Fraunhofer
IIS

JPEG 2000 process for the project "Images for the Future" at the Netherlands Institute of Sound and Vision

Final Report

Dipl.-Ing. Arne Nowak
Dipl.-Ing. Joachim Keinert
Dipl.-Ing. Klaus P. Bönisch

15/10/2009

Document History

Version	Date	Changes	Contributors
V0.1	20.07.09	Document created	nok
V0.2	12.09.09	Initial version	nok, ket, boeniskk
V0.3	22.09.09	Document refined, comments by S&V incorporated, sections about scanning and JPEG2000 MTF added	nok, ket, boeniskk
V0.4	01.10.09	QA section refined	nok, ket, boeniskk
V0.5	02.10.09	Main sections extended and refined	nok, ket, boeniskk
V1.0	03.10.09	Final version of document	nok, ket, boeniskk
V1.1	14.10.09	Requests and comments from S&V incorporated	nok, ket

Contributors

Abbreviation	Name	Organization
nok	Arne Nowak	Fraunhofer IIS
ket	Joachim Keinert	Fraunhofer IIS
boeniskk	Klaus Patrick Bönisch	Fraunhofer IIS

Fraunhofer IIS

Contents

Fraunhofer IIS

1 Introduction

1.1 Overview

Within the Images for the Future programme the Netherlands Institute of Sound and Vision (S&V) plans to scan a large amount of mostly 16 mm reversal film material to a digital format with the main goal to make the films easily accessible for its clients, but also to create a digital copy for preservation purposes. This digital master is a DPX that is stored "on the shelf". To create accessibility other file formats will be used.
On one hand, an XDCAM 422HD (50 Mbit/s) file will be generated of each film for immediate access and use in professional broadcast environments. On the other hand, due to the relatively low quality of the XDCAM files compared to the uncompressed scanned images, the scanned files will be compressed to a higher bit rate JPEG2000/MXF file. These files can then be stored as a managed object and be used in the future to create higher quality deliverables for customers that are not satisfied by the quality level that XDCAM is able to provide. The data rate of the JPEG2000/MXF files is planned to be between 150 and 300 Mbit/s but should be high enough to provide a quality that does not lead to a visually perceivable loss of quality even in the case that image manipulation processes like color correction etc. are performed on the compressed images. The reasons for the restriction of the data rate are mainly availability and costs of managed digital storage space. Both, the XDCAM and JPEG2000/MXF files, will be ingested into a managed digital archive system. Figure 1 gives an overview of the overall process as it has been planned by S&V.

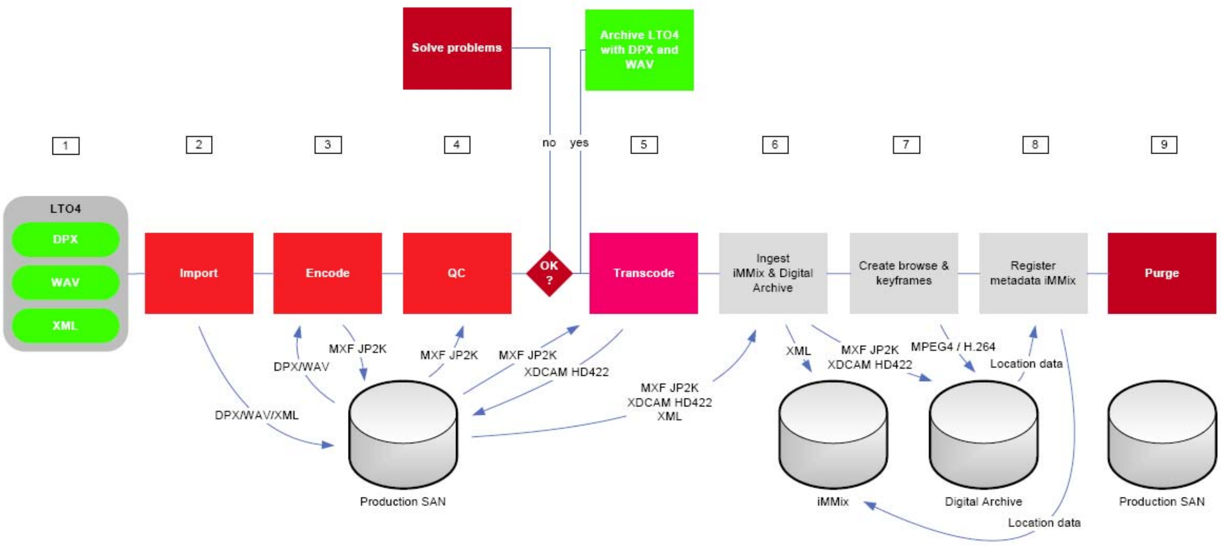


Figure 1: Process planned at the Netherlands Institute of Sound and Vision (diagram provided by S&V)

The scanning is planned to take place at the facilities of external service providers. After scanning the uncompressed DPX, audio and metadata files are provided on LTO-4 digital data tapes. These tapes are used to ingest the material to a storage area network (SAN) system to create the XDCAM HD422 and JPEG2000/MXF files. Additionally, these LTO-4 tapes will be stored on shelves in order to have a possibility to go back to the original uncompressed data in case that this becomes necessary in the future.
A quality control step is included in the process to make sure that especially the image quality of the JPEG2000 compressed material is up to the expectations and requirements of S&V.

1.2 Goals of this consulting project

In order to carry out the described digitization and encoding project S&V needs to define a set of technical requirements, parameters and settings for the scanning and encoding of the film images. Specification documents will be created by S&V for scanning of the films and for realisation of necessary systems and the implementation of the process of encoding the material to JPEG2000/MXF and XDCAM HD422 files. The Fraunhofer Institute for Integrated Circuits IIS has been asked to provide advice and recommendations on the technical details for scanning and encoding based on the nature of the films that are to be processed. This advice encompasses parameters for scanning as well as JPEG2000 encoding and especially the required JPEG2000 data rates that are minimal necessary to achieve the goal of visually lossless compression under the conditions described above.

1.3 Scope of this project

According to the contract between S&V and Fraunhofer IIS from July 14th 2009 the following points have been defined as the scope of this project.

- Definition of formats and data processing paths:
  Experts from Fraunhofer will discuss with other project experts from the Netherlands Institute of Sound and Vision the possible formats and data processing paths. This includes:
    - Color space specifications and recommendation for picture processing including investigation of linear vs. logarithmic quantization
    - Format specifications for MXF and JPEG2000 code streams (definition of JPEG2000 parameter constraints)
    - MXF and metadata assessment for film archive applications
    - Data error detection mechanism

  N.B.: Due to constraints of available storage space mainly data rates for JPEG2000 between 150 Mbit/s and 300 Mbit/s should be considered for this project. JPEG2000 mathematically lossless coding will also be included in the compression tests because it might be reasonable to use it for certain parts of the collection.
  If necessary, the specifications can be used for the definition and realization of the services and systems required for the project. The Netherlands Institute of Sound and Vision

however decides by its own, what part of the specifications they want to use. Fraunhofer takes no responsibility on the correct implementation of the project.

- Definition of quality control measures:
  In this work quality control mechanism are described, which offer structural and visual quality control options. Advice on objective quality control mechanisms on key points in the entire processing chain. The tools to realize this are not part of the quote, but described in chapter 9 of the Annex to the quote (long term strategy for cooperation).
- Compression tests: Fraunhofer will realize quality tests with different compression parameters for the workflow: Concrete conversion from DPX to JPEG2000 with different parameter sets, quality judgment (PSNR and visual blind tests in the IIS theatre), conversion to XDCAM HD422. PSNR-to-bit rate ratio will be analyzed and plotted. All tests will be carried out and analyzed separately for different types of source material to be specified and provided by the Netherlands Institute of Sound and Vision (e.g. BW, color, original, dup neg).

During the first discussion meeting at S&V the particular questions of the MXF format have been exempted from this project's scope because S&V already has an expert for this field. During the discussion meetings S&V and Fraunhofer IIS agreed to put the main focus of this project on the technical details of the film scanning process and the JPEG2000 quality assessment and data rate analysis.

1.4 Deliverables

As a result of this project the following deliverables have been produced.

Documents:
- Workflow and parameter specification document (WORD Format)
- Compression test result document (WORD Format)

Test Data:
- Set of JPEG2000 test data derived from the test data of the Netherlands Institute of Sound and Vision

In order to provide a better view on the whole project the two documents have been united in this final report. The test data set is delivered on DVDs together with the printed version of this report.

1.5 Structure of this document

After the introductory section this report is sub-divided into the main sections film scanning, image compression, subjective quality assessment and objective quality assessment followed by recommendations on these subjects. The report closes with a concluding section.

2 Film scanning

2.1 Overview

The goal of scanning the archive film material is in the first place to preserve as much information as possible from the original film image in order to be able to use the scanned images as universally as possible. This makes the resulting files usable for relatively easy access to the material on a very high quality level without having to rescan the film. For long-term preservation at S&V both the DPX and the original film will be stored under appropriate environmental conditions.
The films will be scanned to the DPX file format. The DPX files will be delivered for further processing on LTO data tapes. The LTO tapes will also be stored by S&V with a long-term perspective as a digital master. The processing consists of

1. compression to a lossy JPEG 2000 profile with a data rate probably in the region of 80 to 300 Mbit/s in the original spatial resolution the film was scanned and
2. compression to the XDCAM HD422 format (MPEG2 Long-GOP) with a data rate of 50 Mbit/s and a spatial resolution of 1920 x 1080. This includes, if necessary, transformation to the ITU-R Rec. 709 color space and a possible color correction with one setting for the complete reel of film (one light scan/correction) to provide a viewable digital version of the film.

When scanning and digitizing motion picture film two major decisions have to be taken:

a) an appropriate spatial resolution has to be chosen that should preferably be higher than the maximum spatial resolution of the film material in order not to lose information;
b) a color representation has to be chosen that provides enough precision for subsequent processing and correction steps and is well defined in terms of allowing to make a (physically) accurate statement regarding the original perceived color present on the film material or to be seen on the screen.

Both decisions depend on several complex factors. They are on the one hand strongly influenced by the type and condition of the original film material. This refers to the question if the material is 16 mm or 35 mm film, color or black and white, negative, print, reversal or dup film etc. and the general condition of the films. Badly faded or otherwise deteriorated material will have to be handled different from films that are in a good condition. On the other hand the decisions will be somehow limited and restricted by the availability and cost of equipment and software and the practicability and efficiency of the overall process. Especially the last point is of high importance in this project since a relatively large amount of around 7000h of films has to be processed.

In the following sub-sections the technical foundations of film scanning, color image processing and file formats are described as a basis for recommendations and options to choose from regarding the scanning process.

2.2 Theory of film scanning

A thorough understanding of the technical principles of film scanning is necessary to take the right decisions how this process should be carried out in detail. In the following sub-sections the most important points will be laid out.

2.2.1 Color

Color imaging with photochemical as well as electronic systems is based on the functional principle of the human visual system (HVS). In the human eye three different types of color receptors, the so called cone cells, can be found. Each type has its own spectral response which makes it most sensitive to light with a within a relatively narrow band of specific wave lengths. These bands represent the perceived colors blue, green and red and the three bands in total cover to visible spectrum. This trichromatic vision means that any perceived color can be represented by a mixture of three linearly independent primary colors. This was discovered by Hermann Graßmann in 1853 [5] and lead to the development of the CIE standard observer color matching functions (Figure 2), the CIE xy chromaticity diagram (Figure 3, left) and the CIE xyY color space in 1931 [6].

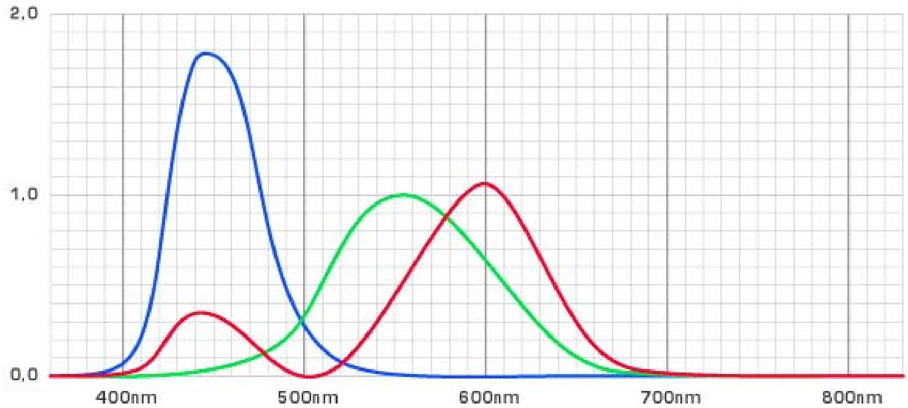


Figure 2: CIE standard observer color matching functions

The left side of Figure 3 shows all colors that can in theory be perceived by the average human eye with the outer curved boundary as the spectral or monochromatic locus (these are the colors of monochromatic light with the wavelength shown in nanometers next to the curve). The right side of Figure 3 shows the set of colors that can be created by a linear combination (or mixture) of the three primary colors that are represented by the circles in the corners of the triangle. This set is called gamut and its size depends on the location of the primary colors in the CIE xy diagram. The resulting RGB color space is completely defined by the color location of its three primary colors and the achromatic color (white point). The white point depends on the illumination and can be described as a color temperature, in this case 6500 Kelvin (D65). Several standard illuminants and corresponding white points are defined and a mathematical conversion between them is possible.

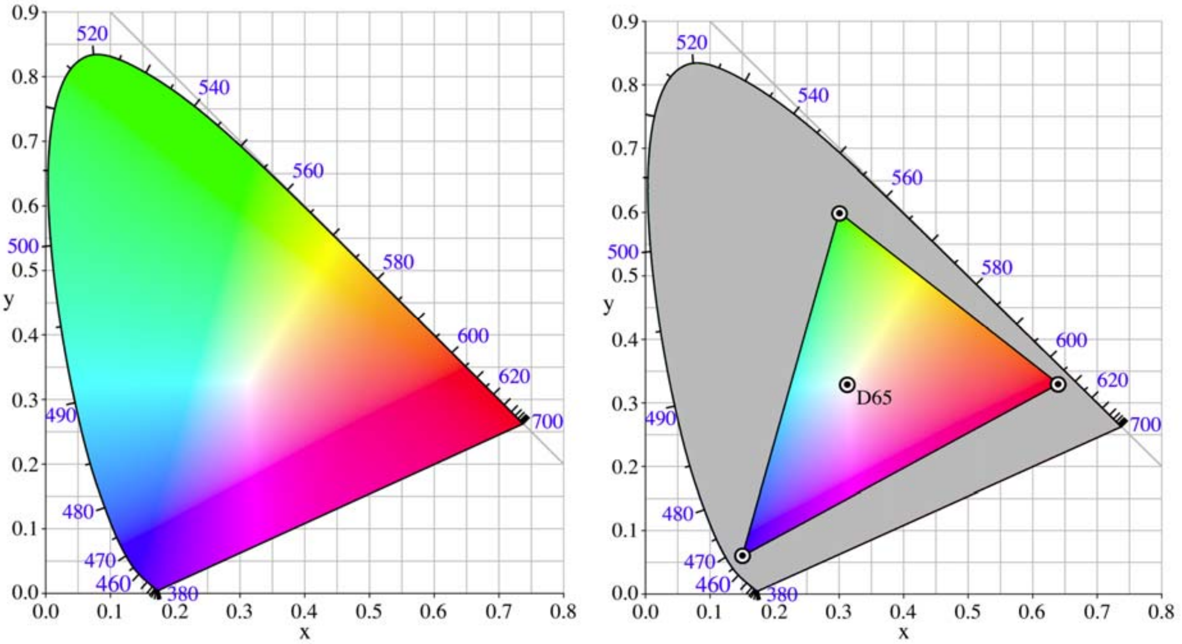


Figure 3: CIE xy chromaticity diagram (left) and RGB color space defined by three primary colors and white point (right)

In a digital RGB representation the color of a picture element (pixel) is described in the form of a three element vector where each element represents the percentage of each primary color in the mixture. To deduce from these three values the originally perceived color the location (CIE chromaticity coordinates) of all the primary colors and the white point in the CIE yx diagram have to be known!

2.2.2 Functional principle of film scanners

A film scanner basically measures the spectral transmittance of the film material with the help of a light source, color filters and a light intensity sensor or a camera (Figure 4).

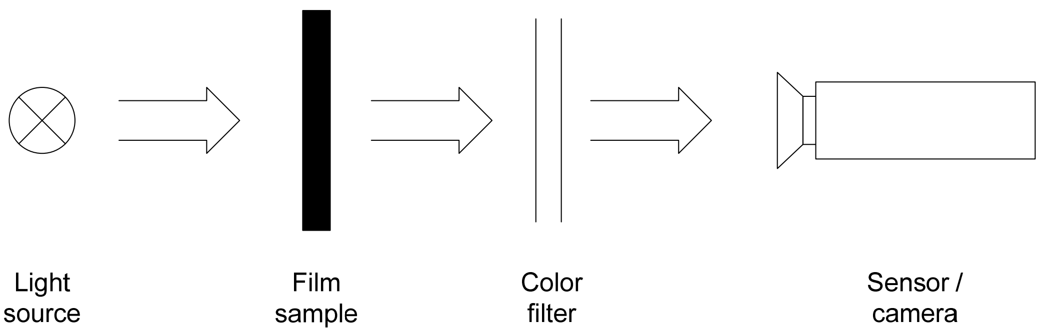


Figure 4: Basic functional principle of a film scanner

In the case of a color scanner there is one color filter for each of the three necessary primary colors. Depending on the technical solution either three images will be recorded by the camera for each original image on the film with the different color filters or the scanner uses three cameras and a beam splitter system. The latest scanner models that use LED light sources have three different LED types for the primary colors that are flashed sequentially to achieve the same goals.
The output values of the three components for one pixel now depend on

- the color spectrum of the light source,
- the spectral transmittance of the piece of film to be scanned at the position of the pixel,
- the spectral transmittance of the color filter in front of the camera, and
- the spectral sensitivity of the imaging sensor in the camera.

Since the transmittance of the film is the parameter that has to be measured all other parameters of the light source, the color filter and the camera have to be known. The CIE chromaticity coordinates for each primary color can be calculated by multiplying the spectrum of the light source with the transmittance of the color filter, the spectral sensitivity of the camera and the CIE color matching functions for each wavelength and integrating the resulting curve over the wavelengths of the visual light spectrum.
The result will usually be a color space that is unique to the scanner that is used. However, the measure RGB values can be transformed to other well-defined color spaces like CIE XYZ or ITU-R Rec. BT.709 using 3D look-up tables (LUT) or matrix transformations. This possibility is usually integrated into the scanner system or the control software and can be changed by the user. To achieve optimal results most scanners provide additional filters that adapt the parameters to specific types of film. Also adjustments of the intensity of the light source and the sensitivity of the camera for each color channels can be made to achieve maximal signal to noise ratios. These adjustments normally also change the color representation so that it is no longer possible to deduce the original perceived color from the film sample from the digital data. Therefore, the details and values of the adjustment should be stored together with the image in order not to lose information.

2.3 Data and file formats for scanning

2.3.1 Density measurement and transfer curves

The digital data that is read from the analog to digital converter of the scanner's camera normally has a linear relation to the intensity of the light that hits the imaging sensor. However, the response of the human visual system to light intensity is highly non-linear. For practical reasons in imaging applications a logarithmic relation can be assumed. This means that the HVS is much more sensitive to a specific absolute change of the luminous intensity at a lower intensity level than it is to the same absolute change at a high intensity level. Especially in a digital system it makes sense to store the intensity values using a non-linear transfer curve to ensure that the available precision (i.e. bit depth) is distributed across the range of input values in a way that the representation is as close as possible to the response of the HSV.
Therefore, when measuring or scanning film not the spectral transmittance is used but instead spectral optical density. Optical density is defined as

$$D = -\log_{10}\left(\frac{I_t}{I_0}\right) = -\log_{10}(T),$$

where D is the optical density, I_0 is the intensity of the light source and I_t is the intensity that is registered at the imaging sensor. T is the symbol for transmittance. For color film the density values depend on the spectral transmittance curve of the color filters for each primary color. The density values described above are commonly called logarithmic values. A logarithmic relation also exists between the exposure of photographic material and the resulting density (Figure 5).

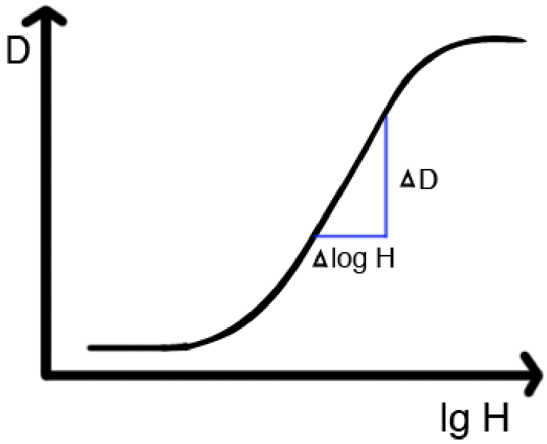


Figure 5: Logarithmic exposure to density diagram

Coming from opto-chemical processes there are several standards for measuring the density of film material. The most commonly used are Status M density and Status A density. Both are defined in the international standard ISO 5-3:1995. Status M density is defined for negative film while Status A density is used for print and reversal film material. Several digital film scanners provide settings for measuring Status M density. Another important term is Cineon Printing Density. Here, no well defined spectral response curves are used. Instead the scanner is calibrated together with a corresponding digital film recorder in order to deliver a specific result on the developed print film. Printing density is a description of the density that "the print film sees on the recorded negative." The system is calibrated for a very specific combination of original negative, recording negative and print film. In the SMPTE Recommended Practice RP180 spectral response curves are described but they are only valid for exactly one combination of material. Using Cineon Printing Density in a film scanner therefore usually does not lead to a well-defined result in a known and defined color space. The characteristics for modern film material are normally published by the manufacturers in data sheets and can be used to define a optimized transfer curves and color space for scanning. These curves have to be stored with the images to ensure correct interpretation of the data. Modern film materials can be automatically recognized using bar codes (key codes) that are put onto the film during its manufacturing. Unfortunately, data sheets and exact knowledge about the properties often do not exist for archive films.
In video and television applications another transfer curve is used to process the linear values from the imaging sensor: a so-called gamma curve. This curve was originally used as a pre-distortion to adapt the video signal to the non-linear properties of a cathode ray tube (CRT) display. The

distortion is applied in the camera and in the case of a CRT display reversed by the CRT itself. The gamma curve can in its simplest form be described as

$$U_{out} = U_{in}^{\gamma},$$

where U_{in} is the voltage of the input signal, U_{out} is the voltage of the output signal and γ is the gamma factor. Figure 6 shows a gamma correction and the transfer curve of a typical CRT display. The term gamma factor photography originally described the slope of the linear part of the log(H) to D curve (Figure 5).

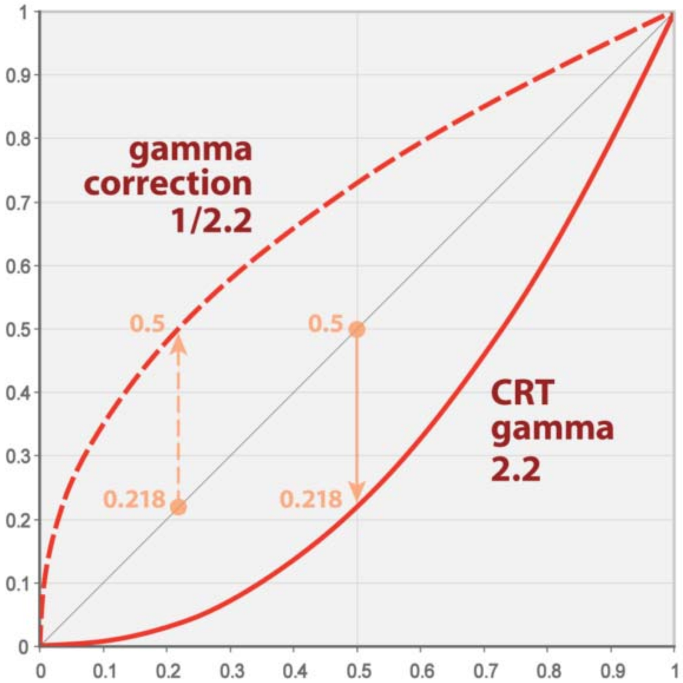


Figure 6: Gamma transfer curve for a typical CRT display

Most digital image formats use gamma corrected values. However, they are commonly misnamed especially in the context of logarithmic representation of film densities as linear value which is not correct. Therefore the term gamma corrected or video gamma values should be used. For additional background information the book by Glenn Kennel [1] serves as a good reference.

2.3.2 File formats

Kodak developed the Cineon file format in the early 1990s for the first digital intermediate system with the same name. This format made use of a logarithmic density representation and also established the printing density concept. As a successor to the Cineon format the compatible DPX format was standardized by SMPTE and is nowadays most commonly used to store scanned film images. This format allows storing logarithmic as well as video gamma values at 10, 12 or 16 bits per component and pixel and also provides means to store essential metadata. Unfortunately, this facility is in many cases not used which leads to unnecessary loss of important information such as

color space, scanner settings etc. The Cineon as well as the DPX format are single image file formats. One file is stored for each scanned film image. The single image files are usually organized in folder structures for each reel or film. The DPX format provides space for additional metadata like frame number, time code, name of the film, a catalogue number and more. These possibilities should be used to keep track of the single images and can aid in organization.
The TIFF format that was developed by the Aldus Corporation (which was later bought by Adobe) is also used in several cases, especially for storing more than 10 bits per component because many software solutions only support 10 bit DPX files. Normally, TIFF files contain gamma corrected data. However, some scanners can store logarithmic values in TIFF files, which often leads to confusion if the necessary metadata is not provided.

2.3.3 Spatial resolution and bit depth

The spatial resolution that is chosen for scanning films is the single factor that has the strongest influence on the amount of uncompressed data that has to be handled, transferred and processed. Therefore, it also has a major influence on the economy of the project. While it is very hard to estimate the actual spatial resolution of a piece of archive film and many different opinions of experts can be found regarding this parameter, it is a fact that information that has been lost due to choosing a too low resolution cannot be regained during any subsequent processing step. The only solution would be to scan the film again. A special problem is imposed by the grain structure of the film. While grain usually has a negative influence on achievable compression ratios it is also a unique element of the original film and often also seen from an artistic point of view or otherwise contributes to the viewing experience. While hard to remove because sometimes present in the same spatial frequency ranges as real image details, grain in some cases also helps to hide other defects like color banding artifacts. It therefore might be sensible to preserve it to some degree.
The practical maximum spatial resolution in the context of this digitization project is 2K. However, the exact details for this term depend on a number of technical considerations that in turn are mostly dependent on the particular scanner model used:

- the actual resolution of the imaging sensor used in a particular model of scanner and the possible output resolutions that can be different due to image processing used the scanner
- the aspect ratio of the film image that is mapped to the image sensor's aspect ratio
- the optical imaging system used to project the film image onto the image sensor and its limitations (e.g. non-existent calibration files for a certain combination of film aspect ratio and scanner resolution)

For most existing 2K scanners this means a maximum usable resolution of 2048 x 1556 samples. While the amount of data for uncompressed material can be easily calculated and is constant no matter how much detail was on the film this is not necessarily true for the compressed images. Due to the transformations from the spatial to the wavelet domain used in JPEG2000 a specific image quality level (in comparison to the uncompressed original data) can be achieved for lower quality material with a lower bit rate using the same digital spatial resolution than for higher quality original material. However, the high spatial resolution leads to much better results when the image has to be scaled and/or cropped to a target format (e.g. from full aperture 2K 2048 x 1556 to an HDTV compatible resolution of 1920 x 1080). Therefore, we suggest to use the

highest economically possible scanning resolution and adapt the compression ratio (automatically) during compression to JPEG2000 according to the resolution of the source material.
It has to be emphasized that in order to achieve maximum quality the scanned DPX images should be provided for compression in the original resolution the scanner generates without any resizing done by the scanning operator except what is performed internally in the scanner in certain operating modes (e.g. 3K → 2K resizing when using a 3K oversampling process). In addition scanners often use adjustable optics to project only a specific area of the film on the imaging sensor. Wherever possible this mechanism should be used to project the image area on the maximal size of the imaging sensor. In some cases special calibration files have to be created for specific formats.

Modern film scanners use analogue to digital converters in their cameras with a quantization bit depth of 14 bits per color or even more. The output of these converters has a linear relation to the light intensity. Because of the non-linear response of the HSV it seems not reasonable to store this data in its linear form but instead to apply a non-linear transfer curve and use this to be able to reduce the bit depth of the signal. Most experts agree that 10 bits per color component are sufficient when the logarithmic transfer curve defined in the Cineon and DPX specifications is used [1] and that the data in this case still provides enough headroom for a certain amount of color correction. Only in cases where major corrections become necessary a higher bit depth makes sense. This is normally the case when major restoration work has to be done. In this case it may make sense to store the original linear bit depth from the scanner, depending on the restoration process.

2.4 Color correction and adjustment of parameters during scanning

From a preservation point of view it may be sensible to scan film in a way that a conclusion to the color impression of the film at the time of scanning can be drawn from the digital data. This implies a color management workflow and a thorough calibration of the scanner. A color space has to be used that is known to encompass all colors that can occur in the film. While this is in principle achievable by the means described in section 2.2, the required effort is very high and most probably the assistance of the system manufacturers will be needed. When long-term preservation is carried out by appropriately storing the original film material and scanning is done mainly to facilitate access to the material a simpler process can be applied. One requirement of S&V is that the JPEG2000 and XDCAM HD422 compressed stored digital images can be viewed and used directly without necessarily having to individually color correct each film. For cost reasons S&V plans to scan most of the material with one scanner setting per film (one light scan) that ideally leads to this usable color rendition. For film in good general condition generally only minor adjustments are necessary while for heavily deteriorated and faded material stronger adjustments become necessary. In order to have a well-defined color representation the scanner should be calibrated and set up to scan to a well-defined commonly used color space and transfer curve with its neutral settings. Necessary adjustments should then be recorded and stored as metadata for each film to be able to retrace what has been changed and to which amount. It is advisable here to use only the most simple and basic means of adjustments to achieve this goal. Namely, these will be the black levels, gain settings and possibly a non-linear correction (gamma adjustment). This information can be stored in a parameter set like the one defined by the ASC Color Decision List [7] that can be understood by many professional systems. In this format the primary color correction is analytically described by

$$out = (i \cdot s + o)^p,$$

where i is the input pixel code value, s is the slope of the curve (gain), o is the offset (black level), p is gamma factor of the curve and out is the color graded pixel code value. Support for ASC CDL is included in many commercially and non-commercially available system and software solutions.

2.5 Conclusion and practical considerations for scanning

In theory, the most elegant and accurate approach would be to scan to a device independent color space like XYZ using a scanner that is calibrated as a color or chromaticity measuring device. All films would then be scanned using the same scanner setting. Necessary adjustments and color correction would only be stored as metadata with the image and the processing steps necessary to apply these corrections would be applied when converting to a distribution format like XDCAM HD422. However, this high-level approach is very complex, implies extensive calibration procedures and results in a large amount of data to be stored because of the required quantization depth that is needed to accommodate especially color-faded films. Apart from that new and nowadays not commonly used downstream processes have to be implemented to apply the metadata-controlled corrections and adjustments when creating distribution formats.
Within this project a reasonable combination of the described theoretical approach and industry-standard scanning processes where corrections are carried out directly for a specific distribution format may make sense. With a focus on practical implementation at the Netherlands Institute of Sound and Vision with regard to the envisaged use cases this could result in the following process outline:

- 2K full aperture (2048 x 1556) as maximum resolution,
- 10 bits per pixel logarithmic Cineon transfer curve,
- chromaticity coordinates of the color primaries for storage in the DPX and JPEG2000 files according to ITU-R Rec. BT.709 (scanner calibrated to this target),
- adjustment of black level and gain in the scanner for each primary color to make best use of the available bit depth,
- storing of these settings as metadata in the DPX and MXF files (to be able to retrace the settings) and
- storing additionally necessary color corrections as metadata only using the method described by the ASC CDL (the correction is applied when converting from the DPX or JPEG2000 files to a distribution format).

In this way the XDCAM HD422 files can be created with a minimal color transformation effort while retaining the maximum information from the original film images. For most practical purposes the gamut of the ITU-R Rec. BT.709 primaries is sufficient, especially when the main distribution route is HDTV.
For film material that would potentially benefit from capturing its full color gamut also other color spaces like XYZ, wide gamut RGB or the P3 RGB color space may be used. However, for a wider gamuts a higher bit depth is necessary to achieve the same color precision, e.g. 12 bits per component in XYZ versus 10 bits in Rec. 709 RGB. [1]

It may be advisable to separate the material for scanning in different categories, like b/w, color, negative, reversal, print film, deteriorated material, material in good condition etc. and use a common set of general settings for each category. The calibration can be optimized when detailed information on the properties of the film material exists as it is often the case with modern materials. For older material where this information does not exist common reasonable settings have to be found and optimized for different categories of material. To develop these settings has to be part of the implementation phase of the digitization project.
The details of the process and settings described above heavily depend on the actual scanner make and model that is used.

3 Image Compression

Archiving high-resolution digital images as described in the previous section requires enormous amount of hard-disk space. Since this is economically not feasible, image compression is typically applied in order to reduce the storage costs.

Technically, two different compression types are distinguished, namely mathematical lossless compression and lossy compression. In the first case, storage space reduction is achieved by elimination of redundant information such that the overall file size is reduced without modifying the image content. Unfortunately, the achievable compression ratios are limited. In particular for images containing huge amounts of grain, they often do not exceed a ratio of 2:1 (See also the results obtained in Section 5.3). Consequently, for further file size reduction, compression algorithms have been developed that eliminate irrelevant information. In order words, they try to only eliminate those image elements that are least relevant for a possible human spectator. Nevertheless, the higher the compression ratios, the higher also the risk that artifacts become visible.[3][4]

In order to apply such lossy and lossless compression to images, different algorithms are available such as H.264, JPEG or JPEG2000. The latter is particularly wide spread in professional applications since it offers several important features such as

- Support of both lossy and lossless compression in one single system
- Inherent scalability, meaning that from a compressed file a small-sizes preview image can be extracted without decoding the overall content. The same holds when an image with full resolution, but reduced quality is sufficient for simple projection or viewing
- Flexible organization of the JPEG2000 codestream in order to fit particular requirements in terms of low-latency decoding, sub-image extraction and simplicity of the resulting file structure
- Precise rate allocation guaranteeing optimized image quality without exceeding an upper file size bound
- Intra-frame only compression enabling the flexible access to individual images
- Support for large bit depths
- Royalty-free algorithm
- Both open-source encoder and decoders are available
- High achievable image quality through the use of the wavelet transform that does not used small-sized image blocks like DCT-based codecs and therefore produces less annoying artifacts.

These features are enabled by the use of a wavelet transform, followed by an arithmetic coding of each subband. These two fundamental blocks are supported by a color transform, adapted quantization, block building, rate control and bit-stream generation.

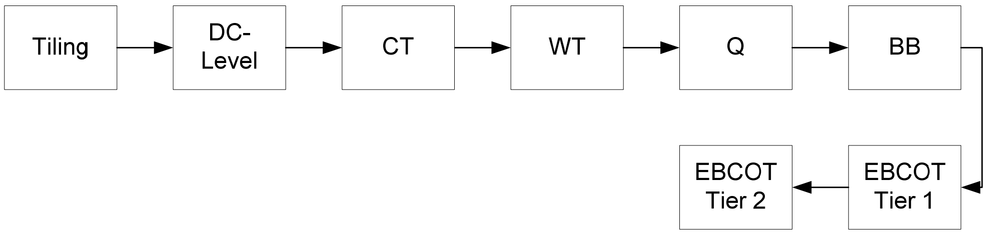


Figure 7: JPEG2000 processing chain

Figure 7 depicts the corresponding JPEG2000 processing chain. In a first step, the input image can be divided into different sub-regions, so called tiles, which are encoded separately. This separation simplifies implementation, is in general however avoided in digital cinema applications if possible.

Next the encoder transforms the input sample data to a nominal dynamic range so that it is approximately centered about zero. In the following, a color transform (CT) can be optionally applied in order to decorrelate the components of color images. JPEG2000 supports both lossless as well as lossy algorithms, whereas the latter one results in better decorrelation.

The discrete wavelet transform (WT) decomposes the input tiles into different frequency subbands in order to eliminate spatial redundancy, improve coding efficiency and allow for resolution scalable bit-streams. JPEG2000 basically offers the choice between an integer transform kernel (5x3) and a floating point one (9x7). The first enables lossless compression, whereas the second delivers better results for lossy operation.

Quantization (Q) is introduced to reduce the number of bit-planes which have to be processed by the following entropy coding units. The block builder (BB) then groups the wavelet coefficients of each subband into code-blocks which can be independently processed by the following units. Common block sizes are 64x64 and 32x32 pixels.

Entropy coding is divided into two parts, the *Embedded Block Coding with Optimal Truncation* (EBCOT) Tier-1 and Tier-2. The first one removes redundancy by arithmetic coding based on dynamic contexts. This process is performed on consecutive bit-planes instead of complete wavelet coefficients and allows thus for quality scalability.

EBCOT Tier-2 arranges the bit streams of the different code blocks into packets and determines the truncation points where they have to be cut in order to obtain best image quality under a given bit-rate constraint. Whereas this leads to better results than simple quantization, it is more calculation intensive.

3.1 JPEG2000 profiles

As mentioned previously, the JPEG2000 compression algorithm offers a huge amount of parameter choices influencing the achievable compression efficiency, required computational effort and exploitable scalability. On the other hand, however, this significantly complicates implementation of a general encoder and decoder risking to negatively impact the compatibility between equipment of different manufacturers.

In order to avoid these difficulties, different JPEG2000 profiles have been proposed. Each of them target a particular application domain. Consequently, it is possible to restrict the parameter choices by imposing predefined values adapted for the application in mind. Note that they only target the JPEG2000 compression without affecting the used color space, audio encoding or MXF wrapping.

Currently, the following profiles interesting for archival applications are available, either in form of a final standard, or in form of a mature draft document:

- DCI 2K and DCI 4K Digital Cinema Profiles.
  These profiles are mainly used for digital playout of cinema movies. Their major purpose is to provide good image quality while simplifying the interoperability is much as possible.
- Scalable 2K/4K Digital Cinema Profile.
  These profiles are an extension of the previous profile in that a DCI compliant file can be extracted with a negligible computational effort. However, in contrast to the DCI profiles, they enable a second quality layer, enhancing thus the scalability and providing for higher bit rates and image qualities if desired.
- Long-term Storage Profile for Cinematic Content
  This profile is very general offering thus a rather huge amount of usage scenarios. In particular it can be used for both lossy and lossless compression, which is not possible with the previously mentioned profiles. Furthermore, more than three components can be stored. However, generation of a DCI compliant file which can be played-back by a standard cinema projector is typically computational intensive.
- Broadcast profiles
  This profiles also target maximum interoperability, offer however a larger choice of admissible bitrates compared to the DCI profiles. Furthermore, they are more adapted to broadcast applications enabling for instance the compression of color-subsampled material. Additionally, also lossless compression is permitted. However, it has to be mentioned that the current standardization document has not reached its final version and changes are still to expect. Also, at this time no commercially available implementations are known to exist.

3.2 Selection of a JPEG2000 profile

In order to select the best suitable profile for archival purposes, the following considerations have to be taken into account:

- Most JPEG2000 encoding and decoding equipment manufactured today that is suitable to accommodate the requirements of the digitization project has been originally developed for use in the digital cinema environment.
- Currently, most real-time playback equipment only accepts DCI compliant JPEG2000 codestreams. In other words, if simple playback of the compressed material is desired, a profile similar to the digital cinema JPEG2000 profiles is a desirable choice.

- Images intended for archival in the Netherlands Institute of Sound and Vision do not fit the typical container sizes for 2K (max. 2048 x 1080) imposed by the digital cinema standards. Instead, their size can attain up to 2048x1556 pixels.

Given these constraints, several possibilities exist as described in the following subsections.

3.2.1 Usage of DCI-4K profile

Since the DCI 2K profile is limited to a container size of 2048x1080 pixels, its use for the intended archival application is not an option due to the occurring image dimensions. Instead a DCI-4K container has to be employed. By this means, the compressed files can be played back directly with a 2K or 4K digital cinema projector. Both options however may require that the image is either padded with black borders or scaled such that either the width of the resulting image equals 4096 pixels or the height amounts 2160 pixels to be compatible with existing encoding and equipment. Since padding adds no additional information to the image, it has no impact on the expected image quality for a given bit rate. However, padding has the disadvantage that the 2K-projection leads to an image whose dimensions are smaller than the nominal 2K resolution. Since the material from S&V is mostly no intended for direct presentation in standard digital cinemas this issue is of less importance. Scaling, on the other hand, is a non-reversible process that possibly impacts both the compression efficiency and the achievable image quality.
The DCI 4K JPEG2000 profile can be used independently from the rest of the DCI specification of SMPTE D-Cinema standards. This means that also other color spaces than XYZ and other wrappings than MXF OP Atom can be used. Note, however, that this requires a playback device being able to handle the alternative MXF wrapping and user defined color transforms. In particular the latter point is typically easily feasible.

The key points of the DCI-4K profiles are:

- Resolution up to 4096x2160
- Data rate up to 250 Mbit/s (1MBit = 1,000,000 Bits) for color films and 200 Mbit/s for black and white films
- Currently, DCI-4K only supports 24fps. An extension for further frame rates is underway, though in a very early stage. It enables any frame rate up to 30 fps.
- One quality layer, between 1 and 6 decomposition levels
- 12 bit per color component, 3 components, no subsampling
- Well defined code stream organization in order to ease interoperability

From this list, the 12 bit restriction seems to be the most critical one, since the scanned material uses 10 bits only. Fortunately, this problem can be solved rather easily by a 2-bit left-shift of the original material. Since lossless compression is typically not possible within the DCI-4K profile, the left shift is expected to have no significant impact on the achievable image quality. This is particularly true, since a scaling of the quantization factors should lead to a quasi identical compressed file.

Benefits:

- Direct playback with both a DCI-2K and DCI-4K D-Cinema systems

- Widely available software and equipment for digital cinema can be used (possibly with small adaptations) for encoding, decoding, playback and transcoding to other formats.

Disadvantages:

- Requires scaling or padding
- In case of padding, image does not cover the complete image area
- Scaling induces an additional computation effort and can cause a quality loss.
- Lossless compression not possible

3.2.2 Scalable 4K Digital Cinema Profile

This profile is basically an extension to the DCI-4K profile. Compared to the previous option, this solution offers the benefit of allowing an additional quality layer. This can be used in two ways:

- Storage of a high bit-rate variant
- Provision of a low bit-rate preview that can be displayed in real-time with less expensive equipment.

Key points of the profile:

- Resolution up to 4096x2160 possible
- Data rate up to 500 Mbit/s
- Currently, only 24fps are supported. An extension for further frame rates is underway, though in a very early stage. It enables any frame rate up to 30 fps.
- Two quality layers, between 1 and 6 decomposition levels
- 12 bit per color component, 3 components, no subsampling
- Well defined code stream organization in order to ease interoperability
- Additional markers for improved error detection capabilities
- Cheap extraction of a DCI-4K compliant code stream with almost no computational effort

Whereas the following image quality evaluations will show that the high-bitrate variant is less interesting for the kind of images contained in the Netherlands Institute of Sound and Vision, a low bitrate proxy for cheap projection seems to be a desirable feature, in particular since the additional effort is minor.

3.2.3 Usage of the Long-term Storage Profile for Cinematic Content

This profile offers the largest flexibility, and thus enables direct compression of the provided content. In particular, both lossy and lossless compressions are available. This, however, has to be paid by the fact that real-time playback requires special solutions, being thus more expensive or even unavailable. Furthermore, it would be recommendable to select only a well-defined subset of admissible parameters in order to simplify the digital archive workflow.

3.2.4 Usage of an Extended Scalable 2K Digital Cinema Profile

As already mentioned, principally the required image extensions inhibit usage of a 2K -DCI or 2K Scalable Digital Cinema Profile. However, a possible solution would be to use a new format that can be transcoded with very little effort into a scalable 2K digital cinema profile. Such an operation would of course induce a cropping operation. Nevertheless, for a fast preview and automatic cropping extracting for instance only the center region of the image could be an acceptable approach. In this case, the JPEG2000 tiling might offer a possibility to perform this directly in the compressed image domain leading directly to a DCI-2K compliant file structure without significant computational effort. It has to be mentioned however, that selection of this option would require a feasibility study in order to be sure that important boundary conditions are not violated.

3.2.5 Definition of an own parameter profile

This variant is definitively the most flexible one. However, it risks incompatibilities with workflows in other archival organizations. Also, implementation will most probably be costly and may lead to a dependence on a single manufacturer of encoding and decoding solutions. Consequently, its selection cannot be recommended without compulsive reasons.

Fraunhofer IIS

4 Subjective quality assessment

The goal of the subjective quality assessment process is to find out what compression ratios can be achieved without the introduction of a visual perceivable quality loss with regard to the specific properties of the films that will be converted to JPEG2000. Since storage space for the compressed files is limited lossy compression has to be used. The bit rate should, according to the requirements of S&V, not exceed a maximum of 300 Mbit/s and should preferably be lower.
For the test provided a set of scanned films as test sequences in the DPX format that shall represent the different types of material that have to be handled in the digitization project in the following categories:

- 16 mm b/w reversal, bad condition (Volleyball)
- 16 mm b/w reversal, relatively good condition (Atletiek)
- 16 mm color reversal, color-faded in bad condition (Televizer)
- 16 mm color reversal, relatively good condition (Voetbal)
- 35 mm color answer print, relatively good condition (Wij Surinamers)

The subjective Quality Assessment (QA) method used is a modification of the Double stimulus Perceived Difference Scale method (DSPDS) already used in June 2001, Los Angeles MPEG Digital Cinema tests.

Goal of the QA:
- "Visually Lossless" compression with JPEG 2000 in four different cases
    - unprocessed color sequences
    - unprocessed gray scale (b/w) sequences
    - sequences with color correction after the compression
    - sequences with a subsequent compression to XDCAM HD422

4.1 Technical arrangement and viewing conditions

The QA is accomplished in the Postproduction-Laboratory of the Fraunhofer IIS, a review room with dark ambient luminance for the projection.
Properties:
Harkness Matt Preview Screen:
- 3.80m width x 1.80m height
- gain of 0.9 for a view angle to 10

Barco Projector DCP 30:
- projector distance to screen of 7.4m

Preferred Viewing Distance (PVD):
- 4.10m from screen

The server used for the playback of the uncompressed sequences is the DVS HD-Station-RGB. Properties:

- resolution of 1920x1080
- bit depth of 10 per color channel
- no color subsampling (4:4:4)
- direct playback of DPX or TIFF files

4.2 Arrangement for the DSPDS method

All sequences consist of the original uncompressed reference in comparison to the compressed sequence arranged in a split-screen composition (butterfly). The reference is positioned randomly at left (A) or right side (B) of the image in Figure 8. The different sequences are presented randomly in succession for each test person, to prevent dependence by the order of the sequences.

Figure 8: Example of one sequence out of a scene with the Butterfly composition

For each scene one sequence consists of the reference in both sides in order to control the reliability of the assessment of each test person (hidden reference). Furthermore for each scene there is one composition with a very high compression ratio, the low anchor, which exhibits easily visible differences in comparison with the reference. By that the test person is given an example for the occurring image quality differences which stabilizes the results. The lower anchor should be recognized by each test person as such.

For the assessment of the sequences the test person uses a Graphical User Interface (GUI) made in Python V5.2 as it is shown in Figure 9.

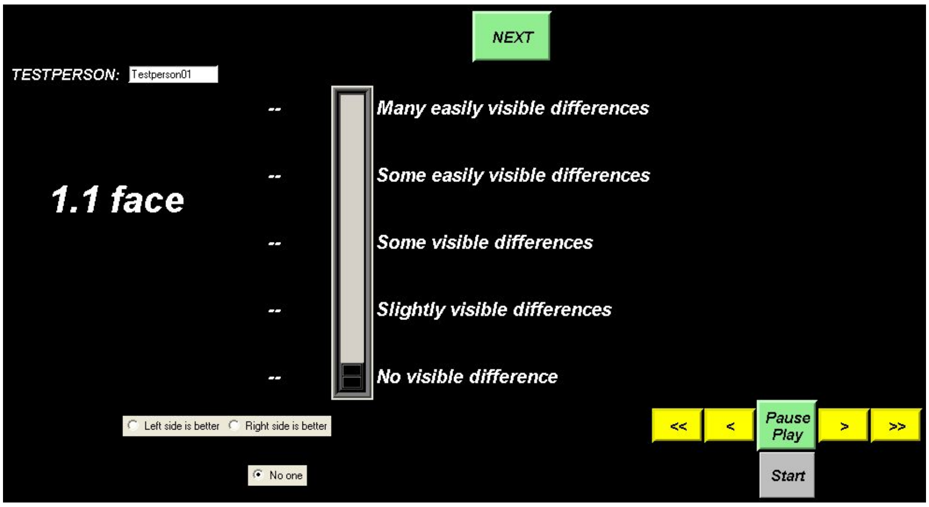


Figure 9: Graphical User Interface of the DSPDS method

Because of the relatively low image quality of the original material, the test person has the possibility to pause the sequence and review it, until a sure decision can be made for the validation.
The test person has to decide if there is a visible difference between side A and B using the 5 stepped adjectival perceived difference scale in Figure 9. In the case of a visible difference noticed, the test person has to decide which side is preferred in terms of image quality according to his subjective opinion. This decision is marked by three possible radio buttons in Figure 9 (left side, right side or no one).

4.3 Test group

Each test person is screened for normal or corrected visual acuity (Landolt test) and normal color vision (Ishihara Test). Before the actual test begins the test person is informed about important issues of the QA and about the proceeding. Next a pre-test session is executed with sequences which are not presented in the relevant actual quality assessment test arrangement anymore. It is recommended to have at least 15 people participating for stable results. In the actual test 22 experts participated and 19 results were used in the analysis (see section 4.5).

In this test only experts are appointed for the assessment. This decision has been made due to the difficult image quality of the material. The term "expert" is considered to be a person that either knows the material used and knows what artifacts to look for, or has professional experience in the domain of image quality. The test lasts about 45 min to 1 hour for each person. The time has to be limited to this amount so that the results are not influenced by loss of concentration and fatigue of the test persons. Experts might bring more motivation and interest to the validation and therefore might show less tiredness.
Since one individual person participates per test session and due to the occurring interaction with the GUI, the "group effect" is avoided and motivation is increased. The "group effect" describes the effect that people can be distracted by each other and the results might be influenced by other group members.

Fraunhofer IIS

4.4 Processing of the image compression

There are 8 different scenes with 7 different sequences with a maximum duration of 10 seconds which differ in the compression ratio and other processing steps. The scenes used for the QA are described in Table 1. The Volleyball scene was not used in the test because it is from a JPEG2000 compression point of view very similar to the Discus scene. Due to constraints in the time each test session needs this scene was omitted.

CINECO_ARRISCAN					
Original scene	*Scene name*	*Frames*	*Properties*	*Processing*	*Projection*
Wij_Surinamers	01.priest 	1781-1855 2448-2549	color, logarithmic, print 35 mm	JPEG2000 compression	LUT
Wij_Surinamers	02. priestXD 	1781-1855 2448-2549	color, logarithmic, print 35 mm	JPEG2000, concluding XDCAM HD422 compression	LUT
Wij_Surinamers	03. house 	2552-2762	color, logarithmic, print 35 mm	JPEG2000 compression	LUT
K66358_Voetbal	04.soccer 	401-641	color, logarithmic, reversal 16 mm	JPEG2000 compression	LUT
2263M_Atletiek	05. discus 	1256-1506	grey value, logarithmic, reversal 16 mm	JPEG2000 compression	LUT
M66143_Televizer	06. face 	1600-1791	color, video gamma, reversal 16 mm	JPEG2000 compression, subsequent color correction	direct
M66143_Televizer	07. faceXD 	1600-1791	color, video gamma, reversal 16 mm	JPEG2000 compression, subsequent color correction, XDCAM HD422 compression	direct
M66143_Televizer	08. politics 	2807-2998	color, video gamma, reversal 16 mm	JPEG2000 compression, subsequent color correction	direct

Table 1: Scenes used for the QA

The different compression ratios are shown in Table 1 to Table 6.
Due to the technical limitation of the server, the images are cropped in HD resolution 1920x1080 and compressed with a maximum of a color depth of10 bits. The color space is RGB according to ITU-R Rec. BT.709 without color subsampling (4:4:4).

The compression is executed by Eyeon Digital Fusion V5.2 with the Fraunhofer IIS JPEG2000 plug-in based on Kakadu V5.2. Then for each compression ratio the butterfly composition is created by combination with the reference. Except for the XDCAM HD422 compression all sequences are decompressed to DPX-format for the playback by the server. The XDCAM HD422 compression is realized by Apple Compressor V3. In this case the relevant JPEG2000 compressed sequences, "priest" and "face", are subsequently compressed to the XDCAM HD422-format. These sequences are then decompressed to TIFF-format for the playback. Consequently the resolving scenes "priestXD" and "faceXD" consist on one side of the reference, which is only compressed with XDCAM HD422, and at the other side the scenes are compressed with JPEG2000 and subsequently with XDCAM HD422.

A special process is used for the color-faded example. In order to assess the influence of compression on subsequent color correction steps the original faded image is compressed to JPEG2000 as it is. Then, a color correction step with exactly the same settings is applied to the original file and the decompressed JPEG2000 file. Both are then compared to each other with the color corrected original image as the reference. The color correction parameters are set in a way to mimic the color correction that was applied by Cineco after scanning and delivered to Fraunhofer separately from the color-faded images.
Because the color-faded images were stored by Cineco in 16 bit TIFF files with a video gamma transfer curve, following the workflow in Figure 10 the sequences "face", and "politics" are first transformed from video gamma to logarithmic and saved to the DPX format with 10 bits per component in order to ensure comparability to the results of the logarithmic transfer curve of all other test sequences.

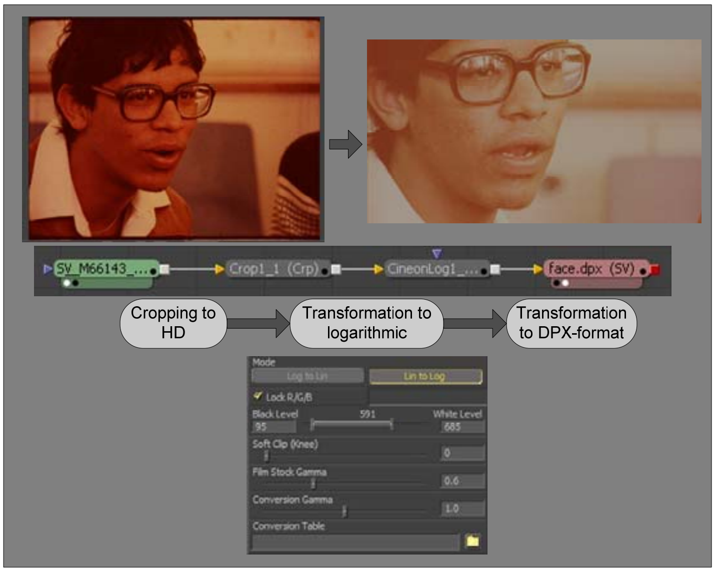


Figure 10: Digital Fusion workflow for the color corrected scene "face"

Subsequently the JPEG2000 compression follows as shown in Figure 11. The sequences are then decompressed to DPX format, color corrected and concluding transformed with video gamma by the same Cineon Log plug-in in Figure 10.

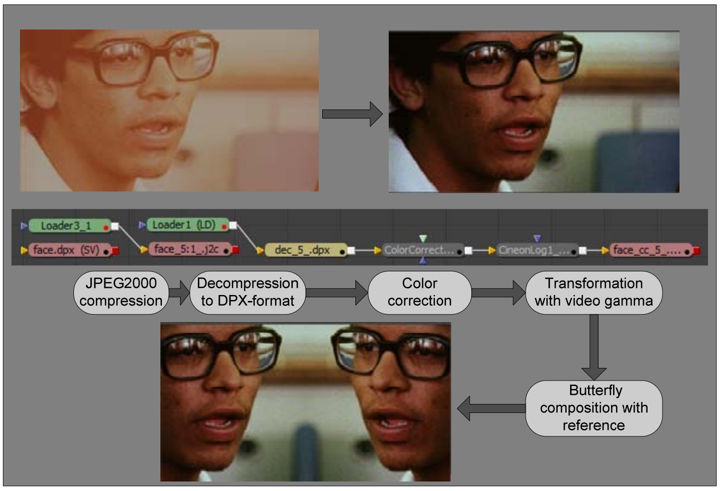


Figure 11: Digital Fusion workflow for the color corrected scene "face"

All other test scenes are directly compressed to JPEG2000 and subsequently decompressed to build the butterfly view and playback on the server.

For the presentation in the Postproduction Laboratory, the logarithmic scenes were projected with using a lookup table (LUT). This is necessary, because the playback server and projector are calibrated to the ITU-R Rec. BT.709 color space in RGB. The lookup table has been calculated using the equation given in [2], but employing a gamma correction of 2.4 according to the ITU standard. Figure 12 shows the resulting transfer curve with code values ranging from 0 to 1023 on the input and output. According to the Cineon and DPX specifications the black level is set to be at input code value 95 and the white point at input code value 685. The range between these values has to be mapped to the complete output rang from code values 0 to 1023. This approach implies that the same setup was used during scanning. This was verified through the images' histograms.

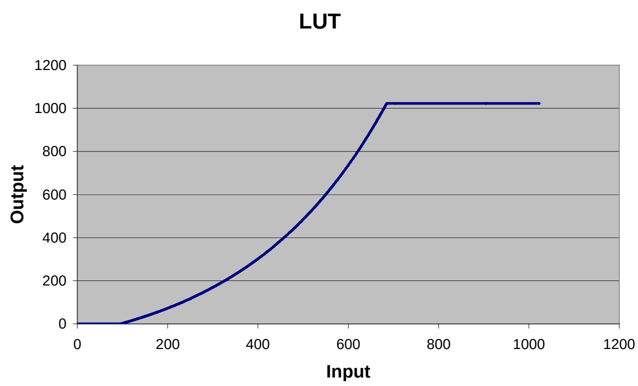


Figure 12: Look up Table for translating the logarithmic DPX files into gamma corrected values for projection.

JPEG2000 settings	
Bit rates	See Tables 3 to 5
Color depth	10 bits per component
Number of quality layers	1
Number of resolution levels	5
Progression order	CPRL
Use two guard bits	
YCC color transformation	On
Visual weights [3][5]	On
Wavelet kernel	9x7(not reversible)
Code block size	32x32
High QStep(2^20)	

Table 2: JPEG2000 settings in Digital Fusion

Colored scenes (priest, priestXD, house, soccer)			
Compression ratio	*Resolution 1920x1080*		*Extrapolated bit rate for a resolution of 2048x1556 in Mbit/s*
	Bit per pixel	*Bit rate for 25 fps in Mbit/s*	
1:1 (uncompr.)	30	1555.2	2390
15:1	2.0	103.7	159.3
20:1	1.5	77.8	119.5
25:1	1.2	62.2	95.6
30:1	1.0	51.8	79.7
40:1	0.75	38.9	59.8
150:1	0.2	10.4	15.9

Table 3: Compression ratio colored scenes

Color corrected scenes (face, faceXD, politics)			
Compression ratio	*Resolution 1920x1080*		*Extrapolated bit rate for a resolution of 2048x1556 in Mbit/s*
	Bit per pixel	*Bit rate for 25 fps in Mbit/s*	
1:1 (uncompr.)	30	1555.2	2390
5:1	6.0	311	478
10:1	3.0	155.5	239
15:1	2.0	103.7	159.3
20:1	1.5	77.8	119.5
25:1	1.2	62.2	95.6
30:1	1.0	51.8	79.7

Table 4: Compression ratio color corrected scenes

Grey value scene (discus)			
Compression ratio	*Resolution 1920x1080*		*Extrapolated bit rate for a resolution of 2048x1556 in Mbit/s*
	Bit per pixel	*Bit rate for 25 fps in Mbit/s*	
1:1 (uncompr.)	10	518.4	796.7
15:1	0.67	34.6	53.1
20:1	0.5	25.9	39.8
25:1	0.4	20.7	31.9
30:1	0.3	17.2	25.6
40:1	0.25	12.9	19.9
150:1	0.07	3.5	5.3

Table 5: Gray value scene

XDCAM HD422 scenes (priestXD, faceXD)		
Compression ratio	*Resolution 1920x1080*	
	Bit per pixel	*Bit rate for 25 fps in Mbit/s*
As before	0.96	50

Table 6: Compression ratio XDCAM HD422 scenes

4.5 Evaluation of the results

The statistical analysis has been made with the formulas of the Recommendation ITU-R BT.500-11. Of 22 test subjects 19 are included in the evaluation, caused by irregular results of 3 test persons for the hidden reference sequences. The test persons are mostly employees of the Fraunhofer IIS, department of image sensors with an average age of 33. Three of the test persons were sent by S&V. The test persons voted in favor to the reference in the question for the preferred image quality compared to the respective compression ratios, in the case of a visible difference noticed.

Figure 13 and Figure 14 indicate the Mean Opinion Score (MOS) over all sequences described in the legend, which is the mean value of the results of all 19 test persons. The MOS is given in the Y-axis divided into 5 adjectival sections. The X-axis shows the different compression ratios of the scenes as listed before in Table 3 to Table 5.

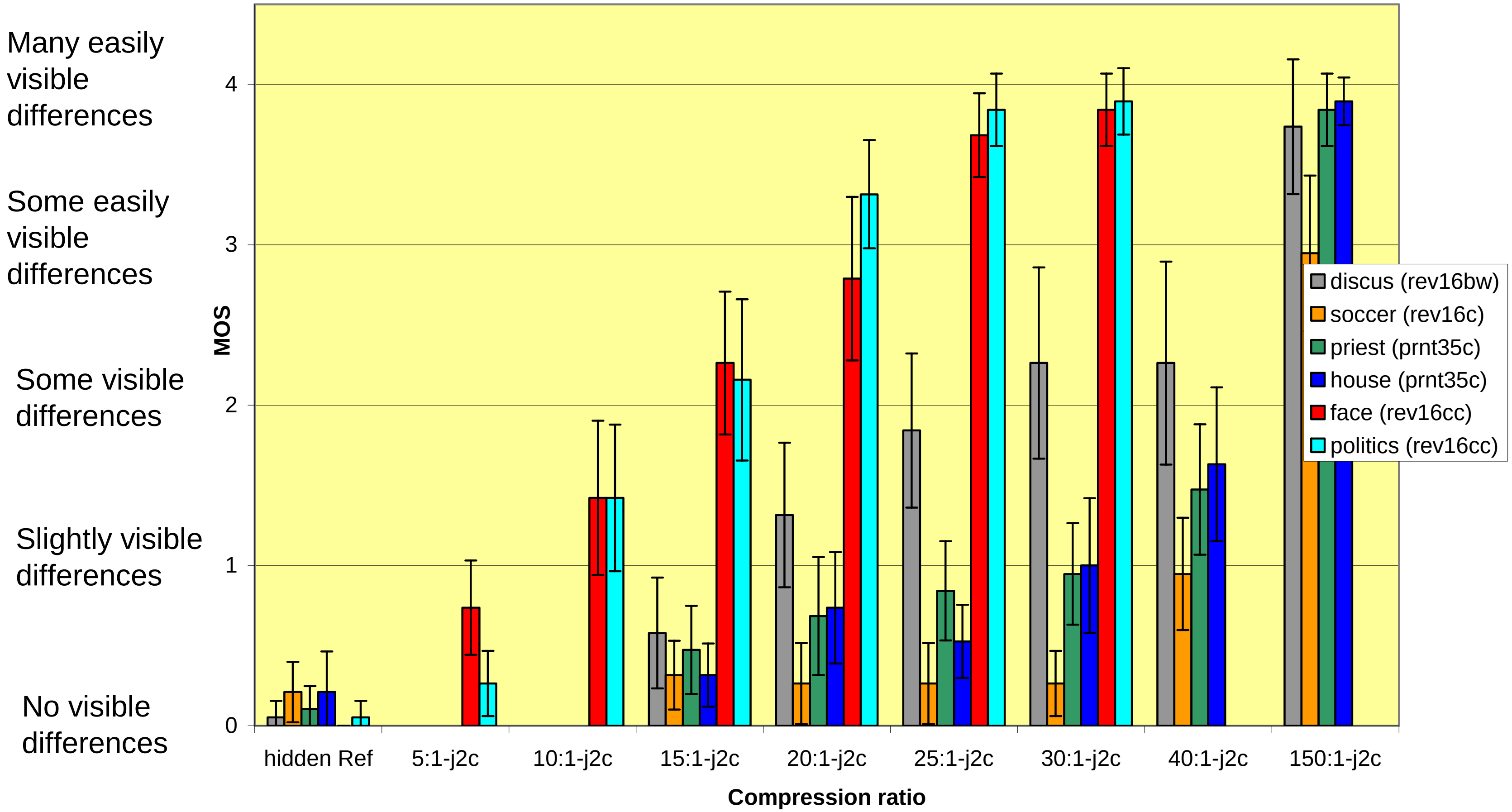


Figure 13: Overall Mean Opinion Score for JPEG2000 compressed sequences

Since however the measured mean value is only an estimation for the "true" Mean Option Score, its validity for every compression ratio is given by the confidence interval. It defines a range, within the "true" MOS value lies with 95% of reliability. That means that if the confidence interval lies beyond the "slightly visible difference" line, we can be sure that with a reliability of 95% there is no significant difference between the compressed material and its reference. On the other hand, if the confidence interval crosses the "Slightly visible differences" mark or lies above, this statement cannot be given anymore. Hence, we say that there is a difference noticeable in comparison to the reference.

The width of the confidence interval has been calculated as standardized in ITU-R BT.500-11 using the following equation[1]:

$$2 \cdot \delta_{jkr} = 1.96 \frac{s_{jkr}}{\sqrt{N}}$$

where s_{jkr} is the standard deviation, N the number of involved test persons. As mentioned previously, its value has been set to 19 for all experiments. In other words, the width of the confidence interval is proportional to the standard deviation, indicating thus not only the region where the "true" values is situated with a probability of 95%, but also gives a hint on the evaluation hardness. The larger the standard deviation, the more differed the answers of the test persons, the less evident the situation has been. Furthermore, it becomes obvious that the width of the confidence interval can be reduced by increasing the number of test candidates if required. However, this would require further effort and thus increased costs. Additionally, it has to be warned that the confidence interval calculation is only an approximation assuming that the answers of the test persons follow a standard Gaussian distribution. Evidently, this is not exact for the performed measurement. Consequently, squeezing the confidence interval will cause that this approximation will gain importance.

With this knowledge in mind, Figure 13 demonstrates that material intended for color correction after the compression step requires a relatively low compression ratio. Already a compression of 10:1 leads to a slightly visible differences, whereas a 5:1 compression seems to induce almost no losses with a reliability of 95%.

On the other hand, material not intended for color correction admits higher compression ratios. The 16mm black and white reversal material, for instance, can be compressed with a ratio of 15:1. The 35mm print film material even can be processed with a 20:1 compression, although 15:1 is definitively a safer choice, since for a 20:1 file size reduction, slightly visible differences cannot be excluded with a reliability of 95%.

The soccer material finally even permits 30:1 compression. However, it has to be emphasized that the material has been verify difficult to evaluate due to difficult content with regards to image compression and quality assessment in form of autumn foliage.

[1] Although statistically not completely correct, we have applied the same formula in order to obtain results that can be compared to other studies.

In addition to these values Figure 13 also shows the measurements for the hidden reference. In other words, it depicts the scenario where the spectators have been shown two times the uncompressed image. As can be seen, not all judges have been acceptable to detect this situation correctly. In particular the fact that the soccer scenes obtained similar votes for the hidden reference and the moderately compressed variants confirms the observation that this material has been very difficult to evaluate.

4.5.1 Evaluation of the XDCAM HD422 results

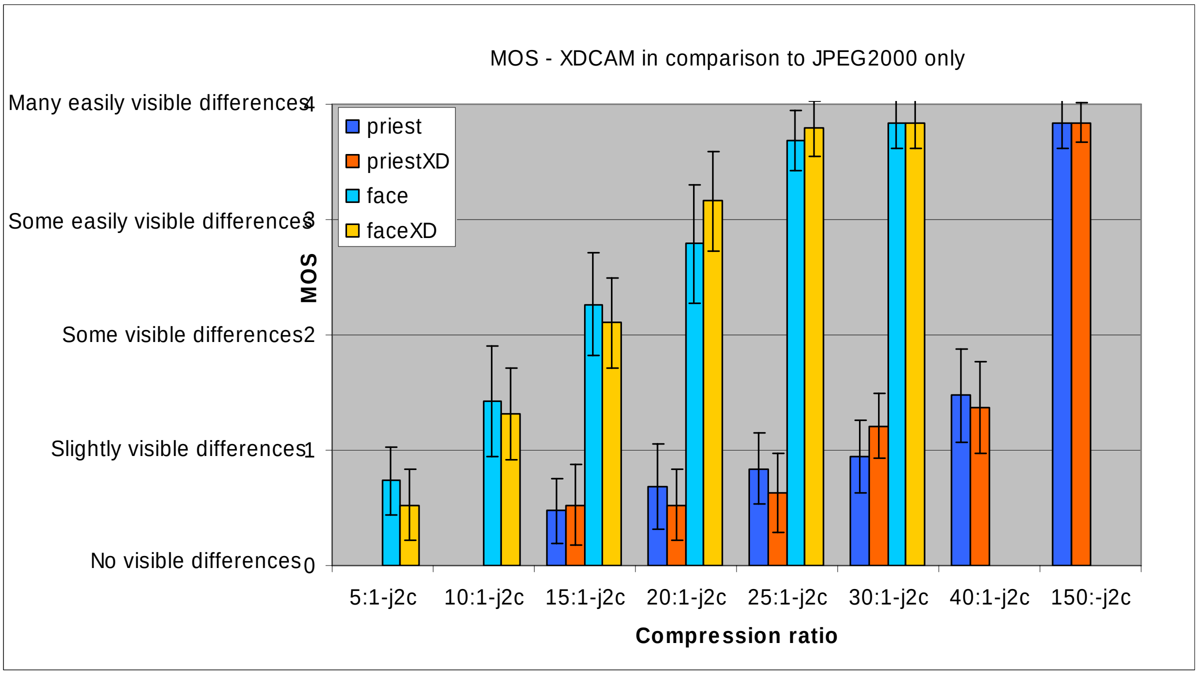


Figure 14: Comparison of "priest", „face" with "priestXD", "faceXD" with XDCAM HD422 compression

Figure 14 shows the MOS value of the scenes "priest" and "face" (blue and cyan bars) in comparison to the same scenes with an additional compression to XDCAM HD422-format (orange and yellow bars). These results shows if there is a degradation of the image quality of the JPEG2000 compressed image after an additional XDCAM HD422 compression in comparison to the XDCAM HD422 compressed reference. The confidence interval of the blue/cyan and orange/yellow bars overlap each other in all compression ratios. Consequently there is no systematically difference with a reliability of 95%. Considering the "Slightly visible difference" at an MOS value of 1, the XDCAM HD422 makes it a little more difficult to notice differences. This may be caused by the compression of the reference, which makes it more difficult to see a difference.

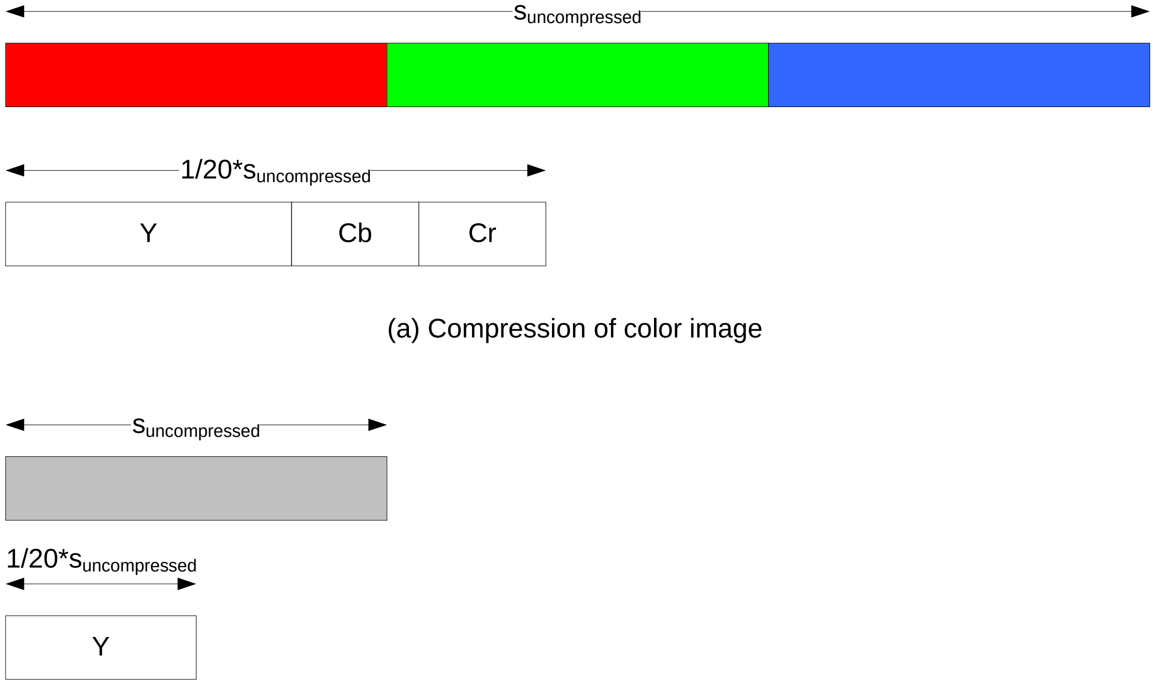
4.6 Conclusion for the subjective quality assessment

The original material has a relatively low quality, due to the typical characteristics of aged film, such as a high noise level, blurring and impurities. The compression with JPEG2000 causes, amongst others, an additional blurring, which for this reason is difficult to perceive besides the characteristics mentioned before. Therefore the participation of experts and the possibility of reviewing and pausing the sequences have been applied to the QA, which has to be pointed out for the interpretation of the results. Consequently the results show the most critical situation for visible differences to notice. In particular, it has to be emphasized that the results have been obtained specifically for the material provided and used during the visual quality assessment. In other words, the transfer to other film material is only possible when it shows similar characteristics like the ones used for this quality session.

Hence, the following recommendations should be kept to prevent a perceptible distortion caused by the image compression:

- For color images in good condition without the necessity to apply additional severe processing steps after decompression a maximum compression ratio of 20:1 should not be exceeded (Figure 13, orange, black and blue bars).
- The recommended compression ratio for images in need of a strong color correction after decompression (Figure 13, red and cyan bars) is a maximum of 5:1. This is, because the color correction seems to enhance the distortion caused by the JPEG2000 compression. Consequently, the color correction makes imperceptible compression artifacts visible.
- Grey scaled images need special consideration, since they cannot profit from any color decorrelation normally performed by the JPEG2000 algorithm. This is schematically depicted in Figure 15. For colored images, the color decorrelation causes that the luminance channel (Y) needs more storage space than 1/20 of one uncompressed color component (Figure 15a). For grey-scale images, on the other hand, only the luminance channel exists causing that a compression ratio of 20:1 would result in a significantly smaller size of the Y-channel, leading to clearly visible artifacts (Figure 15b). Hence, a maximum compression ratio of 15:1 should not be exceeded (Figure 13, grey bars).

Detailed diagrams for all sequences can be found in the annex of this report.



(a) Compression of color image

(b) Compression of grey-scale image

Figure 15: Schematic illustration of compression ratios for color and grey-scaled images. The symbol $s_{uncompressed}$ stands for the file size of the uncompressed image.

5 Objective quality assessment

Since subjective quality assessment is very time consuming and expensive, often objective quality measures are employed in order to quantify the distortions caused by an image processing algorithm or system. Unfortunately, their expressiveness and precision is limited since most of them do not properly take the properties of the human visual system into account. Nevertheless, they present a quick and cheap method in order to get an idea about the approximate image quality resulting after the compression. To this end, two different object quality assessment methods shall be discussed in the following, namely the Peak Signal-to-Noise Ratio (PSNR), and the Modulation Transfer Function (MTF).

5.1 The Peak Signal-to-Noise Ratio (PSNR)

Given a compressed (or otherwise impaired) image and the corresponding original, the peak signal-to-noise ratio (PSNR) delivers an objective quality value by measuring the mathematical distortion. In order to allow comparisons between different images, this mathematical distortion is put into relation with the maximum possible pixel value:

$$PSNR = 10 \lg \frac{(dc\_value)^2}{MSE^2}$$

$$MSE^2 = \frac{\sum_{x,y} \left(I_{compr}(x,y) - I_{ref}(x,y) \right)^2}{n}$$

where dc_value is the maximum possible pixel value amounting 1023 for 10 bit images and $n = \sum_{x,y} 1$ the number of overall pixels. $I_{compr(x,y)}$ is the compressed image, $I_{ref}(x,y)$ the uncompressed reference. Both of them are not gamma corrected in order to correspond to what is seen on the screen by a human spectator. In case of color images, both magnitudes are set to one of the three color components. Alternatively, the luminance calculated by means of the following equation can be used:

$$Y(x,y) = 0.299 \cdot I_{red}(x,y) + 0.587 \cdot I_{green}(x,y) + 0.114 \cdot I_{blue}(x,y)$$

The resulting PSNR value is often called Y-PSNR.

The following graphs show for each frame in the X-axis the corresponding PSNR value on the Y-axis. The different curves correspond to the used compression ratios described in the legend. A high PSNR value indicates a small difference of the compressed image compared to the reference image.

Figure 16 shows the results for the Y channel by means of "discus" and the "soccer" example scenes and Figure 18 shows the results for the "house" example scene. These graphs confirm the assumption of a low PSNR value for high compression ratios.

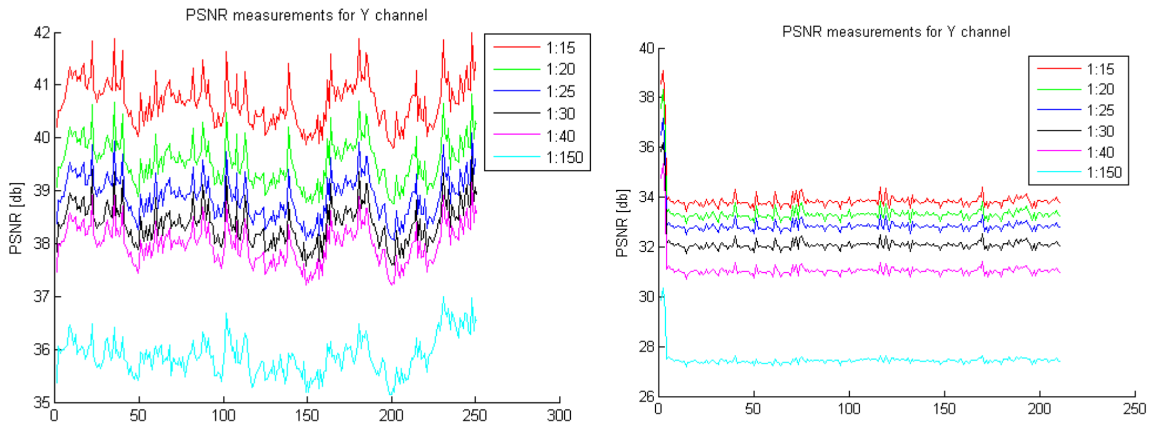


Figure 16: PSNR measurements of the "discus" scene (b/w) on the left and the "soccer" scene (color) on the right. Note that the axes are not scaled identically.

Comparing the different scenes in Figure 16, it is obvious that the PSNR values for "discus" exceed those of the "soccer" scene. In other words the mathematical differences of the compressed images against the uncompressed references are larger for the "soccer" scene than for the "discus" example. On the other, performing the same comparison using subjective quality assessment leads to the results depicted in Figure 17. Obviously the human spectators had much more difficulties in detecting the differences between the compressed and uncompressed images of the "soccer" scene compared to the "discus" example. This is probably due to the fact that the content of the soccer scene has been very noisy because of the moving leaves that appear all over the image's background, making it very difficult for the human eye to detect the occurring differences that were introduced by the compression. This mismatch between the subjective quality assessment and the PSNR value clearly demonstrates the problematic of the PSNR method, since the latter only shows a small correlation with the subjective perception of image quality.

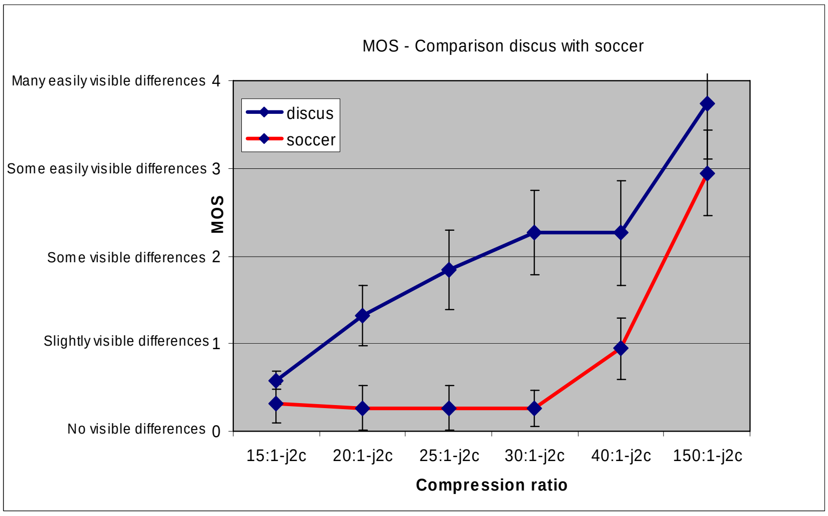


Figure 17: Comparison of the MOS values of "discus" and "soccer"

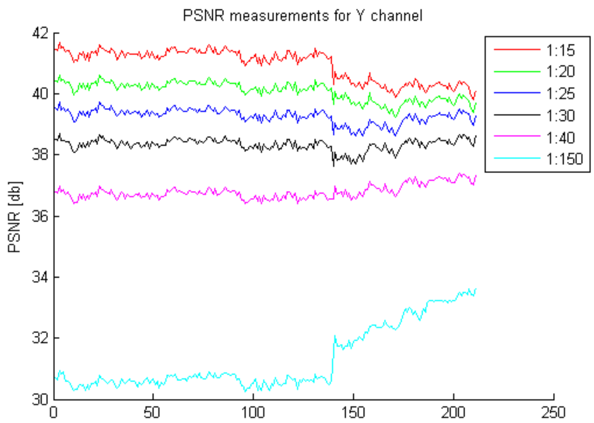


Figure 18: PSNR measurements of the "house" scene

Figure 19 to Figure 21 depict additionally the PSNR value for the red channel. This is particularly interesting for the “face” scene, since in this case also the subjective image quality has been dominated in our observations by artifacts occurring in the red color channel. This can be clearly confirmed by comparing both diagrams in Figure 19. Obviously the overall Y-PSNR is much larger than the red PSNR value.

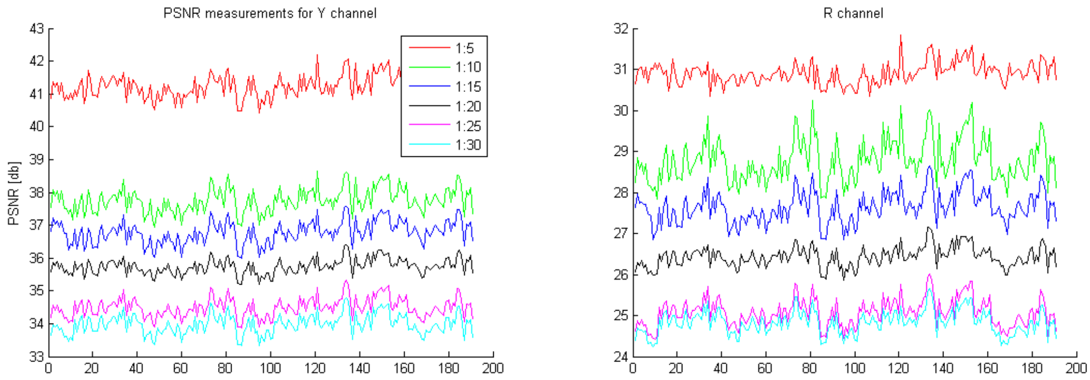


Figure 19: PSNR measurements of the "face" scene

Figure 20 exemplarily shows the PSNR values for a black and white scene. In this case the Y-PSNR value obviously equals the R-PSNR. Note that the peak around image 50 corresponds to a scene change.

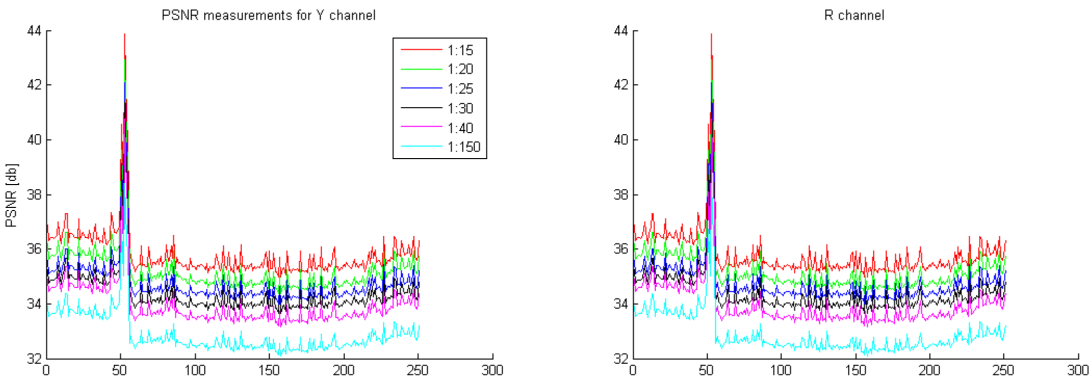


Figure 20: PSNR measurements of the "Volleyball" Sequence

Figure 21 finally shows the PSNR measurements for the soccer scene. Here the rather low PSNR values are striking. A corresponding visual inspection of the images showed that there are indeed some all-over compression artifacts leading to low PSNR values.

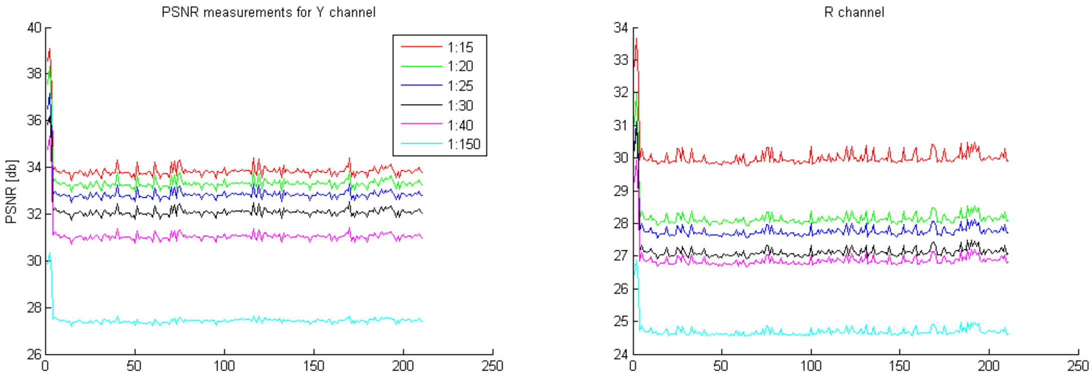
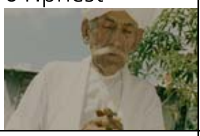
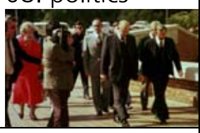
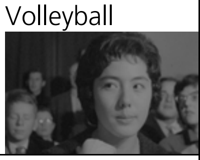


Figure 21: PSNR measurements for the soccer scene

In order to permit for better comparison between the different image scenes, Table 7 finally shows the mean PSNR values for all scenes.

Original scene	**Used scene name**	**Frames**	**Compr. ratio**	**Y-PSNR [dB]**	**R-PSNR [dB]**	**G-PSNR [dB]**	**B-PSNR [dB]**
Wij_Surinamers	01.priest 	1781-1855 2448-2549	1:15	38.81	36.52	34.54	34.87
			1:20	38.05	34.59	33.75	33.54
			1:25	37.36	33.45	33.13	32.71
			1:30	36.68	32.99	32.64	32.28
			1:40	35.46	32.42	31.69	31.51
			1:150	31.57	29.56	28.74	29.16
Wij_Surinamers	03. house 	2552-2762	1:15	40.91	36.43	36.50	37.71
			1:20	40.07	34.56	35.70	36.06
			1:25	39.26	33.19	34.92	35.09
			1:30	38.32	32.12	34.19	34.51
			1:40	36.78	31.06	33.01	33.65
			1:150	31.19	27.21	28.72	30.36
K66358_Voetbal	04.soccer 	401-641	1:15	33.87	30.01	32.16	26.41
			1:20	33.33	28.16	31.28	25.90
			1:25	32.85	27.77	31.00	24.92

Original scene	**Used scene name**	**Frames**	**Compr. ratio**	**Y-PSNR [dB]**	**R-PSNR [dB]**	**G-PSNR [dB]**	**B-PSNR [dB]**
			1:30	32.11	27.14	30.36	24.60
			1:40	31.06	26.88	29.70	23.85
			1:150	27.45	24.69	26.42	22.05
2263M_Atletiek	05. discus 	1256-1506	1:15	40.66	40.66	40.66	40.66
			1:20	39.54	39.54	39.54	39.54
			1:25	38.86	38.86	38.86	38.86
			1:30	38.36	38.36	38.36	38.36
			1:40	37.97	37.97	37.97	37.97
			1:150	35.86	35.86	35.86	35.86
M66143_Televizer	06. face 	1600-1791	1:5	41.21	30.89	41.84	33.89
			1:10	37.74	28.75	37.78	31.21
			1:15	36.77	27.62	36.60	29.95
			1:20	35.73	26.42	35.99	28.37
			1:25	34.48	25.06	35.24	27.78
			1:30	33.97	24.82	34.89	27.23
M66143_Televizer	08. politics 	2807-2998	1:5	39.36	30.09	38.67	34.63
			1:10	37.30	28.88	36.31	31.66
			1:15	35.97	26.94	35.38	29.97
			1:20	34.78	25.24	34.88	29.05
			1:25	33.66	24.30	33.89	27.89
			1:30	33.02	23.66	33.86	27.59
S&V_2656_Indoor_Volleyball	Volleyball 	6731-6931	1:15	35.75	35.75	35.75	35.75
			1:20	35.16	35.16	35.16	35.16
			1:25	34.71	34.71	34.71	34.71
			1:30	34.34	34.34	34.34	34.34
			1:40	33.90	33.90	33.90	33.90

Original scene	Used scene name	Frames	Compr. ratio	Y-PSNR [dB]	R-PSNR [dB]	G-PSNR [dB]	B-PSNR [dB]
			1:150	32.85	32.84	32.84	32.84

Table 7: Mean PSNR values for the selected scenes

Figure 22 finally puts the Y-PSNR values into relation with the subjective test results. As can be seen, both the PSNR values and the MOS values follow the same trend indicating a reduction of the achievable image quality when increasing the compression. On the other had, the PSNR values cannot be used to predict the observed visual quality in a more precise manner. First of all, both the orange, green, grey and blue PSNR curves show a very similar slop, whereas for the subjective quality assessment, the trend of the corresponding scenes clearly differs. Additionally, there seems to be no relation between the detection of a slightly visible difference and the precise PSNR value. This can be particularly well seen when considering for instance the orange and black curve. The answer to the question whether such a threshold could at least be found for a well-defined film type is probably no, although a corresponding proof would require further subjective quality assessment.

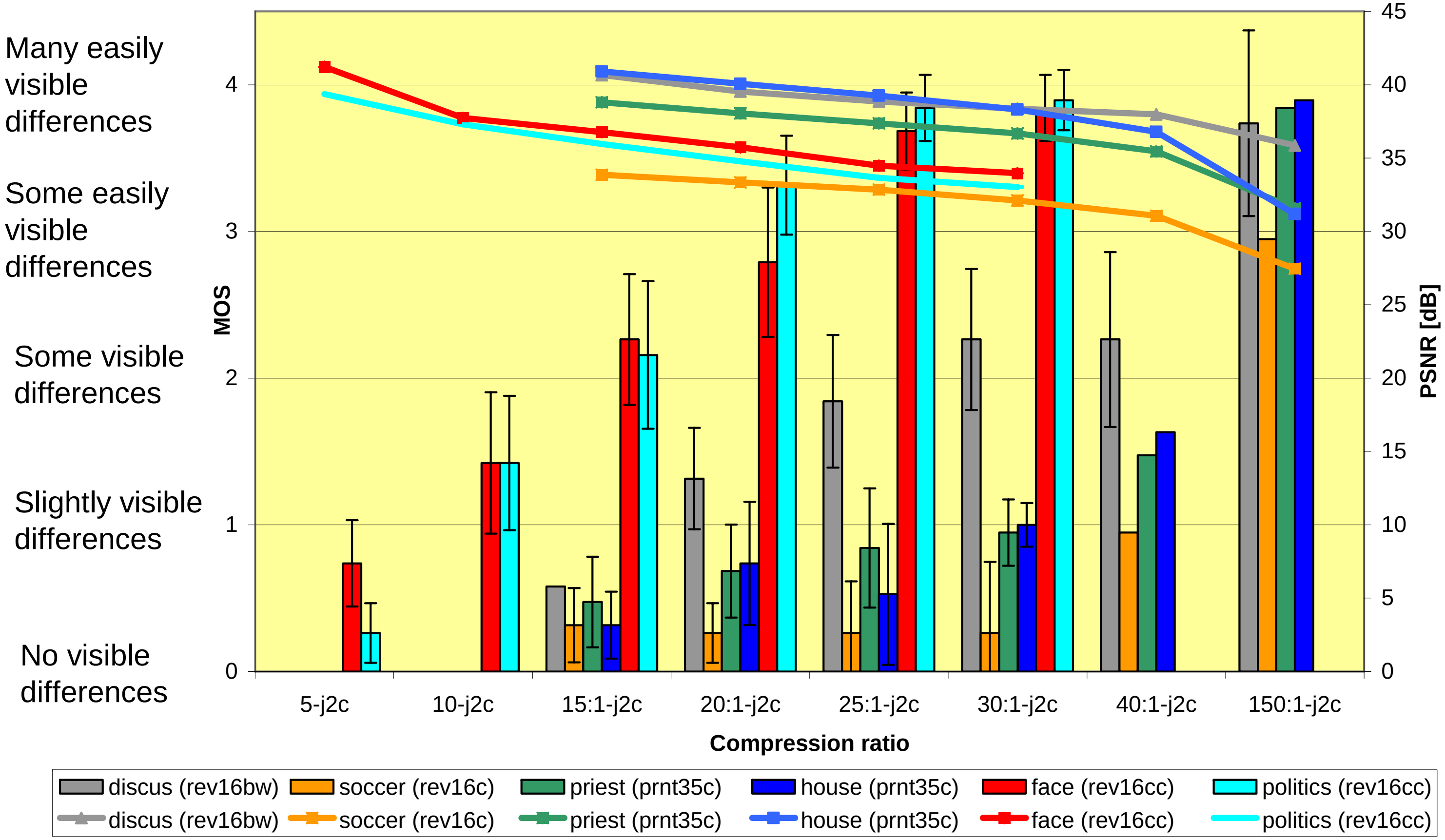


Figure 22: Comparison of the Y-PSNR values with the subjective measurements

5.2 Modulation Transfer Function (MTF) of the JPEG 2000 compression

Since subjective quality assessment is very time consuming and expensive, often objective quality measures are employed in order to quantify the distortions caused by an image processing algorithm or system. Besides the Peak Signal to Noise Ratio discussed in Section 5.1, the modulation transfer function is another possibility to quantify the quality of an image processing system.

However, in contrast to the PSNR, which evaluates the distortion for one precise image, the modulation transfer function characterizes the image processing system itself by defining its spatial frequency response. In other words, it describes for each spatial frequency, how strongly a corresponding sine wave would be attenuated. The higher the attenuation, the lower the observed contrast. The obtained values are normalized such that a very low frequency leads to a transfer value of 1 (or 100%).

Since the modulation transfer function is widely spread in the film and photography industry, its transfer to the JPEG2000 compression would provide archive operators with a familiar quantity of the quality to expect from the overall system. Consequently, a corresponding study has been performed on request of the S&V film archive, in how far this is possible for JPEG2000 compression.

As a result corresponding investigations could demonstrate that the modulation transfer function is inadequate to describe the behavior of the JPEG2000 compression in its entirety. This is because the concept of the modulation transfer function assumes a linear system behavior characterized by the following equations, where I is an input stimulus and f the transfer function:

$$f(I_1 + I_2) = f(I_1) + f(I_2)$$
$$f(\alpha \cdot I_1) = \alpha \cdot f(I_1)$$

In case of the JPEG2000 algorithm, I would be the original, uncompressed image, $f(I)$ the image resulting after compression.

Unfortunately, JPEG2000 violates both equations due to occurring quantization and rate allocation. The quantization, for instance, can be expressed by an integer division of the form

$$q(a) = \left\lfloor \frac{a}{b} \right\rfloor$$

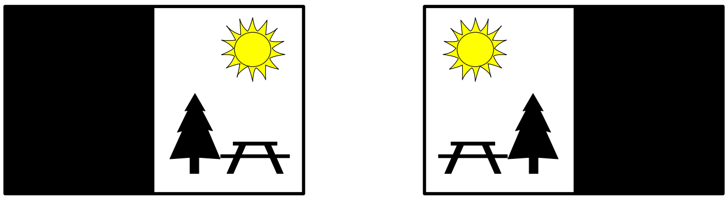
where a is the original pixel value, b the quantization factor, and q the quantized coefficient. Supposing $b = 2, a = 1$ it can be clearly seen that

$$q(2 \cdot a) = 1 \neq 2 \cdot q(a) = 0$$

Similarly, the rate allocation causes that a given image I_1 is saved in the best possible quality by using as many bits as possible for the compressed file without exceeding a given target rate. Assuming now two artificial images as depicted in Figure 23, it can be clearly seen that

$$f(I_1 + I_2) \neq f(I_1) + f(I_2)$$

This is because the combined image $I_1 + I_2$ would require twice the bit rate than image I_1 in order to achieve the same quality. This, however, is not permitted due to the rate constraint forcing JPEG2000 to generate an image whose size does not exceed those of images I_1 and I_2.



I_1 I_2

Figure 23: Two artificial images in order to illustrate the non-linearity of the JPEG2000 algorithm

In other words, JPEG2000 is a non-linear system. This, however, means that the system response to a stimulus I cannot be derived from the impulse response:

$$f(I(x,y)) = f\left(\sum_{n,m} I(x,y)\delta(x-n,y-m)\right) \neq$$
$$\sum_{n,m} f(I(x,y)\delta(x-n,y-m)) = \sum_{n,m} I(n,m) f(\delta(x-n,y-m)) = \sum_{n,m} I(n,m) f_\delta(x-n,y-m),$$

where $\delta(x,y) = \begin{cases} 1 & x = y = 0 \\ 0 & otherwise \end{cases}$ and f_δ the impulse response of the system.

Consequently, also the rules for the transfer function cannot be applied.

5.2.1 Experimental results

The consequences of this fact can be particularly well explained by determining the modulation transfer function using a sequence of grey-scale, 8bit artificial images. Each of them has the same size and consists of a quadrangle with slanted edges as depicted in Figure 24. This allows direct derivation of the modulation transfer function using the imatest tools.[2] The background consists of randomly generated noise whose strength varies between the different images. From the classical understanding, the modulation transfer function measured at the slanted edge should be (almost[3]) independent of the contained background noise.

[2] http://www.imatest.com/guides/modules/sfr

[3] Ideally, the JPEG2000 impulse response should remain an impulse. Pratcially, it will be blurred at little bit.

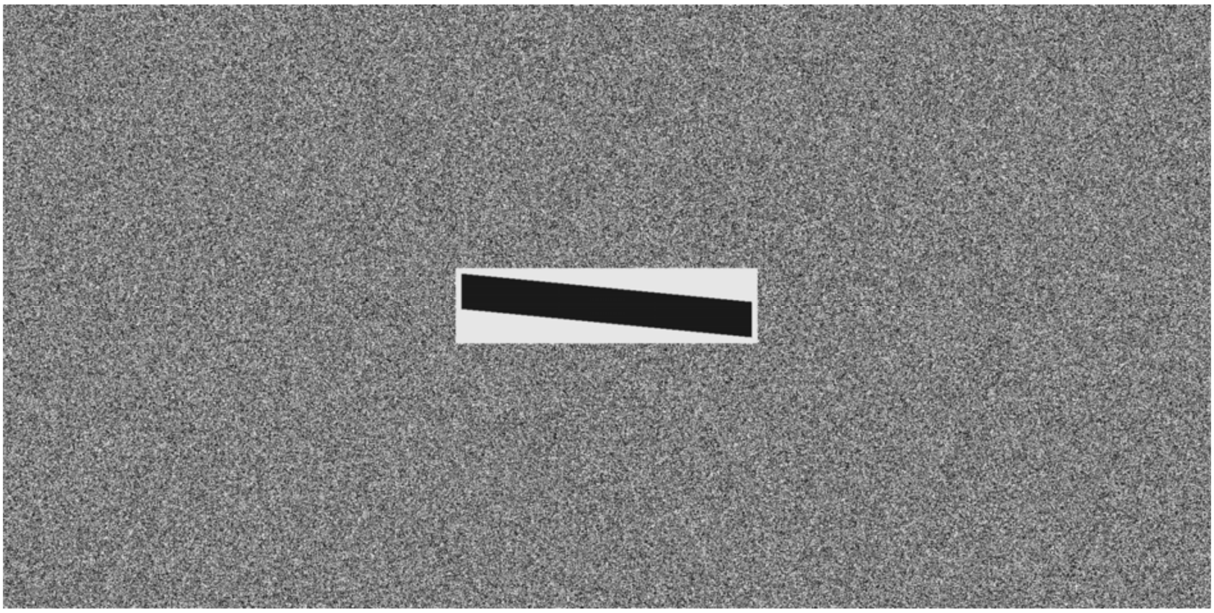


Figure 24: Exemplary artificial test image used for MTF measurements. The slanted-edge rectangular together with the light-grey background stays the same for each image of the test sequence. The noisy background on the other hand varies between the different images.

Figure 25 shows the corresponding results for different compression rates ranging from 0.1 bits/pixel up to 7.0 bits/pixel, while Table 8 establishes the relation between the bit rate and the compression ratio.[4] Note that the depicted MTF curves uniquely characterize the JPEG2000 compression and do not include any impact of a scanner, optical lenses etc. The different curves correspond to different noise variances of the background noise. As can be seen, the MTF strongly depends on the image background. In other words, the MTF depends on the image content. This can be explained by the non-linearity of the JPEG2000 algorithm. For low bit rates and heavy noise, many bits are required to encode the background appropriately. This, however, means, that only very few bits are available for the slanted edge, leading to blurred edges and thus heavy attenuation of the high frequency signal components. On the other hands, if the background is very smooth, it can be compressed very easily. Consequently, many bits remain for encoding the slanted edge, leading to "good" modulation transfer curves.

This effect is of course particularly pronounced for low bit rate constraints. The more bits are spent for the overall file, the less the slanted edge and the noisy background need to compete for the available bits, and consequently, the less the MTF curves depend on the image content.

Note that the decline of the curves for high frequency values cannot be attributed to the JPEG2000 compression, but are rather a problem of limited measurement precision. This can be demonstrated by Figure 26 depicting the MTF for an uncompressed slanted edge. As can be seen, the MTF also shows a decline for high frequencies, since the image granularity is limited by

[4] Comparison with the bit rates given in Table 5 shows that the compression rates are rather low. This comes from the fact that the noisy background is much more difficult to compress than real images, needing thus more bits per pixel.

the image pixel structure. In other words, Figure 26 shows in some sort the MTF of an ideal JPEG2000 compression.

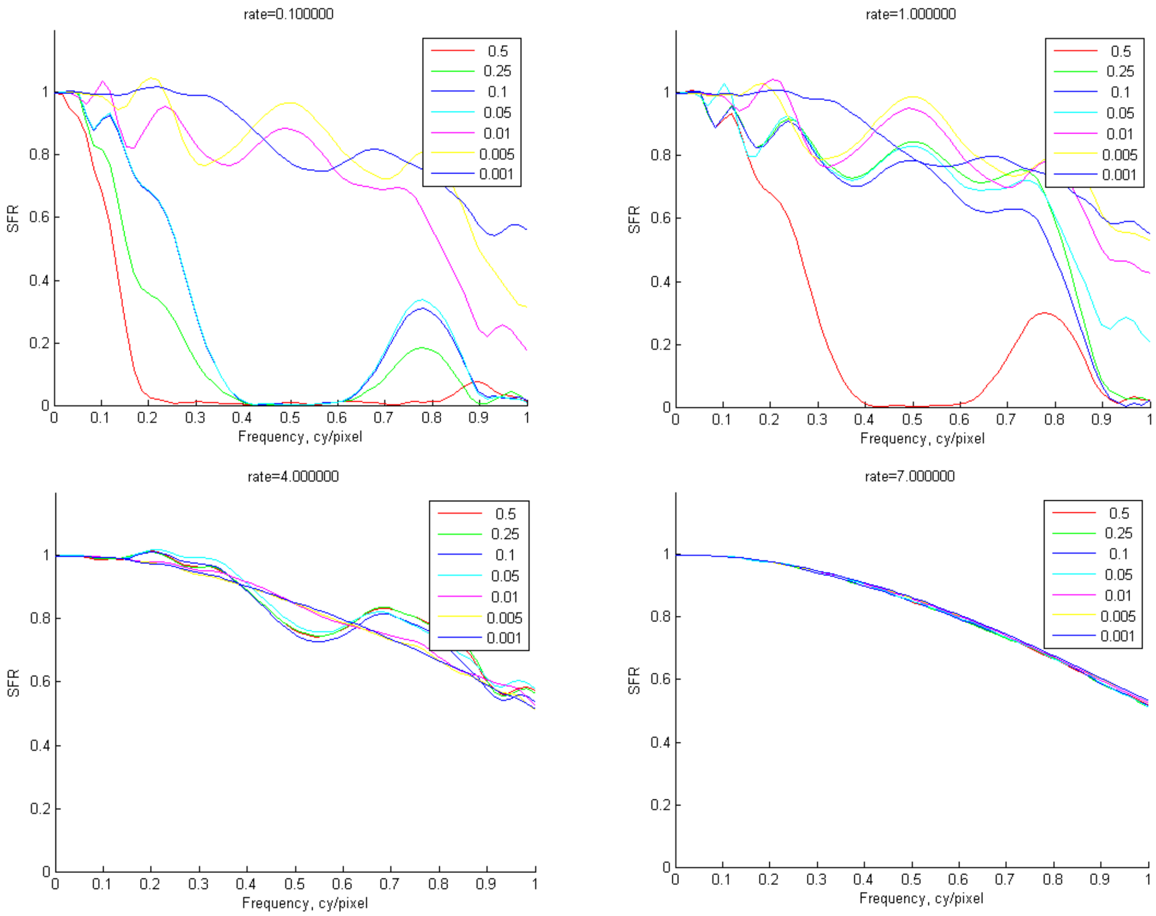


Figure 25: MTF for JPEG2000 compression with constant rate (indicated in the diagram titles). The different lines correspond to different noise variances. The larger the value, the more noise is contained in the background, the more difficult it is to compress.

Rate	Compression ratio
0.1 bits/pixel	1:80
1.0 bits/pixel	1:8
4.0 bits/pixel	1:2
7.0 bits/pixel	1:1.4

Table 8: Compression ratios used for the MTF experiment

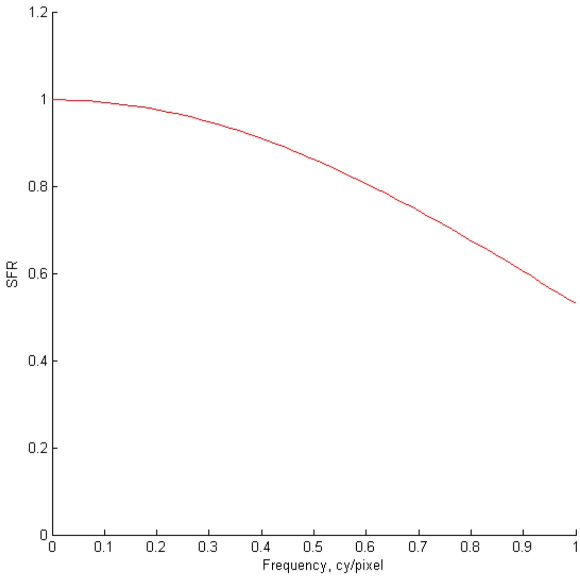


Figure 26: MTF of an uncompressed image

In order to provide a better impression how the decompressed slanted edge looks, Figure 27 depicts two corresponding examples. Obviously the edge gets more blurred when a noisy background is present, although the slanted edge and its light-grey environment are identical in the uncompressed images.

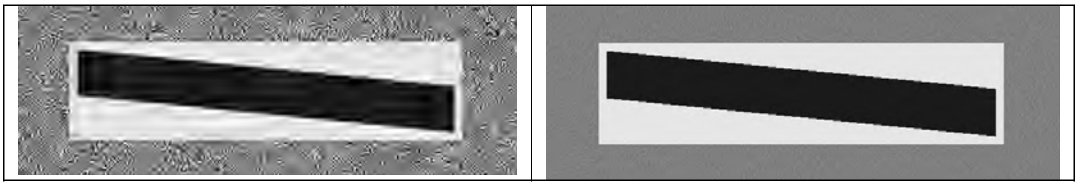


Figure 27: Decompressed slanted edge, both using 0.1 bits per pixel. For the left image, a noise variance of 0.5 has been used; the right image corresponds to variance of 0.001

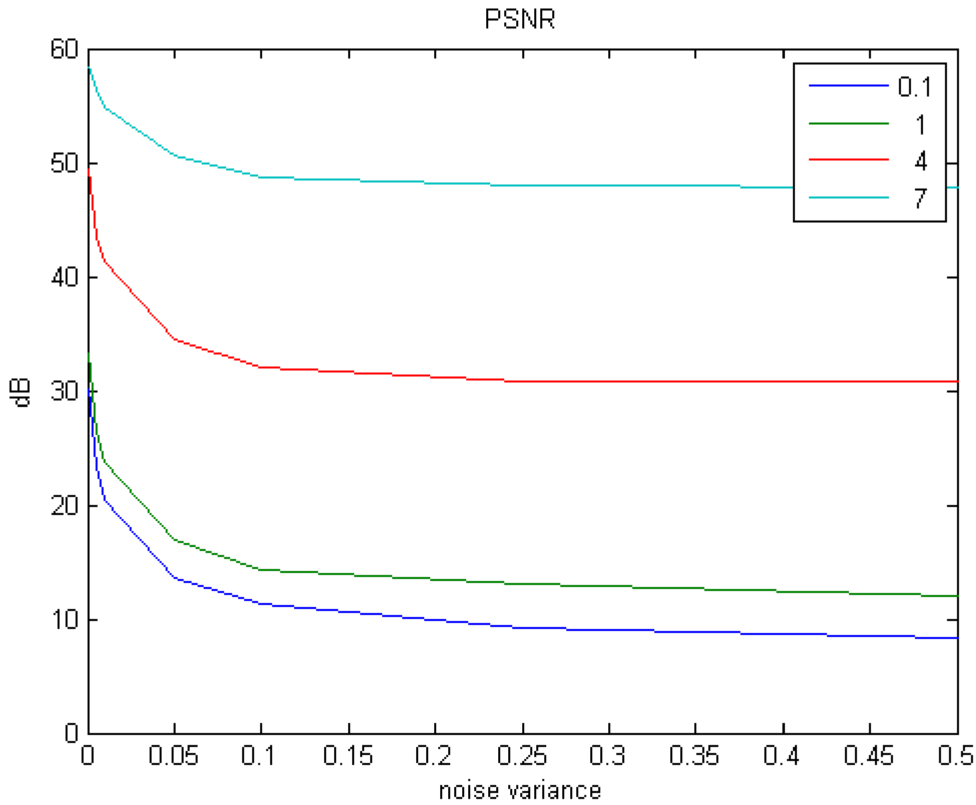


Figure 28: PSNR measurements for the previous scenarios. In this diagram, the lines represent the different target bit rates.

Figure 28 puts the MTF curve in relation to the PSNR values. This demonstrates that for a given bit rate the PSNR decreases with increasing noise variance. This is easy to understand, because the stronger the noise, the more difficult is the image to compress. Furthermore, it can be seen that the lower the target bit rate, the higher the influence of the background on the PSNR. However, in contrast to Figure 25, where the impact of the noise is almost vanished for a bit rate of 7 bits/pixel, the PSNR diagram in Figure 28 still clearly shows this impact. In other words, it seems that from a given bit rate on sufficient bits are allocated in order to properly represent the slanted edge. Additional bits are then particularly beneficial for coding the background.

5.2.2 Impact of the rate allocation

As explained in Section 5.2, the unusual behavior of the Modulation Transfer Function is due to two components of the JPEG2000 algorithm, namely the rate allocation and the quantization. In order to quantify their impact, a further experiment has been performed, disabling the rate allocation and performing only quantization.

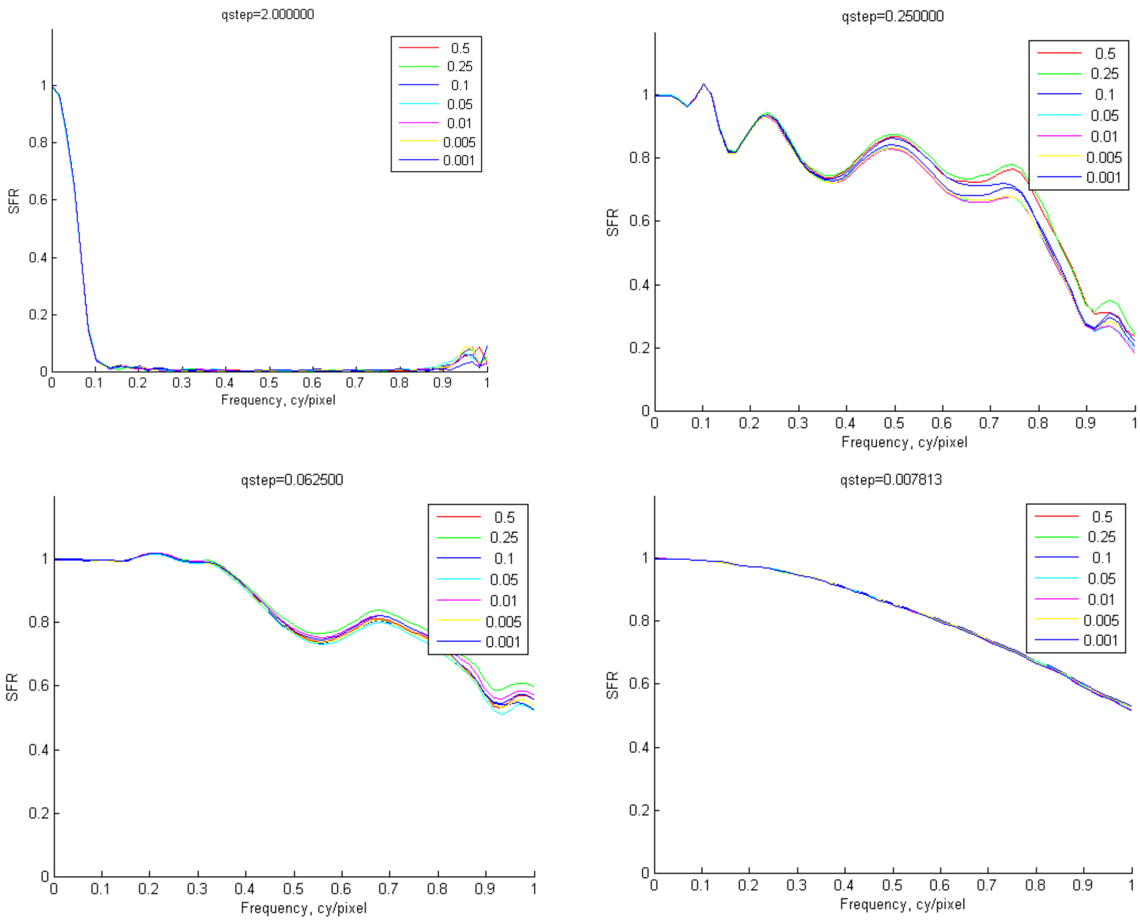


Figure 29: MTF curves for disabled rate allocation. The larger the qstep value, the stronger the quantization. The different curves correspond to different noise variances.

Figure 29 depicts the obtained results. First of all it can be stated that the impact of the noisy background on the MTF curve is far less important than with enabled rate allocation, even if the quantization is very strong and leads to “bad” MTF curves. This confirms the assumption that the rate allocation is an important factor disturbing the non-linearity of the JPEG2000 algorithm. Furthermore, it can be seen that the smaller the quantization the smaller the impact of the noisy background on the MTF curve. This is easily explained by the fact that the smaller the quantization, the smaller the derivation of the integer division from a normal, linear division. And finally, it is also evident that a smaller quantization improves the image quality in that high frequencies are less attenuated.

5.2.3 Conclusion

Although being a widely spread measure for the quality of an image processing system, the modulation transfer function has difficulties in representing the JPEG2000 algorithm in its entirety. The problem is situated in the fact that the transfer function in a particular image point depends on the complete image. Furthermore, since the JPEG2000 algorithm is non-linear, a transfer function determined for one image cannot be transferred to another image. In other words, it misses all important properties that make the modulation transfer function an intuitive tool for measuring the quality of an image processing system. Consequently, its application cannot be recommended.

5.3 Lossless image compression

Since for very important film material a lossy archival cannot be accepted, this chapter presents the results for a lossless compression. To this end, the sequences used for PSNR and subjective quality assessment have been processed accordingly using a Kakadu based compression engine.

Figure 30 depicts the corresponding results by showing the file sizes for the individual sequences. For each of them, the first 175 files haven been selected. Obviously the required storage space is much larger than for lossy compression. Furthermore, it significantly depends on the image content. In particular the black and white images can be compressed much better than the colored ones. However, for a meaningful comparison the size of the b/w DPX files has to be divided by three because all three color components contain the same data. Again the soccer sequence proofs to be very difficult to compress. This corresponds to the performed PSNR investigation where this sequence attained the lowest values.

Nevertheless, in all cases a lossless compression might be beneficial compared to the storage of the 2K DPX files as they come out of the scanner, since the latter occupy approximately 8 Mbyte per frame of film.

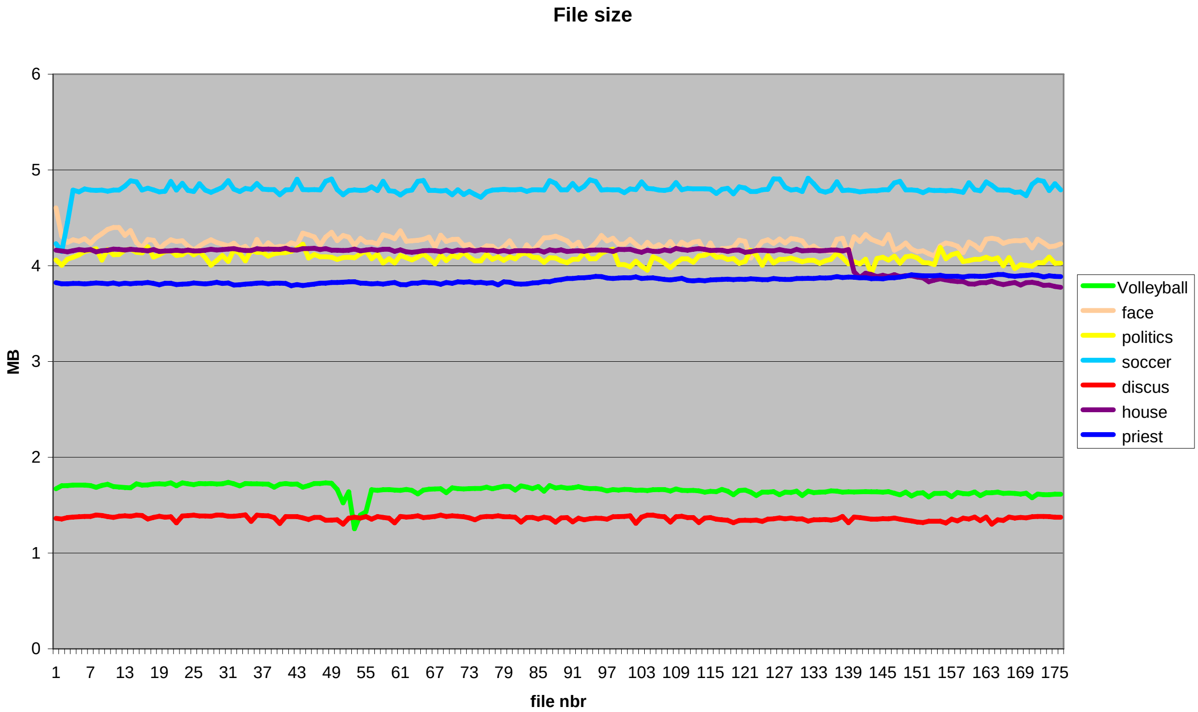


Figure 30: Achievable file sizes for lossless compression

5.4 Quality control measures that can be used during the production process

Since scanning and processing errors must be avoided in order to not impact the value of the archived material and since manual inspection is time consuming and expensive, at least an automatic pre-verification is highly recommended for an efficient archiving workflow. To this end, several possibilities exist as described in the following subsections.
The findings and results for each tool should be stored together with the media data. Additionally, descriptions and version numbers of the QC tools should be stored alongside.

5.4.1 Image quality control and automatic histogram analysis

Although it is difficult to predict the expected PSNR and PAE (Peak Absolute Error) values, these two measurements can be used to detect possible compression errors by defining a corresponding threshold. If actual PSNR falls below this threshold, or the PAE exceeds the defined maximum, a visual inspection should be performed to avoid any compression difficulties. Automatic histogram analysis can be used in order to detect scanner calibration problems. A corresponding example could for instance be found in the priest sequence. Figure 31 depicts the corresponding results for one selected image. From this image it can be seen that a significant color clipping occurred in the red color channel. Looking at this channel individually, this can indeed by observed in the white shirt of the priest. Furthermore, it can be recognized that the green and blue color channels seem to be misaligned. Finally a histogram analysis can help to detect errors that are introduced when a reduced bit depth instead of the desired 10 bit precision was used in an intermediate processing step.

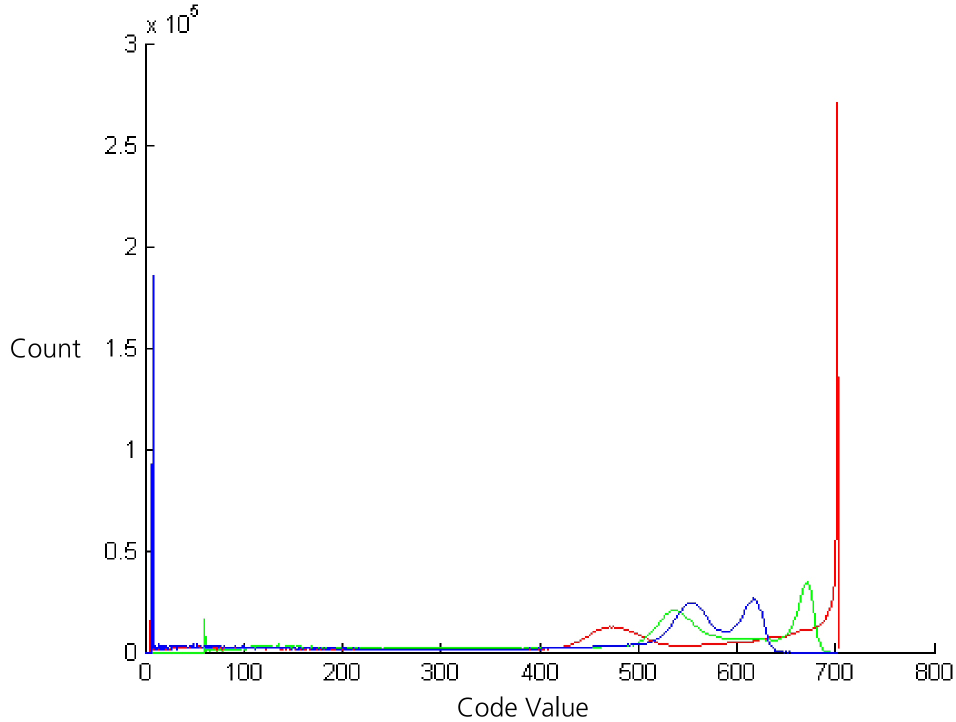


Figure 31: Histogram analysis for image S&V_2.2-1735-9554.00002449 of the priest sequence

5.4.2 Scanner / DPX metadata analysis

Since technical and processing metadata should be recorded during scanning and stored either (preferably) within the DPX files or in a separate file the existence and completeness of this metadata can easily be checked. It should also be verified that the data is according to the specifications for the category of the film, e.g. that the specified color space information and data about scanner settings are present and in a reasonable range.

5.4.3 JPEG2000 codestream analysis

In order to avoid that erroneous hardware or a software bug leads to corrupted image, it is recommended to perform a structural analysis of the JPEG2000 codestream. This can be performed on different levels including:

- verification that the codestream can be parsed and decoded without errors,
- verification that the selected profile parameters are met.

Whereas the latter point requires only very little computation power, decoding is of course more expensive. Independently, both tests shall be performed on a machine different of that for encoding in order to detect possible hardware failures.

5.4.4 MXF structural analysis

Similar to the JPEG2000 codestream analysis the file structure of MXF files can be analyzed and checked for errors in general and conformity to a specific operational pattern or application specification. The verification of the MXF files is strongly recommended and should also include checks of the metadata for completeness and integrity.

5.4.5 Data error detection

No storage media and system is fail-proof in a sense that it is immune to bit errors. Errors can occur not only to data on media but also during such processes like simple copying of files from one medium to another, no matter if hard disk or magnetic tape. Therefore, measures should be taken to be able to detect such errors and to provide redundancy in order to be able to correct the detected errors.
Two main concepts exist to accomplish this goal: the simplest possibility is to separately calculate checksums for single files and to duplicate all files on at least two separate sets of media. Checksums can easily be calculated using standard tools like MD5. Having two separate copies of each file provide for a very high level of redundancy, especially if the second copy is stored geographically separated from the first copy. However, this practice is also very expensive. A more advanced approach to the problem is called forward error correction. In this concept the combination of checksums and redundancy is integrated into the data stream itself. The level of redundancy can be freely chosen according to the properties and error probability of the media or transmission channel. On the file level forward error correction has to be integrated into the

file format or an additional separate wrapper format has to be defined. It is also a computational complex process. Forward error correction is commonly used on the lowest abstraction layers, e.g. integrated into the data stream that is physically written to a tape or DVD medium. One example for forward error correction is the Reed-Solomon code that is used on Compact Disc, DVD and LTO media.
For archiving purposes it is advisable to create checksums of all files and to provide for redundancy of the data separately because this is the least complex possibility.

6 Recommendations

In this section recommendations are formulated to help the Netherlands Institute of Sound and Vision to take decisions on the implementation of the processes. Since the matter is very complex these recommendations are in some cases not definite in the sense that there is only one solution to a problem or issue. The recommendations try to express a practical way for S&V that in some points may include trade-offs compared to the theoretically optimal solution. The final decision on the implementation of the recommendation has to be taken by the Netherlands Institute of Sound and Vision.

6.1 Film scanning

(1) Calibrate the scanner to the chromaticity coordinates of the primary colors described in ITU-R Rec. BT.709. Quantize to 10 bits per color component and use the logarithmic Cineon/DPX transfer curve to make best use of the bit depth.

This calibration ensures a consistent color rendition and provides and easy process for conversion to HDTV formats using simple 1D LUTs while simultaneously preserving most of the color information that can be suspected to be present in the original film image. S&V has to decide if this trade-off is acceptable for its material since the Rec. 709 color gamut is smaller than the best film gamut (see [1]). It is also possible to use this approach for the majority of material and define an extra route for very valuable material.

(2) Scan at the highest reasonable resolution, but at least at the minimal resolution used for distribution. Use the scanner's optical system, if available, to adapt the film area to the imaging sensor size. Don't use digital scaling or resizing at this stage other than what is absolutely necessary because of oversampling and done internally in the scanner!

With HDTV distribution as a goal the minimum resolution for scanning should be in the 1920 x 1080 frame, where the exact resolution depends on the aspect ratio. Since scaling and cropping may be required it is advisable to scan at a higher resolution than theS&V's current HDTV target, e.g. at 2048 x 1556 for 4:3 aspect ratios. This resolution leaves enough room for cropping or pan/scan and when downscaling leads to a much better image quality because alias artifacts can be minimized. Scanning at SDTV or 1K resolution should be avoided since upscaling to HDTV then leads to blurry images. Even if the 16 mm films themselves do not provide the real details of 2K, the grain would give the viewer the impression of details. This grain would more or less disappear when scanning at 1K.

The images should be stored in the native resolution of the scanner. They may be cropped in order to adapt to the aspect ratio of the film but not digital scaling or resizing should be employed prior to the JPEG2000 compression in order not to impair the image quality.

(3) Establish different groups of material that are scanned and processed with a common set of parameters that are optimized for each group.

Possible groups can be 16 mm reversal color, 16 mm reversal b/w, 35 mm negative color, 35 mm negative b/w, 35 mm print film color etc. A distinction should be made between color and b/w film. Also defining additional categories from a non-technical point of view can be reasonable: e.g. films of very high cultural value (highest scanning quality to be used), films of lesser value (stronger emphasis on smaller files and faster process) etc. Categories should be used for scanning as well as for JPEG2000 encoding. Common sets of parameters could include optimized LUTs and scanner calibration files color space transformations, transfer curves etc.

(4) Store black and white image data in a single color component DPX file during scanning.

When storing monochromatic images in RGB files all three channels contain the same information and therefore occupy three times the storage space that is actually necessary. Also the transfer of the files to and from the LTO tape takes three times the time. If the employed scanning or postproduction software is only able to handle three component (RGB) DPX files open source tools can be used to convert to and from single component DPX files.

(5) For color-faded material the black levels and gain settings for each color channel should be adjusted during scanning to use the maximal value range of the 10 bits in the DPX files.

The results of the visual quality assessment clearly show that a much lower compression ratio must be used if the images shall be stored faded as they are. To be able to use higher compression ratios the fading has to be compensated by appropriate level adjustments. This can probably not replace a real image restoration and it may be possible that a serious restoration cannot be based on the lossy compressed JPEG2000 files. After the adjustments the images can be handled with regard to the JPEG2000 compression settings like non-faded material and 10 bits per component will most probably be sufficient. For faded or deteriorating material of very high value it may be reasonable to scan at more than 10 bits per color component and store the data either uncompressed or using mathematically lossless JPEG2000 compression.
The settings for all adjustments should be recorded as metadata in order to be able to retrace them!

(6) Store as much metadata as possible directly in the DPX files.

The most important information are color space specifications and settings of the scanner but also other information become valuable in order to detect problems and facilitate processing. Even if the scanner or other software used during scanning does not support all necessary metadata in the DPX files, commercially or non-commercially available software can be used to add these. Alternatively, the required data can be stored in an extra test or XML file. But this is then only sub-optimal because of the risk of losing the connection between the images and the metadata.

(7) If color correction is necessary use only basic primary color corrections that can be describe with the means of the ASC CDL. Don't apply the color correction to the image (i.e. do not "bake it in") but store the settings for automated application during creation of the distribution formats.

This way the image quality is not impaired by a color correction that may have to be changed in the future. The process is in principle straight-forward but it has to be implemented and adds an extra burden to the planning of the complete workflow. S&V has to decide if this flexibility is needed. If this approach is not followed it is nevertheless advisable to use only the parameters of the ASC CDL and store the settings for retraceability as metadata.

6.2 Image compression / JPEG 2000 encoding

(1) Use JPEG2000 lossy compression ratios appropriate to the material category and derived from the findings of the visual quality assessment session.

As a result of the visual QA a compression ratio of 15:1 can bee seen as a safe choice for all types of non-faded color and b/w material. In general, color material can be compressed with slightly higher compression ratios compared to b/w material because of the correlation of the color components. For material of lower value also higher compression ratios up to 30:1 may lead to acceptable results. The resulting bit rate in Mbit/s can be calculated from the spatial resolution, bit depth, number of color components and the frame rate.

(2) Separate material into different categories for compression.

Not all films have to be compressed using the same settings and compression ratios. It makes sense to have at least a minimal set of two categories for color and b/w films. Additional categories for material of very high and comparably low value might be defined.

(3) Define one common JPEG2000 profile and a basic common parameter set that is used for all films.

From a practical point of view it may be reasonable to use either the 4K DCI profile or the Scalable 4K Digital Cinema profile to store the 2048 x 1556 resolution because they *are widely supported by existing hardware and software solutions. The image should not be scaled to fill the 4K resolution container. Instead it can be padded with black borders that do not impact compression efficiency or a slightly modified profile can be used (but with a possible negative impact on interoperability and compatibility). b/w images can be handled as three component images without impact on compression efficiency. In case of padding, the actual image area within the 4K container has to be stored as metadata in the file. Since S&V's material is not intended to be projected in digital cinemas the padding is not problematic.*

6.3 Quality control measures

(1) Calculate checksums of all files using the MD5 algorithm to be able to check data integrity and apply redundancy where deemed necessary.

The MD5 values should be calculated for the MXF files as well as for the DPX, audio and metadata files that are written to LTO tapes in order to be able to detect errors in the files. The checksum should be stored together in the same place as the original files and in a second copy on separate media or systems.

(2) Establish image quality control mechanisms for that are able to detect problems in scanning and compression.

Histogram analysis should be implemented to detect problems created during scanning such as unevenly set black levels or clipping. An automated PSNR analysis should be used to detect quality problems in compression. The PSNR results should be stored as metadata with the images in order to be able to retrace problems.

(3) Check metadata stored in the DPX files and that is delivered from scanning for existence, completeness and consistency.

Especially essential metadata like color space information and scanner settings are of crucial importance.

(4) Analyze and verify the JPEG2000 codestream.

To make sure the JPEG2000 files are useable the codestream verification is an essential step. Also the compliance of the encoding parameters with the specification for the digitization project should be checked.

(5) Analyze the MXF files.

All MXF files should be analyzed and checked for compliance with the specification for this project. Additionally, it is of crucial importance that all specified metadata is present.

6.4 Metadata

(1) Collect, record and store metadata during scanning. Store this metadata as far as possible directly in the DPX files. Store additional metadata in on XML file per film.

The DPX format provides an extensive set of metadata fields that should be used and filled as far as reasonable. This recommendation encompasses the correct usage of essential internal fields like the image element descriptors, transfer characteristic, colorimetric specification (see SMPTE S286M-2003 for a reference list) as well as other metadata fields (attributes). Where the above mentioned internal fields are ambiguous additional information should be provided in user defined data section. This is especially true for the exact colorimetric specification.
The metadata set should incorporate, but is not limited to, the following elements:

- *The film name and a reference to the S&V catalogue.*
- *Information about the film material itself (gauge, b/w, color, reversal etc.)*
- *Timecode and/or frame number.*
- *Detailed color space specification including a name, a reference to a standard and the chromaticity coordinates of the primary colors and the white point.*
- *Information about the make, model, serial number and software revision of the scanner as well as date of last calibration.*
- *Geometric information about the scanned image, such as dimensions, position of the digitized image area with reference to a fixed point of the film, scanning resolution.*
- *Detailed scanner settings (depending on scanner type), at least black levels and gain values for each color that are applied in addition to the calibration.*
- *All lookup tables (LUTs) that have been applied to the image during scanning.*
- *Detailed information about any additional processing steps the image went through including which systems or software was used.*
- *If color correction has been applied or is to be applied during creation of the distribution formats the settings in a universal format (ASC CDL).*

Relatively large data sets such as LUTs may be stored outside the DPX files in one single (XML) file per image sequence while all other essential data should be stored within the DPX file as far as possible and practicable. While support for all possible metadata may probably not be included directly in the scanner's software commercially and non-commercially available software can be used to add this information.

(2) Transfer the complete metadata that has been recorded during scanning from the DPX and XML files to the MXF files.

Since this information is important for future processing steps it has to be carried along to the MXF files.

(3) Store additional important information in the JPEG2000 MXF files.

The JPEG2000 MXF files that are stored in the managed storage system should contain specific information about the JPEG2000 files. Besides the standard descriptors information about the tools and settings that have been used to encode the files should be stored. Also additional information from quality control processes such as PSNR values may be helpful. From a preservation point of view it may make sense to also store a copy of the descriptive metadata from the archive's catalogue in the MXF file. At least a unique reference to the catalogue record should be stored in the file.
The metadata set should incorporate, but is not limited to, the following elements:

- *JPEG2000 profile used.*
- *JPEG2000 encoder manufacturer, model/product and version including hardware revisions and software versions used.*
- *JPEG2000 encoding parameters.*
- *PSNR for each image as far as it is calculated.*

7 Conclusion

This report gives an overview on the results and findings of the study that has been carried out for the Netherlands Institute of Sound and Vision to define technical details for scanning and JPEG2000 encoding of films. Because of the wide and complex nature of this field it is impossible to cover all technical details in the same depth. The report shows the most important issues and problems in this area and tries to give a guideline in to form of recommendations on how an optimized process can be accomplished. For some issues no definite solution can be presented. The Netherlands Institute of Sound and Vision has to take a decision between several options to find the best overall solution for its particular situation.

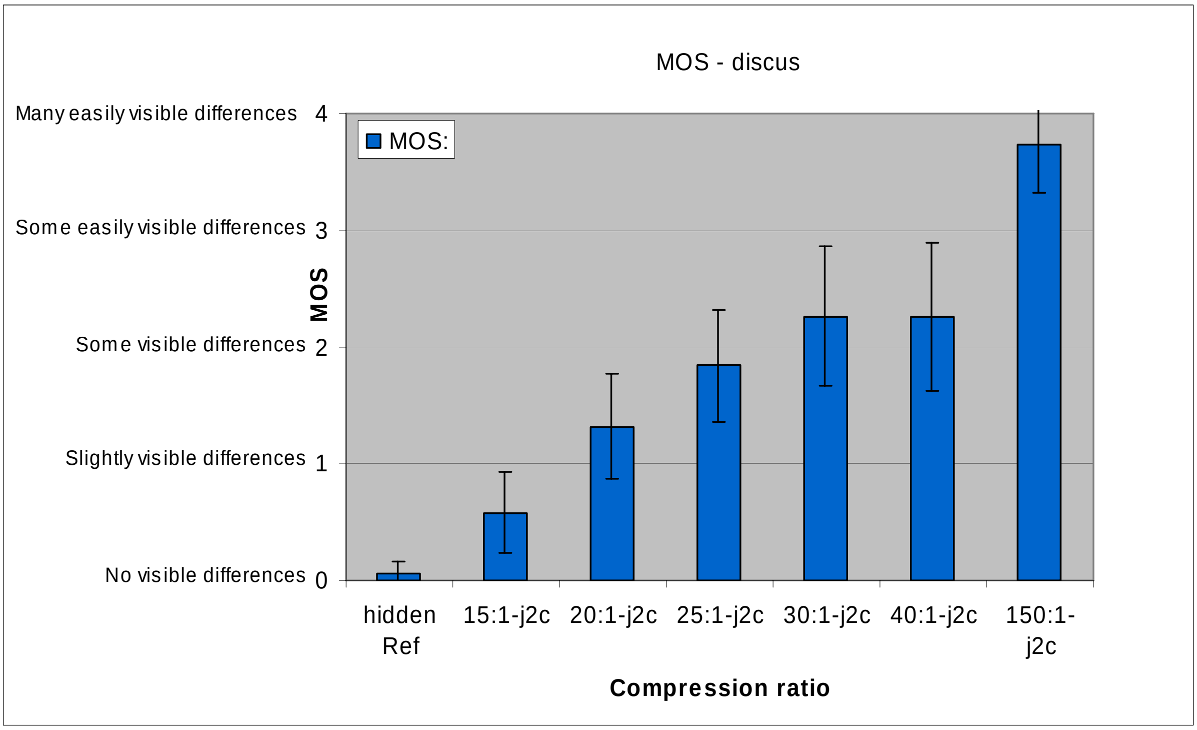
Fraunhofer IIS

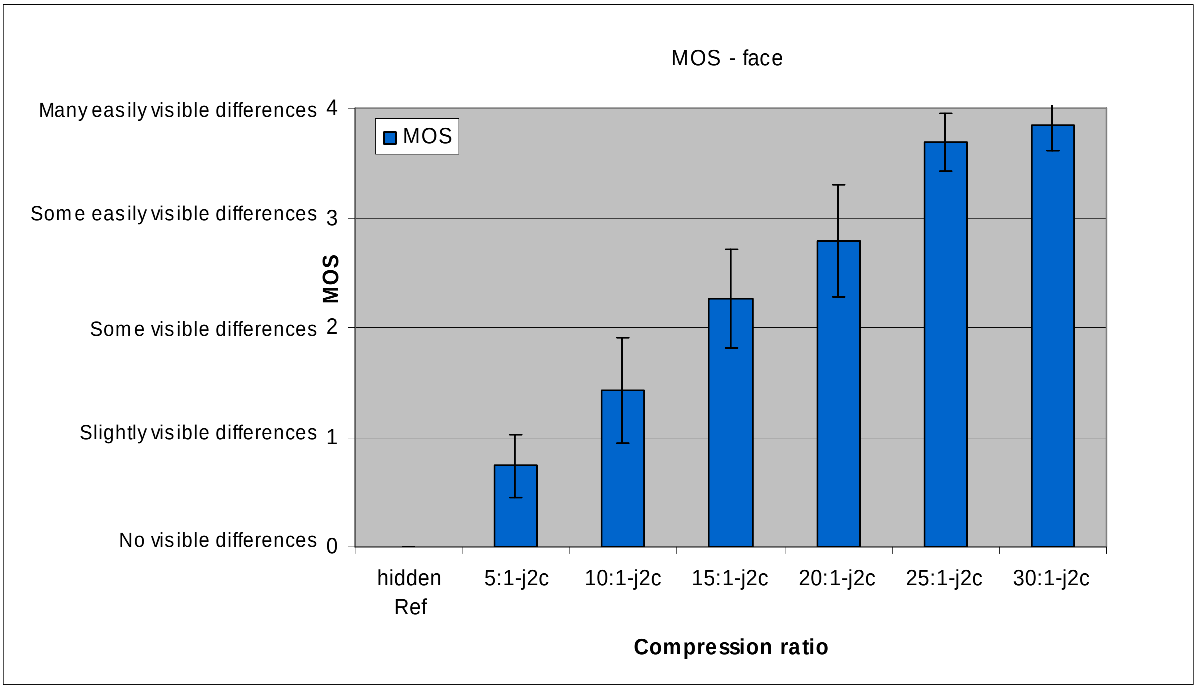
8 References

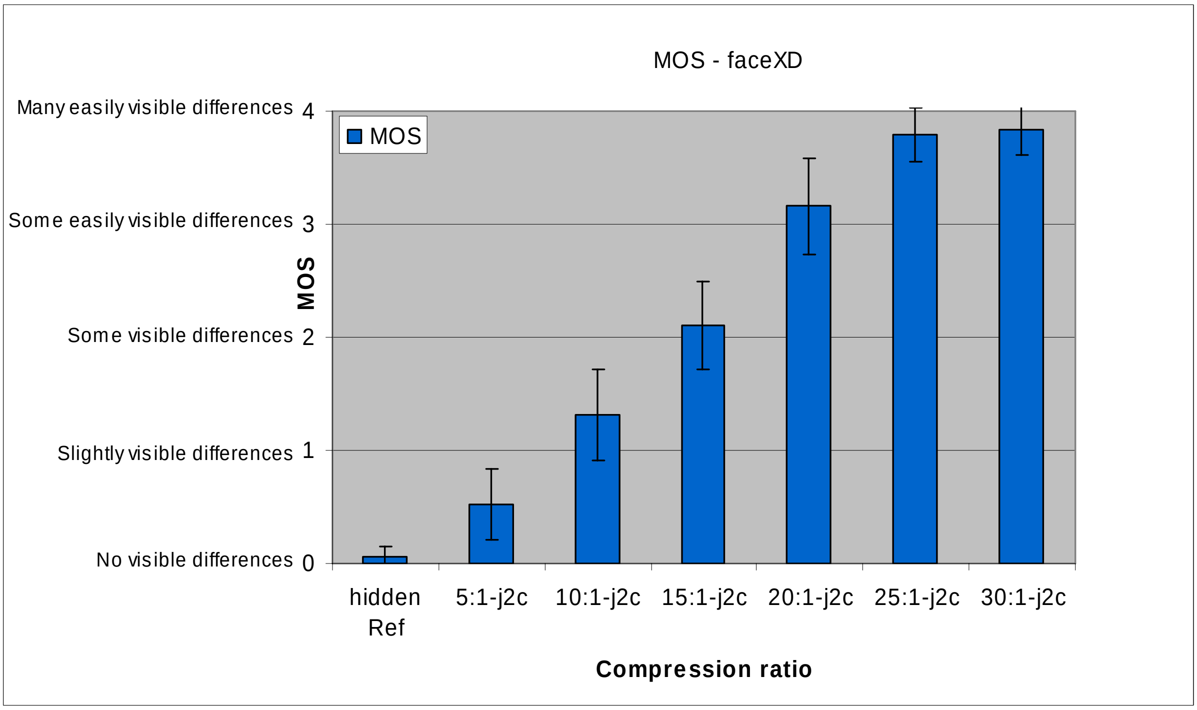
[1] Color and mastering for digital cinema, Glenn Kennel, Focal Press, 2007
[2] Conversion of 10-bit Log From Data To 8-bit Linear or Video Data for the Cineon Digital Film System, Version 2.1, July 26, 1995, Kodak
[3] JPEG2000 Image Compression Fundamentals, Standards and Practice
David S. Taubman, Michael W. Marcelling, 2002
Chapter 16.1
[4] The JPEG 2000 still image compression standard
A. Skodras, C. Christopoulos and T. Ebrahimi
IEEE Signal Processing Magazine, 2001
[5] Zur Theorie der Farbmischung, Hermann Günther Graßmann, in: Poggendorfs Annalen der Physik und Chemie 89, pp. 69–84, 1853
[6] Commission Internationale de l'Eclairage proceedings, Cambridge University Press, Cambridge, 1931
[7] http://www.theasc.com/clubhouse/committee_tech.html

9 Annex: Diagrams

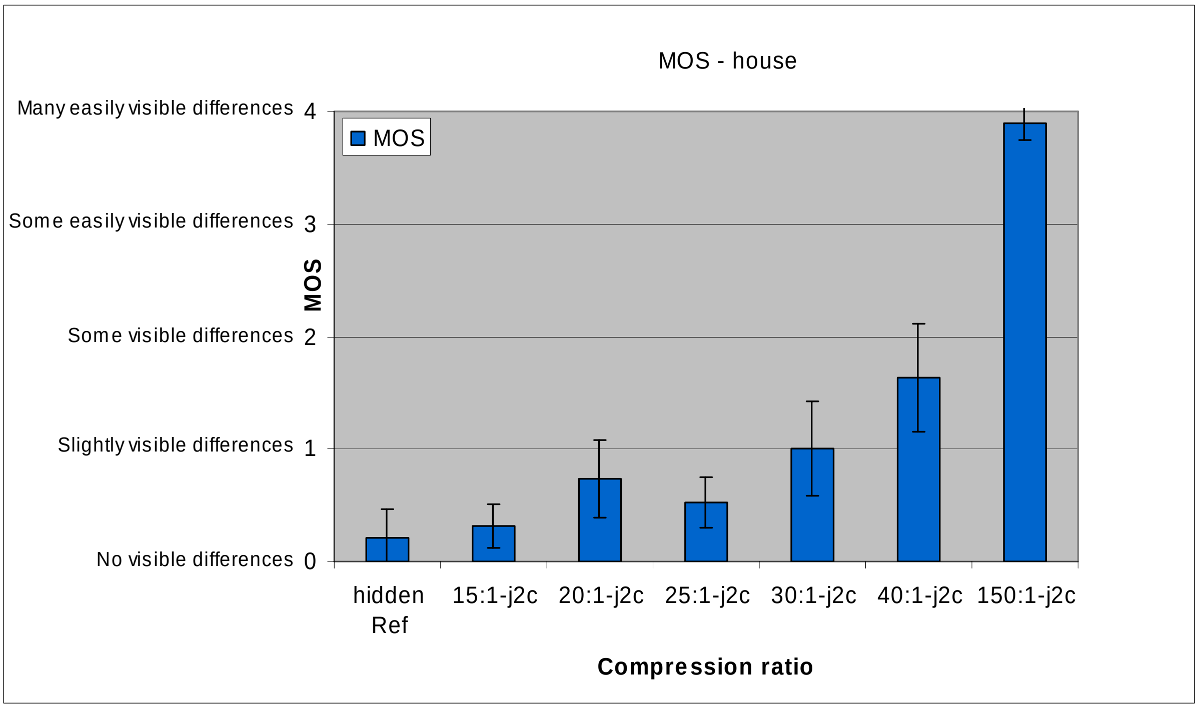
9.1 MOS Diagrams

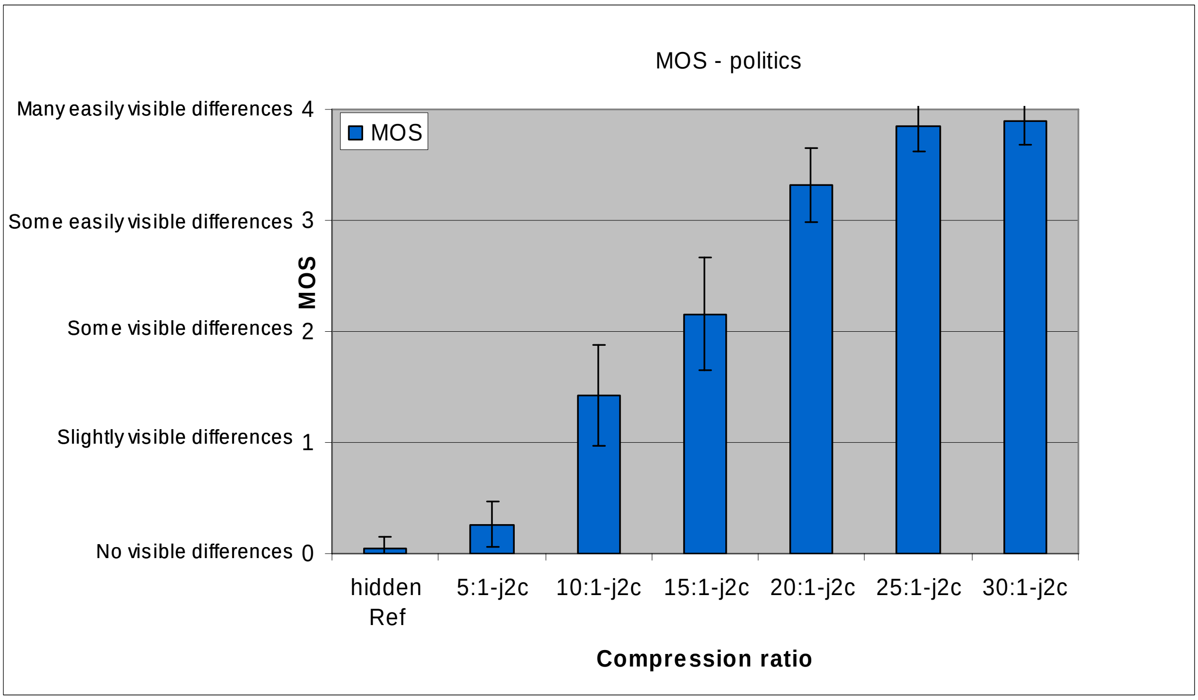


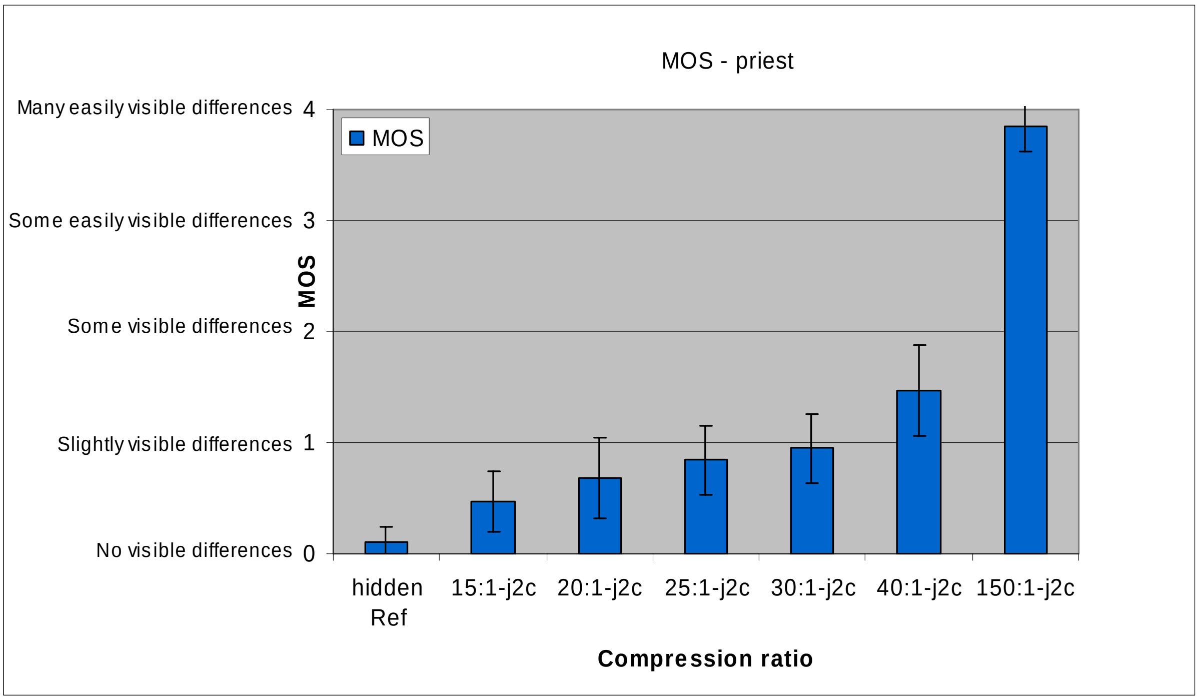

MOS - face
MOS
Many easily visible differences 4
Some easily visible differences 3
Some visible differences 2
Slightly visible differences 1
No visible differences 0
hidden Ref
5:1-j2c
10:1-j2c
15:1-j2c
20:1-j2c
25:1-j2c
30:1-j2c
Compression ratio

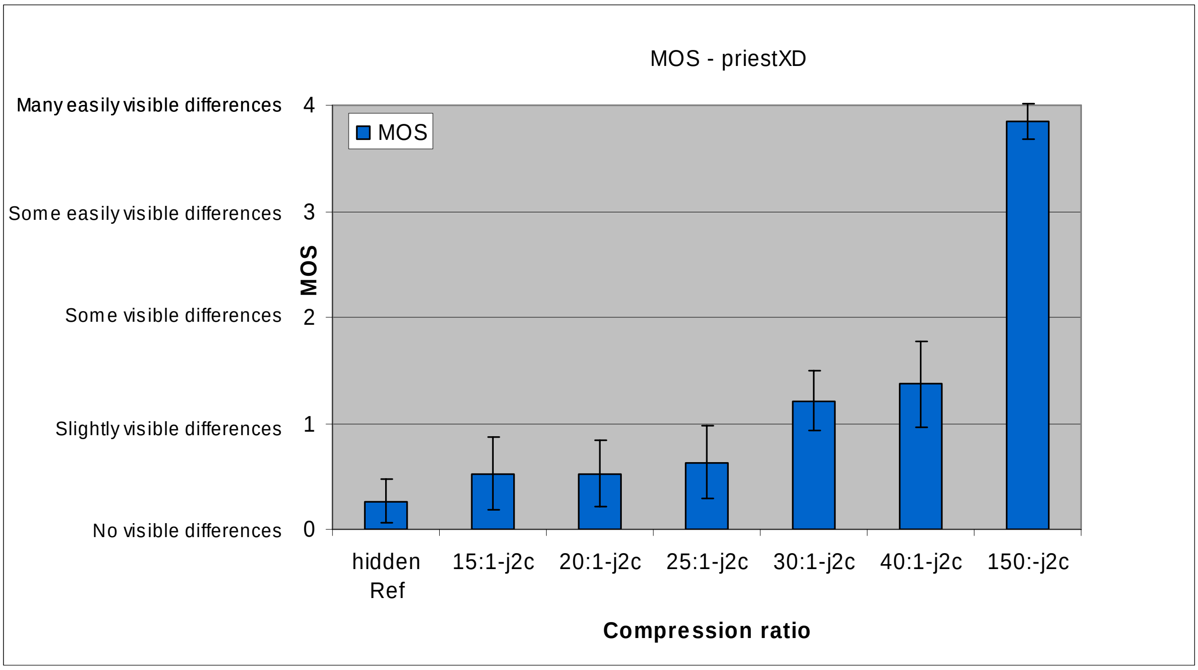

MOS - faceXD
MOS
Many easily visible differences 4
Some easily visible differences 3
Some visible differences 2
Slightly visible differences 1
No visible differences 0
hidden Ref
5:1-j2c
10:1-j2c
15:1-j2c
20:1-j2c
25:1-j2c
30:1-j2c
Compression ratio

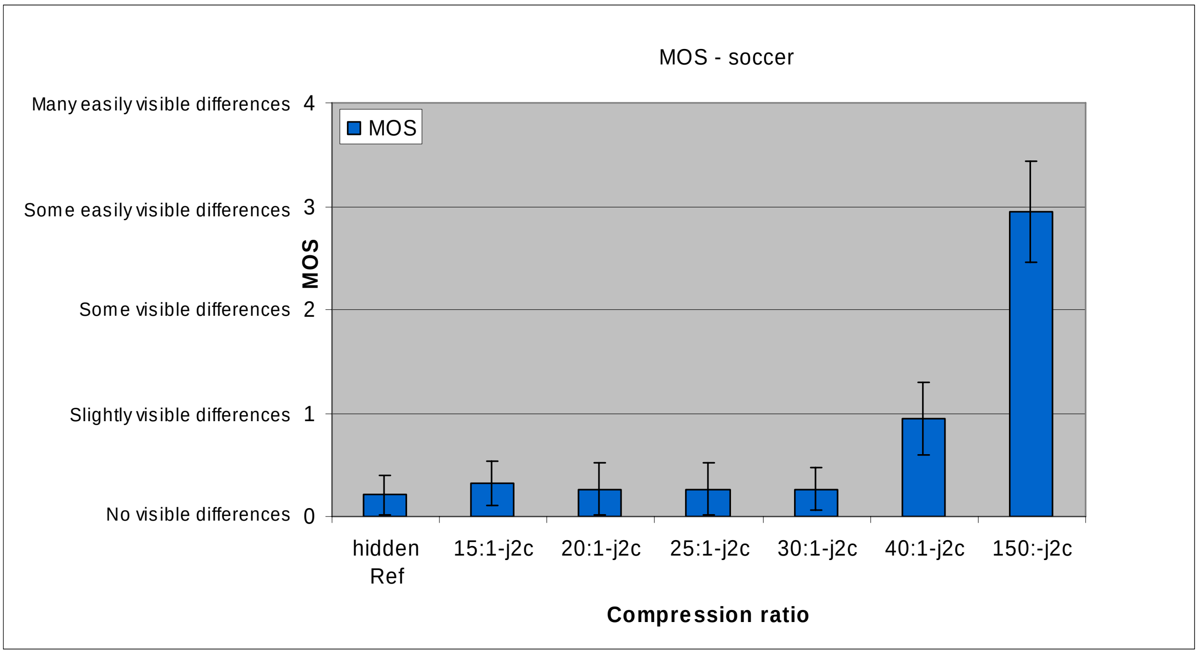
Fraunhofer IIS



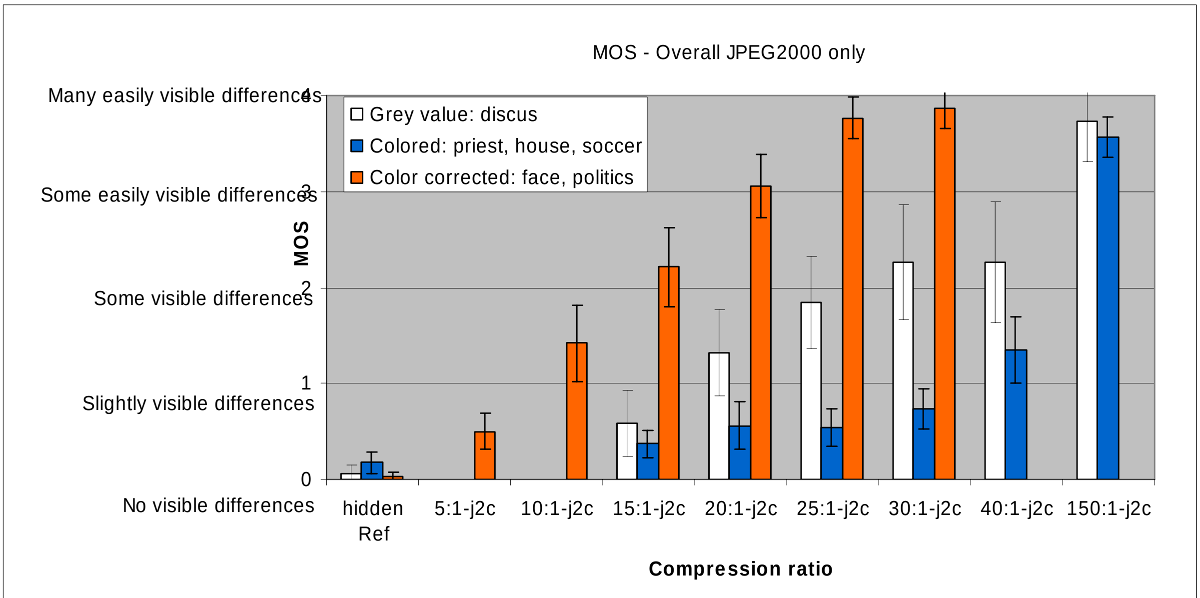



MOS - priest
Many easily visible differences 4
Some easily visible differences 3
Some visible differences 2
Slightly visible differences 1
No visible differences 0
MOS
MOS
hidden Ref
15:1-j2c
20:1-j2c
25:1-j2c
30:1-j2c
40:1-j2c
150:1-j2c
Compression ratio

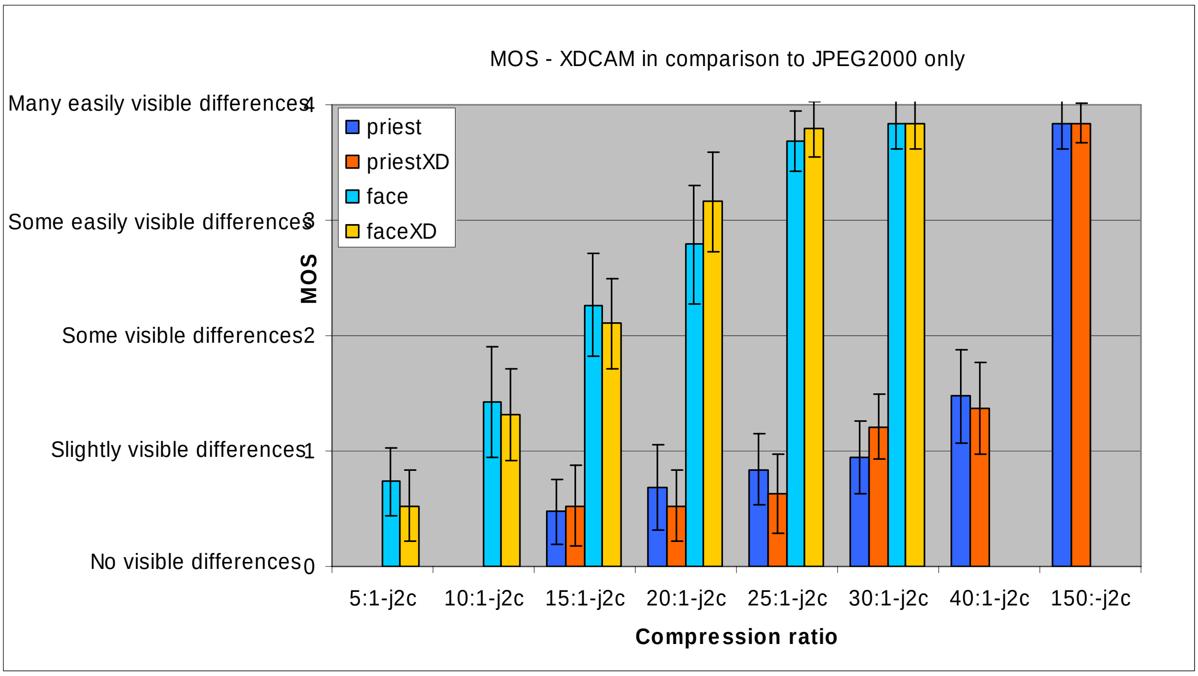

MOS - priestXD
Many easily visible differences 4
Some easily visible differences 3
Some visible differences 2
Slightly visible differences 1
No visible differences 0
MOS
MOS
hidden Ref
15:1-j2c
20:1-j2c
25:1-j2c
30:1-j2c
40:1-j2c
150:-j2c
Compression ratio



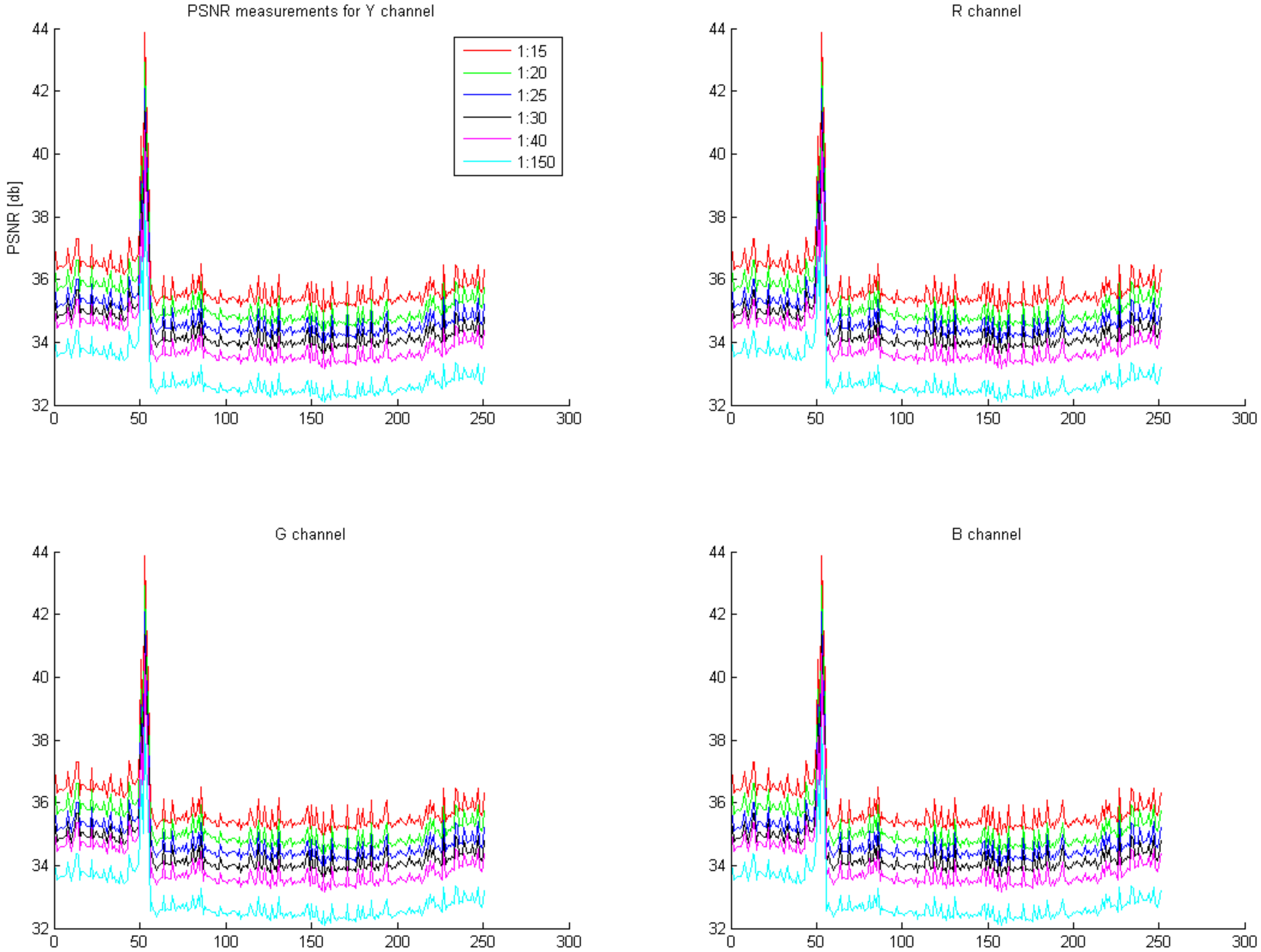
MOS - soccer
Many easily visible differences 4
Some easily visible differences 3
Some visible differences 2
Slightly visible differences 1
No visible differences 0
MOS
hidden Ref
15:1-j2c
20:1-j2c
25:1-j2c
30:1-j2c
40:1-j2c
150:-j2c
Compression ratio



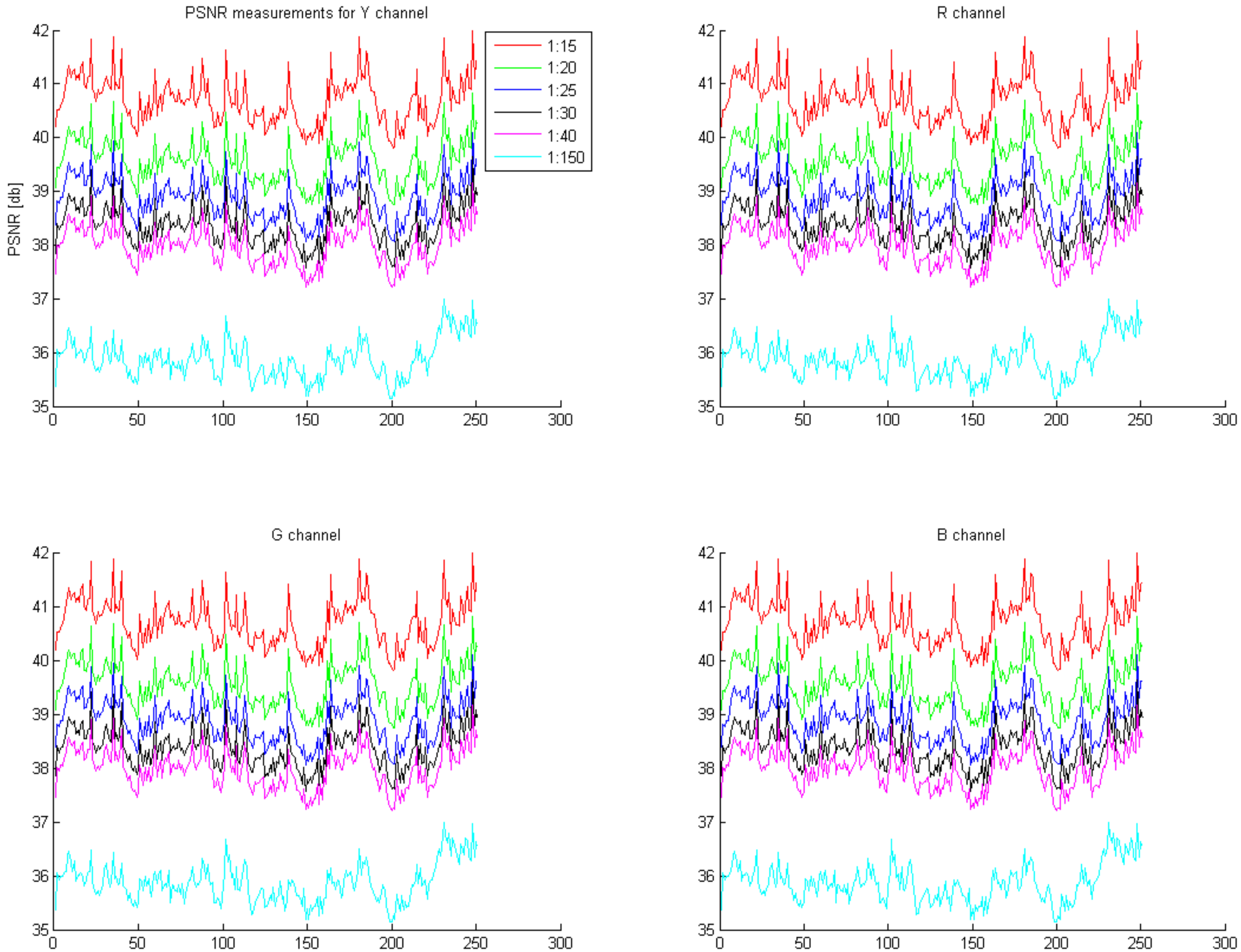
MOS - Overall JPEG2000 only
Many easily visible differences
Some easily visible differences
Some visible differences
Slightly visible differences
No visible differences
MOS
Grey value: discus
Colored: priest, house, soccer
Color corrected: face, politics
hidden Ref
5:1-j2c
10:1-j2c
15:1-j2c
20:1-j2c
25:1-j2c
30:1-j2c
40:1-j2c
150:1-j2c
Compression ratio


MOS - XDCAM in comparison to JPEG2000 only
Many easily visible differences 4
Some easily visible differences 3
Some visible differences 2
Slightly visible differences 1
No visible differences 0
MOS
priest
priestXD
face
faceXD
5:1-j2c
10:1-j2c
15:1-j2c
20:1-j2c
25:1-j2c
30:1-j2c
40:1-j2c
150:-j2c
Compression ratio

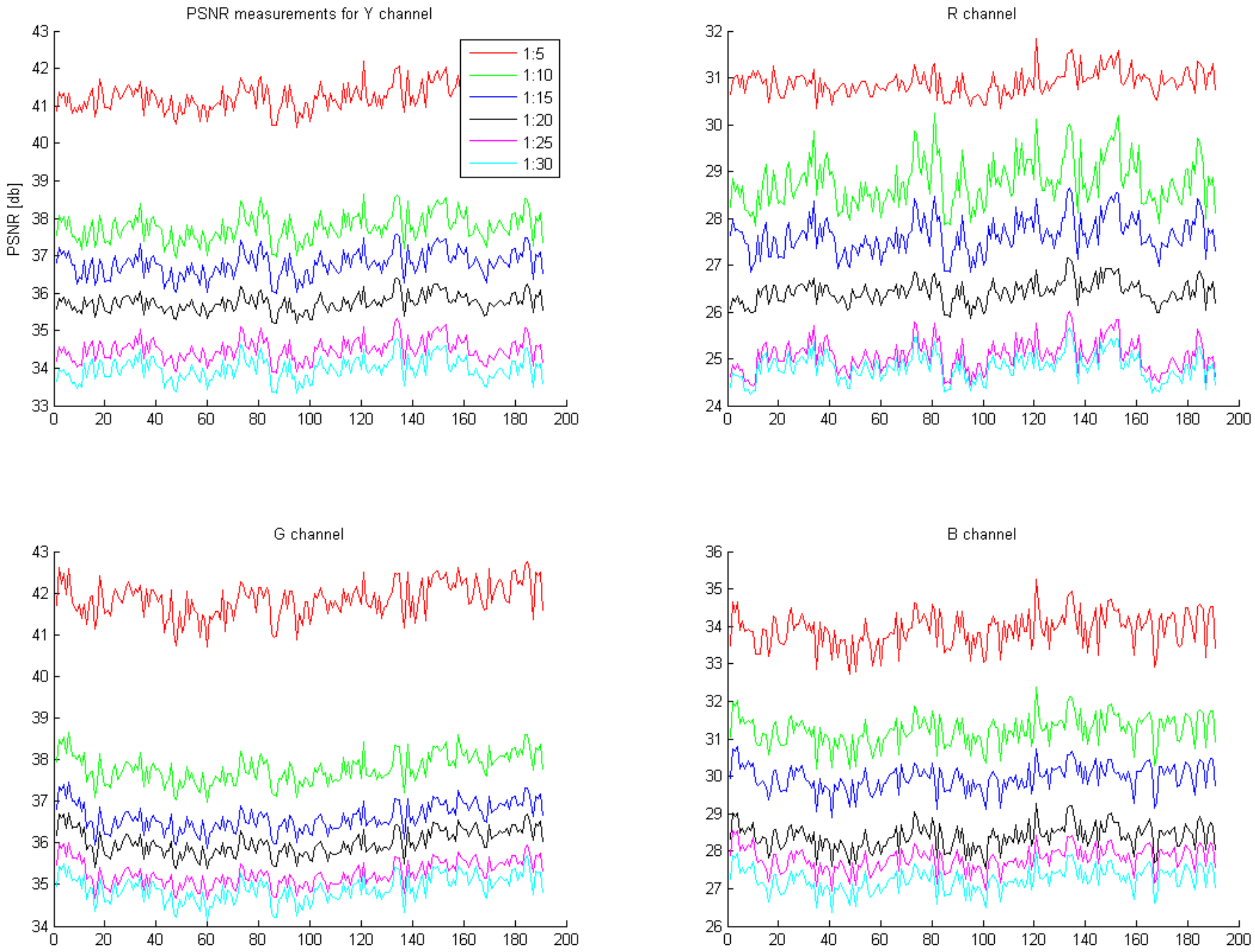
9.2 PSNR S&V_2656_Indoor_Volleyball



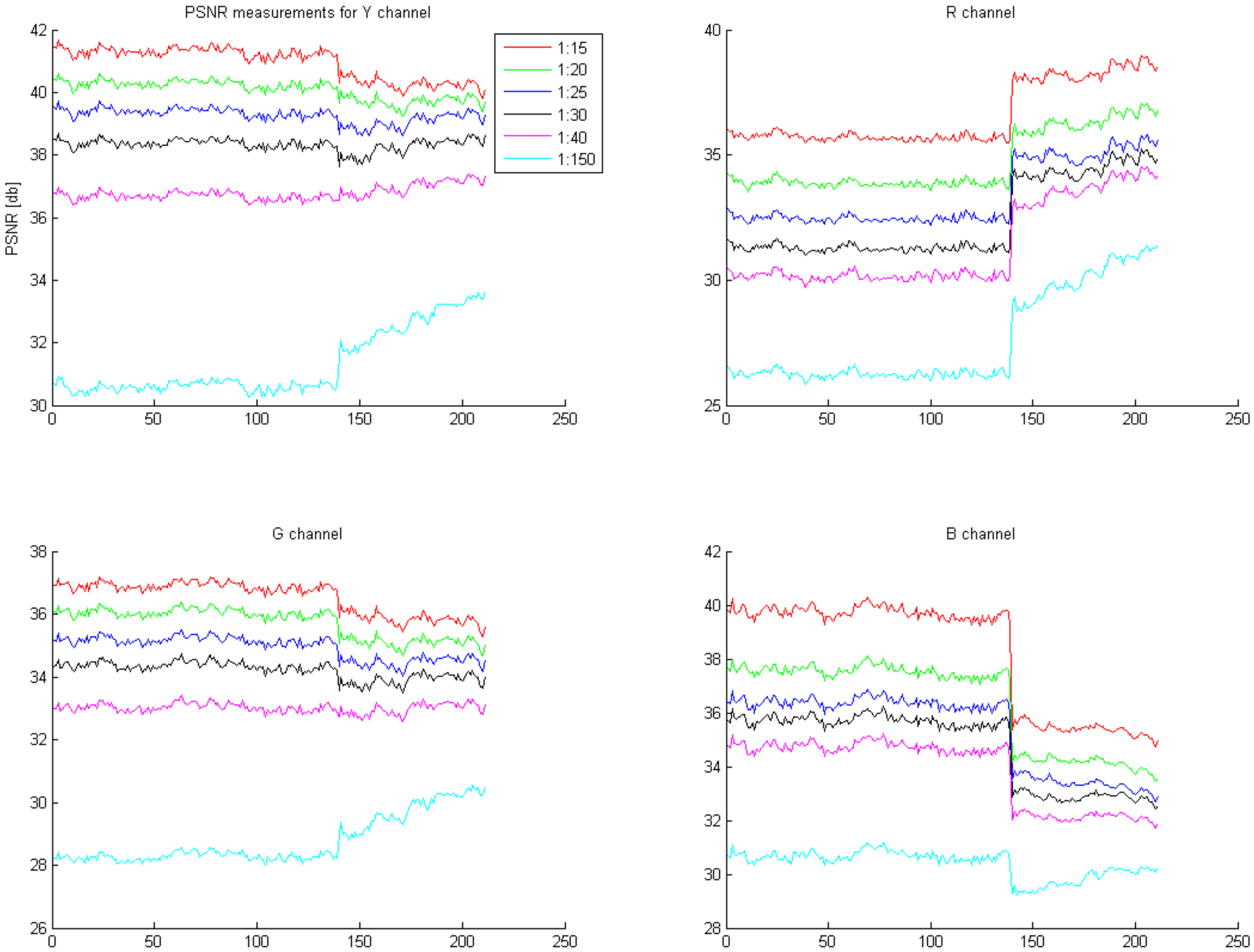
9.3 PSNR Discus



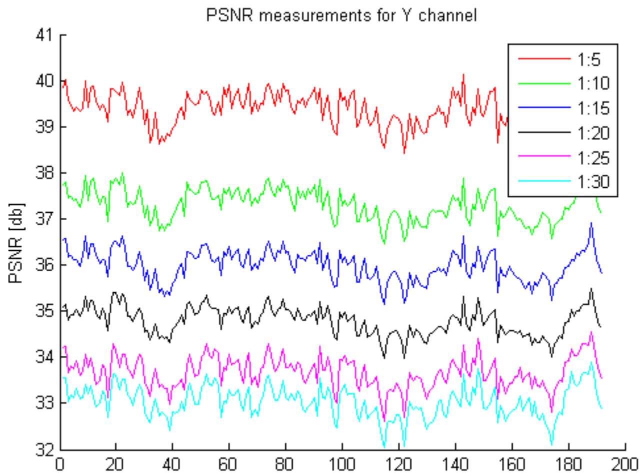
9.4 PSNR face

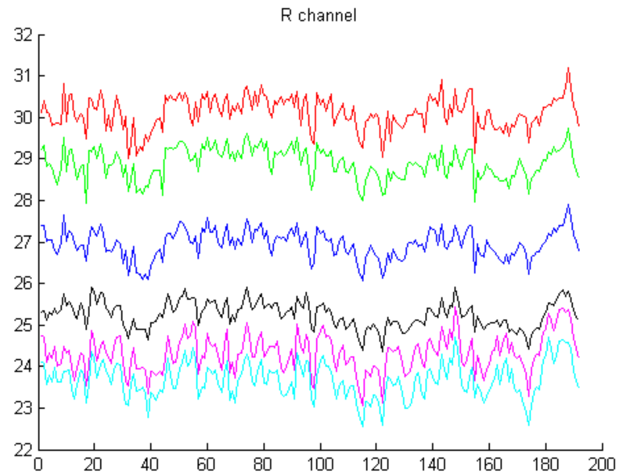

PSNR measurements for Y channel
PSNR [db]
1:5
1:10
1:15
1:20
1:25
1:30
R channel
G channel
B channel

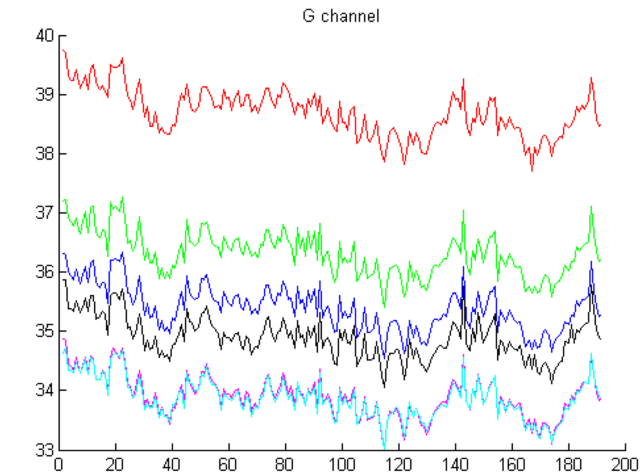
9.5 PSNR house

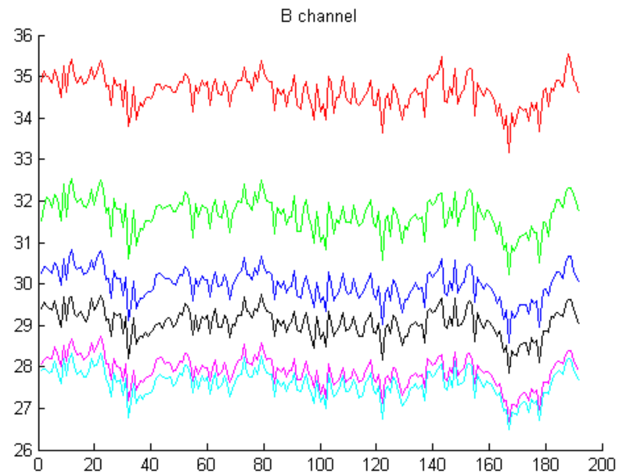


9.6 PSNR Politics

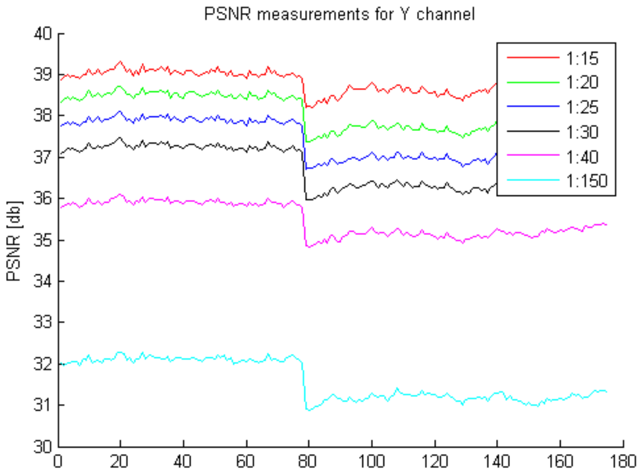


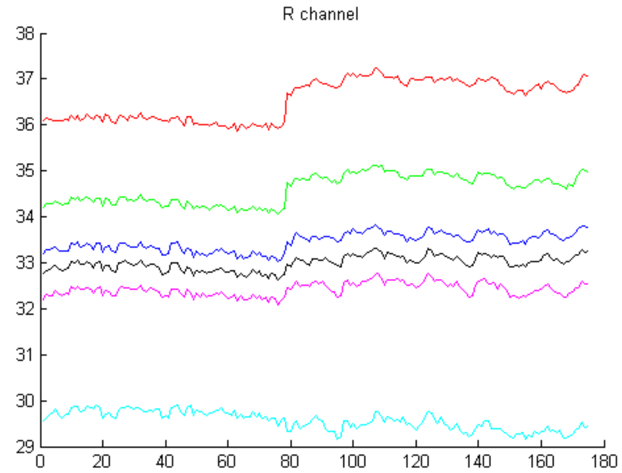


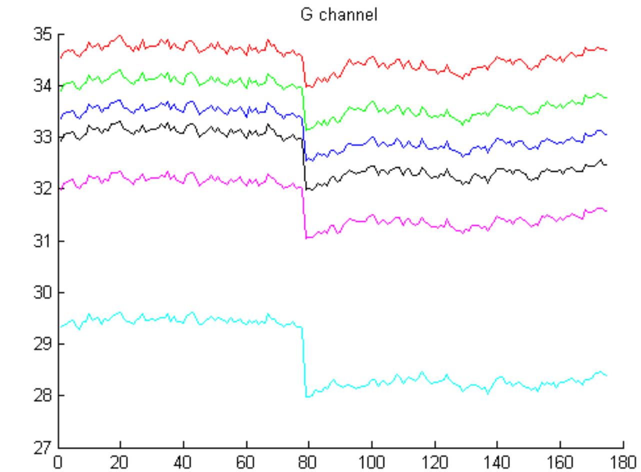


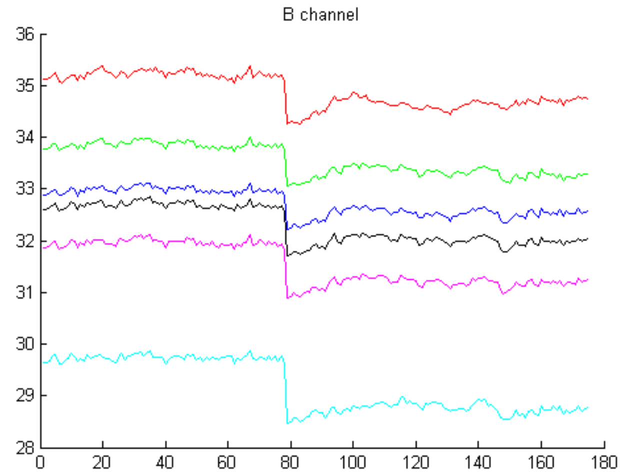


9.7 PSNR Priest

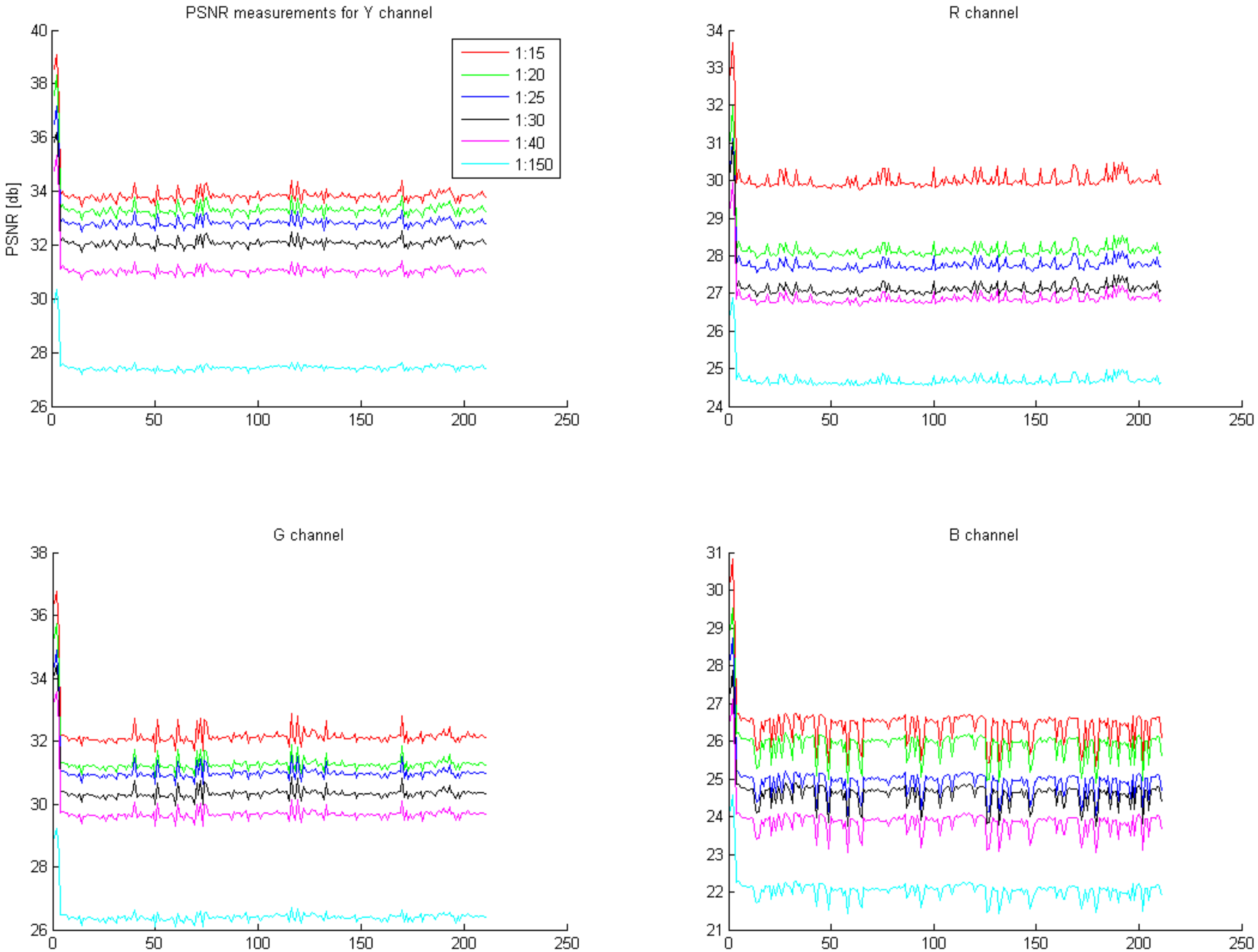








9.8 PSNR Soccer



Fraunhofer IIS
Abteilung Bildsensorik
Am Wolfsmantel 33
D-91058 Erlangen

Dr.-Ing. Siegfried Fößel
Tel. +49-9131-776-5140
Email siegfried.foessel@iis.fraunhofer.de

Dipl.-Ing. Arne Nowak
Tel. +49-9131-776-5154
Email arne.nowak@iis.fraunhofer.de