1. Step 1: Background probability estimation

The background probability estimation is based on the traditional NP [28,30,31] used for object detection. The same process is used for NP, WNP an SANP coding techniques.

Given $x_1, x_2, \ldots, x_N$ is a set of past intensity values for a pixel in consecutive N frames in temporal order, we can estimate the probability density function with pixel intensity x_t at time t by using kernel estimator $K(.)$ as:

$$p_r(x_t) = \frac{1}{N}\sum_{i=1}^{N} K(x_t - x_i) \tag{1}$$

where $K(.)$ is a kernel estimation function.

If we consider the kernel estimator function $K(.)$ to be a Normal function $N(0, S)$ where S is the kernel function bandwidth, then the density estimation can be determined as:

$$p_r(x_t) = \frac{1}{N}\sum_{i=1}^{N} \frac{1}{(2\pi)^{\frac{d}{2}}|S|^{\frac{1}{2}}} e^{-\frac{1}{2}(x_t-x_i)^T S^{-1}(x_t-x_i)} \tag{2}$$

With the assumption of independence between the different color channels d and with different kernel bandwidths σ_j^2 for the jth color channel we can write:

$$\sum = \begin{pmatrix} \sigma_1^2 & 0 & 0 \\ 0 & \sigma_2^2 & 0 \\ 0 & 0 & \sigma_3^2 \end{pmatrix}$$

The density estimation equation then can be reduced to:

$$p_r(x_t) = \frac{1}{N}\sum_{i=1}^{N}\prod_{j=1}^{d} \frac{1}{\sqrt{2\pi\sigma_j^2}} e^{-\frac{1}{2}\frac{(x_{t_j}-x_{i_j})^2}{\sigma_j^2}} \tag{3}$$

In order to estimate the kernel bandwidth σ_j^2 for jth color channel for any given pixel we need to compute the median absolute deviation over the sample for consecutive intensity values of the pixel. The median difference m can be calculated as

$$m = median\ (|x_i - x_{i+1}|) \text{ where } i = 1, 2, \ldots, N-1.$$

We can consider that the intensity pair being consecutive comes from local-in-time distribution. Assuming that the local-in-time distribution is Normal $N(\mu, \sigma^2)$, then the deviation $(x_i - x_{i+1})$ is also Normal $N(0, 2\sigma^2)$. The standard deviation of the first distribution can be computed as:

$$\sigma = m/(0.68\sqrt{2})$$

3.2. Step 2: Background generation

With the NP technique, the calculated probability is compared against a threshold to determine if the current pixel belongs to the background or foreground. If the pixel is considered as background then its value is retained in the background frame. In the proposed WNP technique we also determine the background using a threshold recommended in [30]. Once the current pixel value is detected as background, the value is combined with the synthesized historical pixel intensity of the given sample for that particular pixel. The synthesized historical pixel value M for the sample pixel values can be calculated as

$$M = median(x_1, x_2, \ldots, x_N) + \sigma * Y \tag{4}$$

The synthesized value M is a mimic of the actual pixel intensity generated using the standard deviation σ and a very small normally distributed random multiplier Y where the component wise median of the pixel intensities is used instead of the mean. M is actually a generated value from a normal distribution with the median as the mean and σ as the standard deviation.

The value X of the background pixel can be calculated using weighted average of the recent value x_t and the sample synthesized value M:

$$X = x_t * \alpha + M * (1 - \alpha) \tag{5}$$

In this way, a background frame is generated for the entire frame and then encoded with high quality. The corresponding quantization set up is similar to [6].

The value of α can be adjusted to provide more importance to either the recent value or the historical value trend for a given pixel. We have conducted simulation studies which show that the impact of α is dependent on the video dataset. Please note that when we use $\alpha = 1$ in the proposed method, the proposed method is equivalent to the traditional NP technique as the background is developed solely based on the most recent pixel value. Our studies with other video data sets suggest that finding the right balance between historical and recent pixel value is the key to selecting an appropriate α value. In order to select the α value we have developed an adaptive α selection procedure described in 6 stages:

1) Decide the potential α values

   At this stage we decide a set of potential α_p *where* $p = 1\ldots P$ is the number of α values we are interested in for a given video data. The α values are between 0 and 1 and may be selected at regular intervals.

2) Generate background for each α

   A background frame for each α is generated by applying the background generation method using a number of frames (described earlier). We can denote the backgrounds as B_p where $p = 1\ldots P$.

3) Calculate pixel intensity variation

   With the set of training frames F_q, where $q = 1\ldots Q$ with width W and height H, we find the intensity difference I for each pixel of each frame with the corresponding pixel for each background can be calculated as:

$$I_{pq}(w, h) = |B_p(w, h) - F_q(w, h)| \tag{6}$$

   where $p = 1\ldots P$; $q = 1\ldots Q$; $w = 1\ldots W$; $h = 1\ldots H$

4) Count background pixel detection

   The background detection percentage represents the area of a frame detected as background by the new WNP method using a particular α value. The total number of pixels in a frame is $W*H$. The number of pixels detected as background in a frame using a particular α value can be calculated applying (6) as

$$C_{pq} = Count(I_{pq}(w, h) <= \delta) \tag{7}$$

The small value δ is the threshold for a pixel to be considered as background. Ideally, the value of δ should be "0". However, the WNP develops a synthesized background using historical and actual pixel values which leads to deviation from the actual pixel value and thus a nonzero δ value for some cases. With a significant amount of experimental studies using test videos we have identified that generally a large change in pixel value classifies it as foreground, while only a small change classifies it as background. Experimental results indicate that up to 2% deviation from the highest possible pixel value (i.e. 255) may be used as the δ value.

5) Calculate background detection percentage

   By applying (7) we can calculate the percentage of total pixels detected as background for a particular frame by:

$$U_{pq} = (C_{pq} * 100)/(W * H) \tag{8}$$

The overall background detection percentage for a test video is calculated based on the number of training frames F_q. The overall background detection percentage for the set α_p can be calculated as:

$$U_p = \left(\sum_{q=1}^{Q} U_{pq}\right)/Q \tag{9}$$

6) *Select α for a video*

   The α value for a given video is the α_p value for the maximum U_p value.

$$\alpha = \alpha_p \text{ for } \max(U_p) \text{ where } p = 1, 2, \ldots P \tag{10}$$

Our experimental results show that the α value that helps detecting the maximum percentage of frames as background provides a better background reference frame for that video leading to better coding performance. The α values were found to be positively correlated with the coding performance (PSNR) of any test video. This objective and adaptive selection procedure for the α value provides the decoder with the ability to generate the best background to achieve the best possible coding quality.

3.3. Step 3: Integrate into the HEVC coding scheme

The background generation process described in Step 2 can be incorporated into the coding process using the following coding scheme.

As shown in Fig. 1, the initial training frames are stored in a frame buffer while HEVC is used to code these frames. Once the buffer is full, the WNP background generator develops the background frame using the frames in the buffer. This background frame is then used by the McFIS based video coder as a reference frame. The coding continues until a scene change is detected. On detection of a scene change, a new

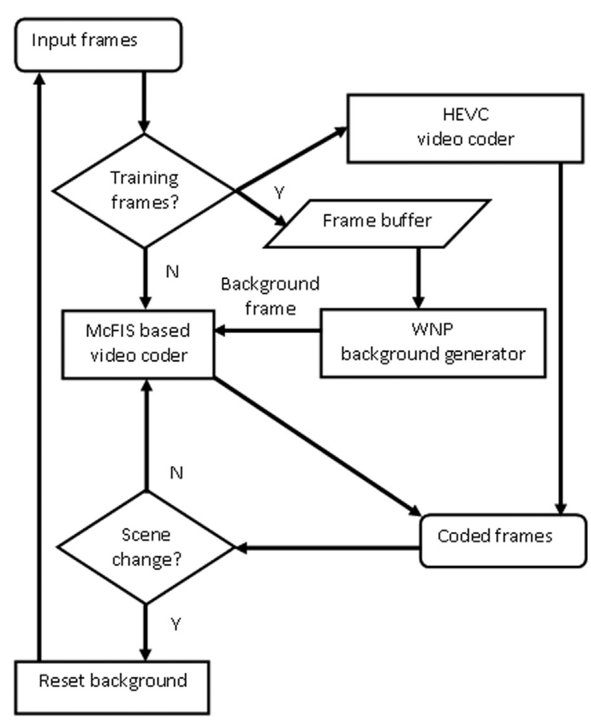


Fig. 1. WNP integrated coding scheme using original frame.

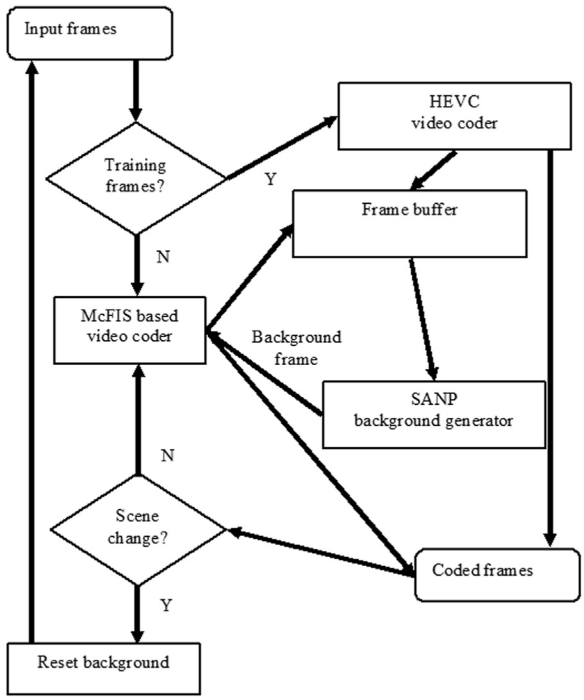


Fig. 2. SANP integrated coding scheme using coded frame.

background frame is generated by populating the frame buffer. In this scheme a HEVC coder is used during the background generation process.

The SANP technique mimics the WNP in generating the background and selecting the αvalue. However in SANP coded frames are used to generate the background frame instead of the original frames as proposed in WNP. The initial background is generated using HEVC coded initial frames and continuously updated with each subsequent coded frame. The fundamental difference is that the WNP technique uses original frames to generate the background frames while the SANP technique uses coded frames.

The SANP background generation process defines a size for a frame set required for generating the background. This set is iteratively updated with each coded frame as shown in Fig. 2 Let us assume that the set size is determined to be 25 frames. At the beginning, HEVC coder is used to code the first 25 frames which are then sent to the frame buffer (size 25). When the buffer is full, the WNP based background generator is used to generate the background frame which is then used by the McFIS coder to encode frame no. 26. The coded 26th frame is then sent to the frame buffer and the 1st frame is discarded from the buffer. A new background is then generated from the frames in the buffer. The new background reference frame is generated by taking the average of the current reference frame and the newly created background frame. This process is repeated for each new coded frame where the oldest frame in the buffer is replaced with the newest coded frame in a first-in-first-out sequence.

Let us consider that the background frame generated by the SANP background generator is denoted as $G$with width y and height z. Each pixel Xin frame Gcan be calculated using the WNP background generation process. Each pixel in the background frame can be referred as $X_{G,y,z}$. Initially this background frame G is the reference frame R. Each pixel of the reference frame can be obtained as:

$$X_{R,y,z} = X_{G,y,z} \tag{11}$$

With each newly coded frame a new background frame G is generated by SANP background generator. The reference frame R is updated with each new coded frame by updating the pixels of the reference frame as:

$$X_{R,y,z} = (1 - \beta)X_{R-1,y,z} + \beta X_{G,y,z} \tag{12}$$

where β is the integration ratio between current background and newly generated background which reduces the impacts of past values of any pixel at a rate of $1/\beta$ with each iteration. The SANP scheme maintains the stability of the background reference frame by incorporating the pixel values of all the past background frames in the new reference frame. In general $\beta = 0.5$ provides a good integration ratio for most applications; however β value can be adjusted to suit specific application requirements. With higher β value, the recent changes are given more importance while generating the background.

When a scene change is detected the background reference frame is reset to initial state. HEVC coder is used until another background reference frame is generated. The SANP scheme eliminates the need for coding the McFIS (i.e. the background frame) as the scheme can be used simultaneously at both the coding and decoding end of the coder. The continuous updating of the background reference frame makes it adaptive to changes in the input frames thus improving the coding performance.

The SANP integration into a coding scheme is shown in Fig. 2. In this scheme the initial training fames are coded using the standard HEVC coder. The coded frames are then used by the SANP background generator to generate a background frame to be used as a reference frame by the McFIS based coder for coding the next new frame. Each coded frame is then used by the SANP background generator to develop a new background frame to be used as reference while coding the next new frame. The key difference between the WNP and SANP schemes is that, in WNP the background is updated only when there is a scene change while in SANP the background is updated with each new coded frame. Due to the continuing background updating in SANP it obtains better RD performance where the background is dynamic however it requires more time for the updating process.

4. Experimental results & discussions

4.1. Experiment setup

The experiments were conducted on a dedicated desktop machine running 64 bit Windows operating system. The program codes for the NP technique were available from the author [28] by personal communication. Each test video has been decoded using the HEVC, NP, MoG and WNP based coder-decoder. For HEVC we used GOP of size 32. For Inter coding we used 2 reference frames. Motion search was performed within a window of ±31. For NP, MoG and WNP, the generated background frame by each technique was used as a reference frame along with the immediate previous frame for coding-decoding purpose. The *standard definition* (SD) test video sequences *Sales*, *News*, *Grandma* and *Table Tennis* are of resolution of 176×144, while the *Silent*, *Paris* and *Container* test video sequences are of resolution of 352×288. We have also used two *high definition* (HD) videos *Blue sky* and *Pedestrian* with resolution of 1088×1092. Table 1 provides detail characteristics of the test video sequences we used. First 25 frames were used as training frames of a scene to generate the background frame for the MoG, NP and WNP techniques.

For the original frame based WNP coding scheme the initial total training frames were chosen to be 25 because it provided a buffering time of <1 s at standard 30 fps. This is acceptable for most of the applications. A set of 25 frames was also large enough to provide a good quality background for all our test video sequences. However, the size of the training frame set can be adjusted according to available buffering time and required background quality. We have used similar coding scheme for MoG and NP where the background frame was generated using 25 frames and then coded as an *I*-frame.

For the SANP based scheme, we have coded first 25 frames using the HEVC technique. These coded frames are then used to generate the reference background frame. For the NP and MoG technique similar

Table 1
Test video characteristics.

Video sequences	Resolution	Camera motion	Content
Sales	176×144	Static	Slow hand movement and lips movement with static background.
News	176×144	Static	Two news readers in the foreground with light lip and head movement. The background contains heavy movements with two ballet dancers dancing.
Silent	352×288	Static	A lady frequently moving one hand in front of a static background uncovering previously occluded background areas.
Paris	352×288	Static	Static background with two people talking in the foreground with lot of complex motions including one person juggling with a ball.
Grandma	176×144	Static	Static background with almost static foreground where an elderly lady is very slowly moving her lips and head.
Table Tennis	176×144	Zoom, Scene change	Very fast moving foreground with mostly static background where a player repeatedly striking a table tennis ball. There are two zoom operations and scene changes in this video.
Container	352×288	Static	Slow moving large container ship in the foreground with highly complex and dynamic background including smaller fast moving boats and water with waves.
Blue sky	1920×1088	Camera motion	Very slow camera motion. Static sky background but very complex foreground where there are complex changes including moving tree leaves and sunlight. The amount of background is comparatively small.
Pedestrian	1920×1088	Static	Highly dynamic foreground where large number of pedestrians are crossing a street section. Static background is formed by building walls where occluded areas are uncovered. Frames consists of larger amount of foreground compared to the background.
Traffic	2560×1600	Static	Smooth motion where road traffic is recorder.
Exit	640×480	Static	People moving through an exit door. Large static background where occluded area is uncovered.

coding schemes were applied where the backgrounds were generated and continuously updated using the coded frames.

We have used 5 different quantization values (QP=28, 24, 20, 16, and 12) for each sequence to obtain different bit rates and the corresponding PSNR values. Then each of the four coding techniques for each scheme was used to code the video sequences.

We have applied the α selection procedure described earlier to select appropriate α for each video sequence. We have set the αselection parameters P=9 (i.e. 9 α values between 0 and 1 were tested to find the best one) and Q=25 (i.e. 25 trainings frames were used). Table 2 shows the amount of background detection (in %) we achieved under various α values for our test videos. Our α selection process successfully detects the most suitable α value for each video thus ensuring that the coding performance is maximized.

4.2. Results and discussions

Our proposed WNP and SANP techniques have shown much better coding performance in terms of PSNR at the same bit rates. Figs. 3 and 4 highlights the performance gains of both the WNP and SANP techniques against comparable schemes of the NP and MoG techniques and the standard HEVC technique for the *Grandma*, *Silent* and *Sales* videos. For our set of test videos WNP and SANP achieved between 0.5 and 1 dB PSNR gains over HEVC and comparable MoG and NP schemes. The performance of the proposed techniques were much better in high bit rate scenarios which makes them more suitable for contemporary applications due to their high bit rate requirements.

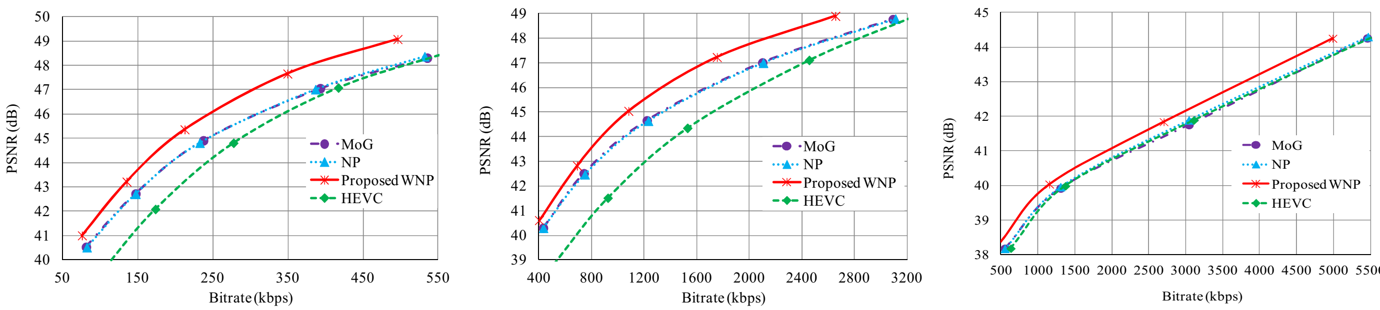
Detail comparative PSNR performances of all the techniques for the set of test videos are provided in Table 3. We observe that the WNP technique performs best for the *Sales*, *Silent*, *Grandma* and *Container* videos compared to MoG, NP, and HEVC techniques. These four videos have some common features including very stable background and less exposer of occuluded areas. The WNP technique was able to generate a background reference frame which did not change much throughout the coding process thus providing better coding performance. The WNP technique was also the best performer for the HD videos *Bluesky* and *Pedestrian* and *Exit*. The SANP performed very close to the WNP in these videos but showed much superior performances compared to HEVC and coded frame scheme based NP and MoG techniques. In the *Bluesky* and *Pedestrian* and *Traffic* HD videos the performances of WNP and SANP is very close to HEVC, MoG and NP methods. The reason for such close RD performance is due to the content of these videos. In *Bluesky* there is complex camera rotating motion and the foreground consists of more than 60% of the frame which gradually keeps increasing over the time. The foreground also has significant motion including moving leaves and changing sunlight coming through leaves. In the *Pedestrian* video the foreground consists of more than 80% of frame area and involves lot of frequent movements where large number of pedestrians and cyclists move around. Due to a smaller background frame area and highly comples foreground movements on these two videos, our proposed methods does not gain significant performance gain over other methods as they utilise background detection efficiently for performance enhancement. However they perform as good as other methods and save computational time over HEVC. For the *Exit* video where there is about 70% background frame area, proposed SANP and WNP outperforms other techniques with significant margin as shown in Figs. 3(c) and 4(c).

The SANP performed the best for the *News*, *Table* and *Paris* videos. These videos are challenging due to their dynamic characterstics. The *News* video contains dynamic background (ballet dancers performing at the back), the *Table* video contains multiple scene changes, and the *Paris* video have several exposers of previously occuluded regions at different stages of the video. The inherent scheme of the SANP technique have the capacity to update the background reference frame quickly thus incorporating the challenging dynamic characterstics of the video background. We observe that WNP performs slightly better than SANP in *Silent* and *Grandma* videos. These two videos have static background with small amount of exposer to previously occluded areas. This characteristic of these two videos does not utilise the strengths of SANP rather performance is suffered slightly by the use of coded frames. WNP works by using original frames and distortion occurs when we encode the McFIS, whereas SANP works by using coded frame and distortion occurs in the coded frames. Thus, for the static background and lengthy scene, WNP should outperform the SANP and for the dynamic background and shorter scene change SANP should outperform WNP.

In order to better understand the effectiveness of the SANP and WNP techniques we present frame by frame performance analysis in Fig. 5. In Fig. 5(a) for the *Silent* video we observe that the WNP shows superior performance trend compared to SANP. Fig. 5(b) shows that WNP has a higher background frame usage as a reference frame during coding leading to better performance of WNP for this video. We observe a contrasting results for the *Table* video. Fig. 5(c) shows that SANP has a better performance trend than WNP which is further

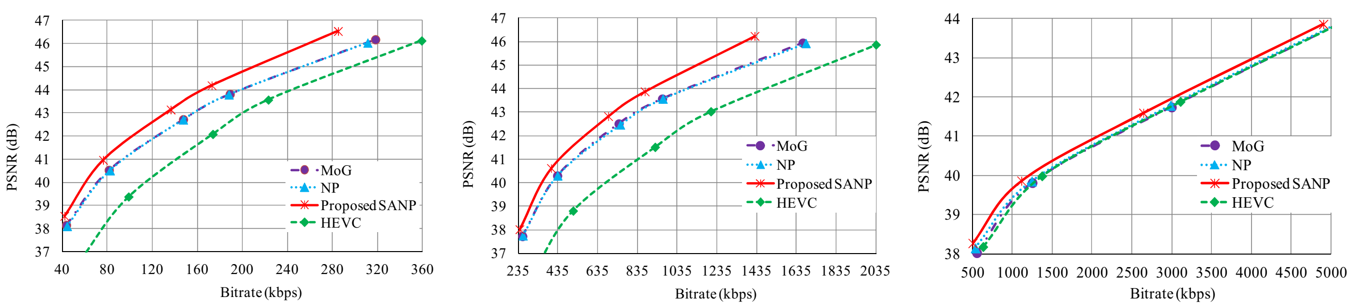
Table 2
Background detection percentage for test videos at various α.

α	Test videos									
	Sales	*Paris*	*Silent*	*News*	*Grandma*	*Table*	*Blue sky*	*Pedestrian*	*Traffic*	*Exit*
0	82.03	76.62	81.56	89.38	96.01	80.61	70.51	68.71	65.84	75.87
0.15	82.12	**76.73**	**81.67**	**89.39**	96.28	**80.73**	**70.73**	**69.12**	**66.12**	**76.35**
0.25	**82.17**	76.68	81.65	89.33	**96.46**	80.48	70.64	68.82	66.04	75.93
0.4	82.01	76.26	81.26	89.12	96.06	79.08	69.22	67.37	65.64	75.16
0.5	81.88	75.72	80.78	88.84	95.06	77.19	66.79	66.93	65.31	74.38
0.65	81.78	74.85	80.09	88.24	92.82	73.41	65.61	66.18	65.11	73.64
0.75	81.68	74.11	79.64	87.81	91.20	70.43	62.83	65.63	64.86	72.81
0.9	81.66	73.52	79.19	87.42	89.10	66.88	60.19	64.79	64.29	71.63
1	82.04	73.96	79.46	88.57	92.52	72.89	68.75	68.55	65.58	75.62



(a) Result for *Grandma* Video (b) Result for *Silent* Video (c) Result for Exit Video

Fig. 3. Coding performance comparison between the WNP and other *original frame* based techniques (a) *Silent* video frame by frame PSNR comparisons using two different schemes (b) Percentages of background frame usage as a reference frame for *Silent* video using two schemes (c) *Table* video frame by frame PSNR comparisons using two different schemes. (d) Percentages of background frame usage as a reference frame for *Table* video using two schemes.



(a) Result for *Grandma* Video (b) Result for *Silent* Video (c) Result for Exit Video

Fig. 4. Coding performance comparison between the SANP and other *coded frame* based technique (a) Result for *Grandma* Video (b) Result for *Silent* Video (c) Result for Exit Video.

proved by the better background usage by SANP as shown in Fig. 5(d). In the *Table* video 3 key changes happens - a) around frame 50 there is a zoom-in, b) around frame 100 there is zoom-in and then zoom-out, and c) around frame 150 there is a scene change. In Fig. 5(c) we have now highlighted these changes and we can observe spikes in RD performances at those frames. After frame 150 there are no more scene change in this test video, and the most significant observation from Fig. 5(c) is that the performance of SANP increases continuously whereas the WNP performance remains flat.

This result is due to the fact that SANP kept updating the background frame with each new frame but WNP used the last updated background frame as no scene change was detected. Fig. 5(d) complements this finding where we can observe that after frame 150 higher amount of background frame was used by SANP for coding resulting in better coding performance.

The SANP technique updates the background with every newly coded frame while the WNP only provides the background reference frame based on the initial set of original frames. The recursive updating process of SANP provides a background reference more consistent with the background changes leading to better coding performance. SANP thus works better for video sequences with high dynamic content such as moving background, unocculusion and scene changes. Similar result trends were observed for the high dynamic *News* and *Paris* videos. On the contrary WNP performed well for videos with stable background such as *Silent*, *Grandma*, *Sales* and *Container*. These results highlight the appropriateness of the two schemes for videos with different characterstics although both of them significantly outperforms HEVC, NP and MOG for different video types.

Another strength of the proposed WNP technique lies in its ability to perform significantly faster than the other major techniques discussed in this study. We observe that throughout our experiment with different test videos and at different quantization values, the WNP method is the fastest among all. Fig. 6 shows the time saving (%) for the MoG, NP and WNP technique relative to the HEVC for different test videos. These results were obtained by applying various quantization values (40 to 8) to the *Sales, News, Silent, Paris, Grandma, Table Tennis, Bluesky* and *Pedestrian* videos, and the results for quantization 24 is presented here. We can observe that the proposed WNP saves

Table 3
Coding performance of tested techniques at different bit rates for the set of test videos.

Test videos	Bit rate (kbps)	HEVC	Coded frame based schemes			Original frame based schemes		
			Proposed NP (coded)	MoG (coded)	Proposed SANP	Proposed NP (original)	MoG (original)	Proposed WNP
Sales	150	40.00	41.90	40.70	**42.70**	42.00	42.10	**42.70**
	300	44.90	45.60	45.70	**46.80**	45.90	46.00	**46.80**
News	100	40.20	39.40	38.80	**39.50**	39.30	39.40	**39.50**
	250	45.90	45.80	45.00	**45.90**	45.60	45.60	45.60
Silent	1000	42.00	43.80	43.00	44.50	43.70	43.80	**44.70**
	1600	44.60	45.70	45.70	46.80	45.80	45.80	**46.90**
Paris	1400	40.50	41.90	40.80	**42.40**	41.50	41.60	42.00
	2400	44.40	45.10	44.40	**45.50**	44.90	44.80	45.30
Grandma	150	41.20	42.40	41.40	42.60	42.80	42.80	**43.80**
	350	46.10	46.20	46.10	46.80	46.55	46.50	**47.70**
Table Tennis	500	43.90	44.00	44.80	**44.90**	43.90	43.95	44.00
	800	47.70	47.80	48.00	**48.40**	47.70	47.00	47.90
Container	1500	43.20	42.90	43.20	43.30	43.30	43.30	**43.80**
	2500	46.30	45.90	46.20	46.30	46.30	46.30	46.80
Blue sky	2100	41.20	41.20	41.20	41.20	41.30	41.30	**41.40**
	4500	47.75	47.85	47.70	47.85	48.15	48.10	**48.20**
Pedestrian	1000	40.50	40.50	40.50	40.55	40.70	40.60	**40.80**
	3500	49.15	49.10	49.15	49.10	49.40	49.30	**49.50**
Exit	1000	39.20	39.30	39.25	**39.80**	39.25	39.30	**39.80**
	4000	42.80	42.81	42.78	43.00	42.82	42.80	**43.20**
Traffic	6000	35.10	35.10	35.00	35.20	35.10	35.10	**35.25**
	18000	42.35	42.30	42.30	42.40	42.35	42.35	**42.50**

about 45–65% of time compared to HEVC. It is also more efficient compared to MoG and NP. Our proposed SANP achieves roughly 20% computational time saving compared to HEVC.

The time saving of WNP is largely due to the more stable background it generates for the reference frame. Generally the encoder-decoder requires less time if it chooses a pixel value from the background reference frame rather than the last frame considered for foreground because no/little motion estimation is required for the background pixel encoding-decoding [3,7].

Inorder to understand the efficiency and performance of the WNP technique, we discuss the results for the *Silent* video in more detail. The first 25 frames were used as training frames for generating the background frame. The 75th, which represents a distant enough frame from the last background-generating frame (25th), was used in our investigation to introduce enough variations between frames. Quantization values were adjusted (around 20) to bring the bit rates for NP, MoG an WNP comparably closer. As shown in Table 4, with similar bit rate the WNP quality (in terms of PSNR) was much higher. The mean error is the mean absolute difference between the 75th original input frame and the decoded frame. The lower error for WNP, shown in Table 4, indicates that the decoded frame produced by the proposed WNP technique is more similar to the original frame than those produced by the MoG and NP techniques.

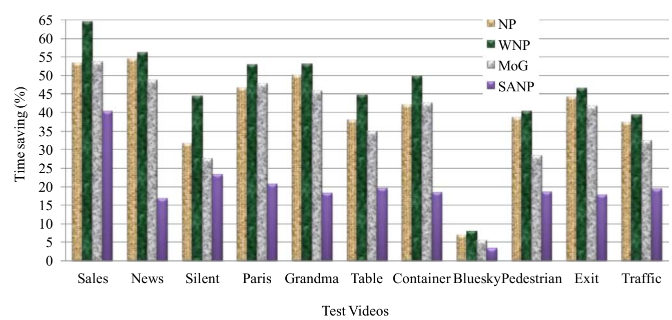


Fig. 6. Mean processing time saving (%) for the different video coding techniques compared to HEVC.

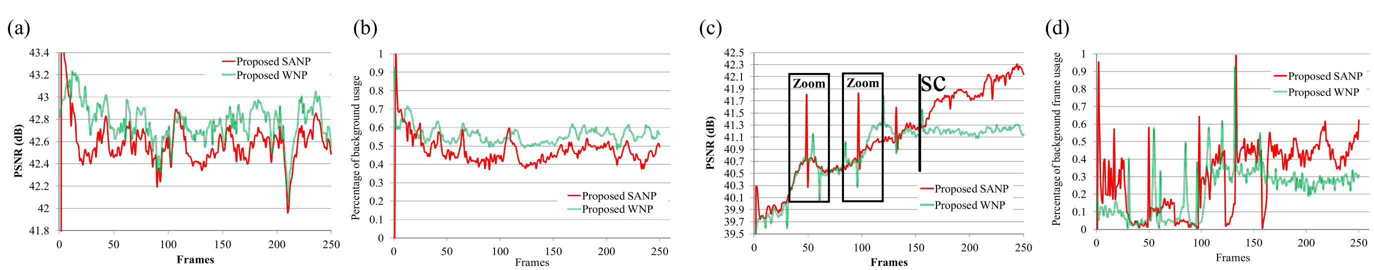


Fig. 5. Frame by frame performance and background usage for SANP and WNP. (a) *Silent* video frame by frame PSNR comparisons using two different schemes. (b) Percentages of background frame usage as a reference frame for *Silent* video using two schemes. (c) *Table* video frame by frame PSNR comparisons using two different schemes. (d) Percentages of background frame usage as a reference frame for *Table* video using two schemes.

Table 4
75th Frame rate distortion performance for *Silent* video.

Technique	Bit rate	PSNR	Mean error
NP	1259.4	44.71	1.1304
Proposed WNP	**1113.7**	**45.12**	**1.0724**
MoG	1242.3	44.72	1.1311

Fig. 7. Input frame (75th) of the *Silent* video.

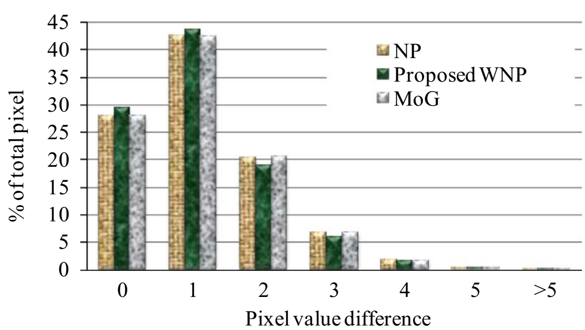


Fig. 8. Input-Output pixel intensity difference (%) for the 75th frame.

The 75th input frame is shown in Fig. 7. The relevant histogram in Fig. 8 shows the absolute pixel value differences between the original and decoded frame for each technique. From the histogram we can observe that the percentage of pixels with “0” and “1” pixel value difference is higher for WNP than NP and MoG while lower for higher pixel value differences. This indicates that the decoded output frame produced by WNP is much similar to the actual input frame. It is clearly evident that the WNP performes more efficiently than MoG and NP. This confirms the lower mean error for WNP shown in Table 4.

Fig. 9, Figs. 10 and 11 show the background frames generated by the NP, MoG and WNP techniques along with the relevant background reference maps showing the sections of the image considered as background. The black regions in the reference maps are foreground. If we inspect the background frames closely with respect to the original 75th frame shown in Fig. 7, we can observe that NP and WNP have been able to better detect the background behind the moving hand compared to MoG. These background frames generated by NP, MoG and WNP were used as reference frames for coding-decoding by the respective coding-decoding technique.

The reference maps in Fig. 9, Fig. 10 and Fig. 11 show the areas considered as foreground (in black) and background during coding-decoding process. By comparing the reference maps we can observe that during coding-decoding of the 75th frame, more background area has been used from the background frame generated by WNP. The additional areas used from the foreground by NP in Fig. 9(b) and MoG in Fig. 11(b) compared to WNP in Fig. 10(b) are highlighted using red circles. This higher usage of the background reference frame has contributed to the better performance and faster processing capacity of the proposed WNP based technique due to no motion search requirements in the background region. Fig. 12 shows the general background frame usage trends for each technique when coding-decoding the frames of the *Silent* video under various quantization values.

From the experimental results it is evident that the WNP and the SANP techniques inherit the *non-parametric* advantages of the NP and provide much superior performances under various testing scenarios. The WNP is more appropriate for coding videos with more stable background by taking the advantages of a background frame generated using original frames. The SANP uses WNP background generation process and works better for coding video sequences with highly dynamic backgrounds by a recursive updating scheme using coded frames.

5. Conclusion

We have developed a new coding technique using the tradiational *non-parametric* (NP) background modelling technique. A novel *weighted non-parametric* (WNP) background model is then developed to suit the video coding applications. Two separate coding schemes based on the original frames and the coded frames are developed to handle video sequences with stable and highly dynamic backgrounds. The proposed WNP based technique adopts the strengths of the well-known NP background modelling technique including automated parameter estimation and better dynamic background detection. The WNP technique generates more stable background by incorporating historic pixel values in the background frame. The balance between

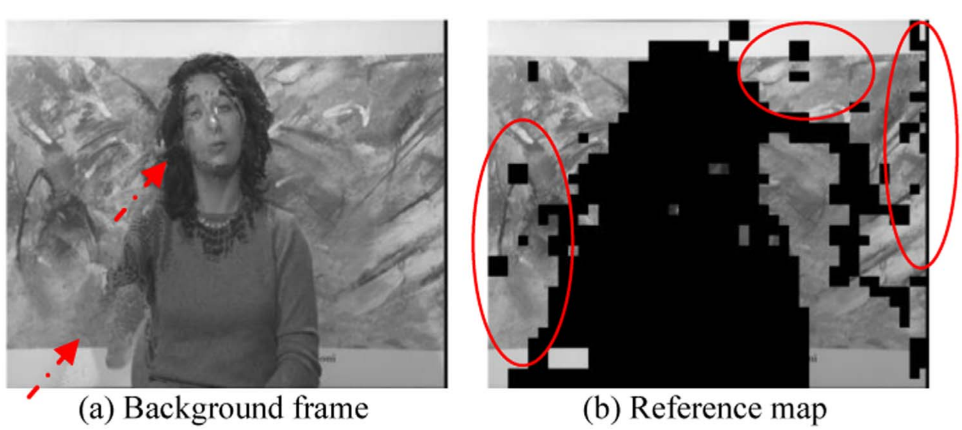


Fig. 9. Background and reference map of 75th frame for NP (a) Background frame (b) Reference map.

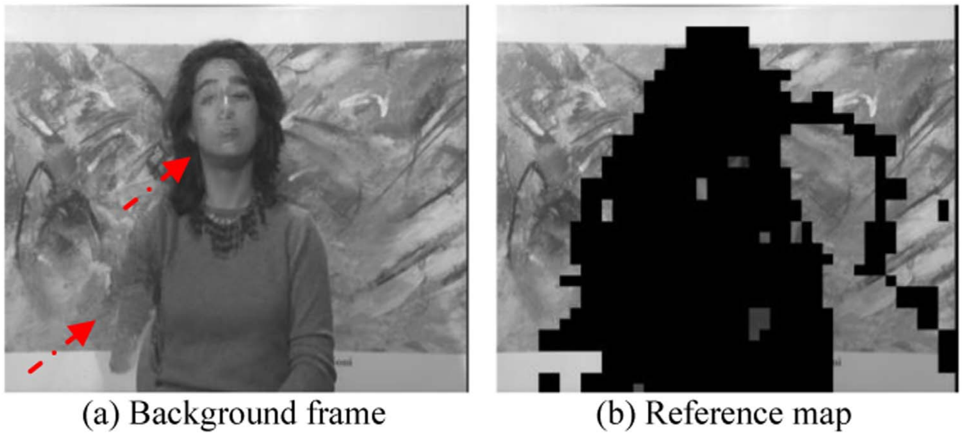


Fig. 10. Background and reference map of 75th frame for WNP (a) Background frame (b) Reference map.

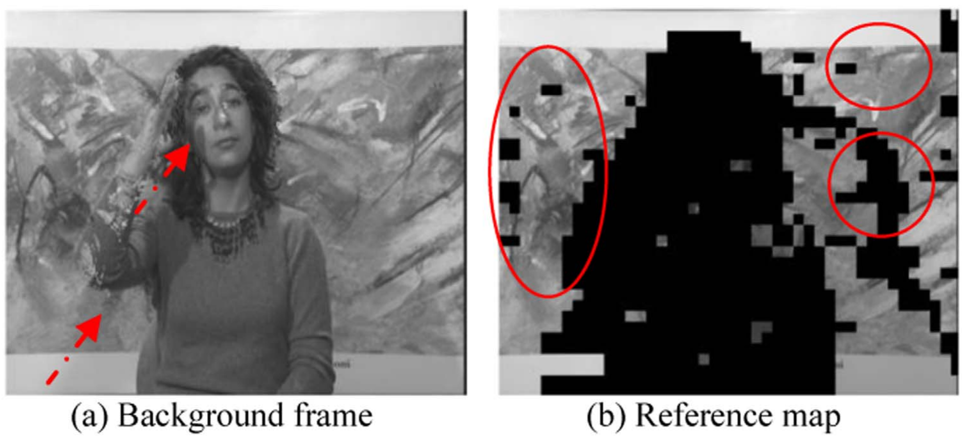


Fig. 11. Background and reference map of 75th frame for MoG (a) Background frame (b) Reference map.

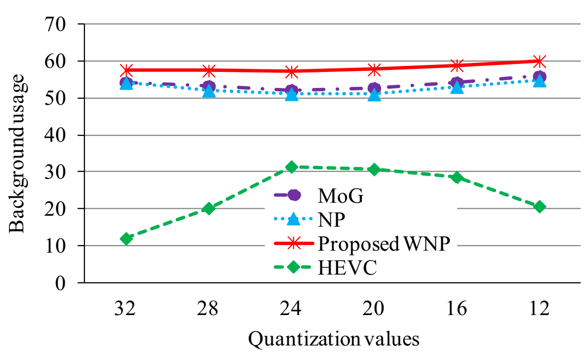


Fig. 12. Background usage (%) as the 2nd reference (LTR) over various quantization for the *Silent* video using Scheme 1.

historic and recent pixel values is maintained adaptively through dynamic ratio selection. The stability of the reference background frame in turn provides more efficient performance for the WNP based coding-decoding technique. The balance between the historic and current pixel value can be obtained using the novel adaptive weight selection procedure presented in this paper. Extensive experimental results presented in this study establish the performance validity of the proposed technique. The SANP scheme showed better performances where the background is dynamic or there is scene changes or there is high exposer of occuleded areas at different stages in the video sequences. The experimental results showed that WNP and SANP provide superior performances for quality (PSNR) compared to similar schemes of NP and MoG, and against the HEVC by up to 1.0 dB. WNP is much faster than these techniques (45–65% faster than HEVC). The study will provide researchers and practitioners with new insights in to improving the performance of the coding-decoding process.

Acknowledgement

This work was supported in part by the Australian Research Council under Discovery Projects Grant DP130103670.

References

[1] G.J. Sullivan, J.-R. Ohm, W.-J. Han, T. Wiegand, Overview of the high efficiency video coding (HEVC) standard,, IEEE Trans. Circuits Syst. Video Technol (TCSVT) 22 (12) (2012) 1649–1668.
[2] T. Wiegand, G.J. Sullivan, G. Bjøntegaard, A. Luthra, Overview of the H.264/AVC video coding standard, IEEE Trans. Circuits Syst. Video Technol. 13 (7) (2003) 560–576 (Jul.).
[3] ITU-T Rec. H.264/ISO/IEC 14496-10 AVC. Joint Video Team (JVT) of ISO MPEG and ITU-T VCEG, JVT-G050, 2003.
[4] J. –R. Ding, J. –F. Yang, Adaptive group-of-pictures and scene change detection methods based on existing H.264 advanced video coding information,, IET Image Process. 2 (2) (2008) 85–94 (Apr.).
[5] M. Paul, W. Lin, C.T. Lau, B. –S. Lee, McFIS: better I-frame for video coding,, IEEE Int. Symp. Circuits Syst. (IEEE ISCAS-10) (2010) 2171–2174 (Jun.).
[6] M. Paul, W. Lin, C.T. Lau, B. –S. Lee, Explore and model better I-frames for video coding,, IEEE Trans. Circuits Syst. Video Technol. 21 (9) (2011) 1242–1254 (Sept.).
[7] Y. –W. Huang, B. –Y. Hsieh, S. –Y. Chien, S. –Y. Ma, L. –G. Chen, Analysis and complexity reduction of multiple reference frames motion estimation in H.264/AVC, IEEE Trans. Circuits Syst. Video Technol. 16 (4) (2006) 507–522 (Apr.).
[8] L. Shen, Z. Liu, Z. Zhang, G. Wang, An adaptive and fast multi frame selection algorithm for H.264 video coding, IEEE Signal Process. Lett. 14 (11) (2007) 836–839 (Nov.).
[9] T. –Y. Kuo, H. –J. Lu, Efficient reference frame selector for H.264, IEEE Trans. Circuits Syst. Video Technol. 18 (3) (2008) 400–405 (Mar.).
[10] K. Hachicha, D. Faura, O. Romain, P. Garda, Accelerating the multiple reference frames compensation in the H.264 video coder, J. Real.-Time Image Process. 4 (1)

(2009) 55–65 (Mar.).
[11] D. Liu, D. Zhao, X. Ji, W. Gao, Dual frame motion compensation with optimal long-term reference frame selection and bit allocation, IEEE Trans. Circuits Syst. Video Technol. 20 (3) (2010) 325–339 (Mar.).
[12] A. Mavlankar, B. Girod, Background extraction and long-term memory motion-compensated prediction for spatial-random-access-enable video coding,, Int. Pict. Coding Symp. (2009) 1–4 (May.).
[13] D. Hepper, Efficiency analysis and application of uncovered background prediction in a low bit rate image coder,, IEEE Trans. Commun. 38 (1990) 1578–1584 (Sept.).
[14] S. –Y. Chien, S. –Y. Ma, L. –G. Chen, Efficient moving object segmentation algorithm using background registration technique,, IEEE Trans. Circuits Syst. Video Technol. 12 (7) (2002) 577–586 (Jul.).
[15] T.Totozafiny, O.Patrouix, F.Luthon, J.–M.Coutellier, Dynamic background segmentation for remote reference image updating within motion detection JPEG2000, in: Proceedings of the IEEE International Symposium on Industrial Electronics, vol. 1, pp. 505–510, Jul, 2006.
[16] T. Sikora, Trends and perspectives in image and video coding, in: Proceedings of the IEEE, vol. 93, no. 1, pp. 6–17, Jan. 2005.
[17] M.Kunter, P.Krey, A.Krutz, T.Sikora, Extending H.264/AVC with a background sprite prediction mode, in: IEEE International Conference on Image Processing (ICIP), pp. 2128–2131, Oct, 2008.
[18] R.Ding, Q.Dai, W.Xu, D.Zhu, H.Yin, Background-frame based motion compensation for video compression, IEEE International Conference on Multimedia and Expo (ICME), vol. 2, pp. 1487–1490, Jun, 2004.
[19] X. Lu, Y. Yuan, P. Yan, Sparse coding for image denoising using spike and slab prior, Neurocomputing 106 (2013) 12–20.
[20] X. Lu, Y. Wang, Y. Yuan, Sparse coding from a Bayesian perspective, IEEE Trans. Neural Netw. Learn. Syst. Vol. 24 (6) (2013) 929–939.
[21] Z. He, T. Ogawa, S. Takahashi, M. Haseyama, X. Zhao, Improvement of video coding efficiency based on sparse contractive mapping approach, Neurocomputing 173 (2016) 1898–1907.
[22] M.Paul, W.Lin, C.T.Lau, B.–S.Lee, McFIS in hierarchical bipredictve pictures-based video coding for referencing the stable area in a scene, in: Proceedings IEEE International Conference on Image Processing (ICIP), pp. 3521–3524, Sept, 2011.
[23] M.Paul, W.Lin, C.T.Lau, B.–S.Lee, Video coding using the most common frame in scene, in: *IEEE International Conference on Acoustics, Speech, and Signal processi*ng (ICASSP), pp. 734–737, Mar, 2010.
[24] M. Paul, W. Lin, C.T. Lau, B. –S. Lee, Pattern-based video coding with dynamic background, EURASIP J. Adv. Signal Process. 2013 (1) (2013) 1–17 (Aug.).
[25] M.Geng, X.Zhang, Y.Tian, L.Liang, T.Huang, A fast and performance-maintained transcoding method based on background modelling for surveillance video, in: Proceedings of the IEEE International Conference on Multimedia and Expo (ICME), pp. 61–66, Jul, 2012.
[26] C.Stauffer, W.E.L.Grimson, Adaptive background mixture models for real-time tracking, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, vol. 2, pp. 246–252, Jun, 1999.
[27] D.-S. Lee, Effective Gaussian mixture learning for video background subtraction, IEEE Trans. Pattern Anal. Mach. Intell. 27 (5) (2005) 827–832 (May.).
[28] A. Elgammal, D. Harwood, L. Davis, Non-parametric model for background subtraction, Comput. Vis. (. 2000) 751–767.
[29] M.Haque, M.Murshed, M.Paul, Improved Gaussian mixtures for robust object detection by adaptive multi-background generation, in: Proceedings of the IEEE International Conference on Pattern Recognition, 2008, pp. 1–4, Dec.
[30] A. Elgammal, R. Duraiswami, D. Harwood, and L. S. Davis, Background and foreground modelling using nonparametric kernel density estimation for visual surveillance, in: Proceedings of the IEEE, vol. 90, no. 7, Jul. 2002, pp. 1151–1163.
[31] A. Elgammal, Background Subtraction: Theory and Practice, Springer Berlin Heidelberg, Berlin, Germany, 2013.
[32] Z. Zivkovic, F. van der Heijden, Efficient adaptive density estimation per image pixel for the task of background subtraction, Pattern Recogn. Lett. 27 (7) (2006) 773–780 (May.).
[33] M. Paul, W. Lin, C.T. Lau, B. –S. Lee, Video coding with dynamic background, EURASIP J. Adv. Signal Process. 2013 (2013). http://dx.doi.org/10.1186/1687-6180-2013-11.
[34] M. Paul, W. Lin, C.T. Lau, B. –S. Lee, Direct inter-mode selection for H.264 video coding using phase correlation, IEEE Trans. Image Process. 20 (2) (2011) 461–473 (Feb.).

Subrata Chakraborty has received Ph.D. degree from Monash University, Australia in 2010. He is currently a Lecturer with the School of Management & Enterprise at University of Southern Queensland. He has previously worked as a Lecturer/Researcher at Monash University, Queensland University of Technology and Charles Sturt University. Dr Chakraborty also worked in the software development sector responsible for developing enterprise scale business systems. His key research interests are in the areas of Video Coding, Computer Vision, EEG Signals analysis, Decision Support systems and Optimization. He has published in international conferences and journals. Dr Chakraborty is a member of IEEE, ACS (CP) and ACM.

Manoranjan Paul received Ph.D. degree from Monash University, Australia in 2005. He was a Lecturer/Assistant Professor (1997–2001) in Ahsanullah University of Science and Technology. He was a Research Fellow (2005–2011) in the University of New South Wales, Monash University, and Nanyang Technological University. Currently he is a Senior Lecturer and Associate Director of the Centre for Research in Complex Systems (CRiCS) in CSU.His major research interests are in the fields of Video Coding, Computer Vision, Hyperspectral Imaging, and EEG Signals analysis. He has published more than 70 refereed publications. He is an invited keynote/tutorial speaker on video coding in IEEE WoWMoM Video Everywhere Workshop 2014, DICTA 2013, and IEEE ICCIT 2010. Dr Paul is a Senior Member of the IEEE and ACS. Dr Paul has served as a guest editor of Journal of Multimedia and Journal of Computers for 6 special issues. Currently Dr Paul is an Associate Editor of EURASIP Journal on Advances in Signal Processing. Dr Paul obtained Research Excellence Award 2013 and 2011 at Faculty and School levels in CSU. Dr Paul obtained an Australian Research Council (ARC) Discovery Project Grant. So far Dr Paul obtained more than $1 M competitive grant money.

Manzur Murshed received the B.Sc. Engg. (Hons) degree in computer science and engineering from Bangladesh University of Engineering and Technology (BUET), Dhaka, Bangladesh, in 1994 and the Ph.D. degree in computer science from the Australian National University (ANU), Canberra, Australia, in 1999. He also completed his Postgraduate Certificate in Graduate Teaching from ANU in 1997. He is currently an Emeritus Prof Robert HT Smith Professor and Personal Chair at the Faculty of Science and Technology, Federation University Australia. Prior to this appointment, he served the School of Information Technology, Federation University Australia as the Head of School, from January 2014 to July 2014, and the Gippsland School of Information Technology at Monash University as the Head of School 2007–2013. He was one of the founding directors of the Centre for Multimedia Computing, Communications, and Applications Research (MCCAR). His major research interests are in the fields of video technology, information theory, wireless communications, distributed computing, and security & privacy. He has so far published 190 refereed research papers and received more than $1 M nationally competitive research funding, including three Australian Research Council Discovery Projects grants in 2006, 2010, and 2013 on video coding and communications, and a large industry grant in 2011 on secured video conferencing. He has successfully supervised 19 and currently supervising 6 Ph.D. students. He is an Editor of International Journal of Digital Multimedia Broadcasting and has had served as an Associate Editor of IEEE Transactions on Circuits and Systems for Video Technology in 2012 and as a Guest Editor of special issues of Journal of Multimedia in 2009–2012. He received the Vice-Chancellor's Knowledge Transfer Award (commendation) from the University of Melbourne in 2007, the inaugural Early Career Research Excellence award from the Faculty of Information Technology, Monash University in 2006, and a University Gold Medal from BUET in 1994. He is a Senior Member of IEEE.

Mortuza Ali received the B.Sc. Engg. degree in computer science and engineering from Bangladesh University of Engineering and Technology, Bangladesh, in 2001 and obtained his Ph.D. degree in Information Technology from Monash Univeristy, Australia, in 2009. Before joining the School of Engineering and Information Technology, Federation University Australia, he worked as a Research Fellow in the Department of Electrical and Electronic Engineering, University of Melbourne from 2009 to 2011. His research interest includes Information and Coding Theory, Multimedia Signal Processing, and Electronic Voting. His works have been published in reputed international journals and conferences.