Constrained Up-Scaling for Direct and Global Image Components

Julian Bader Martin Pätzold Andreas Kolb
Computer Graphics Group, Institute for Vision and Graphics, University of Siegen
Hölderlinstraße 3, 57076 Siegen, Germany
{ julian.bader,martin.paetzold,andreas.kolb }@uni-siegen.de

Abstract

The separation of direct and global illumination components is interesting for many applications in Computer Graphics and Computer Vision, such as BRDF estimation or material classification. However, for full-resolution images, a large number of coded images have to be acquired. For many interactive applications, such as the acquisition of dynamic scenes or video capturing, this is not feasible. In this paper, a new constrained up-scaling technique for separated direct and global illumination images is proposed which requires two to three coded input images, only. Our approach imposes the boundary condition that the sum of the direct and global components equals the fully illuminated image. We work in a predictive-corrective manner where we first use a single-image up-scaling method in order to predict the higher resolution images. Afterwards, the missing higher frequencies are determined using a fully illuminated image. As the distribution of the higher frequencies differs among the various frequency bands, we apply our approach in an iterative way for small up-scaling steps distributing the missing information by minimizing the overall frequencies. We evaluate the up-scaling scheme and demonstrate the improvement compared to single-image approached. As our method aims at minimizing the structured light patterns needed for acquisition, we additionally discuss the performance of existing pattern sets in terms of applicability for dynamic scenes.

Keywords

Computational Photography, Image Processing, Global Illumination

1 INTRODUCTION

The separation of direct and indirect illumination in a scene is an interesting task for both, Computer Graphics and Computer Vision domains. The direct component helps obtaining the distribution of reflected light in a scene. This can be used to derive a model for the bidirectional reflection distribution function (BRDF) [GKGN11]. Also, 3D reconstruction methods are more robust when working only on the direct component [GKGN11, NG12]. The global component gives insight into the scattering behaviour between one or more objects in a scene. This is useful to achieve more photo-realistic renderings of a scene by observing the complex light flow [MYR10]. Also, object recognition benefits from such approaches as the scattering of an object highly depends on the material [GL12]. In 2006, Nayar et al. [NKGR06] presented an approach for separating the direct and global illumination using a structured-light approach. However, for a high resolution separation, a large number of structured-light patterns is needed. Thus, this approach is not feasible for dynamic scenes, such as capturing biometric information of human faces or mobile navigation, or for video capturing. In their work, an alternative method was proposed where only one structured light pattern is sufficient. The disadvantage of this method is that the image resolution decreases dramatically. As a consequence, small scale details of fine textured regions are lost.

In this paper, we propose a new constrained image up-scaling technique for direct and global component images. We employ an iterative prediction-correction approach which uses a single-image up-scaling method in order to predict the higher resolution representation for each component. These up-scaled images are then corrected according to a high resolution fully illuminated image which has been additionally acquired. Fine structures in images, such as textiles or fur, can be recovered although they are not apparent in the low resolution input image. Our contributions are the following:

- A general constrained up-scaling scheme for the transfer of high frequency information.
- The application of the proposed scheme for direct and global component images.

- An evaluation of various low-resolution separation approaches which shows that the resulting image quality is not sufficient for the proposed up-scaling scheme.

To the best of our knowledge, we are the first proposing a constrained up-scaling approach for direct and global component images.

The rest of the paper is structured in the following way: Sec. 2 gives an overview of existing single-image and multi-modal up-scaling techniques. In Sec. 3, the principles of the separation approach of [NKGR06] is reviewed and the impact of the different structured light patterns is discussed. Our constrained up-scaling method is described in Sec. 4. In Sec. 5, the results are discussed. Furthermore, the improvement of our approach compared to the single-image prediction is shown. Finally, Sec. 6 concludes the paper.

2 RELATED WORK

In the past, the problem of image up-scaling has been addressed many times during the past 20 years. Here, a brief overview of the major approaches is given whereas we distinguish between the single-image and multi-modal approaches.

Single-image methods. The single-image up-scaling methods can be divided in to four groups: region-based interpolation, edge-directed interpolation, edge-directed reconstruction and example-based reconstruction.

Region-based interpolation methods apply different interpolation algorithms depending on the region in the image. Thurnhofer and Mitra [TM96] proposed an image interpolation approach which classifies image pixels into three different regions (constant, irregular and oriented). Other methods distinguish only between two regions (oriented, homogeneous) for different interpolation scheme application [ABA01, CHL05]. Zi et al. [ZDLL12] proposed a perceptual motivated segmentation of regions. For the general and transition regions, different kinds of interpolation algorithms are used. The attention region is computed according to an energy formulation optimizing curvature continuity, curvature enhancement and isolevel curve smoothing.

Edge-directed interpolation schemes adjust the interpolation according to the edges present in the image. Li and Orchard [LO01] use the local covariance for an edge-directed interpolation. Alternatively, sharp luminance variations and discontinuities are considered [BGS02]. Su and Willis [SW04] proposed a pixel triangulation scheme in order to use only pixels for interpolation not separated by an edge. Other methods propose the use of the second order derivative in order to determine the interpolation direction [GA08, GA11] and iterative refinement.

Edge-directed reconstruction algorithms try to estimate a high-resolution edge image which is then used to reconstruct the up-scaled image. Tai et al. [TTT06] proposed to use tensor voting for high resolution edge reconstruction. In alternative approaches, edge statistics are learned in order to predict enhanced edge images [Fat07, SXS08, YXYC12]. Another kind of approach [YXXY12, WXM+13] computes sharp high resolution gradients directly from the low resolution images.

Example-based approaches try to build up a dictionary which relate low resolution image patches to missing higher frequencies [YWL+12]. Freeman et al. propose a nearest neighbour search in the training set to find the high frequencies for image patches [FJP02]. Kim and Kwon [KK08] formulated the relationship between low and high frequencies as a linear regression model in order to derive a suitable image patch. Alternatively, a single-hidden-layer feed-forward neural network is used for learning the relation between low and high images frequencies [AB12]. A variety of methods are based on self similarity [GADTH97, EV07, SSU08, GBI09, FF11]. Here, the assumption is that an image contains similar structures at different scale levels.

Multi-modal methods. In [LLK08] and [STDT08], methods are proposed for the combination of low resolution time-of-flight depths images with high resolution RGB images. However, instead of using the RGB information for a constrained up-scaling, both approaches first do a single-image up-scaling of the depth image before fusing with RGB data. Langmann et al. [LHL11, LHL12] describe a fusion approach of time-of-flight and RGB data where several cross bilateral filtering strategies are compared. Additionally, a framework for joint segmentation and super-resolution is proposed. They assume the smooth change of the depth field in homogeneous regions. Fine structures, such as in fur or textiles, can therefore not be recovered.

Direct-Global Separation methods. In 2006, Nayar et al. [NKGR06] proposed a method for separating the direct and global image components using high-frequency structured-light patterns. Gu et al. [GKGN11] extended this approach in order to use multiple light sources. In 2012, two approaches were presented which estimate the light transport in a scene [ORK12, RRC12]. Both are also capable of direct-global separation. However, O'Toole et al. use a complicated setup and long acquisition times. The method of Reddy et al. needs a high number of structured-light patterns. As we focus on the acquisition of dynamic scenes, the latter approaches are not suitable.

The aim of our approach is the reconstruction of small detail which get lost during single-coded-image separation process. As the above mentioned single-image up-scaling methods can handle singularities, such as

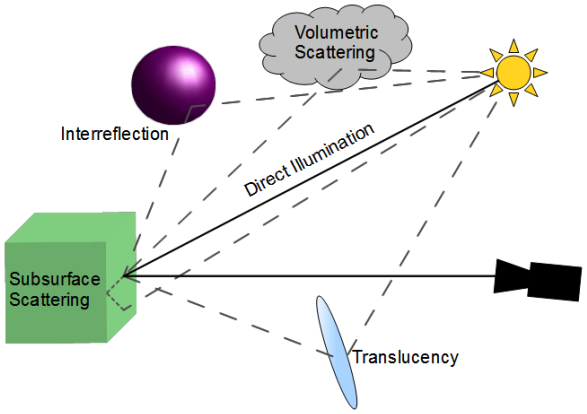


Figure 1: Light can reach the camera in different ways. Light being directly reflected to the camera is called *direct illumination*. *Global illumination* denotes light that is scattered in the scene and travels an indirect path to the camera.

edges or joints, quite well, fine homogeneous structures, such as fur or threads in textiles, cannot be reconstructed. Due to the smoothness assumption of depth data, the multi-modal super-resolution methods for fusion of time-of-flight data also cannot cover these kind of fine details. In contrast, our proposed method uses a single-image up-scaling prediction for the direct and global component images and enforces the sum of these two to be equal to a fully illuminated image.

3 SEPARATION OF DIRECT AND GLOBAL ILLUMINATION

The illumination in a scene can be decomposed into two parts. The first part is the so called direct illumination. This covers all the light which is emitted by a light source and is reflected directly to the scene observer (usually a camera or the human eye). The second part addresses all the light which is first scattered or otherwise redirected and reaches the observer in an indirect path. This is considered as global illumination (cf. Fig. 1).

3.1 Separation Scheme

In 2006, Nayar et. al [NKGR06] proposed a method to separate the direct and global illumination in a scene using structured light. Here, the underlying assumption is that the scattering of light, which is responsible for the global illumination effects, acts as a kind of low-pass filter, i.e. there are no high frequencies present in the global component. Thus, by projecting different high frequency patterns onto the scene, the following relations can be formulated depending on whether a part of the scene is hit by a light or dark source element:

$$L^+(\mathbf{i}) = L_d(\mathbf{i}) + \alpha L_g(\mathbf{i}) + b(1-\alpha)L_g(\mathbf{i}), \quad (1)$$

$$L^-(\mathbf{i}) = bL_d(\mathbf{i}) + (1-\alpha)L_g(\mathbf{i}) + \alpha bL_g(\mathbf{i}). \quad (2)$$

Here $L^+(\mathbf{i})$ refers to the intensity value of a scene element at position $\mathbf{i} = (x,y)^T$ which is directly illuminated, whereas $L^-(\mathbf{i})$ represent the intensity when the element is not directly illuminated. L_d and L_g refer to the direct and global part of the illumination and α is the fraction of projector source elements which are lit. Since digital projectors cannot be tuned completely dark, b accounts for the fraction of a white source element emitted by a "black" projected pixel with $0 \leq b \leq 1$. For a more detailed description of the separation method, the reader is referred to the paper of Nayar et al. [NKGR06].

3.2 Structured-Light Patterns

Nayar et al. [NKGR06] proposed several high-frequency structured-light patterns in order to acquire $L^+(\mathbf{i})$ and $L^-(\mathbf{i})$ for a scene.

Checker-board. The scene is illuminated by a sequence of shifted checker-board patterns. For each pixel in a scene the brightest and the darkest intensity value among all the acquisitions is chosen. With this kind of patterns, one can compute the direct and global component directly at the camera's resolution. However, due to the acquisition of multiple source images, this method is not feasible for the fast acquisition of scenes.

Phase shifted sinusoidal. Each pixel of the pattern can be expressed as $p_0 = 0.5(1+\sin\phi)$ for $\phi \in [0, 2\pi]$. Analogue to the first pattern, two more are generated where ϕ is phase shifted by $\frac{2\pi}{3}$ and $\frac{4\pi}{3}$. In general, the pixels of the first pattern can be chosen arbitrarily, but in order to ensure high frequency, a sinusoidal function which varies in both, x- and y-direction, should be chosen. Using these patterns, only three acquisitions are necessary. However, Gu et al. [GKGN11]stated that this method suffers from serious image artefacts. Therefore, it is not suitable if the capturing of fine details is required.

Stripes. By using a horizontal or vertical stripe pattern, only one acquisition is necessary. Here, the brightest and darkest pixels are determined in small rectangular patches which are aligned perpendicular to the stripes. These extrema are interpolated and averaged in order to generate the minimum $L^+(\mathbf{i})$ and maximum $L^-(\mathbf{i})$. Then, the direct and global components are computed. As only one pattern needs to be acquired, this method is suitable for capturing dynamic scenes. However, due to the averaging process, the resolution of the resulting images is decreased by a fraction of four [NKGR06].

4 A CONSTRAINED UP-SCALING METHOD

The aim of our approach is the constrained up-scaling of direct and local component images according to one

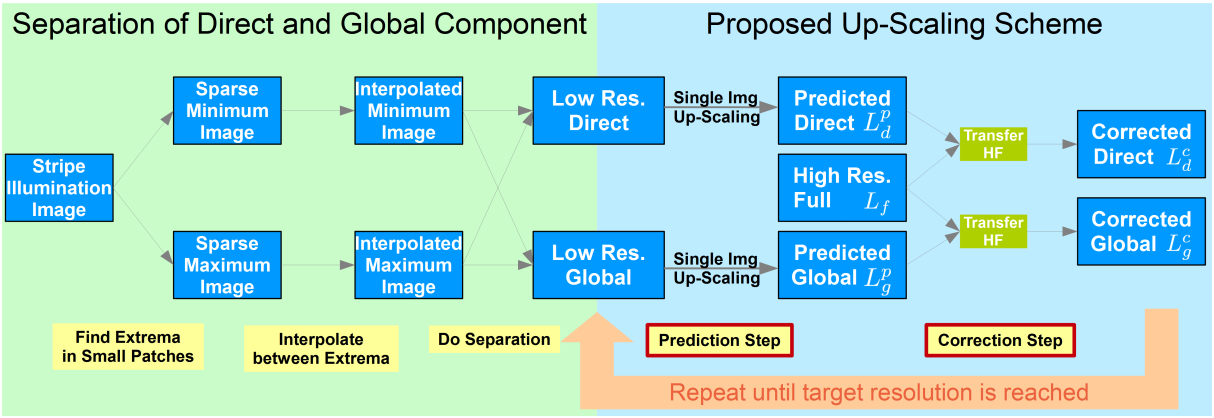


Figure 2: Overview of the complete system. The left side (green) shows the separation of global and direct components according to [NKGR06]. The right side (blue) describes our constrained up-scaling technique.

high-resolution fully illuminated image whereas we focus on the reconstruction on fine structures in homogeneous regions. Before the details are discussed in the following two sections, a rough system overview is given.

As input images, we take the two low resolution global and direct component images retrieved by the stripe pattern separation approach proposed by Nayar et al. [NKGR06]. Additionally we use a high resolution image which is acquired with full illumination. The proposed up-scaling scheme consists of two steps which are applied iteratively until we reach the target resolution (cf. Fig. 2):

1. In the **prediction** step, we estimate the higher resolution image using a single-image up-scaling technique. The first iteration takes the low-resolution direct and global component images as input. For all following iterations, the corrected images from the previous iteration are used.

2. The **correction** step corrects the predicted images according to the main condition that the sum of direct and global images results in the fully illuminated image. In order to decrease the noise level in the up-scaled image, the error to the full image is distributed in such a way that the overall intensity of higher frequencies is minimized.

4.1 Prediction

In general, there are no limitations regarding the prediction step in our up-scaling scheme. As the correction works on the predicted image directly, every single-image up-scaling method could be used. In our experiments, we used the so-called Local Self-Examples approach proposed by Freedman and Fattal [FF11]. This method has two advantages. On the one hand, it can predict the object boundaries quite well. This improves the performance of the correction step, as there are lower intensity errors. On the other hand, the approach increases the resolution of the input images iteratively by nondyadic scales (eg. 5:4 or 3:2). Therefore, in each iteration only a small higher frequency band is added to the image, leading to a more robust correction in terms of frequency distribution.

4.2 Correction

After the prediction step, the image is corrected according to the constraint

$$L_f(\mathbf{i}) = L_d^p(\mathbf{i}) + L_g^p(\mathbf{i}) \tag{3}$$

where $L_f(\mathbf{i})$ is the pixel at position $\mathbf{i}$ under full illumination. $L_d^p(\mathbf{i})$ and $L_g^p(\mathbf{i})$ correspond to the (unconstrained) predicted direct and global pixels respectively. Taking the difference of the sum of the direct and global images and the fully illuminated image

$$D(\mathbf{i}) = L_f(\mathbf{i}) - (L_d^p(\mathbf{i}) + L_g^p(\mathbf{i})), \tag{4}$$

we get the error introduced by the prediction step. Please note that, similar to the *Difference of Gaussians*, $D(\mathbf{i})$ only contains the missing higher frequencies which could not be recovered by the single-image up-scaling approach. In order to correct the predicted images, we distribute the error via two functions $f_d(\mathbf{i}, L_d^p, D)$, $f_g(\mathbf{i}, L_g^p, D)$ so that

$$D_\mathbf{i} = (f_d(\mathbf{i}, L_d^p, D) + L_d^p(\mathbf{i})) + (f_g(\mathbf{i}, L_g^p, D) + L_g^p(\mathbf{i})) \tag{5}$$

holds. These functions should be designed in such a way that they select these higher frequencies which are missing in the corresponding predicted image. If we now assume f_d and f_g to be linear with respect to D, we express our corrected images as follows

$$L_d^c(\mathbf{i}) = L_d^p(\mathbf{i}) + \alpha_d(\mathbf{i}) \cdot D(\mathbf{i}), \tag{6}$$

$$L_g^c(\mathbf{i}) = L_g^p(\mathbf{i}) + \alpha_g(\mathbf{i}) \cdot D(\mathbf{i}) \tag{7}$$

where L_d^c and L_g^c are the corrected direct and global component images, respectively. Here, α_d, α_g represent the weights of distributing the prediction error D

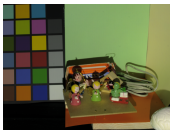
	 Scene 1		 Scene 2		 Scene 3		 Scene 4		 Scene 5	
	corr.	uncorr.	corr.	uncorr.	corr.	uncorr.	corr.	uncorr.	corr.	uncorr.
RMSE (d)	0.0067	0.0078	0.0061	0.0113	0.0088	0.0194	0.0057	0.0101	0.0090	0.0158
RMSE (g)	0.0063	0.0073	0.0057	0.0081	0.0085	0.0115	0.0058	0.0084	0.0088	0.0114
ℓ_1-Norm (d)	11370	14504	12167	20909	16738	36631	9788	16989	16639	30017
ℓ_1-Norm (g)	9503	12133	11334	15698	15399	20194	9633	13476	15084	18611
ℓ_2-Norm (d)	12.79	15.03	11.23	20.77	16.18	35.63	10.49	18.54	16.49	28.95
ℓ_2-Norm (g)	12.02	14.08	10.51	14.89	15.70	21.11	10.66	15.35	16.16	20.99

Table 1: Comparison of five test scenes with different error metrics (root mean square error (RMSE), ℓ_1- and ℓ_2-norm): Low resolution images with separated direct(d) and global(g) illumination were scaled up by a factor of 2 with the approach of [FF11]. The values of the corrected image can be compared with the uncorrected image for each scene.

to both images on a per-pixel level. Now, we need a condition how to set the α-values. As the wrong distribution tends to increase the noise level, a reasonable condition is to choose the weights in such a way that the present high frequencies are minimized. Thus, we can formulate a minimization problem where the sum of the squared Laplacian (Δ) of both images is minimized on a per-pixel level:

$$(\alpha_d(\mathbf{i}), \alpha_d(\mathbf{i})) = \arg\min_{\alpha_d,\alpha_g} \{(\Delta L_d)^2 + (\Delta L_g)^2\}. \quad (8)$$

Furthermore, we satisfy the constrains $0 \leq \alpha_d, \alpha_g \leq 1$ and $\alpha_d + \alpha_g = 1$ in order to eliminate the prediction error D. Solving for $\frac{\partial F}{\partial \alpha_d} = 0$ and $\frac{\partial F}{\partial \alpha_g} = 0$ with $F = (\Delta L_d)^2 + (\Delta L_g)^2$ yields

$$\alpha_d(\mathbf{i}) = \frac{\Delta L_d^p(\mathbf{i})}{\Delta D(\mathbf{i})}, \alpha_g(\mathbf{i}) = \frac{\Delta L_g^p(\mathbf{i})}{\Delta D(\mathbf{i})}. \quad (9)$$

The constraints can easily be satisfied by normalizing each α_d, α_g with the sum $\alpha_d + \alpha_g$. Thus, the resulting correction terms for direct and global component images can be expressed as

$$L_d^c(\mathbf{i}) = \frac{\Delta L_d^p(\mathbf{i})}{\Delta L_d^p(\mathbf{i}) + \Delta L_g^p(\mathbf{i})} \cdot D(\mathbf{i}) + L_d^p(\mathbf{i}), \quad (10)$$

$$L_d^c(\mathbf{i}) = \frac{\Delta L_g^p(\mathbf{i})}{\Delta L_d^p(\mathbf{i}) + \Delta L_g^p(\mathbf{i})} \cdot D(\mathbf{i}) + L_g^p(\mathbf{i}). \quad (11)$$

The direct and global component images L_d^c, L_g^c are then given as input for the prediction step of the next iteration until the target resolution is reached.

5 RESULTS

In order to evaluate our up-scaling scheme and the low-resolution separation methods, a measurement setup consisting of a projector (Panasonic PT-LC76E) and a single-lens reflex camera (Nikon D300) was used. Structured-Light patterns are projected into the scene that is imaged by the camera. The direct and global illumination for the given scene are separated by the approach of Nayar et al. [NKGR06]. In order to get a reliable ground truth, 25 shifted checker-board patterns were used.

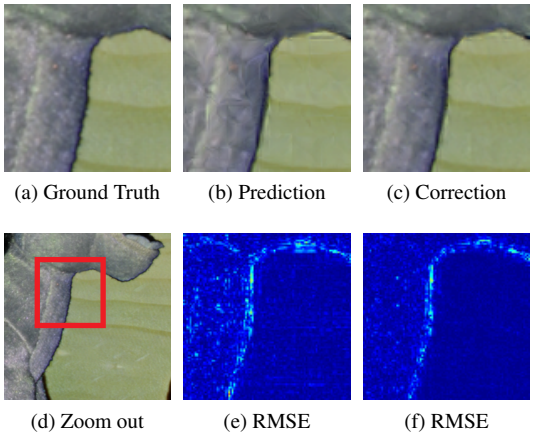
In Sec. 5.1, we evaluate our up-scaling scheme. Therefore, we sampled down the ground truth and used these low-resolution images as input. Five test scenes were created for evaluating the influences of different scene properties, the impact of our correction for several predictors and the influence of the up-scaling factor. Sec. 5.2 shows how our up-scaling scheme performs on low-resolution direct and global component images produced directly using different structured-light patterns. We focussed on patterns which require up to three acquisitions, only.

5.1 Evaluation of the Up-Scaling Scheme

Tab. 1 shows different error values for the test scenes and the separated illumination. The error of the corrected image can be compared with the error of the uncorrected image for each test case. The error values in scene 1 look similar while the error values in scene 3 are reduced by half. Depending on the properties of the scene, the error varies. It is observable that a more detailed scene can be corrected better than a scene with large homogeneous areas.

Influence of the Scene Properties

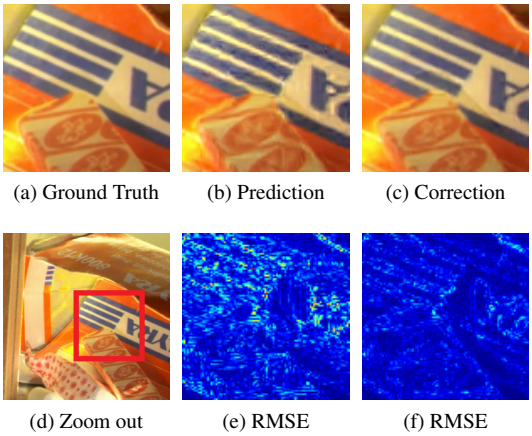
In the Figs. 3, 4 and 5, we show the influence of different kinds of cluttered regions on our up-scaling scheme. We used the approach of Freedman and Fattal [FF11] as the predictor and compare the predicted image with



(a) Ground Truth (b) Prediction (c) Correction

(d) Zoom out (e) RMSE (f) RMSE

Figure 3: Direct component of a stuffed toy in front of a homogeneous background. The high frequencies of the fur are corrected more than the low frequencies of the background.



(a) Ground Truth (b) Prediction (c) Correction

(d) Zoom out (e) RMSE (f) RMSE

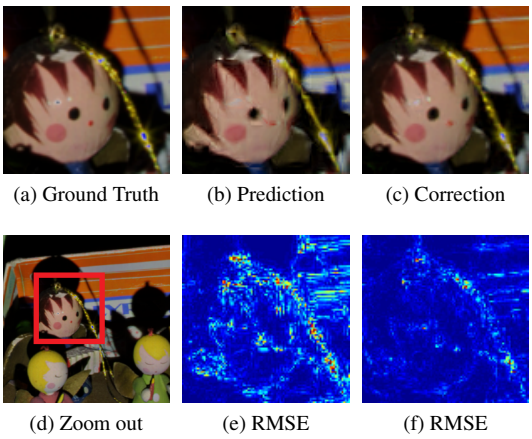
Figure 4: Global component of two pieces of chalk wrapped in paper in front of a printed box. Independent of the 3D object structure, the high frequencies are corrected much better.

the results of the proposed correction scheme. Fig. 3 shows a close-up of a stuffed toy in front of a homogeneous background. As can be seen, our scheme reproduces the fine fur structure more detailed while in the predicted image this region looks smeared out. The homogeneous background is reproduced well with both methods. In Fig. 4, two pieces of chalk wrapped in paper in front of a printed box are shown. Here, in contrast to the stuffed toy, the fine details are not due to fine 3-dimensional structures. The letters on the box and the printed patterns of the paper wrapping the chalk are reconstructed in more detail with our correction. Fig. 5 shows small wooden figures with small scale painting and thin ornamentations. The edges can be reproduced more precisely with the correction scheme compared to the prediction.

These three examples show that the quality of the correction is independent of the scene objects. Only the frequencies in the resulting image of the scene objects are responsible for the impact of the improvement. Fine structures, i.e. high frequencies, in the image lead to a large improvement in quality while areas with low frequencies cannot be corrected much more.

Improvement of the Correction

The up-scaling of the five exemplary scenes was done with three different methods to evaluate the influence of the prediction to the correction. Nearest neighbour interpolation, bicubic interpolation and local self-examples from [FF11] were used as predictors for the up-scaling scheme. Fig. 6 shows a comparison between the RMSE for the different prediction methods and the afterwards corrected images. It can be seen that the correction always improves the predicted image, but depending on the quality of the prediction the correction



(a) Ground Truth (b) Prediction (c) Correction

(d) Zoom out (e) RMSE (f) RMSE

Figure 5: Direct component of small wooden figures in front of a printed box. The correction between different objects is corrected very good. High frequencies are corrected independently of the objects.

cannot compensate the larger error. For example, the correction of the nearest neighbour interpolation, which has the largest error in all scenes, improves the error but still has a larger absolute error than a better prediction without a correction for the global illumination in scene 3 and 5. An additional conclusion is that our proposed correction improves prospective predictions that will have less errors.

Please note that the local self-examples approach produces a higher prediction error than bicubic interpolation for scenes 2 to 5. In contrast to scene 1, these scenes contain cluttered regions. As stated by Freedman and Fattal, their approach is not able of handling these kind of image region and the algorithm spuriously reconstructs edges there [FF11]. This behaviour causes larger error than bicubic interpolation with its smoothing ability.

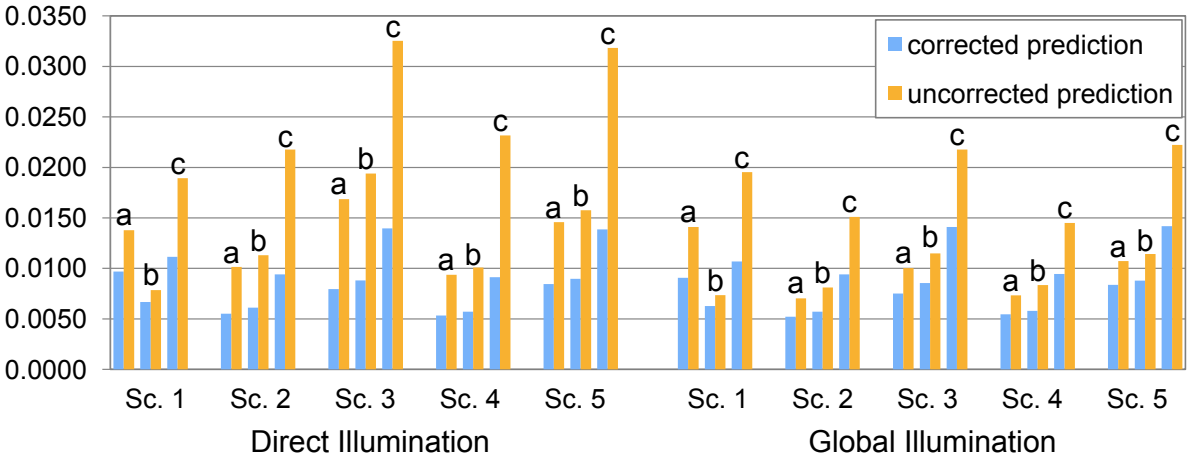


Figure 6: The sorted RMSE for (a) bicubic interpolation, (b) Local Self-Examples [FF11] and (c) nearest neighbour interpolation as predictors shows that the quality of the prediction influences the quality of the correction. The correction improves the quality of the prediction for direct and global illumination in all five scenes.

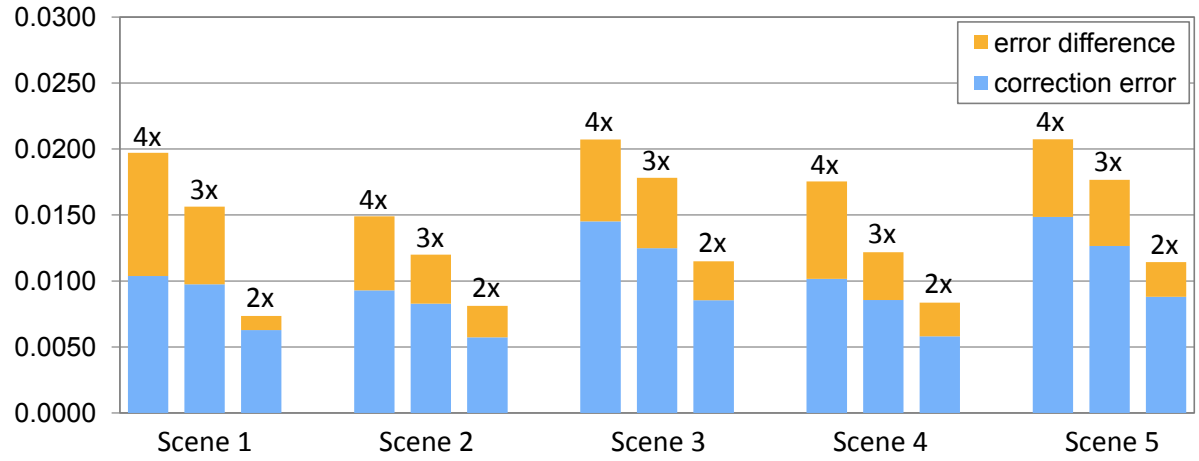


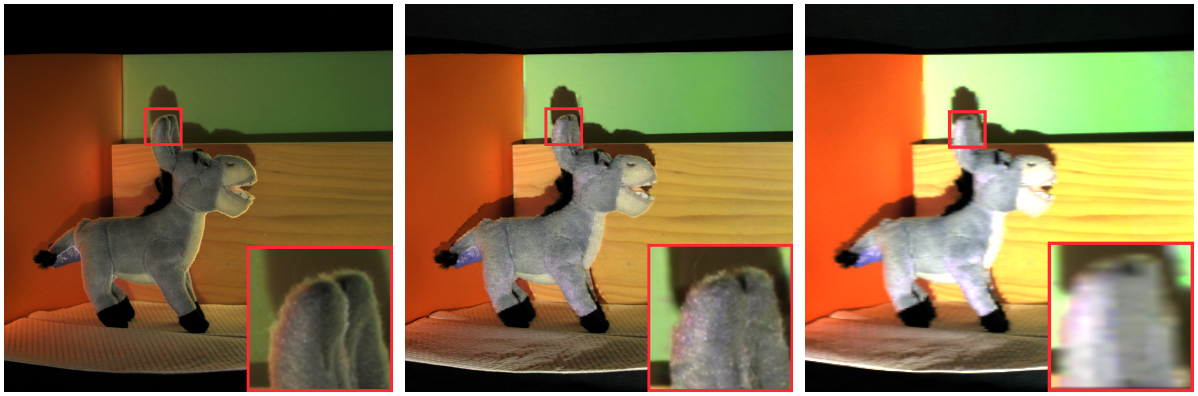
Figure 7: The dependency of the up-scaling factors is shown. For each scene the factors 4x, 3x, 2x are compared (left to right bar). A higher scaling factor leads to a higher error difference between prediction and correction. The global illumination images of the five test scenes are upscaled using [FF11].

Influence of the Up-Scaling Factor

The up-scaling factor is another parameter of influence. Fig. 7 compares the RMSE for the scaling factors of 4x, 3x and 2x from left to right for each global illumination image of the five scenes. The single-image up-scaling method of [FF11] is used as predictor for this comparison. The absolute error introduced by the prediction is higher for larger scaling. But, it can be seen that the improvement of the reconstruction error by our correction scheme gets better for higher scaling factors in all scenes. As our algorithm works in an iterative manner, for higher up-scaling factors, more refinement steps are done. As in each step the predicted image is corrected we are able to better guide the up-scaling process.

5.2 Evaluation of Patterns

We used different structured-light patterns for the separation of the illumination component in order to evaluate their applicability regarding our up-scaling scheme. As the focus is on low acquisition time, only patterns were chosen where up to three acquisition (including the fully illuminated image) are needed. Thus, we tested vertical stripes, horizontal stripes, a combination of both stripe patterns and a high-frequency checker-board pattern where each square had the same dimensions as the width of the stripes. The single-image separation approach proposed by Nayar et al. [NKGR06] was used in order to determine the low-resolution direct and global image components (cf. left part of Fig. 2). As for the stripes patterns a blurring of edges within the direction of the stripes occur, we combined the separation results of both patterns in order to compensate for this effect. However, for this procedure an additional acquisition is required. Therefore, we introduced the high-frequency checker-board pattern. Here, the same separation methods were applied as for horizontal



(a) Global Checker Board (ground truth) (b) Global Stripes (up-scaled) (c) Global Stripes (low resolution)

Figure 8: Comparison of the separation methods using checker-board (a) and stripes structured-light patterns (b). As can be seen, the stripes separation (c) suffers severe artefacts. Our up-scaling scheme is not capable of properly reconstructing the high-resolution images.

	vert. stripes	hor. stripes	comb. stripes	vert. checker	hor. checker	comb. checker	down-sampled ground truth
RMSE (LSE)	0.2219	0.2024	0.2085	0.3083	0.3099	0.3089	0.0140
ℓ_1-Norm (LSE)	29703	26350	27768	41835	42058	41945	1236
ℓ_2-Norm (LSE)	221.88	202.41	208.52	308.26	309.85	308.91	14.01
RMSE (BC)	0.2147	0.1826	0.1948	0.3020	0.3036	0.3015	0.0233
ℓ_1-Norm (BC)	28792	24059	26160	41043	41369	41066	1810
ℓ_2-Norm (BC)	214.72	182.55	194.84	302.03	303.63	301.49	23.30

Table 2: Different kind of patterns were used for a fast separation. The resulting images were fed into our up-scaling scheme with local self-examples (LSE) and bicubic interpolation (BC) as predictor.

and vertical stripes. In order to improve blurring artefacts, the separation results were combined, also. Tab. 2 shows the reconstruction error of our up-scaling scheme for each of these separation methods compared to the down-sample ground truth. As predictors we used bicubic interpolation and local self-examples [FF11]. Our experiments show that the reconstruction results differ depending on whether the down-sampled ground truth or the low-resolution images resulting from the stripes or checker-board patterns are used. As can be seen in Fig. 8, the stripes-pattern separation suffers severe reconstruction artefacts. Besides the wrong color reproduction in homogeneous regions, at diagonal edges stair-case artefacts are present. Furthermore, cluttered regions are subjected to noise. Although our up-scaling scheme is capable of reducing the artefacts in brighter image regions, it cannot recover the fine details. In general, the up-scaling error for the stripes patterns is smaller as for the checker-board. The choice of the predictor does not influence the up-scaling result significantly. Also, the combination of horizontal and vertical stripes could not improve the quality of the input images. By monitoring the weights for the prediction error distribution α_d, α_g, we realized that they were close to 0.5 for all pixels of the direct and global images. This is no surprise, as the noise in both images cannot be reduced significantly by distributing the prediction error.

6 CONCLUSIONS

We presented a new constrained up-scaling approach for direct an global component images. Predictions based on single-image up-scaling techniques are corrected using a high resolution fully illuminated image. It can be seen that our method can visually improve cluttered regions of an image such as fur or fine structured ornamentations. The experiments showed that our algorithm scales with the quality of the prediction as even well predicted images can be improved further. The iterative manner of the method is also capable of improving single-image up-scaling approaches for large up-scaling factors.

Our method is applicable when the acquisition full resolution images using many structured-light patterns is not possible due to limited time-constraints. This is the case for capturing dynamic scenes in Computer Vision tasks, such as biometry of human faces or mobile autonomous navigation. Here, only two images are required. However, by acquiring more than one image our method is prone to motion artefacts and misalignment of the two images. Thus, it might be necessary to apply optical flow or motion compensation. Also, our up-scaling scheme poses additional computational complexity which might not be suitable for real-time analysis of the acquired data.

Furthermore, when using structured-light patterns or variations known in literature for fast acquisition, the quality of the resulting low-resolution images is not sufficient. Obviously, these separation methods are strongly influenced by wrong color reproduction and stair-case artefacts. Thus, these kind of patterns produce fundamentally different direct and global component images than methods which work directly at the resolution of the acquired images.

In future work, we study the impact of different light patterns on the separation results. This will give us useful insight about how to solve the above mentioned discrepancies between low-resolution images and ground-truth data.

7 ACKNOWLEDGMENTS

This work was funded by the German Research Foundation (DFG) as part of the research training group GRK 1564 'Imaging New Modalities'.

8 REFERENCES

[AB12] L. An and B. Bhanu. Image super-resolution by extreme learning machine. In *Proc. IEEE Int. Conf. Image Processing (ICIP)*, pages 2209 –2212, 2012.

[ABA01] C. Atkins, C. Bouman, and J. Allebach. Optimal image scaling using pixel classification. In *Proc. Int. Conf. Image Processing*, volume 3, pages 864 –867, 2001.

[BGS02] S. Battiato, G. Gallo, and F. Stanco. A locally adaptive zooming algorithm for digital images. *Image and Vision Computing*, 20(11):805 – 812, 2002.

[CHL05] M.-J. Chen, C.-H. Huang, and W.-L. Lee. A fast edge-oriented algorithm for image interpolation. *Image and Vision Computing*, 23(9):791 – 798, 2005.

[EV07] M. Ebrahimi and E. Vrscay. Solving the inverse problem of image zooming using self-examples. In M. Kamel and A. Campilho, editors, *Image Analysis and Recognition*, volume 4633 of *Lecture Notes in Computer Science*, pages 117–130. Springer Berlin Heidelberg, 2007.

[Fat07] R. Fattal. Image upsampling via imposed edge statistics. *ACM Trans. Graph.*, 26(3), 2007.

[FF11] G. Freedman and R. Fattal. Image and video upscaling from local self-examples. *ACM Trans. Graph.*, 30(2):12:1–12:11, 2011.

[FJP02] W. Freeman, T. Jones, and E. Pasztor. Example-based super-resolution. *IEEE Computer Graphics and Applications*, 22(2):56 –65, 2002.

[GA08] A. Giachetti and N. Asuni. Fast artifact free image interpolation. In *Proc. BMVC 2008*, 2008.

[GA11] A. Giachetti and N. Asuni. Real-time artifact-free image upscaling. *IEEE Trans. Image Processing*, 20(10):2760 –2768, 2011.

[GADTH97] M. Gharavi-Alkhansari, R. DeNardo, Y. Tenda, and T. S. Huang. Resolution enhancement of images using fractal coding. *Proc. SPIE 3024, Visual Communications and Image Processing*, 3024:1089–1100, 1997.

[GBI09] D. Glasner, S. Bagon, and M. Irani. Super-resolution from a single image. In *Proc. IEEE Int. Conf. Computer Vision*, pages 349 –356, 2009.

[GKGN11] J. Gu, T. Kobayashi, M. Gupta, and S. K. Nayar. Multiplexed illumination for scene recovery in the presence of global illumination. In *Proc. IEEE Int. Conf. Computer Vision (ICCV)*, pages 1–8, 2011.

[GL12] J. Gu and C. Liu. Discriminative illumination: Per-pixel classification of raw materials based on optimal projections of spectral brdf. In *Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR)*, pages 797–804, 2012.

[KK08] K. Kim and Y. Kwon. Example-based learning for single-image super-resolution. In G. Rigoll, editor, *Pattern Recognition*, volume 5096 of *Lecture Notes in Computer Science*, pages 456–465. Springer Berlin Heidelberg, 2008.

[LHL11] B. Langmann, K. Hartmann, and O. Loffeld. Comparison of depth super-resolution methods for 2d/3d images. *Int. Journal of Computer Information Systems and Industrial Management Applications*, 3:635 –645, 2011.

[LHL12] B. Langmann, K. Hartmann, and O. Loffeld. A modular framework for 2d/3d and multi-modal segmentation with joint super-resolution. In A. Fusiello, V. Murino, and R. Cucchiara, editors, *Computer Vision - ECCV 2012. Workshops and Demonstrations*, volume 7584 of *Lecture Notes in Computer Science*, pages 12–21. Springer Berlin Heidelberg, 2012.

[LLK08] M. Lindner, M. Lambers, and A. Kolb.

Data fusion and edge-enhanced distance refinement for 2d rgb and 3d range images. *Int. J. on Intell. Systems and Techn. and App. (IJISTA), Issue on Dynamic 3D Imaging*, 5(1):344 – 354, 2008.

[LO01] X. Li and M. Orchard. New edge-directed interpolation. *IEEE Trans. Image Processing*, 10(10):1521 –1527, 2001.

[MYR10] Y. Mukaigawa, Y. Yagi, and R. Raskar. Analysis of light transport in scattering media. In *Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR)*, pages 153–160, 2010.

[NG12] S. K. Nayar and M. Gupta. Diffuse structured light. In *Proc. IEEE Int. Conf. Computational Photography*, pages 1–8, 2012.

[NKGR06] S. K. Nayar, G. Krishnan, M. D. Grossberg, and R. Raskar. Fast separation of direct and global components of a scene using high frequency illumination. *ACM Trans. Graph.*, 25(3):935–944, 2006.

[ORK12] M. O'Toole, R. Raskar, and K. N. Kutulakos. Primal-dual coding to probe light transport. *ACM Trans. Graph.*, 31:39:1–39:11, 2012.

[RRC12] D. Reddy, R. Ramamoorthi, and B. Curless. Frequency-space decomposition and acquisition of light transport under spatially varying illumination. In *Computer Vision - ECCV 2012*. Springer Berlin Heidelberg, 2012.

[SSU08] N. Suetake, M. Sakano, and E. Uchino. Image super-resolution based on local self-similarity. *Optical Review*, 15:26–30, 2008.

[STDT08] S. Schuon, C. Theobalt, J. Davis, and S. Thrun. High-quality scanning using time-of-flight depth superresolution. In *Proc. IEEE Conf. on Comp. Vis. & Patt. Recogn.; Workshops (CVPRW)*, pages 1 –7, 2008.

[SW04] D. Su and P. Willis. Image interpolation by pixel-level data-dependent triangulation. *Computer Graphics Forum*, 23(2):189–201, 2004.

[SXS08] J. Sun, Z. Xu, and H.-Y. Shum. Image super-resolution using gradient profile prior. In *Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR)*, pages 1 –8, 2008.

[TM96] S. Thurnhofer and S. K. Mitra. Edge-enhanced image zooming. *Optical Engineering*, 35(7):1862–1870, 1996.

[TTT06] Y.-W. Tai, W.-S. Tong, and C.-K. Tang. Perceptually-inspired and edge-directed color image super-resolution. In *Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR)*, volume 2, pages 1948 – 1955, 2006.

[WXM+13] L. Wang, S. Xiang, G. Meng, H.-y. Wu, and C. Pan. Edge-directed single image super-resolution via adaptive gradient magnitude self-interpolation. *IEEE Trans. Circuits and Systems for Video Technology*, PP(99):1, 2013.

[YWL+12] J. Yang, Z. Wang, Z. Lin, S. Cohen, and T. Huang. Coupled dictionary training for image super-resolution. *IEEE Trans. Image Processing*, 21(8):3467 –3478, 2012.

[YXXY12] L. Yu, H. Xu, Y. Xu, and X. Yang. Robust single image super-resolution based on gradient enhancement. In *Proc. Signal Information Processing Association Annual Summit and Conference (APSIPA ASC)*, pages 1 –6, 2012.

[YXYC12] Q. Yan, Y. Xu, X. Yang, and K. Chen. Image super-resolution based on a novel edge sharpness prior. In *Proc. Int. Conf. Pattern Recognition (ICPR)*, pages 1056 –1059, 2012.

[ZDLL12] L. Zi, J. Du, M. Liang, and J. Lee. A perception-motivated image interpolation algorithm. In *Proc. World Congress on Intelligent Control and Automation (WCICA)*, pages 4754 –4759, 2012.