A Joint Intrinsic-Extrinsic Prior Model for Retinex

ICCV 17
Venice, Italy
October 22-29, 2017

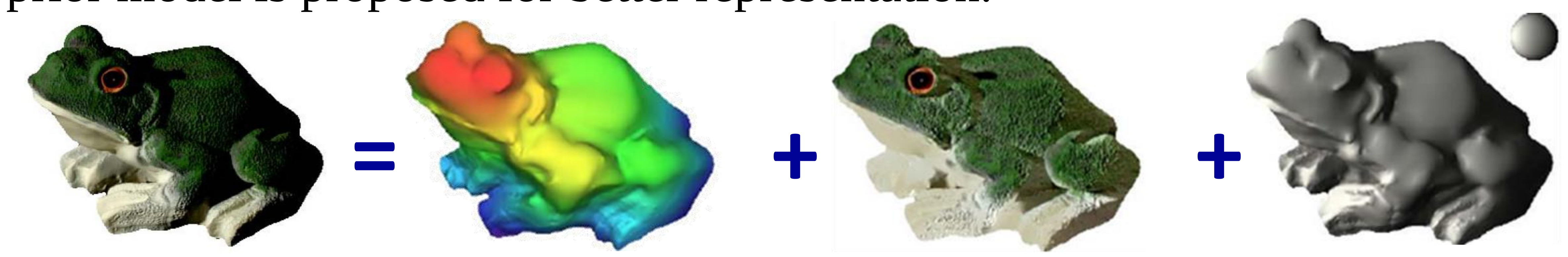
Bolun Cai[1] Xiangmin Xu[1] Kailing Guo[1] Kui Jia[1] Bin Hu[2] Dacheng Tao[3]

[1] School of Electronic and Information Engineering, SCUT, China
[2] Ubiquitous Awareness and Intelligent Solutions Lab, LZU, China
[3] UBTECH Sydney AI Centre, FEIT, USYD, Australia

https://caibolun.github.io/JieP

1. Abstract

We propose a joint intrinsic-extrinsic prior (***JieP***) model to estimate both illumination and reflectance from an observed image. The 2D image formed from 3D object in the scene is affected by the intrinsic properties (***shape*** and ***texture***) and the extrinsic property (***illumination***). Based on a novel structure-preserving measure called ***local variation deviation***, a joint intrinsic-extrinsic prior model is proposed for better representation.



$S = I \cdot R$ | **Shape** $E_s = \|\mathcal{R}_x\|_1 + \|\mathcal{R}_y\|_1$ | **Texture** $E_t = \|\nabla_x R\|_1 + \|\nabla_y R\|_1$ | **Illumination** $E_l = \|I - B\|_2$

2. Local Variation Deviation (LVD)

- **Edge-aware Filters** (*e.g.* BLF, RGF)
  Gibbs phenomenon of local filters result in ringing-effect near the edge.
- **Statistics-based Smoothing** (*e.g.* WMF, LEF)
  For high-frequency signals, local statistics still produce oscillating results.
- **Optimization-based Method** (*e.g.* WLS, RTV)
  They focus on relatively small variance and vulnerable to textures.

To address above problems, a ***local statistical*** magnitude called local variation deviation (LVD) is proposed by a ***global optimization*** function.

$$\mathcal{D}_{x/y} = \left| \nabla_{x/y} I - \frac{1}{|\Omega|} \sum_\Omega \nabla_{x/y} I \right|$$

to enhance the discrimination

$$\mathcal{R}_{x/y} = \left| \frac{\nabla_{x/y} I}{\frac{1}{|\Omega|} \sum_\Omega \nabla_{x/y} I + \epsilon} \right|$$

	A	B	C	D	E	F
$\overline{\mathcal{D}}$	0.073	0.079	0.043	0.013	0.007	0.013
$\overline{\mathcal{R}}$	3.672	3.971	2.168	0.572	0.336	0.637

LVD can be explained intuitively as:

- Case 1: **Flat.** If I is almost constant, $\nabla I \approx 0$ and $\overline{\nabla I} \approx 0$, so $\overline{\mathcal{D}} \approx 0$ and $\overline{\mathcal{R}} \approx 0$
- Case 2: **Texture.** If I changes frequently, ∇I fluctuates more rapidly than $\overline{\nabla I}$, so $\overline{\mathcal{D}} > 0$ and $\overline{\mathcal{R}} \approx 1$
- Case 3: **Structure.** If the patch I changes in accordance, the deviation of ∇I is vary small, so $\overline{\mathcal{D}} \approx 0$ and $\overline{\mathcal{R}} \approx 1$

3. Joint Intrinsic-Extrinsic Prior Model

- **Intrinsic Prior on Shape & Texture**

The prior on ***shape*** is motivated by that the object shape tends to be oriented isotopically in the scene.

$$E_s(I) = \left\| \frac{\nabla_x I}{\frac{1}{|\Omega|} \sum_\Omega \nabla_x I + \epsilon} \right\|_1 + \left\| \frac{\nabla_y I}{\frac{1}{|\Omega|} \sum_\Omega \nabla_y I + \epsilon} \right\|_1$$

Fig 1: The structure of illumination in the real-world

The reflectance contains fine ***texture*** and is piece-wise continuous, so the distribution of gradients is formulated with a Laplacian distribution.

$$E_t(R) = \|\nabla_x R\|_1 + \|\nabla_y R\|_1$$

- **Extrinsic Prior on Illumination**

The ***illumination*** prior is on the visual effect of inverted observed images $1 - S$, which is intuitively similar to haze images.

$$(1 - S) = 1 - I \cdot R = (1 - R) \cdot I + (1 - I)$$

$\xrightarrow{H = 1 - S, J = 1 - R}$ $H = J \cdot T + a(1 - T)$

$\xrightarrow[\text{Dark channel}]{T = I, \alpha = 1}$ $T = 1 - \min_\Omega \left(\min_{c \in \{r,g,b\}} \frac{H^c}{a} \right)$

$\xrightarrow{\text{Bright channel}}$ $B = \max_\Omega (\max_c S^c)$

$$E_l(I) = \|I - B\|_2^2$$

Fig 2: The inverted image of those shown in Fig. 1

- **Joint Optimization**

$$E(I, R) = \|I \cdot R - S\|_2^2 + \alpha E_s(I) + \beta E_t(R) + \lambda E_l(I)$$

Iteratively Re-weighted Least Square

$$\begin{cases} E_s(I) = u_x \|\nabla_x I\|_2^2 + u_y \|\nabla_y I\|_2^2 \\ E_t(R) = v_x \|\nabla_x R\|_2^2 + v_y \|\nabla_y R\|_2^2 \end{cases}$$

where

$$\begin{cases} u_{x/y} = \left(\left| \frac{1}{\Omega} \sum_\Omega \nabla_{x/y} I \right| \left| \nabla_{x/y} I \right| + \epsilon \right)^{-1} \\ v_{x/y} = \left(\left| \nabla_{x/y} R \right| + \epsilon \right)^{-1} \end{cases}$$

Block Coordinate Descent

$$\text{(P1)}\ I_k = \arg\min_I \|I \cdot R_{k-1} - S\|_2^2 + \alpha \left(u_x \|\nabla_x I\|_2^2 + u_y \|\nabla_y I\|_2^2 \right) + \lambda \|I - B\|_2^2$$

$$\text{(P2)}\ R_k = \arg\min_R \|I_k \cdot R - S\|_2^2 + \beta \left(v_x \|\nabla_x R\|_2^2 + v_y \|\nabla_y R\|_2^2 \right)$$

4. Experiments

Tab. 1: Quantitative performance comparison on 35 images with NIQE and ARSIM

Method	HE	BPDFHE	SSR	MSRCR	NPE	GOLW	MF	LIME	SRIE	WVM	Ours
NIQE	3.4475	3.7267	3.3778	3.4295	3.4091	3.3647	3.5335	3.6155	3.4590	3.3594	**3.3409**
ARISM	3.2902	3.3275	3.0469	3.1014	3.0891	3.3243	3.0200	3.1753	2.9930	2.9958	**2.9917**

Tab. 2: Comparison of color constancy with angle error on the Color-Checker Dataset

Method	White-Patch	Grey-World	Gray-Edge	Shades-Gray	Bayesian	CNNs	Gray-World	Grey-Pixel	Ours
Mean (°)	7.55	6.36	5.13	4.93	4.82	4.73	4.66	4.60	**4.32**