

VIDIT: Virtual Image Dataset for Illumination Transfer

Majed El Helou, Ruofan Zhou, Johan Barthas, Sabine Süsstrunk

School of Computer and Communication Sciences, EPFL, Switzerland
{majed.elhelou, ruofan.zhou}@epfl.ch

Abstract. Deep image relighting is gaining more interest lately, as it allows photo enhancement through illumination-specific retouching without human effort. Aside from aesthetic enhancement and photo montage, image relighting is valuable for domain adaptation, whether to augment datasets for training or to normalize input test data. Accurate relighting is, however, very challenging for various reasons, such as the difficulty in removing and recasting shadows and the modeling of different surfaces. We present a novel dataset, the Virtual Image Dataset for Illumination Transfer (VIDIT), in an effort to create a reference evaluation benchmark and to push forward the development of illumination manipulation methods. Virtual datasets are not only an important step towards achieving real-image performance but have also proven capable of improving training even when real datasets are possible to acquire and available. VIDIT contains 300 virtual scenes used for training, where every scene is captured 40 times in total: from 8 equally-spaced azimuthal angles, each lit with 5 different illuminants. <https://github.com/majedelhelou/VIDIT>

1 Dataset Description

1.1 Content

VIDIT includes in total 390 different scenes, each of which is captured with 40 predetermined illumination settings, resulting in 15,600 images. The illumination settings are all the combinations of 5 color temperatures (2500K, 3500K, 4500K, 5500K and 6500K) and 8 light directions (N, NE, E, SE, S, SW, W, NW), each of which is shown in Fig. 3 for a sample image from our dataset. The directions are equally-spaced azimuthal angles, with a fixed polar angle. The dataset includes both indoor and outdoor scenes, and miscellaneous objects with different surfaces and materials. All scenes are rendered with a resolution of 1024×1024 pixels, contain different materials such as metal, wood, stone, water, plant, fabric, smoke, fire, plastic, etc. The illumination settings as well as the depth information are recorded with each rendered images.

VIDIT is split into different mutually-exclusive sets; train (300 scenes), validation (45 scenes), and test (45 scenes). The test set is kept private for benchmarking purposes, while the train and validation sets are made public for both training and method evaluation.



Fig. 1. A selection of scenes captured at 4500K, with the light source placed at (N,NW,SE,S,W,S,NE,W) from left to right then top down. Images are shown down-sampled by 2.

| | Scenes | Light Orientations | Color Temperatures | Total | Availability |
|------------|--------|--------------------|--------------------|--------|--------------|
| Train | 300 | 8 | 5 | 12,000 | Public |
| Validation | 45 | 8 | 5 | 1,800 | Public |
| Test | 45 | 8 | 5 | 1,800 | Private |

Table 1. VIDIT content distribution. The dataset is split into train, validation, and test, the latter kept private for benchmarking. Each scene is captured under 40 different illumination settings, and is rendered with a 1024×1024 pixel resolution.

1.2 Acquisition

We use Unreal Engine 4 to render high-resolution and realistic scenes to build our dataset. The scenes are obtained from a variety of different virtual environments, which are scaled into a uniform reference space before running the illumination and rendering process. The selection of scenes (or fields of view) is done manually to avoid any duplicates or very similar content being captured accidentally by two images. The scene is illuminated with an omni-directional light source, with 5 color temperatures, and the source is positioned at the 8 different azimuthal angles illustrated in Fig. 2. The fields of view with walls or large obstructive objects that do not permit proper illumination from all angles were manually removed from the dataset.

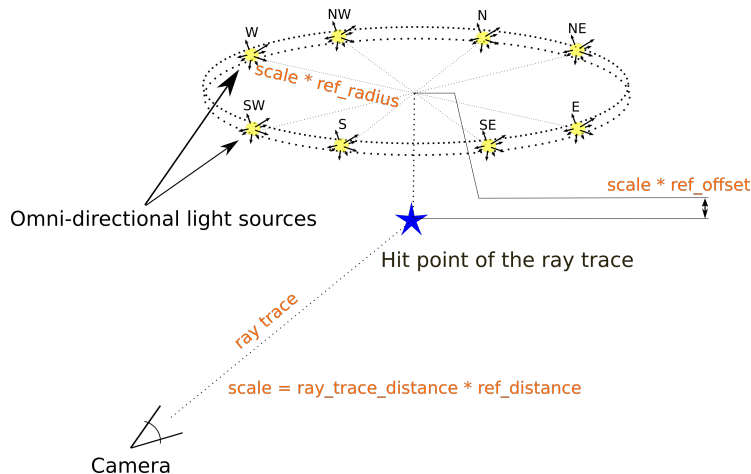


Fig. 2. We place a set of light sources above the hit point of the ray trace to keep the same light settings for each of our captured scenes. We automatically adjust to the environment’s scale by adapting our reference radius, reference offset, and reference camera distance. During capture, we turn on the light source corresponding to the selected cardinal direction (N, NE, E, SE, S, SW, W, NW) and turn off all the others. We also vary the light source’s color temperature for 5 different captures.

2 Related Work

We list a series of recent work related to our dataset which we found relevant.

A Dataset of Multi-Illumination Images in the Wild ICCV’19 [4].

This paper presents a dataset of *interior* scenes only, mostly objects in small and controlled environments. Each scene is captured with a flash from 25 different orientations. The illuminant itself, however, does not vary across the dataset.

Underexposed Photo Enhancement Using Deep Illumination Estimation CVPR’19 [7].

This dataset contains pairs of underexposed images. In opposition to the MIT-Adobe FiveK dataset, which is geared towards general photo enhancement, the authors focus specifically on enhancing underexposed photos. The ground-truth enhanced images are created by photo experts using Adobe Lightroom.

Deep Image-Based Relighting from Optimal Sparse Samples TOG’18 [8].

This paper, similar to [4], also proposes a dataset of scenes with different light directions. The differences are that the images are rendered and that the light directions are randomized.

Single Image Portrait Relighting TOG’19 [6].

This paper aims at relighting human portrait photos. The method is trained on

a dataset consisting of 18 individuals captured under different directional light sources. The capture is under controlled settings, with the individual illuminated by a sphere with numerous lights. Specifically targeted at face relighting, the method does not extend to general scenes.

Deep Face Normalization TOG'19 [5].

This paper partly addresses relighting. It aims at normalizing faces in portrait images, i.e., removing distortions and non-neutral facial expressions, and relighting to simulate an evenly-lit environment. It does not address relighting to different light sources or to a desired light direction.

There is a rich recent body of literature on style transfer methods and on intrinsic image decomposition. To name a few, IIW [1] and SAW [2] contain human-labeled reflectance and shading annotations, BigTime [3] contains time-lapse data where the scenes are illuminated under varying lighting conditions. These works, however, do not provide pairs of scenes captured with the same lighting conditions (color temperature, direction) that allow a quantitative evaluation of illumination transfer from one photo to another as our dataset does.

3 Discussion and Applications

Our dataset can be used in different applications, both for learning and evaluation. It can be used to predict the orientation of the illuminant in a scene, as well as its color temperature. More interestingly, it can be employed in domain adaptation to transform any input image such that it matches the target domain's illumination settings (which is an any-to-one illumination style transfer problem). An even more general and challenging application is what we call the any-to-any illumination style transfer, which consists of mapping an input image with certain illumination settings to match the illumination settings of another input guide image. Both the any-to-one and any-to-any problems can be addressed with a guide input image from the target domain, or with input illumination settings (orientation and color temperature). And all applications can be evaluated with objective metrics, since our dataset includes all of the different illumination settings for every single captured scene.

Some use cases of image relighting methods would be

- **Domain adaptation for data augmentation:** input scenes can be transformed into a variety of illumination settings and added to training datasets to improve the robustness of deep networks to light variation.
- **Domain adaptation for normalization:** input scenes can be transformed from many illumination settings into a single reference one, before they are fed into a pre-trained deep network or a search and retrieval method.
- **Photo editing and shadow manipulation:** a scene can be relit for aesthetic reasons, to match the desired result of the user.

- **Photo montage**: combining multiple images into one necessitates that they all match the same final illumination settings. This can be achieved using an image relighting method.

References

1. Bell, S., Bala, K., Snavely, N.: Intrinsic images in the wild. *ACM Transactions on Graphics (TOG)* **33**(4), 159 (2014) [4](#)
2. Kovacs, B., Bell, S., Snavely, N., Bala, K.: Shading annotations in the wild. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. pp. 6998–7007 (2017) [4](#)
3. Li, Z., Snavely, N.: Learning intrinsic image decomposition from watching the world. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. pp. 9039–9048 (2018) [4](#)
4. Murmann, L., Gharbi, M., Aittala, M., Durand, F.: A dataset of multi-illumination images in the wild. In: *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*. pp. 4080–4089 (2019) [3](#)
5. Nagano, K., Luo, H., Wang, Z., Seo, J., Xing, J., Hu, L., Wei, L., Li, H.: Deep face normalization. *ACM Transactions on Graphics (TOG)* **38**(6), 183 (2019) [4](#)
6. Sun, T., Barron, J.T., Tsai, Y.T., Xu, Z., Yu, X., Fyffe, G., Rhemann, C., Busch, J., Debevec, P., Ramamoorthi, R.: Single image portrait relighting. *ACM Transactions on Graphics (TOG)* **38**(4), 79 (2019) [3](#)
7. Wang, R., Zhang, Q., Fu, C.W., Shen, X., Zheng, W.S., Jia, J.: Underexposed photo enhancement using deep illumination estimation. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. pp. 6849–6857 (2019) [3](#)
8. Xu, Z., Sunkavalli, K., Hadap, S., Ramamoorthi, R.: Deep image-based relighting from optimal sparse samples. *ACM Transactions on Graphics (TOG)* **37**(4), 126 (2018) [3](#)

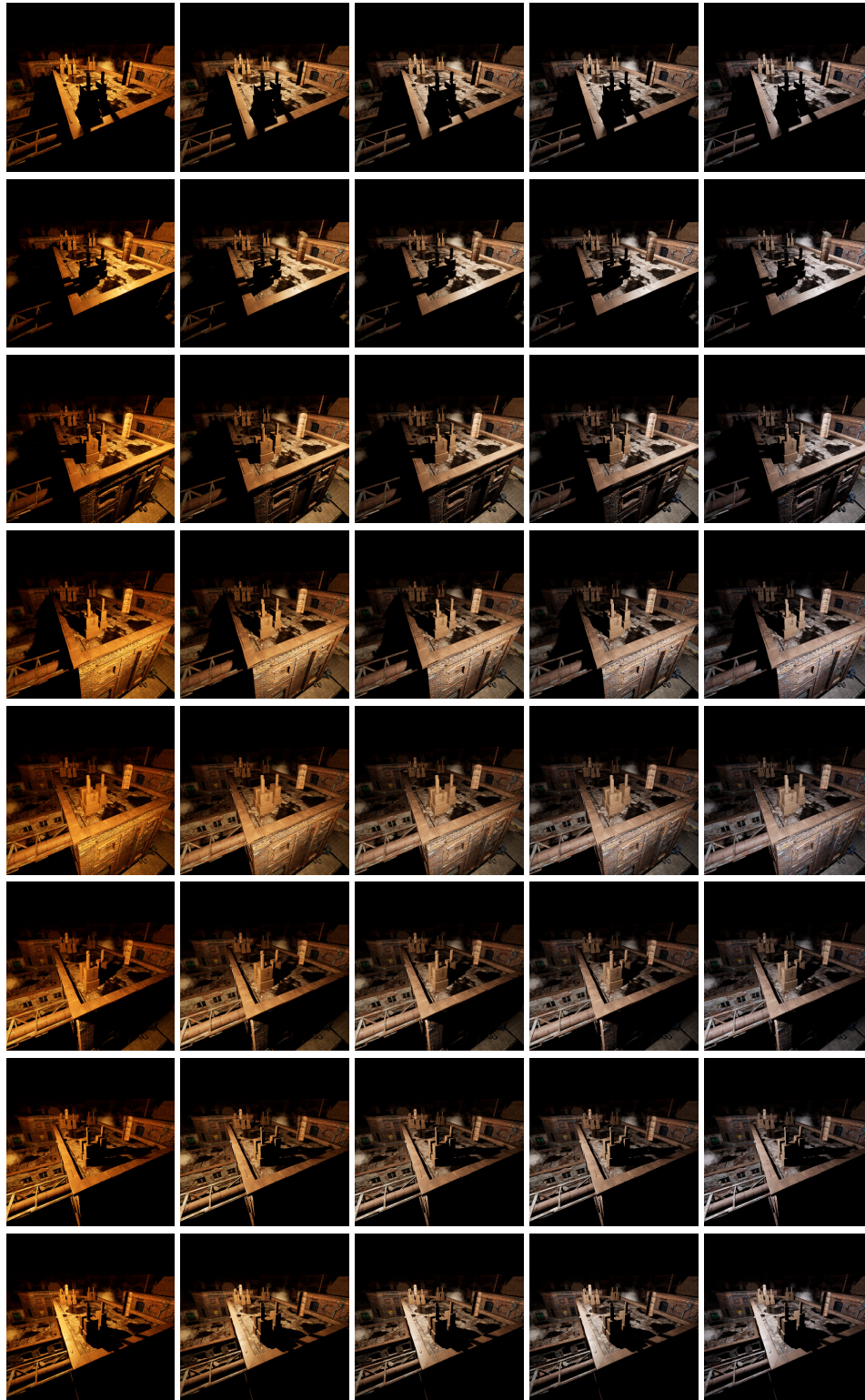


Fig. 3. Example scene captured with 40 different illumination settings. From left to right the illuminant color temperature goes from 2500K to 6500K, and from top to bottom the light source position goes from N to NW.