Influence of study design in psycho-visual image quality evaluation: Why a specific clinical task?

Ljiljana Platiša,¹ Leen Van Brantegem,² Asli Kumcu,¹ Cédric Marchessoux,³ Ewout Vansteenkiste,¹ and Wilfried Philips¹

¹TELIN-IPI-iMinds, Ghent University, Ghent, Belgium
²Faculty of Veterinary Medicine, Ghent University, Merelbeke, Belgium
³Healthcare Division, Barco N.V., Kortrijk, Belgium
CIMI project (ICON 2010-2011)

Color Imaging & Multidimensional Image processing in medical applications
Image data

Tissue 1 (Tiss1)
Gastric fundic glands of a dog

Tissue 2 (Tiss2)
Liver of a foal

Tissue 3 (Tiss3)
Gastric fundic glands of a dog

(1200 x 750 pixels)
Image data

12 reference images
  • 1200 x 750 crops of pathological slides stained with H&E

60 artificially manipulated images
  • One type of manipulation at a time
  • 5 (of 9) most dominant effects selected from pilot study
  • Degradation adjusted using High Dynamic Range Visible Difference Predictor (HDR-VDP)*

Image quality (IQ) research questions

Q1. What are the effects of image degradation & manipulation on the perceived IQ?

Q2. Are those effects influenced by the expertise of the viewer? How?

Q3. Which of the degradations/manipulations influence diagnostic performance? How?
Study A. Technical setup
72 images (cases) per test subject (reader)
Psycho-visual experiment

Readers’ task (single stimulus experiment)

• 5 rating questions per image
• 6 point rating scales

Readers (test subjects)

• Pathology experts, PE … Practicing diagnostic pathologists (#6)
• Pathology students, PS … Veterinary students (#7)
• Imaging experts, IE … Researchers in digital image processing (#11)

Data analysis

• Median opinion scores
• Kruskal-Wallis non-parametric one-way ANOVA (α=0.05)
Image viewing conditions

3MP medical color LCD display

Proprietary web-based interface for displaying images and collecting subject’s responses

No zoom in/out or window/level adjustment allowed

No time limitation

Controlled viewing environment
Single stimulus experiment

The task

You will be interpreting a total of 72 images obtained by Haematoxylin and Eosin (H&E) staining protocol. For each of these images, you are asked to answer the following five questions:

Q1: How would you judge overall quality of the image?
   Very low quality (0) | (1) | (2) | (3) | (4) | (5) Very high quality

Q2: How would you judge the level of noise?
   Very disturbing (0) | (1) | (2) | (3) | (4) | (5) Not disturbing at all

Q3: How would you judge the level of blur?
   Very disturbing (0) | (1) | (2) | (3) | (4) | (5) Not disturbing at all

Q4: How would you judge the contrast?
   Very poor (0) | (1) | (2) | (3) | (4) | (5) Very good

Q5: How would you judge the color saturation?
   Very poor (0) | (1) | (2) | (3) | (4) | (5) Very good
How would you judge overall quality of the image?
Very low quality
Very high quality

How would you judge the level of noise?
Very disturbing
Not disturbing at all

How would you judge the level of blur?
Very disturbing
Not disturbing at all

How would you judge the level of contrast?
Very poor
Very good

How would you judge the color saturation?
Very poor
Very good
Conclusions

- **Pathology Experts** seemed less rigorous in rating image quality (less disturbed by image manipulations) compared to **Imaging Experts** except for Gamma, which was the reverse.

- **Pathology Students** seemed to take a conservative approach.

- Subject expertise could be a critical factor of subjective IQ assessment.

*Platisa et al., “Psycho-visual evaluation of image quality attributes in digital pathology slides viewed on a medical color LCD display,” in Proc. SPIE MI 2013*
Study B. Clinical setup
Image quality (IQ) research questions

Q1. What are the effects of image degradation & manipulation on the perceived IQ?

Q2. Are those effects influenced by the expertise of the viewer? How?

Q3. Which of the degradations/manipulations influence diagnostic performance? How?
1/72
Psycho-visual experiment

Readers’ task
• Lesion detection & localization
• Possible multiple lesions per case
• Rating scale continuous [0, 100]

Image data (72 images per reader)
• 12 cases (5 normal, 7 abnormal)
• 6 treatments (1 REF + 5 manipulations)

Readers (test subjects)
• 6 practicing diagnostic pathologists (PEs)

Data analysis
• Median opinion scores
• JAFROC (not shown here)
Image viewing conditions

3MP medical color LCD display

Proprietary web-based interface for displaying images and collecting subject’s responses

No zoom in/out or window/level adjustment allowed

No time limitation

Controlled viewing environment
Results. Perceived overall image quality (pIQ)

Median opinion scores for each type of image manipulation.

Each box plot indicates
- the median
- the middle fifty (IQR)
- the 1.5 IQR interval (whiskers)

Overall IQ was perceived lower for JPG compressed images.
Why a specific clinical task?
Perceived overall image quality (pIQ) rated by PE

Study A

Study B

Noise
Blur
Contrast
Color

Technical setup

Clinical setup

75%
Concluding remarks

Overall IQ ratings differ between Study A and Study B

Possible reasons

• Technical setup might have caused confusion for pathologists, forcing them to use IQ criteria which they are not comfortable with

• Focusing on specific types of technical image artifacts might have been distracting for pathologists, as they are little familiar with those
Questions?

Thank you!