Information Sciences 281 (2014) 507–520

Contents lists available at ScienceDirect

Information Sciences

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

A novel segmentation based video-denoising method with noise level estimation

Yang Cao [a], Shi Jie Zhang [a], Zheng Jun Zha [b,*], Jing Zhang [a], Chang Wen Chen [c]

[a] University of Science and Technology of China, China
[b] Hefei Institute of Intelligent Machines, Chinese Academy of Sciences, China
[c] University at Buffalo, State University of New York, United States

ARTICLE INFO

Article history:
Received 31 January 2013
Received in revised form 8 May 2014
Accepted 19 May 2014
Available online 2 June 2014

Keywords:
Consistent segmentation
Noise level estimation
Multiplicative noise
Video denoising
Multimedia processing

ABSTRACT

Most state-of-the-art video-denoising algorithms assume an additive noise model, but such a model does not often reflect true conditions experienced in practice. In this paper, two main issues are addressed, namely, segmentation-based block matching and estimation of noise level. Unlike previously reported block-matching methods, the present method uses an efficient algorithm to perform block matching in spatially consistent segmentations of each image frame. To estimate the noise level function (NLF), which describes the noise level as a function of image brightness, a fast bilateral-median-filter-based method is proposed herein. Under the assumption of short-term coherence, this method of estimation is extended from a single frame to multiple frames. Coupling these two techniques together creates a segmentation-based, customised BM3D method that can be used to remove coloured multiplicative noise from videos. Experimental results obtained for benchmark data sets and real videos show that this method significantly outperforms other methods in removing coloured multiplicative noise.

1. Introduction

Noise reduction (denoising) is of crucial importance in multimedia processing [8,26,22,10]. Because digital images and videos are often contaminated by noise during acquisition, compression, and storage, it is desirable to reduce noise for visual improvement or as a preprocessing step for subsequent processing tasks [18,32,9].

A number of video-denoising methods are direct extension of methods used for image denoising. These include block matching and 3D filtering [3], wavelet shrinkage [21], PDE-based [25] and non-local means (NLM) methods [1]. Several methods integrate motion estimation with spatial filtering. For example, an NLM framework integrated with robust optical flow was introduced in a previous study [16]. The idea of sparse coding in a patch dictionary has also been applied to video denoising (e.g. [19,20]). In these cases, denoised image patches are detected by seeking the sparsest solution in a patch dictionary. In another previous study, the problem of denoising patch stacks was converted to the problem of recovering a complete low rank matrix from noisy and incomplete observations [11].

The performance of a denoising method is highly dependent on how closely the real noise fits the noise model assumed by the method. Most state-of-the-art video denoising methods operate under the premise of additive noise. However, such

* Corresponding author. Tel.: +86 551 65591153.
E-mail address: zhazj@iim.ac.cn (Z.J. Zha).

http://dx.doi.org/10.1016/j.ins.2014.05.031

noise is not often encountered in practice. According to a previous study, there are many varieties of noise, such as fixed pattern noise, dark current noise, shot noise, thermal noise, amplifier noise, and quantisation noise [7]. Dark current noise, thermal noise, amplifier noise and quantisation noise are additive noises. These types of noise are independent of image content. Fixed pattern noise and shot noise are multiplicative noises. These types of noise are dependent on image content. Practical image noise is probably coloured multiplicative noise. Here "coloured" refers to noise that is nonzero-valued, which is why most current denoising approaches cannot effectively estimate and remove real noise. This situation prevents noise removal techniques from being applied to practical multimedia applications.

The main contribution of this paper is its proposal of a segmentation-based customised BM3D method that can very effectively reduce the coloured multiplicative noise introduced by today's digital cameras. Two main issues are addressed in this paper: segmentation-based block matching and noise level estimation. An efficient method unlike previously reported block-matching methods, which search for similar blocks in a fixed-size neighbourhood, is introduced for block matching in spatio-temporally consistent segmentations for each image frame. The estimation of the noise level function (NLF), which describes noise levels as a function of image brightness, is key to ensuring the removal of coloured multiplicative noise. In the present method, a bilateral median filter is exploited to estimate the NLF by fitting a lower envelope to the discrete samples measured from image segmentations. To render video processing more efficient, the method of extrapolating information from single frames to multiple frames is performed under the assumption of short-term coherence of segmentations.

Because the proposed method can provide segmentation-based denoising results, it can be used in video enhancement, video decoding, and many other video applications [29,28,13]. For example, Wang et al. presented a novel scheme that could automatically convert a video clip into six-frame comic format [27]. The proposed method can serve as a preprocessing step and form segmented homogeneous regions to provide enough information for the detection of important and reliable key shots.

2. Related work

Several video denoising methods have been proposed over the past few years. Because a detailed review is beyond the purview of this paper, only the work most similar and relevant to the present method is discussed.

Over the past several decades, a variety of denoising methods have been developed for use in imaging. The significance of denoising processes lie in their ability to allow the user to exploit image sparsity [33]. In the frequency and transform domains, when a natural image is decomposed into multiscale-oriented sub-bands [24,6], the user can observe highly kurtotic marginal distributions [5,30]. In this context, image sparsity led to the development of coring algorithms, which were used to suppress low-amplitude values while retaining high-amplitude values [23]. In the spatial domain, image sparsity is formulated as image self-similarity, specifically a quantification of the number and size patches in an image that are similar to one another. NLM methods were proposed by finding patterns similar to a query patch and using the mean or other statistics to estimate the true pixel value [1]. Recently, several improved NLM methods have been proposed by the integration of pre-classification with block matching [34,31]. Other denoising algorithms, such as PDE and region-based denoising, also implicitly formulate sparsity in their representation [25,17]. Sparsity is also present in videos. A number of previously developed video denoising methods have been directly extended from image denoising. Because video sequences usually feature a great amount of temporal redundancy, a new frame can be predicted from previous frames using motion estimation. Several methods integrate motion estimation with spatial filtering for better performance [16]. The advances in sparse representations have shown outstanding denoising results [19,20,29].

The frequency and spatial forms of image sparsity are both part of the BM3D method [3]. BM3D is one of the most effective image-denoising algorithms. This method relies on a two-step paradigm: grouping and collaborative filtering. First, mutually similar 2-D image blocks are stacked into 3-D groups (grouping), and then the groups are filtered through transform-domain shrinkage (collaborative filtering), which simultaneously provides individual estimates for each grouped block. These estimates are returned to their respective locations and eventually aggregated to produce the final denoised image. Because it leverages both the nonlocal and local spatial correlation of natural images, exploiting the abundance of mutually similar patches and the high correlation of image data within each patch, the BM3D method produces high-quality results. This method has been extended to video denoising using the V-BM3D algorithm by exploiting more redundancy information in videos [2].

Although many video denoising methods produce state-of-the art results (e.g. [3,1,11]), most of them assume an additive noise model, which often does reflect conditions encountered in practice. As stated in two previous studies, there are many varieties of noise, such as fixed pattern noise, dark current noise, shot noise, thermal noise, amplifier noise, and quantisation noise [7,14]. Fixed pattern noise is usually modelled by a constant multiplier associated with each site. Dark current noise is due to the thermal energy and usually adds a constant offset to the observed intensity plus a small fluctuation, which can be modelled as Gaussian noise. Shot noise depends on image irradiance, but thermal noise is independent of scene radiance. All other noises are considered Gaussian noise. The noise model of a CCD camera was proposed as follows in a previous study [17]:

$$I = \Phi(L + n_s + n_c) + n_q. \quad (1)$$