Improvements on Multi-View Intrinsic Images of Outdoors Scenes

Sebastien Bonopera Adrien Bousseau George Drettakis

Inria

This document describes an improved implementation of our multi-view intrinsic image algorithm [Duchêne et al. 2015]. This algorithm computes an intrinsic decomposition of multiple images of an outdoor scene captured under the same lighting condition. **This is work in progress, please check this document regularly for potential updates.**

1 Problem formulation

1.1 Image formation model

The intrinsic images model assumes diffuse surfaces and expresses the image values $\mathbf{I}$ at each pixel as the product between the incident illumination $\mathbf{S}$ and the object reflectance $\mathbf{R}$. Formally, the radiance towards the camera at each non-emissive, visible point corresponding to a pixel is given by the equation

$$\begin{aligned} \mathbf{I} &= \mathbf{R} * \int_{\Omega} \cos\theta_\omega \mathbf{L}(\omega) d\omega \\ \mathbf{I} &= \mathbf{R} * \mathbf{S} \end{aligned}$$

where lighting is integrated over the hemisphere Ω centered on the normal at the visible point. $\mathbf{L}(\omega)$ is the incoming radiance in direction ω, θ_ω is the angle between the normal at the visible point and direction ω. Capital bold letters represent RGB color values and $*$ denotes per-channel multiplication.

We can further separate out the incoming radiance into three components: the radiance due to the sun, that due to the sky and that due to indirect lighting. To simplify notation, we define two subsets of the hemisphere: Ω_{sky}, i.e., the subset of directions in which the visible point sees the sky, and Ω_{ind} the subset of directions in which another object is visible, and thus contributes to indirect lighting. We model the sun as a directional light source subject to the visibility term v_{sun} to obtain

$$\begin{aligned} \mathbf{I} &= \mathbf{R} * \Big(v_{sun} \cos\theta_{sun} \mathbf{L}_{sun} + \int_{\Omega_{sky}} \cos\theta_\omega \mathbf{L}_{sky}(\omega) d\omega \\ &\qquad + \int_{\Omega_{ind}} \cos\theta_\omega \mathbf{L}_{ind}(\omega) d\omega \Big) \\ \mathbf{I} &= \mathbf{R} * \Big(v_{sun}\mathbf{S}_{sun} + \mathbf{S}_{sky} + \mathbf{S}_{ind} \Big) \end{aligned} \tag{1}$$

where $\mathbf{R}$ is the object RGB reflectance, $\mathbf{S}_{sun}$, $\mathbf{S}_{sky}$ and $\mathbf{S}_{ind}$ are the RGB incident illumination (or shading) from the sun, sky and indirect lighting respectively, v_{sun} indicates points visible from the sun and as such captures shadows [Laffont et al. 2013].

1.2 Joint recovery of reflectance and shadows

Given the multiple images of the scene, the first step of our method is to reconstruct a 3D model of the scene using multi-view stereo algorithms [Snavely et al. 2006; Furukawa and Ponce 2010], which we call *proxy*. The resulting sparse 3D reconstruction only provides an imprecise and incomplete representation of the scene. Nevertheless, this reconstruction is sufficient to compute plausible sky and indirect illumination at each reconstructed 3D point. The coarse proxy is however unreliable for sun illumination because it typically contains high-frequency features due to cast shadows.

Our key observation is that sun visibility (i.e. shadows) can be estimated jointly with reflectance once all other unknowns have been computed. Assuming known $\mathbf{S}_{sun}$, $\mathbf{S}_{sky}$ and $\mathbf{S}_{ind}$, we rewrite Equation 1 to express reflectance as a function of sun visibility:

$$\mathbf{R}(v_{sun}) = \frac{\mathbf{I}}{v_{sun}\mathbf{S}_{sun} + \mathbf{S}_{sky} + \mathbf{S}_{ind}}. \tag{2}$$

With this formulation, each point now has multiple candidate reflectance values depending on the value of v_{sun}. We can further simplify our problem by assuming binary visibility, i.e. $v_{sun} \in \{0, 1\}$, which yields two candidate reflectances per point. A guiding principle of our approach is to select the value of v_{sun} that favors a small number of reflectances in the scene, a heuristic also used in prior work on intrinsic image decomposition and segmentation [Omer and Werman 2004; Barron and Malik 2013; Bell et al. 2014].

1.3 Progressive recovery of all unknowns

Equation 2 requires knowledge of all shading terms $\mathbf{S}_{sun}$, $\mathbf{S}_{sky}$ and $\mathbf{S}_{ind}$. While $\mathbf{S}_{ind}$ and $\mathbf{S}_{sky}$ can be computed directly from the 3D reconstruction and an estimate of the sky, $\mathbf{S}_{sun}$ requires a known sun intensity $\mathbf{L}_{sun}$, which we don't have yet. However, we can estimate this quantity using Equation 1 and a pair of points with same reflectance and different visibility. Given two points p_1 and p_2 with the same reflectance, with one in shadow and the other in light, we can compute

$$\mathbf{L}_{sun} = \frac{\mathbf{I}_1 * (\mathbf{S}_{sky2} + \mathbf{S}_{ind2}) - \mathbf{I}_2 * (\mathbf{S}_{sky1} + \mathbf{S}_{ind1})}{\mathbf{I}_2 * v_{sun1} * \cos(\theta_{sun1}) - \mathbf{I}_1 * v_{\mathrm{sun}2} * \cos(\theta_{sun2})}. \tag{3}$$

We thus face a chicken-and-egg problem: we need an estimate of v_{sun} to select pairs of points (p_1, p_2) to compute $\mathbf{L}_{sun}$ from Equation 3, but we need a known $\mathbf{L}_{sun}$ to select v_{sun} using Equation 2. Our strategy to cope with this problem will be to first compute a coarse estimate of v_{sun} using complementary cues from the image intensity and the approximate 3D reconstruction. This coarse estimate is sufficient to compute $\mathbf{L}_{sun}$ from many pairs of points in light and in shadow. We then use the now known $\mathbf{L}_{sun}$ to refine our shadow estimate based on Equation 2.

While our original implementation described in [Duchêne et al. 2015] employs two very different shadow classifiers to compute the coarse and fine estimates of v_{sun}, the main novelty of the implementation described in this document is to express the two classifiers within a common Markov Random Fields (MRF) formulation. We implement this formulation over a graph of super-pixels which are created using mean-shift clustering [Christoudias et al. 2002] each super-pixel corresponding to a small region of pixels with similar colors.

2 Coarse shadow classifier

Our goal is to assign a sun visibility label $x_s \in \{0 = shadow, 1 = light\}$ to each super-pixel s of each image. Denoting $\mathcal{X}$ a labeling that assigns a label to all super-pixels, we seak to recover the label configuration that minimizes

$$\underset{\mathcal{X}}{\operatorname{argmin}} \sum_{s \in V} \phi_s(x_s) \; + \sum_{(s,t) \in \mathcal{E}} \phi_{s,t}(x_s, x_t), \;\; x_i \in \{0, 1\}, \quad (4)$$

where V denotes the set of nodes of the graph (i.e. the super-pixels), $\mathcal{E}$ is the set of edges that connect nodes, $\phi_s(x_s)$ is the unary term that measures the cost of label x_s given the appearance of super-pixel s, and $\phi_{s,t}(x_s, x_t)$ is the pairwise term that measures the agreement between the couple of labels (x_s, x_t) of two connected super-pixels (s, t). We now detail the definition of these energy terms and the construction of the connections $\mathcal{E}$.

2.1 Unary term

The goal of the unary term is to predict the shadow label x_s of a superpixel s solely from its local information. In practice we combine two local cues, one based on the reconstructed 3D proxy and the other one based on the intensity of the image.

We first compute the shadow image v_{sun}^{proxy} casted by the reconstructed proxy. Denoting $P\big(v_{sun}^{proxy}(s)\big)$ the ratio of pixels within super-pixel s that are covered by this shadow, we define the cost

$$\begin{aligned} \phi_s^{proxy}(x_s = 0) &= 1 - P\big(v_{sun}^{proxy}(s)\big) \\ \phi_s^{proxy}(x_s = 1) &= P\big(v_{sun}^{proxy}(s)\big). \end{aligned}$$

In other words, we penalize label 0 if the super-pixel is not covered by the shadow of the reconstructed proxy, and vice-versa.

Our second cue is based on the "bright channel" image, i.e. the graylevel image obtained by taking for each pixel the value of its brightest channel [Panagopoulos et al. 2013]. Given the shadow image computed from the reconstructed proxy, we build two brightness histograms, one for the pixels covered by the shadow and one for the pixels in light. We then fit Gaussian mixture models [Prince 2012] on these two histograms to obtain the probability distributions $P(b|light)$ and $P(b|shadow)$ of brightness level b to appear in light (resp. in shadow). In practice we use three Gaussian functions for each mixture model. Denoting $P(shadow)$ the ratio of pixels covered by the shadow of the proxy, and $P(light)$ its complement, we apply Bayes' rule to derive the following cost of a pixel to be in shadow or light given its brightness

$$\begin{aligned} P(shadow|b) &\propto P(b|shadow)P(shadow) \\ P(light|b) &\propto P(b|light)P(light) \\ \phi_s^{brightness}(x_s = 0) &= 1 - P(shadow|b) \\ \phi_s^{brightness}(x_s = 1) &= 1 - P(light|b). \end{aligned}$$

2.2 Pairwise term

The goal of the pairwise term at this stage is to propagate labels from confident super-pixels to ambiguous ones, building on the assumption that neighboring super-pixels with similar colors should have similar labels [Boykov and Jolly 2001]. We encourage color-aware smoothness by penalizing configurations that assign different labels to neighboring super-pixels that have similar color distributions, as measured by the χ^2 distance between their color histograms. We define the following cost that decreases with distance

$$\begin{aligned} \phi_{s,t}^{smoothness}(x_s = x_t) &= 0 \\ \phi_{s,t}^{smoothness}(x_s \neq x_t) &= G_{\sigma=0.15}\big(\chi^2(H_s, H_t)\big) \end{aligned}$$

where G_σ is a Gaussian function of standard deviation σ and H_s and H_t are the Lab histograms of the two super-pixels. We create an edge $(s, t) \in \mathcal{E}$ for each immediate neighbor t of each super-pixel s.

2.3 Optimization

We include our terms in Equation 4 by setting $\phi_s(x_s) = w^{proxy}\phi_s^{proxy}(x_s) + w^{brightness}\phi_s^{brightness}(x_s)$ and $\phi_{s,t}(x_s, x_t) = w^{smoothness}\phi_{s,t}^{smoothness}(x_s, x_t)$. We solve the resulting binary labeling problem using Belief Propagation [Felzenszwalb and Huttenlocher 2006]. Note that GraphCuts [Boykov and Jolly 2001] could also be used at this stage, but this won't be the case once we will augment our formulation to refine the classification, as described next.

3 Refined shadow classifier

While approximate, the coarse shadow classifier gives us sufficient information to select a number of pairs of points in light and in shadow, from which we deduce sun intensity using Equation 3. We now have all the necessary quantities to evaluate Equation 2 for the two possible sun visibility values, which provides us with two candidate reflectances per super-pixel. We denote the two candidates of super-pixel s as $\mathbf{R}_s(0)$ and $\mathbf{R}_s(1)$.

3.1 Pairwise term

We can now exploit our hypothesis that the scene is composed of few reflectances. Given two super-pixels s and t, our goal is to assign them visibility values i and j such that their difference of reflectance $D_{ij} = \|\mathbf{R}_s(i) - \mathbf{R}_s(j)\|$ is minimized. We achieve this goal by augmenting our formulation with additional pairwise terms of the form

$$\phi_{s,t}^{reflectance} = 1 - \big(G_{\sigma=0.04}(D_{ij})\big).$$

However, connecting each super-pixel to all others with such a term is not practical. Instead, we only create edges between super-pixels that have a high probability of sharing the same reflectance for at least one of the label configurations. In practice we select for each super-pixel and each 4 possible label configurations (i, j) the K super-pixels with the smallest D_{ij}. Each super-pixel thus has an edge with cost $\phi_{s,t}^{reflectance}$ with at most $4K$ other super-pixels.

3.2 Optimization

Unfortunately, the pairwise term defined by Equation 5 is not symmetric (i.e. $D_{01} \neq D_{10}$ and $D_{11} \neq D_{00}$), which prevents the use of GraphCuts to solve for a labeling that minimizes the total energy. We thus rely on Belief Propagation [Felzenszwalb and Huttenlocher 2006], which supports such pairwise terms.

References

BARRON, J., AND MALIK, J. 2013. Shape, illumination, and reflectance from shading. Tech. rep., Berkeley Tech Report.

BELL, S., BALA, K., AND SNAVELY, N. 2014. Intrinsic images in the wild. *ACM Transactions on Graphics (Proc. SIGGRAPH) 33*, 4.

BOYKOV, Y. Y., AND JOLLY, M. P. 2001. Interactive graph cuts for optimal boundary & region segmentation of objects in n-d images. In *IEEE International Conference on Computer Vision (ICCV)*, vol. 1.

CHRISTOUDIAS, C. M., GEORGESCU, B., AND MEER, P. 2002. Synergism in low level vision. In *Pattern Recognition, 2002. Proceedings. 16th International Conference on*, vol. 4, IEEE, 150–155.

DUCHÊNE, S., RIANT, C., CHAURASIA, G., LOPEZ-MORENO, J., LAFFONT, P.-Y., POPOV, S., BOUSSEAU, A., AND DRETTAKIS, G. 2015. Multi-view intrinsic images of outdoors scenes with an application to relighting. *ACM Transactions on Graphics*.

FELZENSZWALB, P. F., AND HUTTENLOCHER, D. P. 2006. Efficient belief propagation for early vision. *International Journal Computer Vision 70*, 1 (Oct.), 41–54.

FURUKAWA, Y., AND PONCE, J. 2010. Accurate, dense, and robust multi-view stereopsis. *IEEE Trans. on Pattern Analysis and Machine Intelligence 32*, 8, 1362–1376.

LAFFONT, P.-Y., BOUSSEAU, A., AND DRETTAKIS, G. 2013. Rich intrinsic image decomposition of outdoor scenes from multiple views. *IEEE Transactions on Visualization and Computer Graphics 19*, 2 (February), 210 – 224.

OMER, I., AND WERMAN, M. 2004. Color lines: Image specific color representation. In *Proc. IEEE Computer Vision and Pattern Recognition*, IEEE Computer Society, Washington, DC, USA, 946–953.

PANAGOPOULOS, A., WANG, C., SAMARAS, D., AND PARAGIOS, N. 2013. Simultaneous cast shadows, illumination and geometry inference using hypergraphs. IEEE, vol. 35, 437–449.

PRINCE, S. 2012. *Computer Vision: Models Learning and Inference*. Cambridge University Press.

SNAVELY, N., SEITZ, S. M., AND SZELISKI, R. 2006. Photo tourism: Exploring photo collections in 3d. In *SIGGRAPH Conference Proceedings*, ACM Press, New York, NY, USA, 835–846.