

User-Assisted Intrinsic Images

Adrien Bousseau^{1,2}

Sylvain Paris³

Frédo Durand²

¹INRIA / Grenoble University

²MIT CSAIL

³Adobe Systems, Inc.

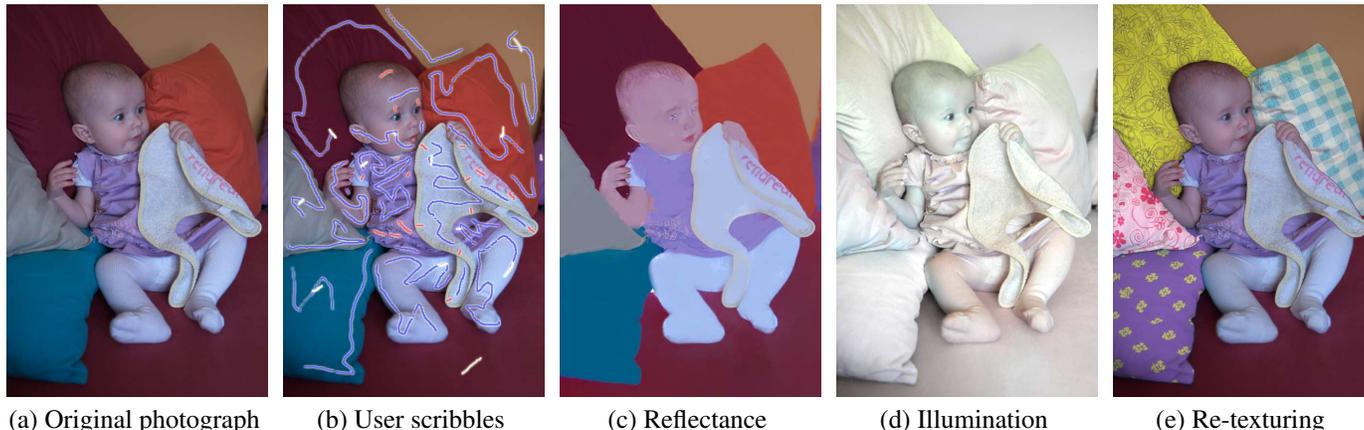


Figure 1: Our system relies on user indications, shown in (b), to extract from a single photograph its reflectance and illumination components (c-d). In (b), white scribbles indicate fully-lit pixels, blue scribbles correspond to pixels sharing a similar reflectance and red scribbles correspond to pixels sharing a similar illumination. This decomposition facilitates advanced image editing such as re-texturing (e).

Abstract

For many computational photography applications, the lighting and materials in the scene are critical pieces of information. We seek to obtain *intrinsic images*, which decompose a photo into the product of an *illumination* component that represents lighting effects and a *reflectance* component that is the color of the observed material. This is an under-constrained problem and automatic methods are challenged by complex natural images. We describe a new approach that enables users to guide an optimization with simple indications such as regions of constant reflectance or illumination. Based on a simple assumption on local reflectance distributions, we derive a new propagation energy that enables a closed form solution using linear least-squares. We achieve fast performance by introducing a novel downsampling that preserves local color distributions. We demonstrate intrinsic image decomposition on a variety of images and show applications.

Keywords: computational photography, intrinsic images, reflectance-illumination separation

1 Introduction

Many computational photography tasks such as relighting, material alteration or re-texturing require knowledge of the lighting and

materials in the scene. Unfortunately, in a photograph, illumination and reflectance are conflated through complex interaction and the separation of those components, called *intrinsic images* [Barrow and Tenenbaum 1978], has long been an open challenge. A pixel can be seen as the per-color-channel product of an *illumination* component, also called *shading*, and a *reflectance* component, also called *albedo*. Given a single image, the problem is severely ill-posed: a dark-yellow pixel can come from, e.g., a white material illuminated by a dark yellow light, or from a dark-yellow material illuminated by a bright white light.

In this paper, we introduce a new image decomposition technique that relies on sparse constraints provided by the user to disambiguate illumination from reflectance. Figure 1(b) illustrates these indications that can correspond to pixels of similar reflectance, similar illumination, or known illumination. Central to our technique is a new propagation method that estimates illumination from local constraints based on a few assumptions on image formation and reflectance distributions. In particular, we reduce the number of unknowns by assuming that local reflectance variations lie in 2D subspaces of the RGB color space. Although this simplification requires color images and cannot handle cases such as a black-and-white checkerboard texture, we show that it can handle a broad class of images with complex lighting. In order to enable fast and accurate solution, we also introduce a novel downsampling strategy that better preserves the local color distributions of an image and enables rapid multigrid computation. Finally, we illustrate the benefit of our image decomposition with image manipulations including reflectance editing and re-lighting.

To summarize, this paper makes the following contributions:

- we describe a user-assisted method to perform intrinsic image decomposition from a single image.
- we introduce a novel energy formulation that propagates user edits based on the local reflectance color distribution.
- we propose a new downsampling scheme that enforces the preservation of color distributions at reduced image resolutions.

1.1 Related Work

The decoupling of reflectance from illumination was introduced by Barrow and Tenenbaum [1978] as *intrinsic images*. The *reflectance* describes how an object reflects light and is also often called *albedo*. The *illumination* corresponds to the amount of light incident at a point (essentially irradiance). Although it is often referred to as *shading*, it includes effects such as shadows and indirect lighting.

Physically-based inverse rendering, such as Yu and Malik [1998], seeks to invert the image formation model in order to recover the lighting conditions of a scene, but requires known geometry.

Using several images of the same scene under different illuminations, Weiss [2001] proposes a method to estimate a reflectance image along with the illumination map of each input image. This approach has then been extended by Liu et al. [2008] to non-registered image collections in order to colorize black-and-white photographs. The use of varying illumination has also been applied to shadow removal in a flash/no-flash setup by Agrawal et al. [2006].

Due to its inherent ill-posedness, decomposition of intrinsic images from a single image cannot be solved without prior knowledge on reflectance and illumination. Based on the Retinex theory [Land and McCann 1971], Horn [1986] assumes that reflectance is piecewise constant while illumination is smooth. This heuristic allows for the recovery of a reflectance image by thresholding the small image gradients, assumed to correspond to illumination. Sinha and Adelson [1993] discriminate illumination from reflectance edges based on their junctions in a world of painted polyhedra: T junctions are interpreted as reflectance variations, while arrow and Y junctions correspond to illumination. Tappen et al. [2005] rely on a classifier trained on image derivatives to classify reflectance and illumination gradients. Despite these heuristics and classifiers, many configurations of reflectance and illumination encountered in natural images remain hard to disambiguate. Shen et al. [2008] propose to enrich these local approaches with non-local texture constraints. Starting from a Retinex algorithm, their texture constraints ensure that pixels that share similar texture will have the same reflectance.

A large body of work has been proposed for the specific problem of shadow removal, either automatically [Finlayson et al. 2002; Finlayson et al. 2004] or based on user interaction [Mohan et al. 2007; Wu et al. 2007; Shor and Lischinski 2008]. The common idea of these methods is to identify shadow pixels, either via boundary detection or region segmentation. Once shadows are detected, they can be removed by color correction or gradient domain filtering. These methods focus on cast shadow removal with clear boundaries, while we also target the removal of smooth shading where the boundary between lit and shaded regions cannot be delimited. Note that although the approach of Finlayson et al. [2002; 2004] relies on the estimation of an illumination-free image, this image is grayscale and does not represent the true reflectance.

Intrinsic images decomposition is also related to other image decompositions. Matting algorithms [Chuang et al. 2001; Levin et al. 2008] aim to separate the foreground and background layers of an image along with their opacity based on user indications. User-assisted approaches have been proposed to separate reflections from a single image [Levin and Weiss 2007]. Automatic decompositions have been introduced to restore pictures corrupted by haze [Fattal 2008], and to perform white balance of photographs taken under mixed lighting [Hsu et al. 2008]. Although all these methods do not directly target the extraction of illumination from a single image, our energy formulation is inspired by the *matting Laplacian* used in the work of Levin et al. [2007] and Hsu et al. [2008]. We rely on a similar assumption that, in natural images, material colors lie on subspaces of the RGB space [Omer and Werman 2004].

1.2 Overview

Our decoupling of illumination and reflectance is based on user-provided constraints and a new propagation model. The user can use sparse strokes to mark parts that share the same reflectance or where illumination does not vary, in particular at material boundaries. The user also needs to provide at least one fixed illumination value to solve for global scale ambiguity.

Based on the assumption that reflectance values are low-rank in local windows, we derive a closed-form equation that depends only on illumination. Adding the user constraints, we can solve a linear least-square system that provides the illumination. Reflectance is simply inferred by a division.

Our least-square solution works best with local windows of medium sizes, which raises computational challenges. We have found that standard multigrid approaches yield unsatisfactory results in our case, because traditional downsampling operators do not respect local color distributions. We introduce a new downsampling scheme for techniques like ours that rely on local color distributions. We show that it enables dramatic speedup and achieves high accuracy.

2 Reflectance-Illumination Decomposition

We first detail our assumptions about the observed image, in particular the fact that reflectance colors are locally planar. We show how it leads to a quadratic energy where illumination is the only unknown, and where user constraints can easily be included.

Image Formation Model As commonly done, we assume that the interaction between light and objects can be described using RGB channels alone. We focus on Lambertian objects and a single light color, although we show later that these hypotheses can often be alleviated in practice to handle colored light.

With this model, the observed color at a pixel is:

$$\mathbf{I} = s \mathbf{L} * \mathbf{R} \quad (1)$$

where s is the *illumination*, a non-negative scalar modeling the incident light attenuation due to factors such as light travel, occlusion and foreshortening, \mathbf{L} is the RGB color of the light, $*$ denotes per-channel multiplication, and \mathbf{R} is the material RGB *reflectance* that describes how objects reflect light. For clarity, we assume that the light is white, (or equivalently that the input image is white balanced). This means $\mathbf{L} = (1, 1, 1)^T$ and Equation 1 becomes:

$$\mathbf{I} = s \mathbf{R} \quad (2)$$

2.1 Low-Rank Structure of Local Reflectance

Our objective is to retrieve the illumination s and the RGB components of \mathbf{R} at each pixel. The difficulty is that Equation 2 provides only three equations, one per RGB channel. In addition, there is a scale ambiguity between illumination and reflectance, that is, if \mathbf{R}_0 and s_0 are a valid decomposition, then $k\mathbf{R}_0$ and s_0/k are also valid for any scalar factor $k > 0$.

We overcome the ill-posedness of the problem by a local low-rank assumption on reflectance colors. We are inspired by a variety of recent studies that show structure and sparsity in the distribution of colors in natural scenes. In particular, Omer and Werman [2004] show that the set of reflectance colors is sparse, which led to practical matting [Levin et al. 2008] and white balance [Hsu et al. 2008] techniques.