EUROGRAPHICS Workshop on Graphics and Cultural Heritage (2019)
S. Rizvic and K. Rodriguez Echavarria (Editors)

Short Paper

Lossless Compression of Multi-View Cultural Heritage Image Data

Max von Buelow[1], Stefan Guthe[1], Martin Ritz[2], Pedro Santos[2], and Dieter W. Fellner[1,2,3]

[1]TU Darmstadt, Germany
[2]Fraunhofer IGD, Germany
[3]Graz University of Technology, Institute of Computer Graphics and Knowledge Visualization, Austria

Abstract

Photometric multi-view 3D geometry reconstruction and material capture are important techniques for cultural heritage digitalization. Capturing images of artifacts with high resolution and high dynamic range and the possibility to store them losslessly enables future proof application of this data. As the images tend to consume immense amounts of storage, compression is essential for long time archiving. In this paper, we present a lossless image compression approach for multi-view and material reconstruction datasets with a strong focus on data created from cultural heritage digitalization. Our approach achieves compression rates of 2:1 compared against an uncompressed representation and 1.24:1 when compared against Gzip.

CCS Concepts

• ***Computing methodologies*** → ***Image compression;***

1. Introduction

In the past, research on cultural heritage digitalization mostly focused on creating reconstructions instead of efficient storage of datasets. Digital copies, commonly called datasets, of cultural heritage artifacts consist of a set of images captured from different camera positions for multi-view geometry reconstructions and with different lighting conditions for material capture. Moreover, they contain the extrinsic and intrinsic parameters for each camera and the light source positions. To achieve the goal of transferring artifacts across generations, it is important to store these datasets at multiple secure locations. This is currently limited by the enormous image file sizes that are caused by capturing the images with high resolutions of up to 100 megapixels and dynamic ranges of up to 16 bit, making storage and transfer inefficient and expensive. In order to reduce storage requirements and transfer times, compression of multi-view and material capture datasets is mandatory. To keep the original quality and avoid losing any information of the original images, the datasets need to be compressed losslessly.

Initially, the individual images are archived either in the proprietary camera file format or an uncompressed representation with generic compression (e.g. Gzip). Due to hardware limitations of the cameras, these image file formats yield sub-optimal compression rates. Industry cameras usually transfer pixel data using a bus system to a standard computer system in uncompressed form. The images are commonly written out to uncompressed file formats due to simplistic interfaces that do not offer features like compression. Usually, Gzip or similar generic compression algorithms are then used to compress the uncompressed image representations, also leading to non-optimal compression rates. Other state-of-the-art image compression algorithms are either lossy or do not support high dynamic range images.

2. Related Work

The lossy JPEG image compression algorithm of Wallace [Wal91] compresses quantized DCT coefficients of an image using Huffman coding. The JPEG 2000 algorithm of Christopoulos et al. [CSE00] uses a discrete wavelet transformations instead that can also be applied losslessly. This algorithm is limited to encode the wavelet coefficients independently using an arithmetic coder. Shapiro [Sha93] showed that the individual subbands of a wavelet transformation can be aligned in a tree structure. He used the fact that insignificant coefficients remain insignificant in higher frequency subbands to perform lossy compression. The PNG algorithm of Boutell [Bou97] is able to compress images with bit depths of up to 16 bit losslessly. For this, he uses local filter techniques whose responses are encoded using a Huffman coder that produces inferior results when compared against other entropy coders.

The lossless approach of Martins and Forchhammer [MF98], usually denoted “LOCO-3D” as it is a 3D extension of the LOCO-I algorithm by Weinberger et al. [WSS96], compresses video frames using a set of different predictors that are sequentially evaluated to optimally predict the neighboring, motion compensated pixels from reference frames to exploit inter-view redundancies. As this algorithm was designed for video compression, a pseudo video sequence must be created prior to the actual encoding. Brunello et al. [BCMR02] extended this to a mathematically well-posed algorithm, called “LOPT-3D”, that uses a weighed sum of previously coded pixels by solving a linear equation system resulting in better

DOI: 10.2312/gch.20191343

compression rates than LOCO-3D. LOPT-3D uses a Golomb-Rice coder as its coding backend. Carotti and De Martin [CD05] used multi-frame Block Motion Compensation to further enhance the results of the LOPT-3D approach. Kamisetty and Jawahar [KJ03] estimated three view relationships using a trilinear tensor to create a prediction followed by a residual coding step. Perra [Per15] encodes multi-view images by minimizing the entropy of the 1D Differential Pulse-Code Modulation (DPCM) and encoding the block-DPCM data using the LZMA dictionary coder backend. A comparison between DPCM, 3D Discrete Cosinus Transformations (DCT) and Principle Component Analysis (PCA) predictions based on estimated disparity values was done by Shah and Dodgson [SD01]. However, all these algorithms need inefficient prior disparity estimation and perform sub-optimal on real-world datasets.

3. Preliminaries

Cultural heritage image data usually consists of a set of images and their camera and light parameters. These images are two-dimensional arrays of a color representation that depends on the camera sensor. Colors are usually represented in the standard RGB (sRGB) color space using three color values for preprocessed images or using a Bayer Pattern with vendor specific bit lengths for unpreprocessed (raw) images. The Bayer Pattern can be seen as a single pixel with four color channels: one red, two green channels and a blue one. The conversion of the Bayer Pattern to a four-channel represenation is visualized in fig. 1.

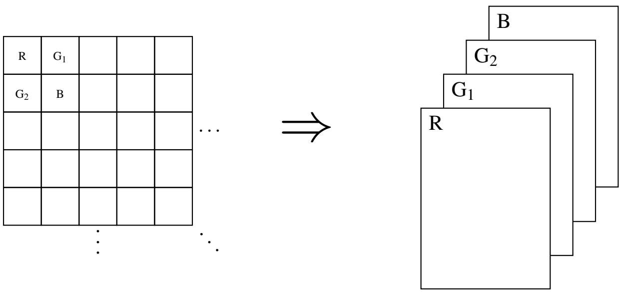


Figure 1: *In the Bayer Pattern, each pixel contains only one of the three color channels. Since the pattern repeats for every 2×2 pixel group, it can be represented as a single, four-channel color image. Due to the high resolution of the camera images, the offset between channels is negligible.*

The sets of images contain different redundancies that can potentially be used for efficient lossless compression:

- Temporal/inter-view redundancies: Pixels from correlating regions (i.e. pixel referring to the same point in an object) in two images commonly tend to have similar values.
- Spatial/inter-pixel redundancies: Adjacent pixels usually have similar values.
- Spectral redundancies: Adjacent frequency coefficients commonly tend to have similar values.
- Coding redundancies: Redundancies caused by inefficient coding of symbol representations.

Spatial and spectral redundancies are caused by the heavy tailed image statistics of natural images [WS00]. Temporal redundancies are caused by multiple images capturing the same object but with different capturing conditions. Coding redundancies occur by allowing trivial random access of the pixels, because files and memory can usually only be addressed in multiples of 8 bit.

To evaluate temporal redundancies, we captured an *optimal* dataset consisting of 10 images with same camera and scene configuration, theoretically having the maximum temporal redundancies possible in realistic conditions. Initial experiments on this dataset showed that it is hard to compress temporal redundancies losslessly, as they need previous expensive motion or light compensation steps and the remaining residual cannot be compressed efficiently. As each image of this dataset consist of the ground truth data with uncorrelated and correlated mean-free noise, the difference between two images has roughly twice the amount of noise. As long as a compressor is better at encoding the ground truth signal than the noise, a prediction based on a single noisy input is going to be worse [Bue19].

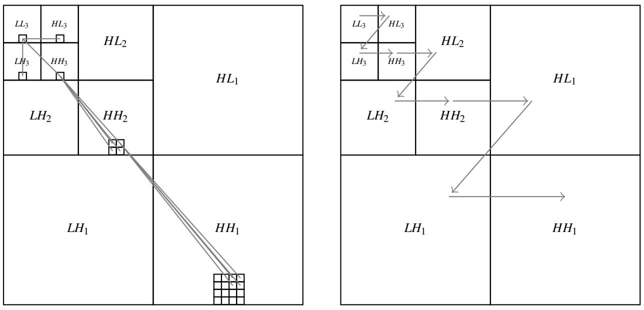
4. Lossless Image Compression

Our method compresses the spectral redundancies for each view independently as follows:

Let L and H be the low and high frequency components of a 1D signal determined by a lossless CDF 5/7 discrete wavelet transformation [CDF92]. Given an input image, our method computes L and H row- and column-wise successively resulting in LL_0, LH_0, HL_0 and HH_0. This process repeats n times on $LL_{1,\dots,n-1}$ to create lower frequency subbands of the image.

This recursive structure of the wavelet transformation allows a topological alignment of each subband, which is defined as follows (see also fig. 2a). Pixels of the subband with the coarsest resolution LL_n are defined to be the root of the tree structure. Pixels at the corresponding position in the remaining three subbands of the same resolution HL_n, LH_n and HH_n are the children of the root node. Due to the self similarity of the signal, high frequency wavelet subbands HL_i tend to be bound by the lower resolution values HL_{i+1}. Thus, the four pixels LH_{i-1} are the children of LH_i. The same applies for HL_i and HH_i.

Given this tree structure, we use a context-adaptive arithmetic coding scheme to encode all wavelet coefficients. The approach performs a context selection based on the parent of each wavelet coefficient. For this, the parent coefficient is divided into two parts with equal amounts of bits. The most significant bits (MSB) are used as the bin for the context selection switching between $m = 2^{n_b}$ models, where n_b is the amount of MSB. We use this model to encode only the n_b MSB of the actual coefficient as the remaining least significant bits (LSB) have too much variation, slowing down the adaptation of the model and encode the LSB in an second encoding step. For this, the actual coefficient's n_b MSB are additionally used to perform a second context selection of further m models for each previous bin resulting in a total number of $m^2 + m$ models. Afterwards, the selected model is used to encode the remaining LSB of the actual coefficient. Assuming a bit depth of 8 bit, the number of models to be allocated is $16^2 + 16 = 272$. For a bit depth



(a) *Tree dependencies of the wavelet tree [Sha93].* **(b)** *Traversal order of the wavelet tree [Sha93].*

Figure 2: *Topological relations between the different subbands. Each coarse subband has four children in the finer subband, except of the coarsest subband. The traversal order is defined to traverse coarse subbands first and then successively each finer subband. This ensures that parents are coded prior to children.*

of 16 bit $256^2+256=65792$ models need to be allocated. All models are based on dynamic frequency tables and the amount of bits for the context selection is limited to 8. However, we expect that on average only m of the m^2 models are used for the second context selection, since coefficients vary only slightly between wavelet levels.

Figure 3 shows the flow chart of the whole process. Again, the four MSB of the current coefficient are encoded with an arithmetic coder that uses a model selected from a list of *parent models* depending on the four MSB of the *parent coefficient*. Then, the remaining four LSB of the *current coefficient* are encoded using another model selected from a list of *current models* depending on each four MSB of the *parent coefficient* as well as the *current coefficient*.

As the root of the wavelet tree, i.e. the coarsest subband, has no parent coefficient, it is encoded using a context-free model. The limit of eight bits for the bin index is used to prevent excessive memory requirements that are caused by the increasing number of models. We use the implementation of Moffat et al. [MNW98] for arithmetic coding and the implementation of Fenwick [Fen94; Fen96] for dynamic frequency tables.

To preserve the reversibility of the compression scheme, all data that was available for compression must be available for decompression as well. Specifically, each parent must be coded prior to its children. Thus, the encoder starts with the root node and ascents from there to the higher frequency subbands as visualized in fig. 2b. The bin index used for the second model selection is available to the decompression algorithm by directly decoding it from the stream.

5. Results

In this section, we evaluate our compression approach on different cultural heritage datasets. The Angel dataset is a multi-view material capture dataset acquired with the CultArc3D scanner of the Competence Center Cultural Heritage Digitization of the Fraunhofer Institute for Computer Graphics Research

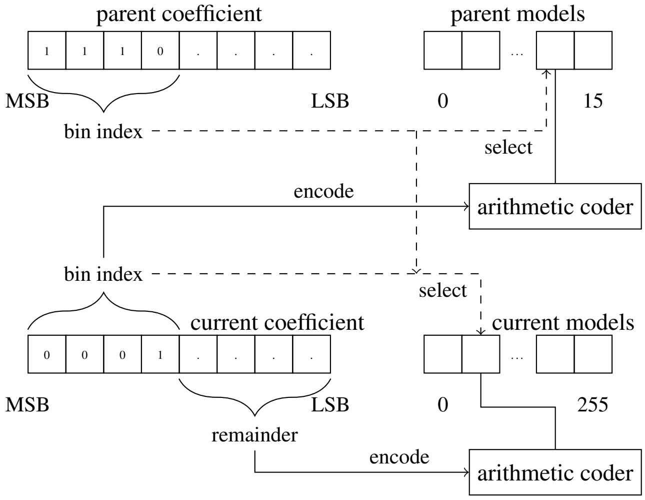


Figure 3: *Flow chart of context-adaptive coding of the wavelet coefficients exploiting the correlations to the topological parent coefficient. Bits used for model selection are marked using a dashed arrow and then connected to the arithmetic coder using a solid line. Bits that are encoded are marked using a solid arrow.*

(a) *Shoe.* **(b)** *Head.* **(c)** *Angel.*

Figure 4: *Example pictures of the individual datasets we used to perform our evaluations. The picture of the Angel dataset is the maximum image of all possible lighting directions.*

[SRT*14; SRFF17]. As image sensor, 18 industry cameras of type *IC11000CU* from NET GmbH with a resolution of 3840×2748 pixels and a bit depth of 12 bit are used. The CultArc3D scanner consist of two rotating arcs of which one is equipped with 9 cameras and the other one with 9 lights at equiangular spacing. Both arcs can move independently of each other and can reach positions at an angular resolution of $1/100°$. For a material capture dataset, 9 individual angular positions are used for each arc in combination, resulting in 9×9 camera and 9×9 light source positions, in total 9^4 images. The Shoe dataset containing 597 images was captured with the CultArm3D scanner [SRT*14] which is using a robotic arm to position an attached *NIKON D610* camera at a resolution of 6080×4028 pixels and a bit depth of 14 bit, controlled by an intelligent next-best view planning algorithm to determine the best poses for 3D reconstruction. The Head dataset, consisting of 252 images, was also captured with the CultArm3D scanner, but using a *Phase One A/S iXG 100MP* with a resolution of $11\,608 \times 8708$ pixels and 16 bit per pixel as imaging sensor. Example pictures from the datasets can be seen in fig. 4.

The original file format of NIKON, Canon and PhaseOne cameras are *NEF*, *CR2* and *IIQ* respectively. These formats are loss-

Approach	Shoe	Head	Angel
Uncompressed	14	16	12
Gzip	9.272	9.411	7.987
PNG	10.434	11.676	7.267
Raw	9.356	9.477	—
Ours	7.659	9.033	5.462

Table 1: *Results of our compression approach compared to state-of-the-art lossless image compression algorithms. All values are in bits per pixel.*

lessly compressed with proprietary algorithms. The images of the NET camera must be read using an API that directly communicates with the camera, producing uncompressed 16 bit TIFF files that contain scaled 14 bit samples. All these cameras encode color information using a Bayer Pattern.

Table 1 shows the results in *bits per pixel*. It can be observed that our approach has compression rates of approximately 2:1 compared to an uncompressed representation. Results of Gzip and the proprietary raw format are similar, indicating that the camera manufacturers possibly use similar compression techniques. Gzip archives compression rates of about 1.5:1 and the rates of PNG range from 1.37:1 to 1.65:1.

6. Conclusion and Future Work

In this paper, we presented a lossless compression approach for cultural heritage datasets that takes advantage of spectral redundancies. Initial experiments showed that temporal redundancies can not be exploited efficiently, having the side-effect that random access within the dataset can be implemented easily. Our approach achieves compression rates of 2:1 against an uncompressed representation and 1.24:1 against Gzip. In the future, we would like to further analyze camera sensor noise and evaluate if it is possible to remove it in a preprocessing step, creating a lossless representation in an optical sense. This would allow for an even more efficient temporal redundancy reduction.

References

[BCMR02] Brunello, Dania, Calvagno, Giancarlo, Mian, Gian Antonio, and Rinaldo, Roberto. "Lossless video coding using optimal 3D prediction". *ICIP (1)*. 2002, 89–92. doi: 10.1109/ICIP.2002.1037966 1.

[Bou97] Boutell, Thomas. *PNG (portable network graphics) specification version 1.0*. Tech. rep. 1997. doi: 10.17487/RFC2083 1.

[Bue19] von Buelow, Max. "Lossless Compression of Structured and Unstructured Multi-View Image Data". MA thesis. Technische Universität Darmstadt, May 2019 2.

[CD05] Carotti, Elias SG and De Martin, Juan Carlos. "Motion-compensated lossless video coding in the CALIC framework". *Proceedings of the Fifth IEEE International Symposium on Signal Processing and Information Technology, 2005*. IEEE. 2005, 600–605. doi: 10.1109/ISSPIT.2005.1577165 2.

[CDF92] Cohen, Albert, Daubechies, Ingrid, and Feauveau, J-C. "Biorthogonal bases of compactly supported wavelets". *Communications on pure and applied mathematics* 45.5 (1992), 485–560. doi: 10.1002/cpa.3160450502 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 1.

[Fen94] Fenwick, Peter. "A new data structure for cumulative probability tables". *Software-Practice and Experience* (1994). doi: 10.1002/spe.4380240306 3.

[Fen96] Fenwick, Peter. "A New Data Structure for Cumulative Probability Tables: An Improved Frequency-to-Symbol Algorithm". *Software: Practice and Experience* 26.4 (1996), 489–490. doi: 10.1002/(SICI)1097-024X(199604)26:4<489::AID-SPE22>3.0.CO;2-S 3.

[KJ03] Kamisetty, Chaitanya and Jawahar, CV. "Multiview image compression using algebraic constraints". *TENCON 2003. Conference on Convergent Technologies for Asia-Pacific Region*. Vol. 3. IEEE. 2003, 927–931. doi: 10.1109/tencon.2003.1273383 2.

[MF98] Martins, Bo and Forchhammer, Søren. "Lossless compression of video using motion compensation". *Proceedings DCC'98 Data Compression Conference (Cat. No. 98TB100225)*. IEEE. 1998, 560. doi: 10.1109/dcc.1998.672302 1.

[MNW98] Moffat, Alistair, Neal, Radford M, and Witten, Ian H. "Arithmetic coding revisited". *ACM Transactions on Information Systems (TOIS)* 16.3 (1998), 256–294. doi: 10.1145/290159.290162 3.

[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.

[SD01] Shah, Druti and Dodgson, Neil A. "Issues in multiview autostereoscopic image compression". *Stereoscopic Displays and Virtual Reality Systems VIII*. Vol. 4297. International Society for Optics and Photonics. 2001, 307–317. doi: 10.1117/12.430829 2.

[Sha93] Shapiro, Jerome M. "Embedded image coding using zerotrees of wavelet coefficients". *IEEE Transactions on signal processing* 41.12 (1993), 3445–3462. doi: 10.1109/78.258085 1, 3.

[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.

[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.

[Wal91] Wallace, Gregory K. "The JPEG Still Picture Compression Standard". *Commun. ACM* 34.4 (Apr. 1991), 30–44. doi: 10.1145/103085.103089 1.

[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.

[WSS96] Weinberger, Marcelo J, Seroussi, Gadiel, and Sapiro, Guillermo. "LOCO-I: A low complexity, context-based, lossless image compression algorithm". *Proceedings of Data Compression Conference-DCC'96*. IEEE. 1996, 140–149. doi: 10.1109/DCC.1996.488319 1.