J. Eur. Opt. Society-Rapid Publ. 2023, **19**, 27

https://doi.org/10.1051/jeos/2023023
Available online at: https://jeos.edpsciences.org

Journal of the European Optical Society-Rapid Publications

Advancing Society with Light, a special issue from general congress ICO-25-OWLS-16-Dresden-Germany-2022

RESEARCH ARTICLE

OPEN ACCESS

Evaluation of a time-gated-single-pixel-camera as a promising sensor for autonomous vehicles in harsh weather conditions

Claudia Monika Bett*, Max Daiber-Huppert, Karsten Frenner, and Wolfgang Osten

Institut für Technische Optik, Universität Stuttgart, Pfaffenwaldring 9, 70569 Stuttgart, Germany

Received 30 November 2022 / Accepted 20 April 2023

Abstract. We propose a time-gated-single-pixel-camera as a promising sensor for image-free object detection for automotive application in adverse weather conditions. By combining the well-known principles of time-gating and single-pixel detection with neural networks, we aim to ultimately detect objects within the scene rapidly and robustly with a low-cost sensor. Here, we evaluate the possible data reduction such a system can provide compared to a conventional time-gated camera.

Keywords: Time gating, Single pixel camera, Compressed sensing, Neural networks.

1 Introduction

Driver assistance systems need to be robust against a variety of different environmental conditions, especially bad weather conditions such as heavy rain, fog or snow, if we envision them to drive our cars autonomously one day. Current systems rely on many different sensors, such as cameras, lidar, radar etc., which produce a significant amount of data. Nevertheless, the performance in bad weather conditions is still often poor. One promising sensor for bad weather conditions is the so-called time-gated camera [1], which filters ballistic object photons. Recently, time-gated sensors in the near infrared (NIR) have been implemented on vehicles by several research groups [2, 3]. Moreover, gated cameras in the short-wave infrared (SWIR) waveband are developed, which promise a good penetration depth in bad weather conditions [4, 5].

Apart from robustness on the hardware side, we also need to very robustly detect objects in real time, combining as much different sensor data as possible. Although computation power seems to be ever increasing, the live readout and interpretation of different sensor data is still challenging – even more so with modern evaluation algorithms such as neural networks. A reduction in data on the recording side while still maintaining all relevant scene information is therefore highly commendable. Such a measurement system can be understood in the framework of compressed sensing (CS) [6]. One possible implementation for a compressed optical sensor is the single-pixel-camera [7], which has been proven to reduce the amount of data by a factor of up to 50 for natural images [7, 8].

In recent years, some pioneering work towards single-pixel 3D-scene detection was reported in literature. Ren et al. first proposed the combination of a single-pixel-camera with time-gating in 2011 [9], whereafter the principle was further developed [10–12] and better decompositions of the CS optimization problem were found [13]. There also exist several studies concerning very low light conditions [14–16]. Interestingly, in single photon counting mode, the number of photons per pattern becomes the limiting factor [16]. More recently, neural networks gradually replace conventional CS reconstruction algorithms, showing better performance with lower compression ratios [8, 16]. Here, a lower compression ratio means less recorded datapoints. Nevertheless, most of the studies focus on shape measurements by reconstruction of the 3D data cube. Mostly, irradiance images and depth maps are reconstructed out of the single-pixel information with the aim of increasing the (depth) resolution and/or decreasing recording time. Although some authors mention the ability to measure through obscuring media or demonstrate it with the help of one obscuring layer in front of the objects [11, 14], few work has been carried out with extended obscuring media. One notable exception is Bashkansky et al., who report single-pixel measurements of static letter objects through heavy fog in the lab with impressive compression below one percent [17]. Moreover, scenes are usually assumed to be static, which relaxes the constraint on recording time. Howland et al. have reported a video with 14 Hz frame rate of a pendulum moving in 3D space [15] and Quero et al. propose a single-pixel sensor for drones with additional four point indirect time-of-flight sensors [18]. Albeit they prove that the sensor can cope with very high background illumination, the indirect time-of-flight

* Corresponding author: bett@ito.uni-stuttgart.de

approach is not suitable to deal with thick extended media. Bashkansky et al. deal with the problem of a dynamic medium in the case of a static object by high pass filtering their data at the cost of reduced reconstruction quality [17].

In this paper, we investigate the possibility to use a time-gated-single-pixel-camera for autonomous vehicles in harsh environmental conditions. Using a sensor in this setting directly leads to two major problems:

1. Illumination power: Due to the extended intermediate medium between illumination/detector and object, only a very small amount of the illumination power ever reaches the sensor. In order to overcome the inherent detector noise, pulse energy of the illumination system should be high. On the other hand, the overall power needs to be low enough to respect eye-safety norms.
2. Rapid scene changes: Driving inherently implies a highly dynamic environment. Accordingly, high frame rates of 25 Hz or more are mandatory. Moreover, the obscuring medium itself is dynamic, such that we have to deal with additional fluctuations in our single-pixel measurements.

We believe that a low compression ratio is the key to solve both problems. If only some few recordings are necessary, we can either measure fast enough to ensure a quasi-static scene during acquisition or even measure in parallel as it was briefly mentioned in [15] (we will elaborate more on this point in Sect. 3.2). Moreover, the overall optical power we need to feed into the scene will decrease with the compression ratio. Apart from the inherent sparsity of all images, we envision two more mechanisms to further reduce the compression ratio when we combine time-gating with a single-pixel detector. On the one hand, time-gated images are even sparser than natural images due to the conic illumination and the time-gating rendering most objects in the scene invisible for any given delay. On the other hand, a reconstruction of an image is not necessary. Albeit an image helps humans to assess the scene with one glance, the machine needs to extract object information only, such as number and class of objects and their positions, in order to safely guide the car.

The basic idea of image-free classification, i.e. direct classification on the single-pixel signal without reconstruction of an image, was given by Davenport et al. very early after CS-theory was formulated, but was mostly unnoticed at the time [19]. Recently, image-free classification picks up interest again, due to the possibility of using neural network classifiers on the single-pixel signal [20, 21]. Yang et al. even proposed a scheme to classify, locate and reconstruct an image out of the single-pixel signal with one single neural network [22].

This paper contains our preliminary study on the possible data compression we can hope to achieve in a time-gated-single-pixel-camera for autonomous vehicles in harsh weather conditions. We will demonstrate that a time-gated-single-pixel-camera is able to robustly detect objects with a minimum of recorded data, i.e. fast enough to even cope with adverse weather conditions and highly dynamic environments.

We will shortly revise the principles of time-gating and a single-pixel-camera in Section 2 and present some early reconstruction as well as classification results in Section 3. Based on the feasible compression ratio found in Section 3.1, we outline a possible setup and estimate its prowess in Section 3.2. Section 4 will wrap up the paper with the concluding remarks.

2 Theoretical background

A time-gated camera consists of a pulsed laser and a camera with a very fast shutter (opening time typically in the nanosecond range). The camera shutter is triggered such that it opens only for a very short time after a laser pulse was sent at a user-defined delay time. The pulse length is typically much lower than the shutter time or gating time of the camera. Independent thereof, the gate can be expressed as the convolution of the detector gate with the laser pulse:

$$G(z) = \int P(z') \cdot G_D(z - z')\mathrm{d}z'. \tag{1}$$

G denotes the gate, G_D the detector gate and P the illumination pulse. Here, we directly express the gate as a range gate, as the range z is proportional to the transit time t via the velocity of light c: $z = ct/2$. Therefore, photons are filtered according to their path lengths through the medium: Only photons with path lengths corresponding to the delay time arrive at the camera while the gate is open. Mathematically this is expressed by the convolution of the gate with the product of atmospheric attenuation and object or medium reflectivity:

$$S\big(x, y, z_{\mathrm{delay}}\big) = P(x, y) \int \beta(z)(\rho(x, y, z) + \rho_{\mathrm{scat}})G\big(z_{\mathrm{delay}} - z\big)\mathrm{d}z, \tag{2}$$

where $S(x, y, z_{\mathrm{delay}})$ denotes the signal intensity detected in Pixel (x, y) at a range gate distance z_{delay}, $P(x, y)$ is the transverse illumination profile, $\beta(z)$ is the attenuation due to the intermediate medium, ρ the reflectivity of the object and ρ_{scat} the reflectivity of the medium, which for simplicity is here assumed to be constant. If the delay time corresponds to the distance between object and camera, all ballistic photons originating from the object will be registered. Most non-ballistic photons or noise photons are not registered though, due to their different path lengths through the medium (see Fig. 1a). For a rectangular gate, the noise contribution from the medium reduces to $\rho_{\mathrm{scat}} \cdot z_{\mathrm{gate}}$ instead of $\rho_{\mathrm{scat}} \cdot z$ for a conventional camera (see Eq. (2)). We directly see that the smaller the gate z_{gate}, the higher the signal-to-noise ratio (SNR).

In order to get the whole scene information, several images with different delay times need to be recorded. Apart from an enhanced recording time, this leads to more data than is strictly necessary (gated images have a lot of dark pixels, see e.g. Fig. 2).

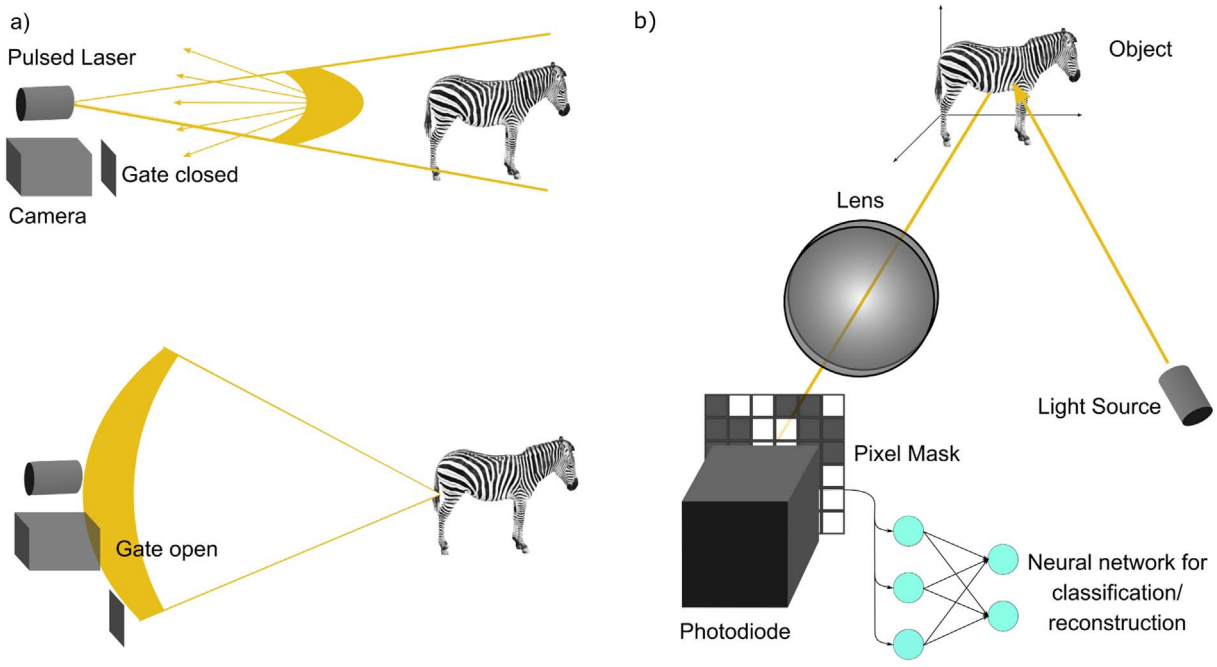

Fig. 1. Operation principle: For time-gating, a laser pulse illuminates the scene. After some delay time t_{delay} the camera shutter is opened for a very short time (nanoseconds). Thereby, the ballistic photons of a certain depth are filtered (a). A single-pixel-camera consists of a photodiode in combination with several (binary) masks (b).

Fig. 2. Examples for image conversion: Original simulated RGB images (top row) and corresponding simulated foggy gated NIR images with active laser illumination (bottom). The original image size is 512 × 512 pixels (13.3° FOV) whereas the gated images are downsampled to 64 × 64 pixels. The attenuation length of the fog was set to 13 m, the gate from 40 m to 60 m, the illumination FWHM to 6.4°, the image blurring to one pixel (gaussian filter) and the background illumination to 1%. RGB images taken from [25].

In a single-pixel-camera, the signal is recorded by a photodiode, which makes it especially suitable in wavelength ranges away from the visible spectrum, where no low-cost cameras exist. The lateral resolution is gained by pixel masks in front of the detector (see Fig. 1b):

$$I_i = \sum_{x,y} M_i(x,y) S(x,y), \tag{3}$$

where I_i is the i-th intensity value registered by the photodiode and M_i represents the i-th pixelmask. Typically, the

masks are projected onto the photodiode with the help of a digital mirror device (DMD) [7]. Due to the inherent sparsity in images [23], the number of masks can be significantly lower than the number of pixels of the reconstructed image. Inherent sparsity can most easily be understood if we consider lossy image compression schemes such as jpeg [24]. The compression ratio is thereby defined as the number of masks K over the total number of pixels N:

$$cr = K/N. \tag{4}$$

In contrast to conventional image compression methods, we want to use neural networks to generate the masks of our single-pixel camera. Thereby, we are not reduced to specific base functions for our compression, but are able to find the optimal basis for our dataset. Moreover, we aim to implement an image-free detection scheme, i.e. we want to directly extract the object information out of the single-pixel signal using neural networks.

Combining equations (2) and (3), the time-gated-single-pixel signal can be expressed as:

$$I_{i,z_{\text{delay}}} = \sum_{x,y} M_i(x,y)P(x,y) \int \beta(z)(\rho(x,y,z) + \rho_{\text{scat}})$$

$$G(z_{\text{delay}} - z)\mathrm{d}z, \tag{5}$$

so that we can evaluate the single pixel signal for each time or range slice z_{delay} separately. As we now use a photodiode as a detector, frame rates of several Gigahertz are feasible. Therefore, the gate can be understood as one time-frame of the photodiode and we can record the whole depth information for one pixelmask with one laser pulse.

3 Results and discussion

In the introduction, we mentioned that a low compression ratio may deliver better preconditions for the real-time data evaluation for autonomous vehicles in harsh weather conditions. Therefore, we carried out several simulation experiments to understand which compression ratio might be feasible for our specific use case. In a first step, we reconstructed images out of the single-pixel information in order to have a reference. As we envision to directly detect objects within the single-pixel signal to further reduce the compression ratio, we additionally trained a classification network.

3.1 Determination of feasible compression ratio via simulations and neural networks

In this section we want to first explain how we simulated an adequate dataset, then proceed to give details about the neural networks and finish with their results.

3.1.1 Creation of dataset

We are not aware of any dataset comprising time-gated images with different delay and gating times in harsh weather conditions. Therefore, we created our own dataset. The data was taken from simulated RGB images of the DENSE dataset [25], for which we employed an algorithm to simulate gated images with fog in the infrared (IR) waveband. For the RGB to IR conversion, we followed Gruber et al. [3]. There, they create IR images out of RGB images by weighting the different color channels such that the visual perception of the new image resembles an image taken with an IR camera. The weighting factors for the different color channels were determined heuristically, such that no clear wavelength dependence can be deduced. We therefore opted to operate our algorithm with their predefined values. The depth data was provided alongside the images by DENSE, such that gated images for different delay and gating times could easily be constructed. We choose a rectangular gating function, which is a good approximation, as long as we are operating with gating times much longer than the pulse width and the rise time of our photodiode. The effect of fog was included by adding noise terms for the noise photons as well as image blurring caused by snake photons. As noise terms, we applied shot noise and Johnson-Nyquist noise. Moreover, the attenuation coefficient β for the signal is calculated following the Beer-Lambert-Law,

$$\beta = \exp(-z/z_a), \tag{6}$$

where z_a represents the attenuation length of the medium. Additionally, an active laser illumination light cone $P(x, y)$ in form of a Gaussian was added and background illumination included. We chose a field of view of 13.3° and downsampled all images to 64 × 64 pixels. Moreover, all datasets were simulated at a range around 50 m with an attenuation length z_a = 13 m, i.e. the objects were situated around four attenuation lengths deep within the medium. We show some exemplary simulated images as well as the original RGB images in Figure 2.

The synthetic DENSE dataset is not labeled. To produce an adequate dataset for the classification task, we added objects of different classes. We fixed the classes to "traffic sign", "human" and "vehicle". A total of ten different object images per class were used, each of these were pasted with a random size at a random position within our background gated images (see Fig. 3 for examples). Therefore, the total number of images amounts to 30 N_{bg}. We created two different datasets, one with a short range interval of 1 m, which doesn't exhibit much background and one with a larger range interval of 15 m and therefore much more background. All relevant dataset parameters are summed up in Table 1.

We split all datasets in 94.5% training data, 5% validation data and 0.5% test data.

3.1.2 Neural network architecture

For the image reconstruction task, we trained an autoencoder-type network. An autoencoder (AE) compresses the data in the encoder part down to a latent vector with size L followed by a subsequent decompression in the decoder to reconstruct the original image [26]. We then used the trained decoder to train our masks such that their single-pixel intensities are linearly mapped into the latent vector space of the autoencoder, i.e. that the original images could be reconstructed from the single-pixel information.

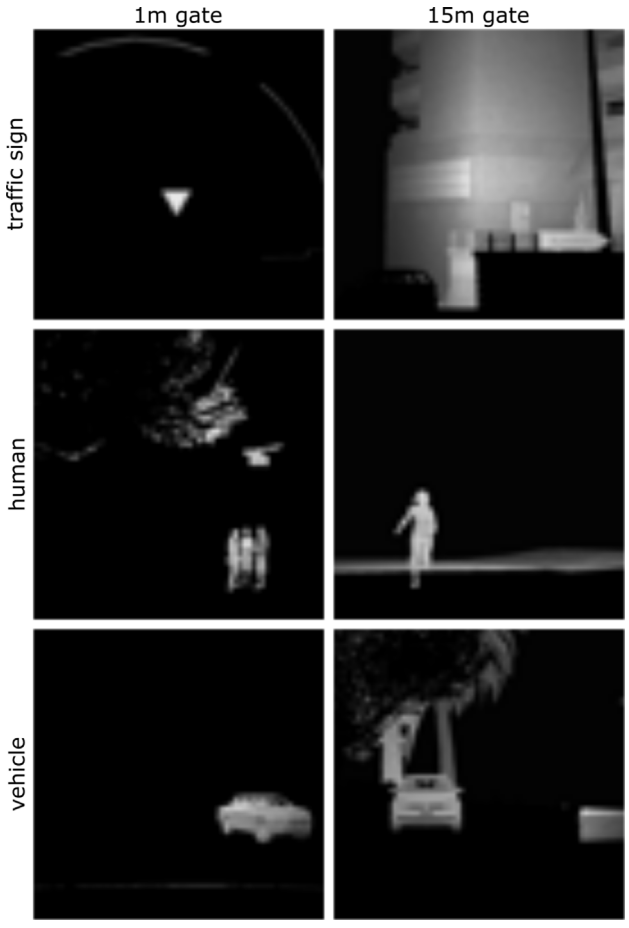

Fig. 3. Examples from the datasets produced for classification: One of ten different object images for each of the three classes "traffic sign", "human" and "vehicle" were pasted with different sizes at a random position within the gated images. We produced two different datasets, one with a gating range of 1 m (left column) and one with a gating range of 15 m around a distance of 50 m. The images with the longer gating range exhibit significantly more background.

Table 1. Simulation parameters for different datasets. N_{bg}: number of background images. ill. = illumination.

Network type	(N_{bg})	Gate (m)	FWHM$_{ill}$ (°)
Reconstruction	7403	20	6.4
Classification	2000	1	13.3
Classification	4000	15	12.5

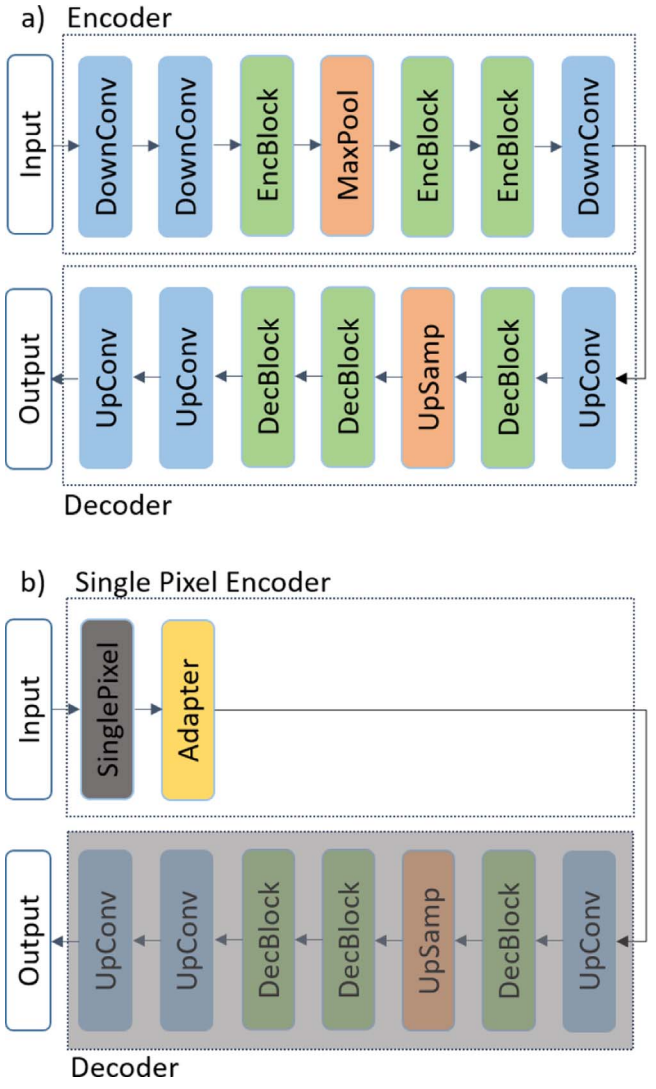

Fig. 4. Schematic drawing of the network architecture of the autoencoder (a) as well as the single-pixel-decoder (b) for 6.25% compression ratio. For lower compression ratios, a down/up-convolutional block as well as an encoder block was added. The encoder/decoder block consists of two convolutional layers with kernel size 3 and 128 filters each as well as a skip connection. As activation function we used LeakyRelu. The weights of the decoder part of the single-pixel-decoder are shared with the autoencoder network and not retrained. DownConv: Convolutional layer with stride 2, EncBlock/DecBlock: Encoder/Decoder block, MaxPool: Maximum pooling layer, UpConv: Convolutional layer followed by a transpose convolutional layer with stride 2, UpSamp: Upsampling layer.

We named this network a single-pixel-decoder (SPD). The mask number K then equals L. We constructed our networks following [8, 26]. A schematic drawing of the network architecture can be found in Figure 4. To generate the single-pixel information, we multiplied the original images with each pixel mask and calculated the sum over all pixels. Thereby, we generate K intensity values as sensor output, one for each mask pattern. As evaluation metric we chose the structural similarity index measurement (SSIM) [27] apart from the mean squared error (MSE), which was used as training loss.

The classification network consisted of two hidden dense layers. The first has 128 nodes, the second 12. As loss function we chose the categorical crossentropy loss and as evaluation metric the accuracy, i.e. the number of correctly classified images over all test images.

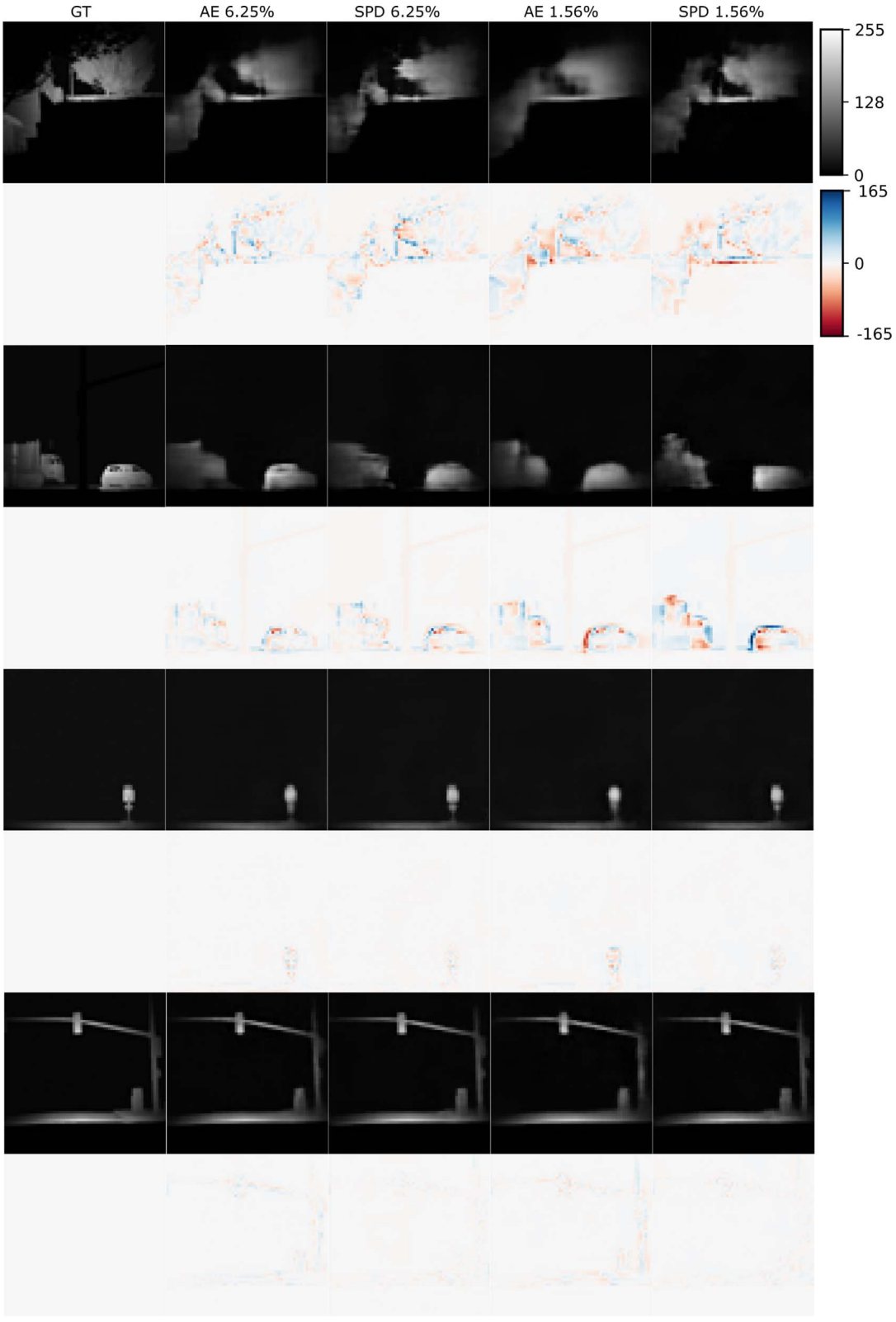

Fig. 5. Examples of reconstruction: Comparison of ground truth (GT) images with reconstructed (rec) results of the autoencoder (AE) and single-pixel-decoder (SPD) networks for two different compression ratios $cr = 6.25\%$ and $cr = 1.56\%$. The first line shows the reconstructed images, whereas the second line shows the difference image $\text{Image}_{\text{rec}} - \text{Image}_{\text{GT}}$. Generally, fine details get neglected for lower compression ratios. Therefore, only the shapes of objects with fine details like e.g. greenery (see first line) get reconstructed. Objects with less high frequency components like traffic signs (last two examples) get reconstructed near perfectly even for the lower compression ratio.

3.1.3 Reconstruction

We trained our reconstruction networks with two different compression ratios $cr = 6.25\%$ and $cr = 1.6\%$. Some exemplary results thereof can be found in Figure 5. The quantitative results for the test data set, i.e. data unseen during training, can be found in Figure 6. While the autoencoder as well as the single-pixel-decoder expectedly perform worse with increasing compression, almost all relevant features stay clearly discernible even with the lower compression ratio of 1.56% which is reflected in the SSIM (see Fig. 6).

Generally, the lower the compression ratio, the more the high spatial frequency components get neglected (see difference images in Fig. 5). This is consistent with CS theory for natural images [23]. Additionally, the network seems to focus more on the brighter part of the images. Please note that the reconstruction of images is not our ultimate goal. In the end, we envision the machine to detect objects directly within the single-pixel information. Therefore, the preservation of relevant features of the object is much more important for us than a pixel-wise accurate reconstruction of the image.

3.1.4 Classification

For three object classification with a short gating interval of 1 m, i.e. negligible background (see Fig. 3 left column for examples), we get a very high accuracy of nearly 100% down to low compression ratios. Performance reduction starts below $cr = 0.5\%$ (see Fig. 7a). Even for $cr = 0.1\%$, which corresponds to only four mask patterns, we can reach a classification accuracy of over 85%. If we increase the gating interval such that background objects are not negligible (see Fig. 3 right column for examples), we see a significant decrease in the overall classification accuracy for all possible compression ratios, even $cr = 100\%$ (see Fig. 7b).

We attribute the mal-classification observed even for the non-compressed signal to our construction of the dataset: The network cannot classify correctly if either the object is pasted unluckily within the background or the background itself confuses the network. Generally, we can observe that the prediction probability, i.e. the probability with which the network associates an image with a specific class, is more ambiguous if the image displays a richer background. One example thereof is given in Figure 8. While the car is correctly classified for an image with simple background (Fig. 8c) with 100% fidelity in the prediction probability even down to 0.5% compression ratio (Fig. 8d), the much richer background in Figure 8a confuses the network even for $cr = 100\%$ (Fig. 8b). We speculate that the background in form of a traffic sign (right bottom corner Fig. 8a) confuses the network for higher compression ratios whereas for lower compression ratios, the features of the tree seem to mimic those of a human (see Fig. 8b). The mal-classification in Figure 8b therefore indicates that our simple network tends to actually learn general features of the classes and not specific ones of the ten objects. This in itself is encouraging but needs further investigation in form of a more sophisticated classification dataset and network architecture.

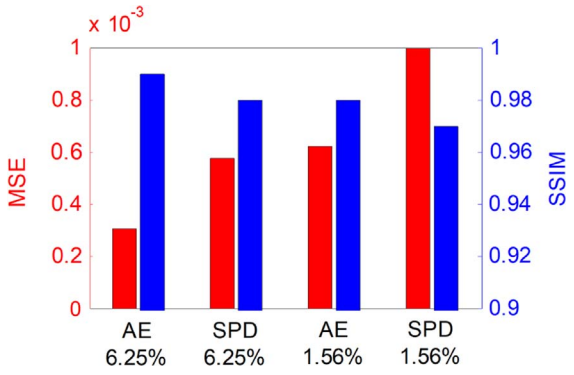

Fig. 6. Quantitative analysis of reconstruction results: The pixel-wise reconstruction performance – expressed by the MSE – decreases with decreasing compression ratio as expected. Moreover, the reconstruction ability of the autoencoder (AE) always outperforms the single-pixel-decoder (SPD). The SSIM on the other hand only decreases slightly with decreasing compression ratio, which indicates that the overall structure of the reconstructed images, i.e. the general shapes of the objects, are unaltered even for the lower compression of 1.56%.

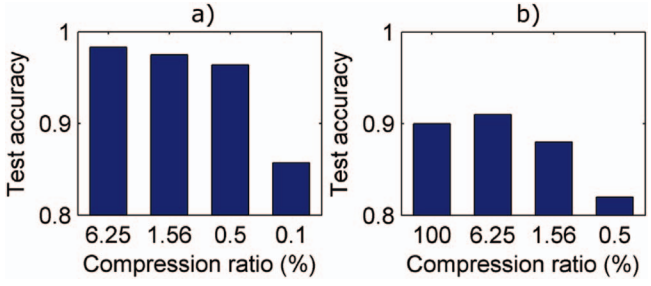

Fig. 7. Quantitative analysis of classification results for short gate of 1 m (a) and longer gate of 15 m (b) around a range distance of 50 m. Whereas the test accuracy is nearly 100% for the short gate down to a compression ratio of 0.5%, it never exceeds 90% for the longer gate, even if the compression ratio is set to one. Nevertheless, the test accuracy only starts decreasing for a compression ratio of 0.5% for the longer gate as well.

3.1.5 Feasible compression ratio

In Table 2, we have summed up the central results from the simulation study concerning a feasible compression ratio. There, we depict the lowest tested compression ratio for which the accuracy for the classification task or the SSIM for the reconstruction task is higher than 95%. The lowest compression ratio of 0.5% is reached for the classification task with a short gating interval. Even for the other tasks though, the compression ratio is around 1.5%. From the results of our simulation study we can therefore deduce, that firstly, for image-free object detection, a short gating interval should be beneficial and secondly that a compression ratio of one percent or lower is feasible.

3.2 Determination of system performance

The results of the last section indicate that a very low compression ratio is sufficient to carry out object detection on

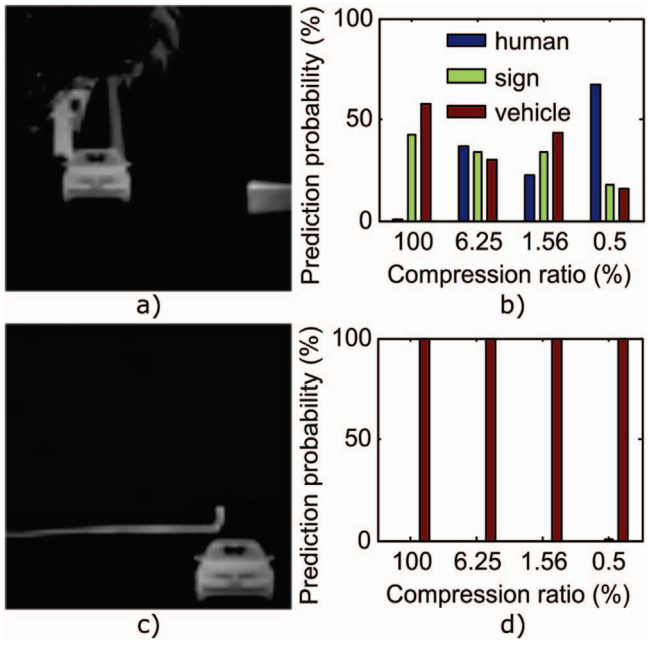

Fig. 8. Two examplary images for one specific vehicle with the long gating range of 15 m and their corresponding prediction probabilities for the three classes: One with significant background (a, b) and one with nearly no background (c, d). For the former, the fidelity of the prediction is low for all compression ratios (b), while for the latter the prediction probability is 100% down to 0.5% compression ratio (d).

the single pixel information. As already discussed in the introduction, this is crucial to deal with the highly dynamic environment as well as keep the illumination power within eye safety constraints. This is near impossible in the visible or near-infrared waveband, we therefore opt for an operating wavelength above 1400 nm. The exact value can be chosen according to the availability of laser sources, for our system we chose 1550 nm. Due to the dynamic environment, we believe a frame rate of $f_{\mathrm{fr}} = 25$ Hz or higher is necessary. In this case, the limiting factor for eye-safety is the total intensity emitted by the system within 10 s [28]:

$$H_{\max} < 10^4 \frac{\mathrm{J}}{\mathrm{m}^2}. \tag{7}$$

The maximum possible energy per measurement can then be calculated to be,

$$E_{\max} = \frac{H_{\max}}{10\,\mathrm{s} \cdot f_{\mathrm{fr}}} \cdot A_{\mathrm{out}}, \tag{8}$$

where A_{out} is the area of the illumination cone once it leaves the housing. If a non-uniform illumination is chosen, it corresponds to an effective area, for a gaussian beam $A_{\mathrm{out}} = 2\pi\sigma^2$. We can now use equation (5) to calculate the energy arriving on the detector. The object is assumed to be lambertian and we additionally allow absorption. We simplify the calculations by assuming a homogeneous illumination with a fill factor $\mathrm{FF}_{\mathrm{ill}}$ compared to the FOV of our camera. The energy per pixel without medium is then distributed via $\mathrm{FF}_{\mathrm{ill}}/N$ with N the total number of pixels. We further neglect the exact pulse form and assume the gate to be small enough that the signal from the object is much higher than the one backscattered from the medium (see introduction). Then the minimal number of photons which must arrive on the detector in order to be registered must be higher than the noise level of the photodiode:

$$\sum_{x,y} M_i(x,y) \cdot E_{\max} \frac{\lambda}{hc} \frac{1}{K} \cdot \frac{\mathrm{FF}_{\mathrm{ill}}}{N} \cdot \rho'(x,y) \exp(-2z/z_a)$$

$$\frac{A_{\mathrm{in}}}{2\pi z^2} \mathrm{QE} > \mathrm{NEP} \sqrt{t_{\mathrm{gate}}} \frac{\lambda}{hc}. \tag{9}$$

Here, ρ' is the pixelwise reflection coefficient considering only absorption as the lambertian nature is expressed via $\frac{A_{\mathrm{in}}}{2\pi z^2}$, where A_{in} is the aperture size of the detector system. QE is the quantum efficiency of the detector and NEP the noise equivalent power of the photodiode.

Let us proceed with an example. According to the simulations in Section 3.1, we set the number of pixels N to 64^2 and the compression ratio to one percent which corresponds to $K = 41$ masks. This directly fixes the pulse repetition rate to $f_{\mathrm{rep\_rate}} = f_{\mathrm{fr}} \cdot K = 1025$ Hz. The gating time is set to 6.7 ns which corresponds to a gate of 1 m (see classification network in Sect. 3.1) and therefore an operating frequency of $1.5 \cdot 10^8$ Hz. The datacube for a range up to 100 m would then be 41×100. As an exemplary photodiode we chose the Hamamatsu 66854-01 [29] with a 2 GHz bandwidth, a NEP of $2 \cdot 10^{-15}$ W/$\sqrt{\mathrm{Hz}}$ and a QE of 95%. We set the diameters of the apertures d_{out} and d_{in} to 50 mm and $z = 100$ m. Then we get:

$$\sum_{x,y} M_i(x,y) \cdot \mathrm{FF}_{ill} \cdot \rho'(x,y) \exp(-200\,\mathrm{m}/z_a) > 1.18 \cdot 10^{-5}. \tag{10}$$

Let us consider for the moment only one illuminated pixel with no absorption and a fill factor of 0.8. Then our system would be able to detect such an object down to 5.6 attenuation lengths. This is comparable to time-gated systems which have been reported to perform down to approximately six attenuation lengths in fog and smoke [17, 30, 31]. If the object is absorptive, the optical thickness of the medium

Table 2. Central results of the simulation study: For each network we depict the lowest compression ratio (cr) for which the metric is higher than 95%. For the reconstruction the metric is SSIM, for classification it is accuracy (Acc). In the case of the longer gating interval, we calculated a weighted accuracy A_w by dividing the accuray of the compressed signal by the accuracy of the signal with no compression $A_w = \frac{A(cr)}{A(cr=100\%)}$. For all networks, the feasible compression ratio is around 1%.

Task	Gate (m)	cr (%)	Metric: value (%)
Reconstruction	20	1.56	SSIM: 97
Classification	1	0.5	Acc: 96
Classification	15	1.56	Acc: 98

needs to decrease accordingly, e.g. for $\rho' = 0.1$ we only get 4.4 attenuation lengths. On the positive side, we took a rather pessimistic view in our calculations: Normally, objects will be much larger than one pixel. Moreover, many materials in traffic are either retroreflective or exhibit a strong backscatter peak, such that the lambertian model does not hold for them. If we want to get more sensitive, we have some degree of freedom in enlarging the output area of the illumination as well as the aperture area of the detection system. One very interesting operation mode would be to record all masks simultaneously instead of sequentially. Instead of using a DMD, the masks could then also be hard coded in front of the individual photodiodes. On the one hand, a large input aperture is then easily realised as the net aperture is $K{\cdot}A_{\mathrm{in}}$. Indeed, instead of increasing the repetition rate, the mask number K then increases the overall sensor area (see also Eq. (9)). On the other hand, we record a true static image. For the frame rate of 25 Hz and a velocity of 100 km/h, objects move over 1 m during one single frame. Obviously, the parallel measurement mode comes at the cost of a more complex calibration routine, as the viewing angles of the different masks will be slightly different.

In our noise calculation, we have estimated the noise floor with the NEP, the true noise floor might be slightly higher due to additional electronics and the rather high operation frame rate. Moreover, we most certainly have got background illumination in the scene. We have not included it in our analysis as we believe it to be small, if a narrowband wavelength filter is used in front of the photodiode. Even if the sun directly shines within our sensor, sun light in the short-wave-infrared (SWIR) region is comparable to our active illumination (0.62 $\mathrm{W/m^2/nm}$ for reference air mass 1.5 spectrum [32]). In heavy obscuring media, also sun light will be heavily attenuated and additionally rather homogeneously distributed such that we ideally only have to deal with a small constant background. One possiblity to reduce background noise is the use of complementary mask patterns [33]. This doubles the masks number but can easily be implemented by measuring boths arms of the DMD in parallel [34]. Another issue with coherent illumination might be speckle noise. We have not considered it in our analysis, as we believe to have enough design freedom in the layout of the detector to decrease the speckle size to below one mask pixel. Moreover, there exist exciting new illumination strategies which not only promise a homogeneous (and even quadratic) illumination profile but also significant speckle noise reduction [35, 36].

In a next step, we plan to implement first a time-gated camera on a vehicle to validate our simulated data with real data. Moreover, we want to implement direct object detection on the single-pixel information, which we are currently working on in another project. Then, all relevant parameters for the time-gated-single-pixel-camera can be fixed and such a sensor system tested.

4 Conclusion

We have introduced the concept of a time-gated-single-pixel-camera as a promising sensor to tackle robust object detection in bad weather conditions for autonomous vehicles. Simulations of the concept in combination with neural networks show good performance. In particular, they prove that as few as 41 masks could suffice. In this case, the masks can either be hard-coded in front of several photodiodes or projected onto them with a single digital mirror device. Due to the live read-out of the photodiodes, a true single-shot detection of the whole scene would then be possible.

Conflict of interest

The authors declare no conflict of interest.

Acknowledgments. We would like to thank the Baden-Württemberg Stiftung gGmbH for the financing of the project.

References

1 Medina A. (1992) *Three dimensional camera and range finder*, US5081530A, United States.
2 Grauer Y., Sonn E. (2015) Active gated imaging for automotive safety applications, in: *Video surveillance and transportation imaging applications*, Vol. **9407**, SPIE, pp. 112–129. https://doi.org/10.1117/12.2078169.
3 Gruber T., Julca-Aguilar F., Bijelic M., Ritter W., Dietmayer K., Heide F. (2019) *Gated2Depth: Real-time dense lidar from gated images*, arXiv. https://doi.org/10.48550/ARXIV.1902.04997, https://arxiv.org/abs/1902.04997.
4 Göhler B., Lutzmann P. (2016) Review on short-wavelength infrared laser gated-viewing at fraunhofer iosb, *Opt. Eng.* **56**, 031203.
5 Willitsford A.H., Brown D.M., Baldwin K., Hanna R.T., Marinello L. (2021) Range-gated active short-wave infrared imaging for rain penetration, *Opt. Eng.* **60**, 013103.
6 Donoho D.L. (2006) Compressed sensing, *IEEE Trans. Inf. Theory* **52**, 1289.
7 Duarte M.F., Davenport M.A., Takhar D., Laska J.N., Sun T., Kelly K.F., Baraniuk R.G. (2008) Single-pixel imaging via compressive sampling, *IEEE Signal Process. Mag.* **25**, 83.
8 Higham C.F., Murray-Smith R., Padgett M.J., Edgar M.P. (2018) Deep learning for real-time single-pixel video, *Sci. Rep.* **8**, 2369.
9 Ren X., Li L., Dang E. (2011) Compressive sampling and gated viewing three-dimensional laser radar, *J. Phys.: Conf. Ser.* **276**, 012142.
10 Li L., Wu L., Wang X., Dang E. (2012) Gated viewing laser imaging with compressive sensing, *Appl. Opt.* **51**, 2706.
11 Sun M.J., Edgar M.P., Gibson G.M., Sun B., Radwell N., Lamb R., Padgett M.J. (2016) Single-pixel three dimensional imaging with time-based depth resolution, *Nat. Commun.* **7**, 12010.
12 Gong W., Zhao C., Yu H., Chen M., Xu W., Han S. (2016) Three-dimensional ghost imaging lidar via sparsity constraint, *Sci. Rep.* **6**, 26133.
13 Li L., Xiao W., Jian W. (2014) Three-dimensional imaging reconstruction algorithm of gated-viewing laser imaging with compressive sensing, *Appl. Opt.* **53**, 7992.
14 Howland G.A., Dixon P.B., Howell J.C. (2011) Photon counting compressive sensing laser radar for 3D imaging, *Appl. Opt.* **50**, 5917.

15 Howland G.A., Lum D.J., Ware M.R., Howell J.C. (2013) Photon counting compressive depth mapping, *Opt. Express* **21**, 23822.
16 Radwell N., Johnson S.D., Edgar M.P., Higham C.F., Murray-Smith R., Padgett M.J. (2019) Deep learning optimized single-pixel lidar, *Appl. Phys. Lett.* **115**, 231101.
17 Bashkansky M., Park S.D., Reintjes J. (2021) Single pixel structured imaging through fog, *Appl. Opt.* **60**, 4793.
18 Quero C.O., Durini D., Ramos-Garcia R., Rangel-Magdaleno J., Martinez-Carranza J. (2020) Evaluation of a 3D imaging vision system based on a single-pixel InGaAs detector and the time-of-flight principle for drones, in: *Three-dimensional imaging, visualization, and display*, Vol. **11402**, SPIE, p. 114020T. https://doi.org/10.1117/12.2558918.
19 Davenport M.A., Duarte M.F., Wakin M.B., Laska J.N., Takhar D., Kelly K.F., Baraniuk R.G. (2007) The smashed filter for compressive classification and target recognition, in: *Computational imaging V*, Vol. **6498**, SPIE, p. 64980H. https://doi.org/10.1117/12.714460.
20 Jiao S. (2018) Fast object classification in single-pixel imaging, in: *Sixth International Conference on Optical and Photonic Engineering (icOPEN 2018)*, Vol. **10827**, SPIE, p. 108271O. https://doi.org/10.1117/12.2502983.
21 Zhang Z., Li X., Zheng S., Yao M., Zheng G., Zhong J. (2020) Image-free classification of fast-moving objects using learned structured illumination and single-pixel detection, *Opt. Express* **28**, 13269.
22 Yang Z., Bai Y.M., Sun L.D., Huang K.X., Liu J., Ruan D., Li J.L. (2021) SP-ILC: Concurrent single-pixel imaging, object location, and classification by deep learning, *Photonics* **8**, 400.
23 Field D.J. (1987) Relations between the statistics of natural images and the response properties of cortical cells, *J. Opt. Soc. Am. A* **4**, 2379.
24 IOS (1994) *Information technology – Digital compression and coding of continuous-tone still images: Requirements and guidelines ISO/IEC 10918-1:1994*, International Electrotechnical Commission (IEC), Genf.
25 driveU. *DENSE dataset*, Universiteat Ulm, Ulm. Access: 15.02.2023, https://www.uni-ulm.de/in/iui-drive-u/projekte/dense-datasets/.
26 Theis L., Shi W., Cunningham A., Huszár F. (2017) *Lossy image compression with compressive autoencoders*, arXiv. https://doi.org/10.48550/ARXIV.1703.00395, https://arxiv.org/abs/1703.00395.
27 Wang Z., Bovik A.C., Sheikh H.R., Simoncelli E.P. (2004) Image quality assessment: From error visibility to structural similarity, *IEEE Transactions on Image Processing* **13**, 600.
28 DIN e.V.(2022) *Safety of laser products - Part 1: Equipment classification and requirements (IEC 60825-1:2014) DIN EN 60825-1:2022-07*, Beuth-Verlag, Berlin.
29 Hamamatsu Photonics K.K. *Hamamatsu InGaAs photodiode 66854-01*, Hamamatsu Photonics K.K., Hamamatsu. Access: 15.02.2023, https://www.hamamatsu.com/jp/en/product/optical-sensors/photodiodes/ingaas-photodiode.html.
30 Christnacher F., Schertzer S., Metzger N., Bacher E., Laurenzis M., Habermacher R. (2015) Influence of gating and of the gate shape on the penetration capacity of range-gated active imaging in scattering environments, *Opt. Express* **23**, 32897.
31 Tobin R., Halimi A., McCarthy A., Soan P.J., Buller G.S. (2021) Robust real-time 3D imaging of moving scenes through atmospheric obscurant using single photon lidar, *Sci. Rep.* **11**, 11236.
32 NREL. *Reference Air Mass 1.5 Spectra*, NREL, Golden. Access: 15.02.2023, https://www.nrel.gov/grid/solar-resource/spectra-am1.5.html.
33 Sun B., Edgar M.P., Bowman R., Vittert L.E., Welsh S., Bowman A., Padgett M.J. (2013) *Differential computational ghost imaging*, Optica Publishing Group, Arlington, Virginia, OSA Technical Digest (online), p. CTu1C.4. https://opg.optica.org/abstract.cfm?URI=COSI-2013-CTu1C.4.
34 Soldevila F., Clemente P., Tajahuerce E., Uribe-Patarroyo N., Andrés P., Lancis J. (2016) Computational imaging with a balanced detector, *Sci. Rep.* **6**, 29181.
35 Laurenzis M., Poyet J.M., Lutz Y., Matwyschuk A., Christnacher F. (2012) Range gated imaging with speckle-free and homogeneous laser illumination, in: *Electro-optical remote sensing, photonic technologies, and applications VI*, Vol. **8542**, SPIE, p. 854203. https://doi.org/10.1117/12.971433.
36 Laurenzis M., Lutz Y., Christnacher F., Matwyschuk A., Poyet J.M. (2012) Homogeneous and speckle-free laser illumination for range-gated imaging and active polarimetry, *Opt. Eng.* **51**, 061302.