On Image based Enhancement for 3D Dense Reconstruction of Low Light Aerial Visual Inspected Environments

Christoforos Kanellakis, Petros Karvelis, and and George Nikolakopoulos

Robotics Group, Department of Computer, Electrical and Space Engineering, Luleå University of Technology, Luleå
{chrkan,petkar,geonik}@ltu.se

Abstract. Micro Aerial Vehicles (MAVs) have been distinguished, in the last decade, for their potential to inspect infrastructures in an active manner and provide critical information to the asset owners. Inspired by this trend, the mining industry is lately focusing to incorporate MAVs in their production cycles. Towards this direction, this article proposes a novel method to enhance 3D reconstruction of low-light environments, like underground tunnels, by using image processing. More specifically, the main idea is to enhance the low light resolution of the collected images, captured onboard an aerial platform, before inserting them to the reconstruction pipeline. The proposed method is based on the Contrast Limited Adaptive Histogram Equalization (CLAHE) algorithm that limits the noise, while amplifies the contrast of the image. The overall efficiency and improvement achieved of the novel architecture has been extensively and successfully evaluated by utilizing data sets captured from real scale underground tunnels using a quadrotor.

1 Introduction

Lately, the Micro Aerial Vehicles (MAVs) have received increased research attention within the robotics community. These platforms are mechanically simple and by providing agile navigation capabilities have increased their popularity in the society, since they are able to fly in different modes, aggressively, smoothly, hover close to a target and perform advanced maneuvers. Until now the majority of consumer grade MAVs have been directed towards the photography-cinematography industry, taking advantage of their payload capacity and stable flight characteristics. Lately, there is an increasing interest from other industries targeting autonomous infrastructure inspection in close proximity with the Region of Interest (ROI). A characteristic example of this trend is the underground mines that have the potential to deploy MAVs (Figure 1) in their operation cycles and by that to reduce the operating costs, increase the productivity, while allowing for an overall increase of the human safety in underground challenging mining conditions [1].

These attributes will have an imminent impact to change the common practices currently utilized in the mining sector such as at: 1) overall mine operation,
2) mine production and 3) safety in the operations. So far, there have been developed multiple systems, limited to remotely operate in open pit mines above ground and assist in stockpile surveying, 3D pit model build, facility monitoring, security inspection and environmental assessment of the mine sites. In all these application scenarios, multiple challenges arise before employing autonomous MAVS, mainly towards the planning and control of these vehicles, such as the narrow passages, reduced visibility due to rock falls, dust, uncertainty in localization, wind gusts and lack of proper illumination.

Within the related literature of MAVs in underground mine operations, few research efforts have been reported trying to address these challenging tasks. In [2] a visual inertial navigation framework has been proposed, to implement position tracking control of the platform. In this case, the MAV was controlled to follow obstacle free paths, while the system was experimentally evaluated in a real scale tunnel environment, simulating a coal mine, where the illumination challenge was assumed solved. In [3] a more realistic approach, compared to [2] regarding underground localization, has been performed. More specifically, a hexacopter equipped with a Visual Inertial (VI) sensor and a laser scanner was manually guided across a vertical mine shaft to collect data for post-processing. The extracted information from the measurements have been utilized to create a 3D mesh of the environment and localize the vehicle. Finally, in [4] the estimation, navigation, mapping and control capabilities for autonomous inspection of penstocks and tunnels using aerial vehicles has been studied, using IMUs, cameras and lidar sensors.

This article focuses on the perception task and aims in proposing novel methods for enhancing the 3D dense reconstructions of low light environments by using MAVs. More specifically, video data recorded on board of the MAV during the mission can be post processed to provide a detailed 3D model of the visited area. The 3D model of the area of interest provides an actual comprehensive visual and geometric information for the asset owner for further analysis, or
the mine inspectors to contextualize the location of the damages found during the inspection task, while the 3D information further facilities the evaluation of defects relative to the neighboring areas [5]. The combination of small scale and agile robotic platforms with advanced computer vision algorithms have the potential to create a powerful tool that is able to address complex tasks and provide better visual data and subsequently enabling a better decision making around the inspected infrastructure.

On the other hand, the quality of an image or a video sequence largely depends on the light conditions of the environment. For example strong light produces images with a wash out effect and weak light produce images that are not visible due to the darkness. For both of the cases, the contrast of the two images is extremely low and needs further modification in order to reveal the details of the images. These typical low light conditions occur in mines, usually due to the lack of proper illumination and in other underground or indoor dark environments, e.g. factories. In all these cases, even if it was possible and realistic to utilize a light on board of the MAV in order to illuminate the surrounding area of interest, this would have drawbacks for the mission design due to: a) the limited weight that the MAV can lift and b) the limited power supply that the MAV can support with the corresponding impact on the flight duration. A characteristic example of such a low light image capturing conditions acquired from a camera on a MAV is displayed in Figure 2.

![Fig. 2: Image of low light conditions in a mine captured by the camera on board of a MAV.](image)

The solution to the described low light capturing conditions will be to employ image processing methods in order to contrast enhance the captured image since higher contrast level images display more detailed images and color differences, when compared to the lower contrast level ones. Contrast enhancement techniques [6] have received great attention mostly because of their simplicity as also their effectiveness. Even now days in the era of Deep Learning (DL),
efforts are being made to enhance images especially in the case of low light conditions [7].

This work aligned with the vision of deploying aerial robotic platforms in the underground mine, focuses on the development of elaborate perception capabilities for MAVs flying in low light environments. The contribution of the proposed work is three-folded. Initially, this article proposes a method for 3D dense reconstruction by using a low cost aerial platform equipped with a single camera that can be considered as a consumable and easily replaceable. Secondly the paper focuses on employing image enhancement methods that are best suitable for low light enhancement methods. Finally, the third contribution stems from the fact that this work is among the few in the field that reports experimental trials on real scale tunnels, demonstrating the concept of enhancing 3D reconstruction of low light areas.

The rest of this article is organized as it follows. In Section 2 the image processing method for low light image enhancement is introduced and also the methods used for the 3D reconstruction are described for a generic clarity. In Sections 3 and 4 the data collection procedure and the results of the proposed novel method are described respectively. Finally, the concluding remarks and future work are presented in Section 5.

2 Methodology

2.1 Contrast Limited Adaptive Histogram Equalization

One of the most prominent and simple techniques like the Histogram Equalization (HE) [8] has been employed to enhance images in low light conditions. This method alters the histogram of the original image in such a way that the resulting image will have a constant histogram. Another method that is based on locally equalizing regions or blocks of the image is the Adaptive histogram Equalization (AHE) [9], [10] method which has the advantage that is adaptive to local information of the image. However both of these methods suffer from the fact that they actually enhance noise, particularly in homogeneous regions of the image, since the histogram in such regions is highly concentrated.

Contrast Limited AHE or (CLAHE) [11] overcomes such problems and the histogram of the new image is clipped in a way that the clipped pixels are re-assigned to each gray level. This is the case in homogeneous or uniform regions of the image where high peaks of the histogram are present. In this case both methods, the AHE or HE enhance the image noise, since a very narrow range of input intensities will be mapped to a wider range of output intensity values. CLAHE on the other hand will enforce a maximum on the counts of the output histogram, thus limiting the amount of contrast enhancement.

The key parameters of the CLAHE method are two: the Block Size (BS) and the Clip Limit (CL). As the threshold CL is increased, the resulting image will be brighter because larger CL will produce a flatter histogram. Finally, as the BS will be larger the dynamic
range will become larger and the contrast of the image is also increased. The
CLAHE method is composed of a number of steps that for clarity will be briefly
described in the sequel:

1. Divide the original image into a number of blocks of size $M \times N$, where $M$
and $N$ are the number of pixels in the $x$ and $y$ direction respectively, and
for each region compute:

$$N_{\text{aver}} = \frac{M \cdot N}{N_{\text{gray}}},$$

(1)

where $N_{\text{gray}}$ is the number of gray levels in the region.

2. The actual clip limit $N_{\text{CL}}$ will be computed as:

$$N_{\text{CL}} = N_{\text{clip}} \cdot N_{\text{aver}},$$

(2)

where $N_{\text{clip}}$ is the normalized $CL$ in the range of $[0, 1]$.

3. The total number of clipped pixels will be defined as $S_{\text{clipped}}$ with the average
number of pixels ($N_{\text{anp}}$) having to distribute uniformly to all gray levels is
defined as:

$$N_{\text{anp}} = \frac{N_{\text{anp}}}{N_{\text{gray}}}. $$

(3)

The contrast limited histogram of the contextual region can be calculated
with the following set of rules:

$$H_{\text{clipped}}(i) = \begin{cases} 
N_{\text{CL}}, & \text{if } (i) > N_{\text{CL}} \\
N_{\text{CL}}, & \text{if } H(i) > N_{\text{CL}} + N_{\text{anp}} \\
H(i) + N_{\text{CL}}, & \text{otherwise}
\end{cases} $$

(4)

where $H$ and $H_{\text{clipped}}$ are the initial histogram and clipped histogram re-
spectively.

4. Redistribute the remaining pixels $N_{\text{rp}}$ by searching from the minimum to
the maximum gray level with the following step: $\text{step} = \frac{N_{\text{gray}}}{N_{\text{rp}}}$. One pixel
will be distributed to the gray level if the number of pixels in the gray level
is less than $N_{\text{CL}}$. The procedure is repeated until all remaining pixels are all
distributed to the new histogram.

5. Apply Histogram Equalization (HE) to the resulting histogram $H_{\text{clipped}}$.

6. In order to reduce some abruptly changes in the resulting histogram apply
linear contrast stretch [8].

An example of the CLAHE method applied to an image of the tunnel is
shown in the following Figure 3.

2.2 Tunnel Reconstruction

The application scenario considered in this work targets the inspection of under-
ground mine tunnels by utilizing aerial robotic platforms. The final outcome
of the inspection mission will be a high fidelity 3D model of the inspected surface
by post processing the collected visual data using the Structure from Motion
Fig. 3: CLAHE method employed to a tunnel image (a) before CLAHE and (b) after CLAHE.
technique (SfM) \cite{12}. The reconstruction can be further analyzed by inspection experts to detect abnormalities or other type of defects speeding up and facilitating the maintenance task, or even overall updates on the a priori mining map libraries.

In the SfM process, different camera viewpoints are used offline to reconstruct the 3D structure. The process starts with the correspondence search step, which identifies overlapping scene parts among input images. During this stage, feature extraction and matching algorithms between frames are performed to extract information about the image scene coverage. Next, it follows the geometric verification using the epipolar geometry \cite{13} to remove false matches. In this approach, it is crucial to select an initial image pair $I_1$ and $I_2$ with enough parallax to perform two-view reconstruction, before incrementally register new frames. Firstly, the algorithm recovers the sets of matched features $f_1$ and $f_2$ in both images and in the sequel estimates the camera extrinsics for $I_1$ and $I_2$ using the 5-point algorithm \cite{14}. Afterwards, it decomposes the resulting Essential matrix with Singular Value Decomposition (SVD) and finally builds the projection matrices $P_i = [R_i | t_i]$ that contain the estimated rotation and translation for each frame. In the final step, by utilizing the relative pose information, the identified features are triangulated to recover their 3D position $X^{3D}$.

Afterwards, the two-frame Bundle Adjustment \cite{15} refines the initial set of 3D points by minimizing the re-projection error and the remaining images are incrementally registered in the current camera and point sets. More specifically, the frames that observe the largest amount of the recovered 3D points are processed by the Perspective-n-Point (PnP) \cite{16} algorithm that uses 2D feature correspondences to 3D points to extract their pose. Furthermore, the newly registered images will extend the existing set of the 3D scene ($X^{3D}$) using multi-view triangulation and in the end a global Bundle Adjustment is performed in the entire model to correct drifts in the process.

3 Data Collection

The proposed novel methodology for 3D dense reconstruction in dark environments has been initially evaluated by using datasets collected from actual flights of a custom designed aerial platform (Figure 1) inside a tunnel under Mjölkuddsberget mountain located at Luleå, Sweden. The selected environment, which resembles an underground mine tunnel, was pitch dark without any external illumination, while the tunnel surfaces consisted of uneven rock formations. The dimensions of the testing tunnel area were 100×2.5×3 m$^3$, capturing the camera sequences, while the MAV was following the path along the tunnel. Furthermore, the tunnel lack the presence of strong magnetic fields, while small particles were floating in the air during the flights. The aerial platform has been equipped with a LED light bar pointing towards the field of view of the camera to illuminate its surroundings. In more details, this light bar was set to different illumination levels (luminous flux per unit area or lux) varying from 4000 lux to
3000 lux. The camera used in the dataset sequences was the FOXEER Box\textsuperscript{1} that was recording with a resolution of 1920×1080 at 60 Frames Per Second (FPS), and with a diagonal field of view of 155°, while Figure 4 depicts snapshots of the field trials during the dataset collection, where the dominating darkness in the surrounding environment is evident.

![Fig. 4: Snapshots of the MAV flight within the tunnel for the dataset collection from Mjölkuddsberget mountain at Luleå, Sweden.](image)

The datasets collected from the field trials have been divided in two cases capturing different locations of the tunnel environment, where Figures 5 and 6 depict some sample image frames from the two datasets, where it should be noted that the dataset1 includes 150 image frames and dataset2 includes 500 image frames that contain both the original as well as the enhanced with CLAHE frames.

## 4 Experimental Results

The described experimental part was designed to demonstrate the performance of the proposed enhanced image based 3D reconstruction scheme for underground tunnel inspection, using the datasets discussed in Section 3. The presented evaluation includes quantitative and qualitative results from the aspects of the 3D reconstruction and image processing. The main goal is to demonstrate the ability to enhance the images fed to the reconstruction pipeline and increase the information that can be extracted, towards a more detailed 3D model generation. The evaluation considers the comparison of the CLAHE enhanced images with the original images by using the state-of-the-art SfM software Colmap \cite{colmap}. In both reconstruction cases the same parameters have been selected regarding the feature extraction, matching as well as the sparse and dense reconstruction, as shown in Table 1.

\footnote{http://foxeer.com/Foxeer-4K-Box-Action-Camera-SuperVision-g-22}
Fig. 5: Representative snapshots from the dataset1 collected during the MAV flight. On the left column original images and on the right column the CLAHE enhanced images.

Fig. 6: Representative snapshots from the dataset2 collected during the MAV flight. On the left column original images and on the right column the CLAHE enhanced images.
Table 1: Colmap parameters used for dataset1 and dataset2

| Camera Model | Dataset1 | Dataset2 |
|--------------|----------|----------|
| Feature Matching | Simple Radial Fisheye | Simple Radial Fisheye |
| –DenseStereo.window_radius | 15 | 15 |
| –StereoFusion.min_num_pixels | 5 | 15 |
| –PoissonMeshing.trim | 7 | 7 |
| –PoissonMeshing.depth | 20 | 20 |

4.1 Dataset1

Initially for the case of the original images, the reconstruction pipeline provided two separate pointclouds. Based on the software documentation COLMAP attempts to reconstruct multiple models if not all images are registered into the same model. Therefore in the original images the SfM pipeline is not able to place the collected images in the same model. The first pointcloud processed 31 image frames and resulted in 300564 points, while the second pointcloud processed 58 image frames and provided 587098 points. Figure 7 a) depicts the 2 resulting pointclouds from the processing of the original images.

For the case of the CLAHE enhanced images the reconstruction pipeline provided a complete pointcloud processing in total 117 image frames, while resulting in 1090067 points. Figure 7 b) depicts the resulting pointcloud from the processing of the enhanced images.

Table 2 presents the total number of points for each of the generated pointclouds as well as the total image frames processed for each case. The proposed method provides a 3D model with an increase of 22.08% compared to the pointclouds generated from the original images. Moreover, the proposed method was able to use 24% more image frames compared to the original images. Based on the quantitative results the proposed method is able to enrich the image content and improve the reconstruction outcome. Generally, the original images fail to provide a single pointcloud as a result and were able only to reconstruct parts of the dataset where the illumination conditions where substantial.

Table 2: Original vs Enhanced images pointcloud summary

| # points | Original | Enhanced |
|----------|----------|----------|
| Pointcloud1 | 300564 | 587098 |
| Pointcloud2 | 587098 | 1090067 |
| # frames | 31 | 58 | 117 |

The generated models have been converted also to a 3D mesh using Poisson surface reconstruction method [17]. The mesh resulting from the proposed method is characterized by improved texture, with brighter colors, compared
Fig. 7: Dense reconstruction of dataset1 using COLMAP. (a) Original images, (b) CLAHE enhanced images

to the same location in the mesh resulting from the original images. Thus, another merit of the proposed method considers the texture enhancement of the 3D mesh. Figure 8 visualizes the different meshes generated from original and CLAHE enhanced images.

4.2 Dataset2

Regarding this case, the reconstruction pipeline processed 110 image frames and resulted in 208525 points, Figure 7a) depicts the 2 resulting pointclouds from the processing of the original images.

For the case of the CLAHE enhanced images the reconstruction pipeline provided a complete pointcloud processing in total 408 image frames, while resulting in 3562538 points. Figure 9 b) depicts the resulting pointcloud from the processing of the enhanced images.

Table 3 presents the total number of points for each of the generated pointclouds, as well as the total image frames processed for each case. The proposed method provides a 3D model with 16× more points compared to the pointclouds generated from the original images. Moreover, the proposed method was able to use 3× more image frames compared to the original images. Based on the quantitative results the proposed method is able to enrich the image content and improve the reconstruction outcome. Similarly to dataset1, the original images
Fig. 8: Instances of the generated 3D meshes using dataset1. (a) and (b) based on original image, while (c) and (d) based on CLAHE enhanced images.

Fig. 9: Dense reconstruction of dataset2 using COLMAP. (a) Original images, (b) CLAHE enhanced images.
provide a single pointcloud as a results only in the parts of the scene where the illumination conditions were substantial.

Table 3: Original vs Enhanced images pointcloud summary

|            | Original | Enhanced |
|------------|----------|----------|
| #points    | 208525   | 3563538  |
| #frames    | 110      | 408      |

Similarly to Section 4.1 the 3D mesh for each case has been generated and is depicted in Figure 10. In this scenario the original images provide slightly smoother mesh but only from the areas with sufficient illumination, whereas the proposed method was able to provide a mesh including bigger part of the inspected tunnel, trading off completeness and accuracy.

Fig. 10: Instances of the generated 3D meshes using dataset2. (a) based on original image and (b) based on CLAHE enhanced images

Based on the results from the datasets presented in this work, the proposed method is able to enhance the information extracted from the image and used by the reconstruction pipeline. The critical part that emphasizes the importance of this study is that it focuses on low cost solutions for 3D model generation applied in underground tunnel environments. This system can be the basis of a robust inspection system structured around aerial robotics and visual sensors.
4.3 Edge detection comparison

In order to evaluate the effectiveness of the proposed method the Sobel edge detection method [8] has been used, with a threshold value of 0.02.

Below a representative example of a tunnel image is depicted with its detected edges before and after the application of the CLAHE method. As one can see significant edge information is absent from the original image (Fig. 7(c)) instead of the processed image where the edge information is significant more (Figure 11(d)).

![Original Image](image1)

![CLAHE Enhanced Image](image2)

![Edge Detection on Original Image](image3)

![Edge Detection on Enhanced Image](image4)

**Fig. 11:** Effect of the Sobel edge detection (threshold=0.02) on the processed image. (a) Original image, (b) CLAHE enhanced image, (c) Edge detection on the original image and (d) Edge detection on the enhanced image.

5 Conclusions

Despite the vast advances and amount of methods for feature extraction and matching methods that are used in the 3D reconstruction pipeline when it comes to tunnel images these methods fail due to the lack of light in the images. This article investigated the effect of the CLAHE method to boost the hidden details in these type of images. The experimental results showed that a significant improvement in image contrast enhancement of tunnel images is achieved through the CLAHE method, which later significantly enhances the quality of the 3D reconstruction. Aerial platforms will have a major role in the upcoming years in the underground mining and the proposed system can be considered among
the first experimental step to address the challenging problem of lacking illumination when using visual sensors in such environments. Future work will focus on further examination of low light image enhancement methods as also Deep learning techniques aiming to employ them in real time localization and mapping of autonomous aerial vehicles in underground tunnel inspection tasks. Additionally, can be merged to online mapping techniques for coarse obstacle avoidance tasks.

References

1. Attard, L., Debono, C.J., Valentino, G., Castro, M.D.: Tunnel inspection using photogrammetric techniques and image processing: A review. ISPRS Journal of Photogrammetry and Remote Sensing 144, 180 – 188 (2018). DOI https://doi.org/10.1016/j.isprsjprs.2018.07.010. URL http://www.sciencedirect.com/science/article/pii/S0924271618302028
2. Schmid, K., Lutz, P., Tomić, T., Mair, E., Hirschmüller, H.: Autonomous vision-based micro air vehicle for indoor and outdoor navigation. Journal of Field Robotics 31(4), 537–570 (2014)
3. Gohl, P., Burri, M., Omari, S., Rehder, J., Nikolic, J., Achtelik, M., Siegwart, R.: Towards autonomous mine inspection. In: Applied Robotics for the Power Industry (CARPI), 2014 3rd International Conference on, pp. 1–6. IEEE (2014)
4. Özaslan, T., Loianno, G., Keller, J., Taylor, C.J., Kumar, V., Wozencraft, J.M., Hood, T.: Autonomous navigation and mapping for inspection of penstocks and tunnels with mavs. IEEE Robotics and Automation Letters 2(3), 1740–1747 (2017)
5. Karpowicz, J.: Uavs in civil infrastructure. Available at http://www.expouav.com/wp-content/uploads/2016/04/free-report-uavs-in-civil-infrastructure.pdf?_ga=2.149425910.1944537166.1536777582-1263854739.1536777582
6. Lal, S., Narasimhadhan, A., Kumar, R.: Automatic method for contrast enhancement of natural color images. Journal of Electrical Engineering & Technology 10(3), 1233–1243 (2015)
7. Chen, C., Chen, Q., Xu, J., Koltun, V.: Learning to see in the dark. arXiv preprint arXiv:1805.01934 (2018)
8. Gonzalez, R.C., Woods, R.E.: Digital image processing. Pearson (2018)
9. Pizer, S.M., Zimmerman, J.B., Staab, E.V.: Adaptive grey level assignment in ct scan display. Journal of Computer Assisted Tomography 8(2), 300–305 (1984)
10. Zimmerman, J.B., Pizer, S.M., Staab, E.V., Perry, J.R., McCartney, W., Benton, B.C.: An evaluation of the effectiveness of adaptive histogram equalization for contrast enhancement. IEEE Transactions on Medical Imaging 7(4), 304–312 (1988)
11. Pizer, S.M., Amburn, E.P., Austin, J.D., Cromartie, R., Geselowitz, A., Greer, T., ter Haar Romeny, B., Zimmerman, J.B., Zuiderveld, K.: Adaptive histogram equalization and its variations. Computer vision, graphics, and image processing 39(3), 355–368 (1987)
12. Schonberger, J.L., Frahm, J.M.: Structure-from-motion revisited. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 4104–4113 (2016)
13. Hartley, R., Zisserman, A.: Multiple view geometry in computer vision. Cambridge university press (2003)
14. Nistér, D.: An efficient solution to the five-point relative pose problem. IEEE transactions on pattern analysis and machine intelligence 26(6), 756–770 (2004)
15. Triggs, B., McLauchlan, P.F., Hartley, R.I., Fitzgibbon, A.W.: Bundle adjustment: modern synthesis. In: International workshop on vision algorithms, pp. 298–372. Springer (1999)
16. Gao, X.S., Hou, X.R., Tang, J., Cheng, H.F.: Complete solution classification for the perspective-three-point problem. IEEE transactions on pattern analysis and machine intelligence 25(8), 930–943 (2003)
17. Kazhdan, M., Hoppe, H.: Screened poisson surface reconstruction. ACM Transactions on Graphics (ToG) 32(3), 29 (2013)