Automated Preservation:
The Case of Digital Raw Photographs

Stephan Bauer and Christoph Becker

Vienna University of Technology, Vienna, Austria
http://www.ifs.tuwien.ac.at/~becker

Abstract. In digital preservation, a common approach for preservation actions is the migration to standardized formats. Full validation of the results of such conversion processes is required to ensure authenticity and trust. This process of *quality assurance* is a key obstacle to achieving scalability for large volumes of content. In this article, we address the quality assurance process for the preservation of born-digital photographs and validate conversions of raw image formats into standard formats such as Adobe Digital Negative. To achieve this, we rely on a systematic planning framework. We classify requirements that have to be evaluated according to their measurement needs. We extend an existing measurement framework using a combination of tools, image similarity algorithms, and purpose-built plugins. By combining metadata extraction, image rendering and comparison, and perceptual-level quality assurance, we evaluate the feasibility of automating the core part of quality assurance that is often the most costly part of preservation processes.

1 Introduction

In our society so strongly relying on digital data and communication for all purposes, digital preservation has become an important challenge: Only a functioning and appropriate technical environment can create a meaningful interpretation of digital objects. In the case of static content, where data structures specify the full semantics of the intellectual meaning of data streams, a common solution is to convert data into different representations to enable rendering in the environment of a certain user group. For preserving images, this corresponds to conversion to a standardised image format with long expected life time and widespread, solid tool support. These *migration* processes need to be fully validated to ensure that the intellectual meaning of the content is fully and authentically preserved. The migration itself can be relatively easily automated and parallelised to achieve scalability on large volumes of content. The key bottleneck lies in verifying the migration results.

Today, professional photographers take literally tens of thousands of pictures per month[1]. They normally rely on a so-called *raw* format that stores the raw

---

[1] http://www.nationalgeographic.com/educator-resources/
presidents-photographer/window-into-white-house/

data recorded by the camera's imaging sensor. The fundamental difference between the raw data and the interpreted data in a JPEG file is that the raw data itself needs interpretation in order to create a meaningful and presentable image, much like a negative film requires developing and printing to produce a photograph. Born-digital photographs captured in raw formats thus present a particular challenge, since the semantics of data captured by the sensors in digital cameras require complex interpretation processes to deliver a meaningful and authentic rendering. Without a trustworthy evaluation of different possible migration paths, no decision and conversion process can be trusted to deliver authentic and usable photographs. There is a vast amount of objects involved in large collections of raw camera data stored in proprietary formats containing specific camera profiles and the inherent uncertainties of migration. Hence, it is clear that such a conversion into standardised digital still image formats can only be successful with reliable and meaningful methods for automated Quality Assurance (QA). QA in this context refers to the process of measuring the significant properties of content and delivering measures of similarity according to specified criteria of interest in a structured form.

This article leverages a preservation planning framework to address this key issue in providing automation and scalability for the preservation of born-digital photographs. We elicit requirements that have to be evaluated and classify them according to their measurement needs. By extending an existing measurement framework using a combination of tools and purpose-built plugins for metadata extraction, image rendering and comparison, and perceptual-level quality assurance, we demonstrate automated QA for preserving digital photographs. We discuss experiment results and outline challenges lying ahead.

The remainder of this paper is organised as follows: Section 2 provides an overview of related work in image preservation and comparison. Section 3 discusses challenges and requirements posed by born-digital photographs. Section 4 demonstrates how to leverage an existing evaluation framework and extend it to achieve automated QA using existing software and new tools. The results of a case study are presented in Section 5. Finally, Section 6 draws conclusions and gives a short outlook on future work.

2   Related Work

In the digital libraries field, the discourse about preservation of images has largely focused on migration strategies converting scanned image content to standardised still formats such as TIFF-6 or JPEG 2000 [8]. A recent contribution discusses issues with JPEG 2000 as a preservation format [10]. Considering born-digital photographs, it must be noted that the term *raw format* does not describe one specific format. Each major vendor has multiple different types, most of which are partially proprietary and not standardized. To increase standardization, Adobe created an open raw format called Digital Negative (DNG).

The format is an extension of the TIFF-6 format and compatible with the TIFF-EP ISO standard [1] submitted for ISO standardisation[2]. The DNG format supports lossless compression as well as uncompressed storage of image data. The suitability of raw formats and DNG as archival still image format for digitised cultural heritage content was discussed in [4]. The authors provide a detailed discussion of the advantages of DNG, but do not propose or perform experiments to evaluate whether migration processes to DNG are preserving the significant properties of the images in question.

In the practitioner's community of digital photography, a certain awareness for the importance of digital preservation is emerging. The initiative *dpbestflow*[3] suggests best practice for storage and migration of digital photos. The recommended strategy is migration to DNG. However, no systematic analysis of the significant properties to be preserved is presented. Some limited experiments have evaluated whether conversion to DNG preserves the image content correctly[4]. However, the small scale of this experiment and the lack of a systematic evaluation approach result in insufficient evidence for drawing conclusions.

The systematic evaluation of preservation actions is the core strength of the preservation planning approach implemented by the planning tool Plato[5] [2]. It relies on controlled experimentation on sample content to enable evidence-based assessment of candidate preservation actions against scenario-specific requirements in a repeatable, well-documented workflow. The challenges currently in the focus of DP research are dominated by an urgent need to make operative preservation actions scalable, so that they can be applied to large volumes of content without the need for human intervention [3]. This requires a combination of tools, procedures and quality assurance mechanisms.

The DP community has systematically analysed the issues and factors influencing digital preservation operations. However, the degree of automation that can be achieved when executing preservation actions such as migration or emulation still strongly depend on the type of content and is in fact low for most object types and virtually non-existent for emulation [3].

Automated QA for image preservation requires methods to measure the similarity of images. Most of the common image similarity metrics focus on error related concepts such as Absolute Error (AE) or Mean Squared Error (MSE). These metrics have a solid mathematic background and can be used for discovering errors in conversion processes. However, they do not properly describe the human perception [11]. A recent promising approach is the structural similarity metric (SSIM) which corresponds more closely to the human visual system [12]. SSIM was originally designed for grayscale images. However, [9] compared the performance of

---

[2] See `http://www.dpreview.com/news/0805/08051503adobeDNG.asp` and
`http://www.iso.org/iso/standards_development/technical_committees/`
`other_bodies/iso_technical_committee.htm?commid=48420`

[3] `http://www.dpbestflow.org/`

[4] `http://www.libraw.org/articles/2-ways-to-nowhere.html`

[5] `http://www.ifs.tuwien.ac.at/dp/plato`

SSIM on different channels in comparison to subjective assessment. They showed its applicability to color channels and propose to apply SSIM on the HSV (Hue, Saturation, Value) color model.

3   Preservation of Born-Digital Raw Images

Multiple interpretation steps are required in order to get a meaningful rendering of the raw data recorded by a camera through an appropriate rendering software [5]: (1) *Demosaicing* interpolates missing color information. (2) *White balance adjustment* defines the ratio between red, green, and blue that shall be rendered as neutral white. (3) *Colorimetric interpolation* maps the arbitrary color values to a certain color space. (4) *Tone mapping* delinearises the data and performs a mapping to the desired bit depth. (5) Finally, *Image enhancements* may include noise reduction, adjustment of color balance and saturation, antialiasing, and sharpening. This process is commonly referred to as *developing* the raw file, analogous to the chemical process of developing exposed film in a dark room. It can be influenced by multiple parameters such as exposure compensation, contrast, and sharpness. There is no *a priori* correct representation of a raw file, and the results of interpretation heavily depend on the parameters and the tool used. The photographer strives to create a subjectively optimal interpretation by manipulating the parameters, thus creating a parameter set. The final result is often a set of derivative files in an image format such as JPEG or TIFF. Given the raw file, the interpretation information, and software capable of interpreting raw files, it must be possible to reproduce an interpretation of the recorded scenery that is exactly identical to the photographer's view.

Our case study is based on a collection of approximately 30.000 photos from about five different cameras. The sample test set contains raw files recorded by different cameras from different vendors using different settings, ranging from a file from a 2002 camera to brand-new raw files. Based on the scenario outlined above, we consider possible preservation strategies. Keeping the original raw files will be sufficient for short periods of time only, since raw capable software is often not supporting older raw file formats. For example, the raw processing engine in *Mac OS* does not support raw files created by *Canon Powershot S45* in version 10.6.7 anymore. However, storing the data in an interpreted format will not preserve the authentic form of the digital image. Normalization of raw files to DNG is potentially an ideal solution combining authenticity with contextual interpretation information, but requires in-depth validation.

Based on the taxonomy described in [3], we can classify the relevant decision criteria that need to be evaluated. Most of the critical criteria belong to the category **outcome object**, which describes desired properties of the outcome of a preservation action. Here, we can distinguish between significant properties of the context and of the content, as discussed in depth below. The characteristics of the chosen **object format** include format adoption, openness of specification, standardisation, stability, compression, the relative file size tendency, embedding of additional metadata and preview images, and whether or not the content is

already interpreted. For scalability reasons, **action runtime** criteria describing the performance and resource usage of conversion processes have to be considered. **Static action** characteristics include the ability to create error reports and performing file format verification and support for XMP sidecar files.[6]

The measurement of these categories has been discussed elsewhere [3], where outcome object properties were identified as the critical obstacle to scalability: They require in-depth analysis of content and generally constitute the dominating factor of trustworthy preservation decisions. The majority of these criteria refer to *significant properties*, i.e. *'the characteristics of digital objects that must be preserved over time in order to ensure the continued accessibility, usability, and meaning of the objects, and their capacity to be accepted as evidence of what they purport to record'* [6]. We will thus focus on the measures required to judge the degree of success in preserving these significant properties by performing fully automated QA instead of the manual validation prevailing to date.

4   Automated Quality Assurance

When transforming raw files into a different format, we need to ensure that this new representation is authentic, i.e. that it represents the same photograph with all its significant properties. Fundamentally, we need to distinguish between QA of content and of context related criteria. Table 1 lists the most important complementary metrics used to verify that content and context are fully retained.

Since the original and the migrated raw file are encoded in entirely different representations models, no formal equivalence relation can be defined. For example, during the course of DNG conversion, specific camera-dependent color matrices are embedded into the DNG file. When creating an interpretation of the DNG file, these are used to perform a mapping from the arbitrary camera color space to the well defined CIE XYZ space [1]. However, we may assume that *two raw files can be called equal if (and only if) they can be transformed into the same interpreted form.* In other words, the original and the migrated raw file can be called equal if they are rendered the same way in a reference rendering environment. In the absence of ground truth for verifying conversion and renderings, we need to rely on a standarized and widely used rendering environment.

The open source command line tool *dcraw*[7] supports full raw interpretation to different uninterpolated as well as fully interpreted derivative formats. For each conversion, dcraw can thus create two pairs of uninterpolated and interpreted image representations in 16-bit high quality TIFF. Subsequent comparison of these using a range of complementary metrics enables automated evaluation. The error-related criteria on the top rows of Table 1 verify that the pixel data remain unchanged. They apply to the technical representation of the uninterpolated image. But as discussed above, equality of uninterpolated data is not sufficient:

---

[6] These standalone files in Extensible Metadata Platform (XMP) format can include metadata which is added by raw capable software, mostly used for interpretation information and descriptive metadata.

[7] `http://www.cybercom.net/~dcoffin/dcraw/`

Table 1. Key content and context properties to be measured

Property	Description
relative AE	describes the percentage of pixels that differ when comparing the uninterpolated images
relative MSE	calculates the mean squared error between two uninterpolated images divided by the bit depth
SSIM Value	the raw data is transformed into fully interpreted 16-bit TIFF, and the brightness of the HSV representation is compared using SSIM
SSIM Saturation	SSIM is applied to the S-channel of HSV for comparison
SSIM Hue	SSIM is applied to the H-channel of HSV for comparison
Exif (exposure)	ExposureTime, FNumber, ExposureProgram, ExposureMode, ISOSpeedRating, SubjectDistance, Flash, FocalLength, ...
Exif (technical)	ImageWidth, ImageHeight, CFAPattern, Make, Model, ...
Exif (location)	GPSLatitude, GPSLongitudeRef, GPSAltitude, ...
Exif (generated)	Metadata added during migration to enable correct rendering: ColorMatrix, AsShotNeutral, BaselineExposure, ...
IPTC	Creator, ContactInfo, SceneCode, Location, City, Province-State, Country, Headline, Description, Keywords, ...
Dublin Core	Title, Subject, Description, Source, Coverage, Creator, Rights, Date, Format, Identifier, Audience, ...
Private Tags	e.g. ColorTemperature, WhiteBalance, InternalSerialNumber, AFPointsSelected, ...
XMP	XMP contains Exif and IPTC metadata as well as custom information as used by Adobe Camera RAW and others

Interpretation parameters, such as the color transformation to apply, constitute an integral part of the semantic structure of the photographs. We thus need to compare fully interpreted images as well. This comparison should be aligned to the human perception and thus relies on the SSIM metric, which is applied to the three HSV channels Hue, Saturation and Value.

An in-depth analysis of the applicability of the commonly used image comparison tool *ImageMagick compare*[8] to the presented case reveals a number of shortcomings. Color systems different from RGB are not readily supported, and parsing the results is error prone as there is no standardised output format. The tool is very sensitive when handling TIFFs. TIFF files with embedded metadata cause numerous warnings and can often not be compared successfully. Stripping metadata in advance may be an option, but increases the overhead and introduces potential error sources. Moreover, color profiles are not correctly handled, as can be demonstrated with test images that use a degenerated color space[9]. Comparing two different images that have the same visual appearance due to their ICC profile are wrongly classified as very different. Finally, the comparison does not support perception-based metrics such as SSIM. On the other hand, Java provides features to perform correct color management and Java Image IO API plugins that handle TIFFs. We have thus implemented all content related

---

[8] http://www.imagemagick.org; (version 6.6.1-5).

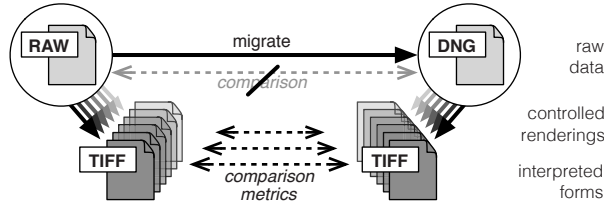[9] http://www.gballard.net/psd/go_live_page_profile/

Fig. 1. Perceptual-level QA relies on controlled renderings for comparing objects

comparison metrics using the Java Advanced Imaging API[10]. The presented solutions are fully integrated in the upcoming release of the planning tool Plato.

The context related properties of a raw file contain different types of metadata information, such as information about the picture taking conditions (Exif[11] as well as private tags, i.e. vendor-specific metadata fields), technical descriptions of the equipment in use, and additional contextual information (IPTC[12], especially Dublin Core). Table 1 lists examples for relevant metadata criteria.

These properties are verified by extracting each field from the source and target objects and comparing the features, thus extending the extraction functionality described in [3]. Specifically, the QA methods included in Plato define evaluator plugins that can measure certain properties. These are defined by a URI-like identifier that contains the measurement domain, a unique property pathname and an optional metric [3]. The identifiers are linked to the corresponding criteria in Plato in order to automatically perform evaluation and report the results. In order to verify the preservation of metadata in a migration process, we have integrated the widely used *ExifTool*[13] for metadata extraction. Each property of interest is assigned an identifier, which is mapped to the corresponding metadata and processed by the evaluation plugin. Metadata can be grouped to cover related fields without the need for explicit descriptions of each specific field. For instance, for verifying if *makernotes* are retained, the comparison tool in its current configuration automatically checks whether all private tags are retained.

5   Experiment Results and Discussion

Space constraints our discussion on two common migration options from RAW to DNG, *Adobe DNG Converter 6.1* (ADC) and *digiKam 0.6.0-svn*. Table 2 shows exemplary results of automated evaluation performed on random samples of different source formats and contrasts these results with manual validation in Adobe Camera Raw (ACR) and dcraw. Problematic migration paths are shaded.

---

[10] http://java.sun.com/javase/technologies/desktop/media/jai/

[11] The 'Exchangeable image file format' describes structure and tags for use in images.

[12] The 'Information Interchange Model' by the 'International Press Telecommunications Council' defines structure and metatags for various file types.

[13] http://www.sno.phy.queensu.ca/~phil/exiftool/

Table 2. Conversion to DNG: Adobe DNG Converter (ADC) and digiKam (dKam)

	Canon Power-shot S45 (CRW)		Canon EOS 5D Mk2 (small CR2)		Canon EOS 40D (CR2)		Nikon D80 (NEF)	
Criteria	ADC	dKam	ADC	dKam	ADC	dKam	ADC	dKam
dimensions (dcraw)	ident.	ident.	ident.	-	ident.	ident.	smaller	ident.
dimensions (ACR)	ident.	larger	ident.	-	ident.	larger	ident.	larger
relative AE	0%	0%	99.9%[a]	-	0%	0%	0%[b]	0%
relative MSE	0	0	0.00791[a]	-	0	0	0[b]	0
SSIM Value	1.0	0.84	0.85	-	1.0	1.0	0.92[b]	1.0
SSIM Saturation	1.0	0.67	0.85	-	0.91	1.0	0.64[b]	1.0
SSIM Hue	1.0	0.76	0.9	-	0.94	1.0	0.94[b]	1.0
manual (dcraw)[c]	ident.	diff.	simil.	-	simil.	ident.	simil.	ident.
manual (ACR)	ident.	diff.	ident.	-	ident.	diff.	ident.	diff.

[a] *dcraw* delivers a color image instead of a grayscale image

[b] comparing the cropped area of the DNG

[c] manual evaluation distinguishes between different, similar and identical

Generally speaking, the Adobe DNG Converter achieves much more consistent solid results in comparison to the free open source tool *digiKam* (dKam). A comparison of the SSIM distances shows that in several cases, differences are introduced that merit closer attention. In the case of the NEF file, the SSIM saturation metric goes as low as 0.64. However, manual evaluation does indicate that results are very satisfying, since the human perception is much less sensitive to changes in saturation than, for instance, to shifts in color space. While the exact correlation of manual evaluation results and thresholds for SSIM on the different channels will depend on context and risk aversion, the SSIM can provide a good indicator of problematic areas that require in-depth validation.

Migrating Canon S45 files with dKam results in files that are equal to the original on the uninterpreted level, but show strong differences when color interpolation is included. The high SSIM value indicates that the grayscaled image itself is retained, but the low SSIM saturation and hue values indicate strong differences in color. Even when manually correcting white balance in Adobe Camera RAW (ACR), reconstruction of the original color appearance is not possible. In-depth analysis reveals that the underlying reason is the incorrect color matrix in the Exif section, which does not properly represent the camera's profile.

The small CR2 file produced by the EOS 5D, which is downscaled to one fourth of the sensor's original size in the camera, is especially difficult to handle. It cannot be converted at all by dKam. For the file created by ADC, dcraw is not able to correctly extract the raw data without color interpolation. Although no official information exists, it is assumed that Canon's sRAW stores intensity values similar to a YCbCr model, whereas the classical RAW data is based on a RGB model [7]. A comparison of the color interpolated images indicates minor differences, but manual evaluation using ACR shows no noticeable differences.

For the EOS 40D photographs, comparing the uninterpolated data shows no differences between CR2 and DNG. For the interpreted rendering, there are

minor differences in the DNG created by ADC. The large color-related SSIM score indicates that the difference is not perceivable, which is confirmed by both manual checks. The DNG produced by dKam gets perfect scores based on the renderings created by dcraw, but manual validation against the rendering of ACR reveals strong differences. A possible explanation is that the DNG file created by the ADC contains two Exif color matrices, whereas the DNG created by dKam contains only the second of these. Given two matrices, a raw converter may perform color mapping by interpolating between these matrices and using additional calibration matrices. Additionally, the white balance information is represented differently by ADC and dKam in this case. As a result, rendering the DNG created by ADC in dcraw produces an incorrect white balance adjustment, as does rendering the DNG created by dKam in ACR.

An unexpected level of difference occurs in the case of NEF files: The resulting images are not even of the same dimensions. The reason lies in the fact that color interpolation is very difficult at the sensor borders, as there are not enough neighbors to interpolate easily. In-camera software often discards these border pixels, while raw processing software may use more processing-intensive algorithms and take those pixels into account. This may result in differently sized images, as seen for the NEF file in our case study. For the DNG created by ADC, the un-interpolated image exported by dcraw is slightly smaller than the export of the original NEF. Only by ignoring the extra rows contained in the original is a direct comparison possible. The uninterpreted image created by ADC is then equal to the original. The color interpolated files still differ, since the DNG created by ADC is slightly brighter. Manual validation using the interpretation rendered by ACR, however, shows that the DNG created by ADC is identical to the original NEF, whereas the DNG created by dKam is slightly brighter. The most reasonable explanation is that ACR correctly respects the exposure metadata, whereas dcraw ignores it.

From the experiment results, a number of observations can be drawn. The correctness and reliability of the intermediate representations used for QA is a crucial element in the evaluation chain. Yet, several shortcomings of existing tools are evident. Nevertheless, we can use each of these tools to discover potential failures in migration. Using the SSIM metrics, we can further distinguish between minor and major differences and classify the difference in terms of structure related (low SSIM Value) and color related (low SSIM Saturation, low SSIM Hue). Comparison metrics can be used more successfully to *falsify* conversions than to *verify* them: False negatives are more likely than false positives. Furthermore, different renderings are required to uncover the variety of potential errors. In this sense, it is advisable to rely on a combination of metrics and interpretation processes to uncover hidden gaps in conversion and evaluation.

6   Conclusion and Outlook

This article has described an automated preservation experiment on born-digital raw photographs that combined controlled experimentation, integration of existing metadata extraction and image comparison tools, and implementations of

state of the art image comparison algorithms. In the case of photographs, metrics and algorithms exist to validate conversion processes, but a direct comparison of the artifacts resulting from preservation actions to the original content is not possible. Hence, we rely on perceptual-level QA to validate conversions. By combining controlled rendering paths and comparison metrics, it is in principle feasible to fully automate preservation operations on non-trivial content. This can be done without violating fundamental requirements such as a full validation of the significant properties of content. However, this validation approach relies on intermediate processing steps with varying reliability. If the interpretation step itself contains conversion errors, this may result in false negative reports of differences between identical files or, even worse, false positive case where a converter creates two identical images (e.g. all black) out of different images. The former case may be identified by cross-checking against different renderings, the latter by checking image properties such as histograms for unusual patterns. Although confidence in the results can be improved by comparing the results of multiple conversion tools, absolute certainty would depend on ground truth. As in many comparable cases in digital preservation, this utter lack of well understood test corpora with known ground truth is a significant impediment against full validation of interpretation and processing algorithms [3]. Such corpora would be a massive enabler for improvement of digital preservation processes.

Current work is focused on developing a modular service deployment of the employed QA mechanisms to prepare large-scale experimentation on the entire data set using a workflow engine. This will be combined with a detailed assessment of the correlation between SSIM and manual validation.

References

1. Digital negative (DNG) specification. version 1.3.0.0. Technical report, Adobe Systems Incorporated (June 2009)
2. Becker, C., Kulovits, H., Guttenbrunner, M., Strodl, S., Rauber, A., Hofman, H.: Systematic planning for digital preservation: Evaluating potential strategies and building preservation plans. IJDL 10(4) (December 2009)
3. Becker, C., Rauber, A.: Decision criteria in digital preservation: What to measure and how. JASIST 62(6) (2011)
4. Bennett, M.J., Wheeler, F.B.: Raw as archival still image format: A consideration. In: Haag, D. (ed.) Archiving 2010, pp. 185–193. Society for Imaging Science and Technology, The Netherlands (2010)
5. Fraser, B., Schewe, J.: Real World Camera Raw with Adobe Photoshops CS4. Peachpit Press, Berkeley (2009)
6. Grace, S., Knight, G., Montague, L.: InSPECT Final Report. InSPECT (December 2009), http://www.significantproperties.org.uk/inspect-finalreport.pdf
7. Kerr, D.A.: The canon sraw and mraw output formats. The Pumpkin (1.1) (March 2010)

8. Kulovits, H., Rauber, A., Brantl, M., Schoger, A., Beinert, T., Kugler, A.: From TIFF to JPEG2000? Preservation planning at the Bavarian State Library using a collection of digitized 16th century printings. D-Lib Magazine 15(11/12) (2009)
9. Okarma, K.: Colour image quality assessment using structural similarity index and singular value decomposition. In: Bolc, L., Kulikowski, J.L., Wojciechowski, K. (eds.) ICCVG 2008. LNCS, vol. 5337, pp. 55–65. Springer, Heidelberg (2009)
10. van der Knijff, J.: JPEG 2000 for long-term preservation: JP2 as a preservation format. D-Lib Magazine 17(5/6) (2011)
11. Wang, Z., Bovik, A.C.: Mean squared error: love it or leave it? - a new look at signal fidelity measures. IEEE Signal Processing Mag. 26(1), 98–117 (2009)
12. Wang, Z., Bovik, A.C., Sheikh, H.R., Simoncelli, E.P.: Image quality assessment: From error visibility to structural similarity. IEEE TIP 13(4), 600–612 (2004)