Intrinsic Scene Decomposition from RGB-D images

Mohammed Hachama hachamam@gmail.com
Bernard Ghanem Bernard.Ghanem@kaust.edu.sa
Peter Wonka pwonka@gmail.com
King Abdullah University of Science and Technology (KAUST)
Thuwal, Saudi Arabia

Introduction

This document contains supplemental material for the submitted paper. The optimization strategy is detailed in Section 1, and more results on the MIT dataset are added [4] (Figures 1, 2, 3, 3, and 5).

1. Optimization strategy

In the paper, we formulated the constraint on the albedo $\mathbf{A}$ and the illumination $\mathbf{L}$ as an energy $E(\mathbf{A}, \mathbf{L})$ (Equation (12)). This energy is minimized using a cyclic block coordinate descent algorithm. We defined the i^{th}-subproblem:

$$\min \quad E_i((\mathbf{a}_i, \mathbf{l}_i)) \quad s.t.\ 0 \leq l_{i,k} \leq 1\,,\ 0 \leq a_{i,k} \leq 1\,, \tag{1}$$

where

$$E_i(\mathbf{a}_i, \mathbf{l}_i) = \sum_{p=1}^{N_p} w_p^i ||\mathbf{y}_p - \mathbf{a}_i \mathbf{n}_i^T \mathbf{l}_i|| + \lambda_a \sum_{j \in \mathcal{N}_i} ||\mathbf{a}_i - \mathbf{a}_j|| + \lambda_l \sum_{j \in \mathcal{N}_i} ||\mathbf{l}_i - \mathbf{l}_j||^2. \tag{2}$$

We solve this problem by applying a fixed-point scheme to find the solution of the Euler-Lagrange equation $\nabla E_i = 0$.

$$\frac{\partial E_i}{\partial \mathbf{a}_i} = \sum_{p=1}^{N_p} \frac{w_p^i\ (-\mathbf{n}_i^T \mathbf{l}_i)\ (\mathbf{y}_p - \mathbf{a}_i \mathbf{n}_i^T \mathbf{l}_i)}{||\mathbf{y}_p - \mathbf{a}_i \mathbf{n}_i^T \mathbf{l}_i||} + \lambda_a \sum_{j \in \mathcal{N}_i} \frac{(\mathbf{a}_i - \mathbf{a}_j)}{||\mathbf{a}_i - \mathbf{a}_j||}. \tag{3}$$

We then write the equation $\frac{\partial E_i}{\partial \mathbf{a}_i} = 0$ as follows:

$$\sum_{p=1}^{N_p} \frac{w_p^i\ (\mathbf{n}_i^T \mathbf{l}_i)^2\ \mathbf{a}_i}{||\mathbf{y}_p - \mathbf{a}_i \mathbf{n}_i^T \mathbf{l}_i||} + \lambda_a \sum_{j \in \mathcal{N}_i} \frac{\mathbf{a}_i}{||\mathbf{a}_i - \mathbf{a}_j||} \sum_{p=1}^{N_p} \frac{w_p^i\ (\mathbf{n}_i^T \mathbf{l}_i)\ \mathbf{y}_p}{||\mathbf{y}_p - \mathbf{a}_i \mathbf{n}_i^T \mathbf{l}_i||} + \lambda_a \sum_{j \in \mathcal{N}_i} \frac{\mathbf{a}_j}{||\mathbf{a}_i - \mathbf{a}_j||}. \tag{4}$$

This equation can also be written as a fixed-point equation:

$$\mathbf{a}_i = \frac{f(\mathbf{a}_i)}{g(\mathbf{a}_i)}, \tag{5}$$

with

$$f(\mathbf{a}_i) = \sum_{p=1}^{N_p} \frac{w_p^i\ (\mathbf{n}_i^T \mathbf{l}_i)^2}{||\mathbf{y}_p - \mathbf{a}_i \mathbf{n}_i^T \mathbf{l}_i||} + \sum_{j \in \mathcal{N}_i} \frac{\lambda_a}{||\mathbf{a}_i - \mathbf{a}_j||}, \tag{6}$$

and

$$g(\mathbf{a}_i) = \sum_{p=1}^{N_p} \frac{w_p^i\ (\mathbf{n}_i^T \mathbf{l}_i)\ \mathbf{y}_p}{||\mathbf{y}_p - \mathbf{a}_i \mathbf{n}_i^T \mathbf{l}_i||} + \lambda_a \sum_{j \in \mathcal{N}_i} \frac{\mathbf{a}_j}{||\mathbf{a}_i - \mathbf{a}_j||}. \tag{7}$$

Finally, we solve this equation using an iterative fixed-point scheme:

$$\mathbf{a}_i^{k+1} = \frac{f(\mathbf{a}_i^k)}{g(\mathbf{a}_i^k)}. \tag{8}$$

Similar equations apply for the estimation of $\mathbf{l}_i$:

$$\mathbf{l}_i^{k+1} = \frac{h(\mathbf{l}_i^k)}{\phi(\mathbf{l}_i^k)}, \tag{9}$$

where

$$h(\mathbf{l}_i) = \sum_{p=1}^{N_p} \frac{w_p^i\ (\mathbf{a}_i.\mathbf{a}_i)(\mathbf{n}_i \star \mathbf{n}_i)}{||\mathbf{y}_p - \mathbf{a}_i \mathbf{n}_i^T \mathbf{l}_i||} + 2\,\lambda_l \sum_{j \in \mathcal{N}_i} (1), \tag{10}$$

$$\phi(\mathbf{l}_i) = 2\,\lambda_l \sum_{j \in \mathcal{N}_i} \mathbf{l}_j^k + \sum_{p=1}^{N_p} \frac{w_p^i}{||\mathbf{y}_p - \mathbf{a}_i \mathbf{n}_i^T \mathbf{l}_i||}\,\mathbf{v}, \tag{11}$$

$$\mathbf{v}_m = \sum_{k=1}^{3} (a_{i,k}\, n_{i,m})\ (y_{p,k} - a_{i,k} \sum_{k' \neq m} l_{i,k'} n_{i,k'}), \tag{12}$$

and $(\mathbf{n}_i \star \mathbf{n}_i)$ is the vector obtained by element-wise product.

The iterative scheme composed of Equation (8) and (9) is used to solve the i^{th}-subproblem (1). The convergence of a fixed-point scheme is usually guaranteed by choosing a contraction application (f/g in our case). We do not present a convergence analysis in this paper. However, we observed good convergence rates in our simulations.

2. Results

References

[1] J. T. Barron and J. Malik. Intrinsic scene properties from a single rgb-d image. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2013. 3, 4, 5, 6, 7

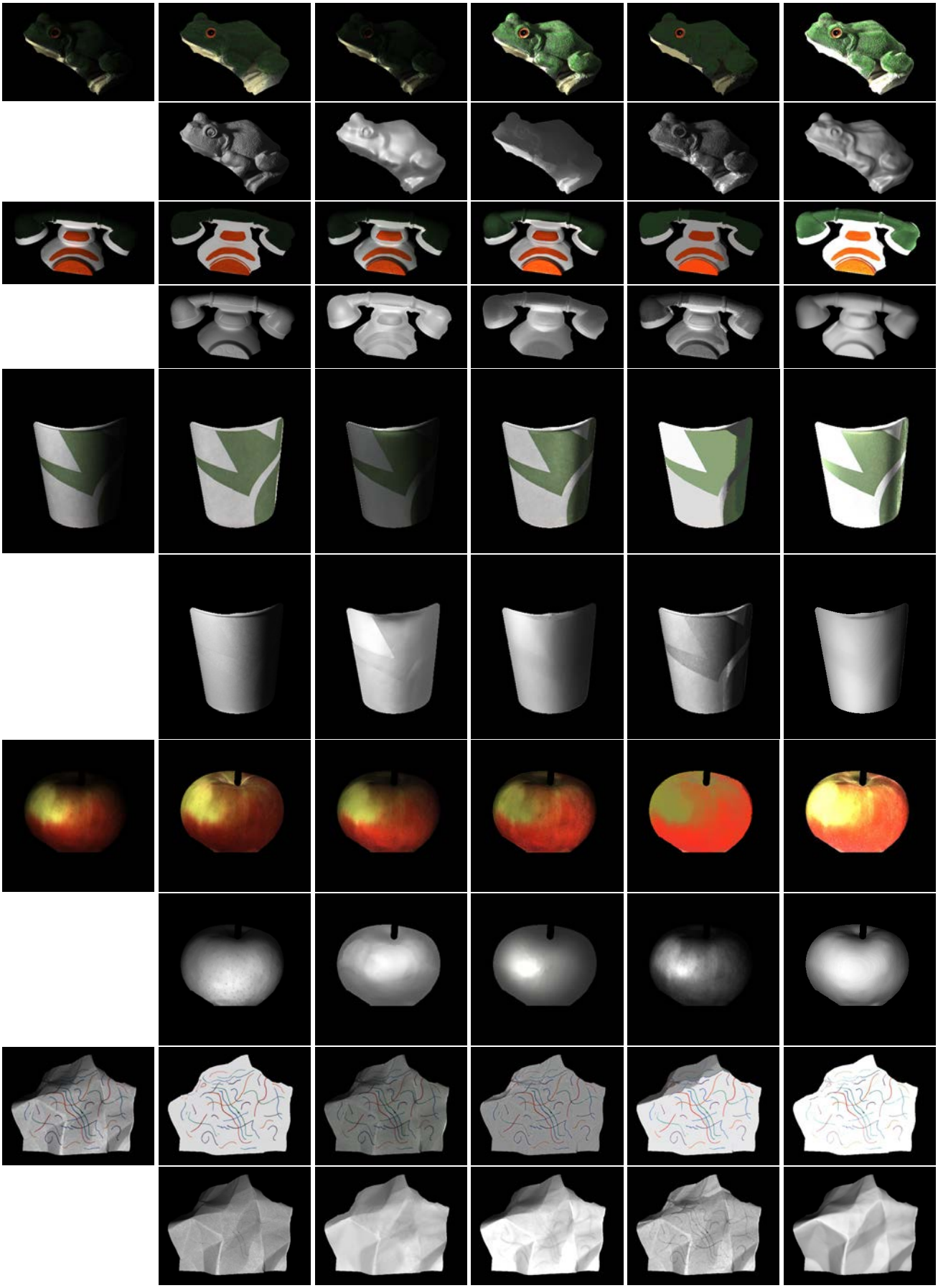
[2] S. Bell, K. Bala, and N. Snavely. Intrinsic images in the wild. *ACM Trans. Graph.*, 33(4), 2014. 3, 4, 5, 6, 7

[3] Q. Chen and V. Koltun. A simple model for intrinsic image decomposition with depth cues. In *IEEE International Conference on Computer Vision*, pages 241–248, Dec 2013. 3, 4, 5, 6, 7

[4] R. Grosse, M. K. Johnson, E. H. Adelson, and W. T. Freeman. Ground-truth dataset and baseline evaluations for intrinsic image algorithms. In *International Conference on Computer Vision*, pages 2335–2342, 2009. 1, 3, 4, 5, 6, 7

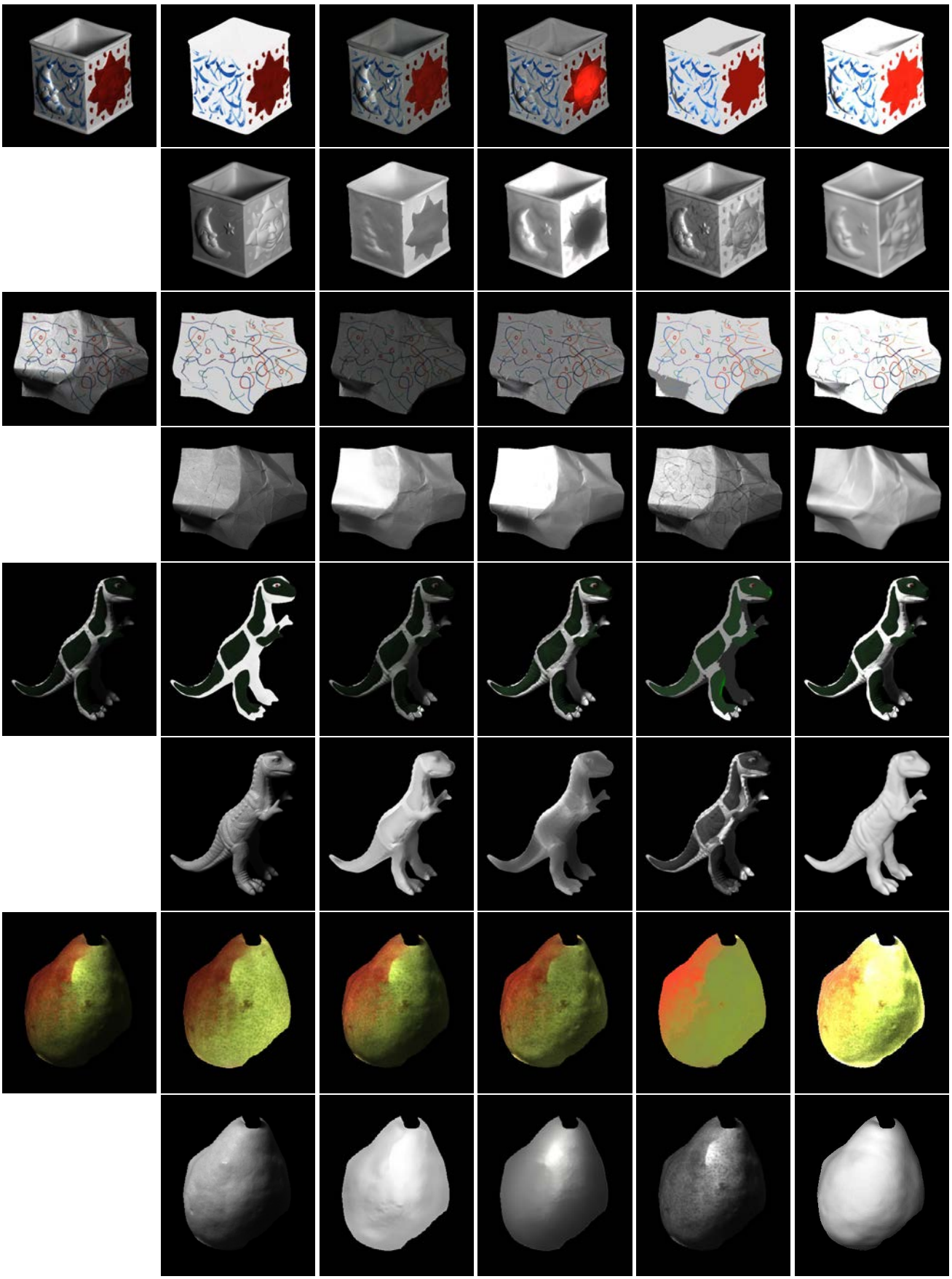
(a) Input (b) Ground truth (c) Barron-Malik [1] (d) Chen et al. [3] (e) Bell et al. [2] (f) Our approach

Figure 1. Intrinsic decomposition of two images from the MIT dataset [4]. (a) Input RGB images. (b) Ground truth albedo and shading images. (c-e) Intrinsic decomposition obtained by state-of-the-art techniques. (f) Intrinsic decomposition obtained by our approach.



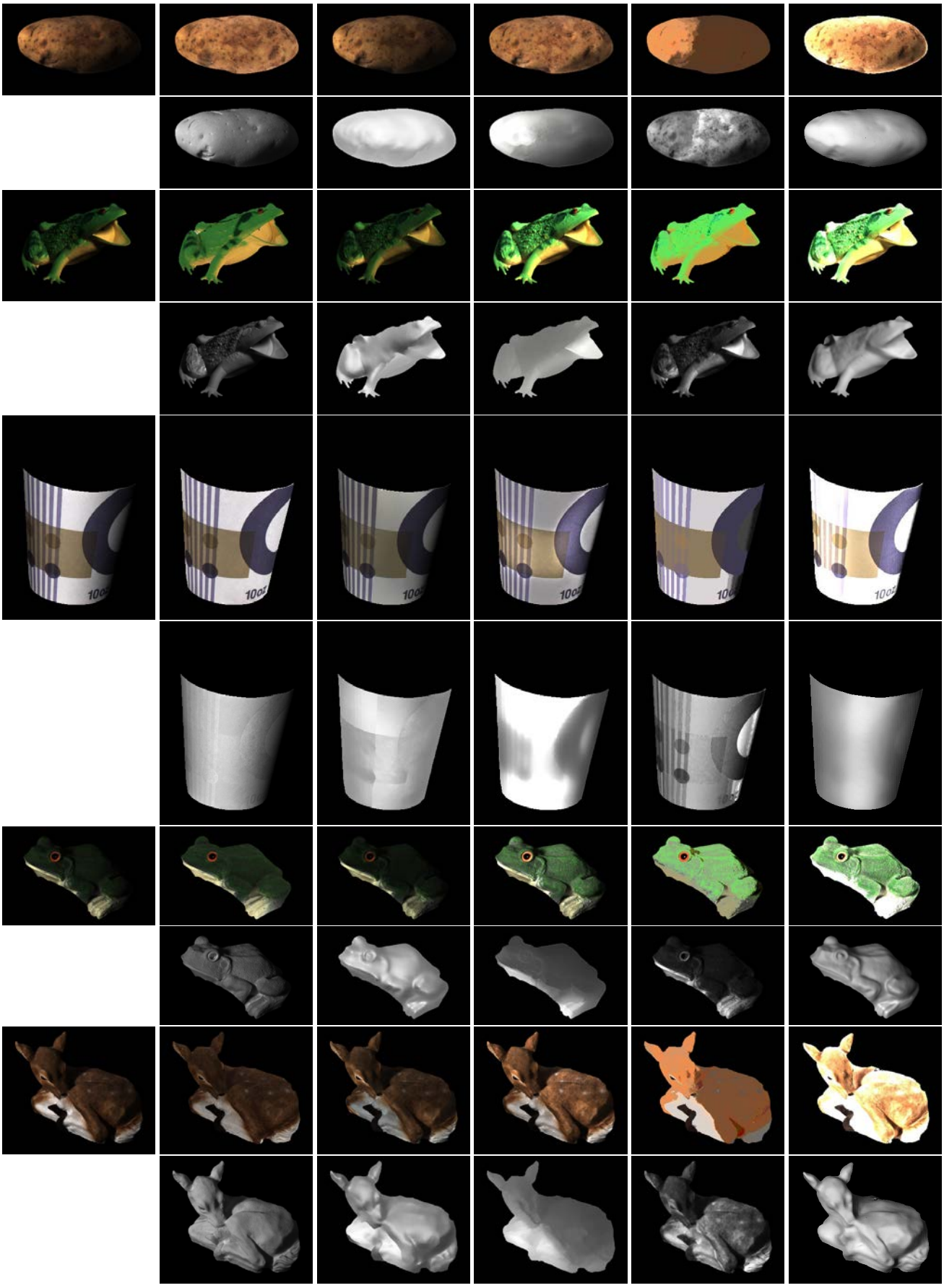
(a) Input (b) Ground truth (c) Barron-Malik [1] (d) Chen et al. [3] (e) Bell et al. [2] (f) Our approach

Figure 2. Intrinsic decomposition of two images from the MIT dataset [4]. (a) Input RGB images. (b) Ground truth albedo and shading images. (c-e) Intrinsic decomposition obtained by state-of-the-art techniques. (f) Intrinsic decomposition obtained by our approach.



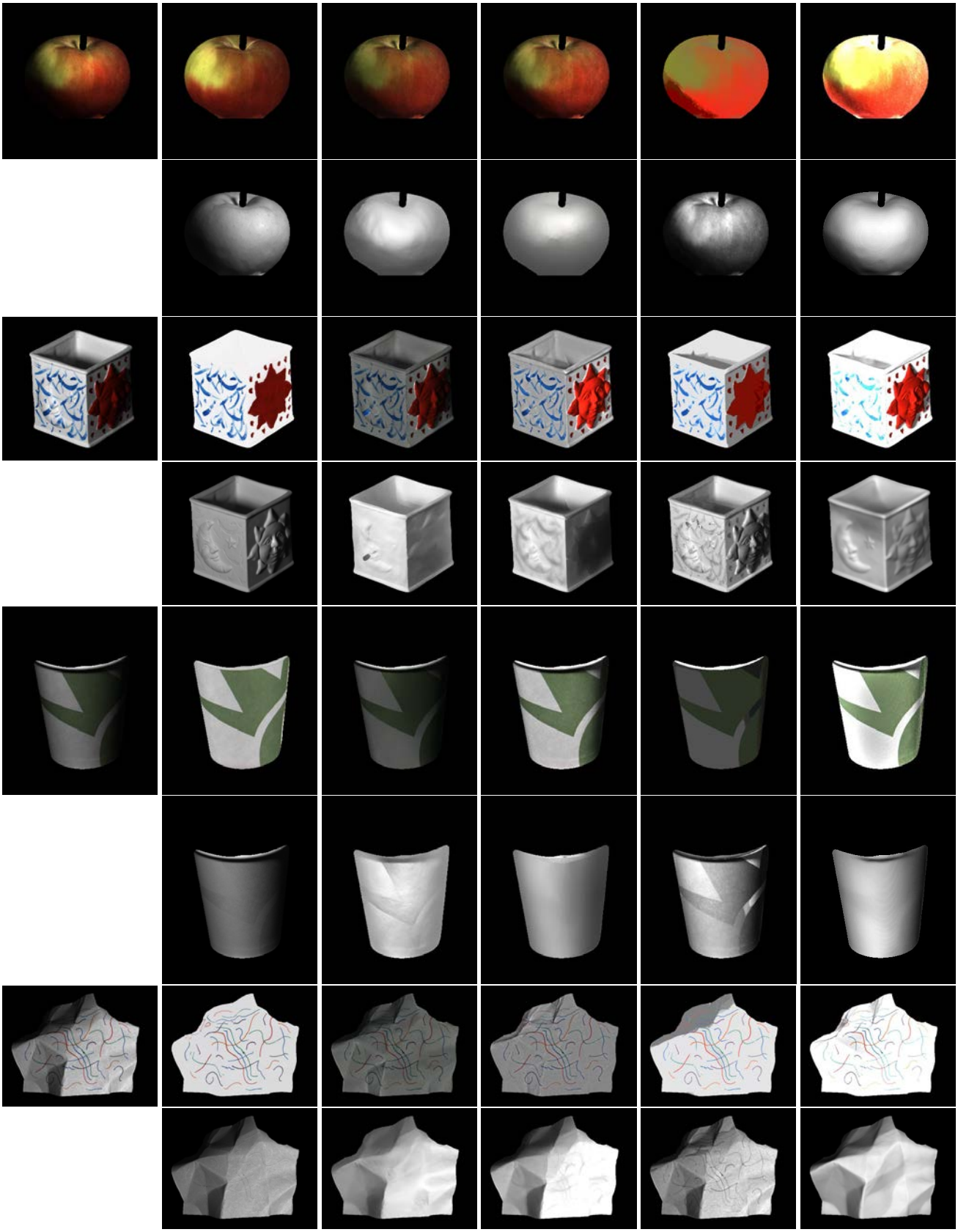
(a) Input (b) Ground truth (c) Barron-Malik [1] (d) Chen et al. [3] (e) Bell et al. [2] (f) Our approach

Figure 3. Intrinsic decomposition of two images from the MIT dataset [4]. (a) Input RGB images. (b) Ground truth albedo and shading images. (c-e) Intrinsic decomposition obtained by state-of-the-art techniques. (f) Intrinsic decomposition obtained by our approach.



(a) Input (b) Ground truth (c) Barron-Malik [1] (d) Chen et al. [3] (e) Bell et al. [2] (f) Our approach

Figure 4. Intrinsic decomposition of two images from the MIT dataset [4]. (a) Input RGB images. (b) Ground truth albedo and shading images. (c-e) Intrinsic decomposition obtained by state-of-the-art techniques. (f) Intrinsic decomposition obtained by our approach.



(a) Input (b) Ground truth (c) Barron-Malik [1] (d) Chen et al. [3] (e) Bell et al. [2] (f) Our approach

Figure 5. Intrinsic decomposition of two images from the MIT dataset [4]. (a) Input RGB images. (b) Ground truth albedo and shading images. (c-e) Intrinsic decomposition obtained by state-of-the-art techniques. (f) Intrinsic decomposition obtained by our approach.