Medical Image Analysis 24 (2015) 41–51

Contents lists available at ScienceDirect

Medical Image Analysis

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

Echocardiogram enhancement using supervised manifold denoising

Hui Wu[a,*], Toan T. Huynh[b], Richard Souvenir[a]

[a] Department of Computer Science, University of North Carolina at Charlotte
[b] Department of General Surgery, Carolinas Medical Center

ARTICLE INFO

Article history:
Received 12 September 2014
Revised 29 April 2015
Accepted 5 May 2015
Available online 21 May 2015

Keywords:
Biomedical image processing
Echocardiography
Image denoising
Supervised learning
Video signal processing

ABSTRACT

This paper presents data-driven methods for echocardiogram enhancement. Existing denoising algorithms typically rely on a single noise model, and do not generalize to the composite noise sources typically found in real-world echocardiograms. Our methods leverage the low-dimensional intrinsic structure of echocardiogram videos. We assume that echocardiogram images are noisy samples from an underlying manifold parametrized by cardiac motion and denoise images via back-projection onto a learned (non-linear) manifold. Our methods incorporate synchronized side information (e.g., electrocardiography), which is often collected alongside the visual data. We evaluate the proposed methods on a synthetic data set and real-world echocardiograms. Quantitative results show improved performance of our methods over recent image despeckling methods and video denoising methods, and a visual analysis of real-world data shows noticeable image enhancement, even in the challenging case of noise due to dropout artifacts.

1. Introduction

Whether for visual enhancement or as pre-processing for downstream algorithms, such as anomaly detection (Qian et al., 2011), object segmentation (Santiago et al., 2013), motion analysis (Huang et al., 2014; Papademetris et al., 2002; Yu et al., 2014), or atlas construction (Duchateau et al., 2011), denoising facilitates visual data interpretation by removing noise, de-emphasizing distractors, and increasing the definition of relevant organ structures. Compared with other cardiac imaging techniques such as magnetic resonance imaging (Huang et al., 2011), denoising is especially important for echocardiograms since the images often suffer from ultrasound imaging noise. Fig. 1 shows sample echocardiograms depicting common cases of ultrasound imaging noise. The images in the first column are relatively clear examples with well-defined cardiac structures. The frames in the second column depict an elevated amount of speckle noise.[1] This can be observed as the granular pattern resulting from the scattering of ultrasound signals, which can introduce discontinuities at the boundaries of larger tissue and, often, a lack or loss of clarity of smaller structures (e.g., heart valves). The third column shows images containing dropout artifacts, which are typically caused by a loss of tight contact between the transducer and the patient, insufficient conductive gel, or extra fluid or fat tissue between the transducer and the heart. Dropout artifacts can result in the loss of visibility in part of the structures of interest, and often affect consecutive frames in a video. These example frames demonstrate a common issue: often, real-world biomedical data contains multiple sources of noise.

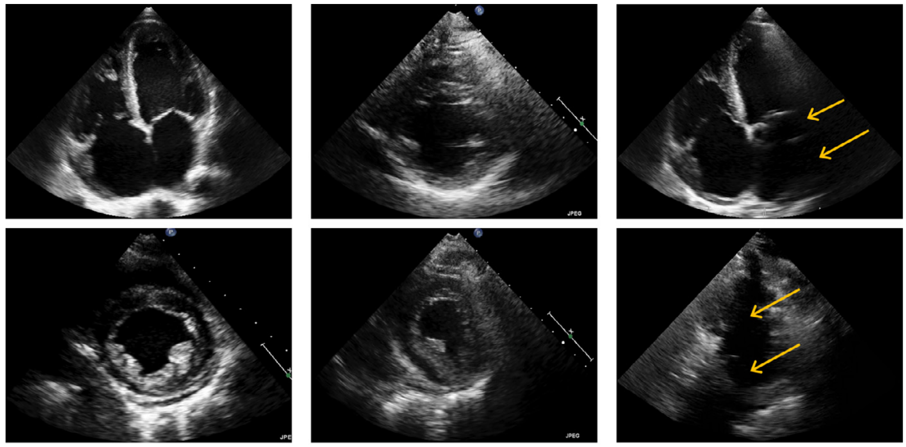
Many existing video denoising algorithms extend single-image algorithms, with some modification to incorporate temporal regularization. In the single-image case, most methods assume a prior underlying model on the noise distribution, such as zero-mean Gaussian noise or a Poisson distribution (Zhang et al., 2008). For single images, even in the biomedical domain, these assumptions are often reasonable and provide adequate denoising. However, as in the echocardiography example shown in Fig. 1, a single statistical model may be insufficient to represent the complexity of multiple, different noise sources.

In this paper, we do not make strong assumptions regarding the noise model, but rather make use of the fact that the number of underlying causes of image change in biomedical video is often quite low. That is, the number of degrees of freedom in such sets is small and often enumerable. For example, cardiac ultrasound frames may vary due to cardiac phase, patient breathing, transducer motion, and imaging noise. In the case of denoising, or other biomedical video analysis tasks, this structure can often provide a more perceptually meaningful basis for regularization than the temporal order of frames in a sequence. However, factoring the causes of image change directly from visual data and recovering this underlying structure is a non-trivial problem due to the high dimensionality of data and low

* Corresponding author. Tel.: +17 046123398.
E-mail address: hwu13@uncc.edu (H. Wu).

[1] For image enhancement, speckles are considered as unwanted visual artifacts. However, it is worth noting that speckle patterns can be useful features for automated tracking algorithms (e.g., Yue et al., 2009).

http://dx.doi.org/10.1016/j.media.2015.05.004
1361-8415/



(a) Clear (b) Speckle Noise (c) Dropout Artifacts

Fig. 1. Biomedical video often suffers from multiple sources of noise. These images show echocardiogram frames that suffer from speckle noise (middle) and dropout artifacts (right) highlighted by arrows.

signal-to-noise-ratio (SNR). To address this problem, we take advantage of the fact that in clinical biomedical settings, there is often available, associated metadata collected during image acquisition that can serve as a proxy for the dominant cause of image change. For example, in echocardiography, electrocardiograms (ECG), which record cardiac electrical activity and, therefore, heart phase information, are usually acquired along with synchronized echocardiograms.

This paper, which extends our previous work (Wu et al., 2013) with a new kernelized model, incorporates motion-relevant context for data-driven video denoising rather than learning this structure directly from noisy image data. Since our approach relies on the underlying relationships between images for denoising, it naturally applies to cases when multiple sources of noise are present.

2. Related work

The literature on video denoising, both within and without the biomedical domain, is vast. Many of the approaches tailored to video are extensions of single-image denoising methods that incorporate some form of temporal regularization. In general, these methods consider noisy images as ideal noise-free images corrupted by random noise drawn from a specified probability distribution and introduce additional constraints to solve the under-constrained task of separating the noise-free image from the noisy input.

Image-based video denoising. One common model is to assume independent and identically distributed, zero-mean Gaussian additive noise, where images are denoised by finding the mean of neighboring pixels, and the neighborhood can be defined spatially and/or temporally based on some notion of similarity (Buades et al., 2005; 2008; Ghoniem et al., 2008; Ren et al., 2012; Xu et al., 2010). For ultrasound image analysis, more complex noise models have been considered, including multiplicative, locally-correlated speckle noise. This model has been incorporated into algorithms that employ non-local means (Coupe et al., 2009) and wavelet shrinkage (Gupta et al., 2007; Khare et al., 2010). One of the most widely-applied approaches, speckle reducing anisotropic diffusion (SRAD) (Yu and Acton, 2002), considers speckle removal as an iterative edge-preserving diffusing process. There have been a number of extensions for ultrasound enhancement (Aja-Fernández and Alberola-López, 2006; Aksel et al., 2006; Krissian et al., 2005; 2007; Yu et al., 2010), which follow the general SRAD framework. These image denoising methods all share the same inherent assumption of a single noise model, whereas in this paper, we consider cases of biomedical image analysis, which exhibit noise models with composite causes, including factors that are difficult to model parametrically.

Manifold denoising. Compared to video denoising methods extended from single-image algorithms, there has been some work that considers the entire video as a whole by considering image-level relationships. Manifold denoising approaches take advantage of the property that related images, when considered as points in a high-dimensional space, lie on or near a low-dimensional manifold. Manifold-based methods denoise images by combining information from nearby images on the manifold, rather than temporal neighbors. One method for manifold denoising learns a global manifold representation of the image set using density estimation, and images are denoised by minimizing the sum of deviation from the learned manifold (Sun et al., 2012). Similarly, there have been denoising approaches based on other unsupervised nonlinear dimensionality reduction techniques, such as Kernel Principal Component Analysis (Kwok and Tsang, 2003), autoencoders (Vincent et al., 2010), and Gaussian Process Latent Variable Model (Gao et al., 2008). These methods use the learned low-dimensional representation of images to denoise via back-projection into image space. There are other methods that do not rely on constructing a global manifold structure, but rather exploit the locally smooth and linear properties of a manifold. Local Principal Component Analysis (PCA) has been used to estimate local manifold directions, and images are denoised by varying the image along the vector in image space in the direction orthogonal to the tangent space (Wang and Carreira-Perpiñán, 2010; Wang et al., 2011), or by minimizing the overall reconstruction error (Gong et al., 2010). The nearest-neighbor graph, used in graph embedding approaches, has been used to approximate the manifold topology and denoise images by graph diffusion, which iteratively reduces the differences between neighboring images (Hein, 2007; Hein and Maier, 2006). An important assumption underlying these manifold denoising methods is that the manifold structure can be well-approximated solely from the input images. However, robustly learning a low-dimensional manifold is a non-trivial problem in presence of composite causes of imaging noise, where the (multimodal) noise variance may dominate data variance.

Denoising with side information. One approach to side-stepping the problem of learning a low-dimensional representation for image data is to incorporate side information. In contrast to the large amount of work done in unsupervised image and video denoising, there is much less that incorporates side information, or metadata. In some methods side information has been used during training to learn a