Evaluation of Image Quality of MRI Data for Brain Tumor Surgery

Frank Heckel\textsuperscript{a,b}, Felix Arlt\textsuperscript{c}, Benjamin Geisler\textsuperscript{a}, Stephan Zidowitz\textsuperscript{a}, Thomas Neumuth\textsuperscript{b}

\textsuperscript{a}Fraunhofer Institute for Medical Image Computing MEVIS, Bremen, Germany; \textsuperscript{b}Innovation Center Computer Assisted Surgery, Leipzig, Germany; \textsuperscript{c}Leipzig University Hospital, Department of Neurosurgery, Leipzig, Germany

ABSTRACT

3D medical images are important components of modern medicine. Their usefulness for the physician depends on their quality, though. Only high-quality images allow accurate and reproducible diagnosis and appropriate support during treatment. We have analyzed 202 MRI images for brain tumor surgery in a retrospective study. Both an experienced neurosurgeon and an experienced neuroradiologist rated each available image with respect to its role in the clinical workflow, its suitability for this specific role, various image quality characteristics, and imaging artifacts. Our results show that MRI data acquired for brain tumor surgery does not always fulfill the required quality standards and that there is a significant disagreement between the surgeon and the radiologist, with the surgeon being more critical. Noise, resolution, as well as the coverage of anatomical structures were the most important criteria for the surgeon, while the radiologist was mainly disturbed by motion artifacts.

Keywords: image quality assessment, magnetic resonance imaging, brain tumor surgery

1. INTRODUCTION

3D medical images acquired by magnetic resonance imaging (MRI) for example are ubiquitous in today’s surgery. They are used in each stage of the clinical workflow for diagnosis, therapy planning, image-guided intervention (e.g., navigated and minimally invasive surgery), and therapy success monitoring. The mere technical availability of the data, however, does not necessarily ensure that the required information is provided. The quality of an image, for instance, is one of its most important relevance criteria.\textsuperscript{1} The requirements on an image’s quality depend on the specific clinical use-case. If they are not fulfilled, the data cannot be used, it is of limited value, or it could even lead to inaccurate diagnosis.\textsuperscript{2,3} Particularly computer-assisted systems, such as automated image analysis or intra-operative navigation, are known to need high quality images in order to provide accurate results.\textsuperscript{4}

This study evaluates the suitability of MRI data in the context of brain tumor surgery. In brain tumor surgery, the extent of surgical resection is linked to patient outcome. On the one hand, all tumor cells need to be removed in order to avoid recurrence. On the other hand, healthy brain tissue and critical risk structures, such as the cerebral cortex, motor, somatosensory, language, and visual function areas, have to be protected from damage.\textsuperscript{5,6} Therefore, high quality images are particularly important for accurate therapy planning and intra-operative navigation. We have analyzed the influence of image quality on the suitability of the images for their specific purpose as well as the agreement between the surgeon and the radiologist.

2. MATERIAL AND METHODS

We have analyzed MRI data from 24 patients in a retrospective study. For each patient, different images (sequences and reconstructions) were available, resulting in a total number of 202 datasets. The data was acquired at Leipzig University Hospital for brain tumor surgery at the Department of Neurosurgery. Both an experienced neurosurgeon and an experienced neuroradiologist have annotated each image in a visual grading analysis using an absolute grading on an adjectival scale.\textsuperscript{7}

The annotation has been done using an intuitive, flexible tool that has been developed using MeVisLab* (see Fig. 1). The ratings and comments as well as meta-information for each image have been stored in an XML file

\*http://www.mevislab.de
Figure 1: The image quality annotation tool that has been used for visual grading analysis.

for further analysis. The following information has been collected: the purpose (“role”) of the image (diagnosis, therapy planning, intra-operative navigation) and its overall suitability for this specific role; the coverage of the overall anatomical region (i.e., the brain; we refer to this as “data coverage”) and the target region (i.e., the specific tumor); general image quality characteristics (signal-to-noise ratio, resolution, homogeneity); and MRI-specific imaging artifacts (motion, ghosting, susceptibility). Images could also be rated as irrelevant. This was done for instance for functional MRI sequences, diffusion weighted images, and sequences without contrast agent, which are not commonly used by surgeons in any of the workflow steps. In addition, the clinicians were allowed to skip images if they were not able to rate them appropriately.

The ratings have been statistically analyzed. Specific quality aspects have also been correlated with the overall suitability. Spearman’s rank correlation coefficient has been used to measure correlation. Finally, the agreement between the surgeon and the radiologist has been analyzed. The overall agreement is computed by

\[ d = \frac{1}{n} \sum_i ||a_i - b_i||, \]

with \(a_i\) and \(b_i\) being the annotation by reader 1 and reader 2 mapped to integers ([0,4] for coverage, signal-to-noise ratio, resolution, and homogeneity; [0,3] for MRI-specific artifacts). Wilcoxon’s signed-rank test has been used to measure the significance of the agreement. The significance level was set to \(\alpha = 0.05\).
3. RESULTS

Figure 2 shows an overview on the different roles of the images and their overall suitability as rated by the neurosurgeon and the neuroradiologist. Obviously, only the surgeon can judge whether a specific image has been acquired for intra-operative navigation, while the radiologist judges the images mostly from a diagnostic point of view. For a surprisingly high number of images the overall suitability has been rated as bad or unusable, particularly by the surgeon, who also seems to be more critical concerning the suitability of the images. The radiologist skipped less images as he also considered images without contrast agent.

From the surgeon’s point of view, the coverage of both the whole anatomical region and the target region was not always sufficient, i.e., important parts were missing in the images (see Fig. 3). This correlates well with the suitability rating. A relatively high number of images shows a bad resolution, signal-to-noise ratio, or homogeneity (see Fig. 4). Again, the surgeon gives worse ratings compared to the radiologist. For the surgeon, the general quality characteristics correlate well with the suitability. For the radiologist, the only parameter that shows a good correlation with the suitability is the presence of motion artifacts (see Fig. 5). Overall, the radiologist is more critical with respect to imaging artifacts compared to the surgeon.

For suitability, coverage, and the general quality characteristics, the surgeon and the radiologist show a rather bad agreement on average (see Figs. 2 - 5). Only for imaging artifacts the agreement is relatively good. Note, however, that disturbing imaging artifacts are rather seldom and most images artifacts were rated as “not present”. The result for the agreement are significant for all evaluated parameters. Note that the evaluation of agreement only considers cases that have been rated by both readers.
4. DISCUSSION

The assessment of the quality of natural and compressed images as well as videos has been a topic of intense research in the past decades. In this context, quality is often defined in a subjective manner as perceived by a human observer. Consequently, there have been many attempts to model human observers for automatic quality assessment with respect to specific tasks. In medical imaging, image quality assessment is an emerging field where most research focuses on image quality control during the acquisition. In recent years, however, researchers also started modeling human observers. In this study, we have analyzed the suitability of medical images for their specific purpose and the influence of image quality on their overall suitability from a clinical point of view based on a specific clinical use-case. In various interviews, physicians stated that “images are almost always good and their quality is not a problem in practice”. In contrast to this subjective impression, our results indicate that there is a relevant number of images that do not meet the required quality standards for the examined clinical use-case. The results also suggest that those standards show a significant gap between radiologists and surgeons, with surgeons being more critical. Coverage and general image quality characteristics, such as resolution and signal-to-noise ratio for example, seem to be the most relevant aspects from the surgeon’s point of view.

The main limitation of this study is the small number of clinical participants with only one representative from each discipline. In addition, the images have been acquired from only one site, which is a university hospital with modern equipment and high standards. As absolute grading studies are known to have a low reproducibility, more readers with different levels of experience have to be included in future studies. Alternatively, relative grading might be used.

The results of our study can act as baseline data for automatic quality assessment tools for neurosurgical interventions. Building automatic image quality assessment algorithms requires either perfect reference images to which new images are compared (full-reference methods) or it is based on the detection of specific types of
distortion, such as blurring or noise (no-reference methods). As reference images are not available for most images acquired in medical imaging, no-reference quality assessment methods are required.

5. CONCLUSION AND FUTURE WORK
This study analyzed the suitability MRI data in the context of brain tumor surgery from a clinical point of view. We have evaluated the influence of image quality and imaging artifacts on the overall suitability and we have compared the results of a surgeon to the results of a radiologist. Our results show that MRI data acquired for brain tumor surgery does not always fulfill the required quality standards and that there is a significant disagreement between the surgeon and the radiologist.
Figure 5: Influence of the most relevant MRI-specific artifacts on the suitability and agreement between disciplines. $\rho$ refers to Spearman’s correlation coefficient.

Further studies with different clinical use-cases, more data from divergent radiology practices, and more experts are necessary to generalize our results. Our future work will also focus on no-reference automatic image quality assessment in order to automatically detect unsuitable image as early as possible in the clinical workflow, as it has previously been proposed by Woodard & Carley-Spencer, Schwaab et al., and Küstern et al., for example.

ACKNOWLEDGMENTS

This work was funded by the German Ministry of Education and Research (BMBF) as part of the project OR.NET (reference number: 16KT1232).
REFERENCES


