Evaluation of Image Quality in Medical Volume Visualization: The State of the Art

Andreas Pommert and Karl Heinz Höhne

Institute of Mathematics and Computer Science in Medicine (IMDM)
University Hospital Hamburg-Eppendorf, 20251 Hamburg, Germany
{pommert,hoehne}@uke.uni-hamburg.de

Abstract. For applications of volume visualization in medicine, it is important to assure that the 3D images show the true anatomical situation, or at least to know about their limitations. In this paper, various methods for evaluation of image quality are reviewed. They are classified based on the fundamental terms of intelligibility and fidelity, and discussed with respect to the question what clues they provide on how to choose parameters, or improve imaging and visualization procedures.

1   Introduction

Volume visualization (VV) of tomographic volume data, as obtained in computer tomography (CT) or magnetic resonance imaging (MRI), is an important aid for diagnosis, treatment planning, surgery rehearsal, education, and research (fig. 1). For clinical applications, it is of course important to assure that the 3D images really show the true anatomical situation, or at least to know about their limitations. Unfortunately, the resulting images are depending on a large number of parameters, including pixel size, filter kernel of the scanner, slice distance and thickness, interpolation method, threshold (or other segmentation parameters), and gradient operators. Variation of these parameters may result in very different images.
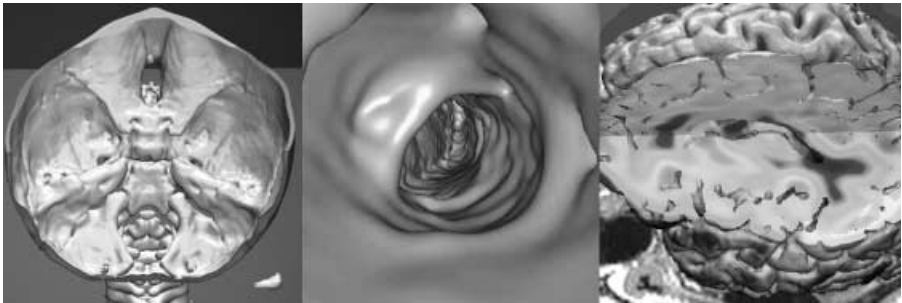

Fig. 1. Examples of volume visualization in craniofacial surgery (left, from CT), virtual colonoscopy (middle, from MRI) and psychiatry research (right, from MRI/PET). But how good are these images?

In medical image computing, validation of image quality is a major concern, as was pointed out in recent papers [13], panels [16], and dedicated meetings [7]. However, compared to the large number of papers dealing with methods and applications of VV, only few papers are focusing on the resulting image quality. Nevertheless, this field is characterized by a multitude of definitions and measures of image quality, goals, investigation methods, and considered processing steps, such that the different approaches are often difficult to compare.

In this paper, the state of the art for the evaluation of image quality in medical VV is reviewed. A classification of methods is developed, and methods are discussed with respect to the question what clues they provide on how to choose parameters, or improve imaging and visualization procedures.

1.1   Aspects of Image Quality

When is an image good or bad? A straightforward definition of image quality is based on the question: How well does an image communicate an information required by an observer? This is called the *intelligibility* of the image [34]. For example, an image used in diagnostic imaging is good if it enables an observer to make the right diagnosis (*diagnostic image quality*). A more technical definition of image quality relates to the question: How much does an image deviate from an ideal image of the scene? This is called the *fidelity* of the image (*technical image quality*) [34]. Intelligibility and fidelity are determined by comparing diagnosis or 3D image to an otherwise determined *ground truth* (figure 2). Both aspects of image quality are discussed in the following.
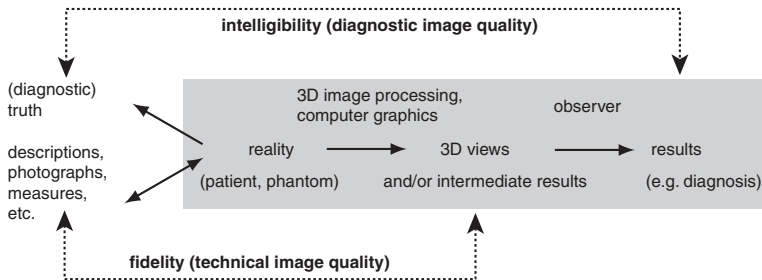

Fig. 2. Evaluation of intelligibility (diagnostic image quality) and fidelity (technical image quality) by comparison to a reference (dashed lines).

2   Intelligibility

In medical imaging, the intelligibility of an image relates mostly to a diagnostic task (fig. 2). Subjective studies include the comparison of different imaging and visualization techniques, such as CT and volume rendering, for applications e.g.

in craniofacial or ENT surgery [1,35]. These papers emphasize the better understanding of spatial relations provided by VV, but give little further insights, since the true anatomical situation is generally not known. The same limitation applies to various papers comparing different processing parameters [26].

Objective investigations of intelligibility for diagnostic tasks are based on blind studies. Image quality can thus be measured in terms of diagnostic accuracy, sensitivity, specifity, or ROC-index. With respect to VV, studies were pioneered by Vannier et al [38], who compared various imaging modalities and visualization techniques (e.g. x-ray, CT, 3D depth shaded, 3D gradient shaded, 3D volume shaded) for various tasks, including diagnosis of craniosynostosis and fractures [14,39]. In most cases, VV compared favourably. Furthermore, it could be shown that VV accelerated the speed of establishing a diagnosis, and improved localization accuracy of the findings.

For screening applications, a high sensitivity is most critical. The sensitivity of virtual colonoscopy, based on CT data, is investigated e.g. by [15,25]. Both studies show a sensitivity similar to that of a real colonoscopy.

In the clinical literature, definition of image quality in terms of intelligibility is generally accepted [6]. However, from a more technical point of view, this definition has some problems. First, results strongly depend on factors outside the image, such as the observer's experience and the task. Second, no measures are at hand for application areas other than diagnostics, such as therapy planning or surgical planning, for which VV methods are most used. Third, observer studies are extremely costly. For certain tasks such as the detection of small signals in nuclear medicine, mathematical model observers were developed [2]. However, the much more complex visual and cognitive tasks involved in understanding perspective 3D images are only little understood so far. Fourth, results of such studies give little or no clues on how to choose parameters, or improve imaging and visualization procedures. An exhaustive testing of all possible settings is hardly feasible, due to the high costs.

3   Image Fidelity

In order to avoid these problems, the more technical definition of image fidelity is used (fig. 2). In a simple case, VV images created using different parameters are compared, without precise knowledge of the anatomical situation. This approach is found in many papers, e.g. [27].

To get at least comparable results, standardized datasets are used, which my be distributed over the Internet [31]. For more thorough investigations, experimental studies may be based on cadavers, phantoms, or simulated data, with known properties. Furthermore, algorithmic or mathematical studies may be carried out.

3.1   Experimental Studies

Anatomical Specimen. In a classic paper [19], 3D images of bone from CT are compared to photographs of the specimen. The investigation also covers

variation of slice orientation, distance, thickness, and scanner type. This way, a first description of artifacts such as pseudoforamina and stairsteps could be obtained. More detailed studies, also including measurements of distances, are presented in [10,20,32]. All these papers also describe effects of variation of a (usually small) subset of the parameter space.

Phantoms. Instead of specimen, artificial phantoms may be used, which can be designed for special purposes. In [40], a cone-shaped phantom is used to investigate step artifacts in spiral CT. In [9], a special phantom is used to investigate the visualization of a stenosis of variable size.

Simulated Data. Simulated data can even easier be adjusted to certain needs, at the cost of loosing some realism. Furthermore, if they are created from a symbolic description, a perfect reference is at hand, which can be used to create error images, showing local deviations [36].

A first question arises how the simulated data are to be created. In some papers, the data are designed to be demanding for visualization algorithms, e.g. by containing high spatial frequencies [22,24,29]. However, more realistic data are obtained by modeling the point spread function of a real tomographic scanner [36], or even the whole physics of image acquisition, as in the *MNI Brainweb* project [12].

Another question is what should be measured. In 2D medical imaging, aspects such as image resolution or signal-to-noise ratio are often used. In volume visualization, other measures such as the accuracy of surface position or surface normal vectors [24,28,36,37] seem to be more appropriate. This way, typical ranges of error, depending on the choice of parameters, could be estimated.

3.2   Algorithmic and Mathematical Studies

Image Space. In some papers, visualization algorithms are studied in detail "on paper", using simple example input. This way, it could be shown that the order of processing steps in volume rendering (classification or interpolation first) has a major influence of the obtained accuracy [43].

A step beyond such qualitative approaches are simple (usually 1D) quantitative models which cover major processing steps such as image acquisition, interpolation, and thresholding [5,33]. This way, it could be shown that the error of surface localization is well below voxel size, provided that a suitable threshold value is used. Furthermore, it could be shown that a poor threshold will likely cause visible artifacts, which can be used for further adjustment [33].

Other approaches include the investigation of the asymptotic error of interpolation functions, based on a Taylor series expansion [3,28].

Frequency Space. In signal and image processing, it is often useful to investigate the response of a system in frequency space [8]. In [3], the quality of

different interpolation filters is studied by comparing their amplitude spectra to an ideal low-pass. This concept is extended in [24], where metrics for smoothing and aliasing are introduced, yielding quantitative (but not very intuitive) descriptions. This approach can also be used for gradient filters [4].

As a major drawback, there is no representation in frequency space for non-linear operations such as thresholding, such that this kind of analysis does not cover all parts of the VV pipeline.

3.3 Predefined Error Bounds

An attractive solution with respect to image fidelity are rendering methods which guarantee the visualization results to be within certain predefined error bounds. One such approach is the *controlled precision volume rendering* [30,42]. However, the controlled precision relates only to a mathematical approximation of the volume rendering integral common to these algorithms, and says little more about the quality of visualization.

With respect to the interpolation of volume data, a new class of interpolation filters is developed in [23]. Under certain assumptions about the data (which may not be met in any case), it is shown that the intensity difference between original and reconstructed function does not exceed a predefined error. While this filter has some practical problems, including high computational costs, it is a promising first step in this direction.

4 Related Fields

There are some other fields closely related to VV which might be of interest here. For segmentation of clinical data, a method for validation without ground truth was developed which is based on a statistical analysis of the results of segmentation by several experts. Roughly, if a segmentation algorithm is within this variation, results are accepted [11]. A tool for validation is presented in [17]. MRI brain images segmented by experts are available from Harvard University [21]. Within the *Insight* project, there are also proposals to use the Visible Human data for this purpose [44].

A special situation occurs for the validation of methods for image registration. Using prepared test datasets, the true transformation can be determined using external markers, which are later removed from the test data, as was done in a well-known study [41]. A comparable situation occurs for the validation of scaling methods [18].

5 Conclusions

In this paper, we presented a brief overview of the methods published so far for an evaluation of image quality in medical volume visualization. This field turned out to be very multifaceted and complex. Nevertheless, some conclusions seem obvious:

- Methods for (objective) evaluation of image intelligibility are the method of choice from a clinical point of view, but provide little help in optimizing visualization procedures, partly due to high costs of such studies. Artificial model observers will very likely not be available in this field for some time.
- For experimental studies of image fidelity, simulated data appear promising, since they may easily be adjusted to certain needs, and provide a means to create error images, precisely showing local deviations, at a rather low cost. Essential is of course a realistic simulation of the tomographic image acquisition.
- Mathematical studies, as well as processing algorithms with predefined error bounds, are currently available for certain steps of the volume visualization pipeline only, especially for interpolation.

So far, no investigations are available which determine how to choose parameters for *all* steps of the volume visualization pipeline in order to achieve certain visualization results, e.g. with respect to size or contrast of the depicted structures. Development of such a "best practice guide" will certainly be a major task in the future.

References

1. Alder, M. E., Deahl, S. T., Matteson, S. R.: Clinical usefulness of two-dimensional reformatted and three-dimensionally rendered computerized images: Literature review and a survey of surgeons' opinions. *J. Oral Maxillofac. Surg. 53*, 4 (1995), 375–386.
2. Barrett, H. H., Yao, J., Rolland, J. P., Myers, K. J.: Model observers for assessment of image quality. *Proc. Natl. Acad. Sci. USA 90* (1993), 9758–9765.
3. Bentum, M. J., Lichtenbelt, B., Boer, M. A., Nijmeijer, A. G., Bosma, M., Smit, J.: Improving image quality of volume rendered three-dimensional medical data. In Kim, Y. (Ed.): *Medical Imaging: Image Display, Proc. SPIE 2707*. Newport Beach, CA, 1996, 32–43.
4. Bentum, M. J., Malzbender, T., Lichtenbelt, B. B.: Frequency analysis of gradient estimators in volume rendering. *IEEE Trans. Visualization Comput. Graphics 2*, 3 (1996), 242–254.
5. Bosma, M. K., Smit, J., Lobregt, S.: Iso-surface volume rendering. In Kim, Y., Mun, S. K. (Eds.): *Medical Imaging 1998: Image Display, Proc. SPIE 3335*. San Diego, CA, 1998, 10–19.
6. Bowyer, K. W.: Validation of medical image analysis techniques. In Sonka, M., Fitzpatrick, J. M. (Eds.): *Handbook of Medical Imaging*, 2, SPIE Press, Bellingham, WA, 2000, ch. 10, 567–607.
7. Bowyer, K. W., Loew, M. H., Stiehl, H. S., Viergever, M. A. (Eds.): *Methodology of Evaluation in Medical Image Computing*. Dagstuhl Seminar Report 301, Internationales Begegnungs- und Forschungszentrum für Informatik, Schloss Dagstuhl, 2001. (ISSN 0940-1121).
8. Bracewell, R. N.: *The Fourier Transform and its Applications*. 3. ed. McGraw-Hill International Editions, Singapore, 2000.
9. Brink, A. J., Lim, J. T., Mang, G., Heiken, J. P., Deyoe, A. J., Vannier, M. W.: Technical optimization of spiral CT for depiction of renal artery stenosis: In vitro analysis. *Radiology 194* (1995), 157–163.
10. Cavalcanti, M. G. P., Haller, J. W., Vannier, M. W.: Three-dimensional computed tomography landmark measurement in craniofacial surgery planning: Experimental validation in vitro. *J. Oral Maxillofac. Surg. 57* (1999), 690–694.

11. Chalana, V., Kim, Y.: A meethodology for evaluation of image segmentation algorithms on medical images. In Loew, M. H., Hanson, K. M. (Eds.): *Medical Imaging 1996: Image Processing, Proc. SPIE 2710.* Newport Beach, CA, 1996, 178–189.

12. Collins, D. L., Zijdenbos, A. P., Kollokian, V., Sled, J. G., Kabani, N. J., Holmes, C. J., Evans, A. C.: Design and construction of a realistic digital brain phantom. *IEEE Trans. Med. Imaging 17*, 3 (1998), 463–468.

13. Duncan, J. S., Ayache, N.: Medical image analysis: Progress over two decades and the challenges ahead. *IEEE Trans. Pattern Anal. Machine Intell. 22*, 1 (2000), 85–105.

14. Fox, L. A., Vannier, M. W., West, O. C., Wilson, A. J., Baran, G. A., Pilgram, T. K.: Diagnostic performance of CT, MPR, and 3DCT imaging in maxillofacial trauma. *Comput. Med. Imaging Graph. 19*, 5 (1996), 385–395.

15. Garry, J. L., Reed, J. E., Johnson, C. D.: Performance of computed tomographic colonography improved by total quality management techniques. In Chen, C.-T., Clough, A. V. (Eds.): *Medical Imaging 2000: Physiology and Function from Multidimensional Images, Proc. SPIE 3978.* San Diego, CA, 2000, 183–194.

16. Gee, J. C., Haralick, R. M., Clarke, L. P., Fitzpatrick, J. M., Haynor, D. R., Ramesh, V., Viergever, M. A.: Performance evaluation of medical image processing algorithms. In Hanson, K. M. (Ed.): *Medical Imaging 2000: Image Processing, Proc. SPIE 3979.* San Diego, CA, 2000.

17. Gerig, G., Jornier, M., Chakos, M.: Valmet: A new validation tool for assessing and improving 3D object segmentation. In Niessen, W. J., Viergever, M. A. (Eds.): *Medical Image Computing and Computer-Assisted Intervention, Proc. MICCAI 2001*, Lecture Notes in Computer Science 2208, Springer-Verlag, Berlin, 2001, 516–523.

18. Grevera, G. J., Udupa, J. K.: An objective comparison of 3-D image interpolation methods. *IEEE Trans. Med. Imaging 17*, 4 (1998), 642–652.

19. Hemmy, D. C., Tessier, P. L.: CT of dry skulls with craniofacial deformities: Accuracy of three-dimensional reconstruction. *Radiology 157*, 1 (1985), 113–116.

20. Hopper, K. D., Pierantozzi, D., Potok, P. S., Kasales, C. J., TenHave, T. R., Meilstrup, J. W., Van Slyke, M. A., Mahraj, R., Westacott, S., Hartzel, J. S.: The quality of 3D reconstructions from 1.0 and 1.5 pitch helical and conventional CT. *J. Comput. Assist. Tomogr. 20*, 5 (1996), 841–847.

21. *Internet Brain Segmentation Repository.* Massachusetts General Hospital, Harvard University, Boston, MA, 2000. `http://neuro-www.mgh.harvard.edu/cma/ibsr/`.

22. Kim, K., Wittenbrink, C. M., Pang, A.: Extended specifications and test data sets for data level comparisons of direct volume rendering algorithms. *IEEE Trans. Visualization Comput. Graphics 7*, 4 (2001), 299–317.

23. Machiraju, R., Yagel, R.: Reconstruction error characterization and control: A sampling theory approach. *IEEE Trans. Visualization Comput. Graphics 2*, 4 (1996), 364–377.

24. Marschner, S. R., Lobb, R. J.: An evaluation of reconstruction filters for volume rendering. In Bergeron, R. D., Kaufman, A. E. (Eds.): *Proc. IEEE Visualization '94.* Tysons Corner, VA, 1994, 100–107.

25. McFarland, E. G., Brink, J. A., Heiken, J. P., Balfe, D. M., Hirselj, D. A., Pilgram, T. K., Argiro, V., Littenberg, B.: Spiral CT colonography (virtual colonoscopy): Multiobserver study of different image display techniques compared to colonoscopy. In Chen, C.-T., Clough, A. V. (Eds.): *Medical Imaging 1999: Physiology and Function from Multidimensional Images, Proc. SPIE 3660.* San Diego, CA, 1999, 106–108.

26. McFarland, E. G., Brink, J. A., Loh, J., Wang, G., Argiro, V., Balfe, D. M., Heiken, J. P., Vannier, M. W.: Visualization of colorectal polyps with spiral CT colography: Evaluation of processing parameters with perspective volume rendering. *Radiology 205* (1997), 701–707.

27. Meißner, M., Huang, J., Bartz, D., Mueller, K., Crawfis, R.: A practical evaluation of popular volume rendering algorithms. In *Proc. 2000 Symposium on Volume Visualization and Graphics.* Salt Lake City, UT, 2000, 81–90.

28. Möller, T., Machiraju, R., Mueller, K., Yagel, R.: Evaluation and design of filters using a Taylor series expansion. *IEEE Trans. Visualization Comput. Graphics 3*, 2 (1997), 184–199.
29. Moorhead, R. J., Zhu, Z.: Signal processing aspects of scientific visualization. *IEEE Signal Processing Magazine 12*, 5 (1995), 20–41.
30. Novins, K. L., Arvo, J.: Controlled precision volume integration. In *Proc. 1992 Workshop on Volume Visualization*. Boston, MA, 1992, 83–89.
31. Pfister, H., Lorensen, B., Bajaj, C., Kindlmann, G., Schroeder, W., Avila, L. S., Martin, K., Machiraju, R., Lee, J.: The transfer function bake-off. *IEEE Comput. Graphics Appl. 21*, 3 (2001), 16–22.
32. Pommert, A., Höltje, W.-J., Holzknecht, N., Tiede, U., Höhne, K. H.: Accuracy of images and measurements in 3D bone imaging. In Lemke, H. U. et al. (Eds.): *Computer Assisted Radiology, Proc. CAR '91*, Springer-Verlag, Berlin, 1991, 209–215.
33. Pommert, A., Tiede, U., Höhne, K. H.: Accuracy of isosurfaces in volume visualization. In Girod, B. et al. (Eds.): *Vision, Modeling, and Visualization, Proc. VMV 2000*, IOS Press, Amsterdam, 2000, 365–371.
34. Pratt, W. K.: *Digital Image Processing*. 2. ed. John Wiley and Sons, New York, 1991.
35. Remy-Jardin, M., Remy, J., Artaud, D., Fribourg, M., Duhamel, A.: Volume rendering of the tracheobronchial tree: clinical evaluation of bronchographic images. *Radiology 208* (1998), 761–770.
36. Tiede, U., Höhne, K. H., Bomans, M., Pommert, A., Riemer, M., Wiebecke, G.: Investigation of medical 3D-rendering algorithms. *IEEE Comput. Graphics Appl. 10*, 2 (1990), 41–53.
37. Udupa, J. K., Gonçalves, R. J.: Imaging Transforms for Volume Visualization. In Taylor, R. H. et al. (Eds.): *Computer Integrated Surgery: Technology and Clinical Applications*, MIT Press, Cambridge, MA, 1995, ch. 3, 33–57.
38. Vannier, M. W.: Evaluation of 3D Imaging. *Crit. Rev. Diagn. Imaging 41*, 5 (2000), 315–378.
39. Vannier, M. W., Pilgram, T. K., Marsh, J. L., Kraemer, B. B., Rayne, S. C., Gado, M. H., Moran, C. J., McAlister, W. H., Shackelford, G. D., Hardesty, R. A.: Craniosynostosis: Diagnostic imaging with three-dimensional CT presentation. *Am. J. Neuroradiology 15*, 10 (1994), 1861–1869.
40. Wang, G., Vannier, M. W.: Stair-step artifacts in three-dimensional helical CT: An experimental study. *Radiology 191* (1994), 79–83.
41. West, J., Fitzpatrick, J. M., Wang, M. Y., Dawant, B. M., Maurer, C. R., Kessler, R. M., Maciunas, R. J., Barillot, C., Lemoine, D., Collignon, A., Maes, F., Suetens, P., Vandermeulen, D., van den Elsen, P. A., Napel, S., Sumanaweera, T., Harkness, B., Hemler, P. F., Hill, D. L. G., Hawkes, D. J., Studholme, C., Maintz, J. B. A., Viergever, M. A., Malandain, G., Pennec, X., Noz, M. E., Maguire, G. Q., Pollack, M., Pelizzari, C. A., Robb, R. A., Hanson, D., Woods, R. P.: Comparison and Evaluation of Retrospective Intermodality Brain Image Registration Techniques. *J. Comput. Assist. Tomogr. 21*, 4 (1997), 554–566.
42. Williams, P. L., Max, N. L., Stein, C. M.: A high accuracy volume renderer for unstructured data. *IEEE Trans. Visualization Comput. Graphics 4*, 1 (1998), 37–54.
43. Wittenbrink, C. M., Malzbender, T., Goss, M. E.: Opacity-weighted color interpolation for volume sampling. In Lorensen, W., Yagel, R. (Eds.): *Proc. 1998 IEEE Symp. Volume Visualization*. Research Triangle Park, NC, 1998, 135–142.
44. Yoo, T. S., Ackerman, M. J., Vannier, M.: Toward a common validation methodology for segmentation and registration algorithms. In Delp, S. L. et al. (Eds.): *Medical Image Computing and Computer-Assisted Intervention, Proc. MICCAI 2000*, Lecture Notes in Computer Science 1935, Springer-Verlag, Berlin, 2000, 422–431.