Video Denoising with local linear denoising and non-local means

EE 367 - Project Proposal

Emil Vardar Elin Byman

February 18, 2022

1 Motivation

Over the course, we have talked a lot about denoising of images. Denoising, in a nutshell, is to reconstruct or estimate the ground truth image from its noisy variant. Over the years, there have been many different suggestions for denoising images. Some of the rudimentary denoising methods are: local-linear smoothing, local-nonlinear filtering, and bilateral filtering as discussed in class [1]. Even though these methods do a decent job, much better denoising algorithms have been developed in the last couple of decades. One of them, which also will be the main algorithm in our project, is the non-local means (NL-means) algorithm. NL-means algorithm has been proven to be a suitable method for image denoising over the last years [1, 2, 3].

One closely related subject to image denoising, which has not been discussed in the class, is video denoising. Actually, in general, noise is even a more significant issue when it comes to videos due to high speed capturing in videos [4]. In other words, many images that have been stacked together with decreased capture time causes an increase in the noise per frame.

When denoising videos, the first thing that comes to mind might be to denoise each frame separately using an image denoising method. That is, use image processing techniques on each frame separately. However, this does not take advantage of the fact that frames close in time are often very similar and could, therefore, be used to improve the denoising. Take, for example, a video with 24 frames per second (fps). This means that we have 24 images per second. How different can an image taken 42 ms after another image be? In many scenarios, not that much! This means that even if there is more significant noise per frame in a video, a larger amount of similar patches can be used to denoise a specific frame. Therefore, in this project, our goal is to use this knowledge about image denoising and expand it to denoise videos. This project will implement two different video denoising algorithms and compare them to state-of-the-art VBM3D (see section 3).

2 Related work

Image denoising is and has been a widely studied field, and as mentioned in section 1 many different algorithms have been proposed over the years. This work will focus on smoothing different frames with non-local means (NL-means) algorithm. NL-means algorithm takes advantage of the high redundancy on a natural image [2]. This means that every small window in an image has many similar windows on other places in the same image [2]. The core idea that separates this method from the earlier methods is that it searches for similar patches for the window in focus in the entire image. In [2], Buades et al. show that the NL-means algorithm successfully reduces noise while keeping the sharp edges.

However, as mentioned in [3], the complexity for the NL-means algorithm when searching for similar windows over the whole image is too large to be able to realize it in any practical application. Therefore, the search area must be limited to a particular patch size [1, 3]. I.e., instead of searching for similar windows in the entire image, the search will be limited to a smaller region in the same image. In [3], Mahmoudi et al. have improved the complexity at no quality cost, which raises hope for using NL-means in video denoising.

In [5], Dabov et al. introduce the so-called VBM3D algorithm for video denoising. In the paper [5], the authors show that VBM3D gives a higher PSNR per frame compared to 3DWTF and WRSTF. They show that the PSNR in their method increased more or less 2 dB/frame (for Gaussian noise with $\sigma = 20$), which is a significant increase. One drawback with this implementation is that the authors only consider Gaussian noise. In our project, we will also follow this frame of mind. However, Ji et al. in [4] show that their robust implementation for video denoising using Low-rank matrix completion can remove 'serious mixed noise from the video data'. Ji et al.'s algorithm is shown to be more effective on mixed noise compared to VBM3D. That Ji et al.'s algorithm can remove mixed noise better than VBM3D makes it more useful in real-world applications since there are many noise factors beyond the Gaussian noise, e.g., Poisson noise [1]. However, we want to underline that our goal is not to implement a video denoising application for mixed noise. Therefore, we will use the VBM3D to compare our results.

In the last couple of years, machine learning approaches have been used for video denoising, as in many other research areas. One of them is [6]. In this article, the authors show that their algorithm, which is based on deep neural network, gives a slightly better PSNR than VBM3D. Even though there are many better methods than VBM3D, we will compare our results to VBM3D in this work, because of easy accessibility (see [7]) and that it works good enough for our purpose.

3 Overview

The project will consist of the following four steps:

1. Implementation of a local linear denoising method in the temporal dimension,
2. Implementation of a non-local means method both in the spatial and temporal dimensions,
3. Comparison of results of each method using different parameters (PSNRs and visual inspection),
4. Comparison of the implemented methods to each other and to an already implemented VBM3D denoising algorithm (PSNRs and visual inspection).

The local linear method will reduce noise in the video by averaging pixels at the same spatial location but in successive frames. The averaging will be weighted using a Gaussian low pass filter and will only consider the frames in a local neighborhood.

The non-local means method reduces noise by selecting a patch around the pixel for each pixel in each frame. This will be done for one pixel at a time, and the current pixel and its patch will be called the main pixel and the main patch from now on. The main patch will be used to search for similar patches in a neighborhood around the pixel in the same frame and successive frames. Ideally, all pixels in all frames would be included in the search, but since that would be too computationally heavy, the search is restricted to a smaller neighborhood both in space and in time (see section 2). A weight will be assigned to the pixel in the middle of each patch being searched. The value of the main pixel will be calculated based on these weights and the corresponding pixel values.

The local linear method will be tested using different parameters, e.g., using different sizes of the Gaussian kernel, and the results from using the different parameters will be compared. The same goes for the non-local means method, where for example, different neighborhood sizes and different patch sizes will be tested, and the results will be compared.

The results of the two implemented methods will also be compared to each other and to the results of a third method, the VBM3D method. This method has already been implemented in [7] and will not be implemented by us.

This project will only consider additive Gaussian noise with zero mean. This is one of the most important sources of noise in most images and videos, and due to the time limitation of the project, no other noise distributions will be considered.

4 Milestones

- -18 February: Read related work and write project proposal.
- 18 February - 2 Mars: Implement the local linear method and the non-local means method.
- 3 Mars - 6 Mars: Compare the results using different parameters and methods.
- 7 Mars - 11 Mars: Prepare poster, write report, finish code for submission.

References

[1] Gordon Wetzstein. Lecture notes in Digital Photography II, The Image Processing Pipeline, January 2022.

[2] Antoni Buades, Bartomeu Coll, and Jean Michel Morel. On image denoising methods. *CMLA Preprint*, 5:19–26, 2004.

[3] Mona Mahmoudi and Guillermo Sapiro. Fast image and video denoising via nonlocal means of similar neighborhoods. *IEEE signal processing letters*, 12(12):839–842, 2005.

[4] Hui Ji, Chaoqiang Liu, Zuowei Shen, and Yuhong Xu. Robust video denoising using low rank matrix completion. In *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 1791–1798. IEEE, 2010.

[5] Kostadin Dabov, Alessandro Foi, and Karen Egiazarian. Video denoising by sparse 3d transform-domain collaborative filtering. In *2007 15th European Signal Processing Conference*, pages 145–149. IEEE, 2007.

[6] Xiangyu Xu, Muchen Li, Wenxiu Sun, and Ming-Hsuan Yang. Learning spatial and spatio-temporal pixel aggregations for image and video denoising. *IEEE Transactions on Image Processing*, 29:7153–7165, 2020.

[7] Y. Mäkinen, L. Azzari, E. Sánchez-Monge, M. Maggioni, A. Danielyan, K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian. *Image and video denoising by sparse 3D transform-domain collaborative filtering.* Tampere University of Technology, 2014 [Online].