Light-field Appearance Editing Based on Intrinsic Decomposition (Supplemental Material)

I. Overview

In the supplementary materials document, we follow the structure of the main paper. We start by discussing our intrinsic decomposition framework in Sec. II. We analyze the impact of different parameter values in Sec. II-A. We later talk about angular consistency and how we extend our decomposition framework for dense light-fields in Sec. II-B and II-C. After an analysis of our intrinsic decomposition methodology, we proceed with the discussion about our appearance editing framework in Sec. III. At the end, we present additional results for our complete appearance editing framework, using ground truth intrinsic layers and our intrinsic decomposition results in Sec. IV.

II. Intrinsic Decompostion

In this section, we discuss the aspects related to intrinsic decomposition of wide-baseline light-fields. We revisit the topics discussed in the method section of the main paper and explain some of the specific concepts in more detail. We start by discussing the significance of parameter weights used in our optimization scheme followed by a discussion on the consistency of our results and the extension of our method to dense light-fields. In the last sub-section we produce additional intrinsic decomposition results.

A. Parameter Analysis

In our intrinsic decomposition method, we use the following optimization model:

$$\arg\min_{r} E(r) = \lambda_r E_r(r) + \lambda_d E_d(r) + \lambda_a E_a(r) + \lambda_s ||r||_1, \tag{1}$$

where E_r, E_d, and E_a are retinex, disparity, and absolute shading scale terms, respectively with their corresponding weights, and λ_s is a regularization parameter to ensure reflectance sparsity. Please refer to the main paper for a detailed explanation of each term in the optimization.

We use fixed weights $\lambda_r = 2$, $\lambda_d = 1$, $\lambda_a = 0.7$, and $\lambda_s = 0.1$ for all results presented in this paper. In this section, we further analyze the range of weights for the individual terms in our optimization scheme. We provide additional results, in this regard, for better understanding. We use three different scene types for this analysis (see Fig. 1). To evaluate the impact of a given weight, we modify its value while the remaining weights have their default values as mentioned above.

The weight parameter λ_r controls the influence of retinex term in the optimization. The retinex term allows shading and reflectance smoothing; however, the reflectance smoothing is inhibited by the edge weights. As a result, a higher value of λ_r results in smoothed shading. By definition, we prefer smooth shading except at occlusion boundaries; however, excessive smoothing leads to loss of geometrical details which otherwise should be captured in shading layer (see Fig. 2).

The weight parameter λ_a restricts shading to be close to absolute gray. A lower value of λ_a allows significant variation in terms of gray value for shading. In such a case, the shading layer is quite close to the grey version of the input image. On the other hand, a higher value of λ_a makes shading less informative by forcing all pixels to be of the same gray value. An ideal gray-scale shading should not have much gray-value variation, but at the same time gray value changes representing changes in illumination and geometry should be preserved (see Fig. 3).

The weight λ_s controls the degree of L_1 regularization applied on reflectance. By applying L_1 regularization, we encourage sparsity in the reflectance layer. A lower value of λ_s gives a reflectance which is less sparse and includes more of shading information. A higher value of λ_s gives a reflectance which is more flat and contains less shading information (see Fig. 4).

The disparity term ensures consistency between intrinsic layers of different views in the light-field. The weight λ_d controls the degree of such consistency application. A higher value for λ_d introduces artifacts by making the occlusion boundaries more smooth (see Fig. 5).

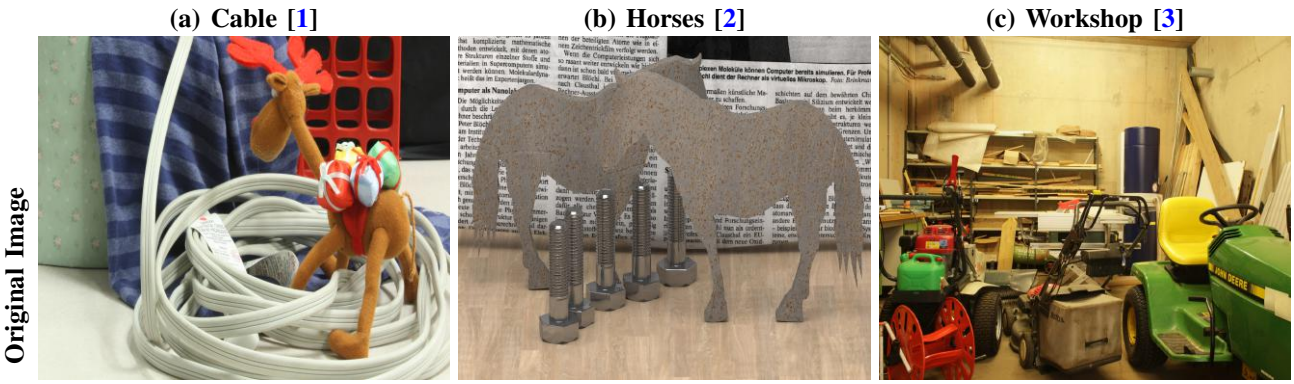

Fig. 1: Original images from, (a) multi-view stereo data (b) small-baseline light-field and (c) wide-baseline lightfield on which we show the effect of different parameter values.

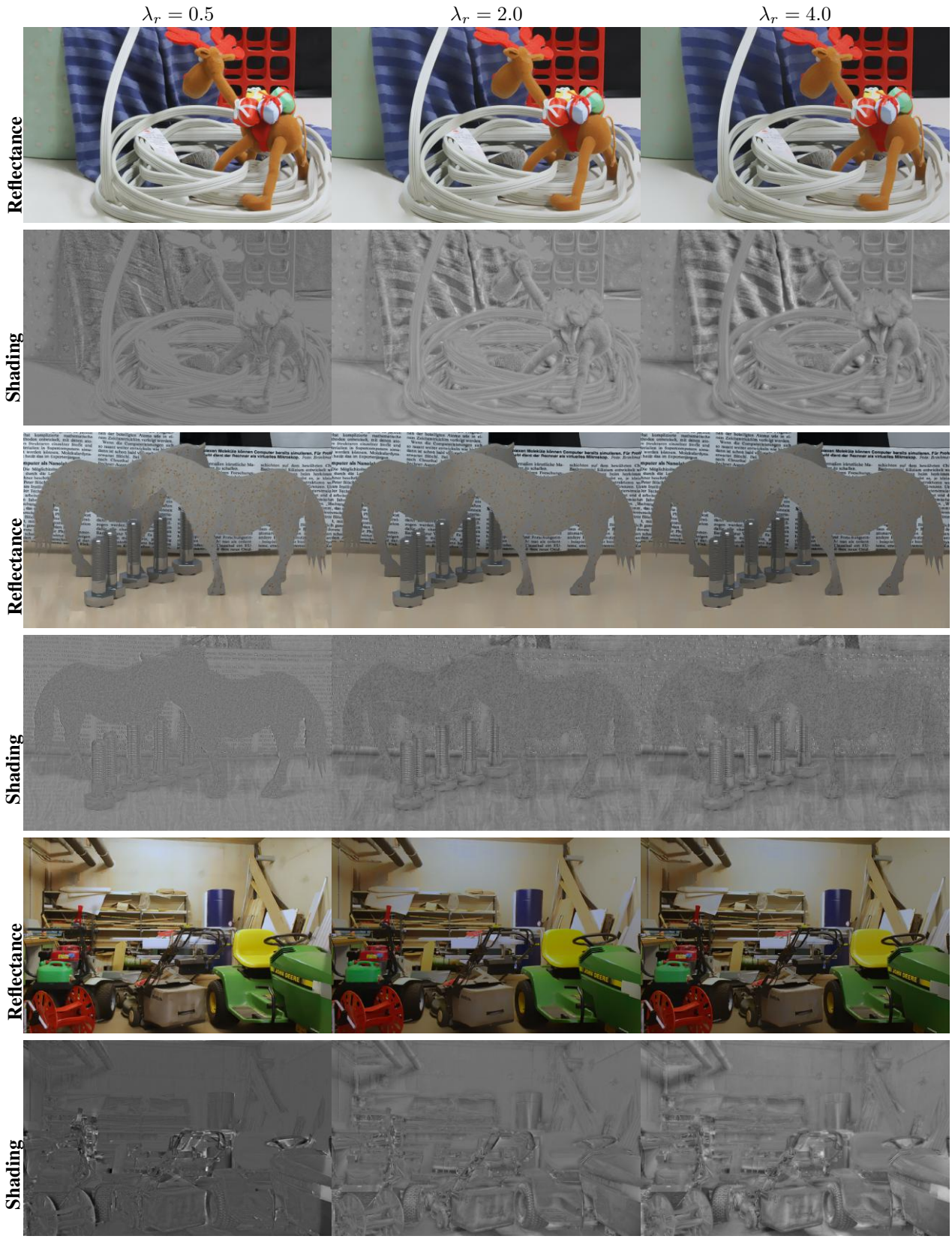

Fig. 2: The effect of λ_r, controlling the significance of *retinex term* on intrinsic decomposition results.

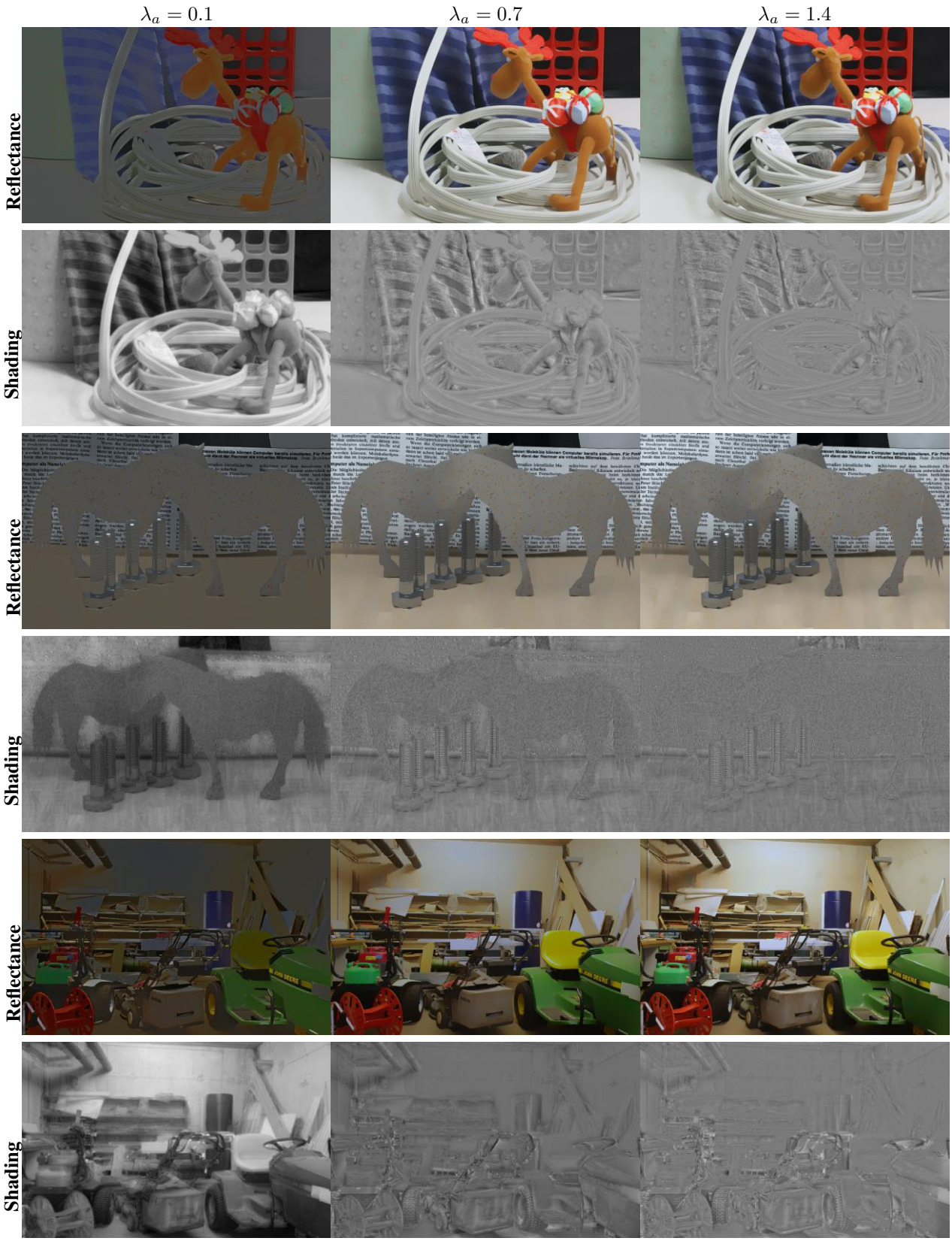

Fig. 3: The effect of λ_a, controlling the significance of *absolute scale term* on intrinsic decomposition results.

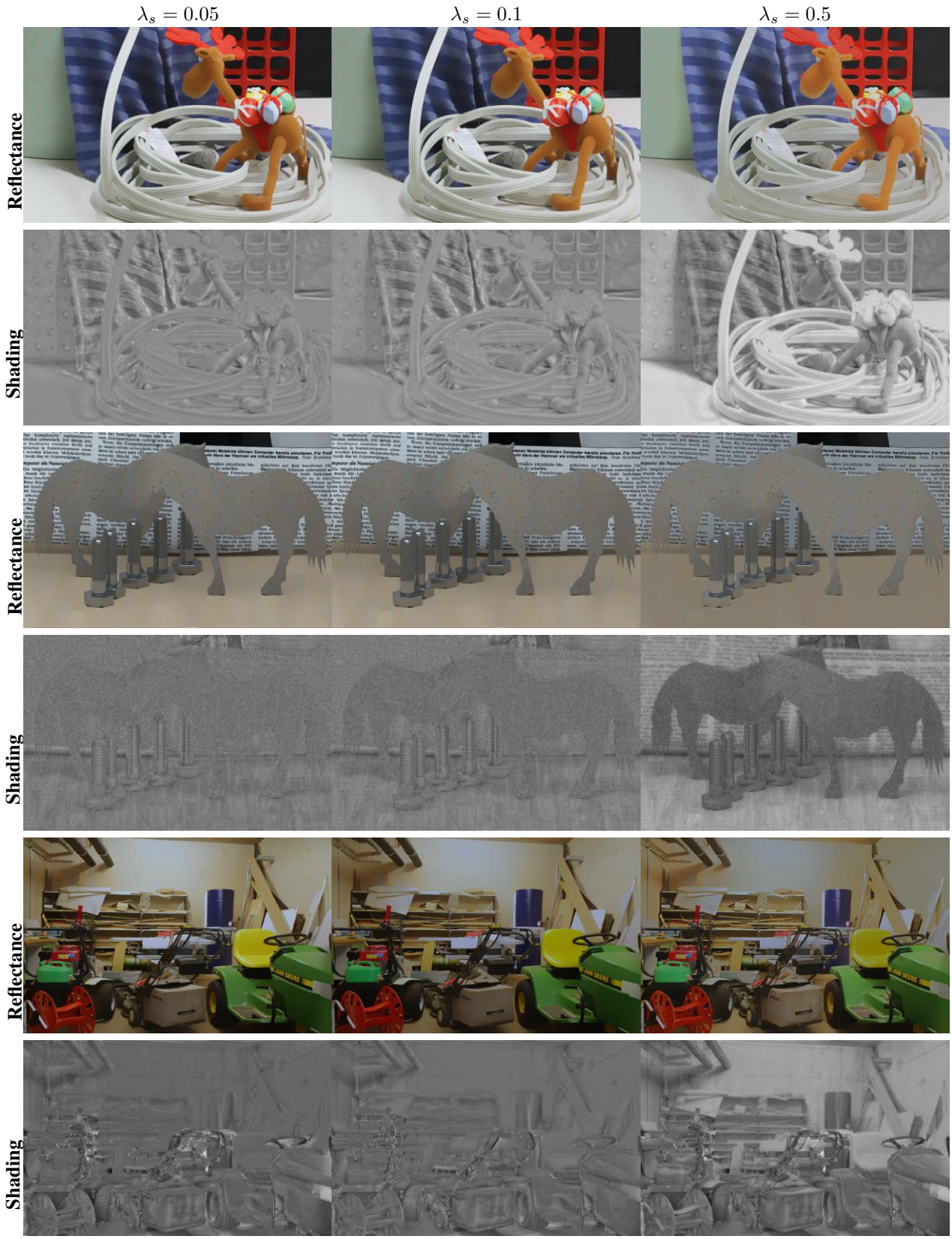

Fig. 4: The effect of λ_s, controlling the amount of L_1 *regularization* of reflectance in intrinsic decomposition.

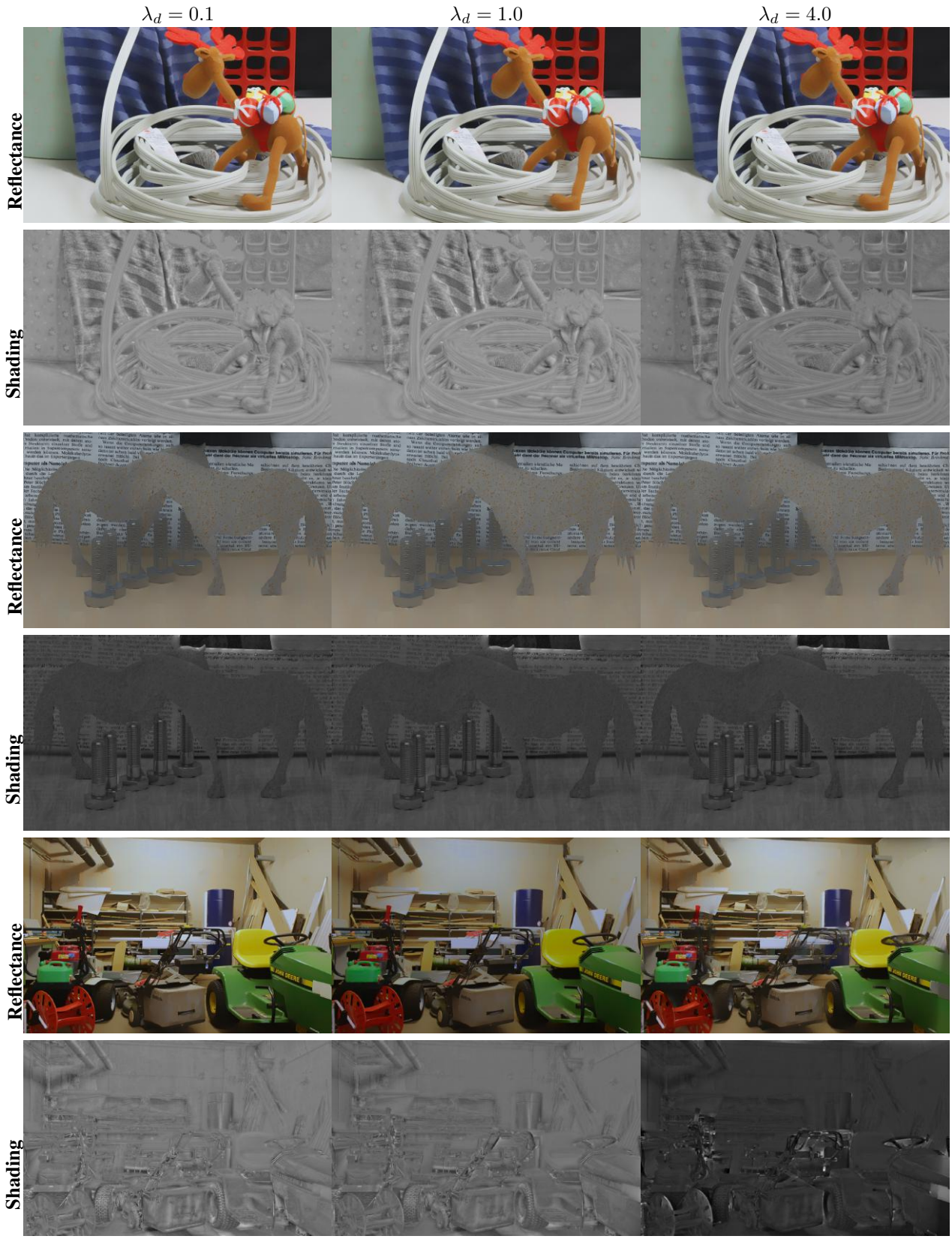

Fig. 5: The effect of λ_d, controlling the significance of *disparity term* on intrinsic decomposition results.

B. Intrinsic Decomposition Consistency

Our intrinsic decomposition algorithm produces consist result for the sparse light-field input. Please note that this consistency is present in both spatial and angular domain. In the main paper, we have presented consistent reflectance result. We further

add to that by presenting consistent results for both shading and reflectance for the given scenes (see Fig. 6 and 7). Consistent result for multi-view scenario is also shown in Fig. 12

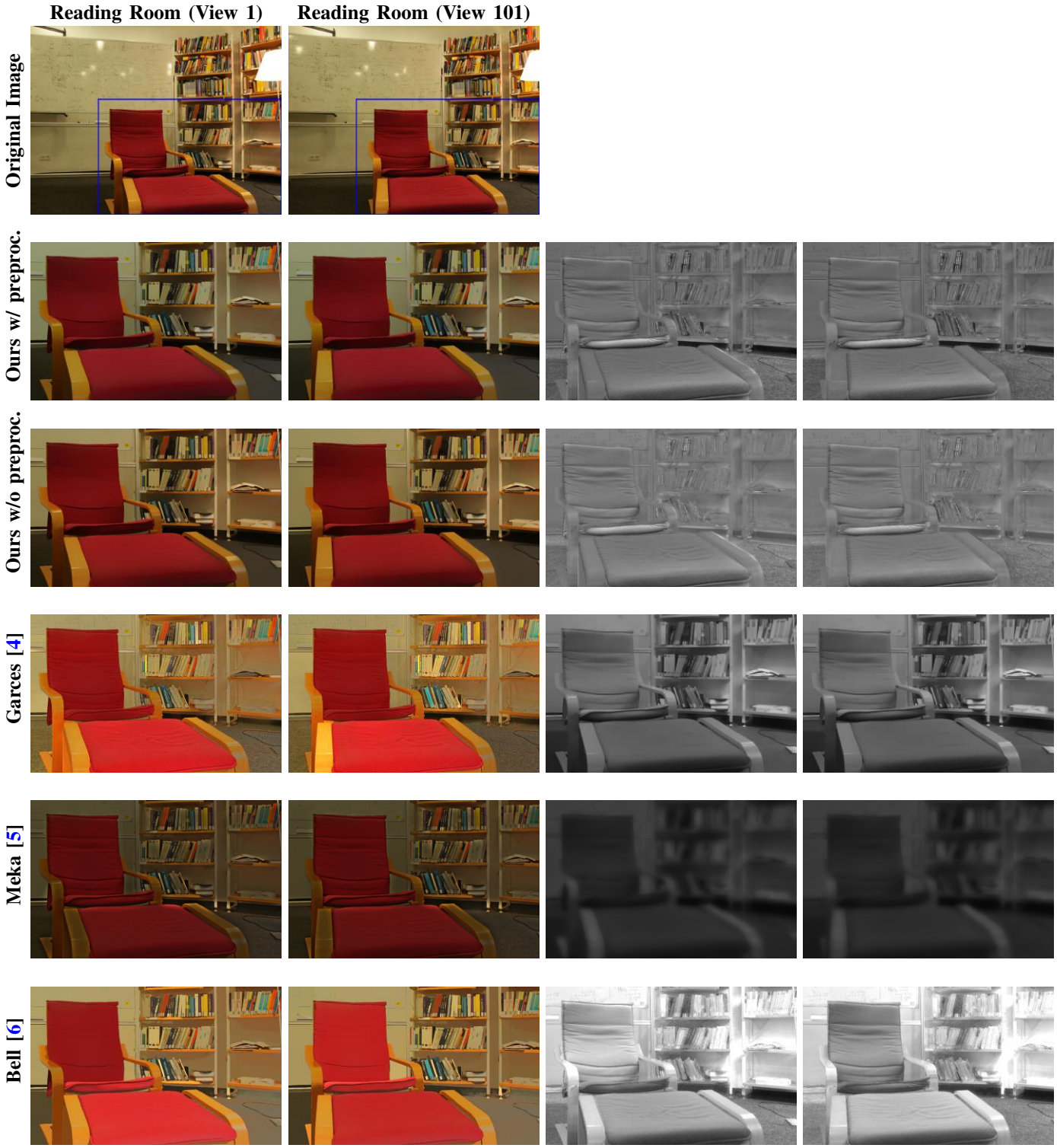

Fig. 6: Comparison of the reflectance and shading layer extraction using our (with/without the pre-processing step consisting of specularity removal and white-balancing), Garces et al., Meka et al., and Bell et al. methods. Wide-baseline light-field of *Reading Room*, composed of 101 views (1 cm baseline between neighboring views) have been used [3] where only the first and last views are shown. As Garces et al. and Meka et al. rely on dense data, all 101 views are used. But for our method, we subsample only 11 equally spaced views (10 cm baseline). For Bell et al. we perform decomposition separately on each view. While no method is perfect, our reflectance results contain much less shading, specularity, and no patch-like artifacts. Our results are consistent not only between views, but also within each view (e.g, the red chair).

Fig. 7: Comparison of the reflectance and shading layer extraction using our (with/without the pre-processing step consisting of specularity removal and white-balancing), Garces et al., Meka et al., and Bell et al. methods. Wide-baseline light-field of *Workshop*, composed of 101 views (1 cm baseline between neighboring views) have been used [3] where only the first and last views are shown. As Garces et al. and Meka et al. rely on dense data, all 101 views are used. But for our method, we subsample only 11 equally spaced views (10 cm baseline). For Bell et al. we perform decomposition separately on each view. While no method is perfect, our reflectance results contain much less shading, specularity, and no patch-like artifacts. Our results are consistent not only between views, but also within each view.

C. Consistent Propagation of Sparse Reflectance

Our intrinsic decomposition method is suitable for wide baseline sparse light-field. However, we can extend this decomposition framework even in case of a wide baseline dense light-field. The approach is inspired from [7] where they use this technique for consistent video filtering for multiple camera arrays. The idea is to do intrinsic decomposition of only few sparse views and then propagate this result to the rest of light-field (see Fig. 8).

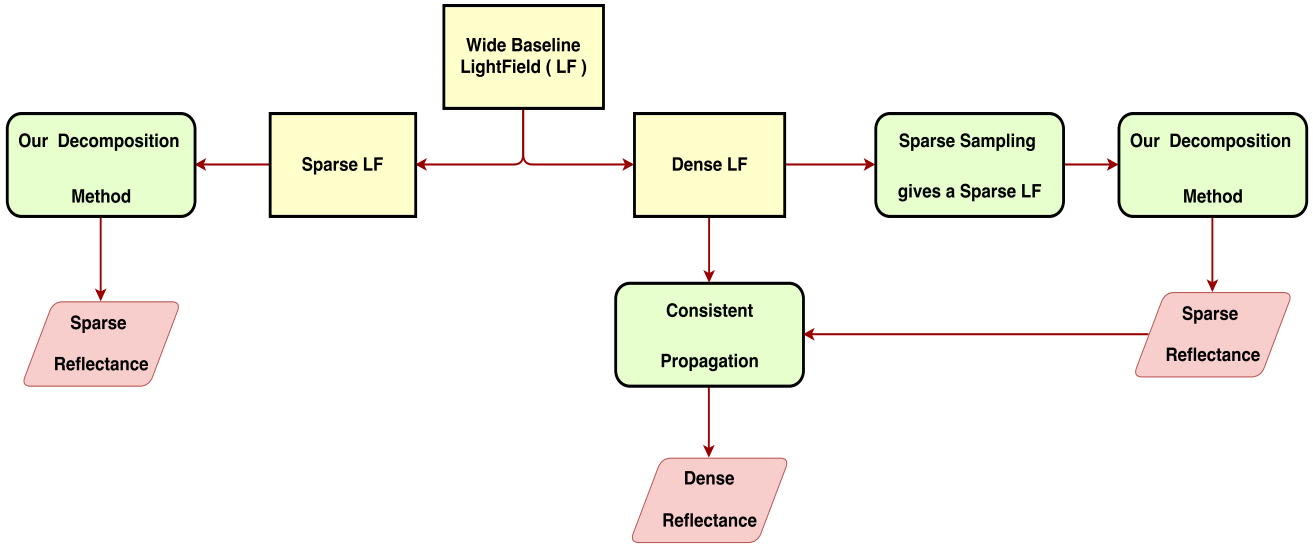

Fig. 8: Flowchart representing our intrinsic decomposition framework for wide baseline light-field. In all the cases, shading is generated using the image formation model discussed in the main paper.

The first step in this case is to perform a sparse sampling. The intrinsic decomposition of sparse samples is described in the method section of the main paper. The sparse reflectance is propagated among all the views to get a dense reflectance output. We extract the dense shading by using the image model described in method section of the main paper. In this section, we describe the methodology used for propagation of sparse reflectance. We first describe the individual input elements and the energy formulation. It is followed by explanation of individual terms in the energy. At the end, we describe how our energy formulation is different from the one used in [7].

1) Energy Formulation for Reflectance Propagation: Let the input views at position a, b and c be represented by I_a, I_b and I_c respectively, where b is an intermediate position. Let's consider that we have sparse reflectance for views at position a and c, given by R_a and R_c respectively. If we find reflectance at b, say R_b, then we can generalize this approach and thus generate a dense reflectance output (see Fig. 9).

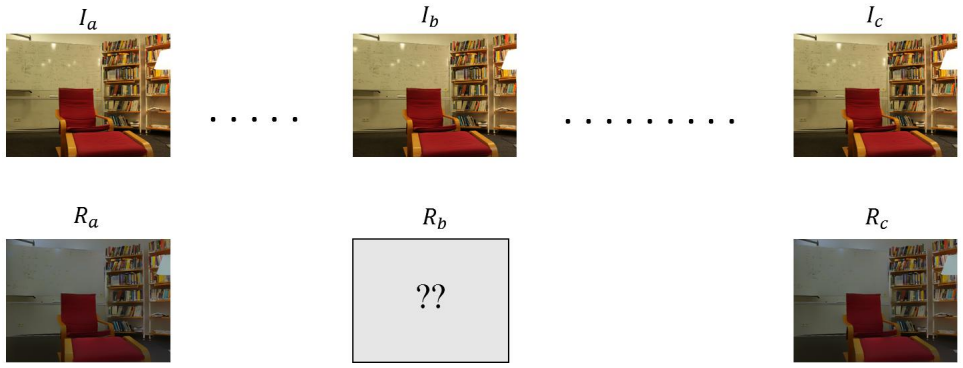

Fig. 9: A pictorial representation showing the input views I_a, I_b and I_c and their corresponding reflectance R_a, R_b and R_c.

An energy minimization scheme to obtain R_b is given as:

$$E(R_b) = \int_{\Omega} (1 - \zeta_{mn})(||\nabla R_b - \nabla I_b||^2) + w_a(||R_b - T_{ab}(R_a)||^2) + w_c(||R_b - T_{cb}(R_c)||^2) \tag{2}$$

where $1-\zeta_{mn}$ are weights representing reflectance edges and w_a and w_c are defined as:

$$w_a = \lambda_a \exp(-\alpha_a ||I_b - T_{ab}(I_a)||^2) \tag{3}$$

$$w_c = \lambda_c \exp(-\alpha_c ||I_b - T_{cb}(I_c)||^2) \tag{4}$$

where T_{ab} is a warp operator that maps view I_a to I_b and T_{cb} is a warp operator that maps view I_c to I_b. The mapped views are given as $T_{ab}(I_a)$ and $T_{cb}(I_c)$, respectively. We use the same operator to map reflectance R_a and R_c to obtain $T_{ab}(R_a)$ and $T_{cb}(R_c)$, respectively. We use simple gradient descent method to find a minimizer for the above energy functional. The energy formulation is based on the idea introduced in [8] and further explored in [7].

2) Energy Terms Significance: The first term, $(1-\zeta_{mn})(||\nabla R_b - \nabla I_b||^2)$ enforces similarity between the gradient of input view I_b and its corresponding unknown reflectance R_b, the similarity is strongly enforced only at reflectance edges. The weight $(1-\zeta_{mn})$ varies between 0 and 1, where values are close to 1 at reflectance edges. An explaination of how these edges are identified is provided in the method section of the main paper.

The second and third terms, $w_a(||R_b - T_{ab}(R_a)||^2)$ and $w_c(||R_b - T_{cb}(R_c)||^2)$, enforce similarity between the unknown reflectance R_b and the mapped neighboring reflectance in pixel value domain. The weights w_a and w_b relax this similarity where mapping is not good enough.

3) Our Contribution in Energy Formulation: The energy formulation proposed above is inspired from the filter transfer for multi camera [7]. However, we have modified this energy for our specific needs.

The first modification is introduced in the form of a weight, $(1-\zeta_{mn})$ for the first term. By introducing such a weight, reflectance is enhanced while it is being propagated to the dense views (see Figure 10). The second modification is making use of two neighboring views, in the angular domain, and their sparse reflectance in the energy formulation. In [7], the authors consider only one previous view in temporal or angular domain. In our case of wide baseline light-field, we consider neighborhood only in the angular domain. By making use of two nearest neighbors, we ensure better consistency and also a faster convergence for energy minimization. One drawback of using two neighbors is to increase the memory overhead. However, for our purpose the advantages in terms of consistency matters more.

Fig. 10: Comparing reflectance propagation for an intermediate view with and without reflectance edge weights (discussed in Sec. II-C1). In the image area marked by red square, we observe that small textures are better preserved when we use reflectance edge weights.

D. Additional Results for Intrinsic Decomposition

We can also generate intrinsic decomposition results for single images by disabling the energy terms specific for multi-view scenario, namely *disparity* and *occlusion* terms (see Fig. 11).

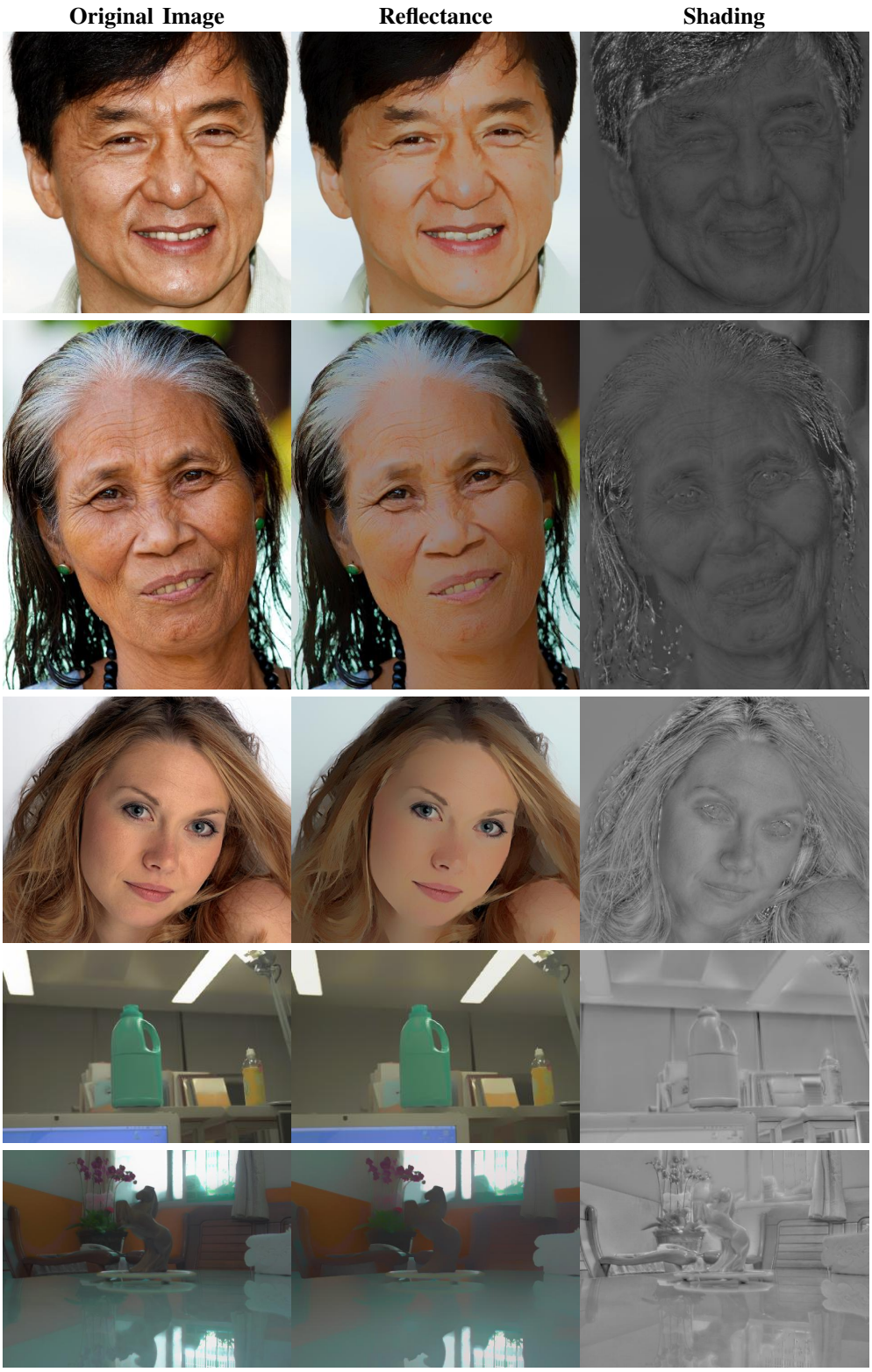

Fig. 11: Intrinsic decomposition results for objects (source [9]) and faces (source [10]).

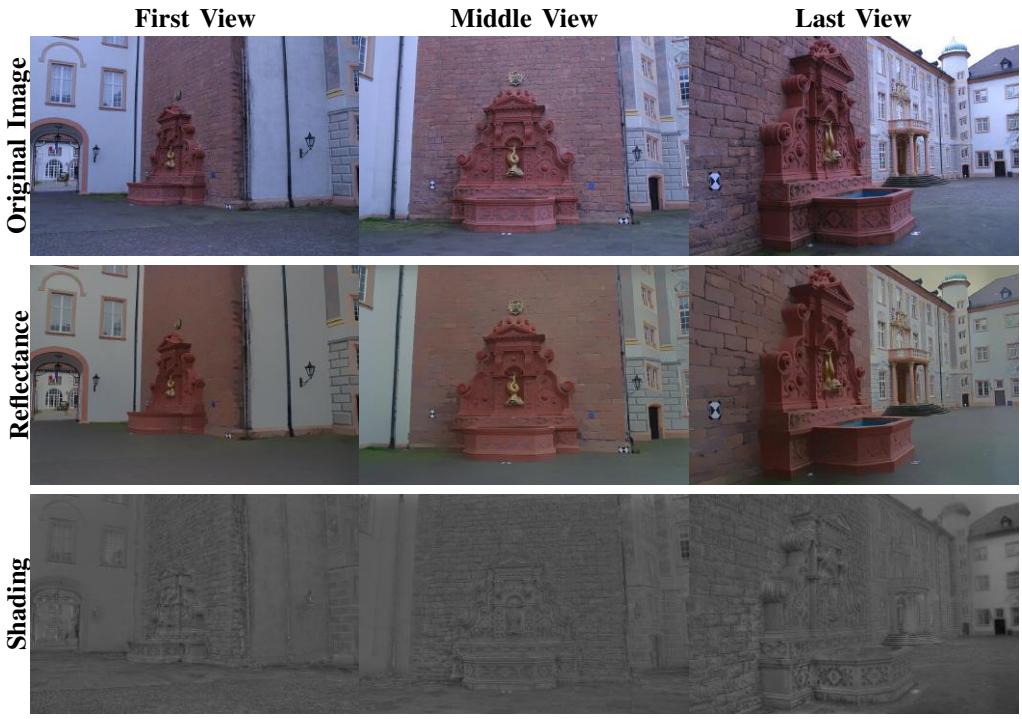

Fig. 12: Consistent intrinsic decomposition results for first, last and middle view of fountain-P11 scene [11].

III. Appearance Editing

In this section, we discuss aspects related to our appearance editing framework. In the main paper, we have explained as to how we achieve different types of appearance editing by sifting intrinsic layers. Here, we add more results supporting our claim that our range of appearance editing is broader than that of [10].

A. Band-Sifting Bounds Analysis

We discuss the range of original band-sift filtering using luminance channel as compared to our sifting based filtering using intrinsic layers. We do this comparison with respect to three different types of material edits, namely: *oily/metallic, shiny glow* and *weathering*

In [10], *oily/metallic* look is achieved by sifting the *luminance* channel of original image using operation: $L(HHP, \kappa_l)$. On the other hand, in case of intrinsic layers we sift the *intensity* of specular layer using operation: $C(HHP, \kappa_c)$. The *shiny glow* look is achieved in [10] by sifting the *luminance* channel of original image using operation: $L(LAP, \kappa_l)$, in our work we sift the *intensity* of specular layer using operation: $C(LAP, \kappa_c)$. In both the case κ_l and κ_c are positive constants greater than 1. In Fig. 13 and 14, we show how the range of editing is increased by using our framework.

In [10], the authors use sifting operation: $L(HLA, \kappa_l)$ for *weathering* effect. In this case, we are more different from the original sifting paper, as we make use of multiple intrinsic layers to achieve similar effect. The idea is to reduce the color and shininess of object by reducing the *intensity* of *reflectance* and *specular layer* and by enhancing the small wrinkles or bumps in the shading layer. The sifting operation for the above is represented by $S(HLA, \kappa_s)$, $R(AAA, \kappa_r)$ and $C(AAA, \kappa_c)$, where $\kappa_s > 1$ and $0 < \kappa_r, \kappa_c < 1$ (see Figure 15).

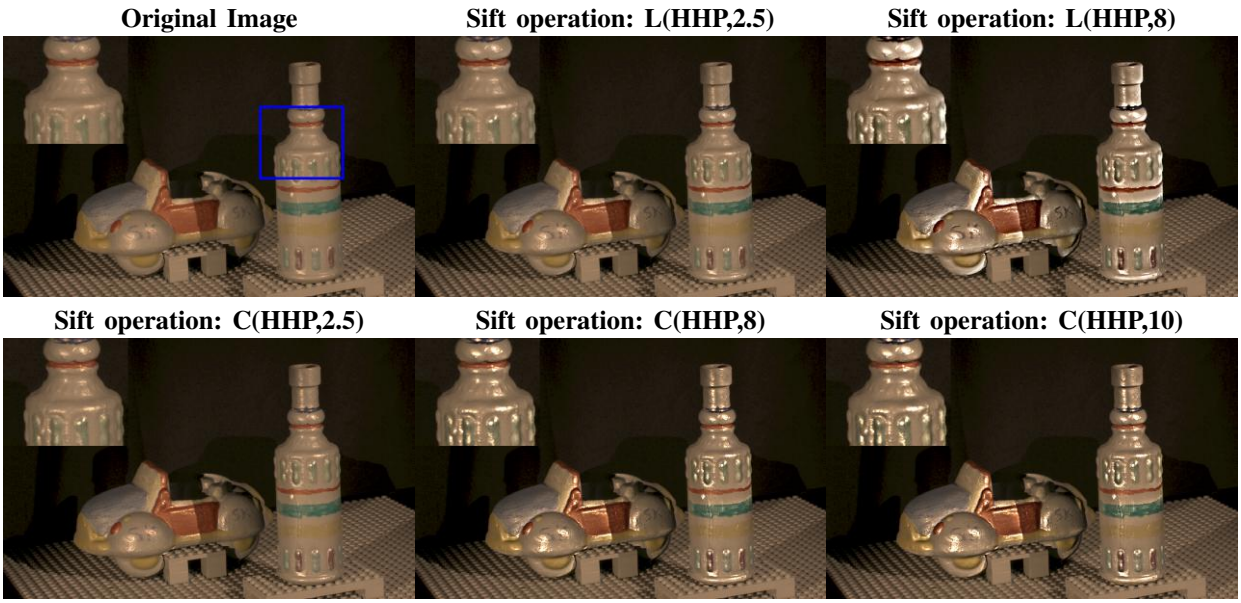

Fig. 13: The oily/metallic effect produced by original band-sifting and our editing framework. Please note that in both cases we start observing visible edits for a factor of 2.5. However we start observing unnatural editing artifacts in case of luminance channel for a κ_l of 8. In case of using specularity we have a larger range to obtain this effect without causing implausible look. The top left of each image shows a zoomed in version of the region marked by the blue rectangle in the original image.

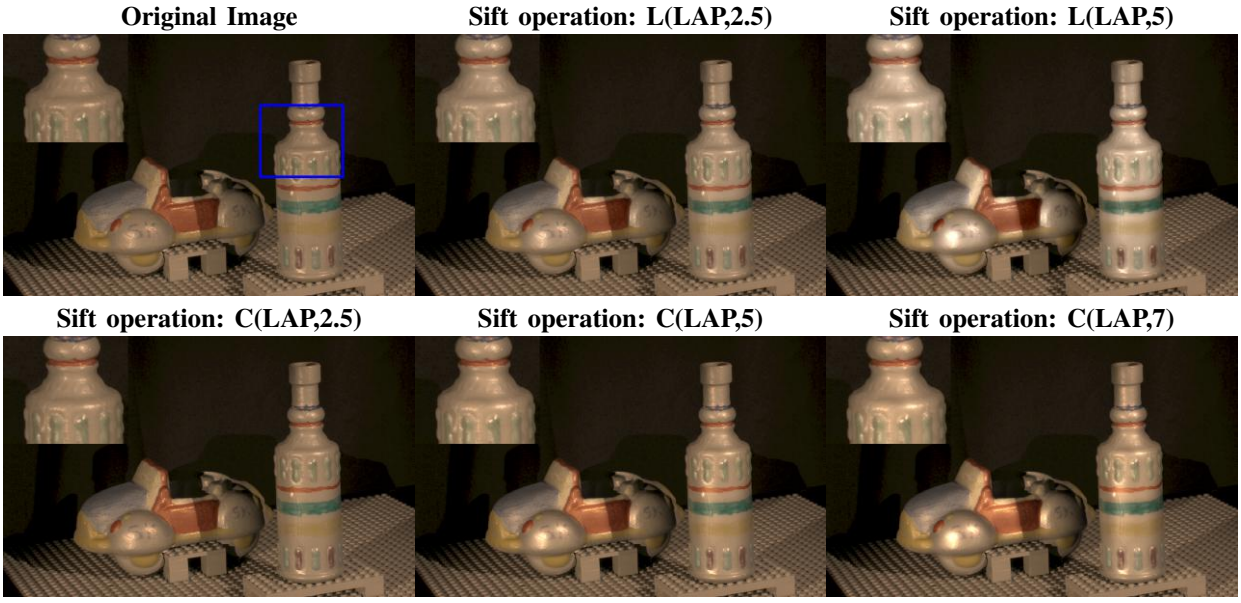

Fig. 14: The shiny glow effect produced by original band-sifting and our editing framework. Please note that in both cases we start observing visible edits for a factor of 2.5. However we start observing unnatural editing artifacts in case of luminance channel for a κ_l of 5. In case of using specularity we have a larger range to obtain this effect without causing implausible look. The top left of each image shows a zoomed in version of the region marked by the blue rectangle in the original image.

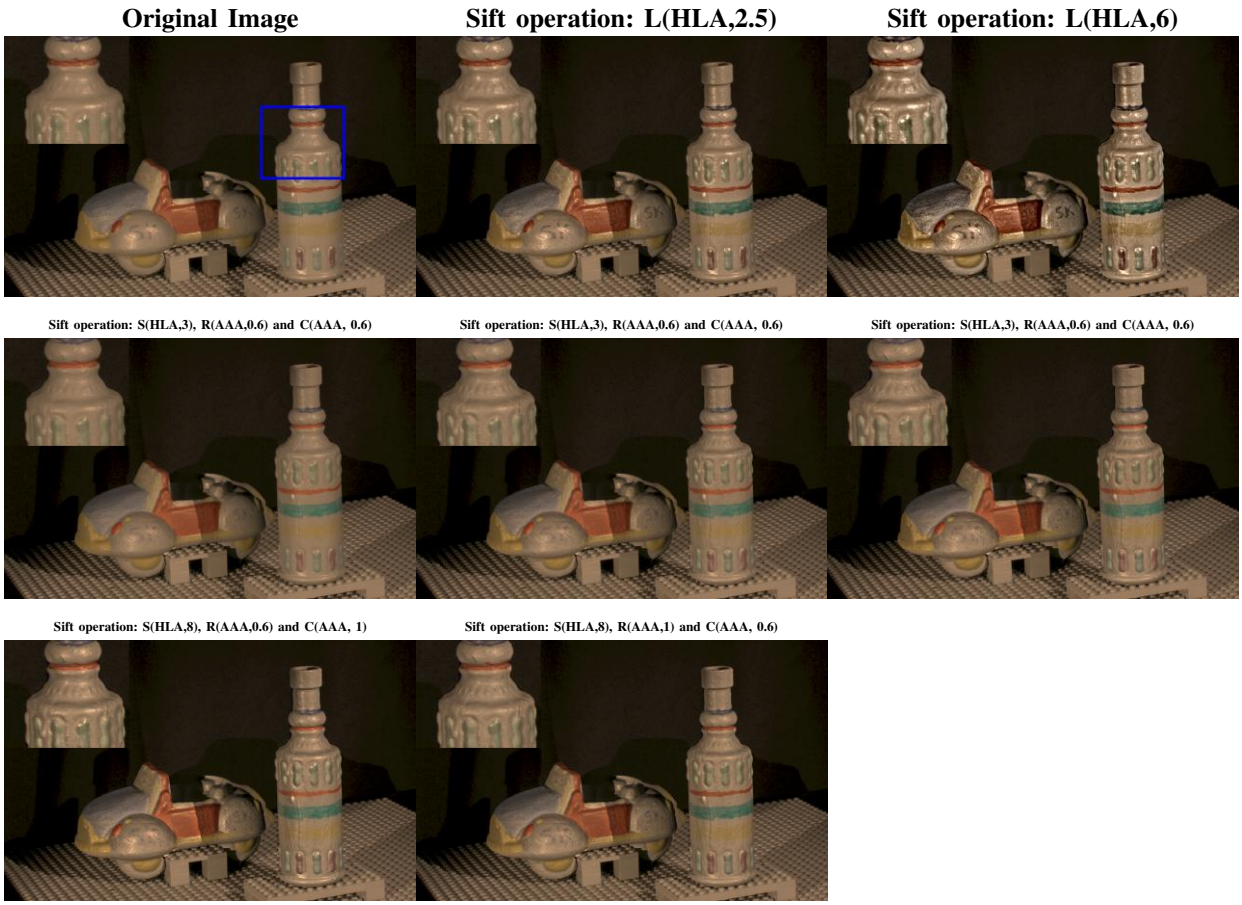

Fig. 15: The weathering effect produced by original band-sifting and our editing framework. In the original sifting we start observing visible changes for κ_l of 2.5. We observe unnatural artifacts for κ_l of 6 and above. In case of our work we have three ranges to modify for each intrinsic layer of shading, reflectance and specularity respectively. We not only achieve a greater range for editing in case of shading layer, but we have more dimensions to edit, considering the additional range of reflectance and specularity. The top left of each image shows a zoomed in version of the region marked by the blue rectangle in the original image.

IV. Additional Results - Full Framework

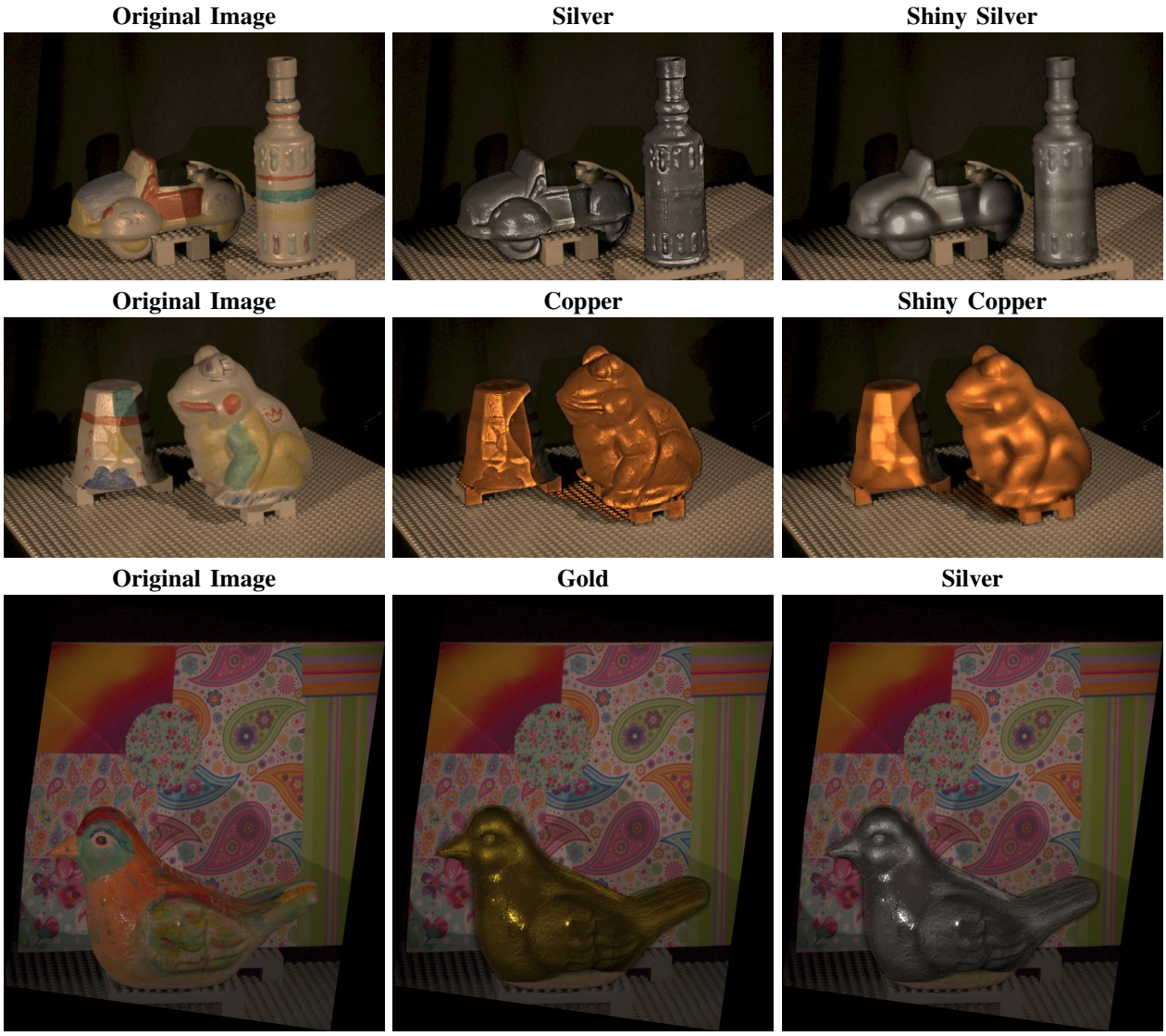

Fig. 16: More results for the Gold, Silver and Copper material edits using ground truth intrinsic layers from [12] and [13].

Fig. 17: Comparison of making a human face old using our framework with that of the original band-sifting [10]. Image Source [10].

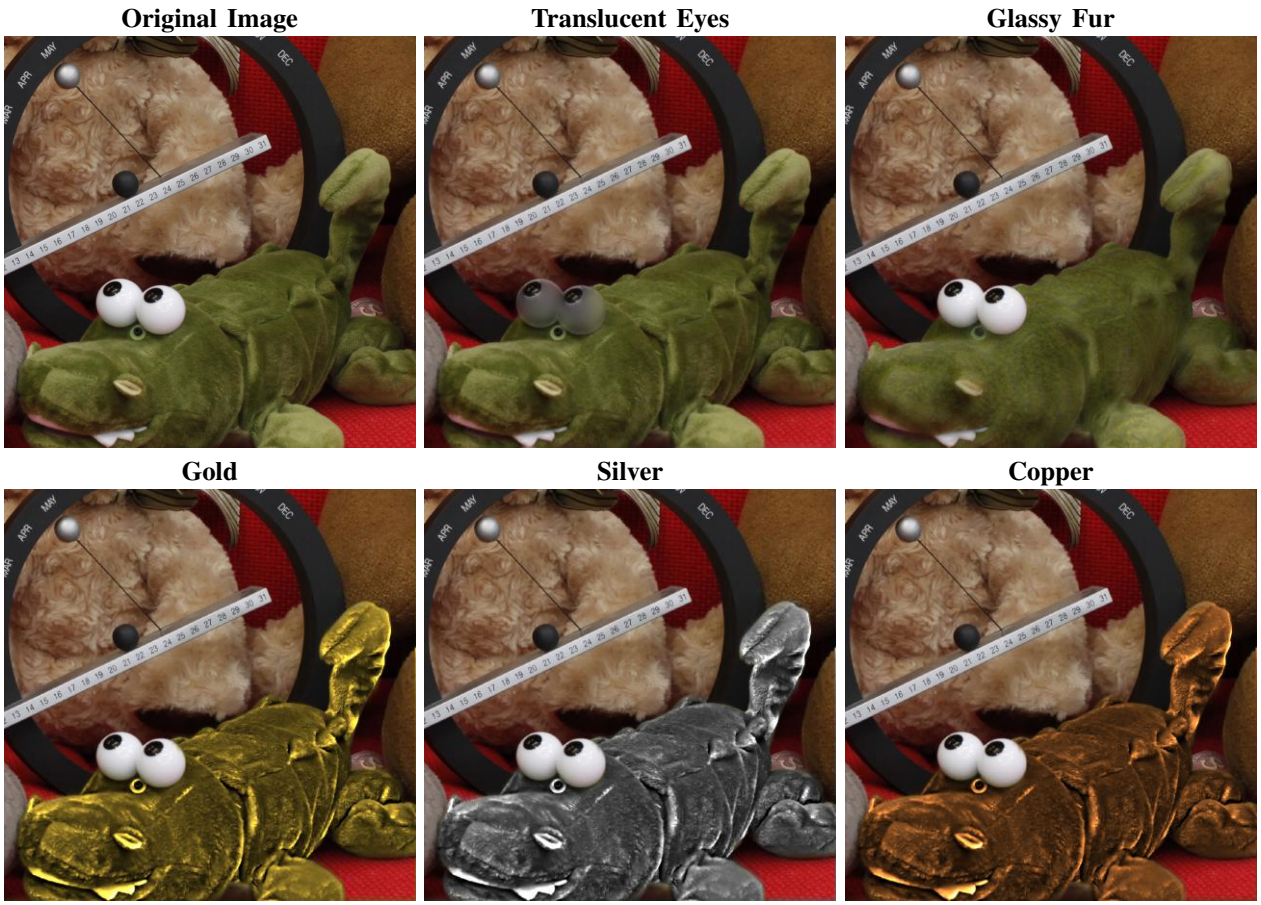

Fig. 18: More results for cropped version of *Couch* scene from Disney dataset and the mask used to apply these effects.

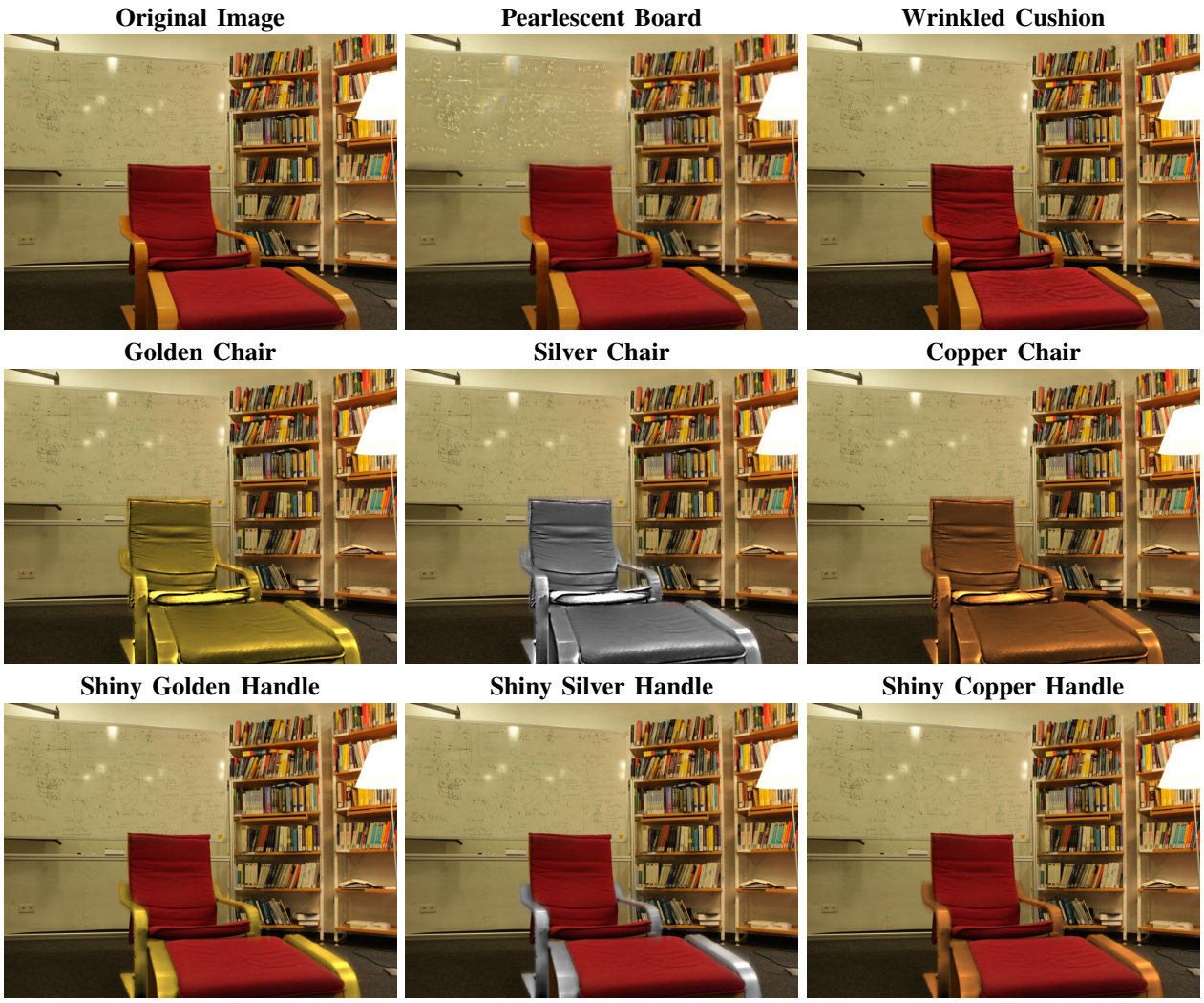

Fig. 19: More results for the *Reading Room* scene from [3] dataset.

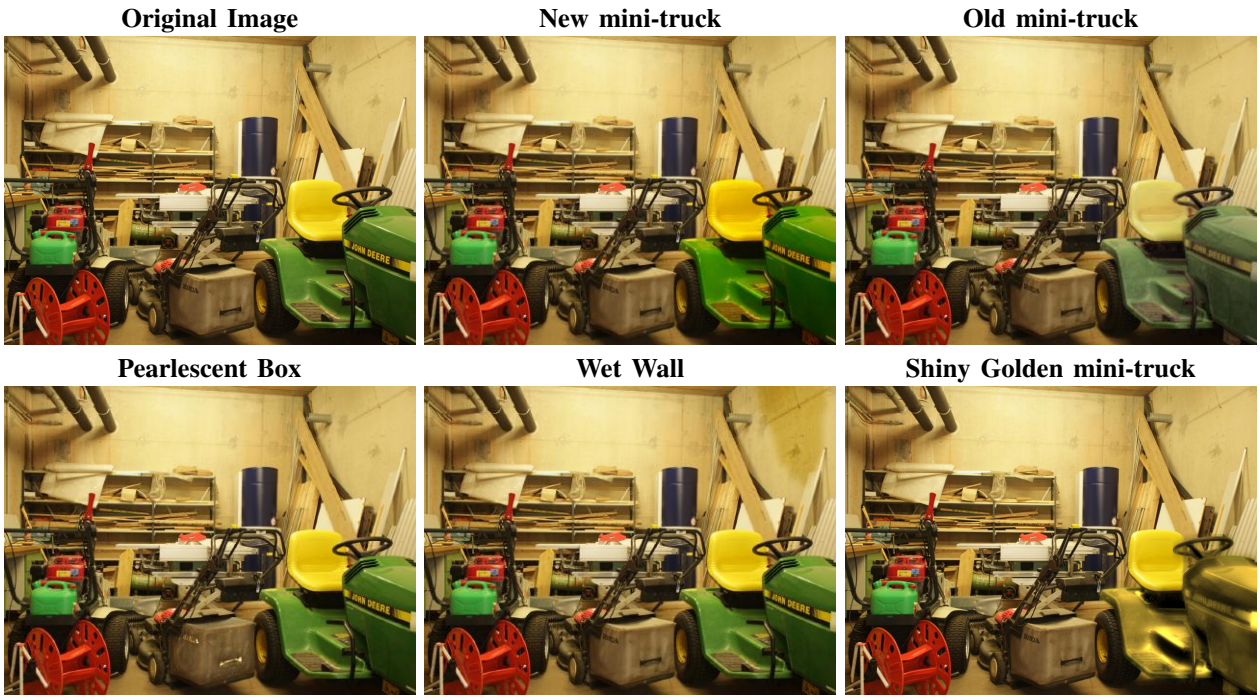

Fig. 20: More results for the *Workshop* scene from [3] dataset.

REFERENCES

[1] D. Scharstein, H. HirschmÃijller, Y. Kitajima, G. Krathwohl, N. Nesic, X. Wang, and P. Westling, "High-resolution stereo datasets with subpixel-accurate ground truth." in *GCPR*, ser. Lecture Notes in Computer Science, vol. 8753. Springer, 2014, pp. 31–42.

[2] S. Wanner, S. Meister, and B. Goldluecke, "Datasets and benchmarks for densely sampled 4d light fields," 2013.

[3] V. K. Adhikarla, M. Vinkler, D. Sumin, R. Mantiuk, K. Myszkowski, H.-P. Seidel, and P. Didyk, "Towards a quality metric for dense light fields," in *Proceedings of the IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2017.

[4] E. Garces, J. I. Echevarria, W. Zhang, H. Wu, K. Zhou, and D. Gutierrez, "Intrinsic light field images," *Computer Graphics Forum*, vol. 36, no. 3, 2017.

[5] A. Meka, M. Zollhöfer, C. Richardt, and C. Theobalt, "Live intrinsic video," *ACM Trans. on Graphics (Proceedings SIGGRAPH)*, vol. 35, no. 4, 2016.

[6] S. Bell, K. Bala, and N. Snavely, "Intrinsic images in the wild," *ACM Trans. on Graphics*, vol. 33, no. 4, 2014.

[7] N. Bonneel, J. Tompkin, D. Sun, O. Wang, K. Sunkavalli, S. Paris, and H. Pfister, "Consistent video filtering for camera arrays," *Computer Graphics Forum*, vol. 36, no. 2, 2017.

[8] N. Bonneel, J. Tompkin, K. Sunkavalli, D. Sun, S. Paris, and H. Pfister, "Blind video temporal consistency," *ACM Trans. Graph.*, vol. 34, no. 6, pp. 196:1–196:9, Oct. 2015.

[9] C.-K. Liang and C. Chung. (2006) Image-based material editing. [Online]. Available: http://chiakailiang.org/project_ibme/

[10] I. Boyadzhiev, K. Bala, S. Paris, and E. Adelson, "Band-sifting decomposition for image-based material editing," *ACM Trans. on Graphics*, vol. 34, no. 5, 2015.

[11] C. Strecha, W. Von Hansen, L. Van Gool, P. Fua, and U. Thoennessen, "On benchmarking camera calibration and multi-view stereo for high resolution imagery," in *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on*. Ieee, 2008, pp. 1–8.

[12] S. Beigpour, A. Kolb, and S. Kunz, "A comprehensive multi-illuminant dataset for benchmarking of intrinsic image algorithms," in *Proc. IEEE International Conference on Computer Vision (ICCV)*, December 2015, pp. 172–180.

[13] S. Beigpour, M. L. Ha, S. Kunz, A. Kolb, and V. Blanz, "Multi-view multi-illuminant intrinsic dataset," in *Proc. British Machine Vision Conference (BMVC)*, 2016.