



**HARVEST4D**

HARVESTING DYNAMIC 3D WORLDS FROM COMMODITY SENSOR CLOUDS

## Publications for Task 6.2

### Deliverable 6.21

Date: 15.01.2016  
Grant Agreement number: EU 323567  
Project acronym: HARVEST4D  
Project title: Harvesting Dynamic 3D Worlds from Commodity Sensor Clouds

## Document Information

<b>Deliverable number</b>	D6.21
<b>Deliverable name</b>	Publications for Task 6.2
<b>Version</b>	1.0
<b>Date</b>	2016-01-15
<b>WP Number</b>	6
<b>Lead Beneficiary</b>	TUDA
<b>Nature</b>	R
<b>Dissemination level</b>	PU
<b>Status</b>	Final
<b>Author(s)</b>	Tobias Plötz, Stefan Roth

## Revision History

Rev.	Date	Author	Org.	Description
0.1	2016-01-05	Tobias Plötz	TUDA	Initial draft
1.0	2016-01-15	Tobias Plötz, Stefan Roth	TUDA	Final version

## Statement of originality

This deliverable contains original unpublished work except where clearly indicated otherwise. Acknowledgement of previously published material and of the work of others has been made through appropriate citation, quotation or both.

---

## TABLE OF CONTENTS

1	Executive Summary .....	1
1.1	Introduction .....	1
1.2	Publications .....	1
2	Description of Publications .....	1
2.1	Overview .....	1
2.2	Towards Datasets of Noise from Image Sensors.....	1
3	References.....	2

---

## 1 EXECUTIVE SUMMARY

### 1.1 INTRODUCTION

This deliverable describes the publications that resulted from Task 6.2, and how they fit into the work plan of the project.

The objective of Task 6.2 is to analyze the distribution of noise from various sensors. Insights gained from this analysis are believed to be beneficial for other upstream applications that make use of the sensed data.

There is so far one publication that is mainly attributable to Task 6.2. As it contains still unpublished material, at the time of delivery it can be found in the restricted section of the website only.

### 1.2 PUBLICATIONS

The following publication can be found on the webpage only:

- T. Plötz, F. Saeedan and S. Roth.  
*Towards Datasets of Noise from Image Sensors.*  
Working paper, TU Darmstadt, 2015

---

## 2 DESCRIPTION OF PUBLICATIONS

### 2.1 OVERVIEW

Harvest4D makes use of various input modalities. Especially images are of great importance since they constitute the input to 3D reconstruction techniques like Structure-from-Motion (see WP5), image registration (see WP4) or material acquisition (see WP7). However, when capturing images, these are inherently degraded by image noise and algorithms that process images are affected by this noise. For this reason, the working paper [Plötz, Saeedan, Roth 2015] aims at quantifying image noise.

### 2.2 TOWARDS DATASETS OF NOISE FROM IMAGE SENSORS

Many algorithms that handle images as input can benefit from a good model of image noise as it allows to incorporate a notion of confidence into an algorithmic framework. Ideally, we would like to have a database of measured image noise in order to model the statistical properties of the

noise distribution. However, measuring noise is a hard problem since it is difficult to separate noise from signal from just a single image.

Therefore, this working paper presents an acquisition procedure that allows to capture pairs of images of the same scene where one of the images has little noise and one is affected by stronger noise. In the end, the difference between both images can be used to measure the amount of noise in the latter image, thus opening the possibility to faithfully model the underlying noise distribution. Moreover, the data could be used to benchmark denoising and noise estimation algorithms. An example of two images showing the same scene with different amounts of noise is shown in [Figure 1](#).

The paper validates the acquisition procedure in a rigorous mathematical framework. In practice, the theoretical model does not apply to the captured images to full extent since, for example, small changes in the scene and illumination result in a residual measurement error. The paper presents and evaluates a post-processing procedure of the captured images to correct the effect of these errors.

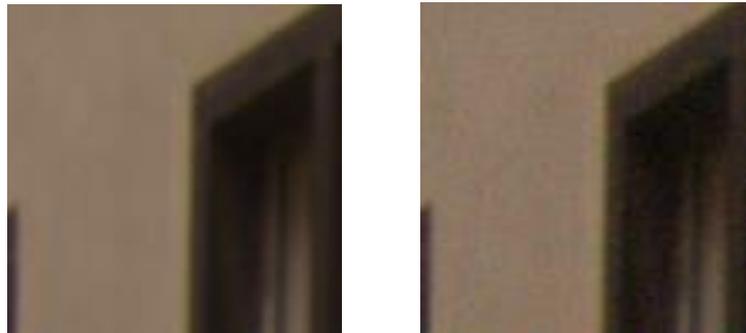


Figure 1: A crop from two images from one scene of the captured dataset. Left: Image taken with low analog gain. Right: Image taken with high analog gain.

---

### 3 REFERENCES

- T. Plötz, F. Saeedan and S. Roth. Towards Datasets of Noise from Image Sensors. Working paper, TU Darmstadt, 2015

# Towards Datasets of Noise from Image Sensors

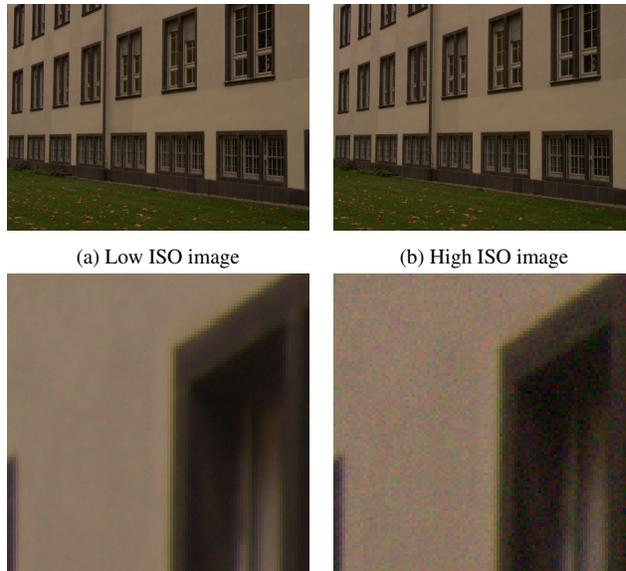
Tobias Plötz      Faraz Saeedan      Stefan Roth  
Department of Computer Science, TU Darmstadt

## Abstract

Noise can significantly degrade the quality of images, hence a multitude of algorithms to estimate and remove noise have been proposed. However, due to the lack of realistic ground truth data, those algorithms are traditionally evaluated on images corrupted by synthesized noise. In this paper we propose means to capture a new dataset that contains pairs of images showing the same scene. One image depicts the scene with little noise, the other with large amounts of noise. Subtracting both allows to study characteristics of the residual image noise and can be used to derive a benchmark for denoising and noise estimation algorithms. We give a theoretical treatment of the acquisition procedure, discuss some challenges encountered in practice, and show how to cope with them.

## 1. Introduction

Noise is inherent to every imaging system. Especially in low-light scenarios, noise often severely degrades the image quality. Although sensor technology has advanced significantly in the recent past, a substantial and inherent amount of noise is caused by the uncertainty of the arrival process of photons. The trend of increasing the number of pixels that are fit onto the sensor further worsens the effect of this shot noise, since the amount of incident light at every pixel decreases with a smaller sensor area being allocated to each pixel. A number of other noise sources within the camera itself further contribute to this challenge. Therefore, a large variety of denoising algorithms have been developed to deal with noise in digital images (e.g. [3, 4, 19]). Even though images with real sensor noise can be easily captured, it is much less straightforward to know what the true noise-free image should be. In the absence of ground truth data it is thus challenging to devise faithful noise models. Without such data, denoising algorithms are typically evaluated by synthetically adding noise to existing, mostly clean images and trying to remove it again [19]. But this only yields an approximation to the true denoising performance in real applications, which is limited by the accuracy of the underlying noise model that is used to synthesize noise. Similar



(c) Zoomed part of the low ISO image (left) and the high ISO image (right)

Figure 1. An image pair of a low and high ISO image from our dataset. We show the JPGs for better display.

challenges exist in noise estimation algorithms [2, 9]. To avoid this, it is thus highly desirable to have a dataset of pairs of real noisy and noise-free images in order to assess the quality of denoising and noise estimation algorithms in a faithful manner.

To the best of our knowledge, there exists only one public dataset for image noise so far [1]. One shortcoming of this dataset is that crucial details about the acquisition procedure are not discussed. When attempting to replicate their acquisition procedure, we encountered numerous practical difficulties, which do not appear to have been addressed. For this reason, we here aim to contribute a novel dataset of image noise with a thorough mathematical treatment of the acquisition protocol as well as significant detail on the practical aspects of the acquisition. The core of our procedure builds on capturing a pair of noisy and noise-free images by taking several images of the same scene with different analog gains (ISO values) and exposure times. The per-pixel intensities in the captured images will follow dif-

ferent noise distributions but, crucially, they share the same true underlying image intensity. Intuitively, the amount of per-pixel noise will increase with the gain and we study this behavior in a rigorous mathematical framework. Overall, the difference between an image captured with a low gain setting and an image with a high gain setting will mainly yield the noise that is present in the latter. Figure 1 shows one of the captured image pairs in our dataset. For this acquisition procedure to produce sensible images it is necessary that the scene as well as the camera do not change in between exposures. However, in practice we found different sources of violations of this assumption, such as misalignment and subtle lighting changes. We describe and discuss a procedure for mitigating these effects.

## 2. Related Work

Since noise is abundant in any imaging system, different **models of sensor noise** have been studied in the literature. A thorough analysis of the different noise sources is provided by Healy *et al.* [11] for CCD image sensors and by El Gamal *et al.* [5] for CMOS sensors. The main observation is that there is inherent noise related to the stochastic arrival process of photons hitting the sensor. This so-called shot noise cannot be eliminated and since it follows a Poisson distribution, its variance is proportional to the mean intensity at a specific pixel and is hence not stationary across the whole image.

Other noise sources originate from the electronics within the sensor chip and can be measured in order to assess the quality of the sensor. Consequently, this led to the development of the standard 1288 of the European Machine Vision Association [6], which provides a standardized protocol for measuring the characteristics of the different noise sources on a sensor. In this protocol the sensor is illuminated with constant irradiation over the whole sensor area and intensity measurements are aggregated *spatially*. This is repeated for different irradiation levels to capture the intensity-dependence of the noise. In contrast, [8] propose a method similar to our capture protocol, where they take multiple exposures of a static scene and then *temporally* aggregate the measurements at every pixel site. Finally [18] studies the dependence of image noise under varying ISO sensitivities and temperatures.

In a practical application it is useful to assess the amount of noise that degrades a certain input image. However, the specific noise characteristics for the camera that took the image might be unknown. Hence, there has been a substantial amount of research in **estimating the noise strength from a single image**. The main idea behind most methods is to search for areas of constant intensity within the image in order to assess the standard deviation of the noise for that intensity. Then a model is fit to represent the noise-level function, *i.e.* the function that relates image intensity

and noise strength. Foi *et al.* [9] use an approach based on wavelet decomposition to estimate pairs of mean intensities and noise standard deviations. A quadratic model is fitted to these measurements to estimate the noise-level function. They also take into account clipping effects due to under- and overexposure. Because measurements of noise are often polluted by high frequency textures, [2] robustify this approach by considering the effect of outlier measurements during the model fitting.

While the above works aim to estimate the noise for raw image data, Liu *et al.* [14] estimate the intensity-dependent strength of noise for images that have already been gone through the camera-internal image processing pipeline. This pipeline can be modeled with a nonlinear camera response function that maps raw input intensities to final output intensities. In [14] a low dimensional basis for processed noise-level functions is computed. The intensity/standard deviation pairs are fitted against this basis to get an estimate of the real noise level function of the image.

Although the image noise variance depends on the underlying intensity, many denoising algorithms ignore this fact and evaluate against artificially generated, stationary noise, where usually i.i.d. Gaussian noise is used. Instead, other works aim to undo intensity-dependent noise [7, 14, 15]. The main idea there is to model the noise distribution as a heteroscedastic Gaussian, whose variance is intensity-dependent. This is valid since the Poissonian components of the total noise can be well approximated by a Gaussian. Other approaches first apply a variance stabilizing transform [16] in order to then employ a denoising method for stationary Gaussian noise. However, the transform may make the noise distribution non-Gaussian.

At the time of writing and to the best of our knowledge, there is no peer-reviewed **benchmark for image noise**. Anaya and Barbu [1] provide the RENOIR dataset, which has not been published to date. Their acquisition technique is highly related to ours as they also take sets of images of a static scene with different ISO values. However, they do not provide much detail on how misalignments within a single acquisition are corrected for, and they convert the raw intensities to 8 bits, thus losing dynamic resolution. Moreover, some of the publicly available images seem to exhibit spatial misalignments; it is thus unclear how reliable the dataset is.

A related benchmark for low-level vision applications is the Microsoft Research Demosaicing Dataset [13]. They downscale high-resolution images to simulate a sensor without a color filter array and use that as ground-truth for the demosaicing task. However, this approach cannot directly be adapted to creating an image noise dataset, as extracting the noise from just a single image will always face the problem of separating noise from signal. Hence, we use two images – one with little noise one that is affected

---

**Algorithm 1** Capture protocol.

---

**Require:**  $K_0, t_{exp,0}$ : Base ISO of camera and initial exposure time

```
 $n \leftarrow \{1, 4, 16, 1\}$   
for  $i = 1, \dots, |n|$  do  
   $K \leftarrow K_0 \cdot n$   
   $t_{exp} \leftarrow t_{exp,0}/n$   
   $I_i \leftarrow$  take exposure  
end for  
return  $I_1, \dots, I_{|n|}$ 
```

---

strongly with noise. Combining both images helps us to separate signal from noise.

### 3. A Dataset of Image Noise

**Capture protocol.** For capturing the dataset, we use the protocol that is outlined in Algorithm 1. The main idea is to capture multiple exposures of the same scene, each with a different combination of ISO value and exposure time. These are chosen in a way as to ensure that the observed intensities remain constant as much as possible, up to noise. We use two different cameras: A Sony A7R with a full-frame sensor and base ISO 100 and an Olympus OM-D E-M10 with a Micro-Four-Thirds sensor and base ISO 200. Both cameras are mirror-less, thus reducing vibrations due to the flapping mirror compared to DSLRs. We mount the camera on a sturdy tripod with a stabilizing weight attached.

Between the individual exposures, the aperture, white balance, the focus and all other camera parameters except for the ISO value and exposure time remain constant. The capture protocol is executed via an Android app that issues all necessary commands to the camera over a WiFi network. In this way, human interaction during the capturing is minimized to limit the amount of camera shake and scene variation.

In the following we analyze our capture protocol from a theoretical perspective, showing that in ideal circumstances it produces identical intensities across the different exposures, up to noise.

**A model of image sensor noise.** Sensor noise in digital CMOS or CCD cameras can significantly degrade a captured image. Hence, models that describe the statistical characteristics of the noise have been developed in the past [11, 6]. We begin by reviewing the basic sources of image noise and then study their dependence on the amplifier gain and the exposure time, since these parameters will be varied within the capturing protocol for our dataset. Figure 2 shows a schematic overview of the imaging process per pixel along with the different noise sources that influ-

ence the final pixel output.

At each pixel site, the photoelectric effect causes incident light in the form of individual photons to be transformed to electrons. The number of incoming photons  $N_p$  during the exposure time  $t_{exp}$  is a random quantity with a Poisson distribution [12] with the mean being linearly dependent on the exposure time:

$$N_p \sim \mathcal{P}(\mu_p) \quad (1)$$

$$\text{with } \mu_p = F_p \cdot t_{exp}. \quad (2)$$

Here,  $F_p$  describes the mean arrival rate of photons per time interval. A basic property of the Poisson distribution is that the variance is equal to the mean

$$\sigma_p^2 = \mu_p = F_p \cdot t_{exp}. \quad (3)$$

The inherent uncertainty in the photon counting process is called *shot noise*, and cannot be avoided by any technical means. The number of electrons  $N_e$  that are generated during the exposure time depends on the number of photons and the quantum efficiency  $\eta$  of the sensor:

$$N_e = \eta \cdot N_p. \quad (4)$$

For the sake of simplicity, in this paper the quantum efficiency describes the overall rate of transformed electrons per incident photon, thus subsuming the actual quantum efficiency of the photo sensing element at the pixel, the fact that not all incoming light hits the sensing element, and others. Also, the quantum efficiency actually depends on the wavelength of the incident light. We can neglect this parameter, as we regard the incoming spectrum to be static between captures.

Consequently, the number of generated electrons can also be described by a Poisson distribution with a mean that linearly depends on  $t_{exp}$ :

$$N_e \sim \mathcal{P}(\mu_e) \quad (5)$$

$$\text{with } \mu_e = \eta \cdot \mu_p = \eta \cdot F_p \cdot t_{exp} \quad (6)$$

$$\text{and } \sigma_e^2 = \mu_e. \quad (7)$$

The photosensitive electronics also produce charge when no light is falling onto the sensor. This effect is called *dark current* [6] and is caused by random thermal fluctuations in the semi-conductor material. The number of “dark” electrons is Poisson distributed as well, and the mean similarly depends on the exposure time:

$$N_d \sim \mathcal{P}(\mu_d) \quad (8)$$

$$\text{with } \mu_d = F_d \cdot t_{exp} \quad (9)$$

$$\text{and } \sigma_d^2 = \mu_d, \quad (10)$$

where  $F_d$  is the generation rate of dark electrons per time unit. Thus, the total amount of observed charges equals

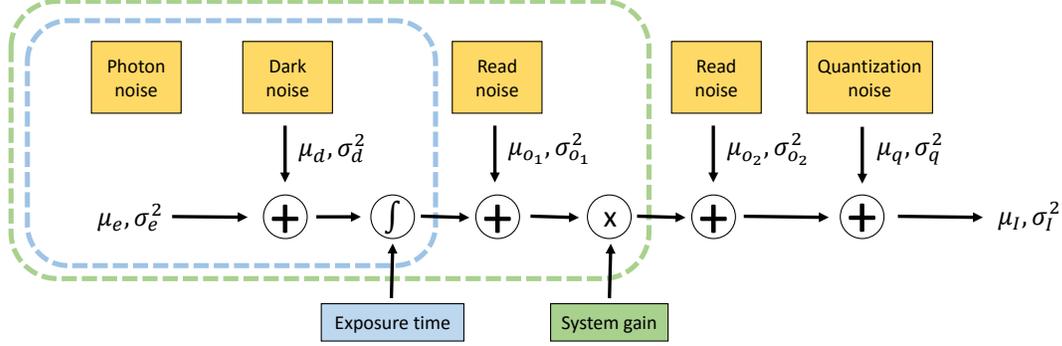


Figure 2. Schematic model of how different noise sources get aggregated in the imaging process. The parts within the blue box are influenced by the exposure time, the parts in the green box are influenced by the system gain.

$N_e + N_d$ . Depending on the sensor technology it is directly converted to a voltage at the pixel site (CMOS sensors [5]) or carried to separate read-out electronics (CCD sensors [11]), where it is converted to a voltage. The voltage is then amplified with a gain factor  $K$  and an analog-to-digital converter is used to subsequently quantize the voltage and convert it to a digital signal. The electronic circuitry adds noise that is independent of the received charge and the exposure time [11]. In this exposition, we split this additional noise in a component  $N_{o_1}$  that is induced before the amplification<sup>1</sup> and a component  $N_{o_2}$  that is induced after the amplification. These noise sources are usually regarded as being Gaussian distributed [6] with mean zero<sup>2</sup>:

$$N_{o_1} \sim \mathcal{N}(0, \sigma_{o_1}^2) \quad (11)$$

$$N_{o_2} \sim \mathcal{N}(0, \sigma_{o_2}^2). \quad (12)$$

The quantization step can be regarded as another noise source. Although being deterministic given a known accumulated signal, quantization can be regarded as uniformly distributed noise  $N_q$  if the signal itself is uncertain [6]:

$$N_q \sim \mathcal{U}(-0.5\Delta_e, 0.5\Delta_e), \quad (13)$$

where  $\Delta_e$  is the amount of charge that is quantized to the same digital output. The mean and variance of the quantization noise are therefore

$$\mu_q = 0 \quad (14)$$

$$\text{and } \sigma_q^2 = \frac{1}{12} \Delta_e. \quad (15)$$

With all these ingredients, we can write the total observed signal as

$$N_I = K \cdot (N_e + N_d + N_{o_1}) + N_{o_2} + N_q. \quad (16)$$

<sup>1</sup>Regarding charge and voltage as convertible, we express  $N_{o_1}$  also as a quantity measured in electrons.

<sup>2</sup>Any non-zero bias is assumed to be detected by the camera logic and corrected for.

Only for pixels that are nearly under- or overexposed the theoretical model is not accurate as it disregards clipping effects. These pixels are, however, easily detectable and can be excluded from further analysis. To validate our capture protocol we now study how the mean and the variance of the observed signal will behave when we increase the gain  $K$  and at the same time decrease the exposure time  $t_{\text{exp}}$  by the same factor. Due to the linearity of the expectation operator the mean signal as a function of gain and exposure time is given by

$$\mu_I(K, t_{\text{exp}}) = K \cdot (\mu_e + \mu_d + \mu_{o_1}) + \mu_{o_2} + \mu_q \quad (17)$$

$$= K \cdot t_{\text{exp}} (\eta \cdot F_p + F_d) + K \cdot \mu_{o_1} + \mu_{o_2} + \mu_q \quad (18)$$

$$= K \cdot t_{\text{exp}} (\eta \cdot F_p + F_d). \quad (19)$$

Where the last equality follows from the fact that  $N_{o_1}$ ,  $N_{o_2}$  and  $N_q$  have zero mean. It is also reasonable to assume that all noise sources are independent and hence the total variance of the signal as a function of gain and exposure time is given by

$$\sigma_I^2(K, t_{\text{exp}}) = K^2 \cdot (\sigma_e^2 + \sigma_d^2 + \sigma_{o_1}^2) + \sigma_{o_2}^2 + \sigma_q^2 \quad (20)$$

$$= K^2 \cdot (\mu_e + \mu_d + \sigma_{o_1}^2) + \sigma_{o_2}^2 + \sigma_q^2 \quad (21)$$

$$= K^2 t_{\text{exp}} \cdot (\eta \cdot F_p + F_d) + K^2 \cdot \sigma_{o_1}^2 + \sigma_{o_2}^2 + \sigma_q^2. \quad (22)$$

Now we have the necessary framework to analyze the proposed acquisition protocol. If we multiply the gain by some constant  $n$  and simultaneously divide the exposure time by  $n$ , from Eq. (19) it becomes apparent that the mean observed intensity indeed stays the same:

$$\mu_I(nK, t_{\text{exp}}/n) = nK \cdot \frac{t_{\text{exp}}}{n} (\eta \cdot F_p + F_d) = \mu_I(K, t_{\text{exp}}). \quad (23)$$

The main reason for this result is that all components of the signal with non-zero mean depend proportionally on the

gain and inversely proportional on the exposure time. For the variance we get a different result

$$\sigma_I^2(nK, t_{\text{exp}}/n) \quad (24)$$

$$= n^2 K^2 t_{\text{exp}}/n \cdot (\eta \cdot F_p + F_d) + n^2 K^2 \cdot \sigma_{o_1}^2 + \sigma_{o_2}^2 + \sigma_q^2 \quad (25)$$

$$= nK^2 t_{\text{exp}} \cdot (\eta \cdot F_p + F_d) + n^2 K^2 \cdot \sigma_{o_1}^2 + \sigma_{o_2}^2 + \sigma_q^2 \quad (26)$$

$$= n \cdot \sigma_I^2(K, t_{\text{exp}}) + (n - 1)(nK^2 \cdot \sigma_{o_1}^2 - \sigma_{o_2}^2 - \sigma_q^2) \quad (27)$$

This result can be interpreted as follows: For pixels with high irradiance the noise is dominated by shot-noise and hence  $\sigma_I^2(nK, t_{\text{exp}}/n) \approx n \cdot \sigma_I^2(K, t_{\text{exp}})$ . The less the irradiance becomes, the higher the weight of the noise before amplification  $\sigma_{iK}^2$  will become and hence  $\sigma_I^2(nK, t_{\text{exp}}/n) \approx n^2 \cdot \sigma_I^2(K, t_{\text{exp}})$ .

#### 4. Residual Errors

The previous results show that our capture protocol is valid and should produce a set of images whose intensity is essentially constant in the mean and whose noise variance increases with increasing gain. However, capturing images in practice does not adhere to this ideal theoretical model. Figure 3b shows the difference image between an image capturing the test target shown in Fig. 3a with the lowest possible ISO value and an image capturing the test target with a high ISO value. From our theoretical analysis in the last section, we would expect the difference image to show the zero mean noise added to the two images. However, the result is not zero-mean. We identified four sources of errors that need to be corrected for in order to relate the intensities of a certain pixel across the different captures:

1. Spatial misalignments during the whole capture procedure induced by small camera vibrations caused by the mechanical shutter. These spatial errors are in the fractions of a pixel.
2. Linear intensity changes due to the fact that scaling the ISO value in the camera does not exactly scale the gain in the same way. The same holds true for the exposure time.
3. The lighting of the scene may change during the capture procedure.
4. In general scenes individual objects may move during the capture procedure.

We now describe how we aim to correct for these errors. In order to reduce the spatial misalignment already during acquisition, we wait for roughly two seconds after an exposure before taking the next image. Correcting the

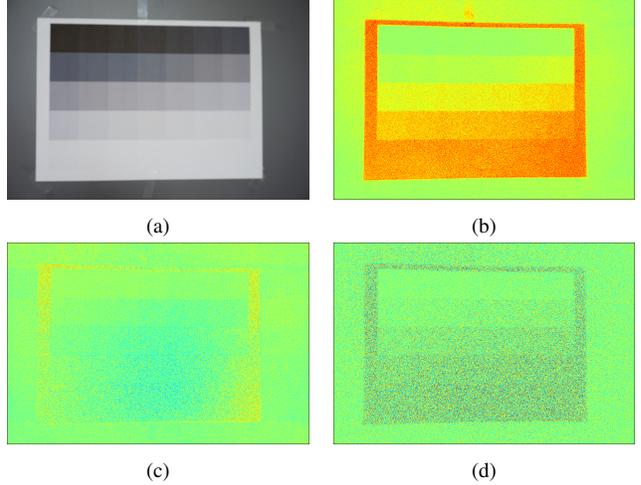


Figure 3. (a) Test target. (b) Difference between high ISO and low ISO image where green means a difference of zero. (c) Difference image after spatial alignment and scale correction and (d) after subsequent low-frequency correction.

residual spatial error is done by employing the method from [10]. It searches for the 2D translation of one image such that the phase correlation to another image is maximized. Subpixel accurate registrations are achieved by upsampling the discrete Fourier transform around the peak of the error landscape. Using a correlation-based alignment has the additional advantage that it is not affected by a linear scaling of one of the input image, in contrast to other measures like Euclidean distance.

Having corrected for the spatial misalignment we next try to undo the linear intensity scaling of the intensity in the high ISO image compared to the low ISO image. To do this, we estimate a linear model that relates the intensities in both images. Then we alter the high ISO image such that this linear effect is accounted for by transforming the intensities in the high ISO image with the estimated linear model. Figure 3c shows the difference image after the linear correction. It can be seen that the intensity dependent bias in the residual is gone.

However, there is still a low frequency pattern on the difference image that we account to small changes in the ambient lighting. To remove this residual bias we add to the high ISO image a low-pass filtered version of the residual image. In detail, we use a broad Gaussian filter with a standard deviation of 20 pixels. Figure 3d shows the final difference image after the low-frequency correction. Now we can see a zero-mean noise image as we expected. Also we see that the variance of the noise increases with the intensity of the mean signal value  $\mu_I$ , as expected.

When applying the acquisition scenario to outdoor scenes, it is likely that some parts of the scene will change due to objects in the scene moving and deforming. In fu-

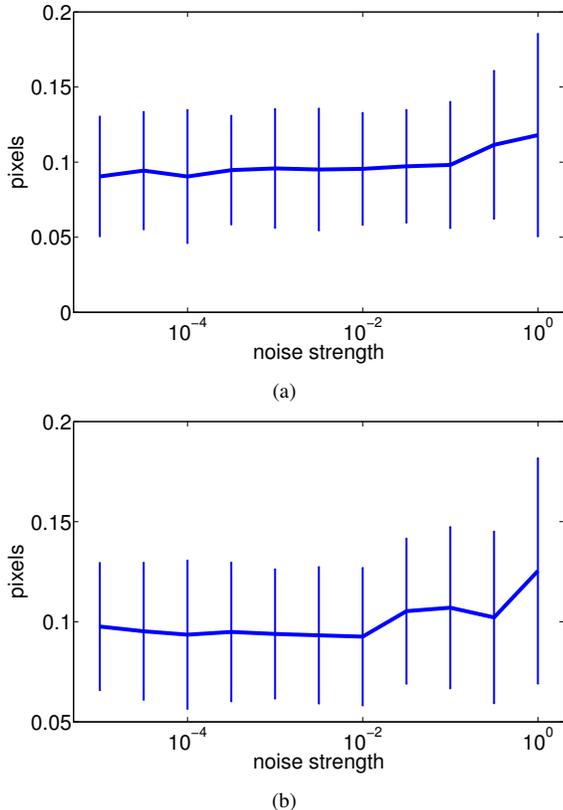


Figure 4. Estimation error of the spatial alignment as a function of the intensity-dependent noise strength. (a) Applying and undoing the translation only or (b) with additional linear intensity scaling.

ture work, we aim to exploit that these scene changes are spatially localized in order to detect them and to exclude them from further analysis. For now we are constraining ourselves to indoor imagery.

## 5. Experiments on Image Noise

In this section we evaluate the post-processing presented in Sec. 4. We also compare noise characteristics estimated from our dataset to the approach from [9].

**Residual errors.** We evaluate the removal of residual errors by simulating the process of spatial misalignments and linear intensity changes. In detail, we sample random translations from a normal distribution with mean zero and standard deviation of 0.3 pixels in both  $x$  and  $y$  direction to represent the spatial misalignment. For the linear intensity changes we apply a linear transformation where the slope is sampled from a normal distribution with mean 1 and standard deviation of 0.002. These values are representative for what we empirically estimated on real data. The translation and the linear scaling are applied to a mosaiced test image. To the transformed images we add Poisson-Gaussian noise

Noise level $[\log_{10}]$	-5	-4	-3	-2	-1	0
S / S	0.0000	0.0001	0.0003	0.0016	0.0531	0.3597
S / T + S	0.0001	0.0001	0.0004	0.0017	0.0539	0.3616
T + S / T + S	0.0147	0.0145	0.0152	0.0162	0.0668	0.3686

(a)

Noise level $[\log_{10}]$	-5	-4	-3	-2	-1	0
S / S	0.0000	0.0000	0.0001	0.0004	0.0177	0.1612
S / T + S	0.0000	0.0000	0.0001	0.0004	0.0180	0.1620
T + S / T + S	0.0041	0.0041	0.0043	0.0045	0.0217	0.1637

(b)

Table 1. RMS errors of linear scale correction.

of different strengths to simulate the effect of capturing the transformed image at a higher ISO value.

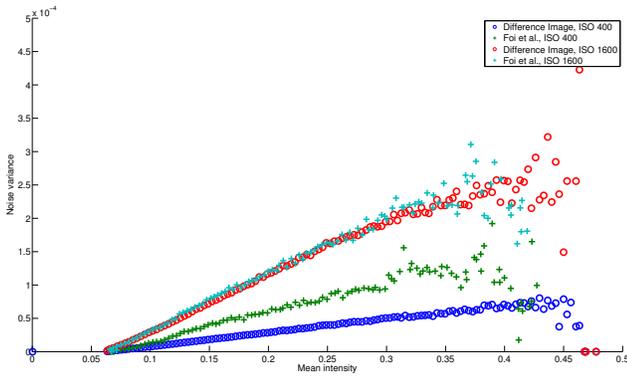
Now we study how well our residual error removal procedure can undo the simulated transformations. First, we look at the spatial registration. Figures 4a and 4b show the mean Euclidean error and standard deviations of recovering the applied translation when applying only the translation or both translation and linear scaling, respectively. As can be seen the estimation error is robust to increasing noise up to a level where the noise variance is equal to the actual intensity. Also, applying the linear scaling to the intensities does not affect the performance, which we contribute to the use of a correlation-based registration method. Interestingly, we found in further experiments that the registration performance decreases when a high-resolution test image is used instead of the low-resolution image of this experiment. We suspect that with little high frequencies the effect of noise will increasingly hurt the registration performance. Resolving this issue is part of future work.

Table 1 shows the root mean squared error of estimating the slope and offset of the linear transformation of intensities. Here, we look at three settings: 1) Applying and undoing just the linear scaling, 2) applying the linear scaling and undoing translation and scaling and 3) applying and undoing both translation and scaling. We make two main observations. First, the error increases again with the noise level, but more drastically than when estimating the translation. Second, estimating the coefficients of the linear transformation gets more difficult when the simulated high ISO image is corrupted by the translation. Hence, future work has to focus on how to robustify this step of our alignment pipeline.

**Noise characteristics.** To show the usefulness of our dataset, we estimate the noise level function on the difference image between a low and high ISO exposure of the scene shown in Fig. 5a. We estimate pairs of mean intensity and noise variances by applying a procedure similar to [9].



(a)



(b)

Figure 5. Comparison of noise-level functions estimated from the difference image between low and high ISO image and using [9] on the high ISO image. The input image is shown in (a) and the noise-level functions in (b).

First, we decompose the low ISO image into approximation and detail wavelet coefficients  $w_{app}$  and  $w_{det}$ , respectively. Then we estimate level sets of the intensity of the low ISO image by binning the approximation coefficients in regular spaced intensity intervals. For each bin we calculate the mean intensity and the variance of the corresponding pixels in the difference image between low and high ISO image.

Figure 5b shows the estimated noise-level functions of our method for a ISO 400 and ISO 1600 image. The low ISO image was taken at ISO 100. We also show the noise-level function estimated with [9] on the high ISO images. According to Eq. (27) the noise variance of the ISO 1600 image should be four times the variance of the ISO 400 image. The noise-level function estimated from the difference image reflects that behavior while the method of [9] overestimates the noise for the ISO 400 image. This emphasizes that the noise estimation based on our dataset is more accurate than estimating the noise just on a single image.

## 6. Conclusion and Outlook

This paper presents a first significant step toward a data set of image noise for noise modeling as well as measuring the performance of noise estimation and denoising algorithms. We presented an acquisition procedure for sets of images, each showing the same scene but with different gains and exposure time. We mathematically argued that all images from the same scene should have the same per-pixel mean intensity and that the variance will increase with the gain. In practice, we encountered residual errors, however. We proposed and partially evaluated a procedure for handling residual errors stemming from spatial misalignment, inaccurate gain changes, as well as lighting changes. Our experiments showed reasonable accuracy of this post-processing, but also revealed that small spatial translations of the camera remain a problem area. In future we thus aim to explore improved spatial alignment methods, e.g. an extension of [17] that can decompose a set of images into a low-rank and sparse error component, where additionally each image undergoes some unknown spatial deformation. We also aim to include more camera models in the acquisition of our planned dataset. Since the most used cameras nowadays are in smartphones, we also aim to use phones featuring the *Camera 2 API* that allows capturing raw images.

**Acknowledgments:** This work was supported by the EU FP7 project “Harvest4D” (no. 323567).

## References

- [1] J. Anaya and A. Barbu. RENOIR - a benchmark dataset for real noise reduction evaluation. *arXiv:1409.8230*, Sept. 2014. 1, 2
- [2] L. Azzari and A. Foi. Gaussian-Cauchy mixture modeling for robust signal-dependent noise estimation. In *ICASSP 2014*, pages 5357–5361. 1, 2
- [3] A. Buades, B. Coll, and J.-M. Morel. Image denoising by non-local averaging. In *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, volume 2, pages 25–28, Philadelphia, Pennsylvania, Mar. 2005. 1
- [4] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian. Image denoising with block-matching and 3D filtering. In *Electronic Imaging '06, Proc. SPIE 6064, No. 6064A-30*, 2006. 1
- [5] A. El Gamal and H. Eltoukhy. CMOS image sensors. *IEEE Circuits and Devices Magazine*, 21(3):6–20, 2005. 2, 4
- [6] European Machine Vision Association. EMVA Standard 1288: Standard for characterization of Image Sensors and Cameras, 2012. 2, 3, 4
- [7] A. Foi. Practical denoising of clipped or overexposed noisy images. In *Europ. Sig. Proc. Conf. 2008*, pages 1–5. 2
- [8] A. Foi, S. Alenius, V. Katkovnik, and K. Egiazarian. Noise measurement for raw-data of digital imaging sensors by au-

- automatic segmentation of nonuniform targets. *IEEE Sensors Journal*, 7(10):1456–1461, 2007. 2
- [9] A. Foi, M. Trimeche, V. Katkovnik, and K. Egiazarian. Practical Poissonian-Gaussian noise modeling and fitting for single-image raw-data. *IEEE T. Image Process.*, 17(10):1737–1754, 2008. 1, 2, 6, 7
- [10] M. Guizar-Sicairos, S. T. Thurman, and J. R. Fienup. Efficient subpixel image registration algorithms. *OSA Optics Letters*, 33(2):156–158, Jan. 2008. 5
- [11] G. Healey and R. Kondepudy. Radiometric CCD camera calibration and noise estimation. *IEEE T. Pattern Anal. Mach. Intell.*, 16(3):267–276, 1994. 2, 3, 4
- [12] G. E. Healy and R. Kondepudy. Radiometric CCD camera calibration and noise estimation. *IEEE T. Pattern Anal. Mach. Intell.*, 16(3):267–276, Mar. 1994. 3
- [13] D. Khashabi, S. Nowozin, J. Jancsary, and A. W. Fitzgibbon. Joint demosaicing and denoising via learned nonparametric random fields. *IEEE T. Image Process.*, 23(12):4968–4981, Dec. 2014. 2
- [14] C. Liu, R. Szeliski, S. B. Kang, C. L. Zitnick, and W. T. Freeman. Automatic estimation and removal of noise from a single image. *IEEE T. Pattern Anal. Mach. Intell.*, 30(2):299–314, 2008. 2
- [15] F. Luisier, T. Blu, and M. Unser. Image denoising in mixed Poisson-Gaussian noise. *IEEE T. Image Process.*, 20(3):696–708, 2011. 2
- [16] M. Mäkitalo and A. Foi. Optimal inversion of the generalized Anscombe transformation for Poisson-Gaussian noise. *IEEE T. Image Process.*, 22(1):91–103, 2013. 2
- [17] Y. Peng, A. Ganesh, J. Wright, W. Xu, and Y. Ma. RASL: Robust alignment by sparse and low-rank decomposition for linearly correlated images. *IEEE T. Pattern Anal. Mach. Intell.*, 34(11):2233–2246, 2012. 7
- [18] O. M. Petteri. Dependence of the parameters of digital image noise model on ISO number, temperature and ashutter time. 2008. 2
- [19] U. Schmidt and S. Roth. Shrinkage fields for effective image restoration. In *CVPR 2014*, pages 2774–2781. 1