MOTION AND REGION DETECTION FOR EFFECTIVE RECURSIVE TEMPORAL NOISE REDUCTION

Mohammed Ghazal, Chang Su, and Aishy Amer

Electrical and Computer Engineering, Concordia University, Montréal, Québec, Canada
Email:{moha_mo,chan_su,amer}@ece.concordia.ca

ABSTRACT

This paper proposes a method to integrate motion and region information into a recursive temporal noise reduction filter for video signals. We also propose an artifact-robust motion detection algorithm suitable for noise reduction. It is based on local low-pass and maximum filters and on noise-adaptive global gray-level stabilization. Region information is obtained from difference frames resulting from the proposed motion detection. The detected motion and regions are then integrated to compute the temporal filter coefficients that reduce both noise and motion blur. Simulation results show that the proposed method increases the performance of recursive temporal filtering and achieves an average gain of 3.6 dB.

Index Terms— Motion analysis, Image region analysis, Object detection, Filtering, Video signal processing

1. INTRODUCTION

Noise in video signals hinders the performance of video processing such as segmentation and compression. Hence, video noise reduction which exploits motion and region information has become increasingly used in modern video systems.

Artifact-robust motion and region information is pivotal for motion-adaptive recursive temporal noise reduction which provides a less computationally demanding alternative to motion estimation and compensation based denoising methods [1] making it more attractive for TV and video systems that do not use motion estimation. However, the problem is how to acquire accurate motion and region information and how to integrate them to yield a stable recursive temporal noise filter in the presence of high motion and noise.

Video signals are temporally correlated in stationary regions in which temporal low-pass filtering can significantly reduce noise. Nevertheless, the correlation is lost in 1) moving regions where temporal filtering is liable to cause motion blur or ghosting or 2) in regions with artifacts such as shadow and illumination changes where low-pass filtering is hindered. Integrating artifact-robust motion and region information can help separate correlated regions and decide how much filtering is suitable in them. Such integration adapts noise reduction to the structure and motion of the object regions.

Algorithms for motion adaptive noise reduction such as [2] pay little attention to obtaining artifact-robust motion information or utilizing stable region information. For example, the method in [2] generates region information through simple thresholding of motion information obtained from wavelet-based noise-reduced frames which can be sensitive to illumination changes and shadows.In contrast, methods that pay attention to the quality of the motion and region information [3] do not adapt to the needs of recursive temporal filtering.

The objective of motion detection (MD) is to indicate the moving regions in video frames. Haan *et. al* [3] detect moving regions by taking the magnitude of frame differencing between the current frame and a reference frame. The MD is efficient, however, it is sensitive to noise and illumination variations. Lee *et. al* [4] present a MD method that considers motion information to be the maximum gray-level difference in three consecutive difference frames and removes false motion due to noise by applying a median filter which does not perform well for Gaussian noise. Note that in [4] and [5] multiple field delays are required. Aach *et. al* [5] model the noise in difference frames as additive white Gaussian noise (AWGN) and perform MD based on a significance test. For robust and accurate MD, a threshold transformation algorithm based on significance invariance is applied. Aach *et. al*'s MD method does not account for artifacts caused by global (e.g., illumination changes) and local (e.g., shadow). Thus, it may suffer from video sequences with complex content.

The contributions of the proposed method are; 1) detection of artifact-robust motion and region information, and 2) integration of motion and region information for stable recursive temporal noise reduction. The remainder of the paper is as follows. Sections 2 presents the proposed approach theoretically. Objective simulation results are discussed in Section 3 and Section 4 concludes the paper.

2. PROPOSED ALGORITHMS

In this paper, an AWGN model is assumed [3]. Let F_k^η be a noisy video frame at time instance k defined by $F_k^\eta =$

This work was supported, in part, by the Fonds de la recherche sur la nature et les technologies du Québec (NATEQ)

$F_k + \eta_k$, where F_k is the noise-free frame and η_k is the added noise component. A pixel in F_k is denoted by $F_k(i,j)$ where (i,j) are spatial coordinates. $\eta_k(i,j)$ is the amount of noise added to $F_k(i,j)$. Fig. 1 shows an overview of the proposed algorithm. First, F_k^η and F_{k-1}^η are fed into the proposed motion detector which outputs the artifact-robust motion information or difference frame D_k. Next, D_k is used to compute region information B_k and the recursive temporal filter coefficients C_k. Then, C_k is updated with binary region information B_k to get the improved coefficients $\hat{C}_k$ based on integration of motion and region information. Finally, $\hat{C}_k$ is used in

$$\hat{F}_k = \hat{C}_k F_k^\eta + (1 - \hat{C}_k)\hat{F}_{k-1}, \tag{1}$$

to get the noise-reduced current frame $\hat{F}_k$ from the noisy current frame F_k^η and the noise-reduced previous frame $\hat{F}_{k-1}$.

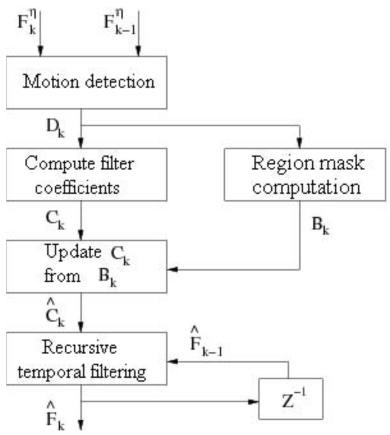


Fig. 1. Proposed integration of motion and region information for noise reduction.

2.1. Proposed Motion Detection

The quality of difference frames D_k may be degraded by three main factors; noise, motion instability, and local changes such as shadow and object velocity. Noise introduces globally scattered artifacts in D_k whereas both motion instability and local changes lead to holes and gaps in the detected moving regions. The basic idea of the proposed MD method (Fig. 2) is to reduce noise and stabilize the gray-levels (i.e., motion difference) in D_k. First, $|F_k^\eta$-$F_{k-1}^\eta|$ is low-pass filtered with a

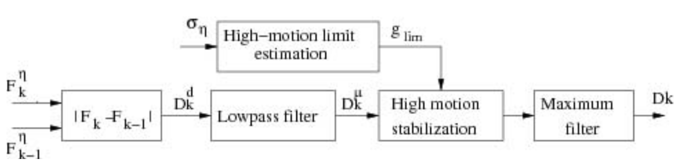


Fig. 2. Block diagram of the proposed MD algorithm.

uniform convolution mask to reduce the effect of noise. Then, we limit the differences in the noise-reduced difference frame D_k^μ to an automatically determined noise-adaptive gray-level limit g_{lim}. This step is necessary to globally stabilize (or filter) the high motion in D_k. High motion may be caused by 1) local changes, 2) object textures and velocities.

It is important to adapt the gray-level limit g_{lim} to noise in order to tune the performance of motion detection and eventually noise reduction. Setting g_{lim} low increases the sensitivity to artifact caused by non-motion related factors (such as shadow) and produces larger regions in B_k, which results in reduced filtering. In contrast, low g_{lim} is good for frames with low noise-levels to reduce motion blur in moving regions.

We adapt g_{lim} to the noise-level using

$$g_{lim} = g_{lim}^c + c \cdot \sigma_\eta, \tag{2}$$

where g_{lim}^c and c are experimentally set to 48 and 2.5, respectively, and σ_η is the estimated AWGN standard deviation. The noise-level is measured in the Peak Signal to Noise Ratio (PSNR) defined by $\text{PSNR}_\eta = 10\log_{10}\frac{g_{max}^2}{\sigma_\eta^2}$ where g_{max} is the maximum possible gray-level (e.g., 255).

It is also important to establish a lower-bound on g_{lim}. This is because a too low g_{lim} will cause many pixels in D_k to be wrongly classified as white (moving) pixels in B_k which will turn off the temporal noise filter. To establish such a lower-bound, we condition $\text{PSNR}_{D^\mu}^{g_{lim}}$ (the PSNR of D_k^μ in Fig. 2 obtained with a given g_{lim}) to be larger than PSNR_η or

$$\text{PSNR}_{D^\mu}^{g_{lim}} = 10\log_{10}\frac{g_{lim}^2}{\text{MSE}_{D_k^\mu}} \geq \text{PSNR}_\eta, \tag{3}$$

where $\text{MSE}_{D_k^\mu}$ is the mean-square error (MSE) between noisy D_k^d and its noise-reduced approximation D_k^μ. When a 3×3 average filter is applied to get D_k^μ, it can be shown that $\text{MSE}_{D_k^\mu} = \frac{1}{72}\sigma_\eta$. Let $g_{lim} = \alpha g_{max}$, (3) is rewritten as

$$\text{PSNR}_{D^\mu}^{g_{lim}} = 10\log_{10}\frac{g_{max}^2}{\sigma_\eta^2} + 10\log_{10} 72\alpha^2 \geq \text{PSNR}_\eta. \tag{4}$$

From (4), we get $\alpha \geq 0.1179$ and $g_{lim}^{min} \geq 30$. Note that as expected $g_{lim}^c > g_{lim}^{min}$.

After stabilizing the high motion in difference frames, a maximum filter is applied to reduce small holes, gaps, and granular blobs inside the moving regions and cause stability around the boundaries between the moving and the static regions. We use the maximum filter after global gray-level stabilization because the maximum filter tends to enlarge the instability of the gray-levels in D_k.

2.2. Proposed Motion and Region Integration

Regions B_k are obtained by thresholding D_k (e.g., using [6]) and initial filter coefficients C_k are computed from motion information D_k using the negative quadratic relation

$$C_k(i,j) = \beta^2\left(\frac{D_k(i,j)}{g_{lim}}\right)^2 - 2\beta\frac{D_k(i,j)}{g_{lim}} + C_{min}, \tag{5}$$

where $\beta = C_{min} - 1$ and $C_{min} = 0.1$ is the minimum filter-coefficient. Since, all gray-levels in D_k after global gray-level stabilization will be limited to g_{lim}, we use it to normalize D_k in (5). We have also tested a linear relation between C_k and D_k and concluded the negative quadratic performs better because it increases C_k with D_k at a faster rate to rely more on the current pixel than the previous pixel (see (1)). Then C_k are updated based on region information B_k to get an updated set of coefficients $\hat{C}_k$ using

$$\hat{C}_k(i,j) = \begin{cases} C_k(i,j) & : B_k(i,j) = 0 \\ B_k(i,j) & : \text{otherwise} \end{cases} \tag{6}$$

The purpose of (6) which integrates D_k and B_k is to determine regions where motion blur is more likely to result from filtering than noise reduction gain. In other regions, the relationship between C_k and D_k will govern how much filtering is performed. For example, if the artifact-robust region detection classifies the pixel as moving ($B_k(i,j) = 1$), noise filtering is turned off by setting $\hat{C}_k(i,j)$ to be 1. Otherwise, if the pixel is classified as stationary, then $\hat{C}_k(i,j) = C_k(i,j)$ and the amount of filtering is governed by (1) and (5). The effect of the proposed motion and region integration using (6) is more visible in sequences with global motion in which D_k identifies both the moving areas and regions with high structure. For example, in the *flowergarden* sequence, despite global motion, the sky is detected as a spatially homogeneous region and used in filtering while the slightest movement in the grass area is considered unsuitable for filtering because of its spatial structure. In other words, (5) and (6) reflect the reliability of motion and the suitability of region.

3. RESULTS

To evaluate the performance of the proposed method, we selected the test video sequences: *Tennis* (zoom in and out camera motion), *Flower garden* (translation camera motion), *Train* (high structure), *Survey* (no global motion with heavy object occlusion) and *Patrol car* (translation and rotation camera motion). The sequences selected represent different types of global and object motion. Each video sequence is corrupted with 25 dB and 30 dB noise-levels. The PSNR gain is calculated as $R = \text{PSNR}_{\hat{F}_k} - \text{PSNR}_\eta$ where $\text{PSNR}_{\hat{F}_k}$ is the PSNR of the noise reduced frame $\hat{F}_k$.

Since our objective is to evaluate the suitability of the proposed motion detection and proposed region integration to temporal filtering, we fix the temporal filter and compare our methods to a recent motion detection method [5] by using its output motion information D_k and region information B_k. The noise reduction gain over time for the 25 dB and 30 dB noisy *Tennis* test sequence is plotted in Fig. 3 (a) and (b), respectively. Note that the performance drop at frames 20 and 50 is due to shot changes. The proposed method outperforms [5] by almost 3.5 dB, indicating that the proposed motion detection is more suitable for motion-adaptive noise reduction than [5]. The same can be said about the results of *Flower garden*, *Train*, and *Survey* test sequences depicted in Figs. 3 (c)-(h).

Table 1 summarizes the performance of the proposed method over all video sequences used. We note the consistent gain increase per noise level.

Table 1. Average gain using the proposed and [5] methods.

Alg.	20 dB	30 dB	40 dB
Proposed	4.2	3.5	3.0
[5]	1.3	0.8	0.5

4. CONCLUSION

This paper proposed an algorithm for integrating motion and region detection tailored for the needs of recursive temporal noise filters. The proposed motion detection deploys, local low-pass and maximum filters and noise-adaptive global gray-level stabilization for motion detection. Region information is obtained from difference frames resulting from the proposed motion detection. The integration of the detected motion and region information is used to compute recursive temporal filter coefficients that reduce both noise and motion blur. The proposed method is found more suitable for recursive temporal noise reduction than the referenced motion detection method with an average gain of 3.6 dB.

5. REFERENCES

[1] B. C. Song and K. W. Chun, "Motion-compensated temporal filtering for denoising in video encoder," *Electronics Letters*, vol. 40, no. 13, pp. 802–804, Jun. 2004.

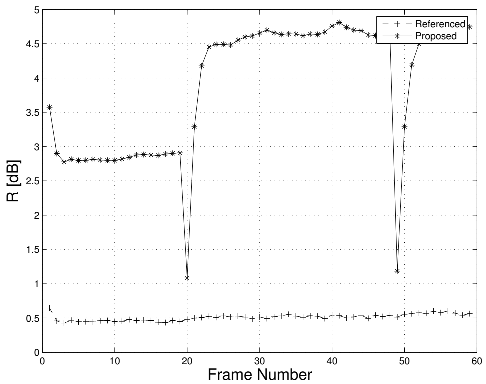
[2] A. Pizurica, V. Zlokolica, and W. Philips, "Combined wavelet domain and temporal video denoising," in *IEEE Conference on Advanced Video and Signal Based Surveillance*, Jul. 2003, pp. 334–341.

[3] G. de Haan, T. G. Kwaaitaal-Spassova, and O. A. Ojo, "Automatic 2-D and 3-D noise filtering for high-quality television receivers," in *Int. Workshop On HDTV*, 1994.

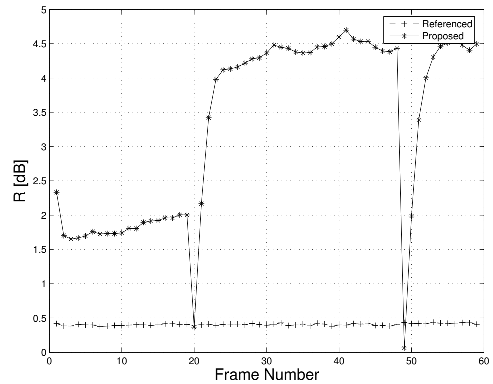
[4] S.-G. Lee and D.-H. Lee, "A motion-adaptive deinterlacing method using an efficient spatial and temporal interpolation," *IEEE Transactions on Consumer Electronics*, vol. 49, no. 13, pp. 1266–1271, Nov. 2003.

[5] T. Aach and A. P. Condurache, "Transformation of adaptive thresholds by significance invariance for change detection," in *IEEE/SP Workshop On Statistical Signal Processing*, Jul. 2005, pp. 637–642.

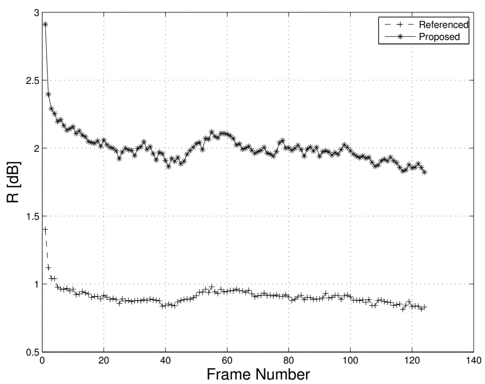
[6] C. Su and A. Amer, "A real-time adaptive thresholding for video change detection," in *IEEE Int. Conf. On Image Processing*, Oct. 2006, pp. 157–160.



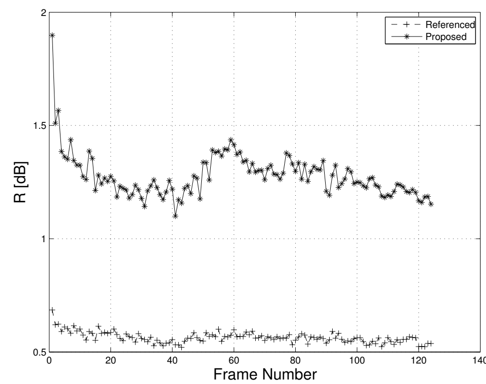
(a) 25 dB Tennis



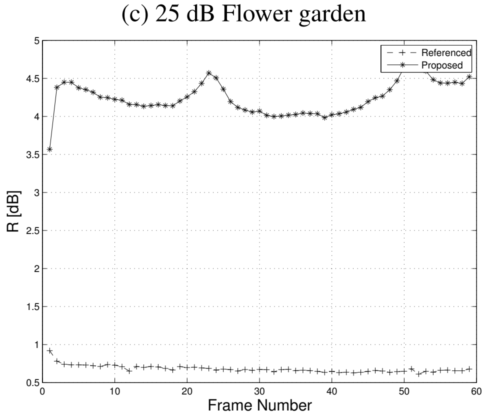
(b) 30 dB Tennis



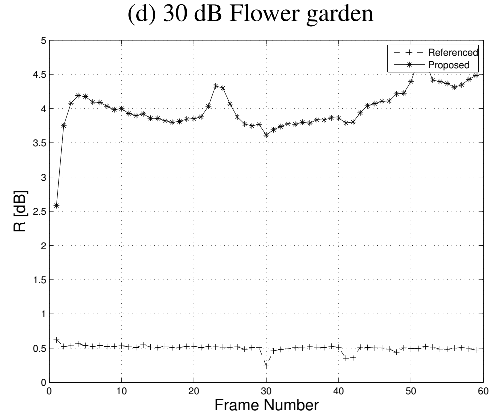
(c) 25 dB Flower garden



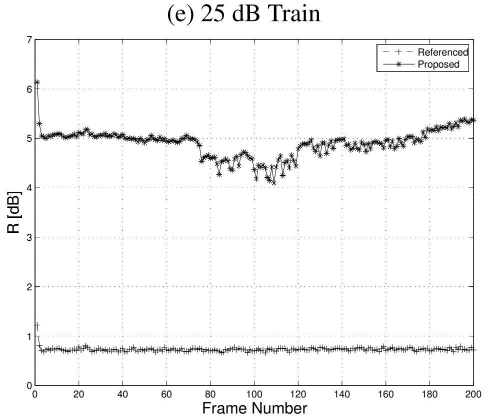
(d) 30 dB Flower garden



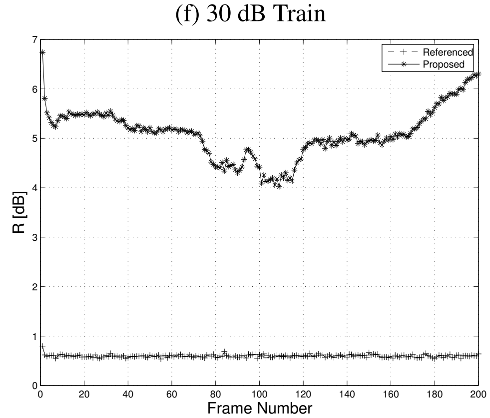
(e) 25 dB Train



(f) 30 dB Train



(g) 25 dB Survey



(h) 30 dB Survey

Fig. 3. Noise reduction gain over time for selected test sequences for proposed and referenced method [5].