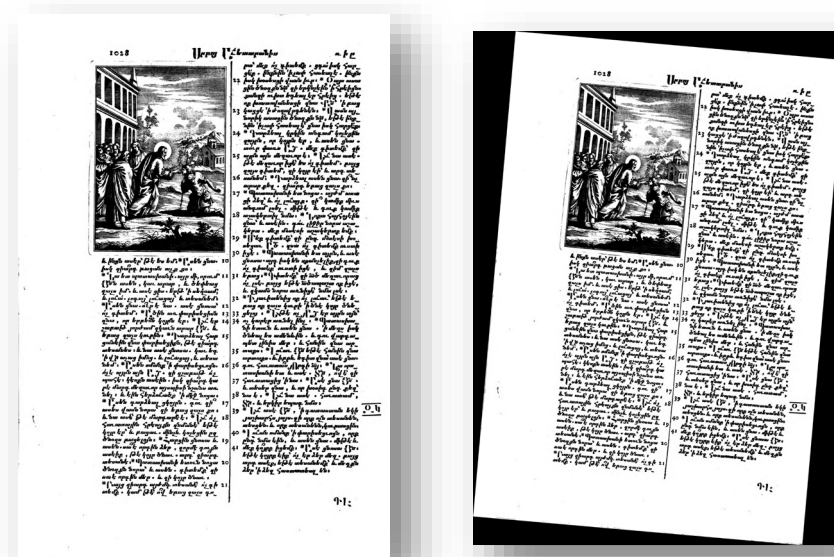


Towards Objective Quality Assessment in Digital Collections

Alexander Schindler

Matchbox

- Supports automated Quality Assurance in Digital Image Collections
- Image processing
- Main Use-Case: Find Duplicates
- Near Duplicate Search:



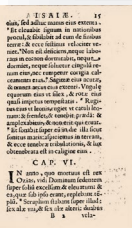
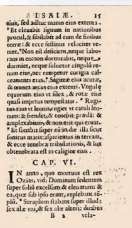
Are same Images?

Have duplicated content?

Which one is better?

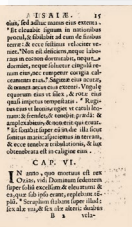
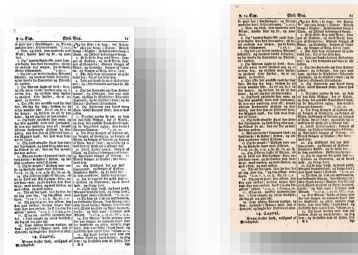
Comparing two Collections

- Find corresponding pages in both collections



Comparing two Collections

- Which collection has higher quality



Good Quality of Digital Collections

Digitization for Public Access

Digitization for Preservation

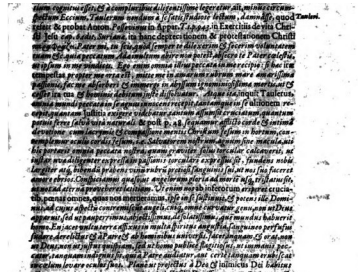
Completeness

- Missing items in collection
- Missing content in one item

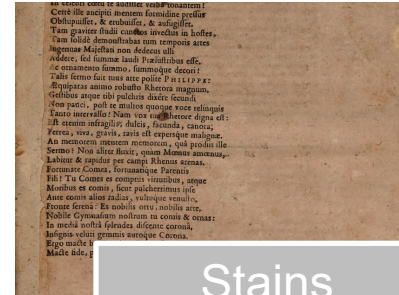


Readability

- Bleed through
- Stains
- Page curl
- Broken characters
- Low contrast



Bleed through



Stains

Post-Processing

also in places of customary communication, where bridges ought to be built, or ferry-boats stationed for the accommodation of the inhabitants, a failure to do so in either case, shall be punished with 40 blows*.

No clause.

* The original work, it is proper to notice, concludes with two supplemental books, containing fifty-seven articles each, relating, however, almost wholly to the Tartar subjects of the empire; these books therefore, upon the same principle that has been acted upon in respect to the other supplemental parts of the work, has been omitted in the present translation.

also in places of customary communication, where bridges ought to be built, or ferry-boats stationed for the accommodation of the inhabitants, a failure to do so in either case, shall be punished with 40 blows*.

No clause.

* The original work, it is proper to notice, concludes with two supplemental books, containing fifty-seven articles each, relating, however, almost wholly to the Tartar subjects of the empire; these books therefore, upon the same principle that has been acted upon in respect to the other supplemental parts of the work, has been omitted in the present translation.

- How does it affect photographs and pictures?
- How many additional information is lost?

Quality Assessment in Digital Collections

- State of the Art / Best Practices
 - **Visual Domain**
 - Peak Signal to Noise Ratio (PSNR)
 - Mean Squared Error (MSE)
 - **Audio Domain**
 - Peak Signal to Noise Ratio (PSNR)
 - Zero Crossing Rate (ZCR)

- Suggested Approach: **Perceptual Similarity Measures**
 - **Visual Domain**
 - Structural Similarity (SSIM)
 - **Audio Domain**
 - Perceptual Evaluation of Audio Quality (PEAQ)
 - Perceptual Evaluation of Speech Quality (PESQ)

Sacrificing Quality

When are compression artifacts perceived disturbant?

Loss-less



50%



75%



95%



Perceived Quality Loss

How do similarity measures relate to preservation needs?

SSIM = 0,95



SSIM = 0,81



SSIM = 0,63



Initial Quality

- How is the initial quality of the collection?
- Is it acceptable for preservation?



Open Research Challenges

- Estimating the initial quality of a collection
 - Missing content
 - Perceived quality
- Estimating the quality difference between two collections
 - E.g., quality degradation after migration to lossy formats
- Applying perceptual similarity measures (PSM)
 - Missing correlations between PSM and digital preservation

AIT Austrian Institute of Technology

Alexander Schindler

alexander.schindler@ait.ac.at