Intrinsic Image Decomposition ... and other things

Carsten Rother
Microsoft Research Cambridge
I3D group
http://research.microsoft.com/en-us/groups/i3d/

Research Themes

“Ultimate” IO This theme explores new designs for cameras for capturing reality, including acquisition of geometry, material properties and light fields, as well as novel display technologies for presenting rich digital content to users.

Recovering and Understanding Reality This theme explores new probabilistic models and algorithms to infer physical properties of the scene such as lighting, materials, and semantics such as object recognition and scene understanding. We also explore applications of such methods, particularly for augmented reality.

Reality Remixed This theme tackles the hard technical and user interface challenges in making augmented reality a reality. Key areas are real-time tracking and reconstruction, augmented reality IO hardware, and rich user experiences that blur the physical and digital.

deForm This theme explores new algorithms for representing and inferring the dynamic nature of the real world, digitally. This includes moving and deforming bodies, hands and faces of users, and our dynamic environments. We also explore new ways in which these dynamic digital representations can be used interactively.
PatchMatch Stereo
[Bleyer, Rhemann, Rother ‘11]

The challenge:
“How can we find the correct slanted plane at each pixel?”
(infinite number of candidate planes)
Algorithm - Intuition

PatchMatch method [Barnes et al., SIGGRAPH09]:

• Relatively large regions of pixels can be modeled by approximately the same plane

• A single (pixel) good guess per plane is enough to succeed - it will be propagated to neighboring pixels

Left image – Sawtooth (Middlebury)  Ground truth disparities  Image consists of 3 planes - ~80,000 guesses for yellow plane
Algorithm

- Random initialization of planes parameters \((a_p, b_p, c_p)\) per pixel \(p\)

- Propagation

  + propagation across views and temporally

- Randomization: sample around current solution

- Total of 3 iterations

- How does it relate to e.g. Particle Belief Propagation?
PatchMatch Stereo
[Bleyer, Rhemann, Rother ‘11]
Intrinsic Image Decomposition
with a Global Sparseness Prior on Reflectance
[Gehler, Rother, Kiefel, Zhang, Schölkopf, NIPS ‘11]

- Decompose an image into its
  - R: reflectance (material property)
  - s: shading + shadows
- Here: no specular highlights

\[ I^c_i = s_i \ R^c_i \]

Pixel-index: i
Color channels: \( c \in \{r, g, b\} \)

Input Image I  
Reflectance R  
Shading s
Motivation

• Interesting and Challenging
• Desired input to other systems (Shape from Shading, Segmentation, Recognition, Image Feature extraction,...)
• Better image manipulation tools, eg [Bousseau et. al., User Assisted Intrinsic Images, Siggraph Asia 09]
Image Database

- **Main Problem**: limited availability of ground truth data
  [Grosse, Johnson, Adelson and Freeman, *Ground-truth dataset and baseline evaluations for intrinsic image algorithms*, ICCV 2009]
Examples

Image

Reflectance

Shading
Model Formulation

- Problem Reduction: search for a scalar “r” only

\[ I^c_i = s_i \, R^c_i \quad \Rightarrow \quad R^c_i = r_i \, \tilde{R}^c_i, \quad \tilde{R}^c_i = I^c_i / \| I_i \| \]
\[ s_i = \| I_i \| / r_i \]

- Pose the problem as an CRF and optimize for MAP

\[
\min_{r,k} w_s E_s(r) + w_r E_r(r) + w_{cl} E_{cl}(r,k)
\]

- Smoothness Prior on Shading
- Gradient Consistency (Retinex)
- Global Sparse Reflectance Prior
Smoothness Prior $E_s(r)$

- Expect the shading to vary smoothly

- 4-connected MRF:

$$s_i = \frac{\| I_i \|}{r_i}$$

$$E_s(r) = \sum_{i \sim j} (\frac{\| I_i \|}{r_i} - \frac{\| I_j \|}{r_j})^2$$
Global Sparseness Prior on Reflectance

- Idea (empirically justified [Omer and Werman ‘04]): Image has often only few materials (reflectance)

\[ E_{cl}(r, k) = \sum_i (r_i \hat{R}_i - \hat{R}_{k_i})^2 \]

\( \hat{R}_{k_i} \) cluster centre
Influence of Global Reflectance Prior

In practice 50-150 clusters (cross-validation)
Gradient consistency

- Retinex (in a nutshell)
  - Gradients are either caused by shading or reflectance edges
- Classify the Gradients into:
  shading (weak edges) or reflectance (strong edges)

\[ E_r(r) = \sum_{i \neq j} ((\log r_i - \log r_j) - \Delta_{ij})^2 \]
Example – Colour Retinex

Input: Turtle

Ground truth
Optimization

Iterate optimization between: r and k

Algorithm 1 Coordinate Descent for solving (3)

1: Select $r^0$ as described in the text
2: $k^0 \leftarrow$ K-Means clustering of \{${r_i^0 \tilde{R}_i, i = 1, \ldots, N}$\}
3: $t \leftarrow 0$
4: repeat
5: $r^{t+1} \leftarrow$ optimize (3) with $k^t$ fixed
6: $\hat{R}_c = \sum_{i: k_i = c} r_i \tilde{R}_i / |\{i : k_i = c\}|$
7: $k^{t+1} \leftarrow$ assign new cluster labels with $r^{t+1}$ fixed
8: $t \leftarrow t + 1$
9: until $E(r^{t-1}, k^{t-1}) - E(r^t, k^t) < \theta$
## Results

Full model better than individual terms

<table>
<thead>
<tr>
<th>comment</th>
<th>$E_s$</th>
<th>$E_{cl}$</th>
<th>$E_{ret}$</th>
<th>LOO-CV</th>
<th>best single</th>
<th>image opt.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Color Retinex</td>
<td>-</td>
<td>-</td>
<td>✓</td>
<td>29.5</td>
<td>29.5</td>
<td>25.5</td>
</tr>
<tr>
<td>no edge information</td>
<td>✓</td>
<td>✓</td>
<td>-</td>
<td>30.0</td>
<td>30.6</td>
<td>18.2</td>
</tr>
<tr>
<td>Col-Ret+ global term</td>
<td>-</td>
<td>✓</td>
<td>✓</td>
<td>27.2</td>
<td>24.4</td>
<td>18.1</td>
</tr>
<tr>
<td>full model</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>27.4</td>
<td>24.4</td>
<td>16.1</td>
</tr>
</tbody>
</table>
## Results

<table>
<thead>
<tr>
<th>Method</th>
<th>LOO-CV</th>
<th>rank</th>
<th>best single</th>
<th>im. opt.</th>
</tr>
</thead>
<tbody>
<tr>
<td>TAP05 [17]</td>
<td>56*</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>TAP06 [16]</td>
<td>39*</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>SHE [14]+</td>
<td>n/a</td>
<td>n/a</td>
<td>56.2</td>
<td>n/a</td>
</tr>
<tr>
<td>SHE [15]×</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
<td>20.4</td>
</tr>
<tr>
<td>BAS [7]</td>
<td>72.6</td>
<td>5.1</td>
<td>60.3</td>
<td>36.6</td>
</tr>
<tr>
<td>Gray-Ret [7]</td>
<td>40.7</td>
<td>4.9</td>
<td>40.7</td>
<td>28.9</td>
</tr>
<tr>
<td>Col-Ret</td>
<td>29.5</td>
<td>3.7</td>
<td>29.5</td>
<td>25.5</td>
</tr>
<tr>
<td><strong>full model</strong></td>
<td><strong>27.4</strong></td>
<td><strong>3.0</strong></td>
<td><strong>24.4</strong></td>
<td><strong>16.1</strong></td>
</tr>
<tr>
<td>Weiss [19]</td>
<td>21.5</td>
<td>2.7</td>
<td>21.5</td>
<td>21.5</td>
</tr>
<tr>
<td>Weiss+Ret [7]</td>
<td>16.4</td>
<td>1.7</td>
<td>16.4</td>
<td>15.0</td>
</tr>
</tbody>
</table>
Results

Input; Turtle

Ground truth

Color-Retinex; LOC-CV; LMSE 67.9

Color-Retinex; image-optimal; LMSE 39.1

Color-Retinex + global term; image optimal; LMSE 24.2

Full model; image optimal; LMSE 22.4

Full model; LOC-CV; LMSE 60.4
Results

Input; Teabag

Ground truth

Color-Retinex; image-optimal; LMSE 22.5

Our model (no edge info); image optimal; LMSE 28.0

Full model; image optimal; LMSE 16.7