

# Intrinsic Image Decomposition using Focal Stacks

Saurabh Saini

saurabh.saini@research.iiit.ac.in, parikshit.sakurikar@research.iiit.ac.in, pjin@iiit.ac.in

Parikshit Sakurikar

P J Narayanan

Centre for Visual Information Technology, Kohli Center on Intelligent Systems, IIIT Hyderabad



Figure 1. Our method takes a focal stack and computes intrinsic images of the scene. Left column shows four sample images from a focal stack. Middle column shows the reflectance and right shows the shading component of the scene as estimated by our proposed method. Note how the shadows and specularities on the chess pieces are separated out in the shading component while reflectance component exhibits textures and uniform color values. (Best viewed in color).

## ABSTRACT

In this paper, we present a novel method (*RGBF-IID*) for intrinsic image decomposition of a wild scene without any restrictions on the complexity, illumination or scale of the image. We use focal stacks of the scene as input. A focal stack captures a scene at varying focal distances. Since focus depends on distance to the object, this representation has information beyond an RGB image towards an RGBD image with depth. We call our representation an *RGBF image* to highlight this. We use a robust focus measure and generalized random walk algorithm to compute dense probability maps across the stack. These maps are used to define sparse local and global pixel neighbourhoods, adhering to the structure of the underlying 3D scene. We use these neighbourhood correspondences with standard chromaticity assumptions as constraints in an optimization system. We present our results on both indoor and outdoor scenes using manually captured stacks of random objects under natural as well as artificial lighting conditions. We also test our system on a larger dataset of synthetically generated focal stacks from NYUV2 and MPI Sintel datasets and show competitive performance against current state-of-the-art IID methods that use RGBD images. Our method provides a strong evidence for the potential of RGBF modality in place of RGBD in computer vision.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from [permissions@acm.org](mailto:permissions@acm.org).

ICVGIP, December 18-22, 2016, Guwahati, India

© 2016 ACM. ISBN 978-1-4503-4753-2/16/12...\$15.00

DOI: <http://dx.doi.org/10.1145/3009977.3010046>

## CCS Concepts

- Computing methodologies → Computer vision problems; Computational photography;

## Keywords

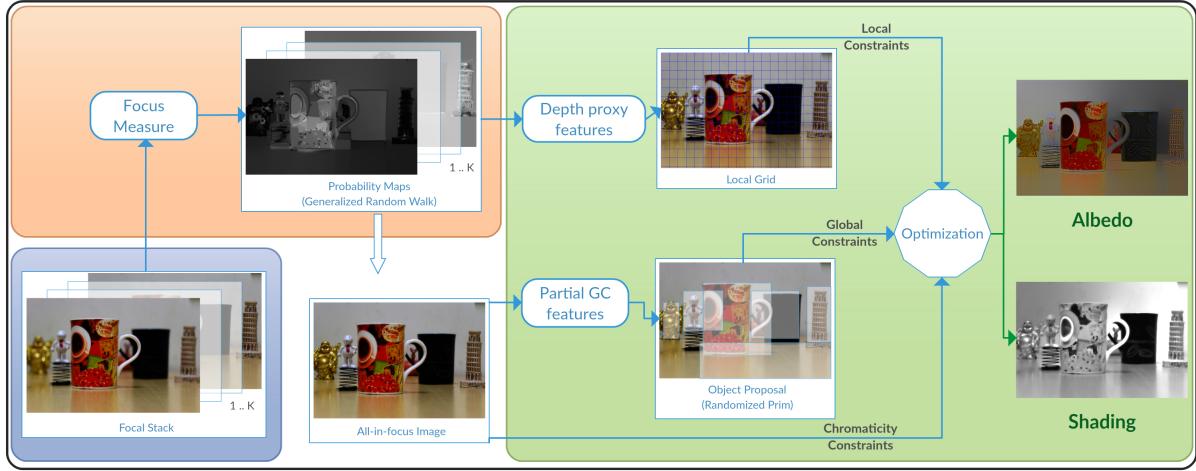
Intrinsic image decomposition; focal stacks; generalized random walk; RGBF, RGBF-IID

## 1. INTRODUCTION

Intrinsic Image Decomposition (IID) involves the decomposition of a given image into its constituent reflectance (albedo) image which is based on the material properties of the scene and the illumination (shading) image based on the lighting conditions. IID is useful in several other computer vision problems such as image colorization [21], shadow removal [18], image enhancement [11], image recoloring [33], image relighting [10] etc.

IID is an ill-posed problem [17]. Decomposing an image  $I$  of a natural scene into its intrinsic components as  $I = R \cdot S$ , where  $R$  represents reflectance and  $S$  represents shading, can have multiple possible valid combinations. For each pixel in image  $I$ , we need to estimate two variables which makes it an under constrained setting. Solutions to IID are thus proposed under various assumptions like color consistency, solution sparsity, lambertian reflection etc. Color consistency assumption can be traced back to the Retinex algorithm [19] which assumes that objects with similar chromaticity indicate similar reflectance. Yet another common assumption is to ignore specular reflections in the given scene. Alternatively several approaches therefore propose to use auxiliary scene information in the form of multiple views, multiple illumination images, scene depth etc. Our method also falls into this category.

We present a method to compute reflectance and shading images for a given scene using focal stacks as the only in-



**Figure 2: Block diagram of the proposed system which consists of three subsystems. Blue block indicates the acquisition stage, orange is focus measure based probability map computation and green block is for the IID optimization (See 3).**

put. We refer to our proposed method as ‘RGBF-IID’ for the rest of the paper. A focal stack is a set of images of the scene captured at varying focal distances. It is a relatively easier way to capture scene structure information compared to other methods which require careful camera positioning, specialized exposures at different lighting conditions or a specialized depth sensor. Capturing a focal stack is possible without any extensive setup up or training. Focal stack capture is mostly an automated process and is possible on devices ranging from high-end DLSR cameras to smartphone cameras. This makes the idea of focal stacks for IID attractive. Moreover, since we directly use images and not specialized sensors, our method can be used for a wide variety of both indoor and outdoor scenes under natural and artificial lighting conditions as we demonstrate later.

The amount of defocus of a given scene point in a focal slice depends on its depth from the camera. We use this observation in conjunction with a robust focus measure and design features to be used as a proxy for depth. We define local and global point neighbourhoods based on this RGBF representation where  $F$  refers to the focus based features. Specifically for RGBF-IID, we compute a dense probability map for each slice corresponding to the contribution of that slice towards the all-in-focus image. We assume that pixels with similar sharpness variation across the stack lie in a similar depth range. Our contributions can be summarized as given below :

- We present a novel way to obtain IID using focal stacks which are easy to capture and universally applicable.
- We show that an easy-to-compute RGBF representation can be used as a substitute for RGBD based solutions. This opens up several exciting research directions for other computer vision problems.
- We construct global constraints for IID using object localization for capturing image semantics and demonstrate results on several challenging, general indoor and outdoor scenes.

## 2. RELATED WORK

We briefly discuss the relevant literature for our proposed RGBF-IID method. We divide this discussion into two parts. In the first part we discuss literature related to focal stack generation and use in computer vision. The second part deals with various solutions for IID.

### 2.1 Focal stacks

Focal stacks have been used earlier in computer vision to compute the all-in-focus image of the scene [2, 23, 32]. This is especially useful for macro imaging where it is very difficult to capture an all-in-focus image in a single shot. Hasi-noff [15] shows that for any given depth range, generating an in-focus image from a focal stack is more time efficient than capturing a single narrow aperture image.

Different parts of the scene come into focus in different focal slices based on their 3D location in the scene. Measuring the amount of focus for each pixel across the stack can thus serve as a proxy for 3D scene information. Depth-from-focus using focal stacks has been attempted by [25] and [29]. Some Depth-from-defocus methods also make use of pairs of differently focused wide aperture images [8, 12].

For fusion of multifocus images we adapt the system proposed by Shen et al. [28] for multiexposure image fusion. They generate dense probability maps for each exposure slice indicating the probability of each pixel belonging to that slice. They use generalized random walks to obtain a high dynamic range output from several low dynamic range inputs. Their framework is useful for solving labeling problems where one desires candidate probabilities for each site.

### 2.2 Intrinsic Image Decomposition

We restrict our discussion here to automatic methods which require no user input. Such methods fall under two categories: those which work on a single image and those which use auxiliary scene data. IID methods which work on a single image require several assumptions as stated earlier. In Tappen et al. [30] image derivatives are used and assumed to be caused by either change in reflectance or shading. Then a generalized belief propagation based framework is used for fi-

nal labeling. Barron [3] uses new priors on reflectance based on sparsity but results are shown only on simple images. In Barron and Malik [4], more priors are introduced on smoothness, entropy, absolute color and illumination in a multiscale optimization framework and improved results are reported. In Bell et al. [6] a new manually annotated dataset is introduced and fully connected conditional random fields are iterated over for reflectance estimation followed by shading optimization. Rother et al. [26] assumes sparse basis for colors and estimates them as latent variables using a gaussian distribution based random field model. Garces et al. [13] uses k-means clustering to estimate clusters with reflectance discontinuities and employs a simple linear system for IID.

The second category of methods leverage extra information from auxiliary data to produce more constraints. Weiss [31] uses multiple images under different illumination conditions. They use natural scene statistics and a laplacian distribution model and optimize it using maximum-likelihood estimation. Liu et al. [21] estimate intrinsic components using multiple images of a given scene. They register the images using SIFT and RANSAC, apply median filtering over the overlapping segments and use the maximum-likelihood method of Weiss [31].

Most relevant approach to ours is by Jeon et al. [17] and Chen and Koltun [9]. Both these works use RGBD information of the scene to compute IID. Chen and Koltun [9] decompose shading into three components caused by direct irradiance, indirect irradiance and illumination color. Different constraints are applied to each component based on point neighbourhood assumptions obtained using the depth data. They pose the final optimization problem using a simple linear least-squares framework. Similar to [9], Jeon et al. [17] also work on RGBD images. They first separate image textures using smoothing filters and then define local and global shading constraints using depth based point neighbourhood estimates. They reformulate these constraints as a Local Linear Embedding (LLE) of sparse features which they propagate to rest of the image using separate propagation terms. We build on their mathematical framework and adapt it for our method. We present these details in sec 3.3.

### 3. METHOD

Figure 2 shows block diagram of our proposed RGBF-IID method. Our method can be divided into three main steps : input acquisition, probability map estimation and intrinsic image decomposition. We discuss these in the three subsequent subsections.

#### 3.1 Input Acquisition

As described earlier, focus stacking is a technique where multiple images of the scene are captured with changes in focus distance between consequent shots. Focus stacking is becoming increasingly easy to use on modern day cameras. Several methods exist to automatically capture a focal stack given a rough range of depths in the scene. MagicLantern [1] is a camera firmware addon for Canon DSLR cameras which enables several image stacking methods directly on these cameras. It allows for discrete control of the amount of focus-ring movement between shots and is a very useful tool for focus stacking. Focus stacking is also possible on mobile phones using either region based focusing such as in [27] or using direct focus distance control of the Android Camera2 API. We use MagicLantern to capture the real world focal

stacks shown in our experiments. The focus-ring movement between consequent images is tuned to capture the entire scene in an exhaustive and exclusive manner i.e. each pixel is in-focus in one and only one focal slice. We also show results on an automatically captured focal stack using the Android Camera2 API on a Nexus 5X device.

We also experiment with synthetic focal stacks created over the MPI Sintel [7] and the NYUV2 [24] datasets. We use the depth map available with each RGB image and first cluster the depth regions present in the scene. We then synthesize a focal stack by setting the focus distance to each of the cluster centers in a sequential manner. Since depth data is available, we use standard lens optics to derive the amount of defocus for each pixel in each focal slice based on the focus distance for the slice. We apply spatially varying defocus kernels in a manner similar to Barron et al. [5] to synthesize the focal slices (See supplementary material).

#### 3.2 Focus based Probability Maps

We compute the all-in-focus image from our focal stack  $F$  by adapting the method of Shen et al. [28] which uses generalized random walks for fusion of multiexposure images. They pose image fusion as a quadratic energy minimization problem and solve it using associated Dirichlet problems for each label. They propose a probabilistic model by assigning color and contrast based node potentials for the image-label graph. We use their notation and adapt their framework for multifocus fusion. Consider graph  $\mathcal{G} = (\mathcal{V}, \mathcal{E})$  where  $\mathcal{V} = \mathcal{L} \cup \mathcal{X}$  and  $\mathcal{E} = \mathcal{V} \times \mathcal{V}$ . Here  $\mathcal{L} = \{l_1, l_2, \dots, l_K\}$  is the set of all labels with  $l_j$  referring to focal slice  $F_j$ .  $\mathcal{X}$  is the set of variables  $x_i$  in the final fused image  $F^*$ . The fused image is composed as

$$F^*(x_i) = F_j(x_i) \mid j = \operatorname{argmax}_k(P^k(x_i)), \quad (1)$$

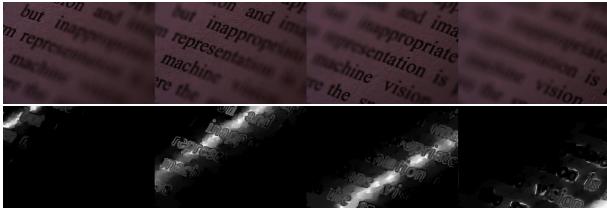
where  $P^k(x_i)$  is the  $k^{th}$  label assignment probability for  $x_i$ . Nodes  $x_i \in \mathcal{X}$  form a 4-connected grid with each node sharing an edge with all nodes in  $\mathcal{L}$ . Also  $\mathcal{E} = \mathcal{E}_{\mathcal{X}} \cup \mathcal{E}_{\mathcal{L}}$  where  $\mathcal{E}_{\mathcal{X}}$  denote edges between grid nodes and  $\mathcal{E}_{\mathcal{L}}$  denotes edges between grid and label nodes. We define compatibility functions  $y$  for  $\mathcal{E}_{\mathcal{L}}$  edges and  $w$  for  $\mathcal{E}_{\mathcal{X}}$  edges. This graph framework is similar to the one frequently used in systems which are posed as a Markov Random Field (MRF). Alternatively MRF based fusion methods could also be used.

$P^k(x_i)$  is proportional to the node potential of  $x_i$  which is obtained by minimizing the energy associated with graph network by solving  $K$  Dirichlet problems. We use parameter values and definitions as given in [28]. The main difference is that we define  $\mathcal{E}_{\mathcal{L}}$  edge compatibility functions as

$$y_k = \theta_k \left[ \operatorname{erf}\left(\frac{f^k}{\sigma}\right) \right]^K$$

where  $f^k$  is a per-pixel focus measure for  $k^{th}$  focal slice and is computed as the local variance in a  $11 \times 11$  patch around each pixel.  $\theta_k$  is the frequency of  $f^k$  and  $\sigma$  is standard deviation of  $f^k$  in  $k^{th}$  focal slice. We define pairwise functions  $w_k$  in a manner equivalent to [28] and solve the Dirichlet equations using generalized random walks.

We then concatenate the obtained probability maps  $P^k$  with normalized image coordinate values  $\bar{x} = \frac{x}{m}, \bar{y} = \frac{y}{n}$  where  $m, n$  are image height and width respectively. The probability maps encode per-pixel focus variations across the stack and the normalized co-ordinates encode spatial local-



**Figure 3:** Top row shows a few images from a sample focal stack with associated probability maps in the bottom row.

ity. Thus our concatenated feature  $d_{xy}$  efficiently captures local neighborhood information required by the IID framework. The major advantage of using GRW instead of MRF here is that this fusion method gives continuous probability maps for each pixel across the stack unlike discrete labels in MRF. We show sample probability maps and the in-focus image in Figure 3

### 3.3 Estimating Reflectance & Shading

We derive inspiration from depth based IID methods [9, 17] for our IID subsystem. These methods use depth and normals obtained from depth to define local and global point neighbourhoods which are then used as constraints in the optimization equation. We use our concatenated proxy-depth features defined in the previous section, to capture neighbourhood information. We build upon the framework of [17] which shows an improvement over [9] and is based on the separation of texture from the image. We use and extend their notation to suit our approach. We briefly discuss their system below and highlight the relevant extensions and modifications.

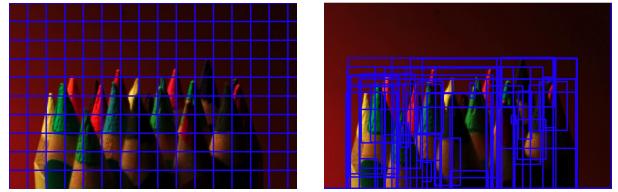
Let  $I = B \cdot T$  where  $T$  is the texture and  $B$  is the textureless base image.  $B$  can be decomposed further as  $B = R_B \cdot S_B$ . Estimating the reflectance  $R_B$  and the shading  $S_B$  involves the minimization of the following energy function

$$f(S_B, R_B) = f^F(S_B) + \lambda^P f^P(S_B) + \lambda^R f^R(R_B)$$

where right hand side terms refer to proxy-depth based energy term, propagation term and reflectance term respectively. The  $\lambda$  parameters refer to the relative weights between the terms. The first term  $f^F(S_B)$  refers to shading constraints such that

$$f^F(S_B) = f_l^F(S_B) + \lambda_g^F f_g^F(S_B)$$

where  $f_l^F(S_B)$  is for local and  $f_g^F(S_B)$  is for global shading consistency constraints. In [17], both these terms are reinterpreted using Local Linear Embedding (LLE) to define local  $\mathcal{N}_l$  and global  $\mathcal{N}_g$  neighbourhoods using depth based features. Similar to their approach, for RGBF-IID we define  $\mathcal{N}_l$  by computing  $\kappa$  nearest neighbours in LLE space using concatenated  $d_{xy}$  features. The underlying intuition is that scene points with similar variation in focus across the stack have a higher probability of lying in the same depth range. Hence, combined probabilities derived from the focus measure and normalized image coordinates will encode local structural properties. For ease of computation we divide the image into rigid  $11 \times 11$  grids and compute LLE between these grids by choosing the feature with has minimum variance as a representative feature for that entire grid. The global shading term  $f_g^F(S_B)$  aims to capture global consis-



**Figure 4:** Left : Non overlapping rigid grids for local constraints. Right: Overlapping flexible sized grids for global constraints.

tency in spite of occlusion and color changes as an object can have multiple colors but shading component should be based on its shape and not texture or color. To enforce global consistency, we capture scene and object semantics by using partial Geometric Context (GC) features [16]. We use color and texture terms from GC features which capture statistics at smaller superpixel scale. In order to use these features for global semantics we collect prime object proposals based on the method of Manen et al. [22]. This takes image superpixel statistics into account and gives window proposals which have higher probability of containing an object. We use these proposals and compute candidate GC features per window based on minimum variance. We use these candidate features to compute LLE based global constraints. This is illustrated in Figure 4.

The second term  $f^P(S_B)$  is used to propagate sparse grid based information computed in the previous step to rest of the pixels. This term comprises of two smoothing factors :

$$f^P(S_B) = f_{lap}^P(S_B) + \lambda_F^P f_F^P(S_B)$$

where  $f_{lap}^P(S_B)$  is the matting Laplacian [20] as defined in [17]. In our approach, we define  $f_F^P(S_B)$  as a smoothing term with the local propagation weights derived from the concatenated  $d_{xy}$  features rather than surface normals as in [17]. The weights are defined as:

$$w_{pq}^F = \exp \left( -\frac{1 - \langle d_{xy}(p), d_{xy}(q) \rangle^2}{\sigma_n^2} \right)$$

The third term corresponding to the reflectance constraint is used directly as described in [17] and is defined as

$$f^R(R_B) = \sum_{\mathcal{P}} \sum_{\mathcal{N}_p} w_{pq}^R (R_B(p) - R_B(q))^2$$

over the set of all image pixels  $\mathcal{P}$  and current pixel neighbourhood  $\mathcal{N}_p$ . Weights are computed based on the angular dissimilarity in the chromaticity between the pixels. We discuss these terms in the supplementary material and refer any interested reader to [17] for further details.

## 4. EXPERIMENTS AND RESULTS

In this section we discuss the three datasets used and the experiments performed. Most of the datasets in this problem domain consist of well lit simple objects and do not capture real world scene complexity and variety. Owing to the lack of any available dataset for IID using focal stacks we manually collected some data for a few scenes. Apart from this we also show qualitative results using NYUv2 and MPI Sintel.

We compare our results with depth based IID frameworks of [17] and [9]. We show that our focal stack based IID

Method	MSE			LMSE			DSSIM		
	reflectance	shading	average	reflectance	shading	average	reflectance	shading	average
RGBF-IID (Ours)	0.0298	0.0679	0.0488	0.0191	0.0390	0.0291	0.2771	0.2762	0.2766
RGBD-IID-Jeon [17]	0.0263	0.0527	0.0395	0.0151	0.0311	0.0231	0.2513	0.2642	0.2577
RGBD-IID-Chen [9]	0.0307	0.0277	0.0292	0.0185	0.0190	0.0188	0.1960	0.1650	0.1810

**Table 1: Quantitative results on MPI Sintel dataset.** Contrary to state-of-the-art RGBD IID methods our proposed method is able to give comparable performance even in the absence of complete depth data with only a few focal stacks.

framework achieves comparable results with these two methods even in the absence of complete depth information. For the experiment with the focal stack we show only our results as the other two methods can not be executed without depth information. We show qualitative comparisons for the other two experiments. For quantitative comparison we use synthetic focal stacks for our method and the provided depth data for the other two methods. We report results using three metrics. Mean Squared Error (MSE) and Local Mean Squared Error (LMSE) [14] and Dissimilarity Structural Similarity Index Metric (DSSIM) as defined by [9].

#### 4.1 Real world focal stacks

We run our first set of experiments on the manually collected focal stacks of the real word scenes captured as discussed in 3.1. All these images are of varying sizes and vary in the number of the stack images captured. Also there is no restriction on the type of scene (indoor and outdoor) or the illumination condition (artificial or natural). We captured approximately 30 focal stacks and we show results on a few of those images in Figure 5. Notice in ‘classroom’ scene how the well lit parts of the scene especially specular regions due to indoor lights, natural lighting near the window, reflections on the floor and on the desk separates into shading image while the reflectance image captures scene albedo as smooth color values. The reflectance image for ‘desk’ scene captures the color of pisa and buddha figurines correctly after removing the dark regions and the lighting reflections respectively in these two cases. In ‘leaves’ even the dark regions under shadow have green reflectance. For the ‘bench’ scene uniform color of the grass blades and background bushes can be seen in reflectance image. In the ‘outdoor’ scene the shadow of the tree is absent in the reflectance image and is captured in shading. One possible failure case is noticeable here as the white flowers on the bush get captured in the shading image owing to being devoid of color and being smaller in size. This also leads to bush being brightly lit while the corresponding color in the reflectance image becomes darker shade of green. We believe this happens because the shading constraints were sparsely computed. In the ‘flower’ image the color of leaves shadowed by the flower petals is visible in the reflectance image.

#### 4.2 NYU dataset

For comparison with the RGBD based IID methods we show results on the artificially generated focal stacks from the testsplit of NYUv2 dataset [24], which is 654 RGB images with  $480 \times 640$  resolution and depth captured using Kinect. We use the depth only to generate the focal stacks for input in our framework. For Jeon et al. [17] (RGBD-IID-Jeon) and Chen and Koltun [9] (RGBD-IID-Chen) we

use the entire data and run using the code provided by authors with the default parameters. Also as [9] divide image into four components as albedo, direct irradiance, indirect irradiance and illumination color. For comparison we combine the direct, indirect irradiance and illumination color channels to generate the shading and we combine albedo and illumination color to generate reflectance image. Our results are comparable to both the methods without using the depth values. This shows that our depth proxy features are strong enough to capture the scene structure information using only a focus measure. The results are shown in Figure 6.

#### 4.3 MPI Sintel dataset

We also show results on Sintel dataset [7]. It is composed of 24 single camera views from a naturalistic open source animated movie. 50 frames for each scene are provided with ground truth depth and albedo in both ‘clean’ pass without particles added and ‘final’ pass with particles. We use the ‘clean’ pass to construct our focal stacks and show the output in Figure 7. As reported by [17], MPI Sintel is not created for IID benchmarking but in the absence of any focal stack based IID benchmarking dataset, we report quantitative results using this dataset. For [17] we run the code provided by the authors on their site with default parameters. The results are obtained on a every fifth frame of MPI Sintel Dataset and the final subset is selected for which both the methods converged (89 images). We use the numbers as reported by [9] in their paper. We can see that our method achieves comparable performance with [17] even though being substantially low in terms of the input information. For [9] their numbers are reported on a larger test subset.

### 5 CONCLUSION

In summary we present a novel method (RGBF-IID) to obtain intrinsic images of a given scene by capturing a set of images at varying focal distances. Compared to RGBD based methods focal stacks are easy to capture and can work on outdoor scenes also. This allows our system to be easily used without any restrictions on scene type. Also in our intrinsic decomposition formulation we capture scene semantics by generating novel object localization based global constraints. Additionally we also provide support to the hypothesis that focus based RGBF modality can be used in place of RGBD for some computer vision applications. In future we would like to explore this idea further and test it on other related problems.

### 6 ACKNOWLEDGEMENT

We thank TCS Research Scholar Fellowship Program for supporting Saurabh Saini and Parikshit Sakurikar.

## References

- [1] Magic lantern. <http://magicleantern.fm/>.
- [2] Aseem Agarwala, Mira Dontcheva, Maneesh Agrawala, Steven Drucker, Alex Colburn, Brian Curless, David Salesin, and Michael Cohen. Interactive digital photomontage. In *ACM Transactions on Graphics*, 2004.
- [3] Jonathan T. Barron. Shape, albedo, and illumination from a single image of an unknown object. In *CVPR*, 2012.
- [4] Jonathan T. Barron and Jitendra Malik. Color constancy, intrinsic images, and shape estimation. In *ECCV*, 2012.
- [5] Jonathan T. Barron, Andrew Adams, YiChang Shih, and Carlos Hernández. Fast bilateral-space stereo for synthetic defocus. In *CVPR*, 2015.
- [6] Sean Bell, Kavita Bala, and Noah Snavely. Intrinsic images in the wild. *ACM Transactions on Graphics*, 2014.
- [7] D. J. Butler, J. Wulff, G. B. Stanley, and M. J. Black. A naturalistic open source movie for optical flow evaluation. In *ECCV*, 2012.
- [8] Subhasis Chaudhuri and Ambasamudram N Rajagopalan. *Depth from defocus: a real aperture imaging approach*. Springer Science & Business Media, 2012.
- [9] Q. Chen and V. Koltun. A simple model for intrinsic image decomposition with depth cues. In *ICCV*, 2013.
- [10] Sylvain Duchêne, Clement Riant, Gaurav Chaurasia, Jorge Lopez Moreno, Pierre-Yves Laffont, Stefan Popov, Adrien Bousseau, and George Drettakis. Multi-view intrinsic images of outdoors scenes with an application to relighting. *ACM Transactions on Graphics*, 2015.
- [11] Elmar Eisemann and Frédo Durand. Flash photography enhancement via intrinsic relighting. *ACM Transactions on Graphics*, 2004.
- [12] P. Favaro and S. Soatto. A geometric approach to shape from defocus. *IEEE TPAMI*, 2005.
- [13] Elena Garces, Adolfo Munoz, Jorge Lopez-Moreno, and Diego Gutierrez. Intrinsic images by clustering. *Computer Graphics Forum (Proc. EGSR 2012)*.
- [14] Roger Grosse, Micah K. Johnson, Edward H. Adelson, and William T. Freeman. Ground-truth dataset and baseline evaluations for intrinsic image algorithms. In *ICCV*, 2009.
- [15] Samuel W Hasinoff and Kiriakos N Kutulakos. Light-efficient photography. *IEEE TPAMI*, 2011.
- [16] D. Hoiem, A. A. Efros, and M. Hebert. Geometric context from a single image. In *ICCV*, 2005.
- [17] Junho Jeon, Sunghyun Cho, Xin Tong, and Seungyong Lee. Intrinsic image decomposition using structure-texture separation and surface normals. In *ECCV*, 2014.
- [18] Vivek Kwatra, Mei Han, and Shengyang Dai. Shadow removal for aerial imagery by information theoretic intrinsic image analysis. In *ICCP*, 2012.
- [19] Edwin H. Land and John J. McCann. Lightness and retinex theory. *J. Opt. Soc. Am.*, 1971.
- [20] A. Levin, D. Lischinski, and Y. Weiss. A closed-form solution to natural image matting. *IEEE TPAMI*, 2008.
- [21] Xiaopei Liu, Liang Wan, Yingge Qu, Tien-Tsin Wong, Stephen Lin, Chi-Sing Leung, and Pheng-Ann Heng. Intrinsic colorization. *ACM Transactions on Graphics (SIGGRAPH Asia 2008 issue)*.
- [22] Santiago Manen, Matthieu Guillaumin, and Luc Van Gool. Prime object proposals with randomized prim's algorithm. In *ICCV*, 2013.
- [23] Hajime Nagahara, Sujit Kothiyurmal, Changyin Zhou, and Shree K Nayar. Flexible depth of field photography. In *ECCV*. 2008.
- [24] Pushmeet Kohli Nathan Silberman, Derek Hoiem and Rob Fergus. Indoor segmentation and support inference from rgbd images. In *ECCV*, 2012.
- [25] S. K. Nayar and Y. Nakagawa. Shape from focus. *IEEE TPAMI*, 1994.
- [26] Carsten Rother, Martin Kiefel, Lumin Zhang, Bernhard Schölkopf, and Peter V Gehler. Recovering intrinsic images with a global sparsity prior on reflectance. In *NIPS*.
- [27] Parikshit Sakurikar and PJ Narayanan. Dense view interpolation on mobile devices using focal stacks. In *CVPRW*, 2014.
- [28] R. Shen, I. Cheng, J. Shi, and A. Basu. Generalized random walks for fusion of multi-exposure images. *IEEE TIP*, 2011.
- [29] Supasorn Suwajanakorn, Carlos Hernandez, and Steven M. Seitz. Depth from focus with your mobile phone. In *CVPR*, 2015.
- [30] M. F. Tappen, W. T. Freeman, and E. H. Adelson. Recovering intrinsic images from a single image. *IEEE TPAMI*, 2005.
- [31] Y. Weiss. Deriving intrinsic images from image sequences. In *ICCV*, 2001.
- [32] Ning Xu, Kar-Han Tan, Himanshu Arora, and Narendra Ahuja. Generating omnifocus images using graph cuts and a new focus measure. In *ICPR*, 2004.
- [33] Genzhi Ye, Elena Garces, Yebin Liu, Qionghai Dai, and Diego Gutierrez. Intrinsic video and applications. *ACM Transactions on Graphics*, 2014.

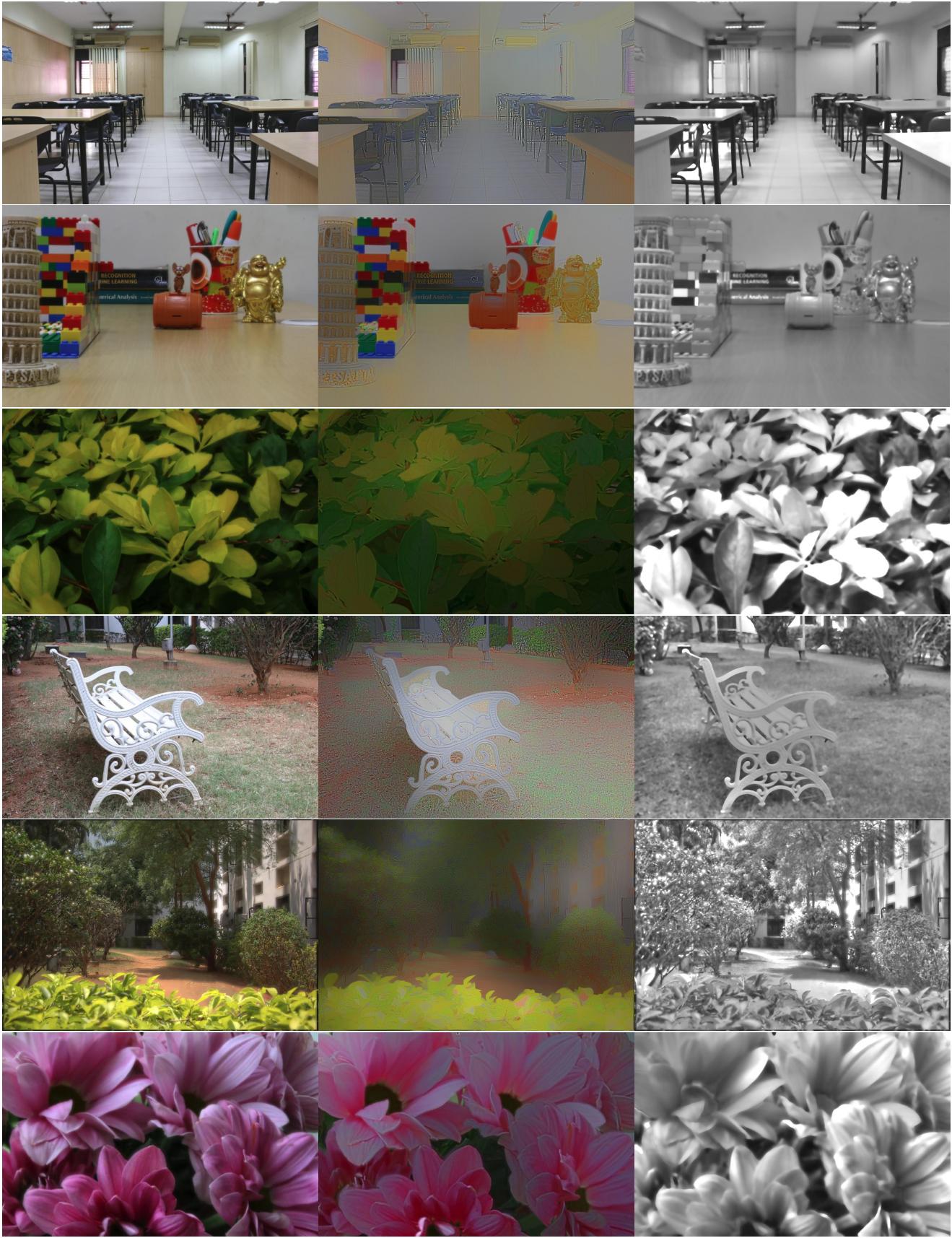


Figure 5: First row is the computed all-in-focus image followed by reflectance and then shading component obtained by our RGBF-IID method. Scenes names (top to bottom) : ‘classroom’, ‘desk’, ‘leaves’, ‘bench’, ‘outdoor’, ‘flowers’. The number of focal stack images used per scene respectively were : 12, 10, 11, 9, 11 and 15.

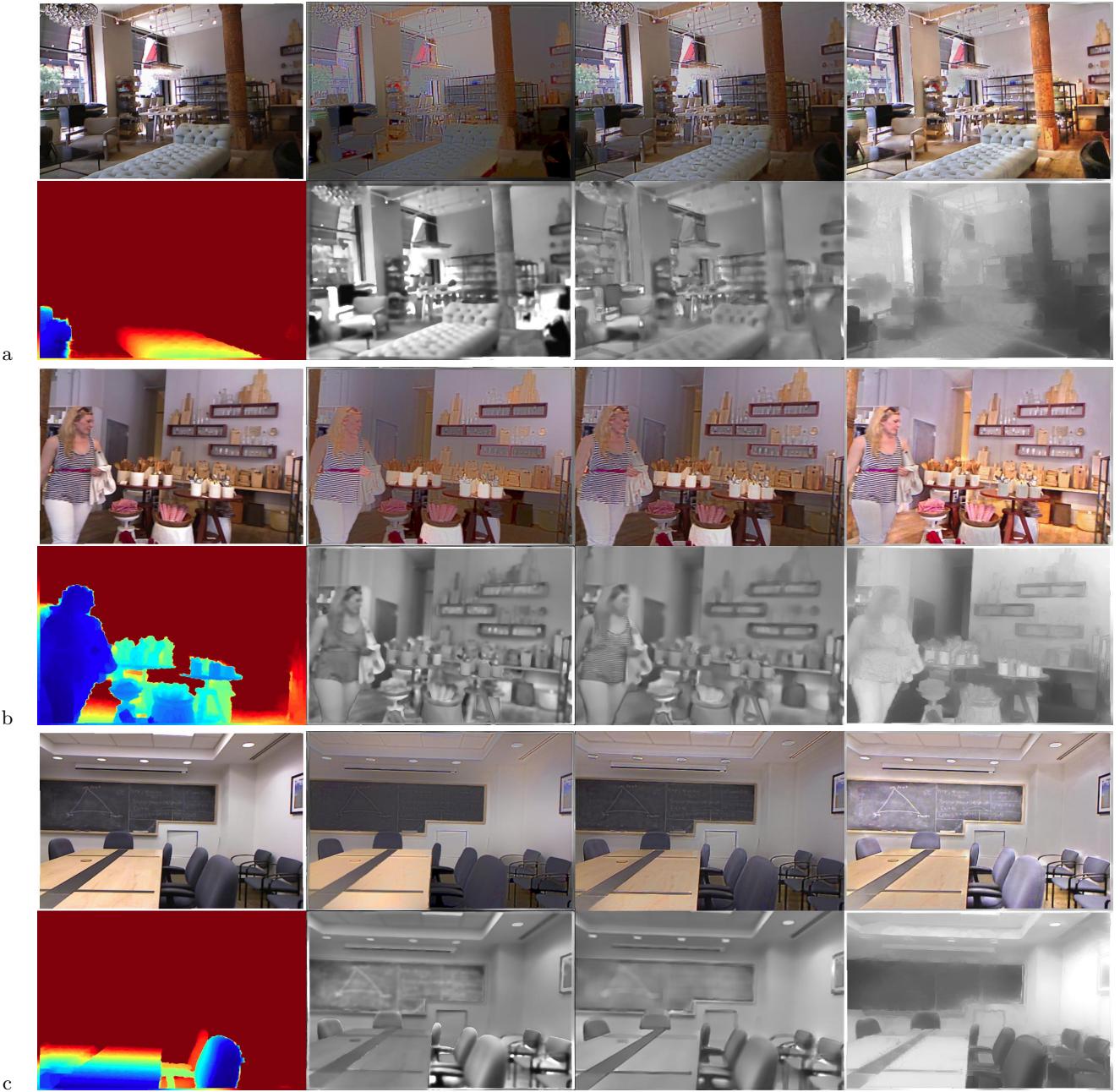


Figure 6: Our RGBF-IID results on NYU synthetic focal stacks. For each scene top row is original image, RGBF-IID reflectance, RGBD-IID-Jeon [17] reflectance and RGBD-IID-Chen [9] reflectance. Bottom row is depth(blue nearer) and corresponding shading results.



Figure 7: RGBF-IID results on MPI Sintel dataset. For each scene the all-in-focus image, RGBF-IID reflectance and shading are shown. More results included in the supplementary material.