CEIG – Spanish Computer Graphics Conference (2021)
A. Chica and L. Ortega (Editors)

Neural Colorization of Laser Scans

M. Comino[1], C. Andújar[2], C. Bosch[3], A. Chica[2] and I. Munoz-Pandiella[4]

[1]Universidad Rey Juan Carlos. Spain
[2]Universitat Politècnica de Catalunya. Spain
[3]Universitat de Vic - Universitat Central de Catalunya, Spain
[4]Universitat de Barcelona, Spain

Abstract

Laser scanners enable the digitization of 3D surfaces by generating a point cloud where each point sample includes an intensity (infrared reflectivity) value. Some LiDAR scanners also incorporate cameras to capture the color of the surfaces visible from the scanner location. Getting usable colors everywhere across 360° scans is a challenging task, especially for indoor scenes. LiDAR scanners lack flashes, and placing proper light sources for a 360° indoor scene is either unfeasible or undesirable. As a result, color data from LiDAR scans often do not have an adequate quality, either because of poor exposition (too bright or too dark areas) or because of severe illumination changes between scans (e.g. direct Sunlight vs cloudy lighting). In this paper, we present a new method to recover plausible color data from the infrared data available in LiDAR scans. The main idea is to train an adapted image-to-image translation network using color and intensity values on well-exposed areas of scans. At inference time, the network is able to recover plausible color using exclusively the intensity values. The immediate application of our approach is the selective colorization of LiDAR data in those scans or regions with missing or poor color data.

1. Introduction

Laser scanners provide a fast and convenient way of acquiring the shape of 3D scenes in a number of applications, including surveying, architecture, building, construction, engineering and cultural heritage. Terrestrial LiDAR equipment (TLS) is often mounted on a rotating support, allowing horizontal sweeping while using a rotating mirror to deflect the infrared beam vertically, so a single scan captures every surface point visible from the scanner location. The distance (range) to a given surface is estimated by measuring the time it takes for the beam to return to the sensor after bouncing on it (time-of-flight).

High-end LiDAR scanners provide accurate geometry data. Moreover, due to the use of an infrared beam, the accuracy of the geometric data is not affected by external lighting conditions. This makes LiDAR scanners much more robust than photogrammetry approaches, in terms of shape acquisition.

Unfortunately, the quality of color data in LiDAR equipment is not on-par with that of geometry data. Although LiDAR scanners embed cameras to capture color, they lack flash units to help illuminate a scene, since these units would only be practical for taking good pictures at distances of over a few meters. Therefore, the quality of the acquired color data heavily depends on existing lighting conditions. As a result, images from LiDAR scans often contain regions with poor/missing color data (Figure 1). Notice that color quality could be improved by using auxiliary lighting kits (such as softboxes) to illuminate the scene more uniformly, but these devices and their supports will appear in the point cloud and thus partially invalidate both color and geometry data. Furthermore, LiDAR color acquisition technology is prone to other color artifacts not directly related to exposure, such as lens flares or geometry and color miss-registrations (Figure 2).

Figure 1: *Color artifacts from a sample LiDAR scan: over-exposed areas (1), under-exposed areas (2), and undesirable shadows cast by scanning equipment (3).*

This contrasts with competing photogrammetric approaches, where flash units, moving illumination devices, and color calibration patterns can be used to get accurate color data, at the expense of requiring a much higher number of photos and much longer post-processing times. For this reason, both technologies are often combined to achieve a final result.

In this paper, we present a deep-learning approach to generate plausible color data from the infrared data. We benefit from

DOI: 10.2312/ceig.20211356

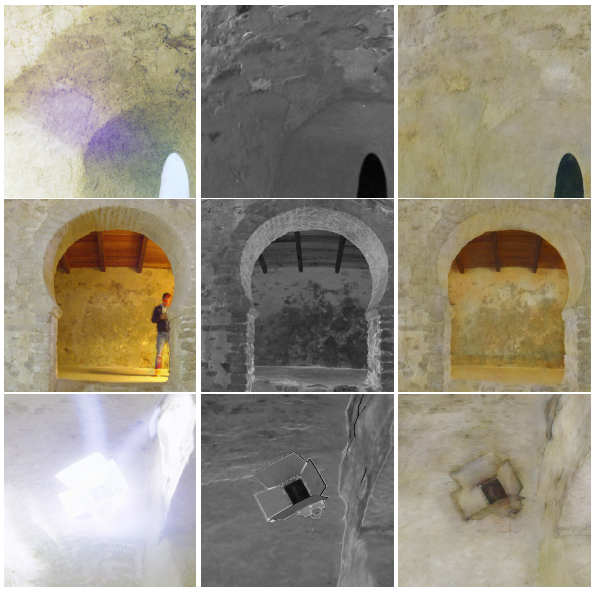

Figure 2: *Color artifacts from LiDAR scans. Top: lens flare artifact. Middle: color-geometry miss-registration (a captured person which does not appear in the geometry). Bottom: Clipped highlight. From left to right: Original color map, infrared intensity (artifact-free) captured by the scanner and color reconstructed with our method.*

the fact that the intensity values measured by the laser beam are nearly illumination independent , specially in indoor scenes. The key idea is to use an image-to-image translation network. The network is trained end-to-end using tiles from selected regions of LiDAR scans containing high-quality color data (i.e. after removing image parts with poor color data). The network learns to predict RGB values just from the (illumination-independent) intensity values. We have tested this approach with actual LiDAR scans. We discuss key issues (preventing border artifacts, importance of the loss metric) to get suitable colors. Our first experiments show promising results, the network being able to fully recover plausible color data, sometimes close to the original color.

2. Previous Work

Colorization Colorization techniques can be classified into scribble-based, exemplar-based, and deep learning based.

In the former, the user must provide some type of local information (usually color scribbles) that allow the algorithm to extend the color to the rest of the image [LLW04]. The result is greatly improved if scribbles and edge information are used to segment the image before spreading the color [HTC*05, LWCO*07].

In the second type, the key is to transfer the colors of an exemplar to the target grayscale image [WAM02, GCR*12, BTP13]. Although this allows to considerably reduce the work on the part of the user, a sufficiently representative exemplar is needed.

Finally, given the possibility of generating pairs of color and grayscale images from a large collection of input images, it is possible to train neural networks to colorize [ZIE16]. It is even possible to combine information provided by the user with this type of techniques [ZZI*17], or to apply more general architectures capable of dealing with the problem of transforming between different types of images [IZZE17].

Laser Scan Intensity The intensity acquired by LiDAR scanners records the power of the reflected beam on the surfaces, often in the infrared range. This intensity might be affected by several factors, and many authors have proposed correction processes to reduce their influence, including incidence angle [KOPW15]. For diffuse surfaces, intensity is typically assumed to follow a cosine law with incidence angle [Bol19], as we do here.

It has been shown that intensity values correlate well with measured luminance [BPMTMML17]. Oishi and Kurazume exploit this to colorize intensity images from laser scans, but requires finding correspondences with color images [OK14]. Color accuracy has been recently evaluated for different scanners in controlled lab conditions [JKR*20]. Such conditions do not show typical illumination changes that happen in real situations, and no solutions are provided to improve color accuracy itself.

3. Our approach

Consider a point cloud of a certain scene captured from a discrete set of scanner locations $Q \subset \mathbb{R}^3$ (our methodology does not require any registration information). LiDAR scanners measure the time-of-flight of infrared pulses to estimate the distance to the surface of the objects in this scene. Consequently, they can provide, for each point $\mathbf{p}$, its associated position (x,y,z) and infrared reflectivity (ir) information.

After capturing the scene geometry, color is usually recorded by stitching a mosaic of photographs. Therefore, for each scan location $\mathbf{q} \in Q$, the associated color information is provided as a registered panoramic texture (either LDR or HDR). Sometimes the color is sampled and provided as (r,g,b) values associated with each point. In this case, we convert these into a texture using a procedure similar to the one used to produce an infrared reflectivity texture, which is explained next.

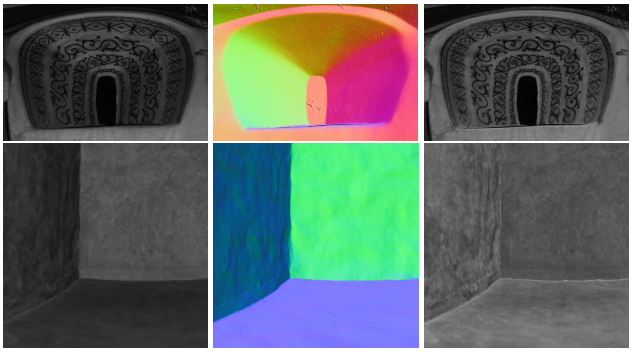

Figure 3: *From left to right: original infrared intensity, estimated normal map and calibrated infrared. We adjust the infrared values using the cosinus between the view direction and the normal of the surface to obtain a more homogeneous signal. Calibrated infrared are the input to our colorization network.*

Figure 4: *Top: Original color. Bottom: masked regions with low-quality color information. Training tiles will only be generated from non-masked areas.*

Recall that our aim is computing a color texture from the (illumination-independent) intensity values. In order to use an image-to-image translation network for this task, we need an infrared texture aligned with the color texture. We follow Comino et al. [CACB17] and raycast a panoramic infrared reflectivity texture from each scan location. Multiple splatting weights are used to ensure the sharpness of the result. Although some LiDAR equipment may already perform some kind of radiometric calibration, we found that for our test device (Leica RTC 360 [web]) the infrared values become much more homogeneous when adjusted using the cosine of the angle between the view ray and the surface normal (Figure 3). Because surface normals are not directly available on raw LiDAR point clouds, we use the method of Comino et al. [CACB18] to estimate them.

In our experiments, we use images of 8192×4096 pixels of resolution and we have on average 20 images (corresponding to 20 scan locations) for each scene. Most neural networks can not directly work on images of such resolution and, therefore, we need to subdivide them into image tiles. For this, we consider two important factors:

- The receptive field of the chosen architecture (explained below) is very large. This means that the influence of border artifacts can potentially be very strong leading to tile discontinuities. In order to alleviate this, we augment our dataset by producing tiles with a huge overlap.
- We want to avoid training on parts of the images which are incorrectly exposed or contain color artifacts. For this, we manually annotate these regions and avoid producing tiles containing them (Figure 4).

More specifically, we regularly generate tiles of 512×512 pixels with a constant stride of 64 pixels skipping any tile which partially overlaps a masked area. On a dataset with 29 scans we produced 64k tiles. Each tile has associated RGB and calibrated infrared information. We further augment these by performing random vertical and horizontal flips at train time.

We learn to generate color (RGB) images from the calibrated infrared textures. For this task, we used the image-to-image translation UNet [RFB15] architecture. This architecture was originally designed for segmentation but it has been extensively used to directly predict the RGB channels by modifying the last layer. In particular, we tuned the Pytorch implementation by Buda et al. [BSM19] and introduced the following changes:

- We substituted the intermediate ReLU activations by LeakyReLUs with negative slope of 0.1, as it is known to produce better learning gradients and ease the learning process.
- We replaced the upsampling deconvolutions by nearest-neighbor upsampling units to avoid upsampling cross artifacts.
- We removed the last sigmoid activation. This is used in classification tasks to output probability values between 0 and 1. Although pixel values are also often limited to this range, sigmoid activations can cause vanishing gradients and hinder the training. Moreover, similar results can be achieved without them.

Our network is trained end-to-end and learns to directly produce RGB values between 0 and 1. As loss function, we use the perceptual loss proposed by Zhang et al. [ZIE*18], in particular its Pytorch implementation using the VGG16. Moreover we add a traditional L1 loss weighted by 0.1. Perceptual losses compare the difference in the activations of the generated and ground truth images after going through a pretrained network (in this case, the VGG16). They are more informed losses than pixel-based ones and have proven to produce sharper results in image generation tasks.

4. Results

All tests were run on a PC equipped with an Intel Core i7-8700 CPU, 32GB of RAM, 4TB Seagate ST4000DM004 HDD, and a NVIDIA TITAN V with 12GB of memory running Ubuntu 20.04. Each epoch processed about 64k images in batches of 4 tiles and took about 90 minutes. All results on this paper were generated after training our models for 16 epochs.

Figure 5 shows how our colorization method behaves both on areas which were used for training and areas which were masked. The reconstructed color looks homogeneous, plausibly resembling the original textures on well-exposed areas, while the prediction on saturated or shadowed areas is consistent. Figure 7 shows ours results on multiple complete panoramas corresponding to different scan locations from the same scene. Other color artifacts are also corrected, as shown in Figure 2.

Notice there is also a lot of consistency between different views. To further illustrate this, in Figure 6 we show the reference and reconstructed colors for two close-ups of mural paintings captured from different locations.

Although our approach alleviates the effect of border artifacts, some smaller color discontinuities can still be observed on the reconstructed color textures in Figure 7. This can easily be addressed by, at prediction time, producing tiles with certain overlap and keeping only their central region, as shown in Figure 8.

Figure 5: *Left: Original images. Middle: Depicting the manually annotated regions (black boxes) from which no training tiles were generated. Right: Predicted colorization for these images. The color is consistent whether the original pixels were masked or not.*

Figure 6: *Mural painting captured from two different locations. Top: reconstructed color. Bottom: original color.*

5. Conclusions and Future Work

Obtaining a good colorization for LiDAR data is a challenging task. In this work, we present a simple strategy to produce a plausible consistent colorization for the different scan locations of a given scene. One major limitation of this approach is reconstructing color just from infrared information. Multiple materials can map to the same infrared intensity and, in order to discern them, the network must look at higher-order features on the infrared channel. The amount of available training data is limited (few scan locations), which translates into a trade-off between correct color reconstruction and generalization capability. In other words, we have observed that some strategies can overfit the training data resulting in more faithful colors at the seen regions while predicting less plausible results for the unseen ones.

Another limitation is the current way of delimiting the training area. In this work, a user must manually annotate the pixel values containing color artifacts. This implies a subjective definition of the color quality. One option would be to automate this process by detecting regions with extreme illumination, both shadowed and saturated, especially when HDR images are also acquired. However, the quality of the images captured by LiDAR equipment is already limited even on correctly exposed areas. This could be solved by learning on high quality photographs taken from high-end digital cameras projected onto the point clouds.

Acknowledgments This work has been partially funded by the Spanish Ministry of Economy and Competitiveness and FEDER under grant TIN2017-88515-C2-1-R, the *Romanesque Pyrenees, Space of Artistic Confluences II* (PRECA II) project (HAR2017-84451-P, UB) and by the EU Horizon 2020, JPICH Conservation, Protection and Use initiative (JPICH-0127) and the Spanish Agencia Estatal de Invesigación (grant PCI2020-111979). The authors would like to thank Prof. Xavier Pueyo for his helpful comments.

References

[Bol19] BOLKAS D.: Terrestrial laser scanner intensity correction for the incidence angle effect on surfaces with different colours and sheens. *International Journal of Remote Sensing 40*, 18 (2019), 7169–7189. 2

[BPMTMML17] BALAGUER-PUIG M., MOLADA-TEBAR A., MARQUÉS-MATEU A., LERMA J.: Characterisation of intensity values on terrestrial laser scanning for recording enhancement. *International Archives of the Photogrammetry, Remote Sensing & Spatial Information Sciences 42* (2017). 2

[BSM19] BUDA M., SAHA A., MAZUROWSKI M. A.: Association of genomic subtypes of lower-grade gliomas with shape features automatically extracted by a deep learning algorithm. *Computers in Biology and Medicine 109* (2019). doi:10.1016/j.compbiomed.2019.05.002. 3

[BTP13] BUGEAU A., TA V.-T., PAPADAKIS N.: Variational exemplar-based image colorization. *IEEE Transactions on Image Processing 23*, 1 (2013), 298–307. 2

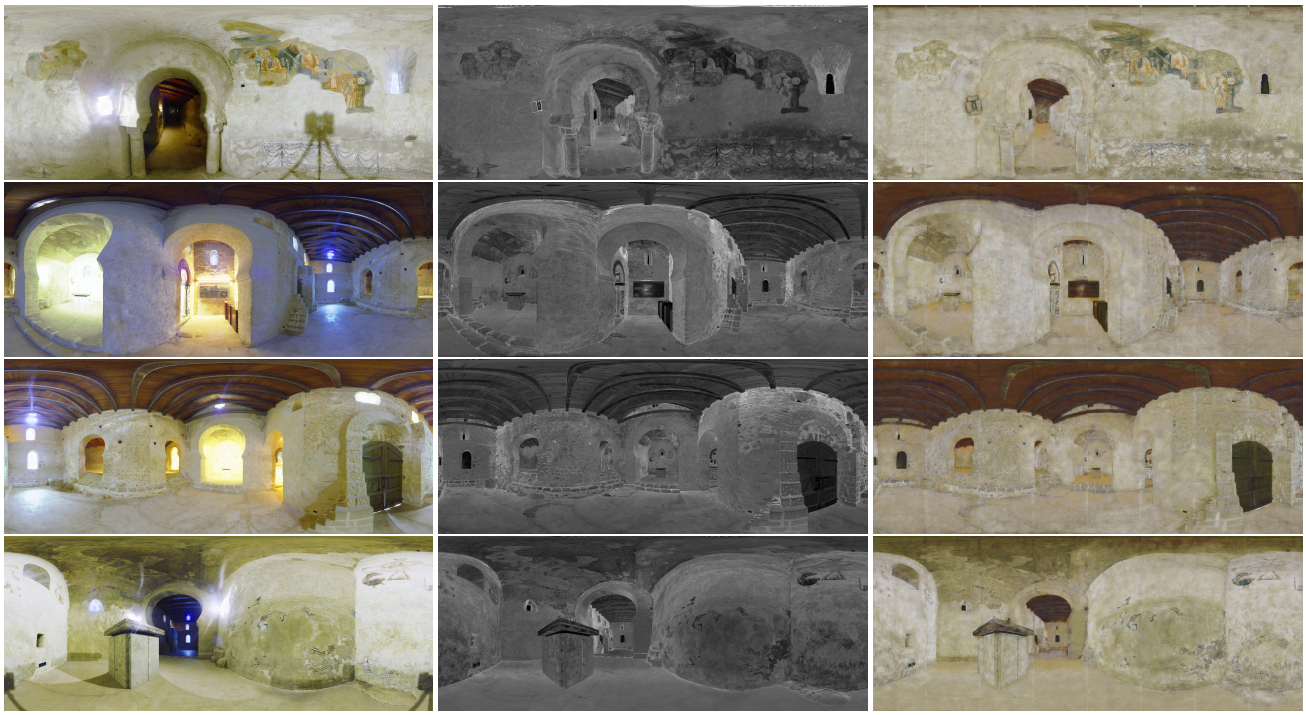

Figure 7: *Results on several scan locations. From left to right: Original color, calibrated infrared and reconstructed color. In this case, the color prediction was done over non-overlapping* 1024×1024 *pixel tiles.*

Figure 8: *Results on several scan locations. Left: Tile-based colorization without overlapping. Right: To prevent border artifacts we produce multiple overlapping tiles and keep their central regions.*

[CACB17] COMINO M., ANDÚJAR C., CHICA A., BRUNET P.: Error-aware construction and rendering of multi-scan panoramas from massive point clouds. *Computer Vision and Image Understanding 157* (2017), 43–54. 3

[CACB18] COMINO M., ANDUJAR C., CHICA A., BRUNET P.: Sensor-aware normal estimation for point clouds from 3D range scans. *Computer Graphics Forum 37*, 5 (2018), 233–243. 3

[GCR*12] GUPTA R. K., CHIA A. Y.-S., RAJAN D., NG E. S., ZHIYONG H.: Image colorization using similar images. In *Proceedings of the 20th ACM international conference on Multimedia* (2012), pp. 369–378. 2

[HTC*05] HUANG Y.-C., TUNG Y.-S., CHEN J.-C., WANG S.-W., WU J.-L.: An adaptive edge detection based colorization algorithm and its applications. In *Proceedings of the 13th annual ACM international conference on Multimedia* (2005), pp. 351–354. 2

[IZZE17] ISOLA P., ZHU J.-Y., ZHOU T., EFROS A. A.: Image-to-image translation with conditional adversarial networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (2017), pp. 1125–1134. 2

[JKR*20] JULIN A., KURKELA M., RANTANEN T., VIRTANEN J.-P., MAKSIMAINEN M., KUKKO A., KAARTINEN H., VAAJA M. T., HYYPPÄ J., HYYPPÄ H.: Evaluating the quality of TLS point cloud colorization. *Remote Sensing 12*, 17 (2020), 2748. 2

[KOPW15] KASHANI A. G., OLSEN M. J., PARRISH C. E., WILSON N.: A review of lidar radiometric processing: From ad hoc intensity correction to rigorous radiometric calibration. *Sensors 15*, 11 (2015), 28099–28128. 2

[LLW04] LEVIN A., LISCHINSKI D., WEISS Y.: Colorization using optimization. In *ACM SIGGRAPH 2004 Papers*. 2004, pp. 689–694. 2

[LWCO*07] LUAN Q., WEN F., COHEN-OR D., LIANG L., XU Y.-Q., SHUM H.-Y.: Natural image colorization. In *Proceedings of the 18th Eurographics conference on Rendering Techniques* (2007), pp. 309–320. 2

[OK14] OISHI S., KURAZUME R.: Manual/automatic colorization for three-dimensional geometric models utilizing laser reflectivity. *Advanced Robotics 28*, 24 (2014), 1637–1651. 2

[RFB15] RONNEBERGER O., FISCHER P., BROX T.: U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention* (2015), Springer, pp. 234–241. 3

[WAM02] WELSH T., ASHIKHMIN M., MUELLER K.: Transferring color to greyscale images. In *Proceedings of the 29th annual conference on Computer graphics and interactive techniques* (2002), pp. 277–280. 2

[web] Leica rtc 360. https://leica-geosystems.com/en-us/products/laser-scanners/scanners/leica-rtc360. 3

[ZIE16] ZHANG R., ISOLA P., EFROS A. A.: Colorful image colorization. In *European conference on computer vision* (2016), Springer, pp. 649–666. 2

[ZIE*18] ZHANG R., ISOLA P., EFROS A. A., SHECHTMAN E., WANG O.: The unreasonable effectiveness of deep features as a perceptual metric. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (2018), pp. 586–595. 3

[ZZI*17] ZHANG R., ZHU J.-Y., ISOLA P., GENG X., LIN A. S., YU T., EFROS A. A.: Real-time user-guided image colorization with learned deep priors. *arXiv preprint arXiv:1705.02999* (2017). 2