Computational Photography Lab.

Find the project web page here:
https://yaksoy.github.io/intrinsicCompositing

Intrinsic Harmonization for Illumination-Aware Compositing Supplementary Material

Chris Careaga
Simon Fraser University

S. Mahdi H. Miangoleh
Simon Fraser University

Yağız Aksoy
Simon Fraser University

ACM Reference Format:
Chris Careaga, S. Mahdi H. Miangoleh, and Yağız Aksoy. 2023. Intrinsic Harmonization for Illumination-Aware Compositing Supplementary Material. In *SIGGRAPH Asia 2023 Conference Papers (SA Conference Papers '23), December 12–15, 2023, Sydney, NSW, Australia.* ACM, New York, NY, USA, 2 pages. https://doi.org/10.1145/3610548.3618178

In this supplementary document, we provide (1) training and inference details for the re-shading network. (2) architectures, training details, and dataset generation process for our albedo harmonization network in Section B, (3) details of our user study in Section C

A SHADING REFINEMENT

Our re-shading network is trained at a resolution of $(384x384)$ with a batch size of 8. For each batch, images are non-uniformly sampled from each of the 3 three datasets. We bias the sampling towards the multi-illumination dataset as the off-the-shelf intrinsic decomposition method is trained on this data and therefore generates reliable shading and albedo estimates. This means less shading information is left in the albedo and the re-shading network learns to generate novel outputs rather than recovering the source illumination conditions from cues in the albedo.

As described in Section 4.1 we use gradient-based optimization with constraints to determine optimal lighting model parameters during inference. During training, this process is too slow and we therefore remove the constraints and solve the optimization with least-squares. We perform least squares using the normals and shading of the foreground region rather than the background. This is possible since the original image is already harmonized, so we know this will give us reliable lighting parameters. Additionally, this allows the model to learn to rely on the Lambertian shading provided as input which results in a controllable lighting refinement network.

All images used in the figures and experiments are sent through our model at 1024-pixel resolution. We found that this was the largest resolution where the off-the-shelf estimators provide consistent results, and our model is able to generate enough details while still being globally coherent across the foreground region.

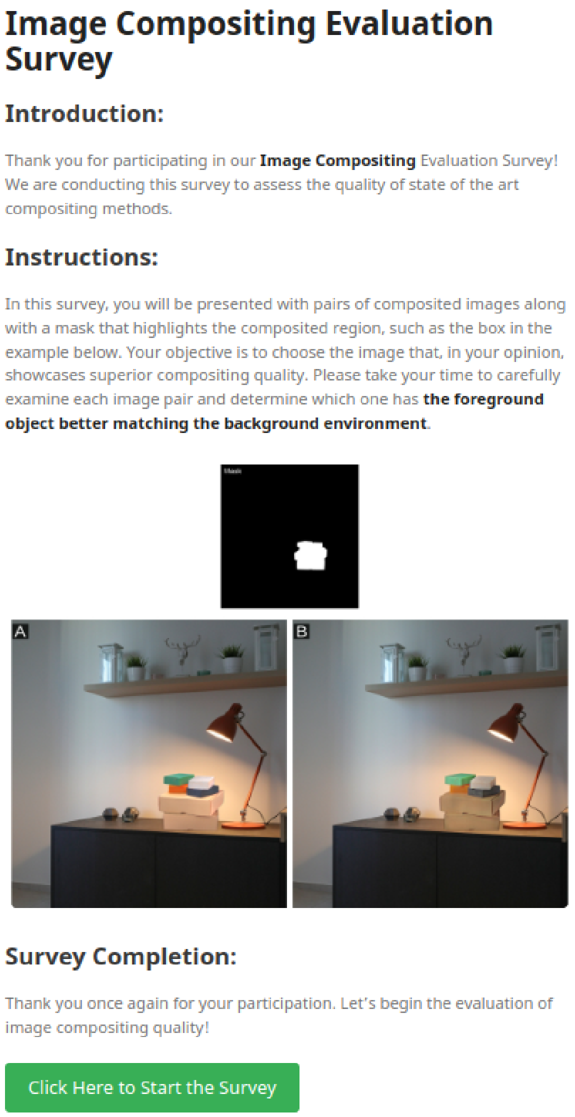

Image Compositing Evaluation Survey

Introduction:

Thank you for participating in our **Image Compositing** Evaluation Survey! We are conducting this survey to assess the quality of state of the art compositing methods.

Instructions:

In this survey, you will be presented with pairs of composited images along with a mask that highlights the composited region, such as the box in the example below. Your objective is to choose the image that, in your opinion, showcases superior compositing quality. Please take your time to carefully examine each image pair and determine which one has **the foreground object better matching the background environment**.

Survey Completion:

Thank you once again for your participation. Let's begin the evaluation of image compositing quality!

Figure 1: The instruction page of our user study. Image credit: Unsplash users Jean-Philippe Delberghe and Miguel Constantin Montes

SA Conference Papers '23, December 12–15, 2023, Sydney, NSW, Australia

ACM ISBN 979-8-4007-0315-7/23/12…$15.00
https://doi.org/10.1145/3610548.3618178

Table 1: Edit operation definition and the parameter ranges used to generate naive composite training images for the albedo matching network.

Operation	Formulation	Parameter range
Exposure	$I' = p_{exp} \cdot I$	$[0.5, 2]$
Saturation	$I' = RGB(h, p_{sat} \times s, v)$ h,s,v=HSV(I)	$[0, 2]$
Color curve	As defined by Hu et al. [2018]	$[0.5, 2]$
White balance	$I' = p_{wb} \cdot I$ (channel-wise)	$[0.3, 0.7]$

B ALBEDO HARMONIZATION

B.1 Training Data

Our albedo harmonization network aims to regress a set of parameters for common image editing operations. To generate our training data we start with the estimated albedo from a natural image and edit a masked region such that it creates a mismatch with the rest of the image. To achieve that we exploit common image editing operations, such as exposure, saturation, and color changes. We select the edit operations and set parameter ranges as summarized in Table 1.

We utilize MS-COCO [Lin et al. 2014] and Davis [Perazzi et al. 2016] dataset as they provide a segmentation mask for each of the objects present in the scene. During training, we randomly select one of the object masks in the scene. Next, we select a random number of edits (between 1-4), then an order for the edit operations, and values for each of the operations, sampled uniformly at random from the pre-specified ranges in Table 1 and apply the edit to the region specified by the mask. Some example training images are visualized in Figure 2.

B.2 Network Architecture

Given a naive composited albedo image and a region mask, our albedo harmonization network regresses multiple sets of parameters, one for each edit operation. We follow Miangoleh et al. [2023] and borrow their image editing network. They use their image editing network to regress image edits that increase or decrease the saliency of a specific region. We update the target ranges for the affine transforms of the estimation head of parameters in Miangoleh et al. [2023] to be $[0.1, 1]$ white balance, $[0, 2]$ saturation, $[0, 2]$ color curve values, and $[0.5, 2]$ exposure.

We train the network using the ADAM optimizer (learning rate of $1e-5$) for 100 epochs with a batch size of 64. A random ordering of the 4 edit operations (out of 24 possible permutations) is sampled for each training batch and provided to the networks as input during training. We also select the permutation at random during inference.

C USER STUDY DETAILS

A screenshot of our user study instruction page is included in Figure 1. We included the following instruction text in our user study, *"In this survey, you will be presented with pairs of composited images along with a mask that highlights the composited region, such as the box in the example below. Your objective is to choose the image that, in your opinion, showcases superior compositing quality. Please take your time to carefully examine each image pair and determine* ***which one has the foreground object better matching the background environment."***

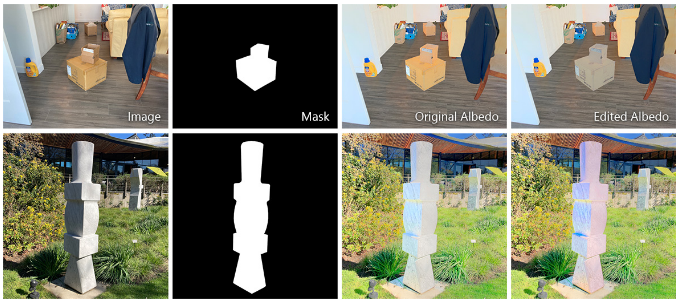

Figure 2: Example images edits used to train the Albedo Harmonization model. We randomly edit the original albedo to create a naive composite albedo that does not match its environment. We train the network to recover the original albedo therefore learning to match the foreground with the environment.

REFERENCES

Yuanming Hu, Hao He, Chenxi Xu, Baoyuan Wang, and Stephen Lin. 2018. Exposure: A White-Box Photo Post-Processing Framework. *ACM Trans. Graph.* 37, 2 (2018), 26.

Tsung-Yi Lin, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C Lawrence Zitnick. 2014. Microsoft COCO: Common Objects in Context. In *Proc. ECCV*.

S. Mahdi H. Miangoleh, Zoya Bylinskii, Eric Kee, Eli Shechtman, and Yağız Aksoy. 2023. Realistic Saliency Guided Image Enhancement. *Proc. CVPR*.

F. Perazzi, J. Pont-Tuset, B. McWilliams, L. Van Gool, M. Gross, and A. Sorkine-Hornung. 2016. A Benchmark Dataset and Evaluation Methodology for Video Object Segmentation. In *Proc. CVPR*.