J. Vis. Commun. Image R. 31 (2015) 14–25

Contents lists available at ScienceDirect

J. Vis. Commun. Image R.

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

Content adaptive video denoising based on human visual perception ☆

Cheolkon Jung *, Tian Sun, Aiguo Gu

School of Electronic Engineering, Xidian University, Xi'an 710071, China

ARTICLE INFO

Article history:
Received 3 September 2014
Accepted 28 April 2015
Available online 22 May 2015

Keywords:
Content adaptive
Adaptive noise filtering
Human visual perception
Noisy frame detection
Motion detection
Nonlocal mean filtering
Perceptual video denoising
Temporal similarity

ABSTRACT

In this paper, we propose content adaptive denoising in highly corrupted videos based on human visual perception. We introduce the human visual perception in video denoising to achieve good performance. In general, smooth regions corrupted by noise are much more annoying to human observers than complex regions. Moreover, human eyes are more interested in complex regions with image details and more sensitive to luminance than chrominance. Based on the human visual perception, we perform perceptual video denoising to effectively preserve image details and remove annoying noise. To successfully remove noise and recover the image details, we extend nonlocal mean filtering to the spatiotemporal domain. With the guidance of content adaptive segmentation and motion detection, we conduct content adaptive filtering in the YUV color space to consider context in images and obtain perceptually pleasant results. Extensive experiments on various video sequences demonstrate that the proposed method reconstructs natural-looking results even in highly corrupted images and achieves good performance in terms of both visual quality and quantitative measures.

1. Introduction

Video signals are inevitably corrupted by noise which invariably degrades the quality of video frames. The restoration of degraded video signals is required to improve the video quality as well as the efficiency of video compression [1].

1.1. Related work

Up to now, many significant achievements have been made by researchers in the field of denoising in images and videos [1]. In [2], Sendur and Selesnick introduced non-parametric statistics of images in image denoising, and employed kernel regression in image denoising. To overcome the inherent limitation of the linear filtering properties in classic kernel regression methods, they developed the nonlinear data-adapted class of kernel regressors. In [3], Fields of Experts (FoE) removed noise in images by learning priors. The key idea was to formulate the priors as a high-order Markov random field (MRF) which was defined over large neighborhood systems. This was implemented by exploiting sparse image patch representations. The resulting FoE modeled the prior probability of an image or other low-level representation in terms of a random field with overlapping cliques, whose potentials were represented as Product of Experts [4]. The transform-domain denoising methods typically assumed that the true signal was well approximated by a linear combination of few basis elements. The basic idea was that the true signal was effectively estimated by preserving the few high-magnitude transform coefficients which conveyed most of the true-signal energy and discarding the rest ones mainly caused by noise. The sparsity of the representation depended on both the transform and the true-signal's properties. Many advanced denoising methods based on multiresolution transforms had been proposed relying on elaborate statistical dependencies between coefficients of typically overcomplete transforms [5–8]. In [9], Dabov et al. proposed the block matching 3D (BM3D) denoising algorithm based on the effective filtering in 3D transform domain by combining sliding-window transform processing with block-matching. They attenuated noise by the shrinkage of the transform coefficients. In this method, block-matching was employed to find blocks that exhibited high correlation to the reference block for all pixels in the image, thus time consuming. Nonlocal image denoising also performed state of the art. The basic idea of nonlocal image denoising was that there exist mutually similar blocks in natural images, and thus averaging them could remove the (random) noise. In this method, the denoised pixel was obtained by the weighted average of the gray values of all pixels in an image. Each weight was proportional to the similarity between the local neighborhood of the pixel being and the neighborhood corresponding to the other image pixels. The optimality of this approach under reasonable criteria was provided

☆ This paper has been recommended for acceptance by Yehoshua Zeevi.

* Corresponding author.
E-mail address: zhengzk@xidian.edu.cn (C. Jung).

http://dx.doi.org/10.1016/j.jvcir.2015.04.015
1047-3203/

in [10]. Although nonlocal image denoising produced state of the art performance, the complexity was quadratic in the number of image pixels. In general, natural images contain much more smooth regions than complex regions. Thus, we perform nonlocal mean filtering only in complex regions to reduce the computational cost. Moreover, signal processing for image denoising have been discussed in the literature [11–23]. Although spatial filtering was able to preserve the motion of video sequences, it failed to remove noise sufficiently and caused the flicker in a wide range of noise distribution where the noise artifacts covered many pixels over a number of levels. In this case, temporal filtering was generally considered better than spatial filtering. However, temporal filtering produced dragging effects on moving objects [11]. To overcome the drawback, the temporal and spatial adaptive filters were proposed [12–14]. To preserve the motion of the video sequences, the spatiotemporal filtering was performed after the motion compensation [15]. If the accuracy of the compensation was insufficient, it was impossible to obtain the high quality video sequences.

1.2. Contributions

In recent years, visual perception studies have discovered that smooth regions corrupted by noise are much more annoying than complex ones corrupted by noise. Also, they have reported that the human visual system (HVS) is more interested in complex regions which contain image details. Thus, it is required to remove noise effectively especially in smooth areas and preserve image details in complex areas as much as possible. It has been reported that applying the same denoising filter to the whole image is not effective in removing image noise and preserving image details simultaneously [24]. Inspired by the previous work, we propose a perceptual video denoising method based on content adaptive filtering which differently performs video denoising in smooth and complex regions. As shown in Fig. 1, the proposed denoising method consists of four main parts: noisy frame detection, content adaptive segmentation, motion detection, and adaptive filtering. We first perform noisy frame detection which determines noisy frames in a video signal by estimating the noise level of the current frame. The noise level is also helpful for the subsequent two steps of content adaptive segmentation and motion detection. In the content adaptive segmentation step, complex regions and smooth regions can be easily distinguished by local statistical properties in noise-free frames. However, it is hard to distinguish between noise and image details in noisy frames due to high frequency characteristics of noise. Thus, we use noise detection results to obtain the amount of noise. Based on them, we define an adaptive segmentation threshold to distinguish between complex and smooth regions. Similarly, in the motion detection step, we determine moving pixels by a block matching in noise-free frames. In noisy frames, the block matching distance is often affected by noise. To deal with this problem, we define an adaptive motion threshold based on noise detection results. Finally, we adaptively remove noise according to contents, i.e. smooth regions or complex regions, using different noise filters in the adaptive filtering step. The main contributions of our work are summarized as follows:

(1) Based on the human visual perception, we provide a perceptual video denoising framework which utilizes different types of denoising filters according to context information, i.e., smooth regions or complex regions.
(2) In complex regions, we perform nonlocal mean filtering in the spatiotemporal domain to remove noise and recover the image details. Image details are successfully preserved

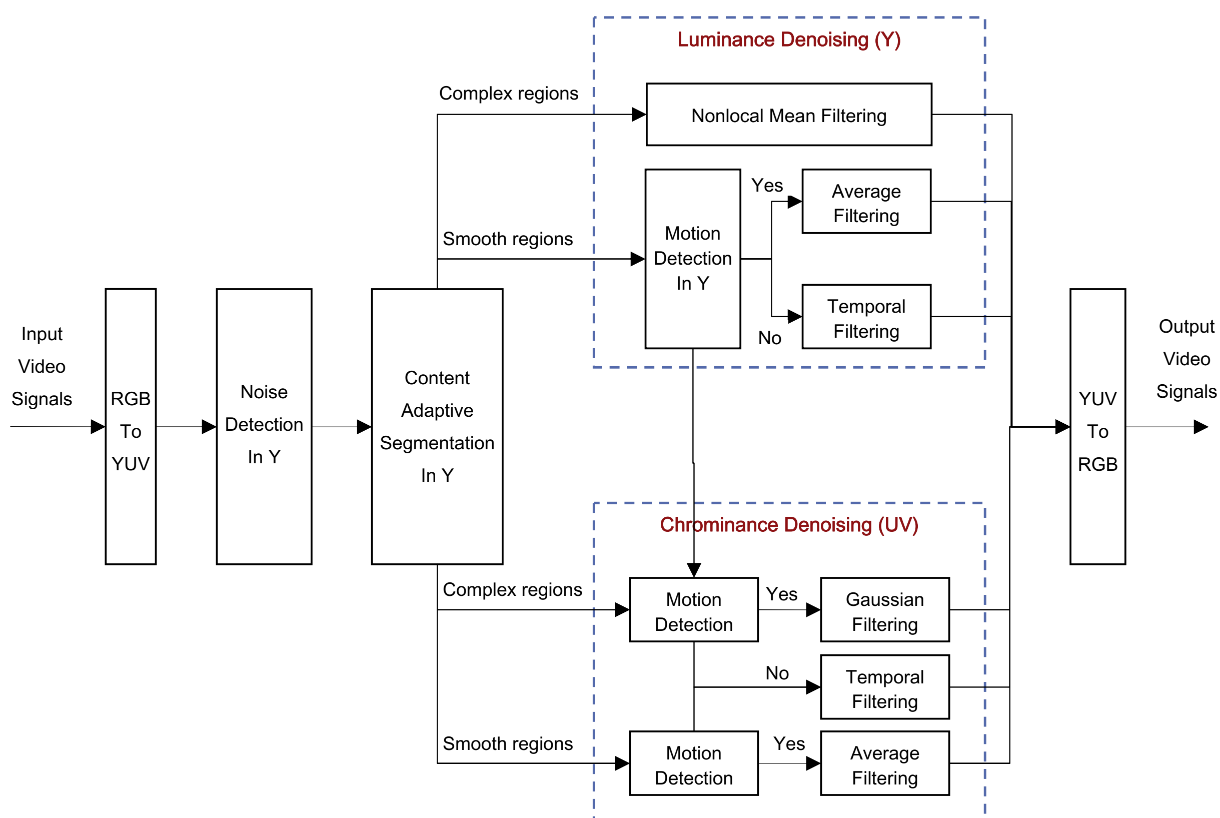

Fig. 1. Block diagram of the proposed perceptual video denoising method.