EUROGRAPHICS 2017
V. Ostromoukov and M. Zwicker
(Guest Editors)

Volume 36 (2017), Number 2
STAR – State of The Art Report

Intrinsic Decompositions for Image Editing

Nicolas Bonneel[1] and Balazs Kovacs[2] and Sylvain Paris[3] and Kavita Bala[2]

[1]CNRS & Univ. Lyon 1 [2]Cornell University [3]Adobe Research

Figure 1: *We evaluate state-of-the-art intrinsic image decomposition algorithms based on their ability to produce seamless, artifact-free results for image edits. To fairly compare different methods, we automatize the image editing process. Left to right, by Poisson-based inpainting of the reflectance layer, we remove a logo on a shirt using the metod of Barron et al. [BM15], we add a picnic blanket over a shadow with the method of Grosse et al. [GJAF09], and add a painting over colored shadows with the method of Bousseau et al. [BPD09].*

Abstract

Intrinsic images are a mid-level representation of an image that decompose the image into reflectance and illumination layers. The reflectance layer captures the color/texture of surfaces in the scene, while the illumination layer captures shading effects caused by interactions between scene illumination and surface geometry. Intrinsic images have a long history in computer vision and recently in computer graphics, and have been shown to be a useful representation for tasks ranging from scene understanding and reconstruction to image editing. In this report, we review and evaluate past work on this problem. Specifically, we discuss each work in terms of the priors they impose on the intrinsic image problem. We introduce a new synthetic ground-truth dataset that we use to evaluate the validity of these priors and the performance of the methods. Finally, we evaluate the performance of the different methods in the context of image-editing applications.

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

1. Introduction

The rich visual world that surrounds us is the result of the complex interplay between light and matter. Light reaches an observer typically after several interactions with physical objects, each of them having a nontrivial effect on its spectrum, different light wavelengths being affected differently depending on the properties of the material involved. Further, cause and effect can be separated by a large distance when an object casts a shadow far away from its actual position. Eventually, all these effects are conflated into a single image. Considered in its full generality, the image formation process may seem impossible to invert because the phenomena involved are too diverse and too complex to be disentangled. Yet, human observers effortlessly identify shadows and recognize object colors under all but the most extreme lighting conditions. *Intrinsic decomposition* of digital images is a task inspired from this ability. Originally, Barrow and Tenenbaum [BT78] sought to characterize properties inherent in a scene such as the color and geometry of objects independent of viewing conditions. With time, the geometric aspects of this original goal became associated with stereo and multi-view reconstruction, and intrinsic decomposition took a more focused meaning becoming synonymous with reflectance estimation. In this manuscript, we follow this trend and define intrinsic decomposition as the task of separating the effect of the scene *illumination* from that of the material *reflectance*. This task relies on a simple image formation model that explains each pixel as the product of two RGB triplets, one for the light color and one for material reflectance. This model analyses only the last bounce of the light transport and makes a number of simplifying assumptions such as "all materials are Lambertian" and "no participating media", but it nonetheless provides a powerful means to reason about light and object colors that has proven to be useful to many image editing applications. For instance, recoloring an object in an image is a nontrivial task even for a uniform-color object because shading variations make

© 2017 The Author(s)
Computer Graphics Forum © 2017 The Eurographics Association and John Wiley & Sons Ltd. Published by John Wiley & Sons Ltd.

some parts bright and others dark, possibly in a discontinuous way if the object exhibits sharp geometric features. The same task with an intrinsic decomposition is straightforward since the reflectance of the object is constant. The rest of this manuscript presents the concept of intrinsic image decomposition in more detail, describes the main existing algorithms to compute such a decomposition, and reviews the most common use cases.

While some approaches consider multiple images as input (from multiple viewpoints [CBLD11, LBP*12, Laf12, LBD13, HWU*14, Duc15, XLL*16], varying illumination [Wei01, MLKS04, Yu16], or different focal distances [SSN16]), images with depth information [BM13, CK13, JCTL14, HGW15], multi-spectral images [SW09], videos [YGL*14, BST*14, KGB14, SYC*14, MZRT16], lightfields [GEZ*16, AG16] or even videos with depth [LZT*12], this document focuses on the use of a single RGB image as input, and emphasizes image editing applications. This is motivated by the wide availability of this kind of data and the need for illumination-aware image editing tools.

Aside from pedagogical content, this document makes the following contributions. First, we evaluate priors commonly used in the literature in the context of realistic and complex scenes. Second, we introduce a small but realistic synthetic ground-truth dataset based on PBRT [PH10], Mitsuba [Jak10] and LuxRender [Ver07] scenes. Third, we evaluate 10 recent methods on real photographs in the context of image editing applications.

2. Problem formulation

The intrinsic decomposition problem seeks to decompose an image into the product of illumination and reflectance layers. This section exposes the motivation behind this problem, as well as various assumptions and priors that numerically help solve for this decomposition.

2.1. Intrinsic Decomposition

In a Lambertian scene, material reflectance does not depend on viewing direction or illumination incidence. This simplifies light transport and allows for the writing of a simplified (yet exact) physically-based image formation model. In this context, we have

$$I(x,\lambda) = \rho(x,\lambda)\, L(x,\lambda) \tag{1}$$

where x is the pixel position, λ the light wavelength, I the rendered image, ρ the diffuse albedo and L is a term which depends on light and geometry. In the following, we will call ρ the reflectance, and L the illumination.

This model holds in the continuous image plane domain, but the spatial filtering and sampling that occur inside real cameras makes the equality in the equation above break on traditional pixel grids when geometry, textures or illumination vary within a single pixel, or in the presence of lens or motion blur. Cameras also often store non-linearly processed pixel values, for instance, to gamma correct, enhance or white balance images.

The inverse problem of forming an image using the model in Eq. 1 is the problem of recovering the reflectance and illumination given an already formed image. This process is called intrinsic decomposition. Eq. 1 makes the intrinsic decomposition problem precisely defined in terms of photometric quantities.

Other techniques also try to understand the role of lighting or textures in images, or relate to intrinsic decompositions:

- Reflectance map extraction [HS79, RRF*15]: This generalizes the intrinsic decomposition problem to the recovery of arbitrary reflectance functions. Rematas et al. offers a deep learning approach that recovers a full hemispherical reflectance function per pixel [RRF*15] – the reflectance map.
- Shadow extraction. This closely related problem consists of detecting and extracting shadows. Under low-frequency lighting conditions, shadows become softer and a precise definition of shadows becomes an issue. Relating this problem to that of intrinsic decomposition, Isaza et al. [ISR12] evaluates intrinsic decomposition methods to detect shadows.
- Light estimation. This tries to uncover the lighting conditions of a scene, for example, by extracting directional light sources [LMGH*13] or environment maps [LE10], from an image. This problem becomes difficult in the presence of localized light sources or inter-reflections, in the absence of a 3d geometric model of the scene.
- Specularity removal. This is a different intrinsic decomposition approach that separates the diffuse from the specular reflection components. While this also extracts *intrinsic images* in the sense of Barrow and Tenenbaum [BT78], the term *intrinsic decomposition* now most often refers to the separation of the diffuse from the illumination components (though exceptions exist [BvdW11]). The interested reader may find further information in the survey of Artusi et al. [ABC11].
- Color constancy. When a colored object is illuminated by light sources of different colors, for example in a sunset, or in indoor lighting, the object appears to humans as having kept its original color. A common photographic operation is to try and compensate for the light source chromaticity – a process often called "white balancing" or color constancy correction. This is typically achieved globally using a grey card or a colored chart [CPCB15]. In the presence of multiple colored light sources at the same time, this operation can be performed locally [HMP*08, BBPD12]. A perfectly local color constancy would recover the illuminant color at each pixel, which would correspond to the chromaticity of the illumination layer $L(x,\lambda)$.
- Texture-Structure decomposition. This separates the high frequency textural elements from lower frequency structures [AGCO06]. The definition of a texture, however, depends on the scale of the observed element. For instance, a forest canopy can belong to the structure if seen from sufficiently close, and it becomes part of textures if seen from sufficiently far. This decomposition has seen applications for intrinsic decompositions [JCTL14, BHY15], and image-based material editing [BBPA15].

A major challenge of intrinsic image decomposition is that the image formation model $I = \rho L$ is ill-posed because if ρ_0 and L_0 satisfy the model, i.e., $I = \rho_0 L_0$, then $a\rho_0$ and $\frac{1}{a}L_0$ also satisfy the model for any nonzero a, including the case where a is spatially varying (Fig. 2). Concretely, this means that the absolute ground-truth decomposition is unattainable unless additional absolute measurements are available, e.g., using a light meter, which is not the case

© 2017 The Author(s)
Computer Graphics Forum © 2017 The Eurographics Association and John Wiley & Sons Ltd.

Figure 2: *The intrinsic decomposition ambiguity. Top row. An input image I is decomposed into the product of reflectance and illumination layers ρ and L using [BPD09]. Bottom row. The product of these altered reflectance and illumination layers ρ' and L' exactly reconstructs the input image I.*

in practice. While absolute values cannot be estimated, relative accuracy can be reached and is useful, e.g., to identify regions of constant albedo or illumination. For this reason, decompositions that differ only by a constant multiplicative factor are considered equivalent [Hor74]. Such solutions provide the same relative estimates since the constant multiplicative factor vanishes when one takes the ratio of any two values. And in this context, the notion of exact solution still exists and corresponds to any decomposition that is equal to the ground truth up to a constant multiplicative factor. This also enables the quantitative evalution of the accuracy. For instance, one can use computer-generated images for which ground truth is available as a benchmark. Instead of directly reporting the error between a given decomposition and the ground truth as is often the case in other contexts, one first searches for the multiplicative factor that minimizes the error, which is the value reported. We will discuss further how to evaluate results later in the manuscript. However, even considering relative accuracy, the image formation model leaves room for too much freedom: the model still holds when a is spatially varying. In fact, twice as many unknown values (ρ and L) as known quantities (I) have to be estimated. To attain satisfactory decompositions, priors and constraints are needed to reduce the number of unknown variables and disambiguate the problem. This represents the heart of intrinsic decomposition research, and is discussed next.

2.2. Common priors

Priors statistically model one's beliefs about intrinsic decompositions and help disambiguating decompositions. A number of priors have been introduced in the literature, as well as assumptions and user constraints, detailed below and summarized in table 8. In the following, we assume the camera sensor response curve has been properly taken care of (for instance, by directly working on raw images or using photometric calibration techniques [KGS05]).

Monochromatic illumination (MI). Often, the illumination layer is assumed to be grayscale. Up to a white balancing step of the input photograph, this corresponds to the use of a single light color and reduces the illumination layer to a single scalar value per pixel $L(x)$ instead of $L(x,\lambda)$. This is the most common assumption (see Table 8), and only few approaches allow for colored lights [CSBC09, BPD09, CCFI14, BM15, TNY15] ([BM13] for RGB-D images). It is interesting to note that the monochromatic assumption often only applies to the light color as seen by the camera. That is, the light itself need not be monochromatic as long as the integral of its energy distribution over wavelengths is the same for each color sensor. The main disadvantage of this assumption is that inter-reflections are most often colored [CRA11], and so, this assumption often fails to capture illumination effects that are due to indirect lighting. However, the monochromatic illumination assumption can be considered as a prior by designing a soft penalty for strongly colored illuminations while still allowing for colored lighting. For instance, Chang et al. [CCFI14] use a Gaussian Process that correlates color channels, and favors grayscale illuminations.

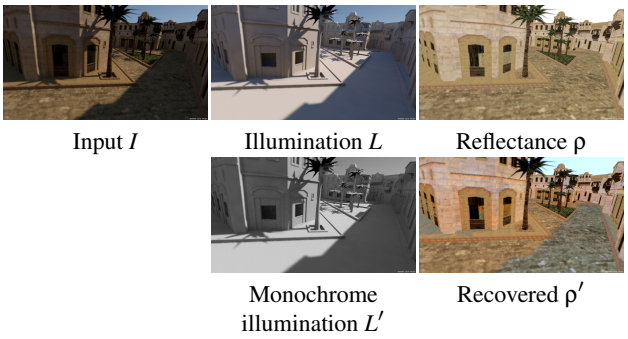


Figure 3: *If the monochromatic illumination assumption is used, colored shadows cannot be removed from the reflectance layer. First row. The input image I is decomposed into reflectance and (colored) illumination layers ρ and L. Second row. We desaturate the illumination layer L', and recover the corresponding reflectance image $\rho' = \frac{I}{L'}$, hence simulating the best intrinsic decomposition achievable assuming grayscale illumination. This new reflectance partly contains the colored shadow. The monochromatic illumination assumption is still widely used to date (table 8).*

Retinex (R). Following a series of experiments involving colored light sources illuminating patterns resembling Piet Mondrian's paintings, Land et al. [LM71] determined that the human eye made sense of relative intensities to form a representation of material reflectance. In this representation, sharp transitions (or edges) are perceived as changes in reflectance properties, while smoother variations are seen as changes in illumination. They coined the term *Retinex*, from *Retina* and *Cortex*, for the eye representation of lightness – a perceptual quantity correlated with reflectance. They further built a *Retinex Machine* reproducing the reflectance of a grayscale color wheel based on a 1D strip of sensors. This approach has been extended in 2-d by Horn [Hor74] who provides numerical tools and physical interpretations of the 2-d Retinex problem. The Retinex model fails for hard shadows or occlusion boundaries since they produce strong edges associated with illumination variations, but it works reasonably well for smooth surfaces. This prior is often implemented by

© 2017 The Author(s)
Computer Graphics Forum © 2017 The Eurographics Association and John Wiley & Sons Ltd.

thresholding gradients [Hor74], using a sparse norm on reflectance gradients [BST*14] or similarly, on differences between adjacent gradients [BHY15], using Gaussian Processes [CCFI14] or other probabilistic frameworks [LB14], or using a quadratic penaly term in a non-linear optimization energy [SYJL11,SY11]. It has also been implemented as a sparse TV norm on shading gradients [CRA11]. The piecewise flatness of reflectance values can be modeled via a smoothness term on the reflectance [BM15], by clustering pixels into superpixels [BHY15, ZIKF15], or using Conditional Random Fields (CRFs) [BHY15].

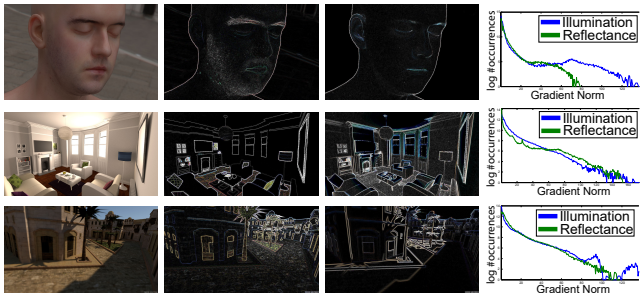


Figure 4: *This illustrates the gradient norm of the reflectance and illumination components. The last column shows the* log*-histogram of gradient occurrences. A smoother illumination layer would exhibit a steeper decreasing curve for illumination gradients than for reflectance gradients. This is however hardly the case in practice: in cluttered scenes, occlusion boundaries dominate and produce strong gradients in both the reflectance and illumination layers.*

Edge is either a reflectance or illumination change (EoI). This states that it is unlikely for a single image edge to result from both a change of illumination and reflectance at the same time [LM71]. This has been used to classify image gradients [Hor74, GJAF09, LSX09, TFA05]. However, this assumption is unlikely to hold at occlusion boundaries: in this case, a silhouette edge separates an occluding surface whose normal is parallel to the image plane (assuming orthographic projection) [BM15] from an occluded surface whose normal can be arbitrary. This change in normal direction often coincides with a change in illumination, and in many cases, these two surfaces will have different reflectances as well. This prior can be seen as a strong version of the Retinex prior.

Chrominance variations are more likely to be reflectance variations (CR). When chrominance varies abruptly, this is more likely due to a change of reflectance than illumination [GJAF09, TFA05, ZTD*12].This can be implemented as a varying gradient threshold for abrupt changes in chrominance [TFA05, GJAF09], by weighting chrominance and luminance differently in the pairwise cost functions used in CRF-based frameworks [BBS14, ZKE15], or by introducing a quadratic penalty term which enforces smooth reflectance only in places where the chrominance is smooth [SY11]. A similar observation concludes that if hue variations correlate positively with intensity variations, this is more likely due to changes in reflectance [JSW10].

Low rank reflectances (LRR). This assumes that locally, within a small neighborhood, reflectance values form a 2-d affine subspace of the RGB space [BPD09]. This prior can be seen as a reflectance smoothness prior.

Sparse reflectance values (SRV). This assumes that most color variations are due to illumination, and that, in fact, few different reflectance values make up a typical image. This assumption is expected to work best for photographs of man-made scenes. This is often implemented via a color clustering step [GMLMG12, BBS14, LYZ15], or a sparsity constraint on reflectance [SY11], or even using superpixels [BHY15]. Alternatively, this prior can be cast in the realm of information theory. Using compression-based complexity measures, Nicola et al. show reflectance has lower complexity than illumination [SF15]. Similarly, Barron and Malik minimize the entropy of the log-reflectance [BM15] to obtain parsimonious reflectance values.

Some reflectance values are more likely (RML). Barron and Malik [BM15] assumes some reflectance values are more likely than others. This is implemented by building a 1-d histogram of log-reflectance values of a ground truth dataset, and using it as a prior. This prior helps disambiguate the overall light color from reflectance colors: a blue pixel will be more likely the result of a white light illuminating a blue object, rather than a magenta light illuminating a cyan object, if cyan reflectances are a priori less likely than blue reflectances.

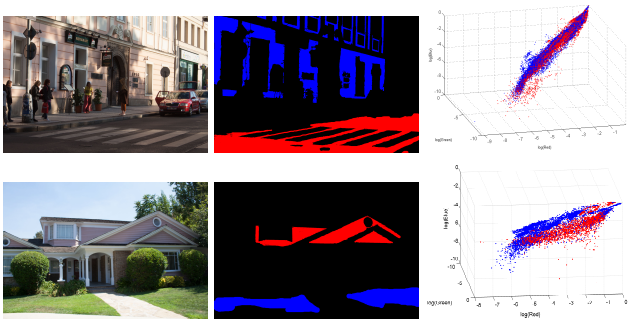
Mean correlates with variance (MV). Under illumination variations, the local mean of pixel values within a neighborhood should vary in the same direction as the local variance [JSW10]. This is due to the fact that illumination acts multiplicatively on reflectance, which can be detected via correlation analysis.

Planckian lighting (PL). Under skylight and a narrow-band camera sensor, it can be shown that pixels belonging to objects of the same reflectance form a single line in log-RGB space when varying the lighting condition [CPCN13, FDL04, LYZ15]. This is often implemented by clustering lines of log-RGB pixel values, similar to the color lines model [OW04] (though not performed in the log-domain), or via entropy minimization [FDL04].

Non-local constraints (NLC). The goal of this prior is to find pixels that are most likely of the same reflectance value within an image. The idea is to compare texture descriptors, such as pixel neighborhoods, and if two descriptors agree, they most likely represent the same structure repeated at a different location [STL08, ZTD*12]. This introduces long-range reflectance constraints. The difficulty lies within the comparison function, which assumes neighbors can be compared in a way agnostic to illumination variations... a chicken-and-egg problem! In practice, simply dividing pixel color values by their intensity (i.e., taking the pixel chromaticity) often serves as a good proxy [STL08, ZTD*12].

User-defined constraints (UC). The difficulty of automatic intrinsic decomposition has led researchers to rely on the user to add constraints. This typically involves asking the user to mark pairs of pixels of similar reflectance or illumination, or brush areas of similar

© 2017 The Author(s)
Computer Graphics Forum © 2017 The Eurographics Association and John Wiley & Sons Ltd.



Input image Reflectance masks log-RGB scatterplot

Figure 6: *A Planckian lighting prior assumes that under skylight, pixels of the same reflectance belong to a single straight line in* log*-RGB space. We use outdoor linear images and plot* log*-RGB color distributions of two constant reflectance surfaces (second column). This hardly results in distinct lines. This prior may produce artifacts when used on images taken in uncontrolled settings.*

reflectance or brush absolute illumination values [BPD09]. These constraints make the decomposition *interactive* instead of *fully automatic*. While this requires effort from users, this often yields better decompositions. This is simply accomplished by adding these constraints in a linear system [BPD09, BST*14] or similarly adding a quadratic penalty term [SYJL11].

Figure 7: *Scribbles help disambiguate intrinsic decomposition. Using a single input photograph, we do not expect fully automatic methods to differentiate a picture of an object displayed on a wall from an actual 3D object. From left to right, input photograph, illumination layer obtained using the method of Bonneel et al. [BST*14] without, and then with user interactions to remove illumination variations within each picture on the wall.*

Data-driven (DD). Similar to user-defined constraints on a specific image, machine learning approaches leverage ground-truth decomposition databases to guide further decompositions. For instance, classifiers have been trained via the output of the 875,833 comparisons across 5,230 photos of the Intrinsic in the Wild dataset [BBS14], which provides a *data-driven prior* often implemented using a Convolutional Neural Network [NMY15, ZKE15, ZIKF15]. The method of Tappen et al. [TAF06] use a Mixture of Experts Estimator to predict gradients (and other local linear constraints) learned from simple images of crumpled paper (see Sec. 4.1), while the earlier method [TFA05] uses an AdaBoost classifier on rendered images of fractals and ellipses. In a simpler way, a ground-truth dataset can be used to learn the hyper-parameters of a model by cross-validation [BBS14, CCFI14, BM15]. Barron and Malik [BM15] also build non-parametric models to construct an absolute color prior using the MIT and OpenSurfaces [BUSB13] datasets.

Human faces. In the specific context of human faces, additional information may be used. In particular, Li et al. [LZL14] uses a known skin reflectance model and a dataset of 3d face geometries.

While this state-of-the-art report focuses on single image intrinsic decompositions for image editing applications, we will briefly mention other specific priors and assumptions that have been introduced in the context of more general intrinsic decompositions.

Temporal consistency. Working on videos, reflectance can be assumed to remain temporally consistent across video frames [LZT*12, BST*14]. RGB-D videos allow for easier tracking via the 3d reconstructed scene [LZT*12], when RGB videos would require an illumination agnostic optical flow in the case of moving shadows [BST*14] (again, a chicken-and-egg problem).

The ambient occlusion model. Using multiple images of the same object under various lighting conditions, Hauagge et al. [HWBS16] replace the illumination term in the intrinsic decomposition by a scaled ambient occlusion term. Ambient occlusion estimates the fraction of the hemisphere visible from any point in the scene, regardless of the lighting conditions.

2.3. Numerical techniques

Various numerical techniques have been investigated to account for (part of) these priors, some of which were described in Sec. 2.2. The mathematical formulation of these priors and of the image formation model matters in practice and is key to designing practical algorithms. We next review a few standard approaches.

A log-transform often conveniently rewrites the product $I = \rho \times L$ into a sum $\log I = \log \rho + \log L$ [LM71, Hor74]. Changing the name of these variables (here, lowercase letters denoting log values), this can be written as $i = r + l$, simplifying a non-linear to a linear relationship. Numerical tools from linear algebra can then be used, particularly when priors can also be expressed as linear relationships (for example, as the result of the minimization of a quadratic energy). This often leads to sparse linear systems [GJAF09]. Interestingly, log-transforms also often render these methods robust to gamma correction as $\log I^\gamma = \gamma \log I$. Priors evoking smoothness or sparsity can often be expressed using gradients – for instance, methods based on the Retinex theory may classify image gradients as belonging to *either* the illumination layer *or* the reflectance layer [TFA05]. In conjunction with the log-transform, the intrinsic decomposition can be advantageously rewritten as $\nabla \log I = \nabla \log \rho + \nabla \log L$.

When priors cannot be easily cast as linear constraints (or quadratic penalty terms), full non-linear solvers have been used, such as l-BFGS [LZL14, BM15].

Alternatively, a probabilistic approach can be taken via CRFs [BBS14]. Here, priors are expressed via probabilistic models as exponentially decreasing functions of some energies, whose joint negative log-likelihood is minimized for. In traditional CRFs, message passing is used to minimize the energy function, though Bell et al. [BBS14] approximate it via high-dimensional filtering.

Finally, with the recent advances in machine learning and the availability of ground-truth datasets, learning-based approaches have

© 2017 The Author(s)
Computer Graphics Forum © 2017 The Eurographics Association and John Wiley & Sons Ltd.

emerged. These methods guide the decomposition by directly classifying image gradients that are propagated using a Markov Random Field (MRF) [TFA05], more generally regress image filters that are propagated via a pseudo-inverse [TFA05], or use full-fledged convolutional neural networks [TNY15] or deep belief networks [TSH12]. Jointly learning depth and intrinsic decomposition via deep convolutional network has seen some success [SBD15], and joint estimation of shape, illumination (as a spherical low-frequency environment map) and intrinsic decomposition performs well [BM15]. Recently, multiple works emerged which learn a reflectance prior from the pairwise judgments of Bell et al. [BBS14] with convolutional neural networks: Narihira et al. [NMY15] learn a lightness classifier, Zhou et al. [ZKE15] integrate the learned priors into Bell et al. [BBS14]'s CRF framework, and Zoran et al. [ZIKF15] solve a quadratic program on super-pixels with these data-driven priors. Other preliminary work on deep architectures for intrinsic decompositions are under investigation [Vit15, SL16, LVVG16].

3. Typical applications

Decomposing an image into illumination and reflectance components has several applications. First, this allows for understanding scenes better. For instance, it could be used to understand the role of the illumination with respect to intrinsic reflectance color in the popular blue-black dress meme [LSHC15] (see Fig. 9). Although the actual dress is blue and black, 30% of people perceive this dress as white and gold, due to a different perception of the illuminant [LSHC15] and other biological factors [RHTP16].

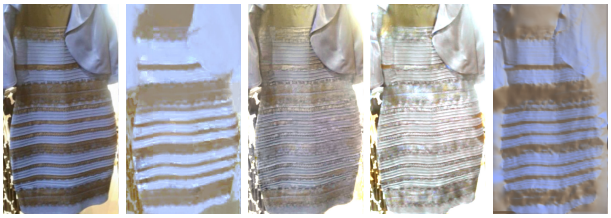


Figure 9: *Left. The decomposition of Bousseau et al. [BPD09] may help understand the color variations of this popular dress meme [LSHC15]. Right. Although being one of the best performing methods, the priors introduced by Barron et al. [BM15] swap the illumination and reflectance layers. In this example, the priors for the reflectance layer perhaps better describe the illumination layer and vice-versa.*

Understanding scenes is what makes computer vision systems powerful. As such, intrinsic decompositions have been used in computer vision: Kong and Black [KB15] use the intrinsic decomposition as a robust feature for transferring depth estimates in videos, Kong et al. [KGB14] use the reflectance for improving optical flow estimates in the context of intrinsic videos, and Ye et al. [YGL*14] use their intrinsic video decomposition to spatially segment videos. Isaza et al. [ISR12] assess intrinsic decomposition algorithms to detect shadows on outdoor scenes, but found that residual textures often remain in the shading layer.

In computer graphics, giving access to illumination and reflectance layers has large potential for artistic image manipulations. Editing the textures of images in a way that preserves illumination variations is often given as an example to illustrate the success of intrinsic decomposition methods [BHY15, BPD09, BM15, BST*14]. Other manipulations of the reflectance layer include color histogram matching [YGL*14], or stylization, for instance with edge detectors [YGL*14]. Similarly, altering lighting conditions is sometimes proposed, though without geometry estimates it is more difficult to illustrate illumination edits that are consistent with the existing geometry. For instance, Bousseau et al. [BPD09] invert the colors of the illumination layer to simulate a night photograph from an input daylight photograph, and Ye et al. [YGL*14] manipulate the histogram of the illumination layer to make diffuse objects look more shiny. Garces et al. [GMLMG12] use a more complex image relighting method [LMHRG10] which estimates rough geometry, and apply it to the illumination layer (and also apply a sepia filter on the reflectance). Li et al. [LZL15] use an intrinsic decomposition for editing the makeup in photographs of faces, but they require an additive layer representing highlights obtained using a previous method of Li et al. [LZL14]. Bonneel et al. [BST*14] use an intrinsic decomposition to composite two videos by combining both illumination layers. Bi et al. [BHY15] integrate 3d objects into images with consistent illumination by estimating an environment map locally using a method of Barron and Malik [BM12]. The joint estimate of an intrinsic decomposition, geometric information and an environment map, allows Barron and Malik [BM15] to locally alter the geometry of objects with displacement maps and change lighting conditions, by re-rendering the object under the new conditions. Liu et al. [LWQ*08] use a gradient domain decomposition on a color image to colorize a grayscale image via color transfer.

Editing results are occasionally illustrated on textureless surfaces, e.g., to alter the color of a uniform object. While it is un-challenging, it is also primarily more easily performed via simple luminance-chrominance adjustments! In fact, a luminance-chrominance decomposition is a correct and valid intrinsic decomposition for textureless uniform surfaces. We hence recommend comparisons against naive baselines, as they often perform reasonably well for many image editing applications.

4. Evaluation and Comparisons

Evaluating intrinsic decomposition methods is not a trivial task. The seminal paper of Grosse et al. [GJAF09] introduced the ground-truth dataset now known as the *MIT dataset*. This paper advocated for the LMSE metric, a mean-square error computed and averaged locally, within pixel neighborhoods. Bell et al. [BBS14] relied on perceptual experiments to determine how two pixels differ in their reflectance, and compare these judgments with the results of automatic algorithms, leading to the WHDR, *weighted human disagreement rate*, metric. Instead, in this paper we evaluate intrinsic decomposition methods based on their applicative ability. That is, we do not need a decomposition to be accurate, as long as it is sufficient to perform a given image editing task. We thus only evaluate the result of the image editing process. This section describes this approach, as well as a more classical evaluation using LMSE on a new ground truth realistic synthetic dataset.

© 2017 The Author(s)
Computer Graphics Forum © 2017 The Eurographics Association and John Wiley & Sons Ltd.

4.1. Datasets

Ground truth datasets are important for evaluating intrinsic decomposition methods, but also to provide data for training approaches based on machine learning. Datasets often precisely meet the assumptions of methods being assessed. This report instead focuses on realistic image editing applications, and we believe slightly violating assumptions is reasonable in this context.

In particular, the MIT dataset [GJAF09] relies on isolated objects with black background to minimize interreflections, and a single directional light source. The reflectance component is obtained by coating the object with a white diffuse paint, while specularities are removed via a polarizing filter. This technique does not allow for colored indirect illumination as the white coating may not reflect the same colored light as the initial object. Sierra used this same technique to extend this intrinsic decomposition database [Ser15]. The MIT dataset has been widely used for benchmarking intrinsic decomposition methods, but has been deemed "not representative of the variety of natural objects in the world" [BM15], and would be hardly useful for assessing methods in ecological contexts. Beigpour et al. have more recently extended this dataset [BKK15, BHK*16] under multiple lighting and viewing conditions using the procedure of Grosse et al. [GJAF09]. The two datasets each contain 5 scenes under 17 illumination conditions, and contain ground-truth depth and specularity information. They consist of two objects resting on a planar surface and remain of moderate complexity. The second dataset features 6 view conditions. A similar approach taken by Tappen et al. [TAF06] uses colors to capture reflectance and illumination independently. In practice, they color a piece of paper using a green pen: this color is invisible in the red channel of the image, and the red channel is taken as the ground-truth illumination layer. They use this technique to build a ground truth database of isolated sheets of crumpled paper.

Bell et al. introduced an extensive crowdsourced dataset of pairwise reflectance comparisons for photos 'in the wild' (i.e., for Flickr images taken from real world settings) [BBS14]. MTurk workers were asked to determine whether random pairs of points share the same reflectance, or if one point has a darker surface color. The IIW dataset comprising 5,230 photos, includes 875,833 reflectance comparisons, and currently is the largest database for benchmarking intrinsic image decomposition algorithms. However, despite its size, since the dataset provides ground-truth data for sparse pairs of points, it is not amenable to evaluating high-frequency errors that often occur in the illumination component, which can lead to artifacts in re-texturing applications.

Beigpour et al. [BSV*13] introduced a synthetic dataset by rendering 3d scenes under various lighting conditions – including Planckian lighting. They provide renderings for 8 isolated objects and 9 complex scenes. While these complex renderings are indeed more complex than the MIT dataset, they consist uniquely of textureless objects. Other scattered datasets exist in specific contexts. The MPI-Sintel dataset offers 23 rendered video sequences used for assessing optical flow methods [BWSB12]. In addition to optical flow ground truth information, this dataset also offers ground truth reflectance and depth (among other data), and has thus been used to assess intrinsic decomposition methods [TNY15]. Fig. 10 illustrates examples from all these datasets.

This report instead provides a dataset of 18 reflectance images from photorealistic scenes from the PBRT [PH10], Mitsuba [Jak10] and LuxRender [Ver07] renderers (see Fig. 5, 11 and supplemental material). The illumination component is then recovered by taking the ratio between the original image and the reflectance layer. However, on realistic images, several of our assumptions break, due to specular components, transmissive surfaces, subsurface scattering and defocus blur. As such, the illumination component may contain artifacts, such as residual reflectance. In addition, artists often simplify complex but small-scale geometries using flat textures on coarser objects. This occasionally yields overly smooth illumination layers, as light transport is not properly simulated on this geometry. They may also add high contrast materials that do not reflect the real world measured materials.

In the context of image editing applications, we further assess recent intrinsic decomposition techniques on a set of 21 real photographs (see Sec. 4.3).

4.2. Ground-truth comparisons

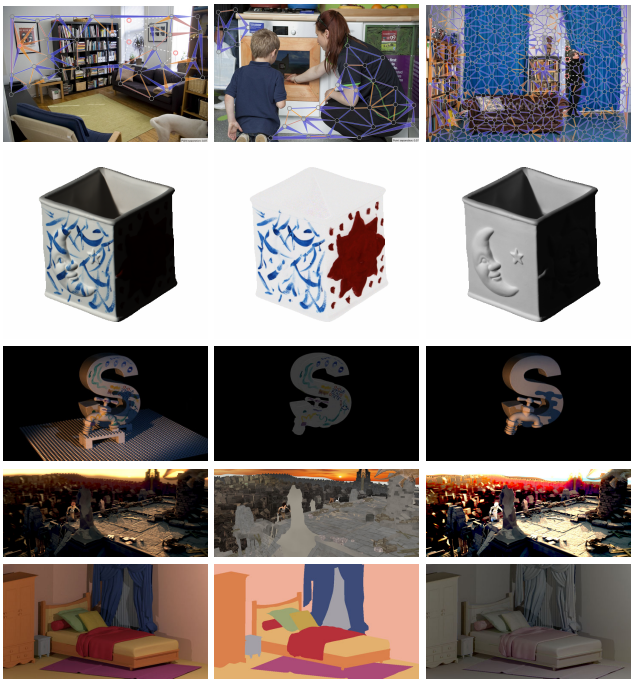
With 9 images from our new realistic ground truth dataset, we first evaluate the results of various algorithms using a classical LMSE metric. For automatic methods, we experimented with several parameter sets (up to 24, for the color Retinex of Grosse et al. [GJAF09]) and for each image, we kept the result minimizing the LMSE. For interactive methods, we manually adjusted parameters interactively and added strokes to visually obtain the best result possible. Fig. 12 plots the LMSE of tested intrinsic decomposition techniques using box-and-whisker plots, and sorts these methods by decreasing average LMSE. In terms of LMSE, the method of Shen et al. [SYJL11] performs best on this benchmark. However, as we shall see in Sec. 4.4, this does not portray an accurate picture of the state-of-the-art in intrinsic image decomposition when one focuses on image editing applications. In fact, piece-wise accurate results – such as faithful reproduction of large areas of constant reflectance – typically yield low LMSE values, while computer graphics applications are less tolerant to localized mistakes. While specularities occasionally yield residual textures in the ground truth shading, we found that this had little impact in the computed LMSE. For instance, on the "Breakfast" scene (Fig. 5), manually correcting the shading layer changes the LMSE by approximately 0.1%.

4.3. Evaluation on image-editing applications

We assess several state-of-the-art methods on image editing applications. We use a database of 21 Creative Commons photographs downloaded from Flickr, spanning different but realistic contexts, and exhibiting interesting illumination and reflectance variations. These include portraits, interior and exterior scenes. For all of these images, we have determined a specific image editing task that an artist could perform via intrinsic decomposition. These include removing a logo on a t-shirt, a tattoo or makeup, smoothing out wrinkles or freckles, or altering a shadow.

To assess multiple intrinsic decomposition methods efficiently, and to fairly evaluate methods with the same image editing operation, we automate these edits. Particular care has been taken to account for the varied dynamic ranges and the intrinsic decomposition scale

© 2017 The Author(s)
Computer Graphics Forum © 2017 The Eurographics Association and John Wiley & Sons Ltd.



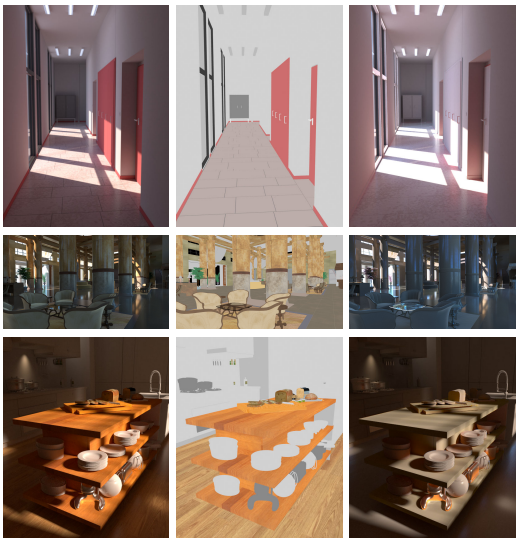
Input image Reflectance Illumination

Figure 10: *Sample images from existing intrinsic decomposition ground-truth datasets. First row. The database of Bell et al. [BBS14] provides reflectance comparison judgments between pairs of points in photographs of real-world settings (mostly indoor photographs, as illustrated by the overlayed graph. Second row. The MIT database [GJAF09] contains simple isolated objects. Their illumination is obtained by painting the object in white, and the reflectance is obtained by taking the image / illumination ratio. Third row. The MIT database has been extended to more complex scenes by Beigpour et al. [BKK15], exhibiting two objects under various illumination conditions. Fourth row. The Sintel dataset [BWSB12], originally for optical flow evaluation, provides reflectance ground truth videos from synthetic renderings. However, taking the ratio image hardly recovers a valid illumination layer. Fifth row. The database of Beigpour et al. [BSV*13] contains 9 complex but textureless synthetic scenes.*

invariance. For instance, replacing a texture should not be performed by directly editing absolute reflectance values, since they may differ from one decomposition to another. Instead, we favored gradient-domain approaches, or filtering operations on the different layers.

Logo removal. We manually determine a rough mask for a logo to be removed, and solve for the Poisson problem $\Delta u = 0$ within the masked domain and $u = \rho$ outside, with Dirichlet boundary conditions. We use the solution of this problem as the modified reflectance ρ', and reconstruct the final image as $I' = \rho' \times L$. This effectively removes the logo on successful intrinsic decomposition results (see Fig. 1).

Shadow removal. We apply the same process as for logo removal, but apply it to the illumination layer (see Fig. 16, fourth row of results, right column).



Input image Reflectance Illumination

Figure 11: *Three of our realistic ground-truth intrinsic decomposition results obtained with LuxRender [Ver07].*

Texture replacement. We inpaint a new texture in the reflectance layer by using Poisson Image Editing [PGB03]. Since complex high-frequency textures may hide artifacts in the processed result – a phenomenon called spatial frequency masking [Dal93] – we choose textures of relatively low frequency content. Further, textures may contain low-frequency illumination variations that do not correlate with the scene geometry. We hence high-pass textures containing residual low-frequency illumination. We avoid complex light-geometry interactions by integrating mostly planar objects on planar surfaces, such as carpets or paintings.

Wrinkles attenuation. Wrinkles are mostly due to shadowing effects on the skin. We manually determine a rough mask for the skin area to be corrected. We blur both the mask and the illumination layer via Gaussian filtering. We linearly interpolate the input illumination with the altered illumination based on the blurred mask, and obtain the final illumination layer L'. We reconstruct the final image as $I' = \rho \times L'$.

We provide the code and data in supplemental material for benchmarking purposes.

4.4. Evaluation

With our set of automatically generated image-processed results for various intrinsic decomposition methods, this section evaluates their success. We deem a method successful if both of the following criteria are met:

- **The effect has been achieved.** That is, if the goal is to remove a logo, the final image should not contain the logo anymore. Indeed, it is easy for a method to be free of any visible artifact, but to miss its primary purpose (e.g., a luminance-chrominance decomposition).
- **The result is realistic.** That is, images do not contain artifacts,

© 2017 The Author(s)
Computer Graphics Forum © 2017 The Eurographics Association and John Wiley & Sons Ltd.

and given a processed and unprocessed images, one cannot determine which one is processed. The method should thus not deteriorate the quality of the final result. Note that this criterion also depends on the realism of the image editing process. We have hence put significant research effort in minimizing artifacts that are inherently due to the automatic image editing process.

Materials. We decompose our dataset of 21 images with the methods of [GJAF09, BPD09, GRK*11, SYJL11, GMLMG12, ZTD*12, BST*14, BBS14, ZKE15, BM15, TNY15], using the author implementations. For the method of Grosse et al. [GJAF09], we evaluate both the grayscale and color Retinex approaches, and keep the best performing result between an L^1 and L^2 gradient reconstruction. For the method of Bonneel et al. [BST*14], we evaluate both the automatic and user-assisted approaches. For each algorithm, we downsample the input images to obtain reasonable computation time, memory usage and robustness, to the largest size the algorithm could handle. We experimented with multiple parameter sets, and kept the best result for each image. We additionally compute baseline decompositions as: 1) "Baseline reflectance" where the decomposed reflectance image is the chromaticity image, i.e. for an input pixel with RGB values (r,g,b), the output reflectance is $\left(\frac{r}{r+g+b}, \frac{g}{r+g+b}, \frac{b}{r+g+b}\right)$. 2) "Baseline illumination" where the decomposed illumination image is constant 1. 3) "Baseline sqrt" where the decomposed illumination image is the square root of the grayscale image, i.e. for an input pixel with RGB values (r,g,b), the output grayscale illumination is $\sqrt{\frac{r+g+b}{3}}$. Our supplemental materials contain the downsampled input for each algorithm, as well as their intrinsic decomposition using the best parameter set.

Results. We note that regarding reflectance editing, most methods fail at completely removing textures from the illumination layer. This results in visible artifacts when removing a logo from a t-shirt or inpainting an object in a photograph (see Fig. 14 and 15). The method of Barron and Malik [BM15] succeeds on few, but difficult, examples (see Fig. 1), and the color Retinex of Grosse et al. [GJAF09] works better on average but very rarely removes textures completely. The user-assisted approach of Bonneel et al. [BST*14] sometimes succeeds but, conversely, tends to leave too much illumination in the reflectance layer. No method succeeds in removing a tattoo better than the best baseline decomposition, but we note that most methods have difficulties dealing with dark gray or black pigments. In fact, when removing textures, a simple gradient-domain inpainting of the input image often produces better results than intrinsic decomposition methods, as no residual texture pollutes the edited image and some illumination information is propagated from the mask boundaries.

The user-assisted method of Bousseau et al. [BPD09] is the only one to succeed in completely removing a strong cast shadow, even on simple geometries. This method handles colored illumination layers, which partly explains this success in addition to user cues. Regarding wrinkles removal, most approaches work reasonably well, but the baseline also succeeds in this case. The method of Narihira et al. [TNY15] does not produce a decomposition such that the product of the reflectance and illumination layers yields the input image, which causes significant artifacts and results in image edits that are worse than the baseline in all cases.

Our supplemental materials provides a combined view of all results. Fig. 16 shows all image edits obtained by the best performing user-assisted and automatic methods.

4.5. Other considerations

As priors are introduced, as well as heavier optimization routines, the speed of intrinsic decomposition techniques rarely meet realtime constraints. In fact, most methods require minutes and even sometimes hours to compute, even on low resolution (<1 mega-pixel) images. In most cases, we downsampled our test images to about 2 mega-pixels for this reason. However, most DSLR cameras now output photographs of tens of megapixels (e.g., Canon EOS series range from 18 to 50 mega-pixels), and even compact and phone cameras often come close to DSLRs in term of pixel resolution (e.g., Samsung Galaxy S7 Edge is 17 mega-pixels or Sony xperia z5 is 23 mega-pixels). With this amount of data, intrinsic decomposition techniques should be able to treat more than half-mega pixel images to be useful for image processing (though they could remain useful for vision applications or image understanding). Notable exception include the GPU framework of Meka et al. [MZRT16] that runs in a fraction of a second.

Fig. 13 illustrates the running time with respect to image resolution for the tested algorithms. We did not time interactive methods for which most time is spent in user interactions. In practice, aside from user interactions, the method of Bousseau et al. [BPD09] takes between 5 and 30 seconds to solve for 0.5 to 1.8 mega-pixel images, and the method of Bonneel et al. [BTS*15], initially designed for videos, approximately takes between 0.2 and 1 second for 1 to 2 mega-pixel images. We do not claim fair nor accurate times, as all the methods we have tested have been run on different machines, various implementations may differ and may not have been optimized for speed, and some methods make use of multi-threading or GPU, but our comparison gives a rough sense of computation times.

We believe speed is an important factor, as slow methods preclude interactive editing applications, fine parameter tuning, their adoption by artists, processing on large data such as frames of a video, and realtime vision applications. This could sometimes just be a matter of engineering and fine implementation tuning.

While certain intrinsic decomposition methods are more robust when working on high-dynamic range (HDR) images, and images with linear camera response, most images available on the web are not HDR images. For instance, even the widely used Flickr image search engine for photograph enthusiasts do not support images of more than 8 bit per pixel and color channel. We have thus evaluated existing methods on non-HDR images. However, HDR images are now more accessible, and intrinsic decomposition methods will likely benefit from this trend.

Finally, we have evaluated algorithms based on available implementations. We advocate for reproducible research and encourage authors to disseminate their code in addition to their research paper.

5. Challenges and future work

In the context of image editing, the quality of most intrinsic decomposition methods is not satisfactory as much reflectance is left in

© 2017 The Author(s)
Computer Graphics Forum © 2017 The Eurographics Association and John Wiley & Sons Ltd.

the illumination layer or vice versa, and in many cases their speed only make them useful for less than mega-pixel images. Intrinsic decompositions have met with limited success even when comparing to synthetic benchmarks. Such ground truth data is incredibly useful for validation and machine learning purposes. In the past, the lack of widely available realistic 3d scenes have constrained researchers to use very simple and unrealistic 3d scenes or isolated-object benchmarks. This may have biased machine learning approaches, and led us to think intrinsic decomposition is a near-solved problem. Recent datasets like IIW [BBS14] have expanded the evaluation to more real world setting using photographs, showing that more progress is required. In fact, in our experiments with 8-bit images, for both synthetic ground-truth comparison and image editing tasks, few methods perform better than a well-chosen baseline. The first challenge is to make intrinsic decompositions suitable for image editing applications, for images of reasonable size. We believe some widely used priors may be harmful for image editing, such as monochromatic illumination that is actually rarely observed in real-world scenes because of effects such as interreflection. Another difficulty with intrinsic decomposition is that it is an intermediate step as far as practical applications are concerned. This makes it difficult to rely on user-driven approaches in practice because they ask users to work on a task that is not the one in which they are interested. Ideally user intervention should not be needed or be as limited as possible, which raises its own challenges. On the other hand, fully automatic techniques are not yet accurate and/or fast enough to be used reliably on real-world high-resolution images.

Current quality metrics do not consider the application for these intrinsic decompositions. Direct comparison with a reference is interesting for many applications, such as intrinsically-guided segmentation or optical flow computation, but it may not be appropriate for image editing purpose. In fact, in this context small local errors degrade the perceived quality of the image edits much more than large low-frequency errors. As such, perceptual metrics are clearly necessary. While metrics like WHDR take a first step towards this goal by evaluating intrinsic decomposition results, more efforts are needed to focus on image editing applications.

Finally, the rise of augmented sensors – whether with additional depth information, lenslets to capture lightfields, or multi-spectral sensors – could alleviate the quality problem. While several approaches use depth information, very few approaches deal with lighfields (to our knowledge, only Garces et al. and Alperovich et al. [GEZ*16, AG16] address static lightfield images) and multi-spectral sensors (to our knowledge, only Shao and Wang [SW09]). We further argue that intrinsic videos can now reasonably be handled via per-frame image intrinsic decompositions followed by temporal regularization [BTS*15]. We believe that benefiting from multiple input modalities could be a promising direction to address our third challenge: the handling of more complex materials, such as transparent and glossy materials. To allow for better evaluation, comparisons and machine learning on complex, photo-realistic renderings with complex materials, we are further extending our ground-truth dataset with other 3d renderings. In this extended dataset, we include multiple lighting conditions, normals, depth, position, irradiance and segmentations. This dataset is currently available at http://liris.cnrs.fr/~nbonneel/intrinsicstar/ground_truth/.

Conclusions. Intrinsic image decomposition is a long standing problem of great importance to computer graphics and vision algorithms. Despite the inherent difficulty of solving this ill-posed problem, significant progress has been made through sophisticated algorithms and detailed datasets. More recently, learning from larger scale datasets is being used to solve the problem. However, when evaluated with the view of specific applications, like image-editing, additional research effort is needed to make intrinsic decomposition sufficiently accurate. We find that user input is often needed to achieve the best decompositions for image-editing applications. We also find that surprisingly simple baselines sometimes can be effective. We introduce new ground truth synthetic datasets, and advocate for the development of perceptual metrics, and more public datasets and algorithms to solve this important and challenging problem.

Acknowledgments

We thank the authors of the intrinsic decomposition methods for having shared their implementation with us and Sean Bell for sharing his evaluation framework. We also thank Julie Digne for initiating the idea of this report, and Adobe for software donations. We would like to thank our funding agencies NSF IIS 1617861, and a Google Faculty Research Award. We acknowledge the use and adaptation of LuxRender scenes from Andrew Price (Kitchen scene), Peter Sandbacka (Hotel Lobby) and Simon Wendsche (School Corridor), PBRT scenes from Jay Hardy (White Room), Guillermo M. Leal Llaguno (San Miguel), Florent Boyer (Villa), Marko Dabrovic and Mihovil Odak (Sponza), and BlendSwap user Wig42 (Modern living room). Some of these PBRT resources were compiled by Benedikt Bitterli, and are availabe in supplemental material. Mitsuba scenes from Johnathan Good (Arabic, Babylonian and Italian Cities).

References

[ABC11] ARTUSI A., BANTERLE F., CHETVERIKOV D.: A Survey of Specularity Removal Methods. *Computer Graphics Forum* (2011). 2

[AG16] ALPEROVICH A., GOLDLUECKE B.: A variational model for intrinsic light field decomposition. In *Asian Conference on Computer Vision (ACCV)* (2016). 2, 10

[AGCO06] AUJOL J.-F., GILBOA G., CHAN T., OSHER S.: Structure-texture image decomposition—modeling, algorithms, and parameter selection. *Int. J. of Comp. Vision (IJCV) 67*, 1 (2006), 111–136. 2

[BBPA15] BOYADZHIEV I., BALA K., PARIS S., ADELSON E.: Band-sifting decomposition for image-based material editing. *ACM Trans. Graph. 34*, 5 (2015), 163:1–163:16. 2

[BBPD12] BOYADZHIEV I., BALA K., PARIS S., DURAND F.: User-guided white balance for mixed lighting conditions. *ACM Trans. Graph. (SIGGRAPH Asia) 31*, 6 (2012), 200:1–200:10. 2

[BBS14] BELL S., BALA K., SNAVELY N.: Intrinsic images in the wild. *ACM Trans. Graph. (SIGGRAPH) 33*, 4 (2014), 159:1–159:12. 4, 5, 6, 7, 8, 9, 10, 14, 15, 17

[BHK*16] BEIGPOUR S., HA M. L., KUNZ S., KOLB A., BLANZ V.: Multi-view multi-illuminant intrinsic dataset. In *British Machine Vision Conference (BMVC)* (2016). 7

[BHY15] BI S., HAN X., YU Y.: An L1 image transform for edge-preserving smoothing and scene-level intrinsic decomposition. *ACM Trans. Graph. (SIGGRAPH) 34*, 4 (2015), 78:1–78:12. 2, 4, 6, 14

[BKK15] BEIGPOUR S., KOLB A., KUNZ S.: A comprehensive multi-illuminant dataset for benchmarking of intrinsic image algorithms. In *Int. Conf. on Comp. Vision (ICCV)* (2015). 7, 8

© 2017 The Author(s)
Computer Graphics Forum © 2017 The Eurographics Association and John Wiley & Sons Ltd.

[BM12] BARRON J. T., MALIK J.: Shape, albedo, and illumination from a single image of an unknown object. *IEEE Comp. Vision and Pattern Recognition (CVPR)* (2012). 6

[BM13] BARRON J. T., MALIK J.: Intrinsic scene properties from a single rgb-d image. In *IEEE Comp. Vision and Pattern Recognition (CVPR)* (2013), pp. 17–24. 2, 3

[BM15] BARRON J. T., MALIK J.: Shape, illumination, and reflectance from shading. *IEEE Trans. Pattern Anal. Mach. Intell. (TPAMI)* (2015). 1, 3, 4, 5, 6, 7, 9, 14, 16, 17

[BPD09] BOUSSEAU A., PARIS S., DURAND F.: User assisted intrinsic images. *ACM Trans. Graph. (SIGGRAPH Asia) 28*, 5 (2009). 1, 3, 4, 5, 6, 9, 14, 15, 17

[BST*14] BONNEEL N., SUNKAVALLI K., TOMPKIN J., SUN D., PARIS S., PFISTER H.: Interactive Intrinsic Video Editing. *ACM Trans. Graph. (SIGGRAPH Asia) 33*, 6 (2014). 2, 4, 5, 6, 9, 15, 17

[BSV*13] BEIGPOUR S., SERRA M., VAN DE WEIJER J., BENAVENTE R., VANRELL M., PENACCHIO O., SAMARAS D.: *Intrinsic image evaluation on synthetic complex scenes*. 2013, pp. 285–289. 7, 8

[BT78] BARROW H. G., TENENBAUM J. M.: *Recovering Intrinsic Scene Characteristics From Images*. Tech. Rep. 157, AI Center, SRI International, Apr 1978. 1, 2

[BTS*15] BONNEEL N., TOMPKIN J., SUNKAVALLI K., SUN D., PARIS S., PFISTER H.: Blind Video Temporal Consistency. *ACM Trans. Graph. (SIGGRAPH Asia) 34*, 6 (2015). 9, 10

[BUSB13] BELL S., UPCHURCH P., SNAVELY N., BALA K.: OpenSurfaces: A richly annotated catalog of surface appearance. *ACM Trans. Graph. (SIGGRAPH) 32*, 4 (2013). 5

[BvdW11] BEIGPOUR S., VAN DE WEIJER J.: Object recoloring based on intrinsic image estimation. In *International Conference on Computer Vision* (2011), pp. 327–334. 2

[BWSB12] BUTLER D. J., WULFF J., STANLEY G. B., BLACK M. J.: A naturalistic open source movie for optical flow evaluation. In *European Conf. on Comp. Vision (ECCV)* (2012), pp. 611–625. 7, 8

[CBLD11] CABRAL M., BONNEEL N., LEFEBVRE S., DRETTAKIS G.: Relighting Photographs of Tree Canopies. *IEEE Trans. on Visualization and Comp. Graphics (TVCG) 17*, 10 (2011), 1459–1474. 2

[CCFI14] CHANG J., CABEZAS R., FISHER III J. W.: Bayesian nonparametric intrinsic image decomposition. In *European Conf. on Comp. Vision (ECCV)* (2014). 3, 4, 5, 14

[CK13] CHEN Q., KOLTUN V.: A simple model for intrinsic image decomposition with depth cues. In *Int. Conf. on Comp. Vision (ICCV)* (2013). 2

[CPCB15] CHENG D., PRICE B., COHEN S., BROWN M. S.: Beyond white: Ground truth colors for color constancy correction. In *Int. Conf. on Comp. Vision (ICCV)* (December 2015). 2

[CPCN13] CORKE P., PAUL R., CHURCHILL W., NEWMAN P.: Dealing with shadows: Capturing intrinsic scene appearance for image-based outdoor localisation. In *IEEE/RSJ International Conference on Intelligent Robots and Systems* (2013), pp. 2085–2092. 4

[CRA11] CARROLL R., RAMAMOORTHI R., AGRAWALA M.: Illumination decomposition for material recoloring with consistent interreflections. *ACM Trans. Graph. (SIGGRAPH) 34*, 4 (Aug. 2011). 3, 4

[CSBC09] CHUNG Y., SHEN C., BAILEY R., CHEN S.: Intrinsic image extraction from a single image. *Journal of Information Science and Engineering 25*, 6 (2009), 1939–1953. 3

[Dal93] DALY S.: Digital images and human vision. 1993, ch. The Visible Differences Predictor: An Algorithm for the Assessment of Image Fidelity, pp. 179–206. 8

[Duc15] DUCHENE S.: *Multi view delighting and relighting*. PhD thesis, Inria / University of Nice Sophia-Antipolis, April 2015. 2

[FDL04] FINLAYSON G. D., DREW M. S., LU C.: *Intrinsic Images by Entropy Minimization*. 2004, pp. 582–595. 4, 14

[GEZ*16] GARCES E., ECHEVARRIA J. I., ZHANG W., WU H., ZHOU K., GUTIERREZ D.: Intrinsic light fields, 2016. arxiv e-print. 2, 10

[GJAF09] GROSSE R., JOHNSON M. K., ADELSON E. H., FREEMAN W. T.: Ground-truth dataset and baseline evaluations for intrinsic image algorithms. In *Int. Conf. on Comp. Vision (ICCV)* (2009), pp. 2335–2342. 1, 4, 5, 6, 7, 8, 9, 14, 16, 17

[GMLMG12] GARCES E., MUNOZ A., LOPEZ-MORENO J., GUTIERREZ D.: Intrinsic images by clustering. *Computer Graphics Forum (EGSR 2012) 31*, 4 (2012). 4, 6, 9, 14, 15, 17

[GRK*11] GEHLER P. V., ROTHER C., KIEFEL M., ZHANG L., SCHÖLKOPF B.: Recovering intrinsic images with a global sparsity prior on reflectance. In *Advances in Neural Information Processing Systems (NIPS)*. Curran Associates, Inc., 2011, pp. 765–773. 9, 14, 16, 17

[HGW15] HACHAMA M., GHANEM B., WONKA P.: Intrinsic scene decomposition from rgb-d images. In *Int. Conf. on Comp. Vision (ICCV)* (2015), pp. 810–818. 2

[HMP*08] HSU E., MERTENS T., PARIS S., AVIDAN S., DURAND F.: Light mixture estimation for spatially varying white balance. In *ACM Trans. Graph. (SIGGRAPH)* (2008), pp. 70:1–70:7. 2

[Hor74] HORN B.: Determining lightness from an image. *Computer Graphics and Image Processing 3*, 4 (Dec. 1974), 277–299. 3, 4, 5

[HS79] HORN B. K. P., SJOBERG R. W.: Calculating the reflectance map. *Appl. Opt. 18*, 11 (1979), 1770–1779. 2

[HWBS16] HAUAGGE D., WEHRWEIN S., BALA K., SNAVELY N.: Photometric ambient occlusion for intrinsic image decomposition. *IEEE Trans. Pattern Anal. Mach. Intell. (TPAMI) 38*, 4 (2016), 639–651. 5

[HWU*14] HAUAGGE D. C., WEHRWEIN S., UPCHURCH P., BALA K., SNAVELY N.: Reasoning about photo collections using models of outdoor illumination. In *British Machine Vision Conference (BMVC)* (2014). 2

[ISR12] ISAZA C., SALAS J., RADUCANU B.: Evaluation of intrinsic image algorithms to detect the shadows cast by static objects outdoors. In *PMC Sensors* (2012). 2, 6

[Jak10] JAKOB W.: Mitsuba renderer, 2010. http://www.mitsuba-renderer.org. 2, 7, 13

[JCTL14] JEON J., CHO S., TONG X., LEE S.: Intrinsic image decomposition using structure-texture separation and surface normals. In *European Conf. on Comp. Vision (ECCV)* (2014). 2

[JSW10] JIANG X., SCHOFIELD A. J., WYATT J. L.: *Correlation-Based Intrinsic Image Extraction from a Single Image*. 2010, pp. 58–71. 4, 14

[KB15] KONG N., BLACK M. J.: Intrinsic depth: Improving depth transfer with intrinsic images. In *Int. Conf. on Comp. Vision (ICCV)* (Dec. 2015), pp. 3514–3522. 6

[KGB14] KONG N., GEHLER P. V., BLACK M. J.: *Intrinsic Video*. 2014, pp. 360–375. 2, 6

[KGS05] KRAWCZYK G., GOESELE M., SEIDEL H.-P.: *Photometric Calibration of High Dynamic Range Cameras*. Research Report MPI-I-2005-4-005, Max-Planck-Institut für Informatik, April 2005. 3

[Laf12] LAFFONT P.-Y.: *Intrinsic image decomposition from multiple photographs*. PhD thesis, Inria / University of Nice Sophia-Antipolis, October 2012. 2

[LB14] LI Y., BROWN M. S.: Single image layer separation using relative smoothness. In *IEEE Comp. Vision and Pattern Recognition (CVPR)* (2014), pp. 2752–2759. 4

[LBD13] LAFFONT P., BOUSSEAU A., DRETTAKIS G.: Rich intrinsic image decomposition of outdoor scenes from multiple views. *IEEE Trans. on Visualization and Comp. Graphics (TVCG) 19*, 2 (2013), 210–224. 2

[LBP*12] LAFFONT P.-Y., BOUSSEAU A., PARIS S., DURAND F., DRETTAKIS G.: Coherent intrinsic images from photo collections. *ACM Trans. Graph. (SIGGRAPH Asia) 31* (2012). 2

[LE10] LALONDE J.-F., EFROS A. A.: *Synthesizing Environment Maps from a Single Image*. Tech. Rep. CMU-RI-TR-10-24, Robotics Institute, Carnegie Mellon University, July 2010. 2

© 2017 The Author(s)
Computer Graphics Forum © 2017 The Eurographics Association and John Wiley & Sons Ltd.

[LM71] LAND E. H., MCCANN J. J.: Lightness and retinex theory. *J. Opt. Soc. Am. 61*, 1 (Jan 1971), 1–11. 3, 4, 5

[LMGH*13] LOPEZ-MORENO J., GARCES E., HADAP S., REINHARD E., GUTIERREZ D.: Multiple light source estimation in a single image. *Computer Graphics Forum* (2013). 2

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

[LSHC15] LAFER-SOUSA R., HERMANN K. L., CONWAY B. R.: Striking individual differences in color perception uncovered by âĂŸthe dressâĂŹ photograph. *Current Biology 25*, 13 (06 2015), R545–R546. 6

[LSX09] LI Y., SHI B., XU C.: Intrinsic image decomposition using color invariant edge. *IEEE Int. Conf. on Image and Graphics* (2009), 307–312. 4

[LVVG16] LETTRY L., VANHOEY K., VAN GOOL L.: Darn: a deep adversial residual network for intrinsic image decomposition. In *arXiv:1612.07899 preprint* (2016). 6

[LWQ*08] LIU X., WAN L., QU Y., WONG T.-T., LIN S., LEUNG C.-S., HENG P.-A.: Intrinsic colorization. *ACM Trans. Graph. (SIGGRAPH Asia) 27*, 5 (2008), 152:1–152:9. 6

[LYZ15] LIU Y., YUAN Z., ZHENG N.: *Intrinsic Image Decomposition from Pair-Wise Shading Ordering*. 2015, pp. 83–98. 4

[LZL14] LI C., ZHOU K., LIN S.: Intrinsic face image decomposition with human face priors. In *European Conf. on Comp. Vision (ECCV)* (2014), pp. 218–233. 5, 6

[LZL15] LI C., ZHOU K., LIN S.: Simulating makeup through physics-based manipulation of intrinsic image layers. In *IEEE Comp. Vision and Pattern Recognition (CVPR)* (2015). 6

[LZT*12] LEE K. J., ZHAO Q., TONG X., GONG M., IZADI S., LEE S. U., TAN P., LIN S.: *Estimation of Intrinsic Image Sequences from Image+Depth Video*. 2012, pp. 327–340. 2, 5

[MLKS04] MATSUSHITA Y., LIN S., KANG S. B., SHUM H.-Y.: Estimating intrinsic images from image sequences with biased illumination. In *European Conf. on Comp. Vision (ECCV)* (2004), pp. 274–286. 2

[MZRT16] MEKA A., ZOLLHÖFER M., RICHARDT C., THEOBALT C.: Live intrinsic video. *ACM Trans. Graph. (SIGGRAPH) 35*, 4 (2016). 2, 9

[NMY15] NARIHIRA T., MAIRE M., YU S. X.: Learning lightness from human judgement on relative reflectance. In *IEEE Comp. Vision and Pattern Recognition (CVPR)* (2015), pp. 2965–2973. 5, 6

[OW04] OMER I., WERMAN M.: Color lines: image specific color representation. In *IEEE Comp. Vision and Pattern Recognition (CVPR)* (2004), vol. 2, pp. 946–953. 4

[PGB03] PÉREZ P., GANGNET M., BLAKE A.: Poisson image editing. *ACM Trans. Graph. (SIGGRAPH) 22*, 3 (2003), 313–318. 8

[PH10] PHARR M., HUMPHREYS G.: *Physically Based Rendering, Second Edition: From Theory To Implementation*, 2nd ed. Morgan Kaufmann Publishers Inc., 2010. 2, 7, 13

[RHTP16] RABIN J., HOUSER B., TALBERT C., PATEL R.: Blue-black or white-gold? early stage processing and the color of 'the dress'. *PLoS ONE 11*, 8 (08 2016), 1–10. 6

[RRF*15] REMATAS K., RITSCHEL T., FRITZ M., GAVVES E., TUYTELAARS T.: Deep reflectance maps. *CoRR abs/1511.04384* (2015). 2

[SBD15] SHELHAMER E., BARRON J. T., DARRELL T.: Scene intrinsics and depth from a single image. *Int. Conf. on Comp. Vision Workshop (ICCVW)* (2015). 6

[Ser15] SERRA M.: *Modeling, estimation and evaluation of intrinsic images considering color information*. PhD thesis, Universitat Autònoma de Barcelona - Computer Vision Center, September 2015. 7

[SF15] STEFANI N., FUSIELLO A.: Recovering Intrinsic Images by Minimizing Image Complexity. In *Smart Tools and Apps for Graphics - Eurographics Italian Chapter Conference* (2015), The Eurographics Association. 4

[SL16] SON H., LEE S.: Intrinsic image decomposition using deep convolutional network. In *SUNw: Scene Understanding Workshop (Poster)* (2016). 6

[SSN16] SAINI S., SAKURIKAR P., NARAYANAN P. J.: Intrinsic image decomposition using focal stacks. In *Indian Conf. on Comp. Vision, Graph. and Image Proc.* (2016), ICVGIP '16, pp. 88:1–88:8. 2

[STL08] SHEN L., TAN P., LIN S.: Intrinsic image decomposition with non-local texture cues. In *IEEE Comp. Vision and Pattern Recognition (CVPR)* (2008), pp. 1–7. 4, 14

[SW09] SHAO M., WANG Y.-H.: Extracting intrinsic images from multispectral. In *International Conference on Wavelet Analysis and Pattern Recognition* (2009), pp. 241–246. 2, 10

[SY11] SHEN L., YEO C.: Intrinsic images decomposition using a local and global sparse representation of reflectance. In *IEEE Comp. Vision and Pattern Recognition (CVPR)* (2011), pp. 697–704. 4, 14

[SYC*14] SHEN J., YAN X., CHEN L., SUN H., LI X.: Re-texturing by intrinsic video. *Information Sciences 281* (2014), 726 – 735. Multimedia Modeling. 2

[SYJL11] SHEN J., YANG X., JIA Y., LI X.: Intrinsic images using optimization. In *IEEE Comp. Vision and Pattern Recognition (CVPR)* (2011), pp. 3481–3487. 4, 5, 7, 9, 14, 16, 17

[TAF06] TAPPEN M. F., ADELSON E. H., FREEMAN W. T.: Estimating intrinsic component images using non-linear regression. In *IEEE Comp. Vision and Pattern Recognition (CVPR)* (2006), vol. 2, pp. 1992–1999. 5, 7, 14

[TFA05] TAPPEN M. F., FREEMAN W. T., ADELSON E. H.: Recovering intrinsic images from a single image. *IEEE Trans. Pattern Anal. Mach. Intell. (TPAMI) 27*, 9 (2005), 1459–1472. 4, 5, 6, 14

[TNY15] TAKUYA NARIHIRA M. M., YU S. X.: Direct intrinsics: Learning albedo-shading decomposition by convolutional regression. In *Int. Conf. on Comp. Vision (ICCV)* (2015). 3, 6, 7, 9, 14, 15, 17

[TSH12] TANG Y., SALAKHUTDINOV R., HINTON G.: Deep Lambertian Networks. In *International Conference on Machine Learning* (2012). 6

[Ver07] VERGAUWEN T.: Luxrender - gpl physically based renderer, 2007. http://www.luxrender.net/. 2, 7, 8

[Vit15] VITELLI M.: Intrinsic image decomposition using deep convolutional networks. In *unpublished* (2015). 6

[Wei01] WEISS Y.: Deriving intrinsic images from image sequences. In *Int. Conf. on Comp. Vision (ICCV)* (2001), vol. 2, pp. 68–75 vol.2. 2

[XLL*16] XIE D., LIU S., LIN K., ZHU S., ZENG B.: Intrinsic decomposition for stereoscopic images. In *Int. Conf. on Image Processing. (ICIP)* (2016). 2

[YGL*14] YE G., GARCES E., LIU Y., DAI Q., GUTIERREZ D.: Intrinsic video and applications. *ACM Trans. Graph. (SIGGRAPH) 33*, 4 (2014), 80:1–80:11. 2, 6

[Yu16] YU J.: Rank-constrained pca for intrinsic images decomposition. In *Int. Conf. on Image Processing. (ICIP)* (2016), pp. 3578–3582. 2

[ZIKF15] ZORAN D., ISOLA P., KRISHNAN D., FREEMAN W. T.: Learning ordinal relationships for mid-level vision. In *Int. Conf. on Comp. Vision (ICCV)* (2015). 4, 5, 6, 14

[ZKE15] ZHOU T., KRÄHENBÜHL P., EFROS A. A.: Learning data-driven reflectance priors for intrinsic image decomposition. *CoRR abs/1510.02413* (2015). 4, 5, 6, 9, 14, 16, 17

[ZTD*12] ZHAO Q., TAN P., DAI Q., SHEN L., WU E., LIN S.: A closed-form solution to retinex with nonlocal texture constraints. *IEEE Trans. Pattern Anal. Mach. Intell. (TPAMI) 34*, 7 (2012), 1437–1444. 4, 9, 14, 16, 17

© 2017 The Author(s)
Computer Graphics Forum © 2017 The Eurographics Association and John Wiley & Sons Ltd.

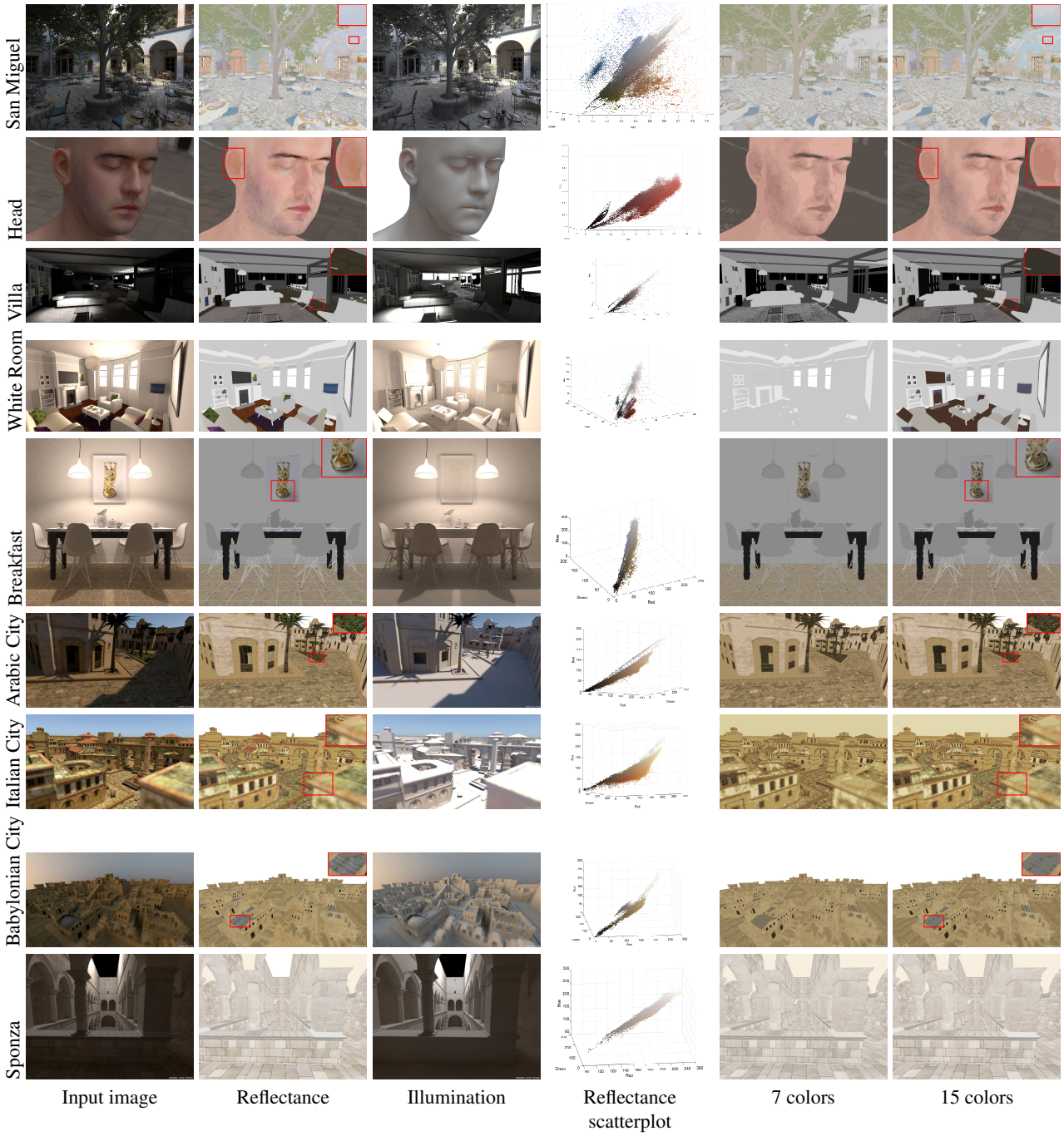


Figure 5: *We evaluate the prior of sparse reflectance values and piecewise reflectance flatness on several realistic synthetic renderings used in PBRT [PH10] and Mitsuba [Jak10]. A 3d RGB scatterplot of reflectance values hardly exhibits clusters, while quantizing reflectance values into up to 15 color clusters still shows some artifacts on complex scenes (no dithering was applied – see insets). The reflectance remains mostly flat for man-made scenes, but fails on the head model. The illumination was computed as the ratio between the input and reflectance images, and may reflect inaccuracies on glossy or refractive objects, or due to subsurface scattering.*

© 2017 The Author(s)
Computer Graphics Forum © 2017 The Eurographics Association and John Wiley & Sons Ltd.

	[FDL04]	[TFA05]	[TAF06]	[STL08]	[BPD09]	[GJAF09]	[JSW10]	[SY11]	[SYJL11]	[ZTD*12]	[GMLMG12]	[BBS14]	[CCFI14]	[ZKE15]	[BM15]	[BHY15]	[TNY15]	[ZIKF15]
MI		?	?	×		×	×	×	×	×	×	×	∼	×		×		×
R				∼		×		∼	∼	∼	∼	∼		∼	∼	∼		×
EoI		×					∼											
CR		∼						∼		∼	∼	∼		∼				
LRR					∼													
SRV								∼			∼	∼	∼	∼	∼	∼		
RML												∼		∼	∼			
MV							×											
PL	×																	
NLC				∼						∼								
UC					∼				∼									
DD		×	×									∼	∼	∼	∼		×	×

Figure 8: *This table illustrates the use of different priors and constraints (whose acronyms are described in Sec. 2.2) as they were introduced chronologically. × represents strict constraints which cannot be violated, while ∼ is a prior or soft constraint. The methods of Tappen et al. [TFA05, TAF06] use a monochromatic illumination constraint, but comparisons shown in two papers exhibit colored illumination [BPD09, GMLMG12]. We expect the use of many priors to improve decompositions but on more limited datasets.*

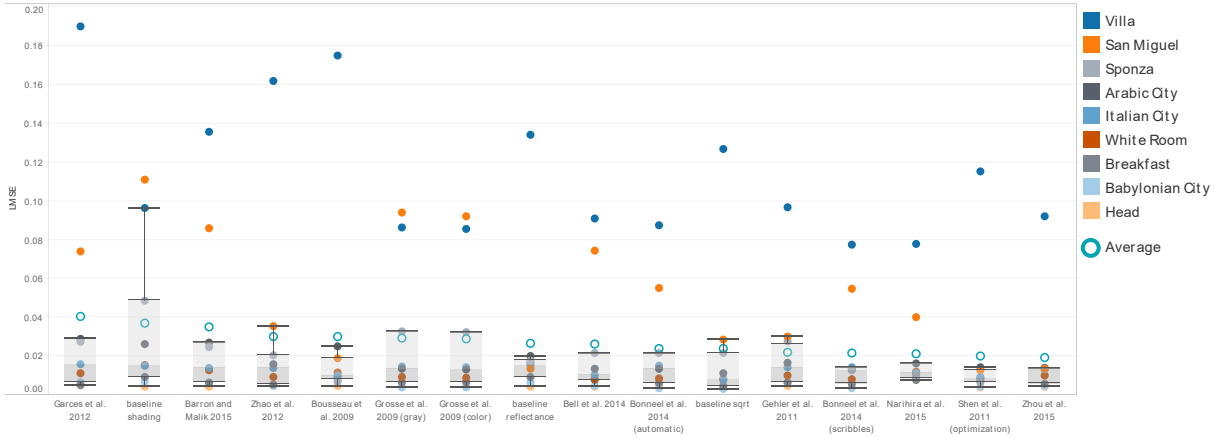


Figure 12: *We compute the LMSE accuracy of various algorithms on our realistic synthetic dataset. These methods are sorted by decreasing average LMSE. The "Villa" And "San Miguel" scenes have consistently lower accuracy (i.e., higher LMSE), while the "Head" and "Babylonian City" scenes have higher accuracy. However, LMSE does not reflect usefulness for computer graphics applications.*

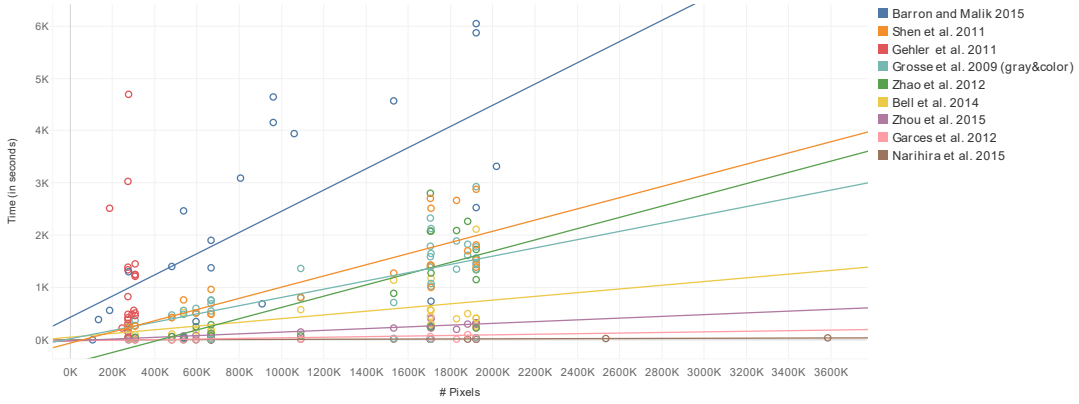


Figure 13: *Computation time for tested automatic intrinsic decomposition methods with respect to image resolution. No trend line is shown for Gehler et al. [GRK*11] since images were resized to roughly the same resolution. For clarity, we merged results of gray and color Retinex [GJAF09] ; however they show a bimodal timing distribution: this is rather due to L1 and L2 reconstructions.*

© 2017 The Author(s)
Computer Graphics Forum © 2017 The Eurographics Association and John Wiley & Sons Ltd.

Figure 14: *Decomposition and application results for various methods. Top to bottom: reflectance, illumination, image-edited result. We add a carpet to the first image, and remove the t-shirt's logo on the second and third images. Additional results can be seen in supplemental materials. The baseline consists of the best of three naive approaches.*

© 2017 The Author(s)
Computer Graphics Forum © 2017 The Eurographics Association and John Wiley & Sons Ltd.

Figure 15: *Decomposition and application results for various methods. Top to bottom: reflectance, illumination, image-edited result. We add a carpet to the first image, and remove the t-shirt's logo on the second and third images. Additional results can be seen in supplemental materials.*

© 2017 The Author(s)
Computer Graphics Forum © 2017 The Eurographics Association and John Wiley & Sons Ltd.

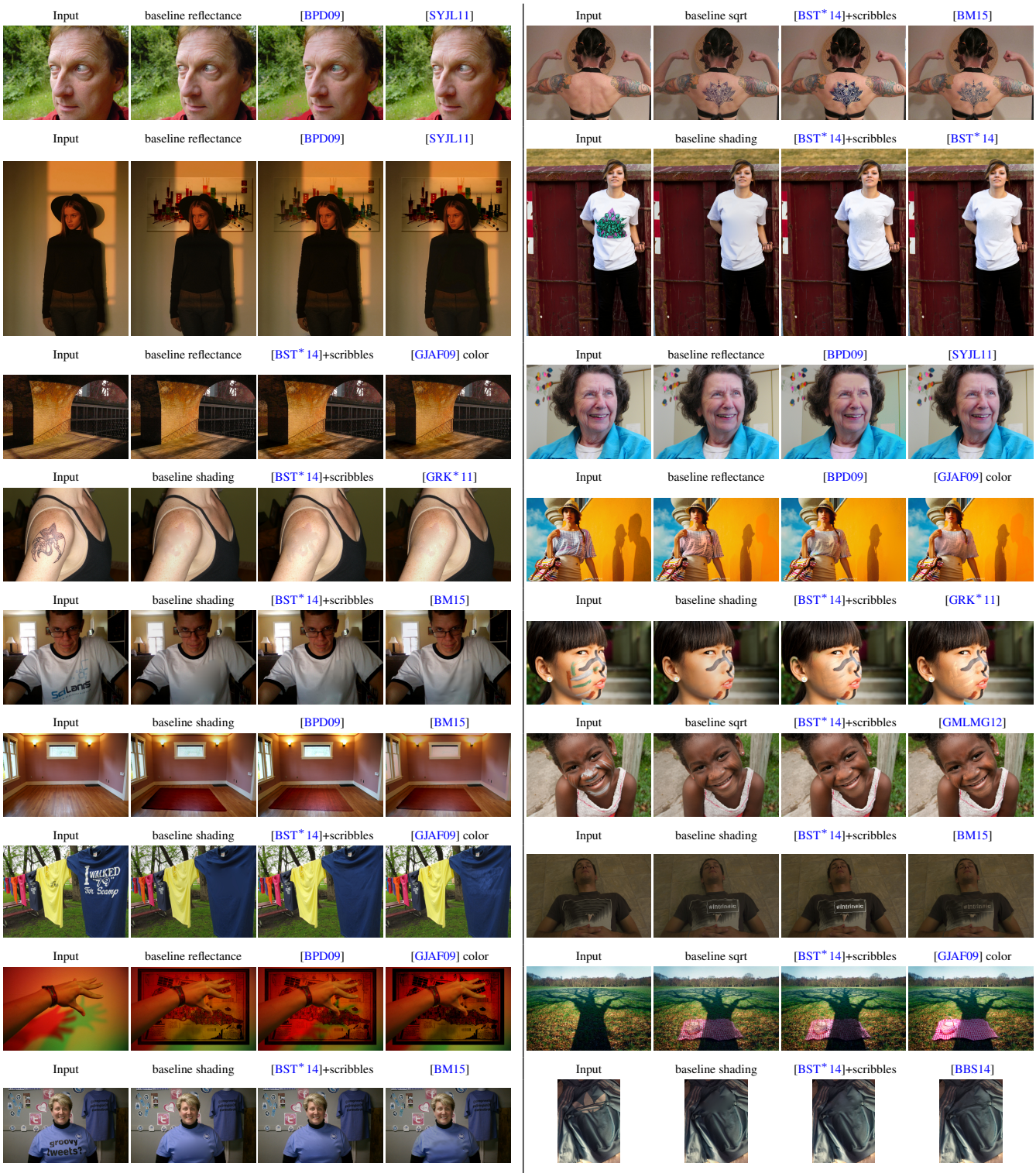


Figure 16: *We illustrate all image editing results achieved using the best user-assisted and automatic intrinsic decomposition methods compared to the best baseline. User-assisted methods include [BPD09, BST* 14], and automatic methods include [GJAF09, SYJL11, GRK* 11, GMLMG12, ZTD* 12, BBS14, BST* 14, TNY15, BM15, ZKE15]. In some cases, many methods perform similarly well.*

© 2017 The Author(s)
Computer Graphics Forum © 2017 The Eurographics Association and John Wiley & Sons Ltd.