CSC2529 Computational Imaging 2022 - Final Project Proposal

Deep Video Denoising for Facial Signal Processing

Xiaoyan Li
xiaoy.li@mail.utoronto.ca

1 Motivation

Over the course, we have learned many effective deblurring and denoising algorithms, such as local-nonlinear filtering, bilateral filtering, Wiener deconvolution, Winer deconvolution + deep learning, and DnCNN Zhang et al. (2017). These algorithms, especially based on CNN, have achieved promising results in image denosing tasks, implying their capability of addressing more challenging denosing tasks for videos. Generally, video denoising can be categorized as spatial video denoising, temporal video denoising, and spatial-video denoising Video denoising (2022). In spatial video denoising methods, denoising is applied to each frame individually; in temporal video denoising methods, denoising is applied to successive frame pairs; and spatial-temporal video denoising methods (i.e., 3D denoising) use a combination of spatial and temporal denoising. Due to the impressive performance of the spatial-temporal method, it has recently been explored by many researchers, such as Claus and van Gemert (2019), Tassano et al. (2020), and Sheth et al. (2020). In this project, one of these advanced algorithms will be implemented using real facial video data. Based on Sun et al. (2011), facial video could be used to simulate finger PPG (finger photoplethysmography), which is called an imaging PPG (iPPG). Therefore, in this project, as an extended application of the denoised facial video, iPPG will be extracted from the denoised facial video and compared to finger PPG. This can also work as an evaluation metric of video denoising when PSNR can not be calculated due to the unavailability of ground truth video.

2 Related Work

Since, in a real implication, the ground truth video might not be available, the unsupervised method would not be suitable. Sheth et al. introduced unsupervised video denoisingSheth et al. (2020). However, they used raw videos that take a lot more space to save than rendered videos, which limited their applications. Claus and Gemert Claus and van Gemert (2019) proposed ViDeNN, which used a CNN for video denoising without prior knowledge of the noise distribution. They used spatial-temporal video denoising methods to spatially denoise each frame and simultaneously combine their temporal information to deal with objects' motion, brightness changes, low-light conditions, and temporal inconsistencies. Tassano et al. (2020) proposed FastDVDnet algorithm, which exhibits several desirable properties, including a small memory footprint and the ability to handle a wide range of noise levels with a single network model. With these attractive properties, in this project, FastDVDnet will be used as the main frame for our facial video denoising.

3 Project Overview

The project will consist of the following steps:

1. Read related papers and understand the FastDVDnet algorithm.
2. Implement FastDVDnet using real facial video data.
3. Extract iPPG from the denoised facial video.
4. Evaluate the denoising results by visual inspection and comparing the extracted iPPG to finger PPG. The iPPG from denoised facial video should be more close to finger PPG than iPPG from the noisy facial video.

The main goal of this project is to let the video denoising algorithm clean the facial video such that it can better reflect the facial blood flow (iPPG). The facial video was recorded by using an iPhone 6s, and finger PPG was collected by using a finger PPG record system (Biopac MP160 monitoring system). To extract iPPG, OpenCV will be used to detect faces from each frame, and a mask will be applied to the facial images to cover the eyes and mouths. For simplicity, the spatial average of all the pixels in the green color channel will serve as the iPPG. There are 60 frames per second, which means the facial video will produce a 60HZ iPPG signal.

The fastDVDnet will be used as the main architecture for our facial video denoising. It includes a number of modified U-Net blocks, which have been shown to have the ability to learn misalignment. This architecture is trained end-to-end without optical flow alignment, which avoids distortions and artifacts, eliminating motion compensation without sacrificing performance Tassano et al. (2020).

4 Milestones

- 16 November -21 November: Read papers and understand CNN-based spatial-temporal denoising algorithms.
- 22 November - 29 November: Implement the FastDVDnet algorithm using the facial video data.
- 30 November - 2 December: Use the denoised facial video for image PPG extraction and evaluate the results.
- 3 December - 7 December: Prepare poster, write report, prepare code for submission.

References

Michele Claus and Jan van Gemert. Videnn: Deep blind video denoising. In *Proceedings - 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, CVPRW 2019*, pages 1843–1852, 2019. doi: 10.1109/CVPRW.2019.00235.

Dev Sheth, Sreyas Mohan, Joshua Vincent, Ramon Manzorro, Peter Crozier, Mitesh Khapra, Eero Simoncelli, and Carlos Fernandez-Granda. Unsupervised deep video denoising, 11 2020.

Yu Sun, Sijung Hu, Vicente Azorin Peris, Stephen Greenwald, Jonathon Chambers, and Yisheng Zhu. Motion-compensated noncontact imaging photoplethysmography to monitor cardiorespiratory status during exercise. *Journal of biomedical optics*, 16:077010, 07 2011. doi: 10.1117/1.3602852.

Matias Tassano, Julie Delon, and Thomas Veit. Fastdvdnet: Towards real-time deep video denoising without flow estimation. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1351–1360, 2020. doi: 10.1109/CVPR42600.2020.00143.

Video denoising. Video denoising — Wikipedia, the free encyclopedia, 2022. URL `https://en.wikipedia.org/wiki/Video_denoising`. [Online; accessed 15-November-2022].

Kai Zhang, Wangmeng Zuo, Yunjin Chen, Deyu Meng, and Lei Zhang. Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising. *Trans. Img. Proc.*, 26(7):3142–3155, jul 2017. doi: 10.1109/TIP.2017.2662206.