Intrinsic images

CVFX @ NTHU

16 April 2015

Outline

Lighting, shading, reflectance

Papers

"Deriving intrinsic images from image sequences,"

- Yair Weiss. ICCV 2001

"Estimating intrinsic images from image sequences with biased illumination,"

- Matsushita, Lin, Kang, Shum. ECCV 2004

"Estimating intrinsic component images using non-linear regression,"

- Tappen, Adelson, and Freeman. CVPR 2006

"User-assisted intrinsic images"

- Bousseau, Paris, and Durand. SIGGRAPH Asia 2009

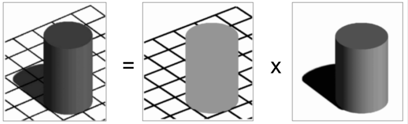
Intrinsic images

An image is decomposed into a reflectance image and an illumination image

- An ill-posed problem $I(x, y) = R(x, y)L(x, y)$

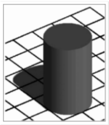
A useful midlevel description of scenes

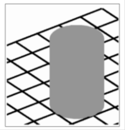
- Viewpoint dependent
- The physical causes of changes in illumination at different points are not made explicit



Advantages of the intrinsic representation

- The task of segmentation may be poorly defined on the input image and many segmentation algorithms make use of arbitrary thresholds in order to avoid being fooled by illumination changes
- On an intrinsic reflectance image even primitive segmentation algorithms would correctly segment the region of an object





Advantages of the intrinsic representation (cont.)

View-based template matching and shape-from-shading would be less brittle if they could work on the intrinsic image representation rather than on the input image

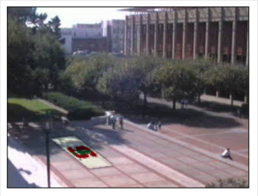
Deriving intrinsic images from image sequences

Given a sequence of T images $\{I(x, y, t)\}_{t=1}^{T}$ in which the reflectance is constant over time and only the illumination changes, can we then solve for a single reflectance image $R(x, y)$ and T illumination images $\{L(x, y, t)\}_{t=1}^{T}$?

$$I(x, y, t) = R(x, y)L(x, y, t)$$

The problem is still ill-posed: at every pixel there are T equations and $T + 1$ unknowns. One can simply set $R(x, y) = 1$ and $L(x, y, t) = I(x, y, t)$.

Example



ML estimator assuming sparseness

Transform the problem into log domain

$$i(x, y, t) = r(x, y) + \ell(x, y, t).$$

To make the problem solvable, we want to assume a distribution over $\ell(x, y, t)$.

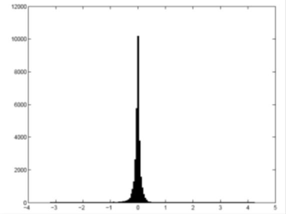
First thought

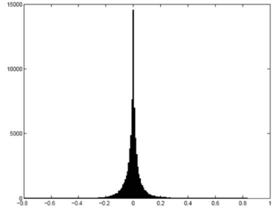
- Illumination images are of lower contrast than reflectance images?
- It is rarely true for the outdoor scenes
- Edges due to illumination often have as high a contrast as those due to reflectance changes

Statistics of natural images

When derivative filters are applied to luminance in natural images (in log domain), the filter outputs tend to be sparse.

- Peaked at zero and fall off much faster than a Gaussian





Lecture videos about natural images

Yair Weiss and Bill Freeman: What makes a good model of natural images? (CVPR 2007)

- Weiss's talk given at UC Berkeley on February 20, 2007
- `http://www.archive.org/details/Redwood_Center_2007_02_20_Yair_Weiss`

From Learning Models of Natural Image Patches to Whole Image Restoration

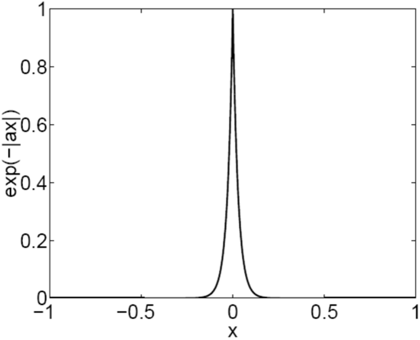
- Zoran's talk given at UC Berkeley on March 1, 2012
- `http://archive.org/details/Redwood_Center_2012_03_01_Daniel_Zoran`

Fit by a Laplacian distribution

Laplacian distribution

$$P(x) = \frac{1}{Z} e^{-\alpha|x|}$$

p.d.f.



How to use the sparseness property?

Assume we have N filters $\{f_n\}$ and we denote the filter outputs by $o_n(x, y, t) = i \star f_n$.

We use r_n to denote the reflectance image filtered by the nth filter $r_n = r \star f_n$.

Claim 1

Assume filter outputs applied to $\ell(x, y, t)$ are Laplacian distributed and independent over space and time. Then the maximum likelihood (ML) estimate of the filtered reflectance image $\hat{r}_n$ are given by

$$\hat{r}_n(x, y) = \mathrm{median}_t \; o_n(x, y, t) \, .$$

Proof of Claim 1

Assuming Laplacian densities and independence yields the likelihood

$$
\begin{aligned}
P(o_n|r_n) &= \frac{1}{Z} \prod_{x,y,t} e^{-\beta |o_n(x,y,t) - r_n(x,y)|} \\
&= \frac{1}{Z} e^{-\beta \sum_{x,y,t} |o_n(x,y,t) - r_n(x,y)|} .
\end{aligned}
$$

Maximizing the likelihood is equivalent to minimizing the sum of absolute deviations from $o_n(x, y, t)$. The sum of absolute values (or L_1-norm) is minimized by the median.

What does Claim 1 imply?

Claim 1 gives us the ML estimate for the filtered reflectance images $\hat{r}_n$. To recover an estimated reflectance $\hat{r}_n$, we solve the overconstrained systems of linear equations

$$f_n \star \hat{r} = \hat{r}_n \,.$$

Over-constrained linear system

N filters

M pixels

$$\begin{bmatrix} f_{11}^T \\ f_{12}^T \\ \vdots \\ f_{1M}^T \\ \vdots \\ f_{N1}^T \\ f_{N2}^T \\ \vdots \\ f_{NM}^T \end{bmatrix} \hat{r} = \begin{bmatrix} \hat{r}_1 \\ \vdots \\ \\ \hat{r}_N \end{bmatrix}$$

NM by M (sparse) | M by 1 | NM by 1

What does Claim 1 imply?

Claim 1 gives us the ML estimate for the filtered reflectance images $\hat{r}_n$. To recover an estimated reflectance $\hat{r}_n$, we solve the overconstrained systems of linear equations

$$f_n \star \hat{r} = \hat{r}_n \, .$$

It can be shown that the pseudo-inverse solution is given by

$$\hat{r} = g \star \left(\sum_n f_n^r \star \hat{r}_n \right)$$

with f_n^r the reversed filter of $f_n : f_n(x, y) = f_n^r(-x, -y)$ and g a solution to

$$g \star \left(\sum_n f_n^r \star f_n \right) = \delta \, .$$

Note

real matrix

A not invertible

$A^T A$ invertible

filter $\begin{pmatrix} \text{real} \\ \text{function} \end{pmatrix}$

f no inverse

$f^r * f$ has inverse

$$\langle f, g \rangle = \int f g \, dx$$

$$(f * g)(t) = \int f(t-\tau)\, g(\tau)\, d\tau$$

Note

$$g * \left(\sum_n f_n^r * f_n \right) = \delta$$

$$g * \left(\sum_n f_n^r * f_n \right) * \hat{r} = \delta * \hat{r} = \hat{r}$$

$$g * \left(\sum_n f_n^r * (f_n * \hat{r}) \right) = \hat{r}$$

$$g * \left(\sum_n f_n^r * \hat{r}_n \right) = \hat{r}$$

$$\boxed{g * \sum_n f_n^r} * \hat{r}_n = \hat{r}$$

recall $(A^T A)^{-1} A^T$

Note

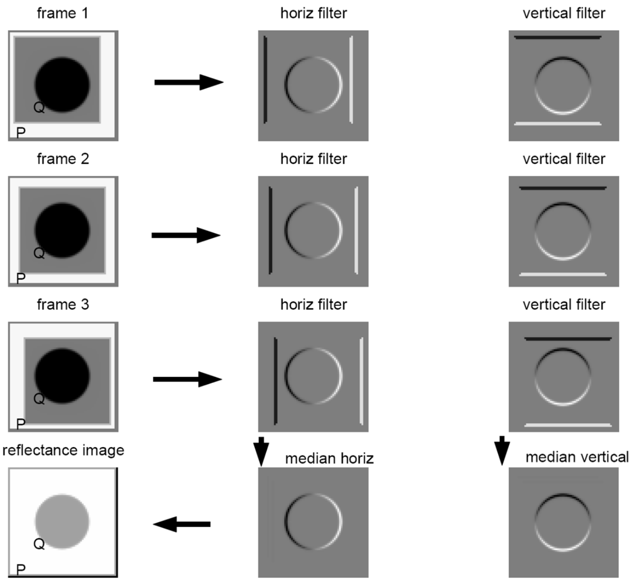
$$G = 1 \,./ \sum_n \mathrm{FFT}(f_n^r) \cdot \mathrm{FFT}(f_n)$$

$$g = \mathrm{IFFT}(G)$$

$$R = \left(\mathrm{FFT}(g) \cdot \sum_n \mathrm{FFT}(f_n^r)\right) \cdot \mathrm{FFT}(\hat{r}_n)$$

$$\hat{r} = \mathrm{IFFT}(R)$$

Example


frame 1
horiz filter
vertical filter
Q
P
frame 2
horiz filter
vertical filter
Q
P
frame 3
horiz filter
vertical filter
Q
P
reflectance image
median horiz
median vertical
Q
P

Claim 2

Let $p_\epsilon = P(|f_n \star \ell(x,y,t)| < \epsilon)$. Then the estimated filtered reflectances are within ϵ of the true filtered reflectances with probability at least

$$P(|\hat{r}_n - r_n^*| < \epsilon) = \sum_{k=1}^{T/2} \left(\begin{array}{c} T \\ k \end{array} \right) (1 - p_\epsilon)^k p_\epsilon^{(T-k)} ,$$

or, equivalently,

$$P(|\hat{r}_n - r_n^*| < \epsilon) = \sum_{k=T/2}^{T} \left(\begin{array}{c} T \\ k \end{array} \right) (1 - p_\epsilon)^{(T-k)} p_\epsilon^k .$$

Proof of Claim 2

If more than 50% of the samples of $f_n \star \ell(x, y, t)$ are within ϵ of some value, then by the definition of the median, the median must be within ϵ of that value. The claim follows from the binomial formula for the sum of T independent events.

Details

Let $p_\epsilon = P(|f_n * \ell(x,y,t)| < \epsilon)$

$$p_\epsilon = P(|f_n * (i(x,y,t) - r(x,y))| < \epsilon)$$

$$= P(|o_n(x,y,t) - r_n(x,y)| < \epsilon)$$

We want to compute

$P(|\hat{r}_n \quad r_n| < \epsilon)$ which measures how good the estimate $\hat{r}_n$ will be

---

Since $\hat{r}_n$ is obtained by $\hat{r}_n = \mathrm{median}_t\, O_n(x,y,t)$

We require more than 50% of $O_n(x,y,t)$ is close enough to $r_n(x,y)$, that is

$$\sum_{k=\lfloor \frac{T}{2} \rfloor + 1}^{T} \binom{T}{k} p_\epsilon^k (1-p_\epsilon)^{T-k}$$

Toy example

Assume $P_\epsilon = P(|f_n * l(x,y,t)| < \epsilon) = 0.85$

about 85% of the filtered illuminations are smaller than ϵ

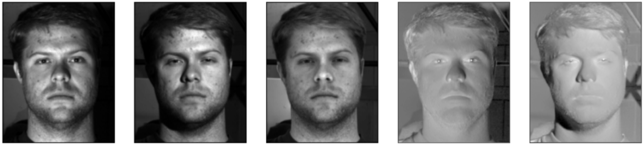
Suppose we use three images

$$\binom{3}{2}(0.85)^2 \cdot (0.15) + \binom{3}{3}(0.85)^3 = 0.9392$$

The probability of the estimated $\hat{r}_n$ being very close to the real r_n is greater than 0.93

Yale face database

- 64 images taken with variable lighting



frame 2 | frame 11 | ML reflectance | ML illumination 2 | ML illumination 11

UC Berkeley webcam

frame 1

frame 11

ML reflectance

ML illumination 1

ML illumination 2

What if the illumination is biased?

Estimating intrinsic images from image sequences with biased illumination

- Matsushita, Lin, Kang, Shum. ECCV 2004

Intrinsic images

$$
\begin{aligned}
I(x,y,t) &= \rho(x,y)L(x,y,t) \\
&= \rho(x,y)\{L_D(x,y,t)+\alpha(x,y,t)\} \\
&= \rho(x,y)\{E(t)g(x,y,t)(\mathbf{n}(x,y)\cdot\mathbf{l}(t))+\alpha(x,y,t)\} \\
&= \rho(x,y)E(t)\{g(x,y,t)(\mathbf{n}(x,y)\cdot\mathbf{l}(t))+\alpha^{'}(x,y,t)\}
\end{aligned}
$$

$\rho(x,y)$: reflectance
$E(t)$: illumination intensity
$g(x,y,t)$: binary shadow map
$\mathbf{n}(x,y)$: surface normal
$\mathbf{l}(t)$: illumination direction
$\alpha(x,y,t)$: ambient light

Review: filtered reflectance and filtered illumination

$$\log \hat{\rho}_n(x, y) = \text{median}_t \{ f_n \star \log I(x, y, t) \}$$

$$\log \hat{L}_n(x, y, t) = f_n \star \log I(x, y, t) - \log \hat{\rho}_n(x, y)$$

$$(\log \hat{\rho}, \log \hat{L}) = h \star \left(\sum_n f_n^r \star (\log \hat{\rho}_n, \log \hat{L}_n) \right)$$

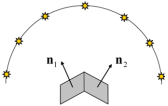
$$h \star \left(\sum_n f_n^r \star f_n \right) = \delta$$

Unbiased illumination samples

For two adjacent pixels with intensities $I_1(t)$ and $I_2(t)$

$$\hat{\rho}_n = \text{median} \frac{I_1(t)}{I_2(t)} = \text{median}_t \frac{\rho_1}{\rho_2} \cdot \frac{E(t)\{g_1 \cdot (\mathbf{n}_1 \cdot \mathbf{l}(t)) + \alpha_1'\}}{E(t)\{g_2 \cdot (\mathbf{n}_2 \cdot \mathbf{l}(t)) + \alpha_2'\}}$$

Assumption: cast shadows do not affect the median

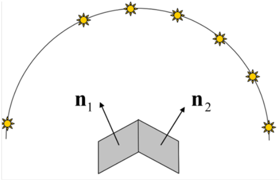
unbiased illumination samples: 

$$\text{median}_{I(t)\in\Omega_t}\ \mathbf{n}_1 \cdot \mathbf{l}(t) - \mathbf{n}_2 \cdot \mathbf{l}(t) = 0$$

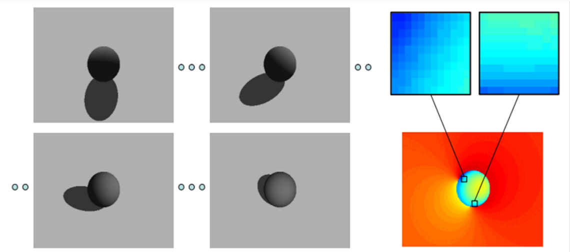
$\hat{\rho}_n = \rho_1 / \rho_2$

Biased illumination

$\hat{\rho}_n \neq \rho_1 / \rho_2$



biased illumination



Hard constraints

Inter-frame constraint (constant reflectance)

$$\frac{I_p(t_i)}{I_p(t_j)} = \frac{L_p(t_i)}{L_p(t_j)}, \quad 0 \le i, j < N, i \neq j\,.$$

N: # of observations

Inter-pixel constraint

$$\frac{I_p(t_i)}{I_q(t_i)} = \frac{\rho_p}{\rho_q} \cdot \frac{L_p(t_i)}{L_q(t_i)}, \quad 0 \le i < N, q \in \omega_p\,.$$

ω_p: neighborhood

$$\sum_{p,i,j:i\neq j} \left(\frac{I_p(t_i)}{I_p(t_j)} - \frac{L_p(t_i)}{L_p(t_j)} \right)^2 + \sum_{p,q,i:q\in\omega_p} \left(\frac{I_p(t_i)}{I_q(t_i)} - \frac{\rho_p}{\rho_q} \cdot \frac{L_p(t_i)}{L_q(t_i)} \right)^2 = 0\,.$$

Flatness

$$e_{pq}(t_i) = \left| \arctan\left\{ \text{median}_t \left(\frac{I_p}{I_q} \right) \right\} - \arctan\left\{ \frac{I_p}{I_q} \right\} \right|$$

$$\xi_{pq}(t_i) = \begin{cases} 1 & (e_{pq}(t_i) < \epsilon : accept) \\ 0 & (e_{pq}(t_i) \geq \epsilon : reject) \end{cases}$$

$$f_{pq} = \left(\frac{\sum_i \xi_{pq}(t_i)}{N} \right)^2$$

Energy minimization based on smoothness constraints

$$
\begin{aligned}
E_\Omega &= \sum_p E_p(\Delta\rho_p, \Delta L_p(t)) \\
&= \sum_p \sum_{q\in\omega_p} \{(\rho_p - \rho_q)^2 + \lambda f_{pq}(t_i)(L_p(t_i) - L_q(t_i))^2\}
\end{aligned}
$$

Hessian matrix

$$
H_p = \begin{bmatrix} \partial^2 E_p/\partial\rho^2 & \partial^2 E_p/\partial\rho\partial L \\ \partial^2 E_p/\partial L\partial\rho & \partial^2 E_p/\partial L^2 \end{bmatrix} = \begin{bmatrix} \sum_{q\in\omega_p} 1 & 0 \\ 0 & \lambda \sum_{q\in\omega_p} f_{pq} \end{bmatrix}
$$

convex

Algorithm

Step 1: Initialization

Step 2: Hard constraints

Step 2: Hard constraints

1. Inter-frame constraint. Update $L_p(t_i)$.

$$L_p(t_i) \leftarrow \sum_{j \neq i} \left(\frac{I_p(t_i)}{I_p(t_j)} L_p(t_j) \right) / (N-1) \tag{16}$$

2. Inter-pixel constraint. Update $L_p(t_i)$ and ρ_p with ratio error β. Letting M_{ω_p} be the number of p's neighboring pixels,

$$\beta_p(t_i) = \left(\sum_{q \in \omega_p} \frac{I_p(t_i)}{I_q(t_i)} \cdot \frac{\rho_q L_q(t_i)}{\rho_p L_p(t_i)} \right) / M_{\omega_p}. \tag{17}$$

   Since the error ratio $\beta_p(t_i)$ can be caused by some unknown combination of ρ and L, we distribute the error ratio equally to both ρ and L in (18) and (20), respectively.

$$L_p(t_i) \leftarrow \sqrt{\beta_p(t_i) L_p(t_i)}, \tag{18}$$

$$\beta_p = \left(\sum_i \beta_p(t_i) \right) / N, \tag{19}$$

$$\rho_p \leftarrow \sqrt{\beta_p} \rho_p. \tag{20}$$

3. Return to 1. unless Equation (10) is satisfied.

Algorithm

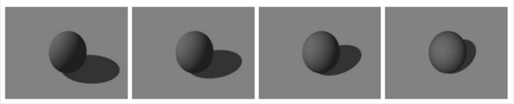
Step 3: Energy minimization by the conjugate gradient method

$$
\begin{aligned}
E_\Omega &= \sum_p E_p(\Delta\rho_p, \Delta L_p(t)) \\
&= \sum_p \sum_{q \in \omega_p} \{(\rho_p - \rho_q)^2 + \lambda f_{pq}(t_i)(L_p(t_i) - L_q(t_i))^2\}
\end{aligned}
$$

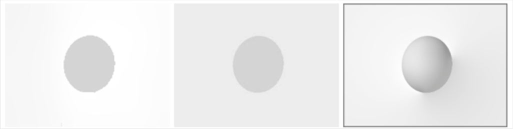
Go back to Step 2 if not converges

Results

input



reflectance



proposed method ground truth ML estimation

proposed method ML estimation

Learning from data

Estimating Intrinsic Component Images Using Non-Linear Regression

- Tappen, Adelson, and Freeman. CVPR 2006

Estimating Intrinsic Component Images Using Non-Linear Regression

Estimate a set of local linear constraints, such as the derivatives, using local image data

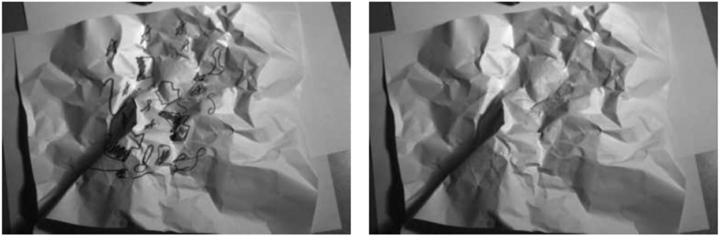
- Estimate the filtered versions of the intrinsic component image
- Use training data to learn to predict the derivatives of the shading and reflectance images, rather than basing the estimates on a simple model of the world.

Solve for the image that best satisfies these constraints, by using a method akin to a pseudo-inverse

- Horizontal and vertical derivatives are differently weighted

Creating shading and reflectance data of real-world surface

- How to create ground-truth demopositions

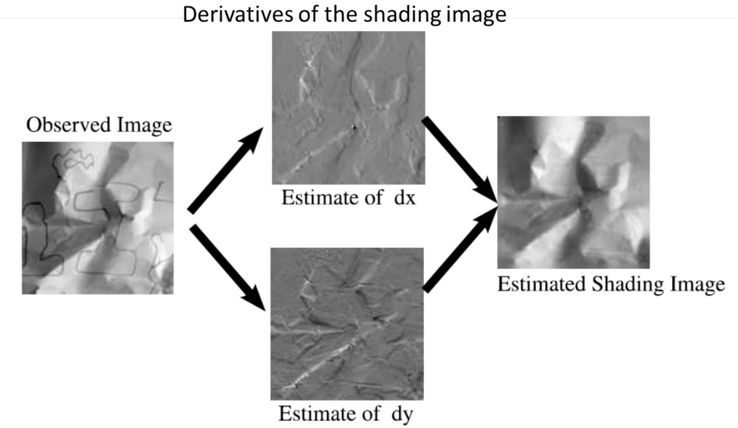


(a) Red Channel (b) Green Channel

- A piece of paper colored with a green marker
- The green channel, containing no markings, is used as the shading image

Locally estimating constraint values

- Using a patch of the observed image to estimate a particular pixel of the filtered intrinsic component image



Learning the estimator

- Training pairs of observed patches and filtered intrinsic components

$$(o_1, c_1) \ldots (o_M, c_M)$$

- Minimize the square error

$$E = \sum_i^M (r(o_i) - c_i)^2$$

$$r(o) = \frac{\sum_{i=1}^N \left(e^{-\sum_j (p_i^j - o^j)^2} \right) f_i{}^T o}{\sum_{i=1}^N e^{-\sum_j (p_i^j - o^j)^2}}$$

choose N prototype patches $\{p_i\}$ and coefficients $\{f_i\}$ using a boosting algorithm

Reconstructing the image

Weighted least squares

$$\hat{x} = (C^T W C)^{-1} C^T W \hat{c}$$

W is block-diagonal

$$C = \left[\begin{array}{c} C_{dx} \\ C_{dy} \end{array} \right]$$

C_{dx} and C_{dy} denote the matrices that express the 2D image convolution with each filter as a matrix

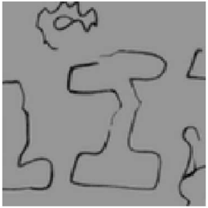
$$\hat{c} = \left[\begin{array}{c} \hat{c}_{dx} \\ \hat{c}_{dy} \end{array} \right]$$

$\hat{c}$ contains estimated derivatives

Results

Observed Image

Shading Image from ExpertBoost



(a) Ground Truth Albedo Image

(c) Estimated Albedo after Adjusting Weights

Application to denoising

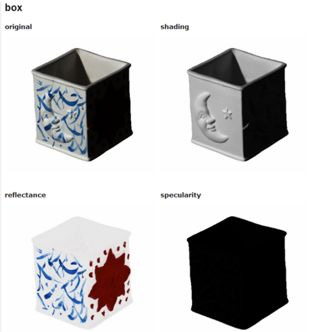
Use different types of image patches and filters to learn an estimator for denoising

User-assisted intrinsic images

(a) Original photograph (b) User scribbles (c) Reflectance (d) Illumination (e) Re-texturing

MIT intrinsic images

- http://people.csail.mit.edu/rgrosse/intrinsic



Summary

Find an estimate of filtered intrinsic image

Reconstruct the intrinsic image from the filtered version