Neural material transfer for improving photogrammetry

Luca Peretti[1,*]

[1]Department of Computer Science, University of Pisa, Pisa, Italy

Abstract

In this work we propose a method to enhance the photogrammetric reconstruction of 3D models by pre-processing the acquired photos with a material transfer network. We generated a dataset of simulating photogrammetric acquisitions. The synthetic dataset is utilized to train a Unet-like architecture to modify input photos by changing the material of the captured object towards a material less susceptible to reconstruction errors. We demonstrated that the material transfer convolutional network permits to improve the photogrammetric reconstructions obtaining more accurate 3D models.

Keywords

Photogrammetry, Material transfer

1. Introduction

Three-dimensional (3D) models are utilized in a variety of contexts, such as architecture, engineering, and entertainment. Although numerous datasets containing millions of 3D models are readily accessible online, creating new models manually is a time-consuming process that requires multiple artists for modeling and texturing. Close-range photogrammetry is a technique that accurately captures objects from photographs. This method involves taking multiple overlapping images from different positions to triangulate the 3D locations of captured surfaces on the objects. Photogrammetry has grown in popularity, thanks to the abundance of available software.

The process involves two major phases: the image capture phase and the subsequent reconstruction pipeline that use the acquired images as input. The photographs need to be captured densely around the object and the result is affected by many factors, such as the acquisition setup and the object geometry. One of the factors that affect the photogrammetric reconstruction process is the lighting setup during image acquisition. Diffuse lighting is a desirable setup, but it is not always feasible to achieve during image capture. Additionally, the material of the captured object may have a non-negligible reflective component, resulting in strongly lit areas such as highlights and reflections that can negatively impact the reconstruction.

To address these issues, we propose a data-driven AI approach that transfers the material of input images to a duller material. By applying material transfer to the input images, it is possible to reduce the impact of specular reflections on the reconstructed model, resulting in more accurate and detailed models.

Ital-IA 2023: 3rd National Conference on Artificial Intelligence, organized by CINI, May 29–31, 2023, Pisa, Italy

l.peretti@studenti.unipi.it (L. Peretti)

2. Method

We generate a synthetic dataset of rendered images simulating photogrammetry setups of the acquisition of different models and materials. The dataset provides the data for the learning based approach, and it also provides a benchmark for photogrammetry reconstructions. For this purpose, we implemented a tool to automatically render multiple overlapping images from given 3D models.

2.1. Dataset

We used 52 models from the ThreeDScans repository, and 8 different physically based rendering (PBR) materials for the generation of the dataset. ThreeDScan consists of realistic and well detailed models, with 2 millions of faces each, which represent statues from collections of various European institutes and museums. To obtain high-quality image captures, we scaled the objects to a unitary cube and positioned the cameras using hemispherical sampling around the object. Each simulated setup consisted of 100 photographs, resulting in a dataset of 41,600 high-resolution images. For each acquired image, we stored the position and direction of the camera used for the capture. The rendered images were split into two categories: reflective materials (M) and dull materials (O). This allowed us to create a dataset for supervising a deep learning model. Reflective material images were used as input for the model, while corresponding dull material images represented the target (a total of 20,800 pairs of images).

For validation purposes, we used four sets of images as a test set. The remaining 17,600 images were split into training and validation sets, with an 80%/20% split, respectively.

2.2. Network architecture

Following recent literature on material transfer [1][2], we implemented the architecture of the net by following the general shape of a UNet. The fully convolutional neural network (CNN) uses skip connections, that link together each layer i with the corresponding layer n-i, where n is the total number of layers. All convolutions use 4x4 spatial filters with stride value of 2. Convolutions downsample and upsample the input by a factor of 2, respectively in encoding and decoding steps. The architecture can be defined as:
Encoder: C64-C128-C256-C512-C512-C512-C512-C512.
Decoder: CD512-CD1024-CD1024-C1024-C1024-C512-C256-C128.
Where Ck denotes a single layer consisting of Convolution, BatchNormalization and activation function ReLU with k filters, and CDk denotes a layer consisting of Convolution, BatchNormalization, Dropout, ReLU with a dropout rate of 50%. Leaky Relu activation function is used for introducing non linearity in the encoding steps, while Relu is used in deconding steps. The model trained with L2 loss, learning rate of 0.0002 with exponential decay turned out to be the best with an Mean Square Errors (MSE) of 0.0003±0.0002 and Structural Similarity Index Measure (SSIM) of 0.97±0.02, averaged over validation test.

2.3. Training

The model has been implemented in PyTorch. The training was performed with a NVidia GeForce RTX 3080 GPU and took 8 hours for the convergence. The selected network requires around 3GB of memory.

2.4. Reconstruction

The net's evaluation was performed by comparing both the inferred and target images, as well as the 3D model reconstructions obtained by preprocessing with material transfer versus unpreprocessed models. The 3D models were reconstructed using Meshroom [3] photogrammetry software, then aligned and filtered according to the available ground truth. Alignment was achieved by registering the ground truth camera poses with the inferred camera poses, while filtering involved removing all the points of the reconstruction point clouds outside the unitary box. Quantitative evaluation of the reconstructed models was performed using the Chamfer distance and Hausdorff distance metrics, computed between the obtained models and ground truths. Additionally, a visual evaluation of the reconstructed models was performed using MeshLab open-source software.

	Recon Points		Camera Used	
	M	O'	M	O'
Test model 1	267387	**310579**	99%	**99**%
Test model 2	295213	**345798**	95%	**98**%
Test model 3	193519	**203778**	73%	**82**%
Test model 4	462977	**476058**	**100%**	**100%**

Table 1
Quantitative evaluation of reconstructed point clouds of the test models. M represents the usage of original image set. O' represents the usage of images preprocessed by using Unet-like architecture. HD is symmetric Hausdorff Distance. CD is Chamfer Distance. Best results in bold.

3. Results

The results of quantitative evaluation are reported in Tables 1-2. The results demonstrate a marked improvement in the reconstruction of 3D models thanks to the use of the material transfer net. Specifically, The use of preprocessed images (O' material) allows the photogrammetric reconstruction to utilize a higher number of images (corresponding to camera shots), compared to the original unpreprocessed images (M material). Additionally, there is a corresponding increase in the number of points in the reconstructed point clouds. This confirms that the use of the neural network for preprocessing resulted in a more detailed representation of the 3D model.

	HD		CD	
	M	O'	M	O'
Test model 1	3.3E-3	**3.1E-3**	5.2E-5	**5.16E-05**
Test model 2	6.6E-3	**6.4E-3**	3.4E-4	**3.39E-04**
Test model 3	8.4E-3	**5.9E-3**	2.8E-4	**2.19E-04**
Test model 4	2.1E-2	**2.1E-2**	2.2E-3	**2.22E-03**

Table 2
Quantitative evaluation of reconstructed point clouds of the test models. M represents the usage of original image set. O' represents the usage of images preprocessed by using Unet-like architecture. Best results in bold.

Reconstructed models were also evaluated visually. Figures 1 and 2 show the comparison between models reconstructed with and without the material transfer step. It can be observed that our method leads to a more detailed reconstruction. Specifically, the holes caused by highlights in the input original images (M) are avoided using preprocessed images.

4. Conclusion

We proposed an innovative approach to improve the accuracy and quality of photogrammetric reconstructions by addressing the negative impact of reflective materials on the acquired objects. The approach involves us-

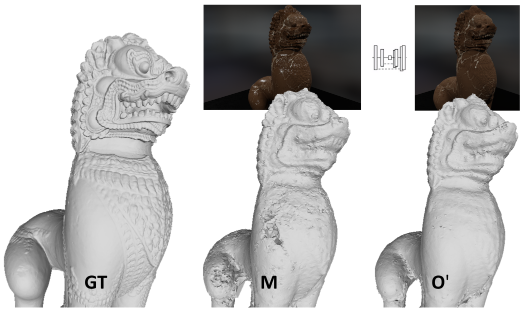

Figure 1: Comparison of photogrammetry reconstructions of *Test model 1*. GT: ground truth. M: original set of images. O': preprocessed set of images. The proposed material transfer method improves the reconstruction.

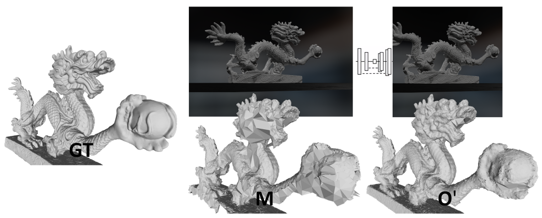

Figure 2: Comparison of photogrammetry reconstructions of *Test model 3*. GT: ground truth. M: original set of images. O': preprocessed set of images. The proposed material transfer method improves the reconstruction.

ing a material transfer method to change the material type of the photographed object from reflective to dull, thereby reducing the impact of specular reflections on the reconstructed model. To achieve this objective, we generated a dataset of images by simulating photogrammetry setups of the acquisition of different models and materials. We used the dataset to train a CNN architecture with skip connections to transfer the material of the input images from a reflective material to a dull one. We empirically demonstrated that preprocessing images using our proposed method improved the results both quantitatively and qualitatively. Our method allows for improving photogrammetry by obtaining accurate and detailed 3D models from photographs.

References

[1] Z. Li, K. Sunkavalli, M. Chandraker, Materials for masses: Svbrdf acquisition with a single mobile phone image, in: Proceedings of the European conference on computer vision (ECCV), 2018, pp. 72–87.

[2] W. Ye, X. Li, Y. Dong, P. Peers, X. Tong, Single image surface appearance modeling with self-augmented cnns and inexact supervision, in: Computer Graphics Forum, volume 37, Wiley Online Library, 2018, pp. 201–211.

[3] C. Griwodz, S. Gasparini, L. Calvet, P. Gurdjos, F. Castan, B. Maujean, G. De Lillo, Y. Lanthony, Alicevision meshroom: An open-source 3d reconstruction pipeline, in: Proceedings of the 12th ACM Multimedia Systems Conference, 2021, pp. 241–247.