

Characterizing visual analytics in diagnostic imaging

C. Lundström^{1,2} and A. Persson¹

¹Center for Medical Image Science and Visualization (CMIV), Linköping University, Sweden

²Sectra-Imtec AB, Sweden

Abstract

Many necessary and desired improvements in healthcare are dependent on progress in medical imaging. As shown in this paper, the challenges targeted by visual analytics (VA) coincide with main challenges for radiologists' diagnostic work. Key prerequisites for VA in this application domain have been identified through analysis of a survey among 22 radiologists at a university hospital. Two major findings are that efficiency is perceived as the most challenging aspect of their diagnostic work and that an exploratory approach is necessary in everyday image review. The presented characterization constitutes a validated input for design of future VA research initiatives within medical imaging.

Categories and Subject Descriptors (according to ACM CCS): H.1.2 [Information Systems]: User/Machine Systems—Human factors J.3 [Computer Applications]: Life and Medical Sciences—Medical information systems

1. Introduction

Medical imaging is a cornerstone for healthcare and opportunities for improved diagnostics steadily arise thanks to technology advances. For the benefit of the patient, the acquired data sets become larger and more complex. It is, however, crucial that the final link in the diagnostic chain – the radiologist's review – develops to match the increased power of the acquisition step [And04]. Thus, there is a strong need to effectively support advanced reasoning through visual interaction, i.e. for visual analytics (VA) solutions.

Knowledge about the application domain is central for successful visualization research. Testimony from radiologists [Rei08] illustrates that in-depth knowledge about the visual interaction prerequisites is still scarce among technical researchers and developers. Our work aims to lay a foundation for future VA methods in this domain by describing and analyzing key characteristics of the VA-related work of a radiologist. Our results are based on a survey among 22 radiologists at a university hospital.

The question can perhaps be raised whether the radiologist's challenge is within the VA scope or not, in particular since it has been relatively sparsely represented in VA research so far. A side-effect of the current work is that diagnostic imaging is shown to indeed be a valid VA application

domain, for example since the key characteristics identified include complex discovery in massive data sets under high efficiency pressure.

2. Related work

The idea underpinning the current work is that human analytic reasoning supported by interactive visual interfaces is a powerful way to tackle large-scale data mining in many scenarios [TC05, KKEM10]. Within the medical sciences much research effort has been spent towards the objective of improving the performance of the radiologist himself, i.e. the human observer. Studies of perceptual processing, sometimes including cognitive aspects, have been employed as means to improve diagnostic performance [MGK05]. Identification of differences between expert and novice users has been proposed as input to improved training programs [Tay07, MMK*08]. Regarding the exploratory nature of image review, Krupinski [Kru03] noted: "...the degree of natural variation in both normal and abnormal structures is quite high, and radiologists will never see all possible variations no matter how long they practice and how many images they see."

The approach in the current work is to characterize image review work as a means to improve technology to better fit

the task at hand. This line of research is related to the field of Cognitive Task Analysis (CTA) [CKH06], but we do not here share the CTA objective to carry out a detailed breakdown of tasks and activities. Rogers [Rog95] studied the interplay between perception and problem-solving in medical diagnosis using “thinking aloud” observations. Expectation based on experience was found to be an important link between perceptual and cognitive processes. Many studies address the medical effect of diagnostic imaging applications [PBL*06, EKH*10] but technology is commonly seen as a black box and VA-relevant conclusions about improved solutions are rare. In contrast, the objective in this paper is to contribute to improved VA methods by paving the way for future research in the form of an extension and refinement of the existing knowledge base.

3. Key characteristics of diagnostic image review

In order to extract the key characteristics of VA work in medical image review, a questionnaire was designed. It was validated by radiologist collaborators and then distributed as a web survey to the radiologists of the University Hospital in Linköping, Sweden. The survey was completed by 22 respondents with 0 to 32 years of experience as specialist physicians in Radiology, in average 9 years. For significance tests a confidence level of 95% was used.

3.1. Cognitive task overview

In this section we first describe the main VA-related parts of a radiologist’s work, namely the tasks pertaining to medical image review. The categorization of the subtasks stems from discussions with practising radiologists including co-authors of this paper. The categorization is not intended to qualify as an in-depth task analysis, but to provide a sufficient overview for the analyses in this paper.

The components identified in a radiologist’s image review work were:

- To plan the examination such that relevant and sufficient image data are acquired (*Plan examination*)
- To localize findings relevant for the primary diagnostic conclusion (*Primary findings*)
- To localize all secondary findings that potentially can affect treatment (*Secondary findings*)
- To characterize the findings found (*Characterize*)
- To determine shape, size and relative positions for different parts of the anatomy (*Geometry*)
- To draw correct diagnostic conclusion, once all relevant findings have been found and characterized (*Diagnosis*)

A prioritized agenda for visualization researchers can be found by determining the main unsolved challenges for the end-users. Therefore, the radiologists were asked in the survey to assess how challenging (difficult and/or time-consuming) the different aspects of their work are. In addition to the subtasks listed above, the challenge *To be efficient*

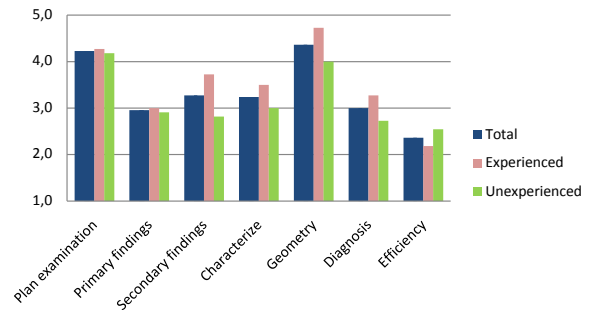


Figure 1: The radiologists’ assessment of how challenging their different tasks are. Low scores means highly challenging, see text for detailed definition. Narrow bars are partial results for experienced and unexperienced radiologists, respectively (see section 3.4).

enough to cope with workload (Efficiency) was added. The results are shown in figure 1. The scale used was a 1-7 score with odd numbers having a semantic description:

1. Extremely great and always present challenge
3. Great and often recurring challenge
5. Significant and sometimes present challenge
7. Insignificant or extremely rare challenge

The results show that radiologists overall see great and often recurring challenges in their work. The Efficiency challenge is perceived as the most problematic. Wilcoxon matched-pairs tests confirm that there is statistical significance in the difference to all other work components except to Primary findings or Diagnosis. A noteworthy result is also that the Diagnosis challenge is important, that is, even when all relevant image features have been found and characterized, the remaining diagnostic task is cumbersome.

Understanding shape, size and relative positions (*Geometry*) is considered the least important challenge. This may be surprising, since this challenge is a major motivation for 3D visualization in medical imaging. An explanation is that radiologists develop an expertise in creating mental 3D models from axial slice images, thus being less dependent on actual 3D visualizations. Nevertheless, the challenge is still significant for the radiologists and it is likely that other physicians have a relatively higher need for 3D.

3.2. Knowledge characteristics and usage

One part of the survey was directed towards investigating usage of the various sources of knowledge available to the radiologists. Five sources were named, *Newly acquired data*, *Old data* (for the same patient), *Reference cases*, *Own experience*, and *Other people’s experience*. “Data” was defined as both textual information and image data. The question posed was: *If you read 100 cases, in how many cases would you*

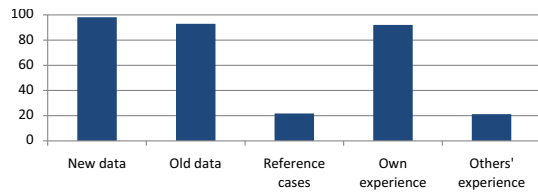


Figure 2: The usage of knowledge sources in percent of diagnostic cases.

use each source of knowledge, respectively? The results are shown in figure 2.

The results show clearly that there are three primary sources of knowledge: New data, Old data, and Own experience. These sources are in play in almost every case. While usage of new data and individual expertise can be expected in virtually any visual analytics domain, a potentially more differentiating characteristic is the fact that prior examinations are very central. A follow-up question was how many times each source would be the least valuable one, judged *a posteriori* for 100 cases. Reference cases were clearly assessed to be the least necessary source, with an average score of 47%, compared to the second ranked source Other people's experience at 23%. The difference between the two was statistically significant (matched-pairs t-test, $p = 0.031$).

The process of building experience was also studied. One question posed was: *How many cases do you need to learn and reliably work with a, for you, new type of images?* Three possible reasons for the novel images was given: a new type of disease, a new type of data acquisition, and a new type of visualization technology. Answers were given in a logarithmic scale, 10^x , $x \in [0, 1, 2, 3, 4]$. The average of x was similar for the three reasons, 1.42, 1.60, and 1.60, respectively. Thus, the radiologists claim to need about 30-40 cases to build sufficient experience for a new type of review.

The level of understanding of the underlying visualization technology is another characteristic of the visual analytics prerequisites. One question addressed this from the aspect of dealing with potential errors: *For the images you work with, how well do you know the visualization technology in terms of limitations and possible artifacts?* The same question was asked for data acquisition technology. The scale used was:

1. Not at all
2. A vague idea
3. Can interpret and deal with the most common situations
4. Can interpret and deal with virtually everything relevant
5. Can write a textbook for physicians
6. Can write a textbook for engineers/physicists

A natural target level for a radiologist would be a score of 4, to ensure that any technology limitation potentially affecting the diagnostic result is known. For visualization technology, the average score was 2.81 and for data acquisition

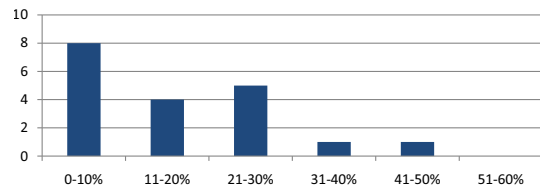


Figure 3: The percentage of cases that has unexpected findings, i.e. that could not be anticipated from the request or other patient history.

3.14. Thus, the radiologists acknowledge that technology understanding is an area for improvement. Using a similar scale, the radiologists were also asked to assess their general knowledge of the two visualization methods *Traditional 2D images* and *Volume rendering (3D)*. The respective average score was 3.44 and 2.60, a statistically significant difference (Wilcoxon matched-pairs test: $p = 0.0078$).

3.3. Exploratory components

A characteristic that highly influences the design of a visualization tool is whether the target is routine tasks or exploratory tasks. A question aimed to throw light upon this issue was: *If you review 100 cases, how many have unexpected findings?* In the instructions, "expected finding" was defined as an anticipated possibility given the request information (the background for the ordered examination) and patient history. As shown in figure 3 the answers ranged between 2-50%, with an average of 19% cases with unexpected findings. Drilling down further, one question read: *Assume that given the request information a list could be produced containing the most probable primary diagnostic conclusions. If you review 100 cases, how many times would the actual primary diagnostic conclusion be a part of the list, if the list had 3, 10, and 50 entries, respectively?* The average answer was 45% (list of top three likely diagnoses), 69% (top ten), and 90% (top 50).

The results point to a duality in the routine vs. exploratory characteristic. On one hand, a majority of cases only have expected findings, thus having a routine trait. On the other hand, as the unexpected is still fairly common, the radiologists need to have an exploratory approach to all cases. Even with a list of the 50 most likely diagnoses, in one case out of ten the conclusion would be something else.

One question concerned the radiologists' view on automatic algorithms: *How high reliability would you require to start using an automatic algorithm?* Note that the usage of the algorithms was described as a complement to the regular review, not as a replacement. The question was put for algorithms targeting detection and characterization, respectively, with reliability defined as the probability that no significant errors occur. The results in figure 4 show that 90% and 99% are the dominating responses.

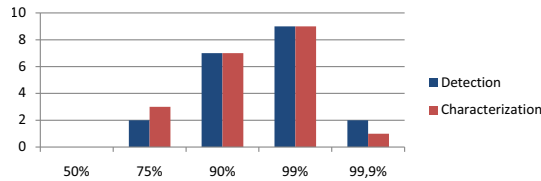


Figure 4: Required reliability (probability that no significant errors occur) for supporting automatic algorithms targeting detection and characterization, respectively.

3.4. Differences within radiologist subgroups

In the analysis of the survey, the respondents were divided into two groups based on the experience level. The experienced group (EG, $n = 11$) had worked 5-32 years as specialist radiologists, the unexperienced group (UG, $n = 11$) had worked 0-2 years. Some differences could be noted, even though they were not statistically significant. The unexperienced radiologists saw greater challenges in their work (figure 1). The average score for UG was 3.17 versus 3.53 for EG. The largest difference was found in localizing secondary findings (UG: 2.82, EG: 3.73). The unexperienced radiologists gave lower assessments of their technology understanding (section 3.2) both for visualization (UG: 2.45, EG: 3.20) and for data acquisition (UG: 2.91, EG: 3.40). Thus, there is no support for the hypothesis that younger generations of radiologists have higher technical competence, but rather indications that experience is what counts also in this area.

A comparison of radiology subspecialties was also done. Whereas significant differences were rare, an exception was the challenge estimation (section 3.1), where subspecialists in *Vascular and interventional radiology* ($n = 5$) saw least challenges, in average 3.63, those in *Thoracic radiology* ($n = 9$) saw greatest challenges, in average 2.92, and the difference was statistically significant (Mann-Whitney U test, $p = 0.047$). No radiologist was in both those groups.

4. Discussion

The picture of diagnostic imaging as a VA subdomain has been confirmed and refined by the results of this study. Apart from the well-known challenge of imaging data overload [And04], the demands on efficient delivery (section 3.1) underlines the relevance of the VA agenda. Furthermore, the fact that diagnosis remains a tough challenge even when all findings are identified points to the ambiguity and conflicts in the imaging data. In summary, the VA prerequisites in diagnostic imaging can be described as:

- Efficiency is the greatest challenge
- There is high dependence on prior data
- An exploratory approach is necessary
- There are deficiencies in the knowledge about underlying technology

An important point of discussion is whether the results of this study can be extrapolated to radiologists in general. Informal discussions with radiologists indicate that image review practice is fairly consistent world-wide, but future work to repeat similar studies elsewhere is needed to confirm this. A result with high potential bearing also outside the medical field is the end-user's limited level of understanding of the visualization technology. This important factor for VA effectiveness is an interesting topic for future work.

The characterization presented is intended to contribute to future VA research in diagnostic imaging. The value of a deeper understanding of the application domain is high in visualization research. The conclusions made can both aid selection of relevant research questions to pursue as well as guide domain-specific tailoring of developed methods.

References

- [And04] ANDRIOLE K.: Addressing the coming radiology crisis: The Society for Computer Applications in Radiology Transforming the Radiological Interpretation Process (TRIP™) initiative. *Journal of Digital Imaging* 17 (2004), 235–243. 1, 4
- [CKH06] CRANDALL B., KLEIN G. A., HOFFMAN R. R.: *Working minds: a practitioner's guide to cognitive task analysis*. MIT Press, 2006. 2
- [EKH*10] ESPINOSA L. A., KELLY A. M., HAWLEY C., INAMPUDI R., CHUGHTAI A., AGARWAL P., KHALATBARI S., MYLES J., KAZEROONI E.: Clinical utility of multiplanar reformation in pulmonary CT angiography. *Am. J. Roentgenol.* 194, 1 (2010), 70–75. 2
- [KKEM10] KEIM D., KOLHAMMER J., ELLIS G., MANSMANN F. (Eds.): *Mastering the Information Age - Solving Problems with Visual Analytics*. Eurographics, 2010. 1
- [Kru03] KRUPINSKI E. A.: The future of image perception in radiology - synergy between humans and computers. *Academic Radiology* 10 (2003), 1–3. 1
- [MGK05] MANNING D. J., GALE A., KRUPINSKI E. A.: Perception research in medical imaging. *The British Journal of Radiology* 78 (2005), 683–685. 1
- [MMK*08] MORITA J., MIWA K., KITASAKA T., MORI K., SUENAGA Y., IWANO S., IKEDA M., ISHIGAKI T.: Interactions of perceptual and conceptual processing: Expertise in medical image diagnosis. *International Journal of Human-Computer Studies* 66, 5 (2008), 370 – 390. 1
- [PBL*06] PERSSON A., BRISMAR T., LUNDSTRÖM C., DAHLSTRÖM N., OTHBERG F., SMEDBY Ö.: Standardized volume rendering for MR angiography measurements in the abdominal aorta. *Acta Radiologica* 47 (2006), 172–178. 2
- [Rei08] REINER B.: One size (doesn't) fit all. *Journal of the American College of Radiology* 5, 4 (2008), 567 – 570. 1
- [Rog95] ROGERS E.: A study of visual reasoning in medical diagnosis. In *Proceedings of Eighteenth Annual Conference of the Cognitive Science Society* (1995), pp. 213–218. 2
- [Tay07] TAYLOR P. M.: A review of research into the development of radiologic expertise: Implications for computer-based training. *Academic radiology* 14 (2007), 1252–1263. 1
- [TC05] THOMAS J. J., COOK K. A.: *Illuminating the Path: The Research and Development Agenda for Visual Analytics*. IEEE Computer Society, 2005. 1