Segmentation-Based Near-Lossless Compression of Multi-View Cultural Heritage Image Data

Max von Buelow[1] , Reimar Tausch[2] , Volker Knauthe[1] , Tristan Wirth[1] ,
Stefan Guthe[1,2] , Pedro Santos[2], and Dieter W. Fellner[1,2,3]

[1]Technical University of Darmstadt, Germany
[2]Fraunhofer IGD, Germany
[3]Graz University of Technology, Institute of Computer Graphics and Knowledge Visualization, Austria
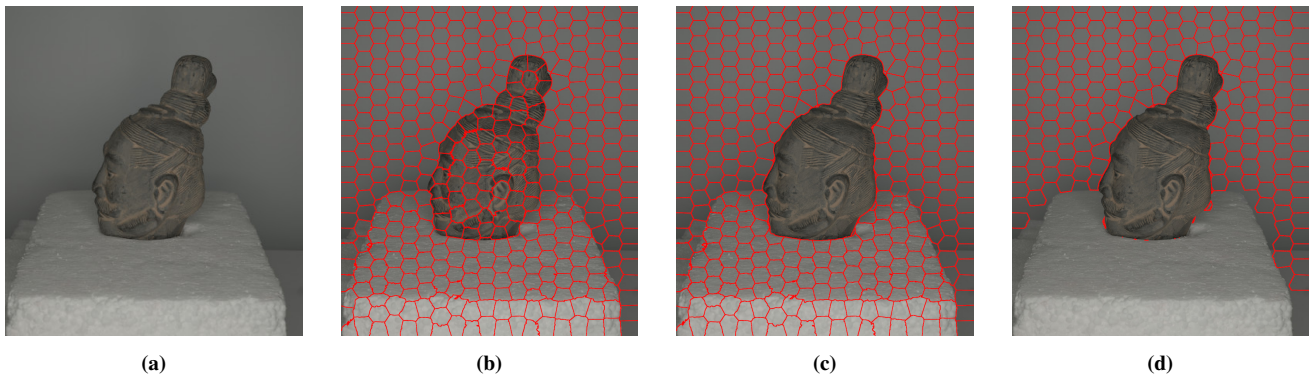
| (a) | (b) | (c) | (d) |

Figure 1: *Steps of our near-lossless compression pipeline. Input image (a) is segmented into superpixels (b) using the SLIC0 algorithm. In (c) the figure-ground segmentation mask is applied to the superpixelization result. (d) shows the final results of homogeneous superpixels that can safely be approximated using planes.*

Abstract

Cultural heritage preservation using photometric approaches received increasing significance in the past years. Capturing of these datasets is usually done with high-end cameras at maximum image resolution enabling high quality reconstruction results while leading to immense storage consumptions. In order to maintain archives of these datasets, compression is mandatory for storing them at reasonable cost. In this paper, we make use of the mostly static background of the capturing environment that does not directly contribute information to 3d reconstruction algorithms and therefore may be approximated using lossy techniques. We use a superpixel and figure-ground segmentation based near-lossless image compression algorithm that transparently decides if regions are relevant for later photometric reconstructions. This makes sure that the actual artifact or structured background parts are compressed with lossless techniques. Our algorithm achieves compression rates compared to the PNG image compression standard ranging from 1:2 to 1:4 depending on the artifact size.

CCS Concepts
• Computing methodologies → Image compression; Image segmentation; Image representations;

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 images and other scene parameters. To safely preserve these artifacts digitally and therefore

their inherent heritage for generations to come, redundant copies across the world are 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 [BGR*19]. This leads to near-lossless approaches that further reduce input data while keeping relevant image parts or frequency spectra lossless.

Our algorithm takes advantage of the design of most stereo algorithms. They usually minimize a data term that is derived from pixel similarity at different disparity levels and is sensitive to edges. Homogeneous regions are regions that do not contain edges or corners and usually occuring in the background of images. They are usually approximated by learned image statistics of neighboring pixels [WS00]. This leads to the observation that homogeneous regions do not explicitly contribute depth information to the reconstruction result. We present an algorithm that automatically detects homogeneous regions in a semi-supervised form and encodes them as parameters of an approximated plane while keeping remaining parts of the image untouched. This is extremely efficient for big homogeneous regions as planes can be stored using only three floating point numbers plus one noise parameter. Additionally, we use a Markov Random Field based figure–ground segmentation in order to speed up our algorithm and further ensure that intra-object homogeneous parts are not affected by the lossy compression.

2. Related Work

Ditigal Cultural Heritage Preservation STEPHENSON [Ste99] review an interactive digital image database for cultural hertiage artifacts in terms of structure, tools and descriptiveness of included metadata. The application of POLITOU, PAVLIDIS, and CHAMZAS [PPC04] use the progressive image transmission technique of the JPEG2000 algorithm to efficiently browse and transmit cultural hertiage images over the internet in a web browser. PAVLIDIS, KOUTSOUDIS, ARNAOUTOGLOU, et al. [PKA*07] give an overview of state-of-the-art algorithms and techniques suitable for 3d reconstructions of cultural hertiage artifacts.

Lossy Compression The most common lossy natural image compression algorithm "JPEG" of WALLACE [Wal91] takes a $3\,\text{bit} \times 8\,\text{bit}$ image and estimates the discrete cosine coefficients of 8×8 pixel patches. Afterwards, it quantizes the coefficients and compresses them using Huffmann coding [Huf52]. The "JPEG2000" algorithm of CHRISTOPOULOS, SKODRAS, and EBRAHIMI [CSE00] 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, VAUTROT, and LUCAS [BVL10] makes use of inter-view redundancies and exploits the positive-sided geometric distribution between pixels of two neighboring images. AYDINOGLU and HAYES [AH94] use the previously computed dispar-

ity values to estimate every second frame. They compensate photometric variations using the "subspace projection technique". The adaptive approach of PERRA [Per15] 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.

Lossless Compression The PNG algorithm of BOUTELL [Bou97] 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, GUTHE, RITZ, et al. [BGR*19] 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 Segmentation The figure-ground segmentation algorithm "GrabCut" of ROTHER, KOLMOGOROV, and BLAKE [RKB04] segments images into multiple regions that have similar image characteristics by minimizing Markov Random Field energy terms. The parameter-free superpixel generation algirithm "SLIC0" of ACHANTA, SHAJI, SMITH, et al. [ASS*12] segments images into a user-defined number of equally sized regions that adapt to color changes.

3. Algorithm

Our image compression algorithm is structured as follows. First, the algorithm computes a superpixelization using the SLIC0 method of ACHANTA, SHAJI, SMITH, et al. [ASS*12] of the given input image. We use the SLIC0 algorithm as it requires no parameter except for the number of superpixels it should generate. After generating the superpixels it applies a pre-computed figure-ground annotation (section 3.1) to the superpixelization. This ensures the homogeneous regions inside the artifact do not get compressed lossy. Regions outside the figure-ground mask can still contain features that may contain important structures supporting stereo algorithms and should therefore stay lossless. In order to fit a plane into a superpixel, we define the coordinate system to be the the horizontal and vertical axis of the image x_i, y_i and the normalized color depth as the third one z_i, resulting in $p_i = (x_i, y_i, z_i)^T$. Now, we use the method of least-squares to fit a plane

$$
\begin{aligned}
& ax + by + cz + d = 0 \\
\Leftrightarrow\ & a/cx + b/cy + d/c = -z, && c \neq 0 \\
\Leftrightarrow\ & a'x + b'y + d' = z, && a' = -a/c,\ b' = -b/c,\ d' = -d/c
\end{aligned}
$$

into the set of pixels $\{p_1, \ldots, p_n\}$ of a superpixel. We apply a singular value decomposition (SVD) to the mean-subtracted superpixel points in order to find a numerically stable least-squares solution. The error between the superpixel and the approximated plane can then be derived by $e_i = |z - z_i|$ for each point p_i in a superpixel.

Once, the algorithm has found an optimal plane, a threshold on the average difference $\sigma < \tau_{\text{avg}}$, where $\sigma = \Sigma_i e_i / n$, is used for each plane to ensure that the plane is a good approximation of the given

superpixel. τ_{avg} depends on the camera sensor as the noise magnitude depends on the response curve of it and must be calibrated for each camera setting (see section 4). In order to calibrate the noise magnitude for the threshold we select a superpixel that is definitely in the background and evaluate its error value e_i and derive the threshold from it. Additionally we introduce a less strict threshold on the peak difference $\max_i e_i < \tau_{max}$ that is used because outliers of a superpixel get averaged out with the prior threshold but could contain important image features. In our implementation it turned out that $\tau_{max} = 0.024$ can be used for all tested camera models. If the error values are small enough, the algorithm discards the input data and only stores the parameters a', b', d', σ, which define the plane and the amount of noise in this region. The whole process is visualized in fig. 1.

3.1. Image Masking

Image masking describes the process of foreground and background segmentation and extraction, where the foreground contains only the object 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 right next to it where the object is placed on or attached to. 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, if a mount partially occludes the artifact's surface but it is applicable to reposition the artifact, e.g. upside down, by changing the mount structure and resolving the occlusion, then the complete surface can be captured and later aligned throughout multiple scan passes by carefully masking out background and mount structure in each image. 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. 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.

To automatically mask arbitrary large image sets suited for photogrammetric 3d reconstruction the Competence Center For Cultural Heritage Digitization [SRT*14; SRFF17] 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 [RKB04], that extends the original image segmentation by graph cut [BJ01] 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 di-
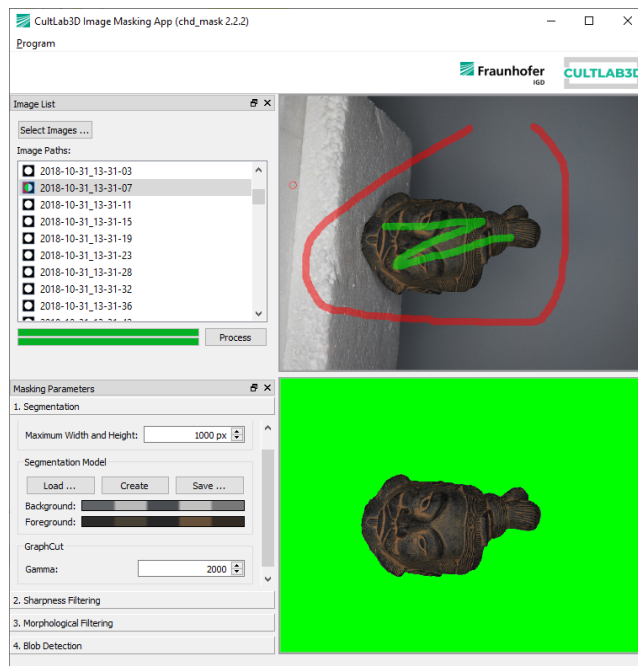

Figure 2: *Screenshot from the CultLab3D Image Masking App. The top right interactive window allows the user to set the initial scribbles. The bottom right window shows the segmentation result.*

mensions from color spaces, such as RGB, Lab and HSV, is formed and iteratively refined on the marked images. Figure 2 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. Especially visible in macro photography and caused by a narrow depth of field, parts of objects become blurry and lead to blurry texture on the 3d model. To address this problem, masked foreground regions can further be reduced to only their sharp parts by thresholding a local sharpness estimate based on the Laplacian operator. The sharpness estimate must be adjusted for each single image independently.

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

3.2. Representation

Our algorithm is able to use arbitrary lossless image compression algorithms for storing lossless regions of the input images. In our implementation we use the PNG algorithm which supports 16 bit images. In order to distinguish between lossless and plane-approximated regions, we define a mask as a one-channel 16 bit image that encodes a 0 for each pixel of a lossless superpixel and an
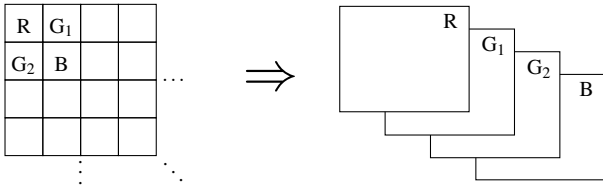
Figure 3: *The Bayer pattern groups four neighboring pixels that can be re-arranged into four independent color channels [BGR*19].*

integral value > 0 denoting the superpixel identifier for each plane-approximated superpixel. Encoding the lossless parts of the mask is almost free, as lossless regions tend to be adjacent to each other. This can easily be exploited by image compression algorithms that usually only encode differences to neighboring pixels. The same applies for plane-approximated regions, where differences are always small, because superpixel identifiers are enumerated continuously in the neighborhood.

The lossless image data is encoded by dividing the Bayer pattern of the camera into four color channels as illustrated in fig. 3. This technique ensures that all adjacent pixels originate from the same type of color filter and enlarges spatial correlations. All plane-approximated pixels are set to zero in the same image. Again, the plane-approximated parts are almost free to encode given the implementation of image compression algorithms.

The plane and noise parameters (a', b', d', σ) of each plane are stored as a list in an uncompressed binary representation as compressing floating point data yields sub-optimal compression rates compared to natural image data and the number of planes are small compared to the lossless parts of the image.

3.3. Decompression

The decompression step of our algorithm works as follows. First the mask image and the image containing lossless superpixels is decompressed using the reverse standard image compression algorithm used in section 3.2. Now, the algorithm copies the lossless pixels of the image if the mask is zero at the same pixel position. Otherwise, it approximates the pixel as a point on the plane $(ax + by + d)/c + \mathcal{N}(0, \sigma^2)$. The algorithm adds mean-free gaussian noise with a variance derived from the average error computed during the plane fitting step because planes at superpixel boundaries tend to create sharp edges due to approximation errors. We do not use de-blocking techniques to avoid loss of sharp edges and corners as the SLIC algorithm does not distinguish between edges that are created because the superpixel reached its size limit or edges that are actually in the original image.

4. Results

In section 4.1 we evaluate the compression rates and run-time performance of our algorithm on a number of datasets that are captured with cultural heritage image acquisition scanners. In section 4.2 we demonstrate how the image masking algorithm handles difficult input images. Additionally, we demonstrate that homogeneous areas
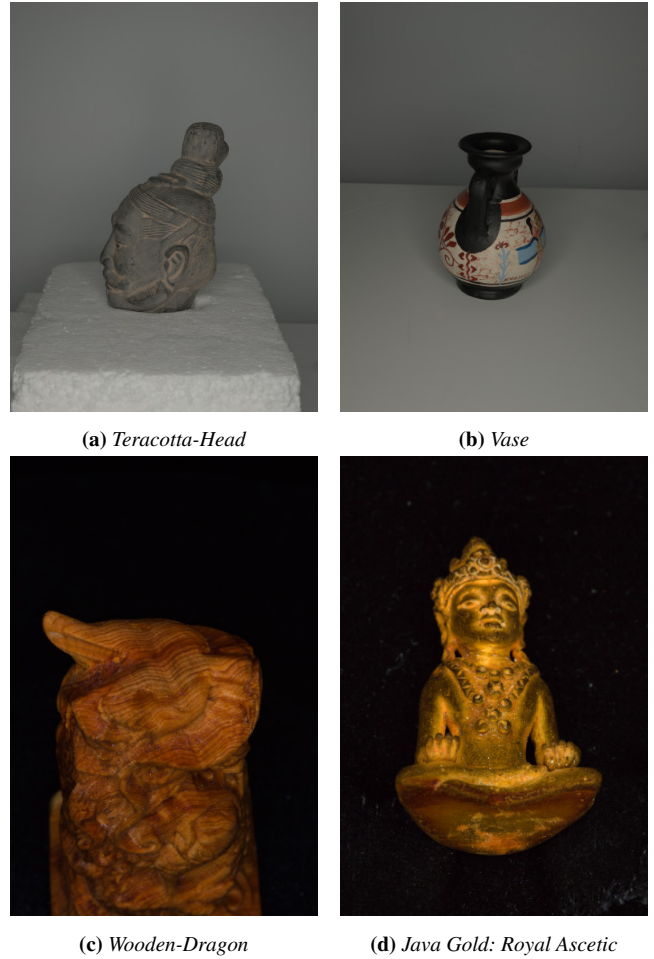

(a) *Teracotta-Head*   **(b)** *Vase*

(c) *Wooden-Dragon*   **(d)** *Java Gold: Royal Ascetic*

Figure 4: *Zoomed in example images from our evaluation datasets.*

do not affect the reconstruction error with an evaluation of the decompressed images against the original images with the Middleburry benchmark in section 4.3.

4.1. Superpixel Count and Compression Rates

A well chosen number of superpixels (parameter n) is essential for optimal compression rates. To obtain these values, we evaluated four multi-view geometry acquisition datasets acquired with the CultArm3D scanning station of the Competence Center Cultural Heritage Digitization of the Fraunhofer Institute for Computer Graphics Research [SRT*14; SRFF17; TDR*20]. These depict the *Teracotta-Head* (252 images) and the *Vase* artifact (259 images) that were scanned with a Phase One iXG 100MP $11\,608 \times 8708$ Pixel camera and a Schneider RS 72mm/iXG lens. The *Wooden-Dragon* (130 images) and the *Java Gold: Royal Ascetic* artifact (540 images) were 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 *Teracotta-Head*, *Vase* and *Wooden-Dragon* datasets are non-

antique testing datasets. 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. Example images from the datasets are shown in fig. 4.

In initial empirical experiments it turned out that the threshold $\tau_{avg} = 0.003$ is useful for the Phase One camera and $\tau_{avg} = 0.0006$ is useful for the Canon camera. These different thresholds need to be calibrated once for each static capturing setup as the amount of noise and the dynamic range depend on different vendor-specific settings like the film speed and the camera response curve.

Figure 5 shows how the compression rate depends on the chosen number of superpixels per image. We found that values between 1024 and 2048 are optimal and values over 2048 have detrimental effects on compression rates. For further experiments we used 1024 superpixels per image. It is noteworthy, that the *Wooden-Dragon* and *Royal Ascetic* images do not suffer as much as the other two data sets from bad compression rates due to smaller number of superpixels per image. This is a result of the different background homogeneities produced by either radial gradients in the *Terracotta-Head* and *Vase* datasets in comparison to the perfectly homogeneous black background of the *Wooden-Dragon* and *Royal Ascetic*. The compression rates are between 1:2.119 for the *Wooden-Dragon* and 1:4.39167 for the *Royal Ascetic*. These results are consistent with the idea of our approach, as the *Royal Ascetic* dataset consists of a smaller object and therefore larger background space for compression.

Table 1 shows the run-time performance of our algorithm in comparison to the PNG compression standard evaluated on a 2.60 GHz Intel Xeon E5-2650 v2.

Dataset	T-Head	Vase	W-Dragon	Java-Gold
Our	648.767	760.307	190.198	558.403
PNG	339.943	427.486	161.783	546.949

Table 1: *Run-time performance of our algorithm and the PNG compression standard for the datasets shown in fig. 4. All values were measured in seconds and resulted for a chosen number of superpixels of 1024.*

4.2. Image Masking Evaluation

The presented CultLab3D image masking app's algorithm is suited to segment mask images even though they feature a suboptimal setup or challenging structures. Figure 6 illustrates the masking of example images that exhibit a subset of such challenges that frequently occur.

For accurate photogrammetric 3d reconstruction many high resolution images from different angles are needed. Especially large and detailed artifacts can be clipped in some of these images. Figure 6c exemplary shows that the presented masking algorithm is able to handle clipped artifact images. Furthermore, artifact imagery may be blurry. Even though blurry images are not optimal for photogrammetric 3d reconstruction Figure 6d illustrates that the CultLab3D Image Masking App is able to mask these images properly.
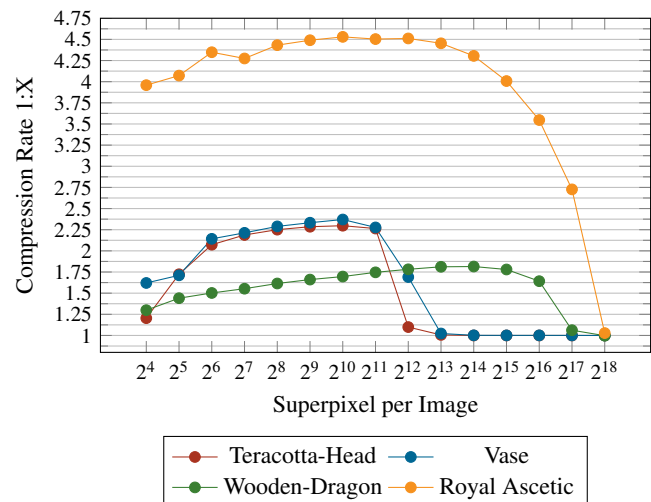

Figure 5: *Compression rates in 1:X in relation to the number of superpixel per image and in comparison to the PNG compression standard for our four datasets.*
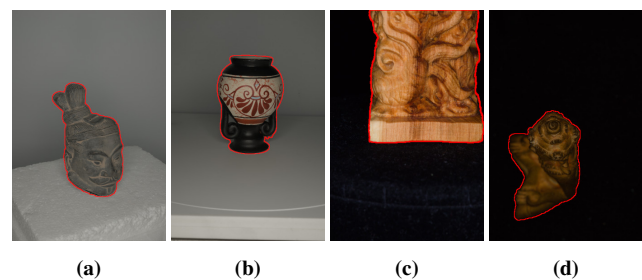

(a)	**(b)**	**(c)**	**(d)**

Figure 6: *Segmentation results from the CultLab3D Image Masking App for example images with structured background (a), holes (b) and clipped (c) or blurry (d) artifact.*

Other challenges in the cultural heritage domain arise from occlusion or the structural properties of individually designed stands for cultural heritage artifacts. Figure 6a illustrates a scenario in which the presented artifact stand features a visible heterogeneous structure. Nevertheless, the presented image masking algorithm is able to perform the masking process properly.

An additional challenge may be posed by the structural properties of the pictured artifact itself. Figure 6b illustrates an example in which the wholes of the vase handles are not excluded by the masking algorithm. This finding suggests that the presented algorithm has issues correctly masking loose objects. However, in practice this issue does not result in relevant performance loss since most cultural heritage artifacts are rather compact. Therefore, even though an ideal masking algorithm would exclude certain areas the implications of this behaviour regarding the compression rate are is insignificant.

Dataset	AAE Whole		AAE Noocc		ARE change		Bpp		Lossy
	orig	our	orig	our	Whole	Noocc	png	our	ratio [%]
Adirondack	19.91	19.97	19.86	19.93	0.29	0.31	4.00	3.90	10.31
ArtL	16.64	16.73	15.91	16.02	0.57	0.72	4.71	4.77	2.65
Jadeplant	41.94	41.72	44.29	44.17	0.53	0.25	4.48	4.25	14.8
Motorcycle	21.46	21.44	21.76	21.72	0.13	0.15	4.74	4.73	1.34
MotorcycleE	21.59	21.60	21.90	21.91	0.02	0.02	4.57	4.55	2.19
Piano	11.72	11.75	11.12	11.15	0.22	0.29	4.00	3.93	12.29
PianoL	13.94	13.90	13.53	13.49	0.27	0.32	4.01	3.90	12.55
Pipes	19.44	19.43	18.65	18.65	0.04	0.03	4.60	4.61	1.4
Playroom	23.59	23.60	21.74	21.77	0.06	0.11	4.28	4.23	8.05
Playtable	19.48	19.44	19.33	19.27	0.23	0.26	4.61	4.55	7.38
PlaytableP	17.98	17.95	17.74	17.70	0.18	0.22	4.62	4.57	7.28
Recycle	12.64	12.97	12.15	12.51	2.64	2.98	3.47	3.39	9.59
Shelves	11.65	11.75	11.61	11.73	0.85	0.97	3.69	3.39	19.97
Teddy	14.91	14.91	13.96	13.96	0.01	0.00	4.92	4.93	0.05
Vintage	36.35	36.55	36.41	36.64	0.55	0.62	3.79	3.51	22.89

Table 2: *Average error rates for the whole image and non occlusion in total numbers (Average Absolute Error) and as change between the original and our compression in percent (Average Relative Error). Compression rates in Bit/Pixel for the data set. The lossy ratio denotes the percentage of superpixels that could be used for our plane based compression method.*

Figure 7: *Example images from the Middleburry benchmark datasets [SHK*14].*

4.3. Reconstruction Quality

To assess the reconstruction quality after compression and decompression we used the 2014 Middlebury benchmark dataset of SCHARSTEIN and SZELISKI [SS02b] and SCHARSTEIN, HIRSCHMÜLLER, KITAJIMA, et al. [SHK*14] to evaluate the average disparity error of the image pairs with and without our near-lossless compression technique applied. Examples for this dataset can be seen in fig. 7. In order to compute the disparity values from the image pair, the algorithm computes the cost volume containing cost values for each possible disparity value at all image positions and perform a local optimization on that cost volume [SS02a]. We do not use any global optimization methods as this would require further assumptions on the dataset. Table 2 shows the average absolute error (AAE) and average relative error (ARE) rates for the whole image and the non occlusion areas for the original images and the images obtained from our compression approach. Furthermore we present the amount of superpixels, that could be successfully used for our plane based compression method in per-

cent. These values can differ widely, as they are depended on available background superpixels. The only dataset leading to a reconstruction quality change over 1.0% was the *Recycle* dataset with a percentage change of 2.64% for the whole image and 2.98% for non occlusion. Non occlusion error changes are higher in almost all cases, as non-occlusions are harder to reconstruct due to the limitation of only perceiving one at a time. The compression rate does not change significantly which would confer a compression rate only at the common level of PNG compression. This is due to two factors that are present in the Middlebury dataset. First, the ratio of background parts is low compared to the foreground background, because these datasets are designed for stereo algorithms. This leads suboptimal compressions. This can be dealt with, by simply using homogeneous backgrounds, when capturing datasets. Second, the images are preprocessed and contain a lower than usual amount of noise, which negatively affects our compression approach. This can be solved by retaining the background noise during preprocessing or not performing preprocessing steps at all. Both drawbacks do not affect the actual cultural hertiage image datasets, which can be seen at the superiour compression rates for them.

5. Conclusion

In this paper, we presented a near-lossless compression algorithm based on superpixels and figure-ground segmentation that makes use of homogeneous background parts of cultural heritage artifact images. Our algorithm uses the well known PNG algorithm as the compression backend and is therefore easy to implement. It archives compression rates of 1:4 compared to the standard PNG algorithm for small artifacts where background parts dominate and not fully cover the image dimensions and 1:2 for artifacts that dominate the image dimensions. Further the algorithm ensures that approximations of lossy parts of the image are bound to a camera-dependant noise threshold. Generally it can be said that background

pixels are hard to compress using standard lossless image compression algorithms due to noise. Handling background pixels with lossy compression algorithm is a very good trade-off between reconstruction quality and compression rates.

Future Work In the future we would like to further evaluate the reconstruction quality on the cultural heritage datasets directly in order to address the issue that compression rates and reconstruction qualities are currently evaluated on different datasets. Furthermore, we would like to automatically detect noise magnitudes of camera images in order to automatically determine the required threshold for our plane-approximation technique.

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

Acknowledgements

References

[AH94] AYDINOGLU, HALUK and HAYES, MONSON H. "Compression of multi-view images". *Proceedings of 1st International Conference on Image Processing*. Vol. 2. IEEE. 1994, 385–389. DOI: 10.1109/ICIP.1994.413597 2.

[ASS*12] ACHANTA, R., SHAJI, A., SMITH, K., et al. "SLIC Superpixels Compared to State-of-the-Art Superpixel Methods". *IEEE Transactions on Pattern Analysis and Machine Intelligence* 34.11 (2012), 2274–2282. DOI: 10.1109/TPAMI.2012.120 2.

[BGR*19] VON BUELOW, MAX, GUTHE, STEFAN, RITZ, MARTIN, et al. "Lossless Compression of Multi-View Cultural Heritage Image Data". *Proceedings of the Eurographics Workshop on Graphics and Cultural Heritage*. GCH. Nov. 2019. DOI: 10.2312/gch.20191343 2, 4.

[BJ01] BOYKOV, Y. Y. and JOLLY, M.-P. "Interactive graph cuts for optimal boundary region segmentation of objects in N-D images". *Proceedings Eighth IEEE International Conference on Computer Vision. ICCV 2001*. Vol. 1. 2001, 105–112 vol.1. DOI: 10.1109/ICCV.2001.937505 3.

[Bou97] BOUTELL, THOMAS. *PNG (portable network graphics) specification version 1.0*. Tech. rep. 1997. DOI: 10.17487/RFC2083 2.

[BVL10] BATTIN, BENJAMIN, VAUTROT, PHILIPPE, and LUCAS, LAURENT. "A new near-lossless scheme for multiview image compression". *Stereoscopic Displays and Applications XXI*. Ed. by WOODS, ANDREW J., HOLLIMAN, NICOLAS S., and DODGSON, NEIL A. Vol. 7524. International Society for Optics and Photonics. SPIE, 2010, 572–582. DOI: 10.1117/12.838908 2.

[CSE00] CHRISTOPOULOS, CHARILAOS, SKODRAS, ATHANASSIOS, and EBRAHIMI, TOURADJ. "The JPEG2000 still image coding system: an overview". *IEEE transactions on consumer electronics* 46.4 (2000), 1103–1127. DOI: 10.1109/30.920468 2.

[Huf52] HUFFMAN, DAVID A. "A method for the construction of minimum-redundancy codes". *Proceedings of the IRE* 40.9 (1952), 1098–1101. DOI: 10.1007/BF02837279 2.

[Per15] PERRA, CRISTIAN. "Lossless plenoptic image compression using adaptive block differential prediction". *2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE. 2015, 1231–1234. DOI: 10.1109/icassp.2015.7178166 2.

[PKA*07] PAVLIDIS, GEORGE, KOUTSOUDIS, ANESTIS, ARNAOUTOGLOU, FOTIS, et al. "Methods for 3D digitization of cultural heritage". *Journal of cultural heritage* 8.1 (2007), 93–98. DOI: 10.1016/j.culher.2006.10.007 2.

[PPC04] POLITOU, EUGENIA A, PAVLIDIS, GEORGE P, and CHAMZAS, CHRISTODOULOS. "JPEG2000 and dissemination of cultural heritage over the Internet". *IEEE transactions on image processing* 13.3 (2004), 293–301. DOI: 10.1109/TIP.2003.821348 2.

[RKB04] ROTHER, CARSTEN, KOLMOGOROV, VLADIMIR, and BLAKE, ANDREW. ""GrabCut": Interactive Foreground Extraction Using Iterated Graph Cuts". *ACM Trans. Graph.* 23.3 (Aug. 2004), 309–314. DOI: 10.1145/1015706.1015720 2, 3.

[SHK*14] SCHARSTEIN, DANIEL, HIRSCHMÜLLER, HEIKO, KITAJIMA, YORK, et al. "High-resolution stereo datasets with subpixel-accurate ground truth". *German conference on pattern recognition*. Springer. 2014, 31–42. DOI: 10.1007/978-3-319-11752-2_3 6.

[SRFF17] SANTOS, PEDRO, RITZ, MARTIN, FUHRMANN, CONSTANZE, and FELLNER, DIETER. "3D mass digitization: a milestone for archeological documentation". *Virtual Archaeology Review* 8.16 (2017), 1–11. DOI: 10.4995/var.2017.6321 3, 4.

[SRT*14] SANTOS, PEDRO, RITZ, MARTIN, TAUSCH, REIMAR, et al. "CultLab3D - On the Verge of 3D Mass Digitization". *Eurographics Workshop on Graphics and Cultural Heritage*. Ed. by KLEIN, REINHARD and SANTOS, PEDRO. The Eurographics Association, 2014. DOI: 10.2312/gch.20141305 3, 4.

[SS02a] SCHARSTEIN, DANIEL and SZELISKI, RICHARD. "A Taxonomy and Evaluation of Dense Two-Frame Stereo Correspondence Algorithms". *International Journal of Computer Vision* 47 (Apr. 2002), 7–42. DOI: 10.1023/A:1014573219977 6.

[SS02b] SCHARSTEIN, DANIEL and SZELISKI, RICHARD. "A taxonomy and evaluation of dense two-frame stereo correspondence algorithms". *International journal of computer vision* 47.1-3 (2002), 7–42. DOI: 10.1109/SMBV.2001.988771 6.

[Ste99] STEPHENSON, CHRISTIE. "Recent developments in cultural heritage image databases: directions for user-centered design". (1999). ISSN: 0024-2594 2.

[TDR*20] TAUSCH, REIMAR, DOMAJNKO, MATEVZ, RITZ, MARTIN, et al. "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 (Jan. 2020). ISSN: 1820 - 4503 3, 4.

[Wal91] WALLACE, GREGORY K. "The JPEG Still Picture Compression Standard". *Commun. ACM* 34.4 (Apr. 1991), 30–44. DOI: 10.1145/103085.103089 2.

[WS00] WAINWRIGHT, MARTIN J and SIMONCELLI, EERO P. "Scale mixtures of Gaussians and the statistics of natural images". *Advances in neural information processing systems*. 2000, 855–861 2.