EE 367 Final Project: BM3D Implementation for Video Denoising

Caelia Thomas

Abstract—Image denoising is a foundational support for video denoising. BM3D specifically is a non-local means denoising method that was developed to be more efficient than classical denoising methods. It's performance was tested and proven in other referenced studies with a PSNR of approximately 30. In the scope of study for this project, the experimental results do not support this fact, and that can be due to a lack of understanding on the author's part.

Index Terms—Computational Photography

1 Introduction

IMAGE denoising is the process of reconstructing or estimating a ground truth image from a noisy image. From a consumer's point of view, this is an important process because of its ability to help correct and aid in taking visually appealing photos under non-optimal conditions. From a research and development perspective, denoising is critical for any process that involves any form of image and information identification.

The fundamental concept behind denoising is that similar pixels in one area can be averaged together in order to reduce noise in said area of an image [1]. Mathematically, a noisy image can be represented in the following form:

$$x = y - \eta \tag{1}$$

where x is the noisy image, y is the ground truth image, and η is noise. The process of denoising then aims to efficiently solve the above equation for the ground truth image. In the most common cases, noise is approximated to have a zero mean and a Gaussian distribution. For the purposes and scope of this project, we therefore focused on Gaussian noise.

There are many variations and improvements of denoising that exist today, and these techniques have different tradeoffs that make them optimal for different uses. The scope of EE 367 has covered several denoising techniques with respect to images, but there is an entire area of video denoising that was not discussed. The motivation and aim of this final project was to test and compare the ability of the BM3D denoising method with other techniques covered in class with respect to video denoising.

2 Related Work

Current denoising methods fall into the categories of classical, transform-based, or convolutional neural network (CNN) based techniques [2]. Classical methods include algorithms like spatial domain filtering, which is when linear or non-linear filters are applied to a noisy image in order to remove spatial domain noise. While they are successful methods for clarifying and smoothing, there remain issues such as over-smoothing and the blurry edges. The most popular filtering method, bilateral filtering, is a popular method that is very successful in smoothing an image while preserving edges. It replaces the intensity value of each pixel with a weighted average of the nearby pixel intensity values. A main issue with this method, however, is that it is a computationally heavy method, so it takes a notable amount of time due to it's inefficiency.

Non-local means method are classified as a transform-based technique, and are considered more effective and efficient than the previously describes techniques. Within one image, there is an extensive amount of similar patches of pixels. Non-local means denoising methods use a weighted average of pixels from the different patches in order to reduce the noise and estimate the pixel intensity values of a specific similar patch. Such an estimation from this style of weighted average results in a notable improvement of noise reduction when compared to local means methods; especially when handling high amounts of noisy in an image.

3 Proposed Method

This section provides an overview of the BM3D method before properly presenting the theory behind it. Other denoising methods are then presented for the sake of comparison.

3.1 Block Matching 3D Arrays (BM3D)

3.1.1 Overview

For the scope of this project, there is a focus on BM3D specifically, which is a non-local method developed by Dabov et al. [3] that uses collaborative filtering on 3D data arrays (referred to as "groups") of 2D image fragments. Collaborative filtering is a procedure that involves a 3D transformation of a pixel group, a shrinking the transform spectrum, and a subsequent inverse transform of the same 3D group. The result of the filtering, after returning the groups to their original positions in the image, is an estimate of the ground-truth image. BM3D was later improved by adding principal

- *C. Thomas is with the Department of Electrical Engineering, Stanford University, Stanford, CA, 94305.*
  E-mail: caelia@stanford.edu

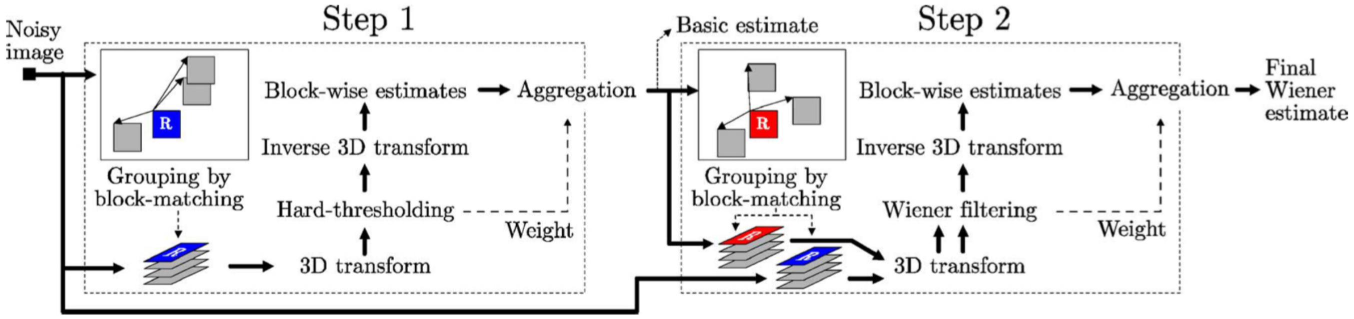

Fig. 1. Flowchart showing the BM3D process in [3]. Operations inside dashed lines are repeated for each processed block.

component analysis (PCA) [4] which groups mutually similar adaptive-shape neighborhoods. BM3D struggles if the noisy image is too sparse, so the benefit of incorporating PCA is that the true signal's sparsity is improved, which therefore improves the effectiveness of BM3D's filtering.

3.1.2 Theory: Block-wise Estimates

The first part of the BM3D method is grouping similar pixel groups before moving onto collaborative filtering. Each pixel-group is referred to as a block, and grouping by block-matching occurs within the noisy image, z. Block-matching is regulated by ensuring that the distance between selected blocks with respect to the reference block (denoted as Z_{x_R}, which is block that is currently being processed). This distance can be represented as

$$d^{ideal}(Z_{x_R}, Z_x) = \frac{\|Y_{x_R} - Y_x\|_2^2}{(N_1^{ht})^2} \tag{2}$$

where the blocks Y_{x_R} and Y_x are blocks being investigated within the ground truth image y. Equation (2) has an exception where if blocks Z_{x_R} and Z_x overlap, then the variance of the equation grows asymptotically. To avoid this problem, the block distance is measured using a coarse prefiltering

$$d(Z_{x_R}, Z_x) = \frac{\|\Upsilon'({}^{ht}_{2D}(Z_{x_R})) - \Upsilon'({}^{ht}_{2D}(Z_x))\|_2^2}{(N_1^{ht})^2} \tag{3}$$

where Υ' is the hard-threshholding operator with threshhold $\lambda_{2D}\sigma$ and ${}^{ht}_{2D}$ is the normalized 2-D linear transform. The result of block-matching is as folllows:

$$S^{ht}_{x_R} = x \in X : d(Z_{x_R}, Z_x) \leq \tau^{ht}_{match} \tag{4}$$

where the fixed τ^{ht}_{match} is the maximum d-distance for which two blocks are considered similar. Stacking the matched noisy blocks $Z_{x \in S^{ht}_{x_R}}$ forms a 3D array denoted as $\mathbf{Z}_{S^{ht}_{x_R}}$. The collaborative filtering of $\mathbf{Z}_{S^{ht}_{x_R}}$ is realized by hard-thresholding in the 3-D transform domain. The true signal group is then represented as

$$\hat{\mathbf{Y}}^{ht^{-1}}_{S^{ht}_{x_R}} = \tau^{ht^{-1}}_{3D}(\Upsilon(\tau^{ht}_{3D}(\mathbf{Z}_{S^{ht}_{x_R}}))) \tag{5}$$

where $\tau^{ht^{-1}}_{3D}$ is the normalized 3-D linear transform.

3.1.3 Theory: Grouping and Collaborative Wiener Filtering

Given the basic estimate $\hat{y}^{basic}$ of the true image from Eq. 5, the denoising is further improved with Wiener filtering. After obtaining and defining the empirical Wiener shrinkage coefficients, the inverse transform $\tau^{ht^{-1}}_{3D}$ produces a group of estimates

$$\hat{\mathbf{Y}}^{wie}_{S^{wie}_{x_R}} ={}^{wie^{-1}}_{3D}(\mathbf{W}_{S_{x_R}}{}^{wie}_{3D}(\mathbf{Z}_{S_{x_R}})) \tag{6}$$

where the block-wise estimates $\hat{Y}^{wie,x_R}_x$ are located at the locations $x \in S^{wie}_{x_R}$.

Each collection of estimates $\hat{Y}^{ht,x_R}_{x \in S^{ht,x_R}_{x_R}}$ and $\hat{Y}^{wie,x_R}_{x \in S^{ht,x_R}_{x_R}}$ together create a complete representation of the ground truth image (note that these expressions represent the block-wise estimates $\forall x_R \in X$). these estimates are then aggregated using weights that are inversely proportional to the total sample variance in the group of estimates. Noting these weights as $w^{ht}_{x_R}$;

$$\hat{y}^{basic}(x) = \frac{\sum_{x_R \in X} \sum_{x_m \in S^{ht}_{x_R}} w^{ht}_{x_R} \hat{Y}^{ht,x_R}_{x_m}(x)}{\sum_{x_R \in X} \sum_{x_m \in S^{ht}_{x_R}} w^{ht}_{x_R} \chi_{x_m}(x)}, \forall x \in X \tag{7}$$

where $\chi_{x_m} : X \rightarrow 0, 1$ is the characteristic function of the square support of a block located at $x_m \in X$. To get the final estimate of the ground truth image $\hat{y}^{final}$ is found by replacing $\hat{y}^{basic}, \hat{Y}^{ht,x_R}_{x_m}, S^{ht}_{x_R}, and w^{ht}_{x_R}$ respectively with $\hat{y}^{final}, \hat{Y}^{wie,x_R}_{x_m}, S^{wie}_{x_R}, and w^{wie}_{x_R}$ in Eq. 7.

3.2 Other Denoising Methods for Comparision

3.2.1 Gaussian

3.2.2 Bilateral

3.2.3 Non-Local Means

4 Experimental Results

For this project, the previously outlined denoising techniques were applied to videos with zero mean Gaussian noise where $\sigma = 0.5$. The project folder contains three possible videos. Each video was split into frames and Gaussian noise was added to each one. Afterwards, each denoising algorithm was applied before the frames were stitched back together.

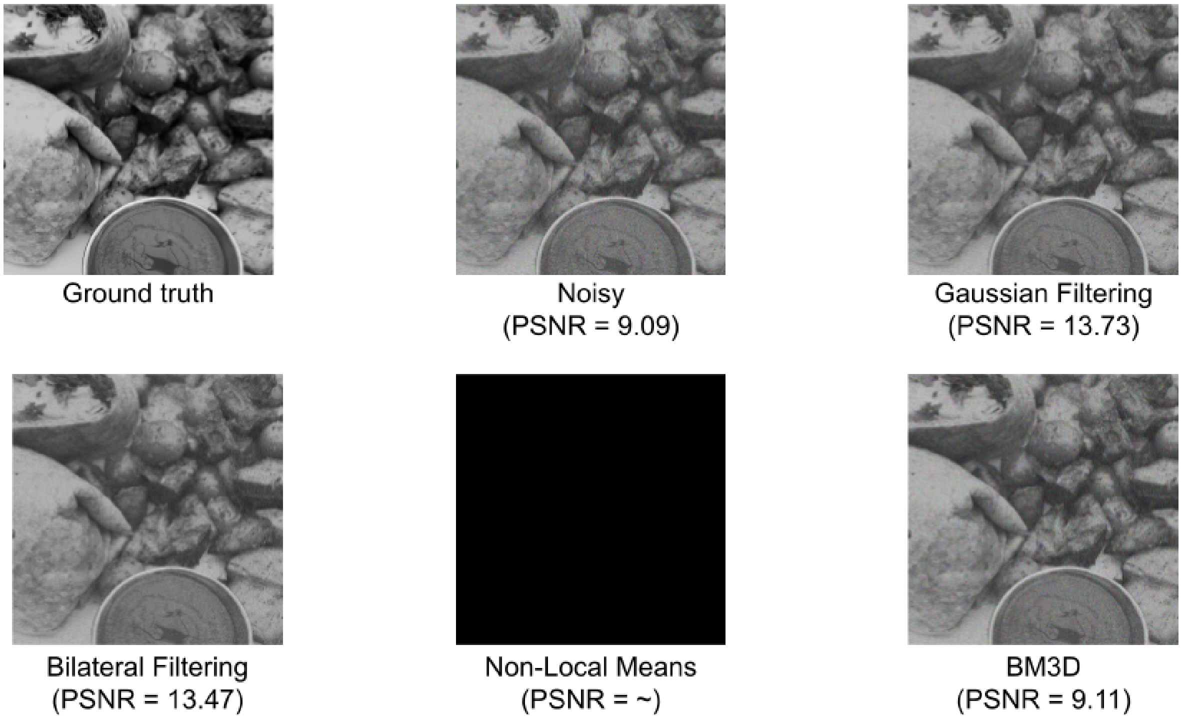

Fig. 2. Experimental results of the project.

4.1 Metrics

Each denoising algorithm was assessed visually as well as with the peak signal-to-noise ratio (PSNR) which can be represented as:

$$PSNR(x,y) = 20log_{1}0(\frac{maxx}{\|x-y\|_2^2}) \tag{8}$$

4.2 Comparison of Techniques

The experimental results of the denoising technique comparison for this project is summarized in Fig 2. According to the experimental results, Gaussian filtering outperforms all the other methods, with Bilateral filtering coming in as the second-best technique. The two filtering techniques produced the cleanest looking images and had the best relative PSNRs.

These results, however, completely contrast with the theory and formulation of all the methods included in this project. What *should* have been observed, was BM3D outperforming the other methods with a PSNR of approximately 30. The Non-Local Means method should have outperformed the filtering methods, and been the second-best performing method included in this study.

The deceiving results in this project could be the result of a lack of understanding regarding the BM3D python package. There also could have been an error in introducing Gaussian noise to the ground-truth image.

5 Conclusion

The aim of this final project was to qualitatively and quantitatively show that BM3D outperforms other classical denoising techniques that were discussed throughout the course of EE367. The experimental results did not reflect this fact, and incorrectly implies that Gaussian and Bilateral filtering are the superior denoising methods in this study.

If this study was to be repeated, more time would be spent toggling the default parameters in the BM3D method, as well as attempting to denoise the video and individual frames in full color.

References

[1] G. Wetzstein (2023), Stanford University Digital Photography II: The Image Processing Pipeline [Powerpoint slides]. Available: https://stanford.edu/class/ee367/slides/lecture4.pdf

[2] Fan, L., Zhang, F., Fan, H. et al. Brief review of image denoising techniques. Vis. Comput. Ind. Biomed. Art 2, 7 (2019). https://doi.org/10.1186/s42492-019-0016-7

[3] K. Dabov, A. Foi, V. Katkovnik and K. Egiazarian, ”Image Denoising by Sparse 3-D Transform-Domain Collaborative Filtering,” in IEEE Transactions on Image

TABLE 1
Summary of Experimental PSNR Values for each Denoising Algorithm

Method	PSNR
Noisy Image	9.09
Gaussian	13.73
Bilateral	13.47
Non-Local Means	–
BM3D	9.11

Processing, vol. 16, no. 8, pp. 2080-2095, Aug. 2007, doi: 10.1109/TIP.2007.901238.

[4] Dabov, Kostadin, et al. ”BM3D image denoising with shape-adaptive principal component analysis.” SPARS’09-Signal Processing with Adaptive Sparse Structured Representations. 2009.

[5] K. Dabov, A. Fol, and K. Egazarian, “Video denoising by sparse 3D transform-domain collaborative filtering,” Proc. 15th European Signal Processing Conference, EUSIPCO 2007, Poznan, Poland, September 2007.