Coherent Intrinsic Images from Photo Collections

ACM Transactions on Graphics
Vol. 31, No. 6, 2012
Pierre-Yves Laffont, Adrien Bousseau, Sylvain Paris, Fredo Durand, and George Drettakis

Presented by Euiwon Nam

School of Electrical Engineering and Computer Science
Kyungpook National Univ.
Abstract

- Decomposition of photo
  - Illumination layer and reflectance layer
  - Under-constrained highly
  - User assistance

- Multi-view stereo
  - Reconstruction of 3D points and normals
  - Estimation to reliably identify reflectance ratios
  - Optimization and coherent solution
Introduction

- Image collections
  - images of scene
    - Variety of viewpoints
    - Captured under different illuminations

- Existing methods
  - Automatic techniques
    - Limited to simple objects
  - Real-world scenes
    - User assistance
  - Fixed of restricted viewpoint
Proposed methods

- Rich information
  - Multiple viewpoints and illuminations in image collection

- Coherent decomposition
  - Large sparse linear system
    - Same reflectance in all images without user assistance
    - Image-guided propagation
      » Augment with term
  - Consideration of radiance ratio between two points

- Identifying pairs of points
  - Similar illumination pairs of points across most conditions
  - Coherence in the reflectance layer
  - Extraction of color variations proper
    - Each image in illumination layer
Related Work

- **Single-image methods**
  - Distinguishment of gradient illumination
  - Texture of same reflectance
  - Chromaticity of same reflectance

- **Multiple-images methods**
  - Median operator as estimator of reflectance
    - Poor decompositions
      - Different normals
      - illumination directions of non uniform
  - Median operator to detect flat surfaces on smooth illumination
  - Timelapse sequences to derive shadow mask and images lit
  - Capture of light probes to estimate reflectance field
- Laser scan and multiple lighting and viewing conditions
  - Performing Relight and estimating Lambertian reflectance
- BRDFs and distant illumination
  - 3D scenes reconstructed with multi-view stereo
- Light probe and multiple images under single lighting condition
- Depth camera to compute intrinsic decompositions
Overview

- Proposed method
  - Collection of photographs \( \{I_i\} \) of given scene captured
    - Different viewpoints and under varying illumination
  - Decomposition into illumination layer \( S_i \) and reflectance layer \( R_i \)
    - each pixel \( p \) and each color channel \( c \), \( I_{ic}(p) = S_{ic}(p)R_{ic}(p) \)

Fig. 2. Our method infers reflectance ratios between points of a scene and then expresses the computation of illumination in all images in a unified least-square optimization system.
Reflectance ratios

- Relations on reflectance between pairs of points
  - Radiance \( I \) towards camera
    \[
    I(P) = R(P) \int_{\Omega} L(P, \hat{\omega}) \, (\hat{\omega} \cdot \hat{n}(P)) \, d\hat{\omega}
    \]  
    where \( L(P, \hat{\omega}) \) is incoming radiance arriving at \( P \) from direction \( \hat{\omega} \), \( \hat{n}(P) \) is normal at \( P \), and \( \Omega \) is hemisphere centered at \( \hat{n}(P) \).

  - Ratio of radiance between two points
    \[
    \frac{I(q)}{I(p)} = \frac{R(q)}{R(p)} \int_{\Omega} L(q, \hat{\omega}) \, (\hat{\omega} \cdot \hat{n}) \, d\hat{\omega}
    \]  
    \[
    \frac{I(q)}{I(p)} = \frac{R(q)}{R(p)} \int_{\Omega} L(p, \hat{\omega}) \, (\hat{\omega} \cdot \hat{n}) \, d\hat{\omega}
    \]
– Visible points for images
  • Equal to between radiance ratio and reflectance ratio
– Difficulty to match in a few images
– Median operators as estimator

\[
\frac{R(q)}{R(p)} = \text{median}_{i \in \mathcal{I}(p,q)} \left( \frac{I_i(q)}{I_i(p)} \right)
\]  

(3)

where median is taken only over images of set \( \mathcal{I}(p,q) \subset \{I_i\} \) in which both \( p \) and \( q \) are visible.

– Ambient occlusion
  • Compensation for differences in visibility
    – Proportion of hemisphere visible from \( p \)
      » Ambient occlusion factor \( \alpha(p) \)

\[
\frac{I(q)}{I(p)} = \frac{R(q)}{R(p)} \frac{\alpha(q)}{\alpha(p)}
\]  

(4)
Selection of constrained pairs

- Subsampling set of all possible constraints
  - normals and distance
  - Discardment of unreliable constraints
    - Simple statistical criterion
- Geometric criterion
  - Selection of candidate pairs on applied constraint
    - Distance $d_n(p, q)$ on normal orientation

$$d_n(p, q) = |1 - \vec{n}(p) \cdot \vec{n}(q)|$$ (5)
- Sampling density functions

**Algorithm 1** Sampling according to 3D distances or normals

1. Estimate the density of distances $f_{original}(d(p, q))$ of all points $q$ to the current point $p$. We use the Matlab `ksdensity` function, which computes a probability density estimate of distances to $p$ from a set of samples $d(p, q)$ by accumulating normal kernel functions centered on each sample.

2. Assign to each point $q$ a sampling probability based on desired distribution $\mathcal{N}(\sigma)$ and the density of distances $f_{original}$:

   $$Pr(q) = \exp \left( - \frac{d(p, q)^2}{2\sigma^2} \right) / f_{original}(d(p, q))$$

3. Select a subset of points according to their probabilities $Pr(q)$ using inversion sampling.
Fig. 3. 2D Illustration of our sampling algorithm for a single point. (a) Given an oriented point cloud, we wish to select N points so that their distances $d_{3D}$ and $d_{\bar{n}}$ to a reference point (black square) follow normal distributions. (b) We first embed the point cloud in a grid and compute Euclidean distances to the cell containing the reference point; the distance is color-coded from blue to red. We infer a sampling probability for each cell based on $d_{3D}$ as described in Algorithm 1, from which we draw N samples to choose the number of points to select in each cell, shown as black numbers. (c) Finally, we sample the corresponding number of points within each cell based on the normal discrepancy $d_{\bar{n}}$. Note that a point can be sampled multiple times if its cell contains too few points.
Photometric statistical criterion
- Detection and rejection such unreliable pairs
  - Less than 50% of radiance ratio values close to median
  - Visible in less than 5 images

\[
\left| \log \left( \frac{I_j(q)}{I_j(p)} \right) - \text{median} \log \left( \frac{I_i(q)}{I_i(p)} \right) \right| > 0.15
\]  

Fig. 4. Analysis of the distribution of radiance ratio (red channel, log scale) between two 3D points (red dots) with similar normals, under varying viewpoints and lighting. The PDF has a dominant lobe, corresponding to (b) and (c) where both points receive approximately the same incoming radiance. In (a), the light is visible from only one of the points and the corresponding radiance ratio falls in a side lobe. (d) shows the point cloud for image (a).
Multi-image guided decomposition

- **Pairwise reflectance constraints**
  - Deducing ratio $Q_j(p, q)$
    - Between illumination of corresponding pixels in image $j$
      \[
      Q_j(p, q) = \frac{S_j(p)}{S_j(q)} = \frac{I_j(p)}{I_j(q)} \frac{R_j(q)}{R_j(p)} = \frac{I_j(p)}{I_j(q)} \text{median}_{i \in I(p,q)} \left( \frac{I_j(q)}{I_j(p)} \right)
      \]
      where $S_j$ is illumination layer of image $j$.
  - Constraint in unknown illumination values by eq.7
    \[
    Q_j(p, q) \frac{1}{2} S_j(q) = Q_j(p, q) \frac{1}{2} S_j(p)
    \]
  - Least-squares sense to get energy $E_{\text{constraints}}$
    \[
    \sum_{j} \sum_{(p, q)} \left[ Q_j(p, q) \frac{1}{2} S_j(q) - Q_j(p, q) \frac{1}{2} S_j(p) \right]^2
    \]
Smoothness

- Translation of matting into local energy
  - Relative color at pixel $x$ with illumination value
    - each channel $S_{jc}(x)$ using affine model

$$
\sum_{c \in \{r, g, b\}} \sum_{y \in w_j^x} (S_{jc}(y) - a_{jc}^x \cdot I_j(y) - b_{jc}^x)^2 + \varepsilon (a_{jc}^x)^2
$$

where $w_j^x$ is 3 x 3 window centered on $x$, $a_{jc}^x$ and $b_{jc}^x$ are unknown parameters of affine model, constant over window, and $\varepsilon = 10^{-6}$ is parameter controlling regularization $(a_{jc}^x)^2$ that favors smooth solutions.

- Energy depends on illumination

$$
E_{smoothness} = \sum_{c \in \{r, g, b\}} \sum_j \hat{S}_j^T M_{jc} \hat{S}_{jc}
$$

where vectors $\hat{S}_j$ stack unknown illumination values in image $j$ and matrices $M_j$ encode smoothness prior over each pixel neighborhood in this image.
- Adding term below to favor illumination values

\[
\sum_{x} \sum_{c \in \{r, g, b\}} (S_{jc}(x) - \frac{1}{3}[I_{jr}(x) + I_{jg}(x) + I_{jb}(x)])^2
\]  

(12)

- Coherent reflectance

- Constraint on illumination for same reflectance

\[
R_m(p) = R_n(p) \Rightarrow \frac{I_m(p)}{S_m(p)} = \frac{I_n(p)}{S_n(p)}
\]

\[
\Rightarrow I_m(p)S_n(p) = I_n(p)S_m(p)
\]  

(13)

- Additional energy term \(E_{coherence}\)

\[
\sum_{p} \sum_{m \in I(p)} \sum_{n \in I(p)} (I_m(p)S_n(p) - I_n(p)S_m(p))^2
\]  

(14)
Solving the system

- Energy $E^k_{coherence}(m, p)$ for point $p$

$$\sum_{\substack{n \in I(p) \\ n < m}} (I_m(p)S_n^k(p) - I_n(p)S_m^k(p))^2 + \sum_{\substack{n \in I(p) \\ n > m}} (I_m(p)S_n^{(k-1)}(p) - I_n(p)S_m^k(p))^2$$

- Single least squares term
  - Plus constant

$$\left( \sum_{\substack{n \in I(p) \\ n < m}} I_n^2 \right) \left( S_m^k - \frac{I_m\left( \sum_{n \in I(p)} I_n S_n^k \right)}{\sum_{n \in I(p)} I_n^2} \right)^2 + \text{constant}$$

where for clarity, we use notation $S_n^k = S_n^k$ when $n < m$ and $S_n^{(k+1)}$ when $n > m$, and omit dependency on $p$. 
Implementation and results

- Intrinsic decompositions
  - Evaluation on a synthetic scene

Fig. 5. Comparison to existing methods and ground truth on a synthetic rendering, generated with path tracing (see text for details). Reflectance and illumination images have been scaled to best match ground truth; sky pixels have been removed.
Fig. 6. Numerical evaluation of five intrinsic decomposition methods. Gray bars indicate Local Mean Squared Error averaged over the three comparison images, while red bars illustrate the standard deviation of LMSE across images.
- Captured scenes
  - Decomposition using 11 and 10 view point

Fig. 7. Results of our decomposition on scenes captured with a flash. Note that the colored residual in the doll illumination is due mainly to indirect light.
- Internet photo collections
  - Famous landmarks

**Table 1.** lists number of images used for each scene

| Scene | Synth. 1 | Captured 2 | Captured 3 | Internet Photo Collections |
|-------|----------|------------|------------|---------------------------|
| $N_d$ | 6        | 5          | 10         | 9                         |
| $N_r$ | 30       | 32         | 48         | 56                        |
| $P_{rec}$ | 100k     | 467k       | 1.3M       | 1M                        |
| $P_{sel}$ | 68k      | 200k       | 199k       | 200k                      |
| $C_{cand}$ | 1.4M     | 1.5M       | 1.5M       | 1.3M                      |
| $C_{pair}$ | 260k     | 724k       | 709k       | 197k                      |
| $C_{coher}$ | 39k      | 105k       | 142k       | 66k                       |

For the second and third columns, $N_r$ and $P_{rec}$ are multiplied by 1,000,000.
Fig. 8. Comparison between our approach and existing single-image methods on picture from online collection.
Fig. 9. Results of our method on internet photo collections. Top: another view of the StBasil scene. The reflectance we extract is coherent with the one shown in Fig. 8. Bottom: the specular objects which cast shadows on the facade are a challenging case for multi-view stereo. Our method is able to extract their shadows despite the lack of a complete and accurate 3D reconstruction.
Fig. 1. Our method leverages the heterogeneity of photo collections to automatically decompose photographs of a scene into reflectance and illumination layers. The extracted reflectance layers are coherent across all views, while the illumination captures the shading and shadow variations proper to each picture. Here we show the decomposition of three photos in the collection.
Analysis and limitations

Fig. 10. Effect of compensating for ambient occlusion on the decomposition of a synthetic image (a). Without special treatment, the reflectance under the arches appears darker (b) because these regions systematically receive less illumination. Correcting the pairwise reflectance constraints by compensating for ambient occlusion (Sec. 4.1) yields a reflectance (c) closer to ground truth (d).
Fig. 11. Influence of the pairwise relative constraints on another image of the “Doll scene”. (a) Without pairwise reflectance constraints, texture cannot be successfully separated from lighting and the resulting illumination layer contains large texture variations. (b) Enabling these constraints allows recovering a smooth illumination on the tablecloth, despite the complexity of its texture.
Fig. 12. Comparison between the decomposition, before and after multi-view coherence in the Florence scene. The coherence constraints between multiple views allow our method to recover a coherent reflectance even under mixed lighting conditions such as this bright sunset with dark blue shadows.
Fig. 13. Given two views of the same scene under different lighting (a, b), we transfer the illumination from one view into the other view (c). We then multiply the transferred illumination by the reflectance layer (d) to synthesize the relit image (f). Transferring the radiance directly fails to preserve the fine details of the reflectance (e).
Fig. 14. We use our lighting transfer to harmonize the illumination over multiple images.
Conclusion

- Coherent intrinsic image decompositions
  - Multiple lighting conditions
  - Reconstruction of 3D point clouds
  - Calibration of camera viewpoints
  - Coherence constraints
    - Improvement extracted reflectance significantly
    - Different lighting conditions
  - Illumination transfer and stable transitions
    - Views with consistent illumination