Data-driven Hallucination of Different Times of Day from a Single Outdoor Photo Supplemental Document

Yichang Shih — MIT CSAIL
Sylvain Paris — Adobe
Frédo Durand — MIT CSAIL
William T. Freeman — MIT CSAIL

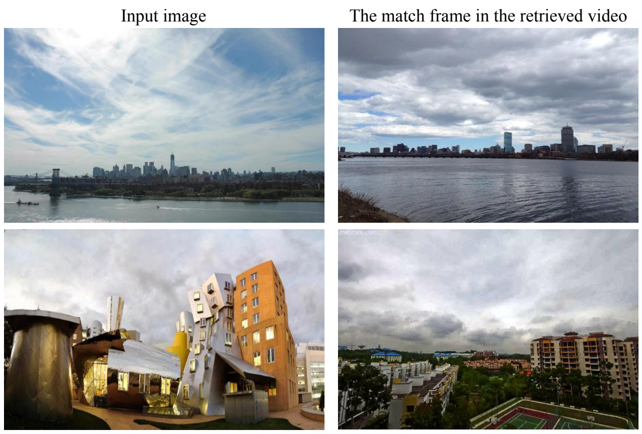

Figure 1: *Video retrieval and match frame results under two different input scenes.*

1 Description of this document

Time-lapse video retrieval and the match frame Figure 1 illustrates the time-lapse video retrieval results and the match frames. The retrieval is based on a standard scene matching technique [Xiao et al. 2010] (Section 5.1) and color statistics (Section 5.1.1).

Locally linear vs affine Figure 2 compares the choice of locally affine model and linear model. Similar to expressivity test, we hallucinate from one input frame to another ground truth frame in a single time-lapse video. We perform the transfer with locally linear and affine model. The difference between the output and the ground truth shows that affine model yields better result.

Expressivity of locally affine transfer Figure 3 illustrates the expressivity of locally affine model under various scenes (Section 6.1), including harbor, lake, skyline, river side. As described in Section 6.1, we take a frame from a time-lapse video as input, and another frame as ground truth. We hallucinate the input to the ground truth frame using the same time-lapse video. The output is visually close to the ground truth, even the lighting between the ground truth and the input frame is very different.

Compare to Deep photo In Figure 5, we compare our results to Deep Photo [Kopf et al. 2008], which uses scene 3D information to relight the image (Section 7.1). We use the input and the result relit at dusk on their web-site. For comparison, we hallucinate the input to "golden hour". Both results are plausible, but we don't need scene-specific data.

Compare to Laffont et al. In Figure 4, we compare our results to Laffont et al [2012], which uses a collection of photos under the same scene for illumination transfer (Section 7.1). They decompose the image into intrinsic and illumination, and then transfer the illumination from one image to another image. In this experiment, they used 17 images for decomposition, and transfer the illumination from a photo under faint light. For comparison, we hallucinate the input to "blue hour". Again, both are plausible, but we only require a single input photo.

2 Accompanying image

We show a montage of our time-lapse videos in montage.jpg (Section 4). For each video, we select one frame at a random time.

3 Accompanying video

We show synthetic time-lapse in the video (Section 7.2). We generate results at different times from a single input, and then linearly interpolate these results to simulate a time-lapse video.

4 Accompanying web page

We show our evaluation on MIT-Adobe 5k dataset [Bychkovsky et al. 2011]. (Section 7)

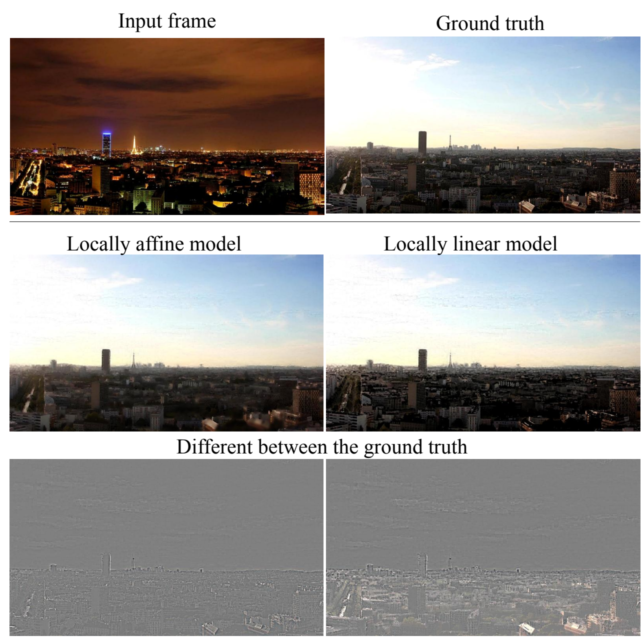

Figure 2: *We show locally affine model is a better choice than linear model. We hallucinate the input to another frame (ground truth) in the same time-lapse video with two different models. The affine model is closer to the ground truth.*

References

BYCHKOVSKY, V., PARIS, S., CHAN, E., AND DURAND, F. 2011. Learning photographic global tonal adjustment with a database

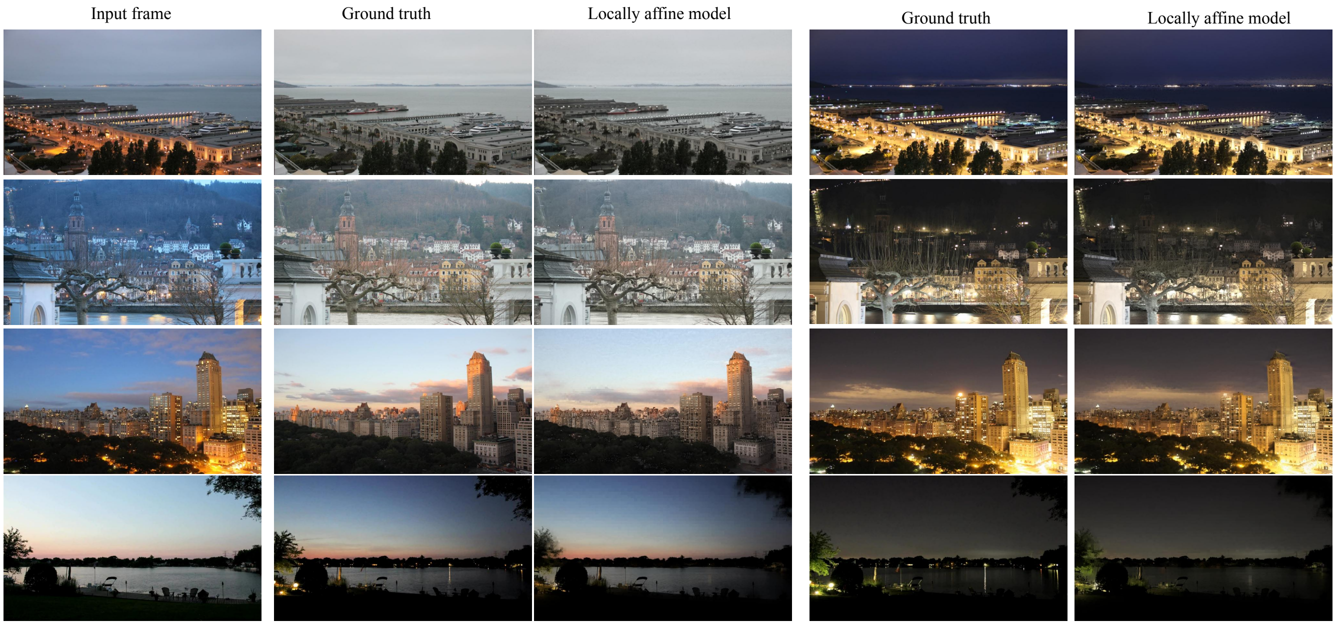

Figure 3: *Locally affine model is expressive enough to model different times of a day. For each row, we pick up a frame from a time-lapse video as input. We choose another ground truth frame from the same time-lapse video as input, and produce the result using our model. Our result is very close to the ground truth and shows our model is expressive for time hallucinations even lighting between input and ground truth is very different.*

of input/output image pairs. In *IEEE Conference on Computer Vision and Pattern Recognition*, 97–104.

KOPF, J., NEUBERT, B., CHEN, B., COHEN, M., COHEN-OR, D., DEUSSEN, O., UYTTENDAELE, M., AND LISCHINSKI, D. 2008. Deep photo: Model-based photograph enhancement and viewing. In *ACM Transactions on Graphics (SIGGRAPH)*, vol. 27, 116.

LAFFONT, P.-Y., BOUSSEAU, A., PARIS, S., DURAND, F., DRETTAKIS, G., ET AL. 2012. Coherent intrinsic images from photo collections. *ACM Transactions on Graphics 31*, 6.

XIAO, J., HAYS, J., EHINGER, K., OLIVA, A., AND TORRALBA, A. 2010. Sun database: Large-scale scene recognition from abbey to zoo. In *IEEE conference on Computer Vision and Pattern Recognition*, 3485–3492.

Input image
(from Laffont et al's paper)

Laffont et al.
(using 17 images)

Our method
(from a single input image)

Figure 4: *Laffont et al. use multiple images at the same scene for intrinsic image decomposition, and then relight the image by transferring illumination to the intrinsic image. We use different data for relighting. We hallucinate the input to "blue hour" to match their result. Laffont's result is directly from their website.*

Input
(from Deep Photo paper)

Deep Photo
(using scene geometry)

Our method
(from single input)

Figure 5: *Deep photo leverages depth map and texture of the scene to relight an image. Our method uses less information and produces plausible looking results. We hallucinate the input to "golden hour" to match their result. We use results directly from Deep Photo project website.*