Interpreter-Based Evaluation of Compressed SAR Images Using JPEG and HEVC Intra Coding: Compression Can Improve Usability

Ulrike Pestel-Schiller and Jörn Ostermann
Institut für Informationsverarbeitung, Leibniz Universität Hannover, Appelstr. 9A, 30167 Hannover, Germany, phone:+49 511 76219580, fax: +49 511 7625333, pestel@tnt.uni-hannover.de

Abstract

Usability-based assessment of SAR systems is time-consuming and costly, if based on human evaluation. We investigate an assessment system which aims at finding algorithms to simulate, complement and partly replace this evaluation. The system component investigated is the coding with spatial and amplitude resolution as parameters. Images of different qualities were generated by JPEG and intra-HEVC for evaluation. SAR experts solved tasks on SAR images with different qualities. A quality metric called *weighted* F_1-score evaluates the solved tasks. Best score is at a data rate of 0.30 bit/pel with spatial resolution of 11 cm using JPEG. Keywords — image evaluation, data compression, SAR

1 Introduction

When transmitting synthetic aperture radar (SAR) data from an airborne platform to a ground station, the channel capacity of the data link tyoically is the bottleneck. One solution is the on-board processing of SAR raw data to image [1] including an operational autofocussing [2], [3] followed by data [4] of the SAR image data. To evaluate such automatic on-board processing chain, typically the entire SAR systems or their components are being compared by their global performance figures like spatial resolution, signal-to-noise ratio (SNR), peak side lobe ratio (PSLR), integrated side lobe ratio (ISLR) [5] and peak signal-to-noise ratio (PSNR) [6] which are only loosely related to usability-based quality. This is motivated by the fact that otherwise a comparison of SAR systems or their components requires an extensive evaluation of the SAR image quality by expert interpreters which is very time-consuming and thus cost-intensive.

In [7] we have presented a usability-based SAR system assessment framework and applied it to find those interpretation tasks whose solvability depends on the investigated system parameters. For example, 'mark vehicles up to 6 m length' or 'mark ships between 30 m and 100 m length' are typical interpretation tasks to be solved. As a side result, test images where no such tasks exist, were identified and excluded from subsequent investigations. In [8] we described first results of the SAR coding component assessment where the coded image versions most preferred by the interpreters were identified.

The assessment system [8] aims at finding digital signal processing algorithms to simulate, complement and partly replace the human evaluation. For obtaining that, in evaluation sessions, expert SAR interpreters were requested to solve a set of representative interpretation tasks on SAR images whose image qualities result from a specific SAR system where the parameters to be investigated were varied.

We decided to begin with the assessment of the SAR coding component for data compression where the spatial and amplitude resolution are the fundamental parameters to be varied yielding image versions with different data rates [9] and corresponding image qualities. The two standardized coding systems JPEG [10] from Joint Photographic Experts Group and High Efficiency Video Coding (HEVC) [11] in intra mode were investigated. In addition to solving the interpretation tasks, the expert were requested to identify the coded image version they would most prefer to work with for each test image.

In this paper, we derive a quality metric based on the well-known F_1-score [12] for evaluating the solved tasks of the interpreter evaluation sessions. The evaluation results are discussed and compared with that SAR image quality the interperters prefer to work with.

The remainder of the paper is organized as follows: In Section 2 the experimental setup of the entire assessment system is described. The evaluation procedure is specified in Section 3. The interpreter based quality metric is derived in Section 4. Results are presented and discussed in Section 5 before Section 6 concludes the paper.

2 Experimental Set up

Fig. 1 shows the block diagram of the setup of the assessment system. A SAR system generates a set of image versions at different image qualities. This paper deals with the assessment of the SAR image coding component which is part of the SAR system. Hence, the spatial and the amplitude resolution are the parameters to be varied. On the one hand each SAR image is going to be evaluated in interpreter sessions, and on the other hand by digital signal processing methods, each resulting in corresponding quality metrics. A long term goal is to find a functional dependency between both metrics.

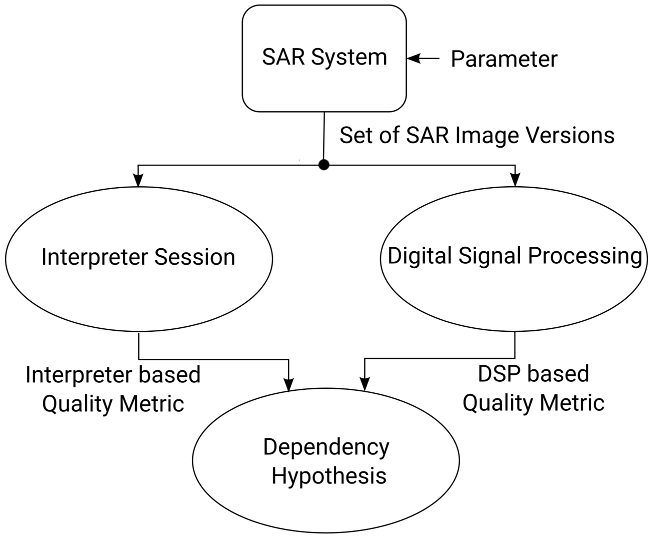

Figure 1 Experimental setup of the assessment system

Fig. 2 shows the SAR system of the experimental setup with its relevant details for the presented assessment. For compensating the monitor nonlinearity, the output of the SAR processing is γ-corrected, followed by a dynamic companding controlled by the statistical values mean m and standard deviation σ of the signal itself. This leads to a SAR Reference Image with an amplitude resolution of 8 bit/pel which is directly well-suited for evaluation. This SAR Reference Image is then converted to images with different spatial resolution levels. The latter are used as input for the standardardized coding systems JPEG and HEVC intra mode, respectively. By variation of the quantization parameters of the investigated coding systems, the amplitude resolution is varied.

Five levels of spatial resolution are analyzed. Level 1 corresponds to the original resolution. The spatial resolutions are reduced by a factor of two in each direction from level to level. For level 1, three image qualities have been investigated, determined by suitably selected fixed HEVC quantization parameters. The data rates of these qualities provide the basis for so called rate ranges. In the following, data rates are given in (number of bits) per (number of pixels of resolution level 1) (bit/pel). The quantization parameters were determined as follows:

- We applied the HEVC coding system HM profile main to the resolution level 1 version of the test images with fixed quantization parameters of 32, 40 and 51 on a scale from 0 (best quality) to 51 (worst quality). This results in similar image qualities for every test image at each rate range but the data rates slightly differ across test images while using the same quantization parameter.
- For each test image we have three distinct data rates.
- Quantization parameters for the HEVC coding system in resolution levels 2 to 5 and quantization parameters for the JPEG coding system in all resolution levels were adjusted such that the data rates are similar to these of the three distinct data rates. This results in three data rate ranges.
- We like to emphasize that not all combinations of spatial resolutions and data rates can be realized by every codec.

In addition to these coded image versions, an image version called unquantized is included for each test image where the HEVC coding system works losslessly.

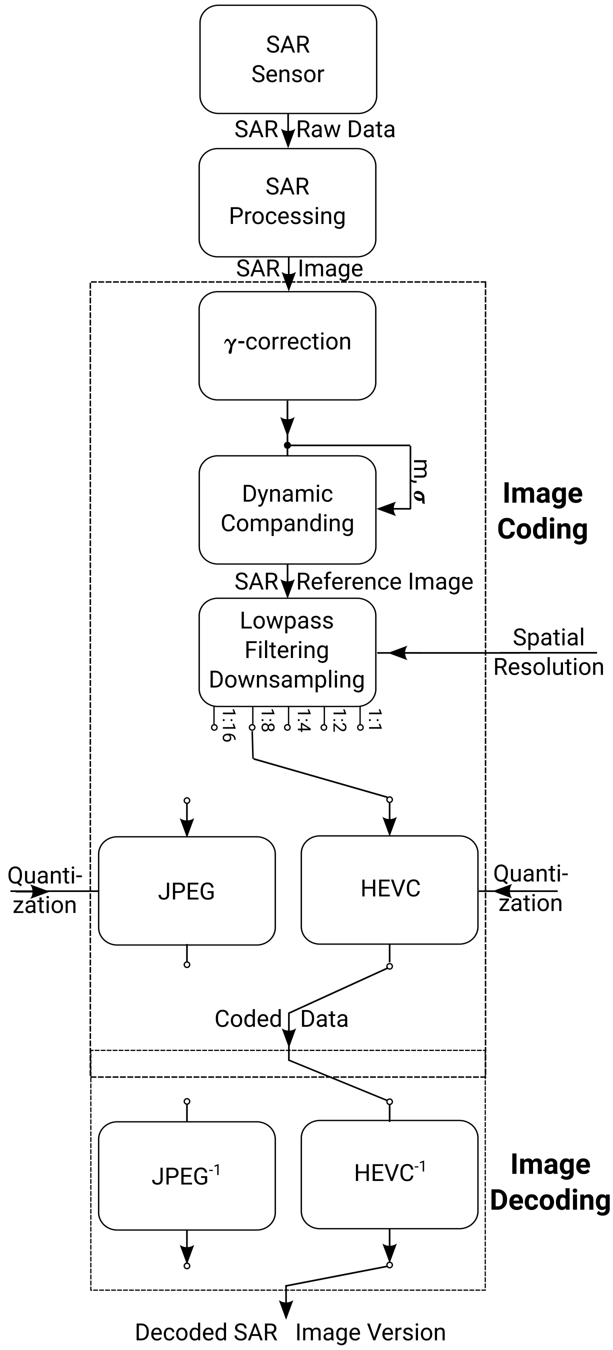

Figure 2 SAR system of the experimental setup with details of coding and decoding

In this work the test image material comprises the two Spot Mode SAR images of airport scenery with spatial resolutions of 11 cm at resolution level 1 [7]. Fig. 3 shows the spatial resolutions over the data rate, marked with crosses or dots. It contains marks of 44 images, comprising 10 unquantized image versions for the two test images and the five resolution levels, 22 image versions coded with the HEVC coding system and 12 image versions coded with the JPEG coding system. The color of the crosses indicate the level of spatial resolution; the three data rate ranges are marked with bars.

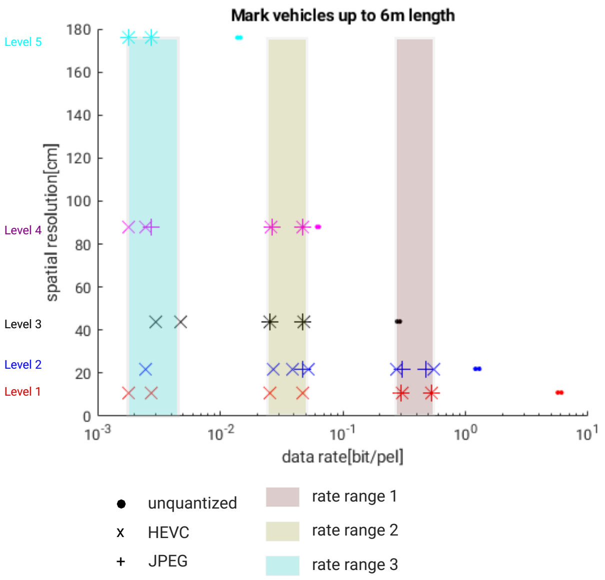

Figure 3 Investigated spatial resolutions and data rates for HEVC intra mode and JPEG using testimages *Airport* (best viewed in color)

Our expert interpreters follow specified interpretation principles according to National Imagery Interpretability Rating Scale (NIIRS) [13] which cover the boundary of solvability over the resolution levels. By proposing associated interpretation tasks, care was taken to avoid any influence of subjective prior knowledge of the interpreter. The current investigations deal with the interpreter session (left path in Fig. 1) wherein expert interpreters were asked to evaluate the selected set of SAR images.
We setup an interpreter based quality metric based on the well-known F_1-score which we call *weighted* F_1-score (see sec. 4). This metric is used for evaluating the solved tasks of the interpreter evaluation sessions. It indicates how reliable interpretation tasks can be solved at given data rates related to image qualities. This allows to compare the investigated coding systems with their corresponding parameter settings and associated data rates.
The results of the *weighted* F_1-score are compared to the other interpreter-based quality metric investigated where every interpreter identifies that SAR image quality he prefers to work with. This dependency results in a relationship between the coded image quality and the corresponding data rate.

3 Evaluation Procedure

An interpreter session was performed using a dedicated GUI where zoom, pan and a dimensioning tool can be adjusted by the interpreters. This enforces the evaluation of the image quality instead of the specific knowledge of each interpreter. The set of SAR image versions was shown to five expert interpreters in the order of increasing data rates. Each decoded SAR image was initially resized to fit the screen. Most interpretation tasks were solved by marking the associated objects. The objects will be marked when the task is solvable with a probability of 80%. In this paper we focus on the most interesting interpretation task *Mark vehicles up to 6 m length*. Different image qualities of such marked vehicles are shown in Fig. 4.
As a second subjective task, the interpreters were asked to identify the image quality they preferably like to work with.

4 Interpreter-based Quality Metric

For the quantitative evaluation of solved tasks of the interpreter evaluation sessions, we derive a quality metric based on the well-known F_1-score [12]. The marking of the objects depends on the individual decision of an 80% probability of being able to solve the task. Thus, for each interpreter an individual F_1-score is calculated.
Using Recall

$$\text{Recall} = \frac{\text{no. of true detected objects}}{\text{no. of existing objects}} \tag{1}$$

and Precision

$$\text{Precision} = \frac{\text{no. of true detected objects}}{\text{no. of (true and wrong) detected objects}} \tag{2}$$

we get the well known F_1-score

$$F_1 = \frac{2 \cdot \text{Recall} \cdot \text{Precision}}{\text{Recall} + \text{Precision}} \tag{3}$$

which is between 0 (worst case) and 1 (best case). The latter case corresponds to a confident detection with F_1-score $= 1$.
Due to a missing ground truth we do not know the number of existing objects. Instead, we describe an individual ground truth for each interpreter. Defining the total number of objects found by this interpreter i in all image versions of one test image as *n true detected objects over all image versions*, we get an individual $\text{Recall}_{\text{ind}}$

$$\text{Recall}_{\text{ind}} = \frac{\text{no. of true detected objects}}{\text{no. of true det. obj. over all image versions}}. \tag{4}$$

We found that wrong detected objects do not occur in the investigated test surrounding. Hence, Precision simplifies to

$$\text{Precision}_{\text{ind}} = \frac{\text{no. of true detected objects}}{\text{no. of true detected objects}} = 1. \tag{5}$$

With Eqs. 4 and 5 we get an individual F_{ind}-score

$$F_{ind} = \frac{2 \cdot \text{Recall}_{\text{ind}}}{\text{Recall}_{\text{ind}} + 1} \tag{6}$$

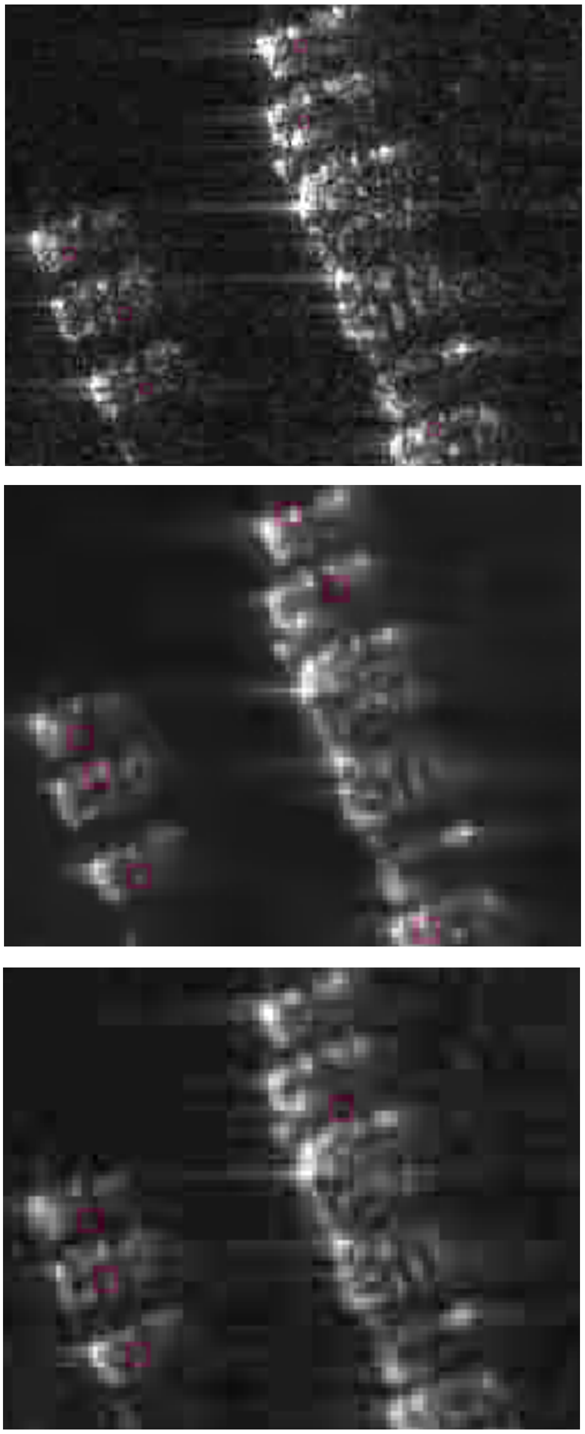

Figure 4 Different image qualities of marked vehicles

Each individual F_{ind}-score is weighted with the *n true detected objects over all image versions* of the associated individual person *i* and after averaging results in the *weighted* F_1-score

$$F_{1\_weighted} = \frac{\sum n_i \cdot F_{ind_i}}{\sum n_i} \tag{7}$$

The *weighted* F_1-score is an adequate quality metric used further on.

5 Results

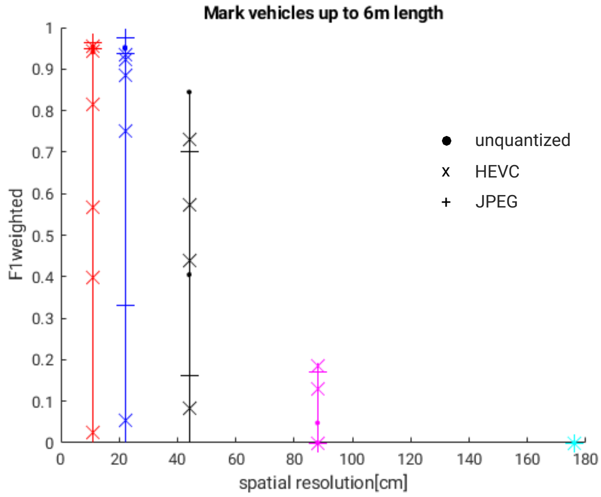

Figure 5 *Weighted* F_1-score over spatial resolution for HEVC intra mode and JPEG

Fig. 5 shows the *weighted* F_1-score over the spatial resolution for both test images.

As mentioned above, three image qualities have been investigated. Because of the fixed quantization parameters in resolution level 1, the data rates for each quality varies from test image to test image. Thus, all together three data rate ranges can be recognized: rate range 1 is located between 0.25 bit/pel and 0.55 bit/pel, rate range 2 between 0.025 bit/pel and 0.062 bit/pel and rate range 3 between 0.002 bit/pel and 0.004 bit/pel.

The *weighted* F_1-score is equal to zero at resolution level 5 which means that no vehicle was detected at the poorest spatial image resolution shown. Resolution level 4 gives very bad results with a *weighted* F_1-score below 0.2. Thus, a spatial resolution higher than 88 cm is neccessary to solve the investigated task. Hence, only level 1, 2 and 3 with spatial resolutions of 11 cm, 22 cm and 44 cm are investigated further on.

Fig. 6 shows the *weighted* F_1-score over the data rate for each test image for the spatial resolution levels 1 and 2. It can easily be seen that the unquantized image versions of the two resolution levels need the highest amount of data in every test image (Fig. 3). Likewise, the image versions achieved with data rates in rate range 3 achieve *weighted* F_1-score below 0.4 which shows that this rate range is too low and will not be considered subsequently.

Fig. 6 shows that at rate range 1 JPEG yields the best *weighted* F_1-scores with level 1 and 2. In contrast to that HEVC achieves higher *weighted* F_1-scores at rate range 2. It is not surprising that the HEVC coding system works much better than the JPEG coding system at the low data rates. But it was not expected that JPEG works even slightly better than HEVC at the high rates. We assume that this is caused by the modulation transfer function of the human eye [10] which is exploited by JPEG in its quantization but not by HEVC.

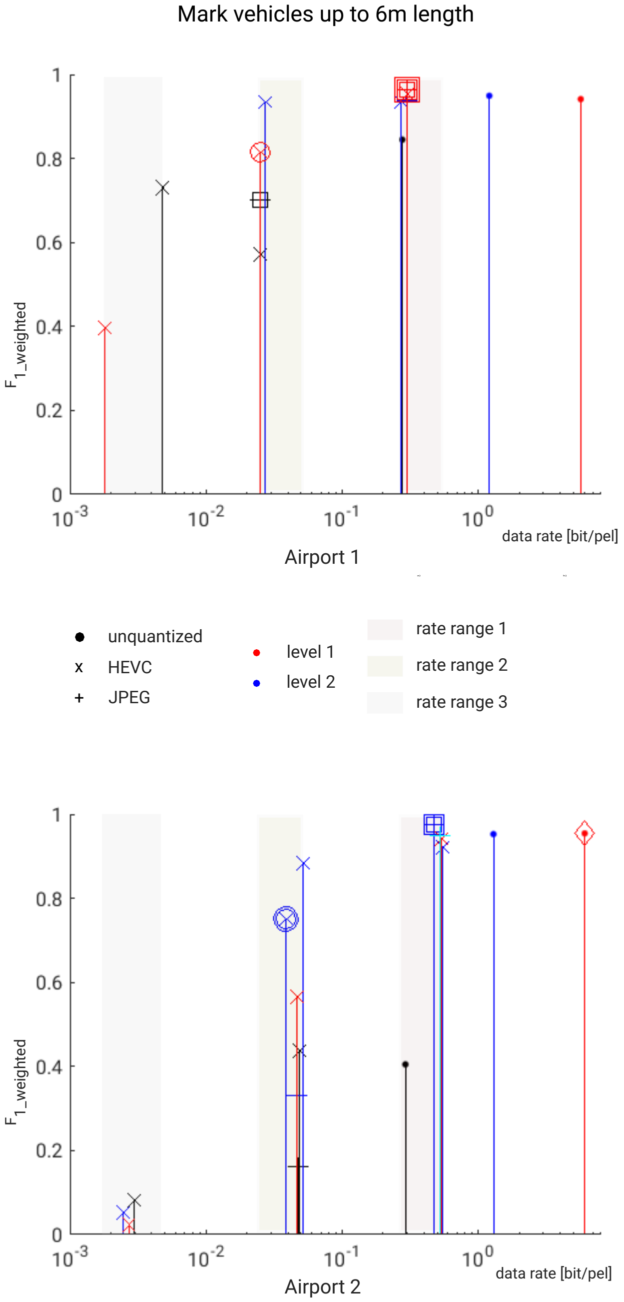

Figure 6 *weighted* F_1-score over data rate for HEVC intra mode and JPEG, level 1 and level 2 and most preferred image qualities

Furthermore Fig. 6 shows the most preferred image quality an interpeter would like to work with. It is marked with a circle, square or diamond, depending on the coding. Multiple votes are indicated by concentric enlarged markers. There is one outlier of the 10 most preferred images in test image *Airport 2* at a data rate above 6 bit/pel for the SAR reference image. For both test images Fig. 6 gives a good impression of the fact that there are two types of voting the most preferred image quality:

- The best *weighted* F_1-score is identically to the most preferred image version which verifies the derived quality metric.
- Image versions with a low rate in rate range 2 are voted as most preferred. In that rate range they are shown at an early stage to the interpreter. Thus, this rate range yields a sufficient image quality for solving the interpretation task without having a very high *weighted* F_1-score.

Thus, compression sometimes improves the usability.

6 Conclusion

In our usability-based system assessment framework, we have assessed the SAR coding component where the spatial and the amplitude resolution have been varied yielding image versions with different data rates. The two standardardized coding systems JPEG and HEVC in intra mode were investigated. Expert SAR interpreters were asked to solve interpretation tasks and to identify for each test image that coded image version they would most prefer to work with. A quality metric based on the well-known F_1-score was derived, called *weighted* F_1-score, for evaluating the solved tasks of the interpreter evaluation sessions.
For the two test images, the best *weighted* F_1-score for the JPEG coding system is at a data rate of 0.30 bit/pel using a spatial resolutions of 11 cm and at 0.47 bit/pel for 22 cm spatial resolution, respectively. The corresponding image versions conform with the majority of votes of the most preferred image version which verifies the derived quality metric *weighted* F_1-score. The second most votes of the preferred image versions was at data rates of 0.025 bit/pel and 0.039 bit/pel, respectively. These rates yield a sufficient image quality for solving the interpretation task without having a very high *weighted* F_1-score.
Finally, we conclude that compression, spatial and in amplitude, sometimes improves the usability.

7 Literature

[1] H. Klausing; W. Holpp: Radar mit realer und synthetischer Apertur: Konzeption und Realisierung, De Gruyter, ISBN 978-3-48-659897-1, 1999

[2] M. Phuong Nguyen: Bewegungskompensation und Autofokussierung von SAR-Rohdaten mit großem Schielwinkel, VDI Verlag, Fortschr.-Ber. VDI Reihe 10, ISBN 978-3-18-383710-6, 2014

[3] A. Sommer; J. Ostermann: Backprojection Autofocus of Moving Ships for Synthetic Aperture Radar, 15th European Radar Conference (EuRAD), Madrid, Spain, September 2018

[4] U. Pestel-Schiller: Coding of SAR Image Data for Data Compression, 10th European Conference on Synthetic Aperture Radar (EUSAR), Berlin, Germany, June 2014

[5] J. C. Curlander; R. N. McDonough: Synthetic aperture radar, Wiley New York, 1994

[6] C. Poynton: Digital Video; HD, M. Kaufmann, p. 40, Waltham, USA, 2012

[7] U. Pestel-Schiller; J. Ostermann: Usability-Based SAR System Assessment: Framework and First Re-

sults, 12th European Conference on Synthetic Aperture Radar (eusar), Aachen, Germany, June 2018

[8] U. Pestel-Schiller; J. Ostermann: Subjective Evaluation of Compressed SAR Images: Sometimes, Compression Improves Usability, 15th European Radar Conference (EuRAD), Madrid, Spain, September 2018

[9] J. Ohm: Digitale Bildcodierung: Repräsentation, Kompression und Übertragung von Bildsignalen Information und Kommunikation, url=https://books.google.de/books id=HaehBgAAQBAJ, 2013, p. 12, Springer Berlin Heidelberg, ISBN 978-3-64279359-2

[10] W. Pennebaker; J. Mitchell: JPEG, Still Image Data Compression Standard, Van Nostrand Reinhold, 1993

[11] ISO/IEC and ITU-T: Recommendation ITU-T H.265 and ISO/IEC JTC1/SC 29 23008-2:2015-05-01 MPEG-H Part 2/: High Efficiency Video Coding (hevc), 2nd Edition, 2015

[12] C. J. Van Rijsbergen: Information Retrieval, http://www.dcs.gla.ac.uk/Keith/Preface.html, Dept. of Computing Science, 2nd edition, University of Glasgow, 1979

[13] IRARS Committee: Civil NIIRS Reference Guide, https://fas.org/irp/imint/niirs.htm (accessed 12 October 2017), April 1996