

Effect of Image Pre-processing on SfM/MVS

Apurv Ashish(396509)¹

¹Department of Mathematical and Computer Science, TU Berlin

apurv.ashish@campus.tu-berlin.de

Keywords: SfM, 3D reconstruction, Feature Matching, Contrast Enhancement, De-noising, Camera-pose Estimation

Abstract: 3D-reconstruction from ordered/unordered set of photographic images has significance in applications that integrates virtual and real data such as augmented reality, scene reconstruction and SLAM algorithms. Structure from Motion is one form of 3D reconstruction, which consists of feature detection, matching, tracking and pose estimation. The key-point detection and matching are one of the critical steps in the SfM pipeline. Since the key-point detection & matching is heavily dependent on the available image quality, hence, image artefacts (if any) present in images can severely affect the key-point detection and matching, thereby adversely influencing the SfM output. In such cases, image pre-processing can enhance the functionality of feature detection algorithm and SfM significantly. The presented paper aims to study the effect of different image enhancement algorithms on the accuracy of SfM pipeline.

1 Introduction

Structure from motion (SfM) (Hartley and Zisserman, 2003) is the method of recreating 3D structure from a set of ordered/unordered 2-D images. SfM, in general, has different application such as 3-D scanning, augmented reality and recreating the structure of Architectural sites from the set of images to name a few.

There are different implementation/algorithms available to compute SfM. The choice of an algorithm is largely influenced by factors such as:

- number of cameras used for reconstruction.
- If the set of images are ordered/unordered

For computing SfM, there are few critical requirements, which must be met to obtain a 3d structure, such as:

- Minimum of two images.
- Overlapping features between images.

Depending on number of images used in 3d reconstruction, SfM can further be classified as 2-view and multi-view(shown in Figure 1 and 2.)

The algorithm used for SfM from 2-View is similar to or can be extended to multi-view reconstruction. SfM requires feature correspondences between images. The corresponding feature points between two images are either computed by feature tracking (like KLT tracker) or feature matching. Generally, the choice between feature matching is pre-

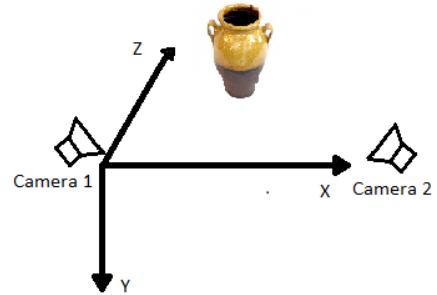


Figure 1: Two-View 3D reconstruction

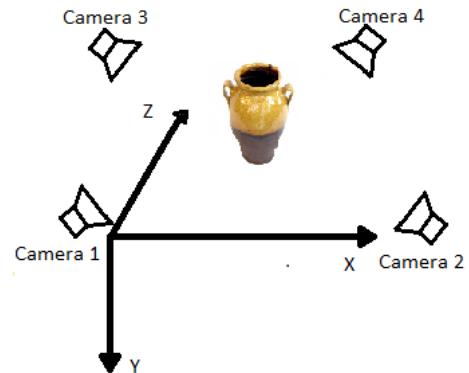


Figure 2: Multi-View 3D reconstruction

ferred in cases where camera positions are far apart.

The matched features are further used to compute the camera poses(generally, relative to first camera) using the fundamental matrix. The fundamental matrix provides the insight into the epi-polar geometry of a pair of camera(shown in figure 3). As briefly

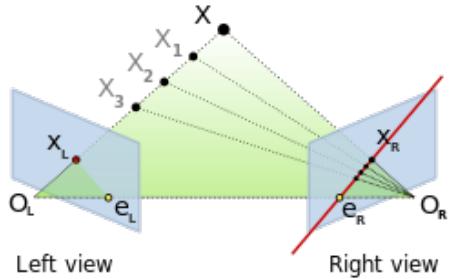


Figure 3: Epi-polar Geometry of two cameras.[source]

discussed above, the key-point detection and matching are one of the most critical and accuracy defining steps of any SfM pipeline. The matching of these key-point features in different images provides the information, as to how the points moves from one image to the another. The information is further used for triangulation of the points and the camera-pose estimation. In general, the images used in 3D reconstruction are taken from digital cameras. Such digital images are subject to a wide variety of distortions during acquisition, processing, compression, storage, transmission and reproduction, and any of these intermediate steps may introduce image artefacts or degrade the quality of the image. Additionally, the image acquisition is also dependent on various factors such as lighting condition, exposure time of camera or motion of the camera or the object itself. Image enhancement techniques, such as denoising and contrast enhancement can substantially improve the quality of the image before it is passed to the SfM pipeline. This paper presents a comprehensive study related to the effect of image-enhancement techniques on the performance of SfM. In section 2, I briefly present the studies already done in this context, section 3 briefly describes the reason behind the noise and low exposure present in the images, section 4 provides the approach and evaluation methodologies and section 5 discusses the results obtained from applying different image enhancement techniques on the publicly available dataset.

2 Literature Survey

In general, the images taken for 3D reconstruction are photographed with special attention of not introducing any unwanted interference or degradation in

the digital image. However, even with utmost care, certain artefacts can appear in an image. Such artefacts can be dealt with specific image pre-processing methods before passing it to an SfM pipeline. In the study presented (Ballabeni et al., 2015), a method was proposed to address the shortcomings of SIFT algorithm due to changes in image luminosity, low contrast blobs in the image. The presented method, comprised of passing an image through the pre-processing pipeline of colour balancing followed by image denoising, grey scale conversion and Wallis filter. This method can be particularly useful in image set of architectural sites, where illumination is often non-uniform. Similarly, in (Kontogianni et al., 2015) effects of using HDR images on the performance of SfM algorithm was presented. The standard images do suffer from the problem of high-low illumination patches in the images. This paper describes the effect of using the standard image and HDR image on the performance of SfM. The results showed an increase in the number of key features increases in the case of toned HDR images in comparison to the standard image.

The use of image pre-processing on SfM has been studied in some details for the cases of Architectural scenes and build environment with non-uniform lighting. However, there hasn't been much research on the general image datasets, such as images available on image-hosting websites. Since the collection of images on these platforms have different sources, hence, the quality of image available also varies widely. In these cases, the effect of image pre-processing techniques such as de-noising and contrast enhancement becomes very useful. The image de-noising and contrast enhancement algorithms can have a profound effect on the performance of SfM pipeline.

There are many basic as well sophisticated methods available for image de-noising and contrast enhancement. In the work presented by (Tomasi and Manduchi, 1998), Bilateral Filters were discussed which had the basic characteristic similar to Gaussian Blurring but are much more adept at conserving structural information[reference]. Similarly, in the variation presented by (Chambolle, 2004) of the original Total Variation Filter (Rudin et al., 1992), the algorithm is efficient in de-noising the image without removing the boundaries of the structure. However, the non-local mean filter (Buades et al., 2005) consider the spatial relationship between pixels, instead, the similarity between different patches and the patch around the target pixel is computed to determine the weighted average of the most similar pixels(inversely proportional to the distance between

its patch and the reference patch of the target pixel). The wavelet-based de-noising presented by (Chang et al., 2000), is mainly a wavelet domain based de-noising method. The method proposes the use of an adaptive threshold for wavelet thresholding the images using Bayes Shrink, which is an adaptive thresholding method that computes separate thresholds for each wavelet sub-band as described. Wiener filter (Orieux et al., 2010) is a Linear-time invariant filter which statistically approximates the output signal by removing noise through minimization of mean square error between the degraded image and the desired image. However, the Wiener filters are unable to reconstruct the signal degraded by the noise but can only suppress them. It is mainly useful if the noise present in the image is an additive type. We can evaluate its effect in images with Gaussian noise.

Similarly, for contrast enhancement novel methods have been proposed which can improve the quality of an image corrupted due to low exposure or non-uniform illumination. The histogram equalization is a technique for adjusting image intensities to enhance contrast level of the input image. Since HE normally changes the brightness of the input image significantly, it makes some of the uniform regions of the output image become saturated with very bright or very dark intensities. In paper proposed by (Ibrahim and Kong, 2007), a new method known as brightness preserving dynamic histogram equalization (BPDHE) was presented. The paper which is largely an extension to HE proposes a method which can produce the output image with the mean intensity almost equal to the mean intensity of the input. The proposed method achieves this by detecting the location of local maximums from the smoothed histogram and further mapping each part to a new dynamic range. Each of these partitions is equalized independently before normalizing the overall image brightness. The normalization is done to reduce the effect of the change in dynamic range on the brightness.

The Contrast Limited Adaptive Threshold Equalization (Pisano et al., 1998), also uses Adaptive Histogram Equalization, however, this method employs clipping limit to suppress(not enhance) the noise during contrast enhancement. This method is especially useful in low-light, hazy images. Another method very useful in low-light condition has been proposed by (Dong et al., 2011). This algorithm proposes inverting the input image and applying image de-hazing algorithms to the inverted image. This algorithm is highly useful in very low-light images and is computationally faster than BPDHE. The use of logarithmic correction is also beneficial in

enhancing the contrast of an image, especially when a large number of pixels are clustered in specific pixel intensities. This function transforms the input image pixel-wise according to the equation

$$\text{output} = \text{gain} * \log(1 + \mathbf{I})$$

after scaling each pixel to the range 0 to 1. The value of gain determines the shifting of the histogram. The other variation of gamma correction is by using Sigmoidal Function (Braun and Fairchild, 1999). This function transforms the input image pixel-wise according to the equation

$$\text{output} = 1 / (1 + \exp^{*(\text{gain} * (\text{cutoff} - \mathbf{I}))})$$

after scaling each pixel to the range 0 to 1.

3 IMAGE ARTIFACTS

An image artefact (Hornak, 1996) is any feature which falsely appears in an image but was not present in the original imaged object. Such artefacts can be seen in images with motion blur, noise, fringes or rings due to the aberration in the lens. The noise present in the image can also be classified as an artefact, however, there is a slight difference in image artefacts and noise. The noise may obscure the features in an image, while artefact may appear falsely as a feature in an image[5]. In spite of being different than one another in few aspects, both noise and artefacts in images have a significant effect on 3D reconstruction from a set of images. The first stage of an SFM pipeline is key-point detection and matching. However, the presence of such artefacts/noise in the input image can render the best key-point detection and matching algorithms inefficient. For example, the blocking artefacts in a compressed image can be detected as a feature although it doesn't contribute towards the scene reconstruction. One such example of blocking artefacts can be seen in the JPEG compressed image of Dataset(Figure 4):

However, it should be noted that image pre-processing can also have undesired effect, since it can distorts or changes the true nature of the raw image used for SFM. This can have influence on the camera pose estimation from the given image dataset. In this paper, we will weigh the effect of image-preprocessing in terms of error in camera pose estimation as compared to the ground truth, sparse and dense cloud and its qualitative aspects.



Figure 4: Blocking artifacts due to JPEG compression

3.1 Image De-noising

Image de-noising is the process of removing undesired signals from the image, while at the same time trying to preserve details and key features. It is worth mentioning, that different type of noise, such as Gaussian Noise, Grain noise, or salt and pepper noise are introduced in the image during the process of image capture. The primary cause of noise in an image is the random distribution of photons as they reach the sensor. The amount of noise present in the image is inversely related to the amount of light reaching the sensor. The fewer photons reaching the sensor, the more apparent their random distribution will be and hence give rise to random noise.

For a given exposure time, in presence of low light conditions, fewer photons per pixel are exposed to light. Since photons are distributed randomly, some pixels may receive more photons in comparison to other pixels. This uneven distribution is the specific cause of photon shot noise which follows a Poisson distribution. Specifically, in low light conditions, where the noise is largely due to low photon counts, a Poisson model of noise can be more appropriate than a Gaussian model (Szeliski, 2010).

Further, in low light conditions, the exposer time increases which introduces thermal noise artefacts in the image. Such noise follows a Gaussian distribution. Grain type noise is the consequence of using a high ISO setting on the digital SLR camera and is also termed as "grain". In images having this type of image noise, a patch of an image which should be smooth have a speckled appearance making it appear as grained.

3.2 Low Contrast

The images used for 3D reconstruction of Archaeological buildings or outdoor structures, in general, can be influenced by low light conditions. Some settings such as museums, art galleries, and poorly-lit store shelves effect the picture adversely due to uneven

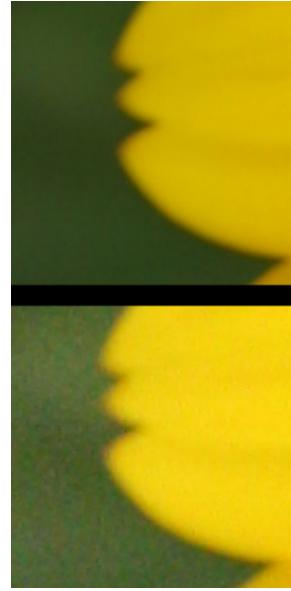


Figure 5: Image noise due to high ISO

or uni-directional lightning. As a result, some of the most salient features may not appear as desired. By redistributing the areas of intense brightness over the entire image, features that might normally be clouded can appear and protrude as we expect of features. The important reason for treating low contrast or unequal illumination as an influential artefact for SfM is the key-point detection phase in the SfM pipeline. According to the original paper [sift], after determining the potential key points, further processing is done to get more accurate results. By using the Taylor series expansion, a more accurate location of extrema is computed, but if the intensity at this extrema is below a threshold (Contrast threshold), it is rejected.

In computing the Difference of Gaussian, the ratio is used to discard the edges (similar to Harris Corner), known as edge threshold. Hence, low-contrast key points are rejected. This is efficient in removing low-interest key points and outliers, however, it can affect the SfM pipeline adversely. In images with low contrast or uneven illumination, certain key points can be rejected and hence will not be included in the Sparse Construction.

4 Methodology

To study the effect of Image pre-processing on SfM, the original dataset has been passed through different image enhancement algorithms before passing it to the SfM pipeline. However, generally, the image datasets publically available doesn't have such arte-

facts present. Hence, the noise has been artificially introduced in the images of Fountain_dense data-set (Strecha et al., 2008) by using ImageMagick software. In the presented result three different type of noise has been introduced in the images

1. Gaussian
2. Shot Noise (Poisson Distribution)
3. Laplace noise.

The Pavillion dataset hasn't been modified and the original images have been used to observe the effects of Image de-noising. For studying the effects of Image denoising on SfM, following algorithms have been used:

- Gaussian Filter
- Bilateral Filter
- Median Filter
- Wavelet Filter
- Total Variation filter
- Tri-state Median
- Wiener Filter

Similarly, in the Herz-Jeus-K7(Strecha et al., 2008), Rathaus and Semper Dataset (Strecha et al., 2006) (Strecha et al., 2004), the contrast of original dataset have been reduced artificially using GIMP photo editor. For studying the effects of Contrast Enhancement on SfM, following algorithms have been used:

- Brightness preserving dynamic histogram equalization (BPDHE)
- Contrast Limited Adaptive Histogram Equalization.
- Fast Efficient Algorithm for Enhancement of Low Lighting Video.
- Logarithmic Gama Level Correction.
- Contrast Adjustment using Sigmoid Function.
- Histogram Equalization on l-channel of LAB color space.

4.1 Structure From Motion

The already pre-processed image is passed to the SfM software "Photoscan Pro" which computes SfM and calculates the Re-projection Error and saves the camera Transformation matrix.

4.2 Evaluation Method

For the evaluation purpose, reference camera's Extrinsic matrix and estimated camera's extrinsic matrix are computed. The error in camera pose estimation is calculated by defining a distance metric between the estimated camera position C_i and the reference camera position C_r . However, the distance cannot be simply calculated by subtracting translation vectors of the C_i and C_r . This is because the rotation vector of both the camera frames are different and are not in a same plane. If the extrinsic matrix can be represented as $[R|t]$, taking the Euclidean distance between t_1 and t_2 will not give the baseline distance metric. This is because t is the position of the world origin on camera coordinates, and the reference frame can be different for each camera. Hence, subtracting vectors with different reference frames will be an invalid operation. Since, the estimated camera position and ground truth are not in same position. Hence, the recovered pose is transformed by applying Rotation, Scaling and then translation. Suppose, the reference camera pose is in coordinate system "A" and estimated camera pose is in coordinate system "B". To compute the baseline distance between the estimated and reference pose, a transformation is required from coordinate system B to coordinate system A. The transformation matrix "T" is calculate by using following relation:

$$A = T \cdot B$$

The rotation, scaling and translation matrix is calculated further from the Transformation matrix. The transformation matrix is represented as below(last row is for homogeneous coordinate.)

$$\begin{vmatrix} a & b & c & d \\ e & f & g & h \\ i & j & k & l \\ 0 & 0 & 0 & 1 \end{vmatrix}$$

Scaling in X,Y,Z direction is calculated as below:

$$S_x = \sqrt{a^2 + e^2 + i^2}$$

$$S_y = \sqrt{b^2 + f^2 + j^2}$$

$$S_z = \sqrt{c^2 + g^2 + k^2}$$

Further the rotation matrix is computed as below:

$$\begin{vmatrix} a/S_x & b/S_y & c/S_z \\ e/S_x & f/S_y & g/S_z \\ i/S_x & j/S_y & k/S_z \end{vmatrix}$$

The translation matrix is equal to the last column of the transformation matrix. After obtaining the Rotation(R), Scaling(S), Translation Matrix(T) and B_old is the coordinate system of the estimated camera pose, the the estimated camera pose in reference coordinate system is calculated as below

$$\begin{aligned} B_{\text{new}} &= R \cdot B_{\text{old}} \\ B_{\text{final}} &= S \cdot B_{\text{new}} + T \end{aligned}$$

The error in the estimated pose and ground truth is presented as root mean square error for all the registered cameras.

5 Result

5.1 Image De-noising

5.1.1 Fountain_Dense Dataset

(Strecha et al., 2008) The 3D reconstruction on the image dataset with three different type of noises was completed through Photoscan Pro(AgiSoft) software. In all the four presented results, the Gaussian and bilateral provides lower camera pose error. The wavelet based filtering method is much efficient in reducing the noise but it also removes the high-frequency details in the image. The non-local mean filter produces the thinnest sparse point cloud with higher re-projection error. The error in camera pose estimation is also higher for non-local mean blurring. The main reason behind it is that these filter overly smooths the image, hence, even removing the structural features. The results in lower number of key-point detected. Gaussian Filter still performs better than all the other filtering methods, because the Fourier Transform of a Gaussian is also a Gaussian, the Gaussian filter does not have a sharp cutoff at the edges of the pass band frequency beyond which all higher frequencies are removed. Instead, it has smooth decline that becomes ever lower as the frequency increases. This means that it act as a low pass filter, but also allow in higher frequency details to a certain extent. Similar reason can be given for the performance of bilateral filter.

Fountain_Dense dataset with Gaussian Noise

The Fountain_dense dataset used in this iteration has be induced with a Gaussian noise using ImageMagick with an attenuation factor of 0.5. The resolution and color depth of the original image dataset has been maintained.



Figure 6: Images after filtering with different smoothing filters

Dataset : Fountain_Dense with Gaussian Noise			
Filter	Sparse Point	Dense Point	Reprojection Error
Bilateral	29,994	626,827	0.6029
Gaussian	35,335	622,808	0.5854
Non-local Mean	19,639	618,646	0.7033
Total Variation	29,036	644,739	0.6439
Wavelet denoising	27,093	617,349	0.6428
original	12820	727848	0.742

Figure 7: SFM Result and Re-projection Error

Dataset : Fountain_Dense with Gaussian Noise					
Bilateral	Gaussian	Non-local Mean	Total Variation	Wavelet denoising	original
0.142329	0.147952	0.1524	0.1769	0.008676	0.137137

Figure 8: Error in Camera-Pose Estimation

Fountain_Dense dataset with Poisson Noise

The Fountain_dense dataset used in this iteration has been induced with a Poisson noise using ImageMagick with an attenuation factor of 1.5. The resolution and color depth of the original image dataset has been maintained. This type of noise has been introduced to imitate the effect of Shot noise.

Dataset : Fountain_Dense with Poisson			
Filter	Sparse Point	Dense Point	Reprojection Error
Bilateral	6,694	560,151	1.32
Gaussian	9,219	571,424	1.092
Non-local Mean	6,485	560,300	1.342
Total Variation	9,070	577,058	1.769
Wavelet denoising	6,787	592,930	1.658
original	6532	556033	1.357

Figure 9: SFM Result and Re-projection Error

Dataset : Fountain_Dense with Poisson Noise					
Bilateral	Gaussian	Non-local Mean	Total Variation	Wavelet denoising	original
0.000042	0.006069	0.038943	0.082234	0.080522	0.257045

Figure 10: Error in Camera-Pose Estimation

Fountain_Dense dataset with Laplace Noise

The Fountain_dense dataset used in this iteration has been induced with a Laplace model noise using ImageMagick with an attenuation factor of 0.5. The resolution and color depth of the original image dataset has been maintained.

Dataset : Fountain_Dense with Laplacian Noise			
Filter	Sparse Point	Dense Point	Reprojection Error
Bilateral	11,102	761,628	0.5965
Gaussian	10,798	751,549	0.604
Non-local Mean	12,344	723,203	0.734
Total Variation	10,996	752,433	0.665
Wavelet denoising	11,622	756182	0.6435
original	11746	754859	0.615

Figure 11: SFM Result and Re-projection Error

Dataset : Fountain_Dense with Laplacian Noise					
Bilateral	Gaussian	Non-local Mean	Total Variation	Wavelet denoising	original
0.086154	0.067425	0.125744	0.000105	0.054871	0.000739

Figure 12: Error in Camera-Pose Estimation

Fountain_Dense dataset with Grain Noise

This type of noise is present due to high ISO settings/amplification of the camera. This type of noise can easily be verified by zooming into the image and observing grain like texture rather than a smooth texture.

5.1.2 Pavilion Dataset[12]

This dataset contains the images of Chair taken in different light conditions. The two conditions used in the presented results are of night time and mid-day time. The image set of night light and day time are shown in figure below:

Chair Sequence(Low Light)

The images taken during night time are infected with shot noise due to high exposure time as well as the

Dataset : Fountain_Dense with Grain Noise			
Filter	Sparse Point	Dense Point	REPRo
Bilateral	49,841	655,710	0.3805
Gaussian	51,385	682,695	0.3881
Non-local Mean	11,690	731,121	0.6484
Total Variation	10,334	766,317	0.5404
Wavelet denoising	9,457	797,993	0.3798

Figure 13: SFM Result and Re-projection Error

Dataset : Fountain_Dense with Grained Noise					
Bilateral	Gaussian	Non-local Mean	Total Variation	Wavelet denoising	original
0.042125	0.00537	0.041973	0.160358	0.105028	na

Figure 14: Error in Camera-Pose Estimation



Figure 15: Image taken in night



Figure 16: Image taken during day

presence of salt and pepper noise, which can be observed as presence of dead pixels(bright pixel in dark region and vice versa in all the three color channels). This images data is preprocessed with image de-noising filter as used in previous dataset(also including Wiener filter and tri-state median filter). The result is presented below:

Chair Sequence(Day light)

The images taken during day time are infected with additive noise. The result is presented below:

The effect of smoothing on this dataset doesn't

Dataset : Pavillion Chair(Night)				
Filter	Sparse Point	Dense Point	Reprojection Error	
Bilateral	5,631	112,014	0.806	
Gaussian	13,108	126,640	0.753	
Non-local Mean	4,797	75,539	0.892	
Total Variation	7,690	90,671	0.864	
Wavelet denoising	5,357	91716	0.874	
Tristate Median	3,971	74,880	0.9132	
Weiner	8,068	118,297	0.776	
original	5421	116182	0.841	

Figure 17: SFM Result and Re-projection Error

Dataset : Pavillion Chair(Night)								
Bilateral	Gaussian	Non-local Mean	Total Variation	Wavelet denoising	Tristate Median	Weiner	original	
0.000154	0.000063	0.000917	0.000739	0.000373	0.002267	0.000094	0.000067	
92	92	91	91	92	78	92	92	

Figure 18: Error in Camera-Pose Estimation

Dataset : Pavillion Chair(Day)								
Filter	Sparse Point	Dense Point	Reprojection Error					
Bilateral	6,927	123,588	0.489					
Gaussian	13,732	136,328	0.428					
Non-local Mean	5,840	121,590	0.549					
Total Variation	7,625	101,568	0.534					
Wavelet denoising	7,010	121,825	0.5134					
Tristate Median	6,838	119,028	0.512					
Weiner	9,288	132823	0.464					
original	6544	122200	0.514					

Figure 19: SFM Result and Re-projection Error

Dataset : Pavillion Chair(Day)								
Bilateral	Gaussian	Non-local Mean	Total Variation	Wavelet denoising	Tristate Median	Weiner	original	
0.000195	0.000095	0.000204	0.000203	0.000107	0.000154	0.000058	0.000081	
100	100	100	100	100	100	100	100	

Figure 20: Error in Camera-Pose Estimation

have a positive influence on the result of SFM. The Gaussian filter in particular provides the worst result. This is because of the over smoothing of the images, hence, it filters out even the structural details from the picture which leads false matching and hence incorrect pose estimation. However, if the blurred image is again processed through with contrast enhancement algorithm such as Clahe, the error in pose estimation reduces. This is more desirable if the original image is on a low contrast side.

Building Sequence(Low-light)

This dataset consist of the image of building taken in extremely low light condition and hence infected with



Figure 21: Image after over-smoothing, blurred details

shot noise due to high exposure time. It should also be noted that due to low light conditions, the sensors are exposed for a longer period of time and hence, along with Shot noise, such images also have additive Gaussian noise(due to heating of sensors). In this dataset, the number of cameras registered have been shown along with the RMS error. In this dataset also, the Gaussian Blur proves to be more efficient method of improving SFM results in low light noisy conditions.

Dataset : Pavillion Building(Night)								
Filter	Sparse Point	Dense Point	Reprojection Error					
Bilateral	1,463	23,346	0.631					
Gaussian	2,317	29,298	0.562					
Non-local Mean	1,291	18,290	0.649					
Total Variation	1,483	18,886	0.693					
Wavelet denoising	1,304	20,047	0.574					
Tristate Median	1,664	20,769	0.646					
Weiner	1,987	24038	0.5885					
original	1,632	18,734	0.678					

Figure 22: SFM Result and Re-projection Error

Dataset : Pavillion Building(Night)								
Bilateral	Gaussian	Non-local Mean	Total Variation	Wavelet denoising	Tristate Median	Weiner	original	
0.000137	0	0.000116	0.000149	0.000044	0.000639	0.00006	0.000214	
36	30	36	36	36	38	45	36	

Figure 23: Error in Camera-Pose Estimation

Building Sequence(Day-light)

This image sequence consist of images taken in broad daylight.

As in the chair dataset in daylight, the Gaussian blurring proves inefficient. However, other filtering methods provides a lesser inaccurate pose estimation.



Figure 24: Image from dataset Pavilion Dataset(Day light)

Dataset : Pavillion Building(Day)			
Filter	Sparse Point	Dense Point	Reprojection Error
Bilateral	1,528	26,498	0.4834
Gaussian	7,137	68,082	0.408
Non-local Mean	1,444	28,864	0.604
Total Variation	2,576	29,595	0.542
Wavelet denoising	1,601	30,774	0.504
Tristate Median	1,749	33,231	0.447
Weiner	4,661	58367	0.447
original	1528	32122	0.473

Figure 25: SFM Result and Re-projection Error

Dataset : Pavillion Building(Day)							
Bilateral	Gaussian	Non-local Mean	Total Variation	Wavelet denoising	Tristate Median	Weiner	original
0	0.000723	0.000011	0	0	0.000003	0.000327	0.000002
35	94	30	30	29	30	94	43

Figure 26: Error in Camera-Pose Estimation

5.2 Observations - Image Denoising

The noticeable effect of pre-processing as observed from above results can be summarized as follows:

1. The Gaussian filter performs better than all the presented filters in presence of Shot noise(which follows Poisson distribution).
2. This can be attributed to the fact that for larger variance, the Gaussian distribution approximates Poisson Distribution[13].
3. The bilateral filter, non-local mean filter and Total Variation filter also provides improvement, but in case of Shot noise the improvement is not satisfactory in comparison to Gaussian filter.
4. The re-projection error increases but that can be attributed to the increase in the number of Tie points. On using lesser number of images, the re-projection error decreases without improving the output of SFM.

5. The Wiener filter also provides better results in low light conditions and improves the pose estimation error. This can be attributed to the fact that Wiener filter essentially outputs a statistical approximation of required output by filtering a noisy spectrum with additive noise. These images are taken in low light and due to sensor heating contain additive noise.

6. Images taken in sufficient lighting condition, may have uniform noise distribution and further image de-noising techniques(presented here) doesn't improve the output of SFM.

5.3 Contrast Enhancement

In this section the influence of contrast enhancement on sparse point reconstruction and camera pose estimation has been shown. The first two dataset, i.e., Herz-Jesu-K7 and Semper, the image contrast has been lowered using GIMP image editor, whereas the MVE Kermit hasn't been tempered.

5.3.1 Dataset - Herz-Jesu-K7

As can be seen from the histogram presented below(Figure 27), after reducing the contrast of the original image with GIMP(by factor of 40), the image has no representation in the lighter and darker tones of the histogram. Histogram equalization on the LAB color space marginally improves the spreading of the intensity distribution. The gamma correction improves the spreading of the histogram but also introduces clipping on the lighter tone in red-channel of the image. Since the value of alpha in gamma correction is 0.9, hence, the histogram doesn't shifts entirely to darker tone. In comparison, CLAHE performs better than other techniques, since it prevents brightness saturation and also provides a better spread of the histogram.

5.3.2 Dataset - Semper

In this dataset similar to Herz-Jesu-K7, the contrast was reduced using GIMP by a factor of 40. After contrast reduction the image has a narrow histogram with a peak at one particular intensity(Figure 30). The histogram equalization proves inefficient in this case. Since in HE, the intensity values are mapped so as pixels are more distributed over all the available intensity. However, a single peak or large clutter of pixel on a single pixel intensity makes HE vulnerable. BPDHE algorithm spreads the histogram, but due to cluttering of pixels around single intensity, the spread is not uniform on both sides of histogram. However, CLAHE and Logarithmic operator

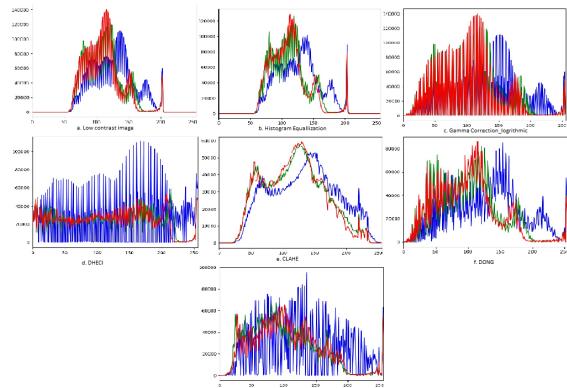


Figure 27: Histogram after different contrast enhancement methods

DATASET - Herz-Jesu-K7			
ALGORITHM	TIE POINTS	DENSE POINTS	REPROJECTION ERROR
CLAHE	8,408	647,618	0.7824
BPDHE	8,594	647,633	0.8184
DHECI	8,719	637,485	0.7771
DONG	8,651	654,404	0.7584
GAMMA_LOG	8,755	654,660	0.7701
HE	8,686	653,091	0.7832
No Processing	8,766	652,500	0.7647

Figure 28: SFM Result and Re-projection Error

Camera Pose RMSE (DATASET - Herz-Jesu-K7)							
Original	BPDHE	CLAHE	DHECI	DONG	Gamm_Log	HE	Gamma_Sigmoid
0.0386	0.0610	0.0388	0.1703	0.0607	0.0178	0.0811	0.0212

Figure 29: Error in Camera-Pose Estimation

provides much better improvement in the histogram. The logarithmic operator has better histogram spread than CLAHE because the logarithmic operator enhances the low intensity pixel values, while compressing high intensity values into a relatively small pixel range. The improvement can also be seen as slight improvement in the pose estimation from the unprocessed image.

5.3.3 Dataset : Kermit MVE

The kermit dataset was used as original without any editing. The image in the original dataset are underexposed as can be seen from the histogram. The logarithmic gamma correction performs better in comparison to previous dataset, as it also re-scales the intensity of the pixels and hence number of pixels with low intensity is increased. The BPDHE algorithm doesn't perform comparable to CLAHE or gamma correction, as it doesn't change the overall brightness of the original image. However, the error in camera pose estima-

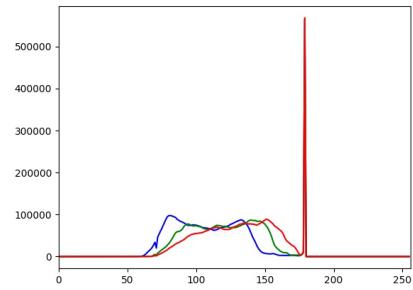


Figure 30: Histogram of reduced contrast Image

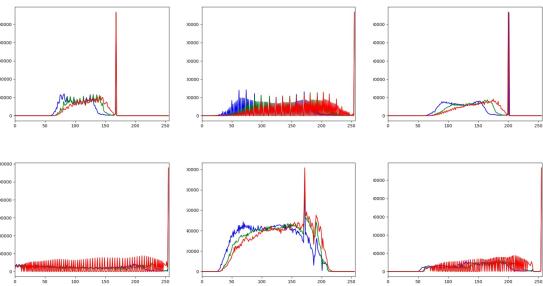


Figure 31: Histogram after Contrast Enhancement, from top to bottom and left to right: HE, Gamma Logarithmic, DONG, DHECI, CLAHE, BPDHE

Dataset - Semper			
ALGORITHM	TIE POINTS	DENSE POINTS	REPROJECTION ERROR
CLAHE	4,597	297,833	0.293
BPDHE	4,732	299,580	0.301
DHECI	4,804	295,460	0.313
DONG	4,645	297,027	0.287
GAMMA_LOG	4,587	302,370	0.293
HE	4,724	298,911	0.298
No Processing	4,728	299,184	0.288

Figure 32: SFM Result and Re-projection Error

Dataset - Semper						
BPDHE	CLAHE	DHECI	DONG	GAMMA_LOG	HE	Original
0.005391	0.011	0.006	0.006	0.012	0.015	0.009

Figure 33: Error in Camera-Pose Estimation

tion is minimum in case of DONG, this maybe because the algorithm doesn't changes the original image in terms of brightness and the shape of histogram but only removes the clipping in the blue channel in the darker tone of histogram.

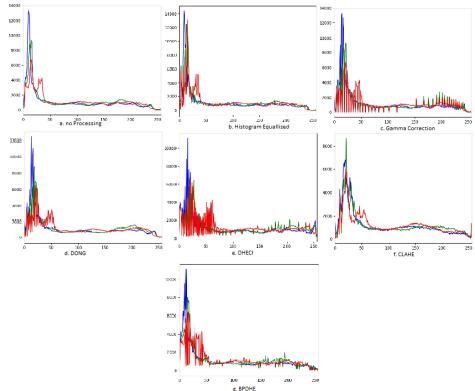


Figure 34: Histogram after different contrast enhancement methods

Dataset : Kermit MVE			
ALGORITHM	TIE POINTS	DENSE POINTS	REPROJECTION ERROR
CLAHE	364	15,754	0.656
BPDHE	397	15,326	0.69
DHECI	386	14,448	0.695
DONG	390	14,278	0.7991
GAMMA_LOG	409	15,345	0.702
HE	389	15,717	0.693
No Processing	392	15,026	0.744

Figure 35: SFM Result and Re-projection Error

Dataset : Kermit MVE						
BPDHE	CLAHE	DHECI	DONG	Gamm_log	HE	Original
0.230	0.214	0.212	0.319	0.304	0.216	0.290669

Figure 36: Error in Camera-Pose Estimation

5.3.4 Dataset: Rathaus[14]

The Rathaus dataset was used as original without any editing. The image in the original dataset is bit underexposed as can be seen from the histogram. The original images are slightly towards the lighter shade of histogram.

After the contrast enhancement the histogram improves and has more spread, especially in case of CLAHE and BPDHE. However, the pose estimation doesn't show any improvement and the error in pose estimation is more or less similar to the unprocessed dataset.

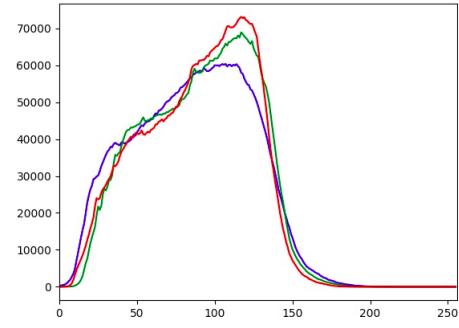


Figure 37: Histogram of original image

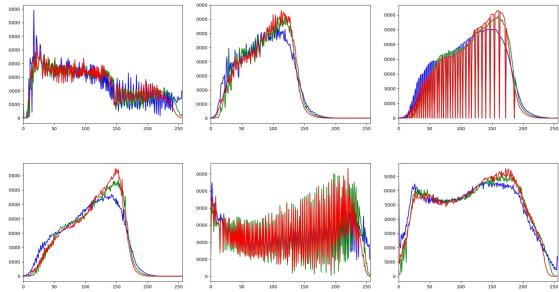


Figure 38: Image after Contrast Enhancement, from top to bottom and left to right: BPDHE, HE, Gamma Logarithmic, DONG, DHECI, CLAHE

Dataset - Rathaus			
ALGORITHM	TIE POINTS	DENSE POINTS	REPROJECTION ERROR
CLAHE	18,319	703,344	0.408
BPDHE	16,814	700,357	0.500
DHECI	17,797	685,998	0.442
DONG	19,283	705,336	0.409
GAMMA_LOG	19,485	706,801	0.414
HE	19,554	711,632	0.421
No Processing	19,761	713,341	0.414

Figure 39: SFM Result and Re-projection Error

DATASET - Rathaus							
CAMERA ID	BPDHE	CLAHE	DHECI	DONG	Gamm_log	HE	Original
Camera 0	71.435	71.436	71.435	71.436	71.436	71.436	71.435
Camera 1	72.453	72.453	72.453	72.453	72.452	72.452	72.452
Camera 2	52.258	52.254	52.258	52.253	52.255	52.254	52.257
Camera 3	21.271	21.261	21.276	21.256	21.261	21.260	21.273
Camera 4	2.766	2.769	2.763	2.772	2.769	2.770	2.764
Camera 5	30.464	30.472	30.459	30.479	30.473	30.474	30.462
Camera 6	46.896	46.908	46.888	46.917	46.908	46.910	46.893

Figure 40: Error in Camera-Pose Estimation

5.3.5 Dataset: Pavillion(ECC, 2016)

Outer Sequence-Building(Night)

This dataset contains image sequence of building taken during low-light conditions. As previously

stated in the comparison of used contrast enhancement techniques, we can observe that contrast enhancement improves the output of SFM in terms of camera pose estimation. It is also worth noting that as expected DONG out-performs all the other presented techniques. This largely is due to the fact that the large part of this dataset has very low-light images. These images occupy the darker tone of the histogram and the contrast overall is pretty low. Although, other techniques also show improvement over the non-processed image, however, DONG, owing to its characteristic of improving the dark/lowlight images proves more efficient.

Dataset - Building(Night)		
ALGORITHM	TIE POINTS	DENSE POINTS
CLAHE	1,374	18,489
BPDHE	1,695	26,745
DHECI	1,212	24,814
DOONG(0.35)	1,495	21,108
DONG(0.7)	1,365	19,470
GAMMA_LOG	1,175	17,376
HE	1,574	19,104

Figure 41: SFM Result

Dataset : Building Night							
BPDHE	CLAHE	DHECI	DONG(0.35)	Gamm_log	HE	Original	Dong(0.7)
0.000001	0.000177	0.000182	0.000357	0.0002	0.000394	0.000214	0.000147
48	36	30	55	44	50	36	56

Figure 42: Error in Camera-Pose Estimation, second row show the number of registered cameras.

Dataset: Chair Sequence-Chair(Day)

This dataset consist of the images of chair taken during the day-time(natural lighting). However, the histogram is largely towards the lighter tone with clipping in blue and green channel. In this dataset, a generic histogram equalization provides more accurate pose estimation than other CE methods. Since, the number of pixels representing the lighter tone are larger in number, hence, a re-distribution of pixel on simple interpolation provides acceptable results. Similarly, we can observe that the logarithmic histogram equalization also provides more accurate result than other sophisticated methods. However, since the logarithmic histogram equalization provides much better result in cases where most pixel represents a particular intensity, hence, in this case it is not comparable to HE as the histogram is spread in the lighter tone of histogram rather than being cluttered in fewer intensities.

Dataset - Chair Day		
ALGORITHM	TIE POINTS	DENSE POINTS
CLAHE	6,845	117,380
BPDHE	6,511	121,764
DHECI	6750	121601
DONG	6,535	119,918
GAMMA_LOG	6,662	119,895
HE	6,601	122,528

Figure 43: SFM Result

Dataset : Chair Day						
BPDHE	CLAHE	DHECI	DONG	Gamm_log	HE	Original
0.000174	0.000051	0.000123	0.000111	0.000226	0.000062	0.000081
99	92	100	100	100	100	100

Figure 44: Error in Camera-Pose Estimation, second row show the number of registered cameras.

Dataset: Chair Sequence-Chair(Sunset)

This is the same image sequence as the previously discussed dataset, however, the lighting condition is similar to that of sunset. In slightly low light condition, the simple HE fails to provide an accurate result. In this dataset, DONG performs better than the other methods but the number of registered cameras are lower than BPDHE. This is because DONG over brightens the images, which lowers the contrast of theses images.

Dataset - Chair Sunset		
ALGORITHM	TIE POINTS	DENSE POINTS
CLAHE	5577	96,295
BPDHE	5,877	106,487
DHECI	5,864	110,299
DONG	5,665	112,619
GAMMA_LOG	5,728	123,299
HE	5,453	115,845

Figure 45: SFM Result

Dataset : Chair Sunset						
BPDHE	CLAHE	DHECI	DONG	Gamm_log	HE	Original
0.000955	0.000298	0.000141	0.000074	0.000145	0.002731	0.000067
100	92	92	92	92	100	92

Figure 46: Error in Camera-Pose Estimation, second row show the number of registered cameras.

Dataset: Chair Sequence-Chair(Night)

The original image dataset performs better than the pre-processed images. In this dataset also, DONG performs better on images which are not registered in other methods. DHECI provide better result than

DONG, but fails to provide same accuracy as original dataset. The main reason behind this can be the images which get registered in DHECI but fails to register in original dataset. Hence, the overall pose error increases for DHECI.

Dataset - Chair (Night)		
ALGORITHM	TIE POINTS	DENSE POINTS
CLAHE	4,192	75,246
BPDHE	4,724	104,003
DHECI	5,517	106,443
DOONG(0.35)	4,146	96,574
GAMMA_LOG	4,710	83,331
HE	3,032	52,941

Figure 47: SFM Result

Dataset : Chair Night						
BPDHE	CLAHE	DHECI	DONG	Gamm_log	HE	Original
0.000044	0.000344	0.000552	0.000001	0.001085	0.000018	0.000067

Figure 48: Error in Camera-Pose Estimation, second row show the number of registered cameras.

5.4 Observation

1. Contrast enhancement can improve the result of SFM, since, the matching algorithm filters out feature points on the basis of contrast threshold.
2. In cases when there is a cluster of the pixels around particular intensity, the logarithmic operator provides better results than other presented techniques.
3. CLAHE improves the brightness of images, but in images already having pixels in the lighter tone of the histogram, it's better to employ BPDHE.
4. DONG performs better than all other presented algorithms on low light and Hazy image.
5. In the dataset, where image sequence consists of low light as well as sufficiently lit images, DONG over improves the brightness of the image. This can be controlled by lowering the value of alpha, however, this will affect the processing of the images which are taken in very low light conditions.
6. In datasets, where the histogram is largely populated in the lighter tone, simple HE can provide a better result than other presented algorithms.

6 Conclusion and Future Work

In this project, different image enhancement algorithms were used to observe their effect on the performance of SfM. It is worth noting that image enhancement has a significant effect on the performance on the SfM. However, any generalization cannot be drawn. The majority of the dataset used during this project were artificially inflicted with noise and low exposure. Hence, obtaining the datasets with above-mentioned artefacts and then using image enhancement can provide many insightful results. Further, the combination two or more image pre-processing methods have to be tested for its effect on SfM's performance.

REFERENCES

- (2016). *Computer Vision - ECCV 2016, 14th European Conference on Computer Vision, Amsterdam, Netherlands, October 8-16, 2016, Proceedings*, Lecture Notes in Computer Science. Springer.
- Ballabeni, A., Apollonio, F., Gaiani, M., and Remondino, F. (2015). Advances in image pre-processing to improve automated 3d reconstruction. *The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, 40(5):315.
- Braun, G. J. and Fairchild, M. D. (1999). Image lightness rescaling using sigmoidal contrast enhancement functions. *Journal of Electronic Imaging*, 8(4):380–394.
- Buades, A., Coll, B., and Morel, J.-M. (2005). A non-local algorithm for image denoising. In *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, volume 2, pages 60–65. IEEE.
- Chambolle, A. (2004). An algorithm for total variation minimization and applications. *Journal of Mathematical imaging and vision*, 20(1-2):89–97.
- Chang, S. G., Yu, B., and Vetterli, M. (2000). Adaptive wavelet thresholding for image denoising and compression. *IEEE transactions on image processing*, 9(9):1532–1546.
- Dong, X., Wang, G., Pang, Y., Li, W., Wen, J., Meng, W., and Lu, Y. (2011). Fast efficient algorithm for enhancement of low lighting video. In *2011 IEEE International Conference on Multimedia and Expo*, pages 1–6. IEEE.
- Hartley, R. and Zisserman, A. (2003). *Multiple View Geometry in Computer Vision*. Cambridge University Press, New York, NY, USA, 2 edition.
- Hornak, J. (1996). *The Basics of MRI*. Rochester Institute of Technology.
- Ibrahim, H. and Kong, N. S. P. (2007). Brightness preserving dynamic histogram equalization for image contrast enhancement. *IEEE Transactions on Consumer Electronics*, 53(4):1752–1758.

- Kontogianni, G., Stathopoulou, E., Georgopoulos, A., and Doulamis, A. (2015). Hdr imaging for feature detection on detailed architectural scenes. *The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, 40(5):325.
- Orieux, F., Giovannelli, J.-F., and Rodet, T. (2010). Bayesian estimation of regularization and point spread function parameters for wiener–hunt deconvolution. *JOSA A*, 27(7):1593–1607.
- Pisano, E. D., Zong, S., Hemminger, B. M., DeLuca, M., Johnston, R. E., Muller, K., Braeuning, M. P., and Pizer, S. M. (1998). Contrast limited adaptive histogram equalization image processing to improve the detection of simulated spiculations in dense mammograms. *Journal of Digital imaging*, 11(4):193.
- Rudin, L. I., Osher, S., and Fatemi, E. (1992). Nonlinear total variation based noise removal algorithms. *Physica D: nonlinear phenomena*, 60(1-4):259–268.
- Strecha, C., Fransens, R., and Van Gool, L. (2004). Wide-baseline stereo from multiple views: a probabilistic account. In *Computer Vision and Pattern Recognition, 2004. CVPR 2004. Proceedings of the 2004 IEEE Computer Society Conference on*, volume 1, pages I–I. IEEE.
- Strecha, C., Fransens, R., and Van Gool, L. (2006). Combined depth and outlier estimation in multi-view stereo. In *Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference on*, volume 2, pages 2394–2401. IEEE.
- Strecha, C., Von Hansen, W., Van Gool, L., Fua, P., and Thoennessen, U. (2008). On benchmarking camera calibration and multi-view stereo for high resolution imagery. In *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on*, pages 1–8. Ieee.
- Szeliski, R. (2010). *Computer Vision: Algorithms and Applications*. Springer-Verlag New York, Inc., New York, NY, USA, 1st edition.
- Tomasi, C. and Manduchi, R. (1998). Bilateral filtering for gray and color images. In *Computer Vision, 1998. Sixth International Conference on*, pages 839–846. IEEE.