Depth of Field Segmentation for Near-Lossless Image Compression and 3D Reconstruction

MAX VON BUELOW, Technical University of Darmstadt, Germany
REIMAR TAUSCH, Fraunhofer IGD, Germany
MARTIN SCHURIG, Fraunhofer IGD, Germany
VOLKER KNAUTHE, Technical University of Darmstadt, Germany
TRISTAN WIRTH, Technical University of Darmstadt, Germany
STEFAN GUTHE, Technical University of Darmstadt, Germany and Fraunhofer IGD, Germany
PEDRO SANTOS, Fraunhofer IGD, Germany
DIETER W. FELLNER, Technical University of Darmstadt, Germany, Fraunhofer IGD, Germany, and Graz University of Technology, Institute of Computer Graphics and Knowledge Visualisation, Austria

Over the years, photometric 3d reconstruction gained increasing importance in several disciplines, especially in cultural heritage preservation. While increasing sizes of images and datasets enhanced the overall reconstruction results, requirements in storage got immense. Additionally, unsharp areas in the background have a negative influence on 3d reconstructions algorithms. Handling the sharp foreground differently from the background simultaneously helps to reduce storage size requirements and improves 3d reconstruction results. In this paper, we examine regions outside the Depth of Field (DoF) and eliminate their inaccurate information to 3d reconstructions. We extract DoF maps from the images and use them to handle the foreground and background with different compression backends making sure that the actual object is compressed losslessly. Our algorithm achieves compression rates between 1:8 and 1:30 depending on the artifact and DoF size and improves the 3d reconstruction.

CCS Concepts: • **Computing methodologies → Image compression**; **Reconstruction**; *Image segmentation.*

Additional Key Words and Phrases: depth of field estimation, near-lossless compression of image data, digital image archiving

1 INTRODUCTION

To achieve state of the art digitization of cultural heritage objects, large datasets are mandatory. For an accurate 3d reconstruction these need to consist of a wide array of different information, mainly consisting of high resolution

Authors' addresses: Max von Buelow, max.von.buelow@gris.tu-darmstadt.de, Technical University of Darmstadt, Karolinenplatz 5, Darmstadt, Germany, 64289; Reimar Tausch, reimar.tausch@igd-extern.fraunhofer.de, Fraunhofer IGD, Fraunhoferstr. 5, Darmstadt, Germany, 64283; Martin Schurig, martin.ralf.schurig@igd.fraunhofer.de, Fraunhofer IGD, Fraunhoferstr. 5, Darmstadt, Germany, 64283; Volker Knauthe, volker.knauthe@gris.tu-darmstadt.de, Technical University of Darmstadt, Karolinenplatz 5, Darmstadt, Germany, 64289; Tristan Wirth, tristan.wirth@gris.tu-darmstadt.de, Technical University of Darmstadt, Karolinenplatz 5, Darmstadt, Germany, 64289; Stefan Guthe, stefan.guthe@gris.tu-darmstadt.de, Technical University of Darmstadt, Karolinenplatz 5, Darmstadt, Germany, 64289 and Fraunhofer IGD, Fraunhoferstr. 5, Darmstadt, Germany, 64283; Pedro Santos, pedro.santos@igd.fraunhofer.de, Fraunhofer IGD, Fraunhoferstr. 5, Darmstadt, Germany, 64283; Dieter W. Fellner, d.fellner@gris.tu-darmstadt.de, Technical University of Darmstadt, Karolinenplatz 5, Darmstadt, Germany, 64289 and Fraunhofer IGD, Fraunhoferstr. 5, Darmstadt, Germany, 64283 and Graz University of Technology, Institute of Computer Graphics and Knowledge Visualisation, Inffeldgasse 16c, Graz, Austria, 8010.

1556-4673/2022/11-ART1 $15.00
https://doi.org/10.1145/3500924

images and other scene parameters. To safely preserve these artifacts digitally and therefore their inherent heritage for generations to come, the storage of copies across the world is essential. This goal however is limited by the sheer amount of necessary data per object. For these reasons, large transfer and storage costs are induced.

We focus on the compression of image data, as it comprises the largest part of the data with well known statistical properties and therefore can yield the most promising storage space reduction results. When capturing images, they are either saved in their camera's proprietary file format or are left uncompressed. Cameras however are not able to produce satisfactory compression rates due to hardware limitations and generic compression algorithms are not suitable for the required compression rates. Other state of the art image compression algorithms suffer from limitations induced by camera noise [21]. This leads to near-lossless approaches that further reduce input data while keeping relevant image parts or frequency spectra lossless.

The current practice of museums and their digital archives is to store cultural heritage image datasets depending on their original captured file format. If only lossless JPEG representations are available, efficient storage is trivial. If, on the other hand, lossless raw image formats are available, there are magnitudes of additional storage requirements. In the latter case, most museums use, a lossy representation, namely JPEG2000 despite of its data loss, to store the data with the negative effect of losing potentially relevant information. In this paper we solve this issue of losing relevant data related to the artifact itself and combine it with the application of standard lossy compression techniques for less important parts of the scene, i.e. the background.

Therefore our algorithm takes advantage of the design of most stereo algorithms. These usually minimize a data term that is derived from pixel similarity at different disparity levels and is sensitive to edges. Homogeneous and blurred regions are regions that do not contain edges nor corners These regions usually occur in the background of images. They are usually approximated by learned image statistics of neighboring pixels [23]. This leads to the observation that homogeneous regions do not explicitly contribute depth information to the reconstruction result and may contribute faulty approximated values to the reconstruction. We present an algorithm that automatically segments the background (i.e. the homogeneous region) from the foreground, which contains important information about the cultural heritage using a Depth of Field segmentation approach. This segmentation mask is then used to compress the image data with standard image compression backends. We also apply the same segmentation on the 3d reconstruction algorithm in order to show that removal of blurry background parts has positive impact on the reconstruction result. This finding additionally supports the claim of lossy compression techniques in these image regions.

2 RELATED WORK

Ditigal Cultural Heritage Preservation. Stephenson [19] reviews an interactive digital image database for cultural heritage artifacts in terms of structure, tools and descriptiveness of included metadata. The application of Politou et al. [15] use the progressive image transmission technique of the JPEG2000 algorithm to efficiently browse and transmit cultural heritage images over the internet in a web browser. Pavlidis et al. [13] give an overview of state-of-the-art algorithms and techniques suitable for 3d reconstructions of cultural heritage artifacts.

Lossy Compression. The most common lossy natural image compression algorithm *JPEG* of Wallace [24] takes a 3x8 bit image and estimates the discrete cosine coefficients of 8x8 pixel patches. Afterwards, it quantizes the coefficients and compresses them using Huffmann coding [8]. The *JPEG2000* algorithm of Christopoulos et al. [7] applies a discrete wavelet transformation to the image and stores quantized coefficients with an arithmetic coder. JPEG2000 also supports lossless image compression by applying a reversible integer wavelet transformation to the image.

Near-Lossless Compression. The *Multiview Image Compression Algorithm* of Battin et al. [2] utilizes inter-view redundancies and exploits the positive-sided geometric distribution between pixels of two neighboring images.

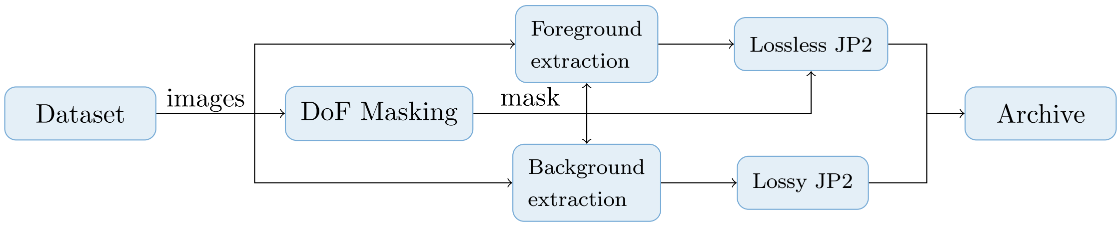


Fig. 1. Overview over the proposed algorithm. First, the images from the dataset are used to create a Depth of Field mask for each image. Then, the mask is used to separate the foreground and the background. The foreground and the mask itself is then encoded using the lossless JPEG2000 algorithm. The background is encoded using the lossy JPEG2000 algorithm. Both compressed images are then saved to the archive.

Aydinoglu and Hayes [1] use the previously computed disparity values to estimate every second frame. They compensate photometric variations using the *subspace projection technique*. The adaptive approach of Perra [14] aims to minimize the entropy of the Differential Pulse-Code Modulation (DPCM) for each block that is small enough that residual image encoding can be omitted. The DPCM coefficients are encoded using the LZMA algorithm. von Buelow et al. [22] use a joint technique of superpixels and figure-ground segmentation in order to apply lossy and lossless compression techniques to the image, depending on the homogeneity of the superpixel content. This work also demonstrates that big noisy background parts still contain enough entropy to make lossless compression inefficient.

Lossless Compression. The PNG algorithm of Boutell [3] evaluates several local filters (i.e. differences) on neighboring pixels, encodes the filter's identifier with the lowest response and compresses its response using Huffmann coding. The PNG algorithm is able to encode images with up to four color channels and a color depth of 32 bit and can therefore be used to encode raw camera sensor data. The compression algorithm of von Buelow et al. [21] uses a wavelet-based compression scheme and re-arranges the Bayer pattern into different color channels. The results are evaluated on dedicated cultural heritage datasets.

Image Masking and Depth of Field Estimation. The figure-ground segmentation algorithm *GrabCut* of Rother et al. [16] segments images into multiple regions that have similar color distributions by iteratively refining user-annotated color distributions and enforcing local homogeneous labelling with a Markov Random Field. Nasse [12] estimates the Depth of Field with respect to camera, lens properties and the focus distance. The standard ISO/TS 19264-1 [9] defines metrics for image sharpness analysis that usually differs with focus. Burns [6] specifies an approach that estimates sharpness with slanted edge image targets.

3 ALGORITHM

The main idea of our algorithm is loosely based on the idea of von Buelow et al. [22]. It distinguishes between the image *foreground* and the image *background*. The *foreground* contains regions that contain relevant information, i.e. the recorded artifact. Therefore, it should be compressed losslessly in order to keep relevant information. In contrast to that, the *background* contains regions that are less or not relevant at all. These can be compressed by lossy techniques without negative effects on the performance of a 3d reconstruction algorithm. This differentiation between *foreground* and *background* constitutes an effective compromise between storage comsumption and originality of the resulting representation.

This section is structured as follows. Section 3.1 describes a lossless image representation that is able to compress digital camera raw files using the JPEG2000 standard. In the following, this representation is extended to use lossy techniques in the background. Therefore, the algorithm computes a *Depth of Field* (DoF) segmentation

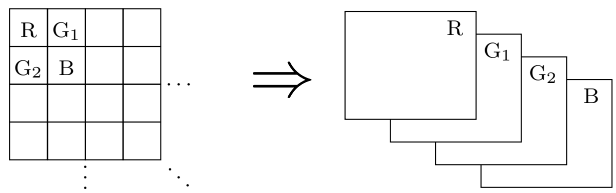


Fig. 2. The Bayer pattern groups four neighboring pixels that can be re-arranged into independent color channels [21].

(section 3.2). This segmentation is fundamental for the following compression step as is effectively separates the lossless part from the lossy part. Section 3.3 describes the algorithm's actual encoding step, which introduces an additional technique to compress the background part with lossy techniques. Finally, section 3.4 shortly describes the decompression. Figure 1 gives an overview over the process.

3.1 Lossless Master-File Representation

The highest quality image representation that is usually obtainable from digital cameras is the so called *raw format*, which is an ambiguous term used for different vendor-specific proprietary image representations. These image representations have in common that they support high dynamic range images, i.e. bit depths of between 8 bit and 16 bit, that are not affected by any post-processing routines.

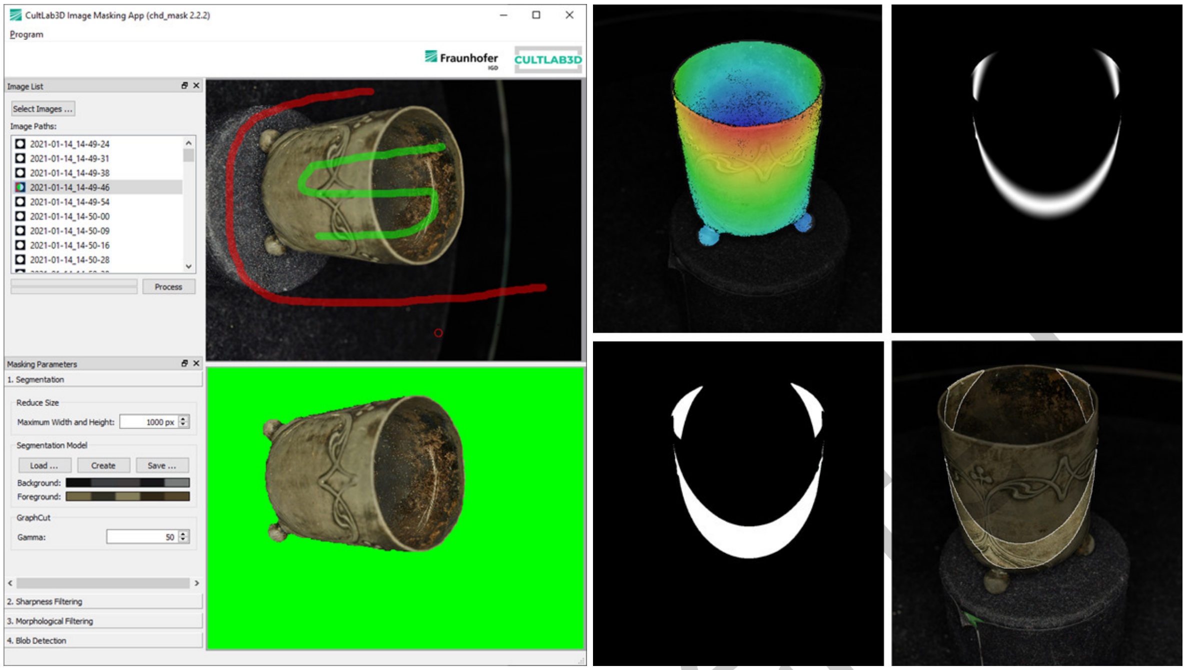
In the following we handle images in a capturing dataset independently and do not make use of any inter-view redundancies due to its inferior performance in high-quality image datasets [21].

Compression Backend. Generally, our algorithm design allows using arbitrary state of the art image compression algorithms that support encoding 16 bit images with at least 4 color channels losslessly. In contrast to the work of von Buelow et al. [22], we decided to use the JPEG2000 algorithm, as its lossless wavelet coding technique generally achieves superior compression rates compared to the PNG algorithm. Although the JPEG2000 format is in general less popular and wide-spread compared to other image formats, it has a high acceptance in the cultural heritage domain [5, 11]. A second advantage is that JPEG2000 also has a lossy mode that uses an irreversible mother wavelet and quantization. Therefore, JPEG2000 is also beneficial for the background encoding, which is described later.

Image Encoding. As most cameras encode color information using the Bayer pattern, our algorithm first re-arranges the pixel data in order to ensure that neighboring pixels originate from the same type of color filter. This is important, as image compression algorithms highly depend on redundancies that come from similar adjacent pixel values. Therefore, the lossless image data is encoded by dividing the Bayer pattern of the camera into four color channels as illustrated in fig. 2. This re-arranged representation is fed into tho JPEG2000 compression backend. In order to ensure the lossless mode in JPEG2000, we disable the *multiple component transform* option and set the target compression rate to 1:1. Additionally, we perform for testing purposes a pixel-wise comparison of the decompressed image and the original image data in order to ensure the reversibility of the compression.

3.2 Image Masking

Image masking describes the process of foreground and background segmentation and extraction, where the foreground contains only the region of interest and the background is usually removed. The background often consists of the blurred environment behind the object and more challenging, the surface or mount attached to the foreground object. Especially in the cultural heritage domain, fragile artifacts are often mounted on and stabilized by their individually designed stands. Concerning the process of photogrammetric 3d digitization,



(a) (b)

Fig. 3. (a) A graphical user interface for defining initial scribbles for color masking and (b) steps from our fully automatic depth-based masking using depth of field projection.

these mounts may partially occlude an artifact's surface. This issue can be resolved by repositioning the artifact eighter in the same or a different mount, e.g. by placing it upside down. Then the complete surface can be captured and later aligned throuhout multiple scan passes. To prevent the mount structure, which then has an inconsistent positioning relative to the artifact, from confusing 3d reconstruction algorithms it is carefully masked out together with the background. Furthermore, this enhances the runtime performance of photogrammetric algorithms by reducing the amount of data to be processed. This is a tedious but common image preprocessing step for photogrammetry that can enhance the resulting 3d model while reducing the computation time. Image masks restrict computationally expensive steps during the 3d reconstruction, (such as feature detection, dense image matching and texture mapping,) to only those image regions containing parts of the object of interest. Especially when the artifact had to be repositioned throughout multiple scan passes, the utilization of image masks is essential for a clean registration between the different image sets. During this feature based alignment process across multiple passes, masking can remove the confusion with otherwise reappearing visible structures, such as the surface of a table or mount under the object, as we show later in the result section. Therefore, most state-of-the-art photogrammetry software solutions, such as Agisoft Metashape, offer the option to provide binary masks along with the actual image set as input. We analyzed two different types of input masks in combination with image compression: color-based and depth-based masks.

Color-Based Masking. To automatically mask image sets of arbitrary size suited for photogrammetric 3d reconstruction the Competence Center For Cultural Heritage Digitization [17, 18] of the Fraunhofer Institute for Computer Graphics Research developed an image masking application that requires only little initial user input.

The approach relies on differences in color and contrast and is based on GrabCut [16], that extends the original image segmentation by graph cut [4] with an iterative energy minimization. It is carried out as follows.

With scribbles the user approximately marks foreground and background on one or more images of the set. From the marked areas a Gaussian Mixture Model (GMM) on multiple dimensions from color spaces, such as RGB, Lab and HSV, is formed and iteratively refined on the marked images. Figure 3a shows the initial user input on one image and on the left side the retrieved means of the typically five components of the GMMs for background and foreground visualized in RGB. This refined segmentation model can be saved and then applied to all other images of the set. Optionally, the aggressiveness or connectivity of the masking can be adjusted by additionally applying a morphological filter chain of subsequent erosion and dilation operations. The resulting blobs can as well be filtered by size and position.

Datasets prepared with the CultArm3D scanning station [20] are normally easy to mask by colors because the station contains a turntable to place the object on and a black or white solid background shield.

Depth of Field Masking. Especially visible in macro photography and caused by a narrow depth of field (DoF), parts of objects become blurry and lead to blurry textures on the 3d model. To address this problem, masked foreground regions can further be reduced to only their sharp parts by thresholding images with respect to depth, more precisely, the distance from the focus plane. The focus plane is parallel to the image sensor and placed on the object of interest. It can also cut through the object causing a depth mask estimation as shown in fig. 3b. This projection of a DoF mask on the original image can be automatically performed if camera intrinsics, extrinsics and 3d object surface information is available. Therefore, we use an incremental approach, where first the object is in parts reconstructed without using image masks. Then masks are produced by thresholding rendered depth images of the reconstructed surface from the aligned corresponding camera perspectives. By using a fixed focus distance and a prior calibration step for retrieving precise camera instrinsics, the rendered masks overlay with the corresponding original images. Therefore, they can be used for incremental refinement, masking and compression.

To ensure optimal DoF masking, not only the camera angle must be precisely aligned, but also the DoF itself must be properly sized around the focus plane. As the distance from this plane increases, the image quality usually decreases. The depth of field describes the acceptably sharp area around this plane. With a properly selected DoF the human eye is not able to detect any difference between the smallest recognizable sharp dot at optimal focus distance and its blurred version at near and far plane [12].

The maximum size of this blurred dot at near and far plane is defined by the circle of confusion. In addition to the focus distance s and the circle of confusion c, the depth of field depends on the focal length f and the used aperture A and can be approximated as follows.

$$\mathrm{DoF} = \frac{sf^2}{f^2 - Ac\,(s-f)} - \frac{sf^2}{f^2 + Ac\,(s-f)} \tag{1}$$

In eq. 1 the minuend describes the far plane and the subtrahend the near plane. The two planes are the farthest and nearest still acceptable sharp planes, respectively. In order to obtain the final segmentation mask for compression and 3d reconstruction, we apply a threshold to this depth masks that converts the depth map to a binary label image that represents the foreground and background regions.

Depth of Field and Lossy Compression. To describe the quality differences more accurately within the depth of field, the modulation transformation function (MTF) is used in ISO/TS 19264-1 [9]. MTF shows resolution and contrast information simultaneously so that a camera system can be evaluated based on the requirements for a specific application. Summary metrics can be used to condense and simplify the comparability of MTF. The two most common metrics are MTF50 and MTF10. MTF50 generally reflects the overall details in an image

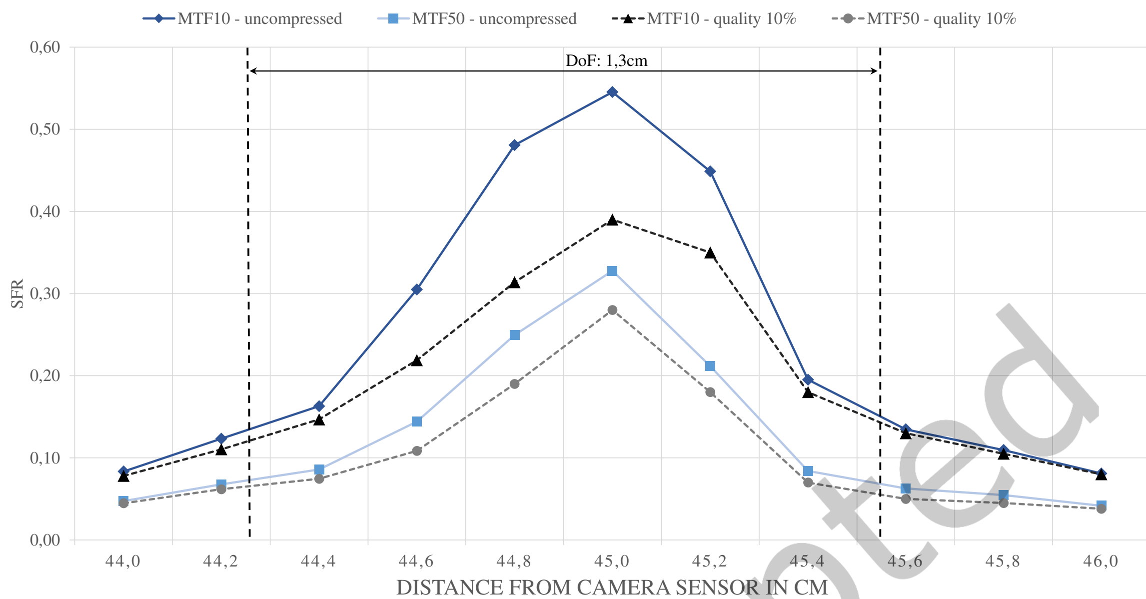


Fig. 4. Evaluated MTF quality metric on a captured SFR target at different distances around a fixed focus distance of 45 cm. For comparison, each image is processed with and without compression.

well. MTF10, the limiting resolution, depicts the smallest detail in an image. Thus, MTF10 captures higher image frequencies, while MTF50 captures average image frequencies.

Figure 4 shows MTF50 and MTF10 for the complete depth of field of 1.3 cm determined in 2 mm steps around the optimal fixed focus distance of 45 cm. At each distance step, one image of an SFR target [6] was captured with and without using image compression and then evaluated for their MTF summary values. In this example, a Phase One iXG 100MP camera with a Schneider RS 72mm/iXG lens was used. A high JPEG compression at 10 % quality was optionally applied, to emphasize the effect of image compression on image quality.

Within the depth of field, a clear influence of the compression on the image details can be observed. Especially MTF10, which captures fine details, is strongly influenced by the compression. However, while approaching the near and far plane, the curves without and with compression converge. The more blurred the image areas become, the smaller is the negative influence of compression on the image. Especially outside of the depth of field, there is hardly any difference noticeable. This supports our approach of applying lossy image compression only on masked areas of the image that are outside the depth of field.

3.3 Near-Lossless Representation

Lossy image compression algorithms generally aim to remove small frequency changes that do not contribute to the image, e.g. noise. A problem lies in the distinction between noise and sharp edges, which have similar statistics. As regions outside the DoF do not contain any edges and this is less useful for the 3d reconstruction as mentioned in section 3.2, it is save to apply lossy techniques to these regions.

In this section, we therefore extend the lossless representation from section 3.1 to handle foreground and background parts separately. First, we introduce an additional channel to the lossless representation that encodes a binary label for each pixel that identifies foreground and background parts. Additionally, we set each pixel from the background part to zero. This way, we retain the adjacency information within the foreground while saving storage space in the background as the underlying arithmetic coder automatically encodes these zero values with negligible cost. Similarly, the introduction of the label channel will have little effect on the compression rates as these binary labels are locally very homogeneous.

Background Compression Backend. While it is also possible to use an arbitrary image compression algorithm for the background, the choice of the lossy compression technique is specific to the application. We chose the JPEG2000 algorithm for two reasons. Firstly, the JPEG2000 algorithm has both lossless and lossy capabilities and thus does not introduce a further dependency. Secondly, the compression and quantization of wavelet coefficients is theoretically equal to blurring images. Since the background is already physically blurred we do not expect negative implication other than noise reduction.

Background Encoding. Similar to the foreground encoding, we set each pixel of the foreground to zero in order to filter out parts that are already compressed losslessly. Unfortunately this technique produces artifacts at the segmentation boundary as the quantization of high detail wavelet-coefficients effectively blurs sharp edges at the mask boundary and the zeroed out foreground pixels influence the neighboring background pixels. In order to address this issue we apply a morphological erosion onto the mask which makes sure to include a support-region to the background. We chose the support-regions size according to the number of wavelet levels (5 by default) used in the JPEG2000 algorithm.

3.4 Decompression

The decompression step of our algorithm works as follows. First, the lossless foreground region and the segmentation mask of the image are decompressed using the reverse standard image compression algorithm used in section 3.3. While decompressing, our algorithm already takes care of extracting the mask that was stored in a separate channel to a separate data structure. Now, the lossy image is decompressed and its decoded pixel values are written to the decompressed image, where the mask denotes a background flag. This way, the algorithm effectively decompresses the faulty but required support region from section 3.3, but it is later ignored as it is already represented by the foreground part.

4 RESULTS

In section 4.1 we give an overview of our datasets that are captured with cultural heritage image acquisition scanners used for further evaluation. In section 4.2 we evaluate the compression rates and run-time performance of our algorithm on the datasets with different types of masks. Section 4.3 evaluates errors on 3d reconstructions given the same maskes used for compression. This section also give a short insight into the results of the lossy background compression.

4.1 Datasets & Segmentation Results

We evaluated our compression algorithm on four multi-view geometry datasets acquired with the CultArm3D scanning station of the Competence Center Cultural Heritage Digitization of the Fraunhofer Institute for Computer Graphics Research [17, 18, 20]. These depict the *Tin Cup* (610 images), the *Apple* artifact (300 images) and the *Mars Venus Amor* dataset (1620 images) on the one hand. These were scanned with a Phase One iXG 100MP 11608×8708 pixel camera and a Schneider RS 72mm/iXG lens. On the other hand, there is the *Java Gold: Royal Ascetic* artifact (540 images), which was scanned with the Canon EOS 5DS R 8688×5792 pixel camera and a Canon EF 100mm f/2.8L Macro IS USM lens. Both cameras capture with a color depth of 14 bit.

The *Tin Cup* data set shows a cup originating from the Art Nouveau. The *Apple* dataset is a non-antique testing dataset. The Mars, Vernus and Amor dataset was scanned at the Museumslandschaft Hessen Kassel (mhk), Germany, in October 2020 the depicted statue was carved by the German sculptor Leonhard Kern in the 17th

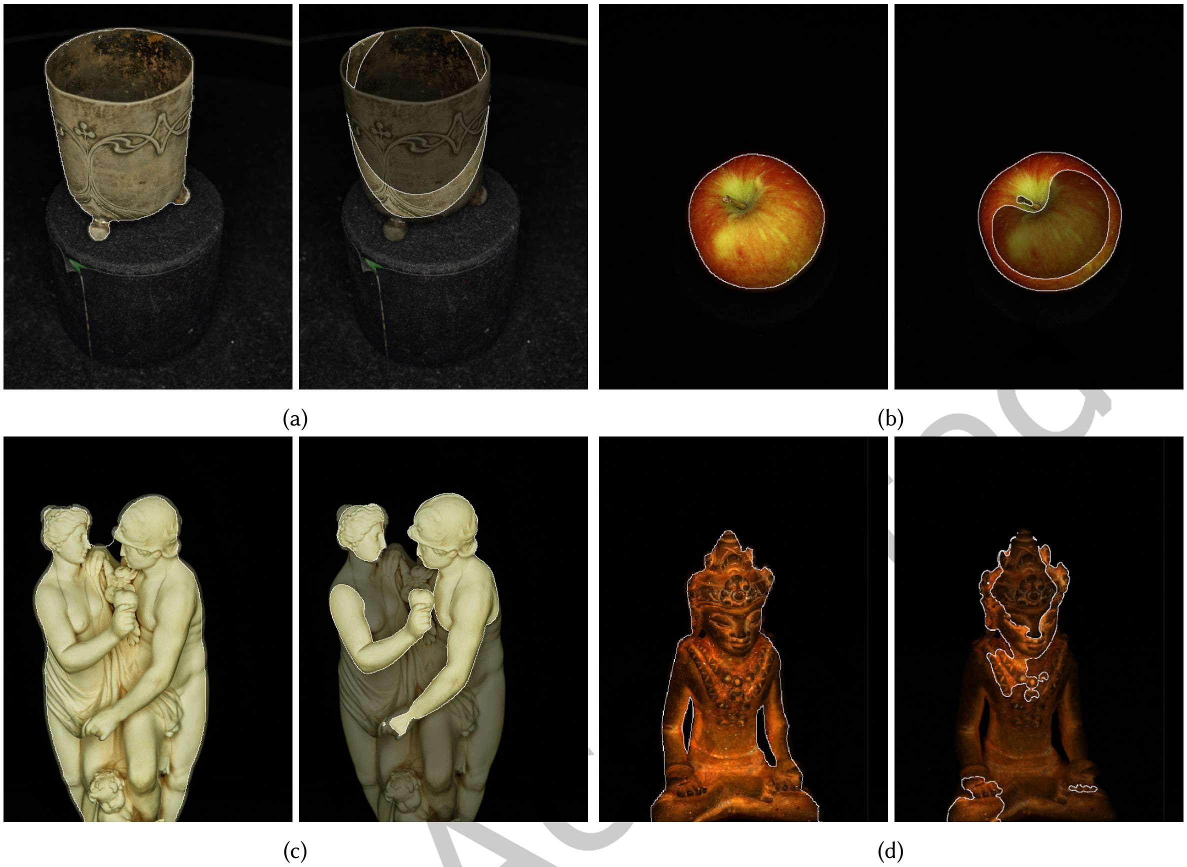


Fig. 5. Segmentation results based on color (left) and DoF (right) on representative example images from the following datasets. (a) shows the Tin Cup, (b) the Apple, (c) shows the Mars Venus Amor dataset and the Royal Ascetic is shown in (d). The highlighted areas in white show the foreground area.

century. The *Java Gold* dataset was captured at the Reiss Engelhorn Museum (rem) in Mannheim, Germany in August 2019 and shows a meditating royal ascetic found at the island Java in Indonesia.

Figure 5 illustrates these datasets by example images and their respective DoF based image masks. These image masks give an impression of the proportion of the image foreground that is in focus. They show that large parts of the foreground are out of focus. These images visualize how our approach has a significant potential for a reduction of losslessly compressed image area compared to a segmentation based on fore- and background. Furthermore, the example image of the Royal Ascetic shows, that these masks may be of complex shape, raising requirements for specialized algorithms. Nevertheless, the user has always the option to visualize the masks with our segmentation app which makes our approach transparent and verifiable.

Generally, for our application, the segmentation result is defined to be *good*, if the reconstruction result improves and the compression rate increases. Both criteria are elaborated in the following sections and limited in both directions. If we remove too much information in the image, compression rates become superior and

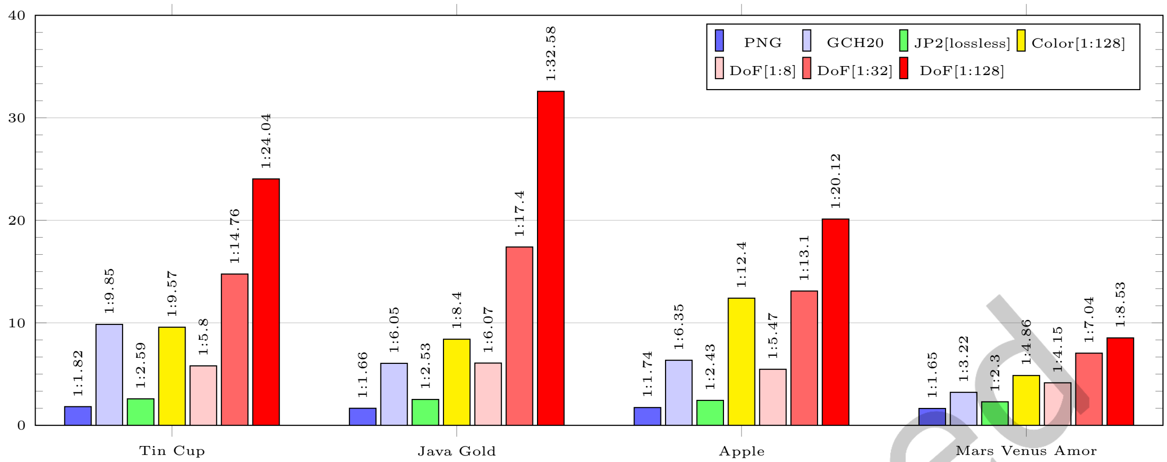


Fig. 6. Compression rate comparison of the PNG algorithm, lossless JPEG2000 and our approach with color and DoF masks. DoF masks are evaluated with lossy background target rates 1:8, 1:32 and 1:128. Color masks are evaluated with lossy background target 1:128 as a comparision.

reconstruction quality becomes inferior. If we remove too less information both criteria become inferior as the data retain more entropy and outliers influence the reconstruction result negatively.

4.2 Compression Rate & Run-Time Performance

As mentioned in section 1, currently there is no specialized image compression implementation for cultural heritage image datasets or other stereo datasets except from our GCH20 algorithm [22]. Digital archives tend to overcome the issue of storage by discarding potential important image data by applying lossy image compression algorithm. Therefore, the JPEG2000 algorithm became the de-facto standard for archiving.

We evaluated the previously described datasets on different compression algorithms. First, we used the PNG algorithm as it is a wide spread lossless compression standard that is capable to encode raw images as it supports 16 bit images. Based on the PNG algorithm, we also compare against our former cultural heritage image compression algorithm *GCH20* [22]. Similar to that, we then present the plain backend compression standard JPEG2000 we use for our current implementation. Finally, we compare the data with our near-lossless approach on different masks and rate configurations on the background. We use the DoF masks and rate configurations 1:8, 1:32 and 1:128 to evaluate how our algorithm performs with different aggressive background compression. Secondly, we evaluate the color masks with a 1:128 lossy target compression rate configuration to compare how different masks result in the final compression rate.

As a comparison metric, we use the outcoming compression rate $r = \frac{s_c}{s_u}$ between the resulting storage size s_c of the individual compressed image dataset and the uncompressed representation's size s_u in bit. In the following we format the rate r as $r = 1{:}\frac{1}{r}$ for better interpretability. The uncompressed size s_u is derived from the image dimensionality ($w \times h$), number of channels (c, usually 4 for the bayer pattern) and the bit depth b and computed as follows.

$$s_u = w \cdot h \cdot c \cdot b \text{ bit} \tag{2}$$

Results show that the lossless version of JPEG2000 outperforms the lossless PNG algorithm on every dataset and requires about 75 % storage consumption. Lossless JPEG2000 has a compression rate of 1:2.45 on average and performs similarly on our datasets. However, our near-lossless algorithm performs differently on each dataset.

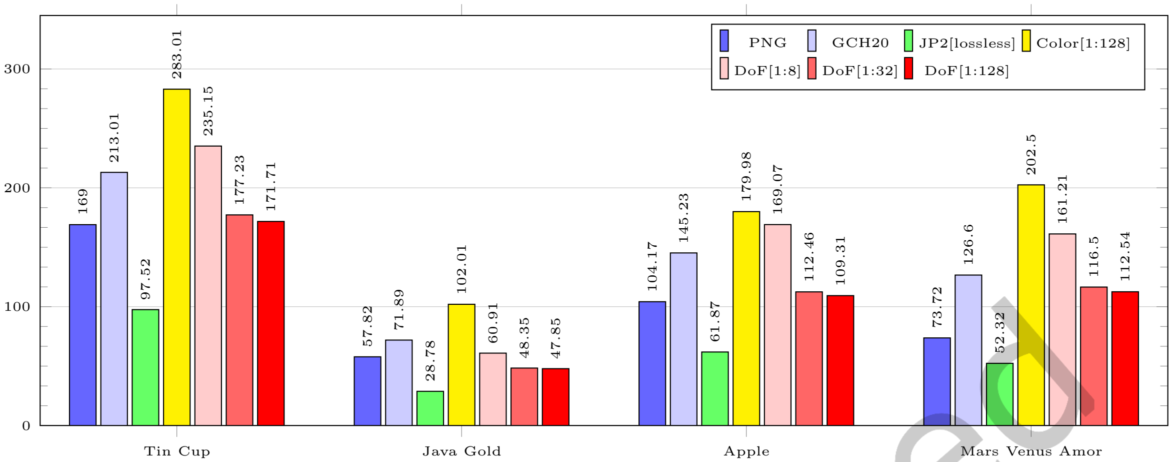


Fig. 7. Run-time performance comparison of the PNG algorithm, lossless JPEG2000 and our approach with color and DoF masks under different background rates. Numbers are in minutes.

We achive a compression rate between 1:6 and 1:24 on *Tin Cup*, between 1:6 and 1:32 on *Java Gold*, between 1:5 and 1:20 on the *Apple* and between 1:4 and 1:9 on *Mars Venus Amor*. This can be explained by different artifact sizes that result in different foreground-background coverage. If the foreground occupies a higher fraction of the image, the compression rate will converge with the lossless JPEG2000 algorithm. The *Mars Venus Amor* dataset also has a bigger DoF, which makes its foreground masks bigger. The comparison with color masks that use a rate configuration of 1:128 shows, that using color masks is as efficient as compressing with DoF masks at a rate configuration of 1:16 in most cases. When comparing with our GCH20 algorithm, it is on par with our DoF masks with a 1:8 target compression rate. The color mask results are consistent with our GCH20 algorithm but slightly better because of the usage of the more efficient JPEG2000 algorithm. Figure 6 shows the compression rates of the selected compression algorithms on our datasets.

Run-Time Performance. We compare the run-time performance of the PNG standard, our previous implementation, the JPEG2000 standard and our approach. Therefore, we use the OpenJPEG implementation and libpng. Figure 7 shows our run-time performance measurements. These measurements were taken on a *2.60 GHz Intel Xeon E5-2650 v2* CPU. Generally, our algorithm achieves run-time performance similar to PNG, whereas lossless JPEG2000 is approximately twice as fast. Our previous GCH20 implementation is slightly less performant than PNG on the presented datasets which corresponds to the findings in our previous work. Nevertheless, run-time performances are usually not a limiting factor for archiving purposes.

4.3 Reconstruction Quality

We visually analyzed the quality of the resulting 3d reconstruction on the four datasets from fig. 5 with the different masking approaches in combination with the image compression and without any further manual post-processing. As a quality criteria the point cloud confidence and the texture sharpness are used.

Point Cloud Confidence. Concerning photogrammetry, the computationally expensive processing step that completes the 3d geometry before the meshing and texturing, is the dense point cloud generation. This step is based on depth maps calculated using dense stereo matching on neighboring camera pairs. Depth maps are calculated for the overlapping image pairs considering their relative exterior and interior orientation parameters

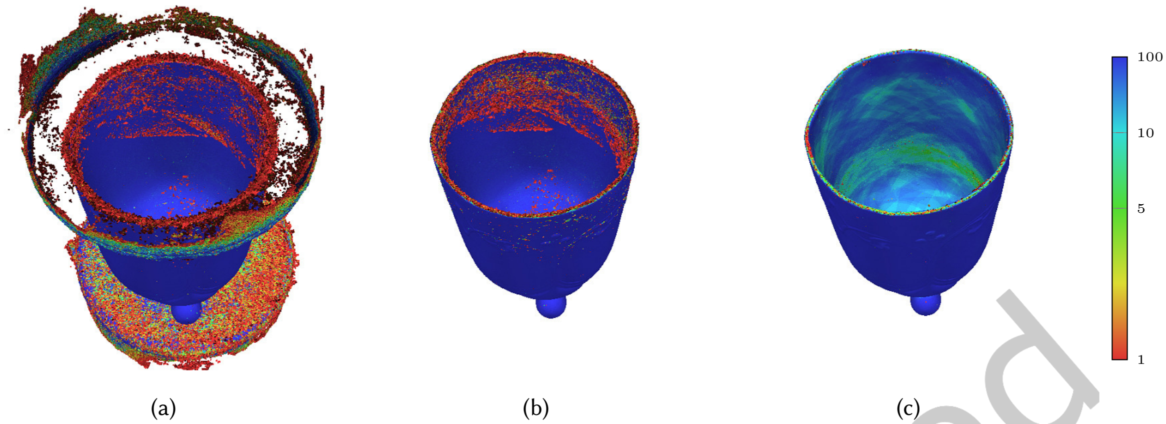


Fig. 8. The confidence point cloud of the Tin Cup dataset reconstructed with different masks. (a) shows the confidence without any masking. The reconstruction of (b) used a color mask and (c) used a DoF mask.

estimated with prior bundle adjustment step. Multiple pairwise depth maps generated for each camera are merged together into a combined depth map, using excessive information in the overlapping regions to filter wrong depth measurements. For every point in the final dense point cloud the number of contributing combined depth maps is recorded and stored as a confidence value [10]. Thus, the confidence value describes the quantity, how often a point was seen and triangulated. It is important to notice, that it does not describe the precision of the triangulation. For example, many unsharp camera images of the same point could lead to a high confidence, although the local precision would be low.

Figure 8 shows the confidence point cloud on the Tin Cup dataset. The dataset was captured in two passes by repositioning the object once upside-down. If no masking was used during the 3d reconstruction, then the dense point cloud contains a lot of noise and background features. It can be observed that the mount surface, was partially reconstructed as well, introducing noise at bottom and top of the object. If color masking was used, then background noise is removed, while the confidence on the object of interest remains. This visualizes the positive effect of early image masking on the photogrammetric 3d reconstruction process as discussed in section 3.2. However, complex areas that are partially self-occluded or located around edges, are still affected by noise. This noise can be further eliminated by applying the DoF masking refinement. The confidence of the point cloud surface inside the cup then drops because the DoF masking further reduces the size of the image areas used for feature extraction and triangulation to the sharp regions of the image. That leads to an effective reduction of the number of triangulations for a certain point. This effect is less visible on the surface outside of the cup, because it is visible on both repositioned passes and thus has a higher overlap and redundancy. If a sufficient coverage and overlap is provided by the dataset, then the effect of an aggressive DoF masking is advantageous. Our experiments show that noise during the photogrammetric reconstruction process is not only introduced by the background, but also by unsharp image areas that are part of foreground but outside the DoF. Hence, the effect of lossy image compression outside the DoF is negligible for the dense point cloud, because those areas are not supposed to be used for the final 3d reconstruction result due to their proneness to noise.

Figure 9 illustrates the influence of DoF masking and color masking on the point cloud confidence. It can be observed that masking effectively reduces low confidence points that are often considered as outliers or point cloud noise (fig. 8), while increasing the amount of points with sufficient confidence. This is because masking reduces the feature extraction to only the sharp surface parts on the object and eliminates confusion with the background, jumping edges and unsharp parts. In the unmasked examples the confidence, i.e. triangulation counter, can reach confidence values greater than 100 because unsharp features are also matched multiple times. However, those triangulations are of lower precision and thus do not improve a already saturated dense cloud quality. For example, if a single point was already precisely triangulated from more than 10 sharp images, matching it with another 100 additional unsharp images will not improve, if not decrease, the point's average precision. As for the Tin Cup, the results show that even after cutting away the oversaturated part of the histograms the completeness of the point cloud is still ensured (fig. 8). Furthermore, the impact of masking is strong on oversaturated simple models as the Apple (fig. 9c) and less significant on complex models as the ivory Mars Venus Amor (fig. 9d), where on both sides, low confidence and oversaturated confidence are only slightly reduced.

Texture Quality. Similar to the confidence, the used masks also have an effect on the final texture projected onto the mesh generated from the dense point cloud. The projection of the texture is done in two steps. First, low frequencies from different images are blended together to prevent edge formation and visible seems. Then, high frequencies are taken from a selected single image that is as perpendicular as possible to a given surface. Here, image areas outside the DoF are also likely to be selected. However, as shown in figure fig. 4, these areas do not always contain high frequencies and thus can lead to blurred textures. This can be observed in the close-up comparison of figure fig. 10b. In the left image, generated with color masks, there are significantly fewer details than in the right image, generated with DoF masks. Here the low frequencies outside the DoF are masked out and successfully excluded from the detailed texture mapping.

Our experiments show that the application of DoF masking leads to a significant increase of texture details. As discussed in section 3.2, those fine details are prone to lossy image compression, but are preserved by our locally lossless compression approach.

Lossy Background Compression. Figure 11 shows a representative example image from the *Tin Cup* dataset that is marked in the unsharp area of the image. Figure 11b and fig. 11c show the same 800 %-zoomed area compressed with different background rate configurations 1:8 and 1:128. Although, both zoomed regions are very similar, it can be observed that less noise remains in the 1:128 version because more blurring was applied to it. This demonstrates the influence of the background rate configuration and shows that the influence on unsharp parts is very small. However, the compression performance is superior.

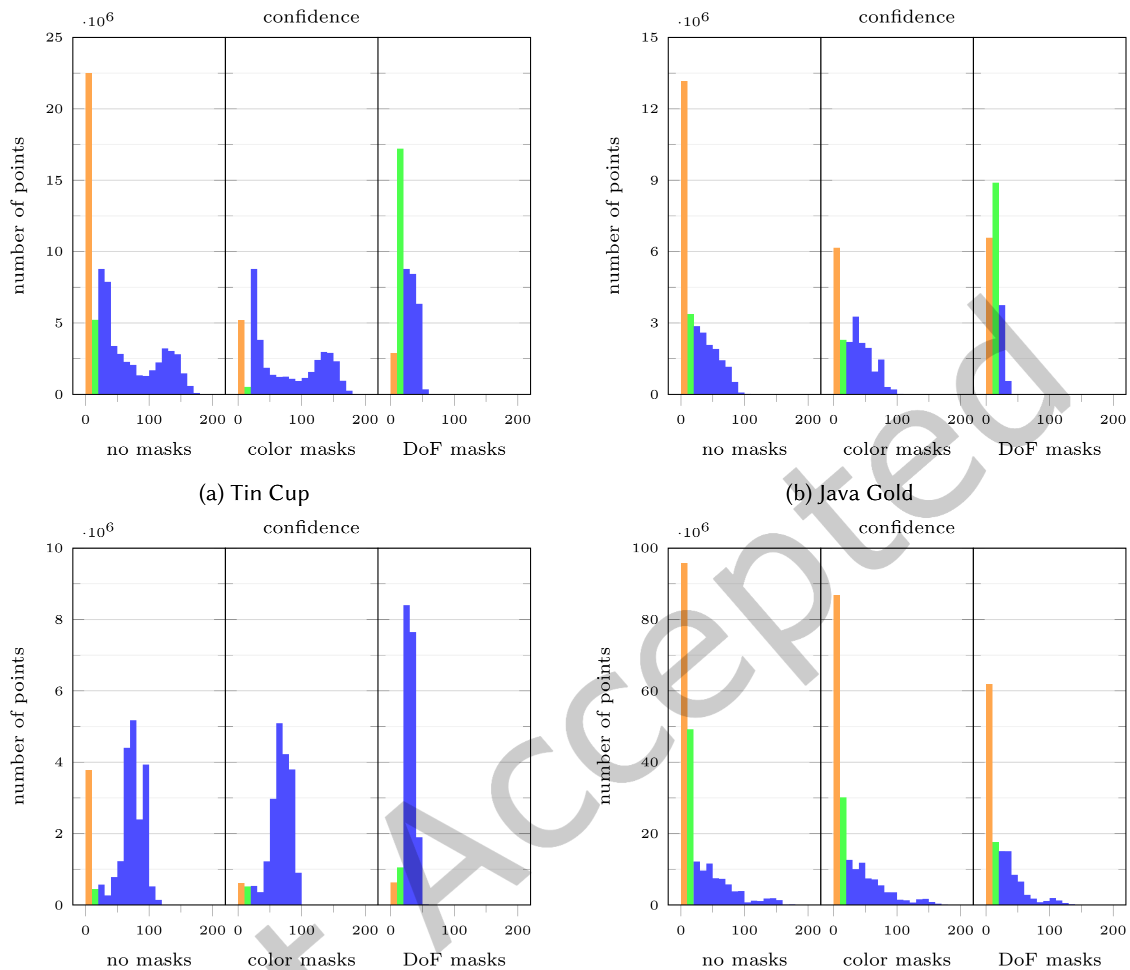
5 CONCLUSION

In this paper, we presented a near-lossless compression algorithm based on a Depth of Field segmentation that makes use of unsharp background parts of cultural heritage artifact images and showed the positive effect on 3d reconstructions. Our algorithm uses the standard JPEG2000 algorithm as compression backend and is therefore easy to implement. It achieves compression rates of between 1:8 and 1:30 compared to an uncompressed representation and can handle raw image files without demosaicing and other pre-processing. Further, the algorithm can work with different rate configurations for lossy background compression and allows a lossless mode. Generally it can be said that homogeneous and unsharp regions are hard to compress using standard lossless image compression algorithms due to noise but are perfectly suited for lossy compression algorithms. Handling these regions differently also improves the overall accuracy of 3d reconstructions on these images.

Future Work. In the future we would like to further evaluate our algorithm outside of the cultural heritage domain as similar use-cases exist in other areas.

SOURCE CODE

The source code for this paper is available at https://github.com/maxvonbuelow/maskcomp.



(a) Tin Cup (b) Java Gold (c) Apple (d) Mars Venus Amor

Fig. 9. Histograms showing the confidence values from the point clouds reconstructed from our datasets with no masks applied, with color masks and DoF masks. Orange bars highlight low confidence points with 1 to 10 matches, green bars are points with sufficient confidence with 11 to 20 matches and all other confidence values are marked in blue.

ACKNOWLEDGEMENTS

Part of the research in this paper was funded by DFG (Deutsche Forschungsgemeinschaft) project 407 714 161. We thank the anonymous reviewers whose comments helped improve this manuscript.

REFERENCES

[1] Haluk Aydinoglu and Monson H Hayes. 1994. Compression of multi-view images. In *Proceedings of 1st International Conference on Image Processing*, Vol. 2. IEEE, 385–389. https://doi.org/10.1109/ICIP.1994.413597

[2] Benjamin Battin, Philippe Vautrot, and Laurent Lucas. 2010. A new near-lossless scheme for multiview image compression. In *Stereoscopic Displays and Applications XXI*, Andrew J. Woods, Nicolas S. Holliman, and Neil A. Dodgson (Eds.), Vol. 7524. International Society for

(a) (b)

Fig. 10. The textured model mapped with different masks. (a) shows a comparison overview of the entire object - on the left with color masks and on the right with DoF masks. (b) shows the same comparison in a close-up view.

(a) Overview (b) 1:8 (c) 1:128

Fig. 11. Effect of different background target compression rates options for the JPEG2000 format. (a) shows an overview image from the Tin Cup dataset. (b) and (c) show the same 800 % zoomed area at the marked point in (a), where the outside-DoF area of (b) was compressed with a rate of 1:8 and (c) with a rate of 1:128.

Optics and Photonics, SPIE, 572 – 582. https://doi.org/10.1117/12.838908

[3] Thomas Boutell. 1997. *PNG (portable network graphics) specification version 1.0.* Technical Report. https://doi.org/10.17487/RFC2083

[4] Y. Y. Boykov and M.-P. Jolly. 2001. Interactive graph cuts for optimal boundary region segmentation of objects in N-D images. In *Proceedings Eighth IEEE International Conference on Computer Vision. ICCV 2001*, Vol. 1. 105–112 vol.1. https://doi.org/10.1109/ICCV.2001.937505

[5] Paolo Buonora and Franco Liberati. 2008. A Format for Digital Preservation of Images. *D-Lib magazine* 14, 7/8 (2008), 1082–9873. https://doi.org/10.1045/july2008-buonora

[6] Peter Burns. 2000. Slanted-Edge MTF for Digital Camera and Scanner Analysis. *Society for Imaging Science and Technology: Image Processing, Image Quality, Image Capture, Systems Conference*, 135–138.

[7] Charilaos Christopoulos, Athanassios Skodras, and Touradj Ebrahimi. 2000. The JPEG2000 still image coding system: an overview. *IEEE transactions on consumer electronics* 46, 4 (2000), 1103–1127. https://doi.org/10.1109/30.920468

[8] David A Huffman. 1952. A method for the construction of minimum-redundancy codes. *Proceedings of the IRE* 40, 9 (1952), 1098–1101. https://doi.org/10.1007/BF02837279

[9] ISO/TS 19264-1 2017. *Photography — Archiving systems — Image quality analysis — Part 1: Reflective originals*. Standard. International Organization for Standardization, Geneva, CH.

[10] Agisoft LLC. 2020. *Agisoft Metashape User Manual - Professional Edition, Version 1.6*. Technical Report. Agisoft LLC.

[11] D-Lib Magazine. 2011. JPEG 2000 for Long-term Preservation: JP2 as a Preservation Format. *D-Lib Magazine* 17, 5/6 (2011). https://doi.org/10.1045/may2011-vanderknijff

[12] Hubert Nasse. 2010. *Depth of Field and Bokeh*. Technical Report. Carl Zeiss Camera Lens Division.

[13] George Pavlidis, Anestis Koutsoudis, Fotis Arnaoutoglou, Vassilios Tsioukas, and Christodoulos Chamzas. 2007. Methods for 3D digitization of cultural heritage. *Journal of cultural heritage* 8, 1 (2007), 93–98. https://doi.org/10.1016/j.culher.2006.10.007

[14] Cristian Perra. 2015. Lossless plenoptic image compression using adaptive block differential prediction. In *2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 1231–1234. https://doi.org/10.1109/icassp.2015.7178166

[15] Eugenia A Politou, George P Pavlidis, and Christodoulos Chamzas. 2004. JPEG2000 and dissemination of cultural heritage over the Internet. *IEEE transactions on image processing* 13, 3 (2004), 293–301. https://doi.org/10.1109/TIP.2003.821348

[16] Carsten Rother, Vladimir Kolmogorov, and Andrew Blake. 2004. "GrabCut": Interactive Foreground Extraction Using Iterated Graph Cuts. *ACM Trans. Graph.* 23, 3 (2004), 309–314. https://doi.org/10.1145/1015706.1015720

[17] Pedro Santos, Martin Ritz, Constanze Fuhrmann, and Dieter Fellner. 2017. 3D mass digitization: a milestone for archeological documentation. *Virtual Archaeology Review* 8, 16 (2017), 1–11. https://doi.org/10.4995/var.2017.6321

[18] Pedro Santos, Martin Ritz, Reimar Tausch, Hendrik Schmedt, Rafael Monroy, Antonio De Stefano, Oliver Posniak, Constanze Fuhrmann, and Dieter W. Fellner. 2014. CultLab3D - On the Verge of 3D Mass Digitization. In *Eurographics Workshop on Graphics and Cultural Heritage*, Reinhard Klein and Pedro Santos (Eds.). The Eurographics Association. https://doi.org/10.2312/gch.20141305

[19] Christie Stephenson. 1999. Recent developments in cultural heritage image databases: directions for user-centered design. (1999).

[20] Reimar Tausch, Matevz Domajnko, Martin Ritz, Martin Knuth, Pedro Santos, and Dieter W. Fellner. 2020. Towards 3D Digitization in the GLAM (Galleries, Libraries, Archives, and Museums) Sector. Lessons Learned and Future Outlook. *The IPSI BgD Transactions on Internet Research* 1 (January 2020).

[21] Max von Buelow, Stefan Guthe, Martin Ritz, Pedro Santos, and Dieter W. Fellner. 2019. Lossless Compression of Multi-View Cultural Heritage Image Data. In *Proceedings of the Eurographics Workshop on Graphics and Cultural Heritage (GCH)*. https://doi.org/10.2312/gch.20191343

[22] Max von Buelow, Reimar Tausch, Volker Knauthe, Tristan Wirth, Stefan Guthe, Pedro Santos, and Dieter W. Fellner. 2020. Segmentation-Based Near-Lossless Compression of Multi-View Cultural Heritage Image Data. In *Eurographics Workshop on Graphics and Cultural Heritage*, Michela Spagnuolo and Francisco Javier Melero (Eds.). The Eurographics Association. https://doi.org/10.2312/gch.20201294

[23] Martin J Wainwright and Eero P Simoncelli. 2000. Scale mixtures of Gaussians and the statistics of natural images. In *Advances in neural information processing systems*. 855–861.

[24] Gregory K. Wallace. 1991. The JPEG Still Picture Compression Standard. *Commun. ACM* 34, 4 (1991), 30–44. https://doi.org/10.1145/103085.103089