Signal Processing: *Image Communication* 29 (2014) 80–95

Contents lists available at ScienceDirect

Signal Processing: *Image Communication*

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

Enhancing dynamic videos for surveillance and robotic applications: The robust bilateral and temporal filter

Andry Maykol Pinto*, Paulo G. Costa, Miguel V. Correia, A. Paulo Moreira

INESC Technology and Science, Faculty of Engineering of University of Porto, Rua Dr. Roberto Frias, s/n 4200-465 Porto, Portugal

ARTICLE INFO

Article history:
Received 4 May 2013
Received in revised form
7 October 2013
Accepted 22 November 2013
Available online 1 December 2013

Keywords:
Spatiotemporal filter
Robust statics
Surveillance
Dynamic environment
Robotic applications

ABSTRACT

Over the last few decades, surveillance applications have been an extremely useful tool to prevent dangerous situations and to identify abnormal activities. Although, the majority of surveillance videos are often subjected to different noises that corrupt structured patterns and fine edges. This makes the image processing methods even more difficult, for instance, object detection, motion segmentation, tracking, identification and recognition of humans.

This paper proposes a novel filtering technique named robust bilateral and temporal (RBLT), which resorts to a spatial and temporal evolution of sequences to conduct the filtering process while preserving relevant image information. A pixel value is estimated using a robust combination of spatial characteristics of the pixel's neighborhood and its own temporal evolution. Thus, robust statics concepts and temporal correlation between consecutive images are incorporated together which results in a reliable and configurable filter formulation that makes it possible to reconstruct highly dynamic and degraded image sequences.

The filtering is evaluated using qualitative judgments and several assessment metrics, for different Gaussian and Salt–Pepper noise conditions. Extensive experiments considering videos obtained by stationary and non-stationary cameras prove that the proposed technique achieves a good perceptual quality of filtering sequences corrupted with a strong noise component.

1. Introduction

Automated surveillance systems usually resort to stationary sensors to monitor the environment; however, the number of research works that propose surveillance applications based on non-stationary cameras is increasing. The presence of noise in videos affects subsequent image processing phases, such as three-dimensional reconstruction, registration, classification of objects, motion segmentation and analysis, tracking, identification and recognition of humans. Thus, denoising is an extremely important pre-processing phase that is used to improve the perceptual appearance of images; however, a trade-off between noise reduction and data preservation is important to enhance the characteristics of images that are relevant for high level algorithms.

Despite recent improvements in the color filter array [6,32], realistic surveillance sequences often have a low signal-to-noise ratio (SNR) [30] since they are commonly corrupted with noise that can be approximated by Gaussian and Salt-and-Pepper noise [8]. This paper presents an image denoising technique called robust bilateral and temporal filter (RBLT) that meets the visual requirements of a surveillance system based on a mobile robot. This name was chosen

* Corresponding author. Permanent address: Departamento de Electrotecnia da Faculdade de Engenharia da Universidade do Porto, Rua Dr. Roberto Frias, s/n 4200-465 Porto, Portugal.
Tel.: +351 22 508 14 00; fax: +351 22 508 14 40.
E-mail address: andry.pinto@fe.up.pt (A. Maykol Pinto).

0923-5965/$ - see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.image.2013.11.003

because the technique follows and enhances the bilateral filter (BL) formulation of Tomasi and Manduchi [31]. The original formulation [31] is completely reformulated using robust estimation principles, for instance, non-quadratic estimators are incorporated into the filter formulation. This makes the filter more robust to the presence of outliers (usually, noise components) while preserving structural information of the image sequence. These two aspects are the major requirements for a denoising technique used in robotic-based surveillance systems. In addition, a temporal component is also incorporated into the filter formulation, which increases the filter's ability to remove strong noise components. The major advantage of using spatial and temporal information in videos is the possibility of achieving filtering performances that otherwise would hardly be obtained. This means that temporal information can be used to infer the noise component that corrupts the current image if there is a reliable correlation for the brightness evolution of pixels between consecutive images. Therefore, the RBLT measures the reliability of temporal information of the brightness evolution and uses spatial information about the neighbors, which are weighted using non-quadratic norms to evaluate the similarity and to estimate the free-noise value of each pixel.

Contributions of this paper include the following:

1. An innovative spatiotemporal filtering technique that can be used by stationary and non-stationary surveillances or robotic applications.
2. A measurement of the photometric similarity based on robust error norms.
3. A temporal filtering component based on the *temporal coherence* assumption with a self-evaluation mechanism to detect and treat violations of this assumption.
4. A filtering technique with a performance less influenced by outliers and by the type of noise that corrupts the sequence.
5. Filtering with a better trade-off between noise reduction and data preservation which is especially recommended for denoising images with low SNR. Thus, the filter does not create ghosts or strange artifacts in the denoised image that compromise the segmentation process (especially, the motion segmentation).
6. A filtering technique that enhances videos corrupted by Gaussian and Salt-and-Pepper noise, in a very competitive manner when compared to other state-of-the-art techniques (especially designed for each type of noises).
7. Extensive qualitative and quantitative evaluation by considering several baseline filters.

The experimental results include the analysis of the proposed denoising technique in several contexts: a comparison to state-of-the-art methods, and an evaluation and discussion of the behavior of the RBLT in real and practical surveillance applications. Therefore, important conclusions are obtained about the usefulness of the filter for different types of videos and noise. In previous experiments, the bilateral filter and the spatiotemporal versions of the Gaussian average and the median filter are considered baseline methods. The performance of the RBLT method is evaluated using sequences corrupted by Gaussian and Salt–Pepper noise. The quality of distorted images is validated using subjective visualizations and several objective assessment metrics, namely, the root mean square error (RMSE), the signal-to-noise ratio (SNR) and especially the peak signal-to-noise ratio (PSNR) and the structural similarity (SSIM). Experimental considerations indicate that filtering corrupted sequences using the "robustification" of the temporal correlation between consecutive images is computationally rewarding and represents an alternative to the state-of-the-art techniques. The filtering properties and edge preserving capabilities of the proposed filter can lead to a potential accuracy enhancement of motion segmentation and contribute to future developments in automated surveillance applications.

The paper is organized as follows. Section 1.1 presents a brief review of the latest spatiotemporal filtering techniques. Sections 1.2 and 1.3 introduce the Gaussian and bilateral formulation, respectively. Section 2 presents the proposed RBLT filtering technique in detail: Section 2.1 shows the formulation based on robust estimation methods and Section 2.2 describes the incorporation of the temporal contribution. Section 3 presents the experimental results of the RBLT: the proposed method is compared to state-of-the-art video denoising techniques in Section 3.1. Later, a set of experiments is conducted in order to evaluate the performance of the RBLT under realistic surveillance videos: Section 3.2 shows the filtering results of videos captured by a stationary surveillance application; Section 3.3 presents the filtering results for videos obtained by a mobile robotic-based surveillance system. Finally, the major conclusions of this research are presented in Section 4.

1.1. Related work

In the literature, many researchers have proposed denoising methods based on the average filtering, median filtering [27,37], non-linear diffusion [17], non-linear total variation [13,33], non-local means filters [35], wavelet [16,1], multi-scale filtering [16,38], morphological operators [20,11] and bilateral filtering.

The bilateral filtering is commonly used in computer vision due to its denoising properties and ability to preserve data. Several bilateral-based formulations have been proposed [29,38,18,37] over the last few years. In fact, an extension of the bilateral filter based on multi-resolution with wavelet transform sub-band mixing is proposed by Zeinab and Yasser [16]. Their technique achieves better results than the BL filter. The improvement was 0.1029 dB (SNR), 1.1292 dB (PSNR) and 3.3809 (MSE) for a Gaussian noise with a standard deviation equals to five. Shyam Anand and Sahambi [1] use the undecimated wavelet transform to obtain the coefficients. Then, a bilateral filter is applied to the transformed approximation in order to preserve relevant edges and to remove the noise. The results obtained by the technique are visually acceptable since the denoising level appears to be high. However, their research fails at performing a qualitative comparison between other filters and, therefore, it is difficult to quantify the improvement of the proposed approach. Liu et al. [14] have demonstrated that the domain parameter of the bilateral filter should be adjusted to $1.95\sigma_n$ in order