Universidad Zaragoza
1542

Trabajo Fin de Máster

Practical Intrinsic Image Decomposition

Autora

Elena Garcés García

Director

Diego Gutiérrez Pérez

Escuela de Ingeniería y Arquitectura

Septiembre 2012

Appendix

Appendix A

Extended results

Figures A.1 to A.9 (From left to right) First column, input image and scatter plot of pixel data in the (a,b) plane (Lab color space). Second column, k-means segmentation according to (a,b) pixel coordinates; third column, final clustering yielded by our method taking into account spatial information (both, second and third rows, are depicted in false color). Last columns, the resulting shading and reflectance intrinsic images.

Figure A.10 Our intrinsic shading and reflectance for an image inspired by Dong et al. 2011. Our technique can also be used in conjunction with others, such as the image-based material modeling technique recently presented by Dong et al. 2011

Figures A.11 to A.15 Comparison with the state of the art methods in intrinsic image decomposition.

Figures A.16 to A.17 Intrinsic images for the MIT dataset. First column, input image. Second and third columns, ground truth shading and reflectance. Last columns, our resulting shading and reflectance. The gamma of the images has been corrected for visualization purposes.

Figure A.1: *Lollipop (original image by Thalita Carvalho, flickr.com)*

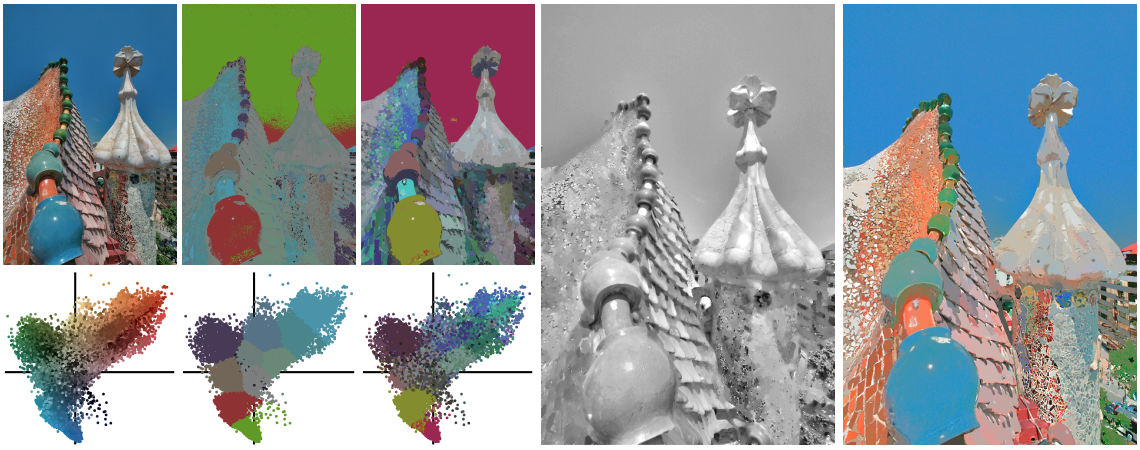

Figure A.2: *Batlló house (original image by lukasz dzierzanowski, flickr.com)*

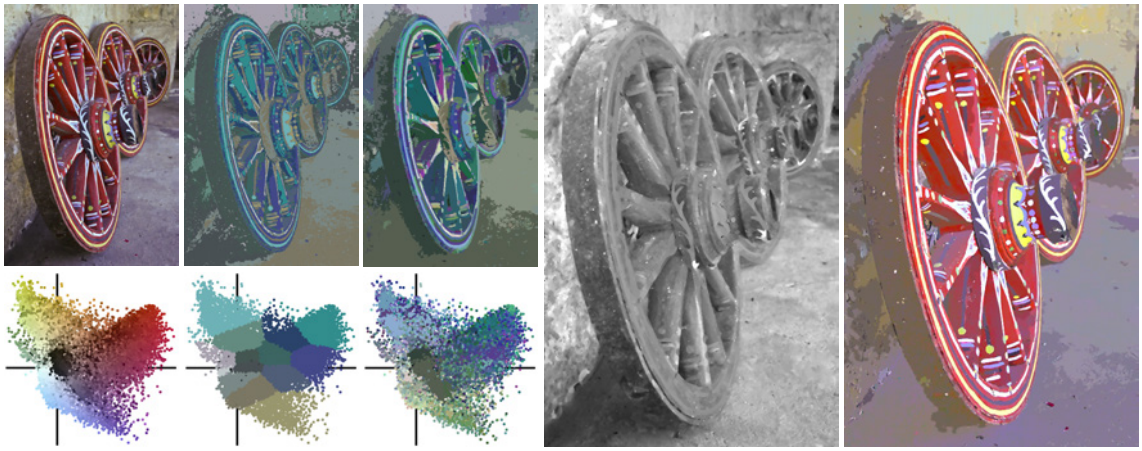

Figure A.3: *Wheels (original image by Angela Smith Kirkman)*

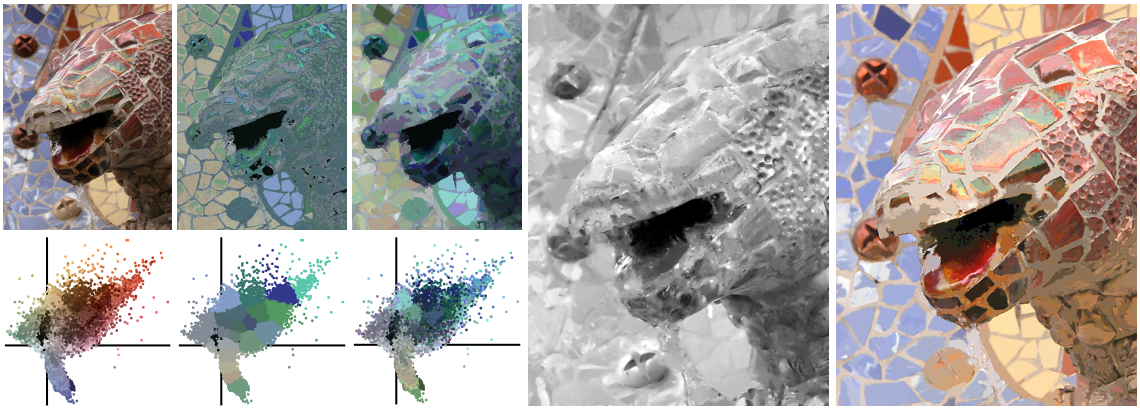

Figure A.4: *Dragon (original image by Jordanhill School D&T Dept, flickr.com)*

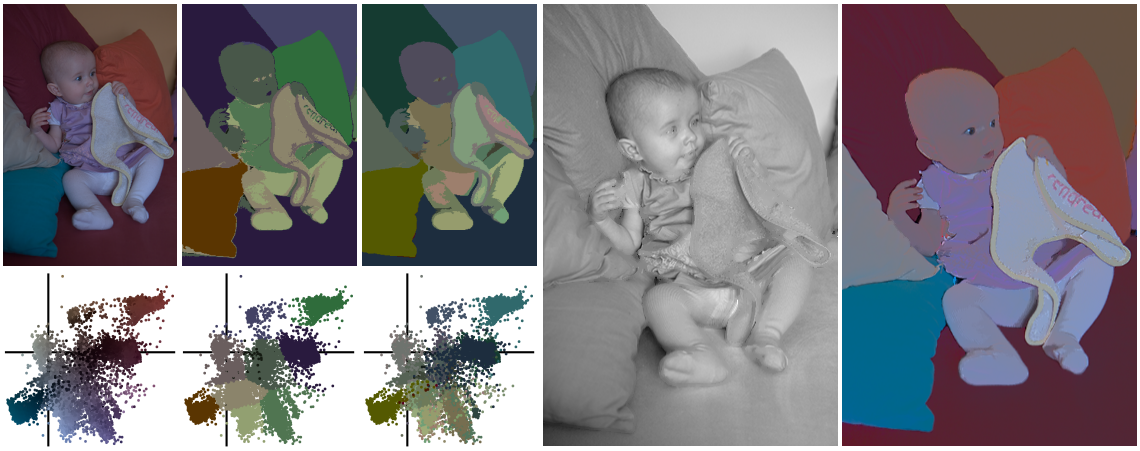

Figure A.5: *Baby*

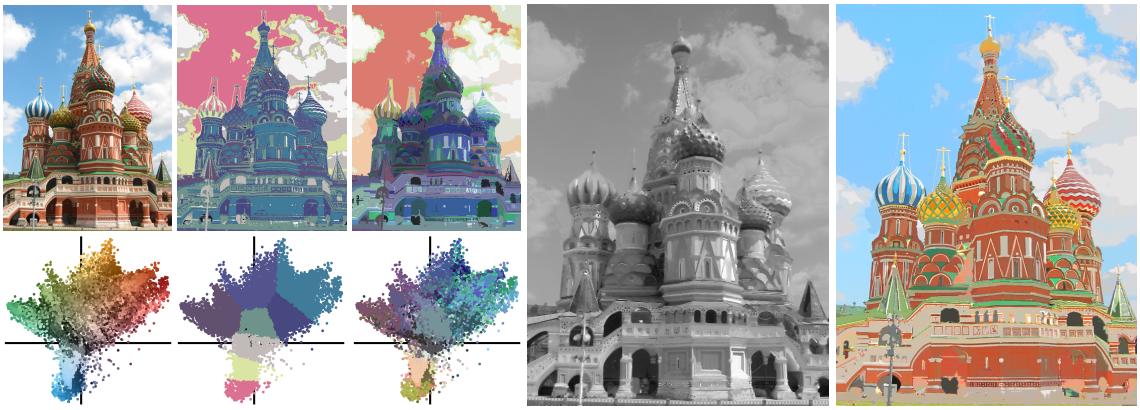

Figure A.6: *St. Basil (original image by Captain Chaos, flickr.com)*

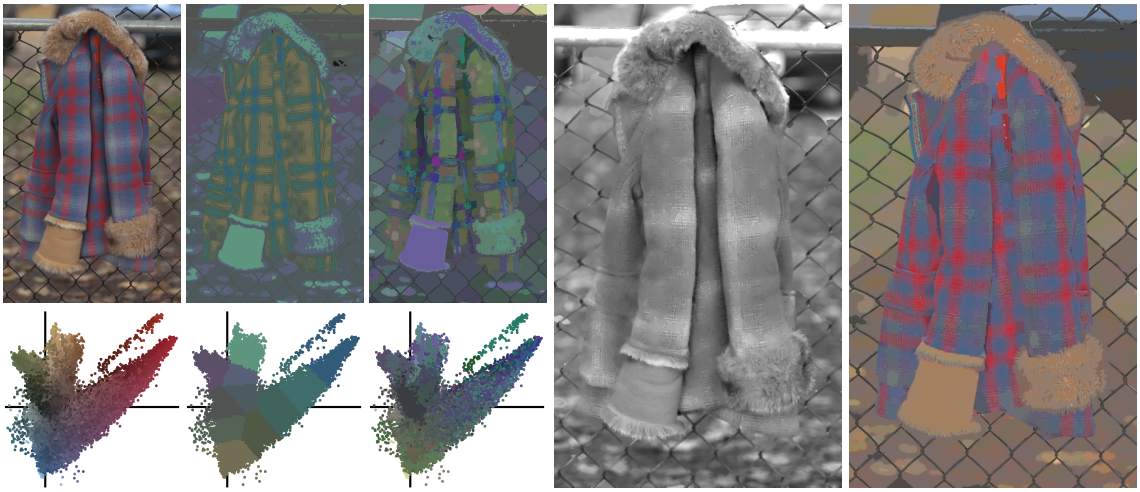

Figure A.7: *Coat*

Figure A.8: *Clown*

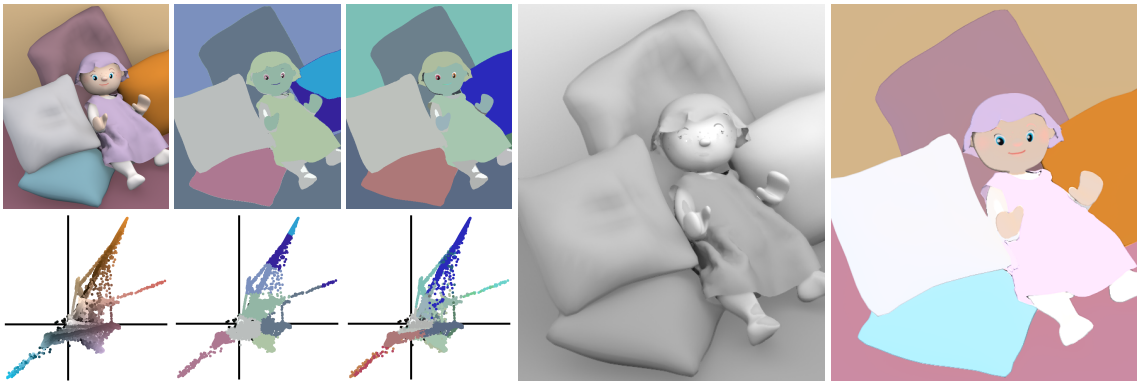

Figure A.9: *Synthetic*

Figure A.10: *Left: Input texture image. Middle: Our intrinsic reflectance. Right: Our intrinsic shading.*

Figure A.11: *Baby and St. Basil (original image by Captain Chaos, flickr.com)*


Input image (*clown*)
User strokes [Bousseau et al. 2009]
[Tappen et al. 2005]
[Shen et al. 2008]
[Bousseau et al. 2009] (33 strokes)
[Shen et al. 2011] (auto)
Our work

Figure A.12

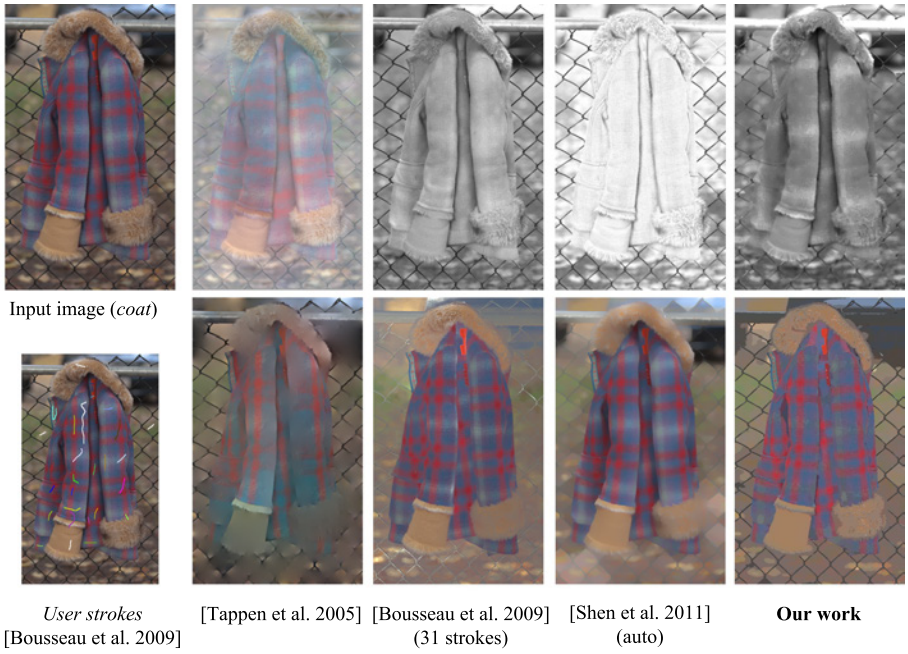

Figure A.13

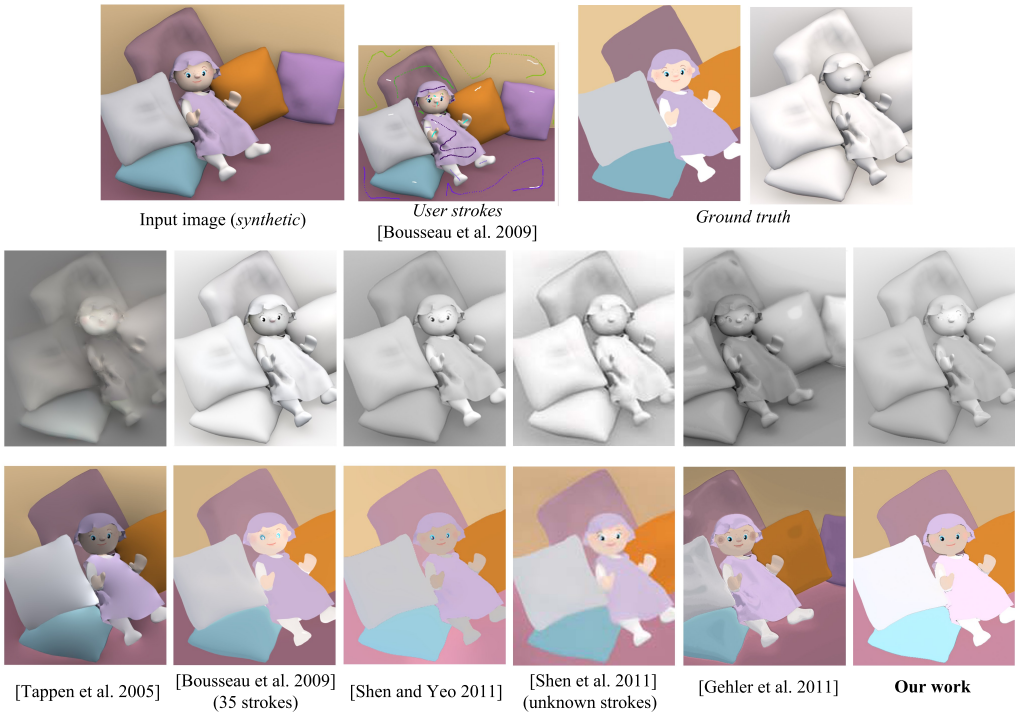

Figure A.14

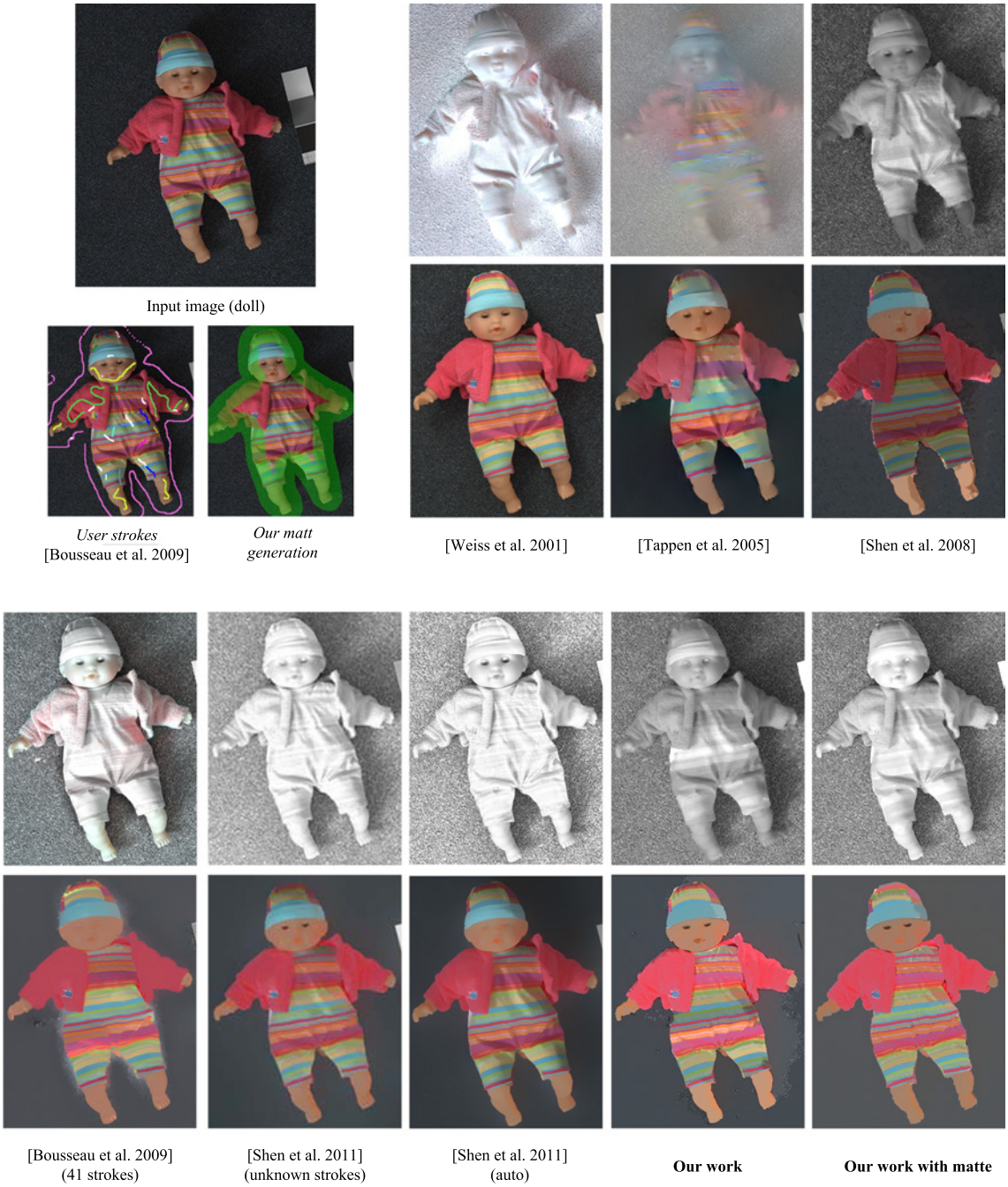
Input image (doll)
User strokes
[Bousseau et al. 2009]
Our matt
generation
[Weiss et al. 2001]
[Tappen et al. 2005]
[Shen et al. 2008]
[Bousseau et al. 2009]
(41 strokes)
[Shen et al. 2011]
(unknown strokes)
[Shen et al. 2011]
(auto)
Our work
Our work with matte

Figure A.15

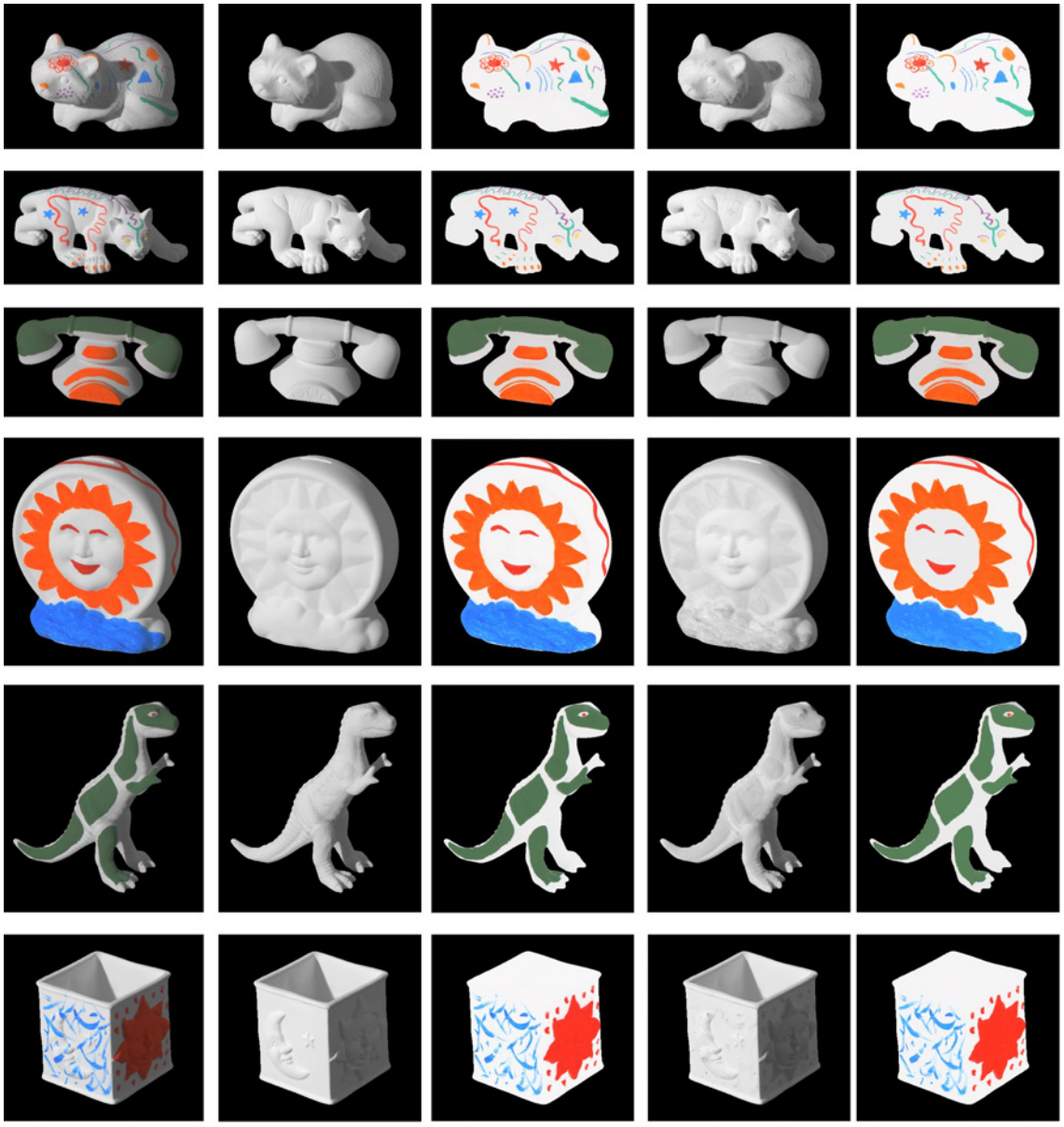

Figure A.16

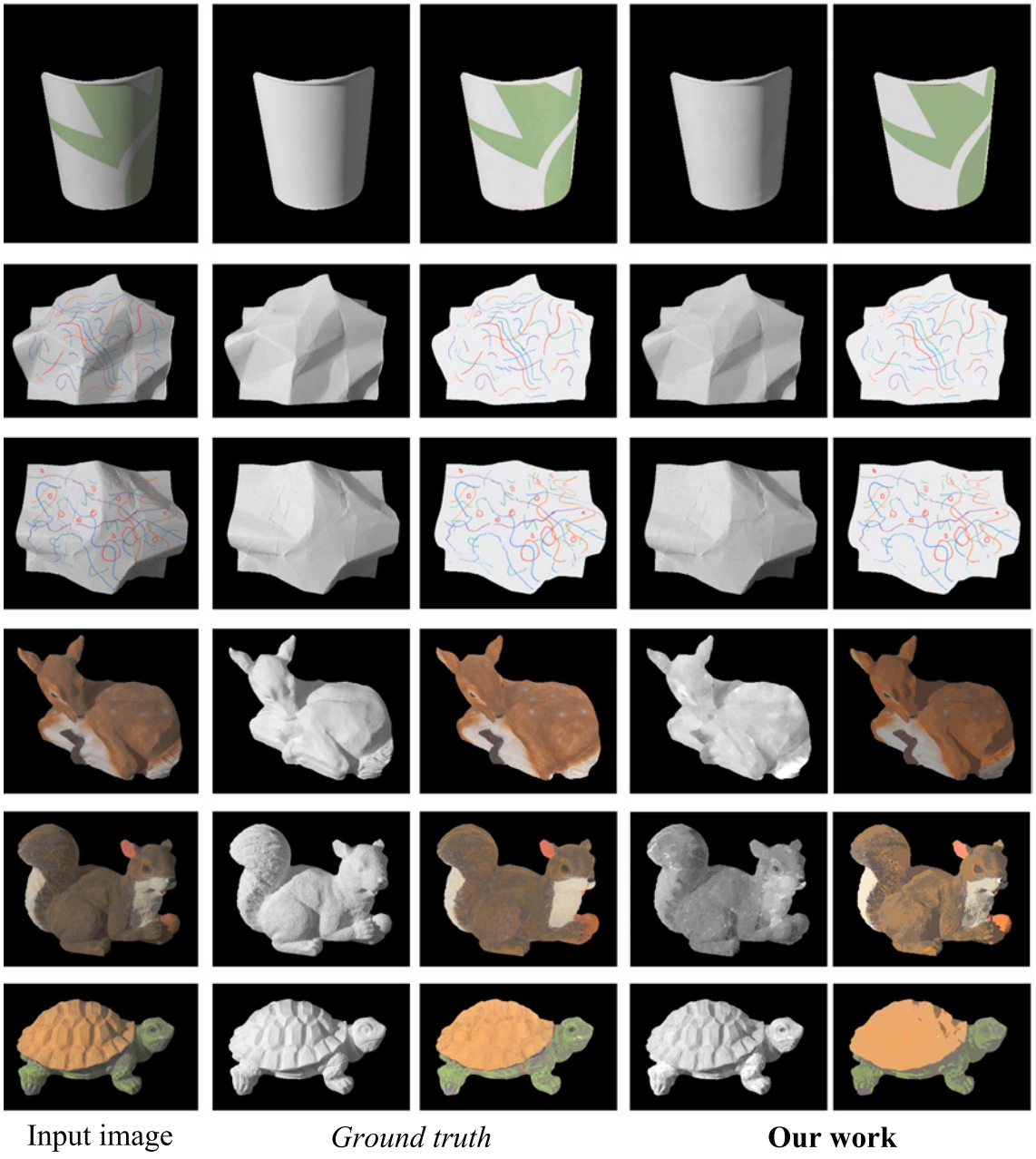

Figure A.17

Appendix B

EGSR 2012 - Intrinsic Images by Clustering

Eurographics Symposium on Rendering 2012
Fredo Durand and Diego Gutierrez
(Guest Editors)

Volume 31 (2012), Number 4

Intrinsic Images by Clustering

Elena Garces[1], Adolfo Munoz[1], Jorge Lopez-Moreno[1,2] and Diego Gutierrez[1]

[1]Universidad de Zaragoza, Spain, [2]REVES / INRIA Sophia-Antipolis, France

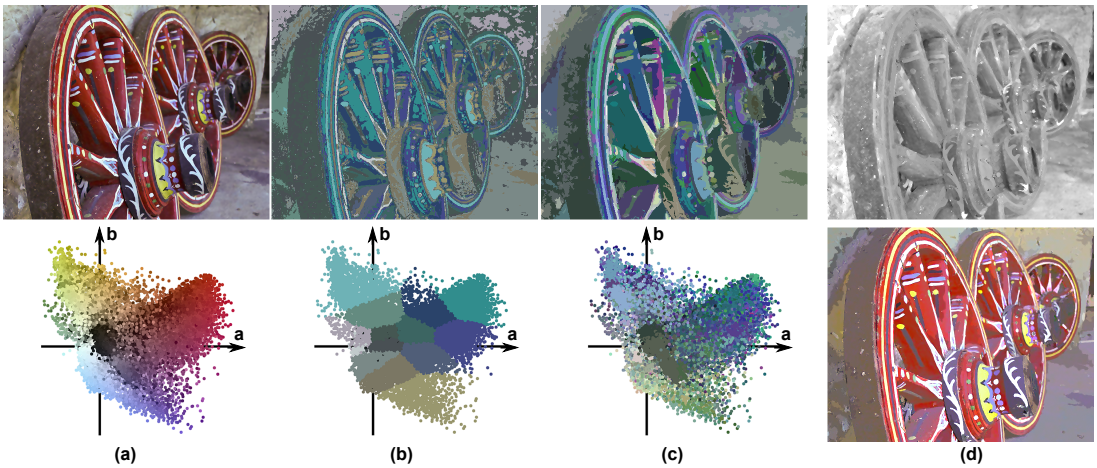

Figure 1: *Our algorithm decomposes an input image into its intrinsic images without user interaction.* ***(a)*** *Input image and scatter plot of pixel data in the (a,b) plane (Lab color space).* ***(b)*** *k-means segmentation according to (a,b) pixel coordinates.* ***(c)*** *Final clustering yielded by our method, taking into account spatial information (both (b) and (c) are depicted in false color).* ***(d)*** *The resulting shading and reflectance intrinsic images. The whole image is shown in Figure 5.*

Abstract

Decomposing an input image into its intrinsic shading and reflectance components is a long-standing ill-posed problem. We present a novel algorithm that requires no user strokes and works on a single image. Based on simple assumptions about its reflectance and luminance, we first find clusters of similar reflectance in the image, and build a linear system describing the connections and relations between them. Our assumptions are less restrictive than widely-adopted Retinex-based approaches, and can be further relaxed in conflicting situations. The resulting system is robust even in the presence of areas where our assumptions do not hold. We show a wide variety of results, including natural images, objects from the MIT dataset and texture images, along with several applications, proving the versatility of our method.

Categories and Subject Descriptors (according to ACM CCS): I.3.3 [Computer Graphics]: Picture/Image Generation—

1. Introduction

The problem of separating an input image into its intrinsic shading and reflectance components [BT78] is extremely ill-posed. However, many applications would benefit from the disambiguation of a pixel value into illumination and albedo, such as image relighting or material editing. This problem is usually formulated as the input image I being a per-pixel product of its unknown intrinsic shading S and reflectance R, so the space of mathematically valid solutions is in fact infinite. Existing methods therefore need to rely on additional sources of information, such as making reasonable assumptions about the characteristics of the intrinsic components, having multiple images under different illuminations or asking the user to add image-specific input.

In this paper, we describe a new algorithm that works on a single off-the-shelf image and requires no user strokes. We do make some reasonable assumptions, in the form of flexible constraints. We formulate the decomposition of an input image into its shading and reflectance components as a linear system that exploits relations between clusters of similar reflectance. Classic Retinex approaches assume that i) reflectance is piecewise constant, and ii) shading is spatially smooth (C^0 and C^1 continuity) [LM71, Hor74]. Based on this, a number of authors have proposed different approaches [FDB92, KES*03, STL08, GRK*11]. In this work we first find clusters of similar reflectance in the image following the observation that *changes* in chromaticity usually correspond to changes in reflectance.

We then relax the second Retinex assumption that shading is spatially smooth in two ways: we assume only C^0 continuity on the shading, and only *at the boundaries* between clusters (as opposed to the whole image), and describe this as a set of linear equations. Our linear system is completed by additionally preserving reflectance between clusters even if they are not contiguous, and adding a regularization term to make it more stable.

Our main contribution is a novel algorithm for intrinsic images decomposition which deals with a wider range of scenarios than traditional Retinex-based algorithms, yields better decompositions than existing automatic methods from single images, and offers an attractive trade-off between quality and ease of use, compared with techniques requiring either significant user input or multiple input images. We present an exhaustive comparison against most existing techniques (see supplementary material), which we will make public along to our source code (once the paper is published). Last, we show compelling example applications of retexturing, relighting and material editing based on our results.

Like all existing methods that deal with this ill-posed problem, our work is not free of limitations: Our C^0 assumption is a simplification that breaks for some occlusion boundaries and sharp edges, which translate into inaccurate equations in the system. However, given our robust formulation which usually translates into a few thousand equations, these inaccurate equations represent a very small percentage, and our method generally handles these situations well.

2. Related Work

Automatic Some automatic methods rely on reasonable assumptions about the nature of these two terms, or the correlation between different characteristics of the image. Horn [Hor74] presents a method to obtain lightness from black and white images, using pixel intensity information and assuming that lightness corresponds to reflectance. He further assumes that the reflectance remains locally constant while illumination varies smoothly (as described by the Retinex theory [LM71]). Funt et al. [FDB92] extend this approach to color images, and propose the analysis of chromaticity variations in order to identify the boundaries of different reflectance areas. They enforce integrability of the shading at these boundaries and propagate their values to their neighboring pixels by diffusion, solving the subsequent Poisson equation with a Fourier transformation. This was later extended by Shen et al. [STL08] with global texture constraints, forcing distant pixels with the same texture to have the same reflectance. This constraint greatly improves the performance of the standard Retinex method, although it relies on objects with repeated texture patterns and may yield posterization artifacts due to the wrong clustering of distant pixels. The related method by Finlayson and colleagues [GDFL04] is mainly oriented to remove shadows by minimizing entropy, but does not to recover intrinsic images. The work by Jiang et al. [JSW10] assumes that correlated changes in mean luminance and luminance amplitude indicate illumination changes. By introducing a novel feature, *local luminance amplitude*, the authors obtain good results, although limited to images of relatively flat surfaces and objects from the recently published MIT dataset for intrinsic image decomposition [GJAF09]. This actually simplifies the problem since such objects are treated in isolation, avoiding the problem of occlusion boundaries at the outlines. Recently, Gehler and colleagues [GRK*11] proposed a probabilistic model, based on a gradient consistency term and a reflectance prior, which assumes that reflectance values belong to a sparse set of basis colors. The problem is formulated as an optimization of the proposed energy function. The method yields good results, although again limited to isolated objects from the MIT dataset. Our linear system formulation allows for much faster computational times (up to a thousand times faster), and generalizes well over a wider range of images (including both natural and texture images).

User intervention Another set of techniques rely on assumptions *and* user intervention. Bousseau and colleagues [BPD09] simplify the problem by assuming that local reflectance variations lie in a 2D plane in RGB space not containing black, which may not be compatible with certain texture or grayscale images. This assumption is also used in the work by Shen and Yeo [SY11], who further consider that neighboring pixels in a local window with similar intensity have also similar reflectance. In addition, Bousseau's method requires that the user define constraints over the image by means of three different types of strokes: constant-reflectance, constant-illumination and fixed-illumination. Their method produces very compelling results, although creating the appropriate strokes for each particular image (from 15 to 81 for the figures in the paper) may be far from intuitive for unskilled users. The same set of user tools is employed in the recent work by Shen and colleagues [SYJL11], who use an optimization approach that further assumes that neighboring pixels with similar intensity have similar reflectance values. In the context of mate-

rial editing, Dong et al. [DTPG11] assume input images of globally flat surfaces with small normal perturbations and lit with a directional light, and require user strokes for optimal decompositions. In contrast, our method is almost fully automatic (usually a single parameter is needed) and requires no user strokes.

Multiple images Last, another strategy consists of incorporating additional information, either from other sources or from multiple images. Tappen et al. [TFA05] classify the derivatives of the image as produced either by illumination or albedo. Ten classifiers are obtained from training Adaboost [FS95] with a set of computer generated images containing only reflectance or illumination components. They further refine their approach by introducing a new training set of real-world images and including a method to weigh the response to these classifiers [TAF06]. Despite these advanced techniques, several configurations of illumination and reflectance remain very difficult to decompose and additional techniques like Markov Random Fields (MRF) and Belief Propagation (BP) are necessary in order to yield good solutions. Weiss [Wei01] uses a large sequence of images of the same scene (up to 64 images, and no less than 35, taken in controlled settings), where the reflectance remains constant and illumination varies in time. Also using multiple images, Laffont and colleagues [LBD11] leverage multi view stereo techniques to approximately reconstruct a point cloud representation of the scene. After some user intervention, illumination information computed on that point cloud is propagated in the image. Their method decomposes the illumination layer into three components: sun, sky and indirect light. Last, the concept of *intrinsic colorization* is introduced by Liu et al. [LWQ*08]; to colorize a grayscale image, their method recovers the needed reflectance component from multiple images obtained from the web, in order to transfer color from there. All these techniques require multiple images as input, sometimes captured under controlled settings, while our approach simply takes an off-the-shelf single image.

3. Algorithm

The desired decomposition consists of separating an image into two components (images): one representing reflectance information, and another containing the illumination or shading. We use RAW or linearized RGB values as input. For a Lambertian scene, the problem can be simply formulated as:

$$I(x,y) = S(x,y) * R(x,y) \tag{1}$$

where $I(x,y)$ is the input image, $S(x,y)$ is the shading image, $R(x,y)$ represents reflectance information and $*$ is a per-channel Hadamard product. Our goal is to obtain $S(x,y)$ and $R(x,y)$, given $I(x,y)$. We make the problem tractable with a few assumptions well-grounded on existing vision and image processing techniques. While of course our assumptions

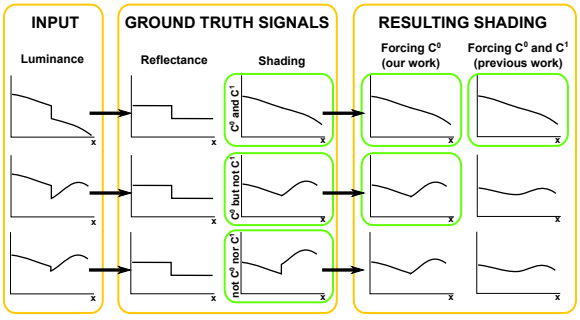

Figure 2: *Intrinsic shading estimation when both the shading and the reflectance present a discontinuity at the same point (as in some occlusion boundaries). Left column: Three different input luminance signals. Middle columns: Ground truth intrinsic signals. All three input signals are the result of multiplying the same reflectance with three different shading signals, presenting different continuity characteristics. Right columns: Results assuming both C^0 and C^1 continuity on the shading, compared to C^0 only (our method). Notice how our algorithm leads to an accurate result in two of the three cases, while yielding less error in the most unfavorable case.*

may not always be accurate throughout the whole image, they allow us to devise a method that works very well on a large range of images while keeping our algorithm simpler than other approaches.

Assumptions Horn made the key observation that, for grayscale images, sharp reflectance changes cause intensity discontinuities in the luminance of an image [Hor74]. Our first assumption relies on the later generalization to color images by Funt et al. [FDB92], who associate changes in reflectance with changes in chromaticity. We first leverage this correlation between reflectance and chromaticity values by detecting regions of similar chromaticity in the input image, which are assumed to approximate regions of similar reflectance. We implement this as a soft constraint, though, which we relax in specific cases (see Sections 3.1 and 3.2).

Furthermore, existing Retinex-based techniques (see for instance [FDB92, KES*03, STL08]) assume that shading is a smooth function, therefore being both C^0 and C^1 continuous. However, there are a number of particular cases (such as some occlusion boundaries) in which this assumption does not hold. Our second assumption relaxes this restriction by imposing only C^0 continuity *at boundaries* between the regions previously detected. This allows us to handle a wider variety of cases correctly; in cases where the smooth shading assumption does hold, our method naturally maintains C^1 continuity as well (see Figure 2). In cases where this assumption breaks, our method still provides a more accurate reconstruction of the intrinsic signals. Last, as previous works, we assume a white light source and a correctly white balanced input image.

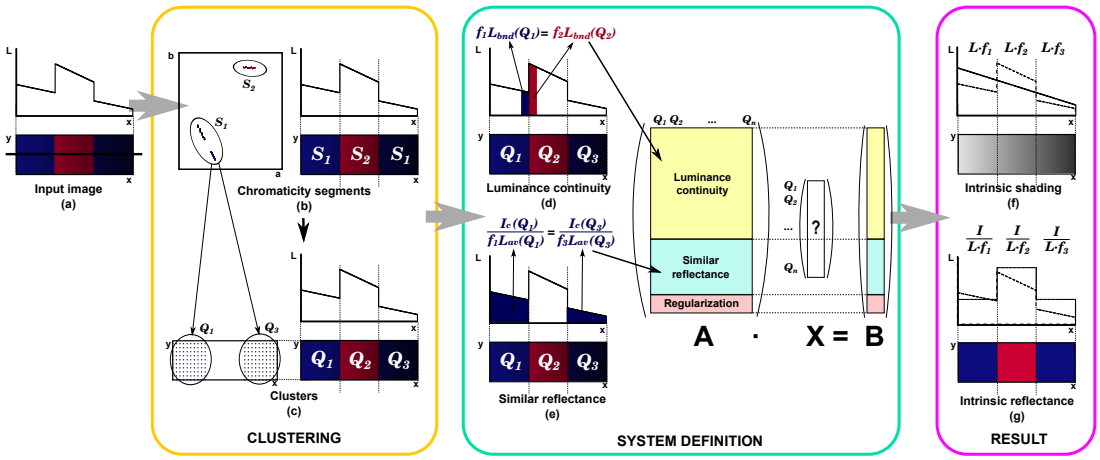

Figure 3: *Overview of the algorithm for the simple case of three colored patches and a continuos shading gradient.* ***(a)*** *Input image, with a plot of the pixels luminance along a scan line.* ***(b)*** *Initial k-means segmentation. Left: A scatter plot of the (a,b) coordinates (Lab color space) shows two segments of different chrominance (S_1 and S_2). Right: These segments belong to different parts of the image, with S_1 split in two image areas (labeled accordingly in the figure).* ***(c)*** *Subsequent clustering. Left: Segments are further divided into clusters of contiguous pixels. The example shows S_1 being clustered into Q_1 and Q_3 in image space. Right: clusters labeled in the image.* ***(d)*** *Enforcing luminance continuity on the boundaries between two clusters yields a large number of equations for the linear system.* ***(e)*** *Clusters originally belonging to the same segment maintain similar reflectance properties (the example shows Q_1 and Q_3, both belonging initially to S_1). This yields another set of equations. The final system is completed with a regularization term.* ***(f)*** *Result: Intrinsic shading. It is a continuous signal, as described by the equations in (d).* ***(g)*** *Result: Intrinsic reflectance. Q_1 and Q_3 share the same reflectance, as described by the equations in (e). Please refer to the text for further details on the equations and their notation.*

Overview Figure 3 shows an overview of our algorithm, applied to patches of different colors with a continuous shading gradient. It works in two steps, which we term *clustering* and *system definition*. First, we segment the input image according to chromaticity. We then subdivide the resulting segments, and obtain a set of clusters of connected pixels with similar chromaticity. This clustering is then refined to better approximate reflectance (as opposed to chromaticity) discontinuities, according to our first assumption.

Based on this clustering, we then build a linear system of equations defining the connections and relations between the different clusters, as well as the different constraints. One set of equations describes the C^0 continuity in the shading at cluster boundaries (our second assumption). We then make the observation that all clusters originally coming from the same segment should in principle maintain similar reflectance, even if they are not contiguous. This is similar to the observation made by Shen et al. [STL08]; however, we improve this in two important ways: first, we do not need to rely on texture information; second, we work at cluster level, as opposed to pixel level, which translates into a more stable solution. This yields our second set of equations. The system is completed with an additional regularization term. By solving the resulting linear system we obtain the intrinsic shading image; reflectance is obtained by means of a simple per-pixel RGB division (Equation 1), as previous works [BPD09]. The next sections describe these steps in detail.

3.1. Clustering

We aim to divide the image into clusters of similar chrominance properties. Given our assumptions, the boundaries between those clusters will indicate reflectance discontinuities. This is a reasonable task, given the reduced set of reflectance values in natural images [OW04]. This reduced set was also leveraged in recent work by Bousseau et al. [BPD09], who further assumed that reflectance colors lie in a 2D plane not intersecting the origin. Several existing segmentation techniques, such as Mean Shift, the graph-based method by Felzenszwalb and Huttenlocher [FH04] or its subsequent modification [GGLM11] have been thoroughly tested, but unfortunately none would yield satisfying results for our purposes. We thus have designed a novel two-step clustering strategy, specially tailored for the problem of intrinsic images decomposition. For the sake of clarity, we refer to the first step as *segmentation*, and to the second as *clustering*.

Segmentation We first segment the image according to chromaticity values, regardless of the spatial location of the pixels. We define our segmentation feature space as $\mathcal{F} = \{\beta, a, b\}$ where (a,b) are the chromatic coordinates of the input image in *CIELab* space, and β is a feature defined to handle strong blacks or whites (these are defined as pixels

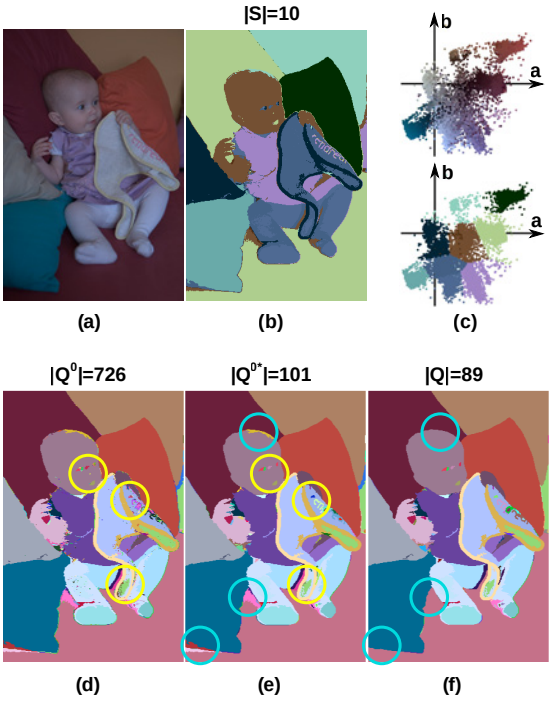

Figure 4: *Our segmentation-clustering process. (a) Input image (b) Result of the first segmentation step, yielding 10 distinct segments in $\mathcal{S}$. (c) Top, scatter plot of the input image in the (a,b) plane. Bottom, scatter plot of the first segmentation step. (d) Initial clustering (726 clusters in $\mathcal{Q}^o$). (e) Merging small clusters. (f) Final cluster set $\mathcal{Q}$ after merging smooth boundaries (89 clusters). The yellow and blue circles highlight areas were the effects of these last two steps are clearly visible. Notice how noise is eliminated (yellow circles), as well as smooth gradients due to shadows (blue circles).*

with very low chromaticity values and very low or high luminance). These values would be difficult to segment properly in a chromaticity-based algorithm, and usually describe important reflectance features. For each pixel in the image, we define β as:

$$\beta = \begin{cases} -\mu & \text{if } (|a| < \lambda) \;\&\; (|b| < \lambda) \;\&\; (L < L_{min}) \\ +\mu & \text{if } (|a| < \lambda) \;\&\; (|b| < \lambda) \;\&\; (L > L_{max}) \\ 0 & \text{otherwise} \end{cases} \tag{2}$$

where $\mu = 10^5$, $\lambda = 0.20\max(|a|,|b|)$, $L_{min} = 0.15\max(L)$ and $L_{max} = 0.95\max(L)$. For this initial segmentation, we use the k-means implementation from Kanungo et al. [KMN*04]. Gehler and colleagues [GRK*11] also used k-means for their global sparse reflectance prior, which along with their shading prior and their gradient consistency term, fit into their global optimization system. In contrast, we use this segmentation to drive a simple and efficient system of linear equations. Note that the high μ value in the definition of β in equation 2 effectively forces the algorithm to create different segments with only strong black (or white) pixels. Except otherwise noted, we set $k = 10$ as the number of segments, but in our implementation it is left as a user parameter. The result of this step is a set of segments $\mathcal{S} = \{S_i\}$ (see Figures 4.a and 4.b). These will guide the clustering step of the process, and help define global reflectance constraints between disconnected areas of the image during the *system definition* stage of the algorithm (Section 3.2).

Clustering The previous segmentation defines global relations between (possibly disconnected) regions of the image. We now take into account local constraints by considering spatial contiguity. We first subdivide each segment $S_i \in \mathcal{S}$ into a set of *clusters* of contiguous pixels (8-neighborhood in image space), obtaining $\mathcal{Q}^o = \{Q_i^o\}$ (Figure 4.c). This set $\mathcal{Q}^o$ may contain very small clusters (due to quantization, aliasing or smooth varying textures), which could potentially later make our system less stable, or pairs of connected clusters where changes in chromaticity do not correspond to changes in reflectance (maybe due to shadows [GDFL04]).

Merging small clusters: Given a cluster Q_r^o containing less than p pixels, we locate its neighbor cluster Q_s^o with the closest average chrominance and merge them together: $Q_s^{o*} = Q_r^o \cup Q_s^o$. For the results in this paper we use $p = 10$. This process is iterated until no more small clusters remain (Figure 4.d).

Merging smooth boundaries: Since chrominance and reflectance are not always exactly related, the k-means algorithm might yield over-segmented results. Given two adjacent clusters Q_r^{o*} and Q_s^{o*}, we average RGB pixel differences across the common border, obtaining a scalar d. The clusters are merged into Q_{rs} if $d < D$. The threshold D is set to 0.01 times the maximum pixel value in the image.

The result after these operations is our final cluster set $\mathcal{Q} = \{Q_i\}$ (see Figure 4.e).

3.2. System definition

The previous step has yielded a set of clusters separated by reflectance discontinuities. We now describe how to estimate the intrinsic shading from this initial clustering. We define a per-cluster factor f_i that, multiplying the luminance of the pixels of the cluster, will result into the intrinsic shading:

$$S(x,y) = f_i L(x,y) \tag{3}$$

where $(x,y) \in Q_i$. Instead of using expensive optimization techniques, we build a linear system of equations where f_i are the unknowns of our system. This system is built from three sets of equations as described below.

Luminance continuity We first enforce C^0 luminance continuity at the boundaries between clusters, in effect assigning abrupt changes at such boundaries to reflectance variations. Given the boundary between two clusters Q_r and Q_s:

$$f_r L_{bnd}(Q_r) - f_s L_{bnd}(Q_s) = 0 \tag{4}$$

where $L_{bnd}(Q_r)$ represents the luminance of the pixels in cluster Q_r at the boundary with cluster Q_s (and vice versa for $L_{bnd}(Q_s)$, see Figure 3.d). Last, f_r and f_s are the unknowns that force luminance continuity. In practice, we make Equation 4 more robust and obtain $L_{bnd}(\cdot)$ by averaging the luminance values of several pixels in a small window to each side of the boundary. We set the width of this window to three pixels for all the images in the paper.

However, applying exactly Equation 4 leads to an unstable behavior of the linear system; instead, we rewrite it in log-space:

$$\ln(f_r) - \ln(f_s) = \ln\left(\frac{L_{bnd}(Q_s)}{L_{bnd}(Q_r)}\right) \tag{5}$$

which leads to a more stable system and avoids both the trivial solution $f_i = 0$ and solutions with any $f_i < 0$. We apply Equation 5 to each pair of contiguous clusters.

Clusters of similar reflectance All clusters in $\mathcal{Q}$ coming from the same segment $S_i \in \mathcal{S}$ should in principle maintain similar reflectance. For each pair of clusters $\{Q_r, Q_s\} \in S_i$ we then have one equation per-channel with $c = \{R, G, B\}$:

$$\frac{I_c(Q_r)}{f_r L_{av}(Q_r)} = \frac{I_c(Q_s)}{f_s L_{av}(Q_s)} \tag{6}$$

where $I_c(r)$ is pixel average of the input image for all the pixels of the cluster Q_r and $L_{av}(r)$ is the average luminance of cluster Q_r (with an analogous definition for Q_s). We again reformulate this in log-space:

$$\ln(f_s) - \ln(f_r) = \ln\left(\frac{I_c(Q_s)L_{av}(Q_r)}{I_c(Q_r)L_{av}(Q_s)}\right) \tag{7}$$

However, clusters of the same chromaticity may actually have different reflectance, in which case the corresponding equations should not be included in the system. We adopt a conservative approach, and turn to the L coordinate to distinguish between different reflectances (e.g. light red and dark red). We define a threshold T_L of 5% of the maximum luminance of the image, and apply Equation 7 across clusters only if $|L_{av}(Q_r) - L_{av}(Q_s)| < T_L$.

Luminance regularization Last, we add a regularization equation to the system as follows:

$$\sum_i \ln f_i = 0 \tag{8}$$

With these three steps, we have posed our problem of obtaining intrinsic images as a system $\mathbf{AX} = \mathbf{B}$, where the number of equations equals the number of cluster boundaries plus the reflectance similarity equations plus the regularization equation. The elements of the unknown vector $\mathbf{X}$ are $x_i = \ln(f_i)$. In order to ensure the numerical stability of the solution, we solve the equivalent system $\mathbf{A^T AX} = \mathbf{A^T B}$ by means of a Quasi-Minimal Residual method (QMR) [BBC*94]. We found that using a standard Jacobi pre-conditioner yields good results in our context. From the solution vector $\mathbf{X}$ we can trivially obtain the final $f_i = \exp^{x_i}$ for each cluster Q_i. Last, we account for the fact that our algorithm assigns each pixel to a given cluster (thus creating hard borders between clusters), while in contrast all the input images present some blur at the edges due to the imaging process; we therefore apply a 5×5 median filter at the cluster boundaries to obtain the final shading image.

4. Results and Discussion

Most of the results shown in this paper have been generated from 16-bit RAW images, although in general our algorithm works well on linearized, 8-bit images. Larger versions of the results, along with detailed clustering and scatter plot data, are included in the supplementary material. Using our unoptimized code, our algorithm works at interactive rates e.g. an average image such as *clown* (see Figure 9), which results in 834 clusters, takes around 5 seconds on an Intel Core i5-2500 CPU at 3.30 GHz. In our tests, the most complex images may have up to 3000 clusters, resulting in about 15 seconds of processing time. Figure 5 shows how our algorithm successfully deals with the varied geometrical and texture complexities of several challenging web images, producing satisfying results (more examples can be found in the supplementary material). *Dragon* in the last row illustrates how 8-bit images may present pixel values close to zero, which cause numerical instabilities and are hard to disambiguate for any intrinsic images decomposition algorithm. This problem is greatly ameliorated with 16-bit images.

In Figure 9 we include an exhaustive comparison against state-of-the-art methods. We have tried to gather together all the published results common to most methods, but not all methods report results on all images. Compared to other automatic, single-image approaches, Tappen et al.'s technique [TFA05] shows perceivable reflectance artifacts in the shading image. Shen et al.'s method [STL08] suffers from posterization artifacts in the recovered reflectance (see for instance the *doll* image or the sky in *St. Basil*); additionally, almost all the shading images show severe continuity artifacts.

Shen and Yeo [SY11] provide a somewhat limited selection of images in their paper, both of which appear in our figure. It can be seen how in *baby*, the shading shows inconsistencies such as the exaggerated contrast between the legs and floor, while the eyes have been wrongly assigned to the shading layer.

The method of Gehler et al. [GRK*11], although producing reasonable results for the MIT dataset, tends to retain reflectance in the shading layer for natural images (see for instance *baby* image in Figure 9 or *synthetic doll* in the supplementary material). Moreover, their optimization function is computationally expensive, taking several hours, while our technique takes seconds to finish.

Last, the *doll* result using Weiss's approach [Wei01], de-

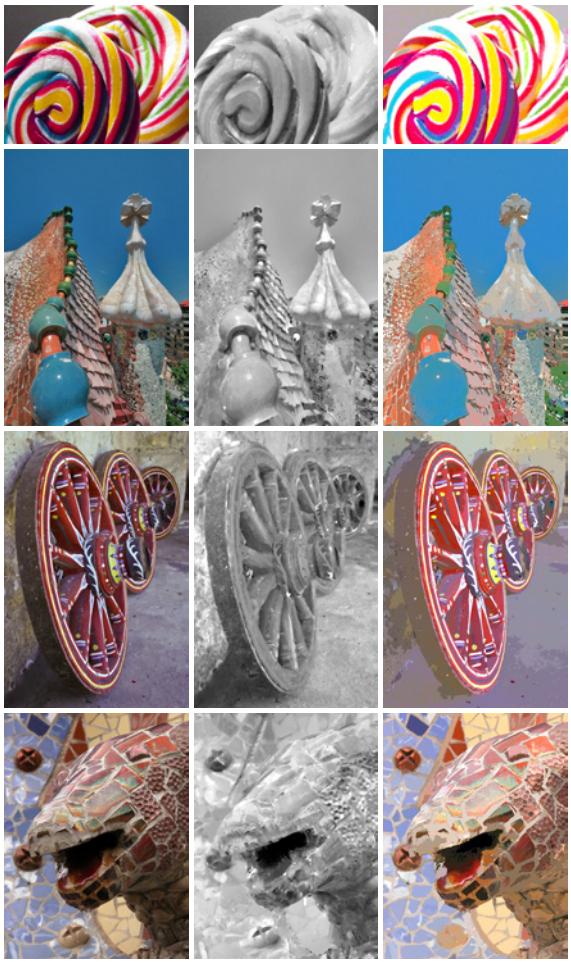

Figure 5: *Intrinsic images obtained with our method (using 8-bit input images). Left column: Input image. Middle column: Intrinsic shading. Right column: Intrinsic reflectance. First row: lollipop,* k = 10 *(original image by Thalita Carvalho, flickr.com). Second row: Batlló house,* k = 12 *(original image by lukasz dzierzanowski, flickr.com). Third row: wheels,* k = 16 *(original image by Angela Smith Kirkman). Last row: dragon,* k = 22 *(original image by Jordanhill School D&T Dept, flickr.com)*

spite using 40 images taken under carefully controlled settings, shows clear shading residuals in the reflectance layer, and wrong gradients in the shading image.

Our work yields results on-par with the user-assisted techniques by Bousseau et al. [BPD09] and Shen et al. [SYJL11], without requiring any user strokes. In *St. Basil*, our solution shows some artifacts especially visible in the black areas assigned to reflectance. Bousseau's result is free from such artifacts, although at the expense of requiring 81 user strokes, divided in three different kinds. Although the authors show that their method is robust to small perturbations applied to such strokes, placing them from scratch is probably challenging for unskilled users. In contrast, our automatic method does a better job at assigning the doll's eyelashes to the reflectance layer Furthermore, Bousseau's method is not well fitted for texture images with rich reflectance variations, as recently demonstrated by Dong and colleagues [DTPG11]. Our approach is free from such restriction, and it handles those cases well (see Figure 6 and Figure 10 in the supp. material).

Applications Accurate decomposition into intrinsic images can play an important role in a broad range of applications such as material editing or image relighting. Figure 8 shows some applications using our intrinsic decomposition. For retexturing we modify the reflectance layer and use the same normal-from-shading approach to deform the new textures. The relighting example, in this case, uses an image-based relighting algorithm [LMHRG10] that modifies the shading image before multiplying it by a sepia-shifted reflectance. Last, the wheel on the right shows another example of global reflectance manipulation. Furthermore, our technique can also be used in conjunction with others, such as the image-based material modeling technique recently presented by Dong et al. [DTPG11].

MIT dataset We have also tested our method using images from the MIT image dataset provided by Grosse et al. [GJAF09]. This dataset is designed to test intrinsic image decomposition methods, providing ground truth images for a variety of real-world objects. Figure 6 shows some examples of our results, compared against Color-Retinex, the combination of Weiss [Wei01] and Retinex and the recent approach by Dong et al. [DTPG11]. Note that Weiss's technique requires more than 30 input images, while the algorithm by Dong et al. requires user strokes. As we can observe, our algorithm obtains near optimal results in these cases. Please refer to the supplementary material for more results.

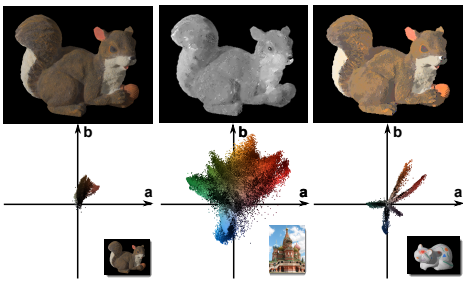

Figure 7: *A challenging case for our algorithm. Top row, from left to right: input image, and our intrinsic shading and reflectance. Notice the shadow close to the tail in the reflectance image. Bottom row: Comparison of scatter plots. From left to right: squirrel, St. Basil and raccoon. Notice the lack of chrominance variation in squirrel, compared to the other two plots.*

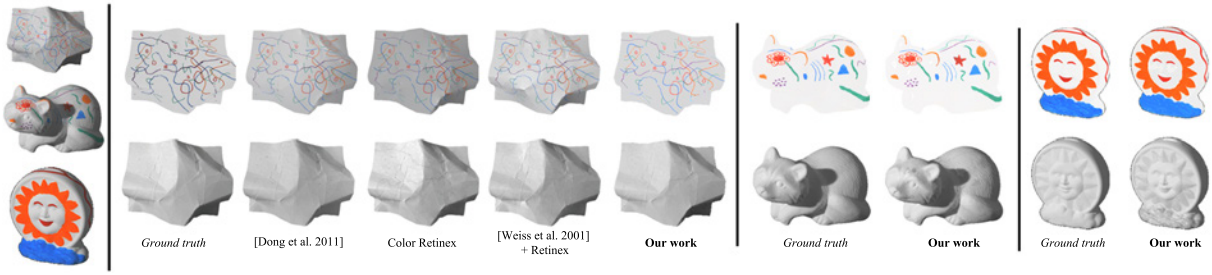

Figure 6: *Results using the MIT image dataset provided by Grosse et al. [GJAF09]. From left to right: Input images (paper, raccoon and sun); comparison with previous works for paper; results for raccoon and sun compared to ground truth. Top row: Intrinsic reflectance. Bottom row: Intrinsic shading.*

Other MIT images are much more ill-posed for any intrinsic images algorithm, presenting extremely complex combinations of shading, smooth varying textures and poor chromaticity. Figure 7 shows our results for the case of *squirrel* (*deer* can be found in the supplementary material). On the top row, we show the original image along with our recovered shading and reflectance. Although our method yields reasonable results, some artifacts are clearly visible. This is mainly due to the extremely poor chromaticity variation of the original image, which hampers our segmentation step. The bottom row shows scatter plots of *squirrel*, along with *St. Basil* and *raccoon* for comparison purposes. Note how, despite its overall lack of chromaticity, *raccoon* presents clear distinct areas in the (a,b) plane, due to the scribbles painted on its surface. To the best of our knowledge, there is no published method that can successfully deal with images such as *squirrel* or *deer*.

Figure 8: *Image edits accomplished using our intrinsic decompositions. Left: Re-texturing by editing the intrinsic reflectance. Center: Relighting of the previous result editing the intrinsic shading (sepia effect by editing intrinsic reflectance). Right: Global color edits on the intrinsic reflectance.*

Occlusion boundaries Our method lets us handle a wide variety of images even in the presence of occlusion boundaries or sharp edges, where our assumptions do not hold. This is due to our robust linear system formulation: The number of equations describing C^0 continuity at an offending occlusion boundary is usually very small compared with the total number of equations in the system, which diminishes their influence in the result. For instance, the system solved for *St. Basil* is made up of 3557 equations, from which only 267 belong to occlusion boundaries (a mere 7%). In all the images shown in this paper, we only found one case (*doll*) where the percentage of offending equations was higher than 10% (48 over 282 equations, 17%). This image presents a unique combination of reflectance distributions that makes it especially challenging for our algorithm. It shows a very uniform background (which translates into very few clusters adding equations to the system), whereas the doll itself has lots of patches of different reflectance in contact with such background, due to the striped pattern of its clothes (adding a relatively high number of occlusion boundary equations). Note how, in contrast, *St. Basil* presents most of the colorful patches inside the building itself (few equations describing occlusion boundaries with the sky). Even though the automatic results provided by our system may be considered satisfactory, we can improve them by simply creating a matte of the doll, which can be easily done with existing image editing tools, and then solve the two resulting images (doll and background) as separate problems. Figure 9, bottom row, shows the result of directly applying our method, and the improved results with our quick matting profile created in *Photoshop*©. Artifacts due to the inherent difficulty of this image can be seen across all methods.

5. Conclusions and Future Work

Decomposing an image into its intrinsic components is still an open problem with multiple potential applications. Automatic methods such as ours need to rely on reasonable assumptions or additional sources of information. On the other hand, existing user-assisted methods remain challenging for the average unskilled user, given the difficulty in telling apart the confounding factors of reflectance and shading in some situations. Our problem formulation is less restrictive than traditional Retinex-based methods, and allows us to relax our initial assumptions in certain cases. We have shown a wide variety of results, not only on natural scenes, but on the MIT dataset and even texture images as well. Additionally, we have also provided a thorough comparison with previous

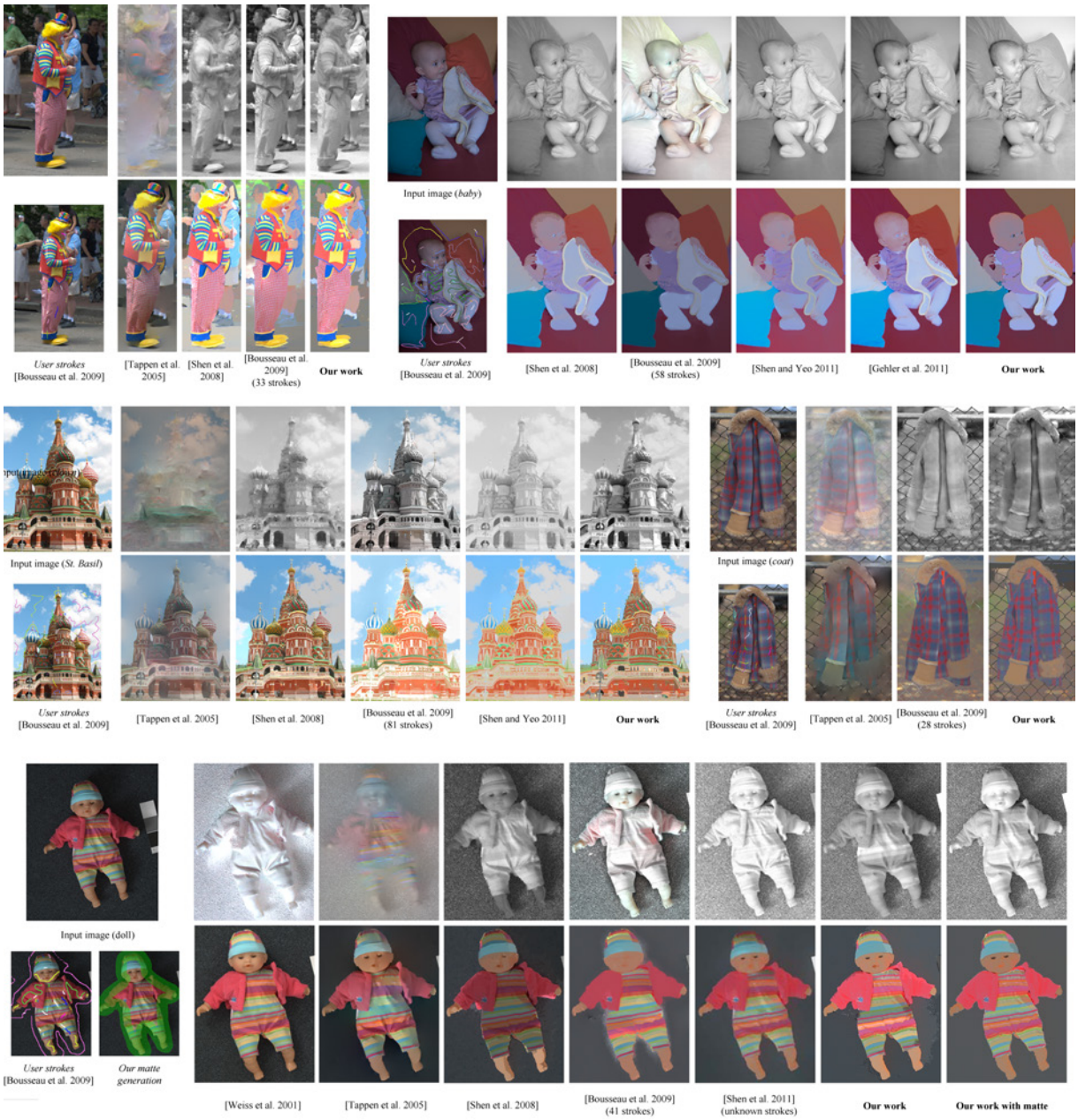

Figure 9: *Top left: clown. The methods by Shen et al. [STL08] and Tappen et al. [TFA05] introduce obvious artifacts in the shading. Our result is comparable with Bousseau et al. [BPD09] without requiring user strokes. Top right: baby. Our shading image is comparable with the one obtained by Bousseau et al. [BPD09] which needs 58 user strokes. Note that the result by Shen et al. [STL08] has quantized the leg of the baby while our result keeps the continuity in the shading. The shading by Shen and Yeo [SY11] is not totally homogeneous and contains reflectance information. Moreover, our reflectance image successfully captures the facial features, outperforming the other methods. Middle left: St. Basil. The methods by Shen et al. [STL08] and Tappen et al. [TFA05] introduce obvious artifacts in the shading. Bousseau et al.'s method produces great results, although it requires 81 user strokes (original image by Captain Chaos, flickr.com). Middle right: coat. Our result without any user strokes is not far from the result obtained by Bousseau et al. [BPD09] using 28 strokes. Bottom: doll. The automatic methods by Shen et al. [STL08] and Tappen et al. [TFA05] fail to obtain an homogeneous shading on the legs.*

approaches. Our algorithm produces better results than other automatic techniques on a broad range of input images, and on-par compared to user-assisted methods, but without the challenging task of providing the right strokes.

A potential line of future work would be to use our results as input to a simplified user interface, where it would be simpler to fix remaining artifacts. Also, our work could inspire and benefit from further research on segmentation methods. In conclusion, we believe that our approach offers an attractive trade-off between accuracy of the results, ease of use and efficiency.

Acknowledgments We would like to express our gratitude to the anonymous reviewers for their valuable comments. Thanks also to Martin Kiefel and Shen Jianbing for sending us images. This research was partially funded by the European Commission, Seventh Framework Programme, through the projects GOLEM (Marie Curie IAPP, grant agreement no.: 251415) and VERVE (Information and Communication Technologies, grant agreement no.: 288914), the Spanish Ministry of Science and Technology (TIN2010-21543) and a generous gift from Adobe Systems Inc. Elena Garces is also funded by a grant from the Gobierno de Aragón.

References

[BBC*94] BARRET R., BERRY M., CHAN T. F., DEMMEL J., DONATO J., DONGARRA J., POZO R., EIJKHOUT V., VAN DER VORST H., ROMINE C.: *Templates for the Solution of Linear Systems: Building Blocks for iterative Methods, 2nd Edition*. SIAM, 1994. 6

[BPD09] BOUSSEAU A., PARIS S., DURAND F.: User assisted intrinsic images. *ACM Transactions on Graphics (Proceedings of SIGGRAPH Asia 2009) 28*, 5 (2009). 2, 4, 7, 9

[BT78] BARROW H., TENENBAUM J.: Recovering intrinsic scene characteristics from images. *Computer Vision Systems* (1978), 3–26. 1

[DTPG11] DONG Y., TONG X., PELLACINI F., GUO B.: AppGen : Interactive Material Modeling from a Single Image. *ACM Transactions on Graphics (Proceedings of SIGGRAPH Asia 2011)*, 2 (2011). 3, 7

[FDB92] FUNT B. V., DREW M. S., BROCKINGTON M.: Recovering shading from color images. In *ECCV-92: Second European Conference on Computer Vision* (1992), Springer-Verlag, pp. 124–132. 2, 3

[FH04] FELZENSZWALB P. F., HUTTENLOCHER D. P.: Efficient graph-based image segmentation. *International Journal of Computer Vision 59* (2004), 2004. 4

[FS95] FREUND Y., SCHAPIRE R. E.: A decision-theoretic generalization of on-line learning and an application to boosting. In *European Conference on Computational Learning Theory* (1995), Springer-Verlag, pp. 23–37. 3

[GDFL04] GRAHAM D. FINLAYSON M. S. D., LU C.: Intrinsic images by entropy minimization. In *Proc. 8th European Conf. on Computer Vision, Praque* (2004), pp. 582–595. 2, 5

[GGLM11] GARCES E., GUTIERREZ D., LOPEZ-MORENO J.: Graph-based reflectance segmentation. In *Proceedings of SIACG 2011* (2011). 4

[GJAF09] GROSSE R., JOHNSON M. K., ADELSON E. H., FREEMAN W. T.: Ground-truth dataset and baseline evaluations for intrinsic image algorithms. In *International Conference on Computer Vision* (2009), pp. 2335–2342. 2, 7, 8

[GRK*11] GEHLER P. V., ROTHER C., KIEFEL M., ZHANG L., SCHÖLKOPF B.: Recovering intrinsic images with a global sparsity prior on reflectance. In *NIPS* (2011), p. 765. 2, 5, 6

[Hor74] HORN B. K.: Determining lightness from an image. *Computer Graphics and Image Processing 3*, 4 (Dec. 1974), 277–299. 2, 3

[JSW10] JIANG X., SCHOFIELD A. J., WYATT J. L.: Correlation-based intrinsic image extraction from a single image. In *Proceedings of the 11th European conference on Computer vision: Part IV* (Berlin, Heidelberg, 2010), ECCV'10, Springer-Verlag, pp. 58–71. 2

[KES*03] KIMMEL R., ELAD M., SHAKED D., KESHET R., SOBEL I.: A variational framework for retinex. *International Journal of Computer Vision 52* (2003), 7–23. 2, 3

[KMN*04] KANUNGO T., MOUNT D. M., NETANYAHU N., PIATKO C., SILVERMAN R., WU A. Y.: A local search approximation algorithm for k-means clustering. *Computational Geometry: Theory and Applications 28* (2004), 89–112. 5

[LBD11] LAFFONT P.-Y., BOUSSEAU A., DRETTAKIS G.: *Rich Intrinsic Image Separation for Multi-View Outdoor Scenes*. Research Report RR-7851, INRIA, Dec. 2011. 3

[LM71] LAND E. H., MCCANN J. J.: Lightness and retinex theory. *Journal of the Optical Society of America 61*, 1 (1971). 2

[LMHRG10] LOPEZ-MORENO J., HADAP S., REINHARD E., GUTIERREZ D.: Compositing images through light source detection. *Computers & Graphics 34*, 6 (2010), 698–707. 7

[LWQ*08] LIU X., WAN L., QU Y., WONG T.-T., LIN S., LEUNG C.-S., HENG P.-A.: Intrinsic colorization. *ACM Transactions on Graphics (Proceedings of SIGGRAPH Asia 2008)* (2008), 1–9. 3

[OW04] OMER I., WERMAN M.: Color lines: image specific color representation. In *Conference on Computer vision and pattern recognition* (2004), CVPR'04, IEEE Computer Society, pp. 946–953. 4

[STL08] SHEN L., TAN P., LIN S.: Intrinsic image decomposition with non-local texture cues. *Computer Vision and Pattern Recognition, IEEE Computer Society Conference on 0* (2008), 1–7. 2, 3, 4, 6, 9

[SY11] SHEN L., YEO C.: Intrinsic images decomposition using a local and global sparse representation of reflectance. In *Computer Vision and Pattern Recognition* (2011), IEEE, pp. 697–704. 2, 6, 9

[SYJL11] SHEN J., YANG X., JIA Y., LI X.: Intrinsic images using optimization. In *Computer Vision and Pattern Recognition (CVPR)* (2011), IEEE, pp. 3481–3487. 2, 7

[TAF06] TAPPEN M. F., ADELSON E. H., FREEMAN W. T.: Estimating intrinsic component images using non-linear regression. In *Conference on Computer Vision and Pattern Recognition - Volume 2* (2006), CVPR '06, IEEE Computer Society, pp. 1992–1999. 3

[TFA05] TAPPEN M. F., FREEMAN W. T., ADELSON E. H.: Recovering intrinsic images from a single image. *IEEE Transactions on Pattern Analysis and Machine Intelligence 27* (2005), 1459–1472. 3, 6, 9

[Wei01] WEISS Y.: Deriving intrinsic images from image sequences. *Computer Vision, IEEE International Conference on 2* (2001), 68. 3, 6, 7