

Feature-enhancing zoom to facilitate Ki-67 hot spot detection

Jesper Molin^{*a,d,e}, Kavitha Shaga Devan^{b,d}, Karin Wårdell^{b,d}, Claes Lundström^{c,d,e}

^aDept. of Applied IT, Chalmers University of Technology; ^bDept. of Biomedical Engineering, Linköping University, ^cDept. of Science and Technology, Linköping; ^dCenter for Medical Image Science and Visualization, Linköping University; ^eSectra AB, Sweden

ABSTRACT

Image processing algorithms in pathology commonly include automated decision points such as classifications. While this enables efficient automation, there is also a risk that errors are induced. A different paradigm is to use image processing for enhancements without introducing explicit classifications. Such enhancements can help pathologists to increase efficiency without sacrificing accuracy. In our work, this paradigm has been applied to Ki-67 hot spot detection. Ki-67 scoring is a routine analysis to quantify the proliferation rate of tumor cells. Cell counting in the hot spot, the region of highest concentration of positive tumor cells, is a method increasingly used in clinical routine. An obstacle for this method is that while hot spot selection is a task suitable for low magnification, high magnification is needed to discern positive nuclei, thus the pathologist must perform many zooming operations. We propose to address this issue by an image processing method that increases the visibility of the positive nuclei at low magnification levels. This tool displays the modified version at low magnification, while gradually blending into the original image at high magnification. The tool was evaluated in a feasibility study with four pathologists targeting routine clinical use. In a task to compare hot spot concentrations, the average accuracy was $75 \pm 4.1\%$ using the tool and $69 \pm 4.6\%$ without it ($n=4$). Feedback on the system, gathered from an observer study, indicate that the pathologists found the tool useful and fitting in their existing diagnostic process. The pathologists judged the tool to be feasible for implementation in clinical routine.

Keywords: cell nuclei, color deconvolution, zoom interaction, clinical evaluation

1. INTRODUCTION

Accurate grading of breast cancer from histology samples is a challenging task for the pathologist due to the heterogeneous nature of the disease. Ki-67 is a commonly used immunohistochemical (IHC) assessment to quantify the proliferation rate of tumor cells¹. Ki-67 scoring has been criticized due to its subjective nature as well as a lack of standardized protocols². This method lacks consistency and uniformity and thus gives rise to inter and intra-observer variability^{3,4}. Despite this downside, the usage of Ki-67 as a prognostic indicator for breast cancer is supported by recent research results^{5,6}. Ki-67 assessment using hot spots is becoming increasingly prevalent in clinical use. The Swedish Pathologists Association⁷ and the Royal College of Pathologists⁸ in London have suggested using hot spots to score the Ki-67 index for breast cancer and endocrine tumors respectively.

While most histological grading are still performed using conventional light microscopy, the increased usage of whole slide imaging (WSI) within telepathology^{9,10} and to some extent for primary diagnostics¹¹⁻¹³ makes it feasible to introduce digital image analysis methods (DIA-methods) to perform image analysis tasks and to aid pathologists in their diagnostic process. Current DIA-methods are able to produce quantitative assessments at an acceptable level for IHC-stained digitally scanned WSI:s under the condition that a suitable region of interest is selected by a human operator¹⁴. Specific methods for Ki-67 assessment of selected regions of interest have been studied^{15,16} and FDA-approved algorithms been introduced to the market¹⁷. However, the selection of hot spots has been given less attention even though it is cumbersome. To be able to see and compare all different regions of a whole slide image at the same time, the

* mjesper@chalmers.se; phone +46 73 818 31 32; <http://www.t2i.se>.

pathologist often need to zoom out to 1X magnification (10 microns/pixel) depending on the size of the screen used and the size of the tissue sample. At that magnification, many cell nuclei are smaller than a pixel and will not be visible. Using a larger screen with increased resolution could in some cases be a solution but that is not always convenient or cost effective.

Another potential solution is to turn to image analysis methods. A common paradigm for medical image analysis is to strive for algorithms that explicitly make the requested classifications, such as "healthy" or "diseased". Even powerful algorithms seldom reach sensitivity and specificity higher than 80-90%, which means that the errors made must be detected and dealt with. The approach often taken is then to create a visualization where the physician can verify the automated results. For example, this paradigm is employed in the method proposed by Roullier et al.¹⁸, a method that shares our target of cell proliferation analysis. There are two important drawbacks of this paradigm. One risk is that the physician puts too much trust in the algorithm results, leading to incorrect conclusions. The other risk is that the physician spends much time to verify the results, which counteracts the efficiency gained from automation. These drawbacks are fundamental hurdles for bringing explicitly classifying algorithms into clinical practice.

In our work, a different paradigm has been employed: to use image processing for enhancements without introducing explicit classifications. Our aim is not to let the algorithm do the task of the physician, but to make the physician's work easier and faster, which makes them well suited for introduction to clinical practice. Generally, it can be noted that humans are not completely replaced when some kind of automation is implemented, their role just changes¹⁹. In line with the conclusions of Dekker and Woods²⁰, our design process is based on accepting that humans cannot be completely excluded, and instead maximize the benefit from human strengths such as the perceptual system's tremendous capacity.

2. METHOD

The main focus of our approach was to improve the pathologist's capability to select Ki-67 hot spots by enlarging positive nuclei to make them discernable at low magnification while at the same time using algorithms that took as few decisions as possible. It was also important to use algorithms that were easy to understand for the pathologist in order to increase their capability of detecting errors. The underlying idea is that of a simple algorithm with full transparency towards the user is to be preferred over a black-box algorithm where it is impossible to understand when and how it fails, even if the latter has a slightly lower error rate.

But simplicity is relative; algorithms that are simple for an algorithm designer can be quite complicated for a pathologist because of the difference in training and expertise. To overcome this difference we wanted to create a system that could be learned through the experience of using it. However, this implies a quite strict limitation onto the system: the output from the system must be consistent in between uses. Therefore any kind of adaptive image operation could not be used.

The choice fell on using a color deconvolution algorithm²¹ which in a first step separates the image into a brown and a blue channel representing the Diaminobenzidine (DAB) and Hematoxylin staining respectively. The brown channel contained mostly positive nuclei, and the blue channel contained other artifacts such as negative nuclei, cytoplasm and stroma. Since the color deconvolution separated out the positive nuclei, the enlargement effect was achieved by filtering the brown channel using a Gaussian kernel with a sigma size of 25 pixels, which caused the positive nuclei to almost double in size. To compensate for the decreased intensity caused by the Gaussian filter, the contrast of the brown channel was then increased using a fixed linear look up table. The blue channel was discarded as it did not provide any visual enhancement for the hot spots. Note that none of the above methods take any statistical properties of the input image into account, it just expects an image stained with DAB and Hematoxylin. Small normal variations in what color DAB and Hematoxylin represent in the color space does not have large impact on the end result since pathologists are used to compensate for variations in staining quality.

As a second step, dyadic subsampled pyramids for different magnification levels (20X, 10X, 5X, 2.5X, etc.) of the modified brown channel and the original image were created. For each magnification level, a specific mix of the modified brown channel and the original image was created. The mix at each level was chosen so that the low magnification levels would be dominated by enlarged brown cells and the high magnification levels stay unmodified, the levels are given in Table 1.

Table 1. Mix between the modified brown channel and the unmodified image for different zoom levels.

Zoom level	20X-10X	10X-5X	5X-2.5X	2.5X and below
Modified brown channel	0%	25%	50%	75%
Original image	100%	75%	50%	25%

By viewing the mixed images with an image viewer, a tool for finding and verifying hot spots was created. At low magnification levels, the modified image helped to detect hot spots by making positive nuclei visible. With a simple zoom interaction using the mouse, the hot spot could quickly be verified by zooming into the original image. This provides a mechanism for learning since there is a smooth transition between the processed image and the real image. The zooming effect is visualized in Figure 1 and the difference between an unprocessed and a processed image is given in Figure 2.

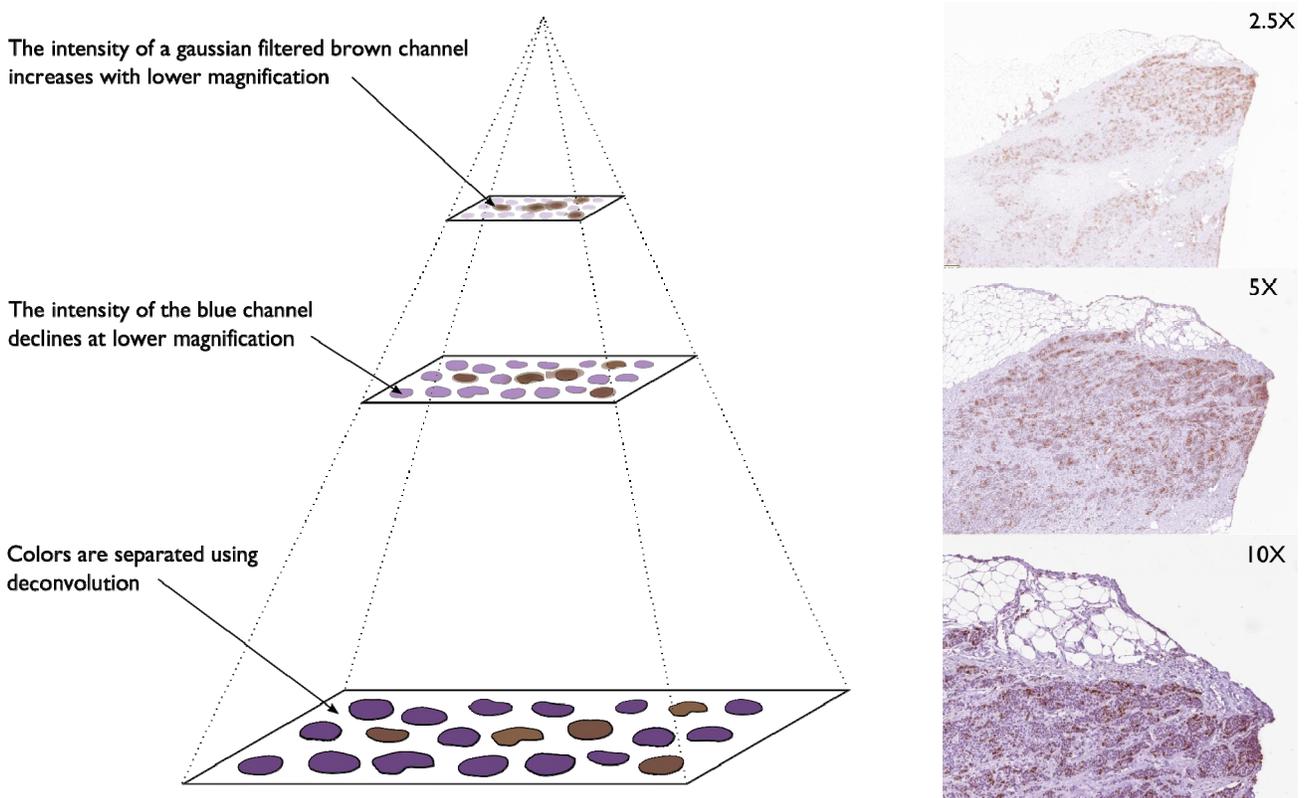


Figure 1. The algorithm used to create the zooming in effect. The original image was separated into a brown and a blue channel. The brown channel was then filtered with a Gaussian kernel to double the size of the Ki-67 positive nuclei in order to become visible at low magnification.

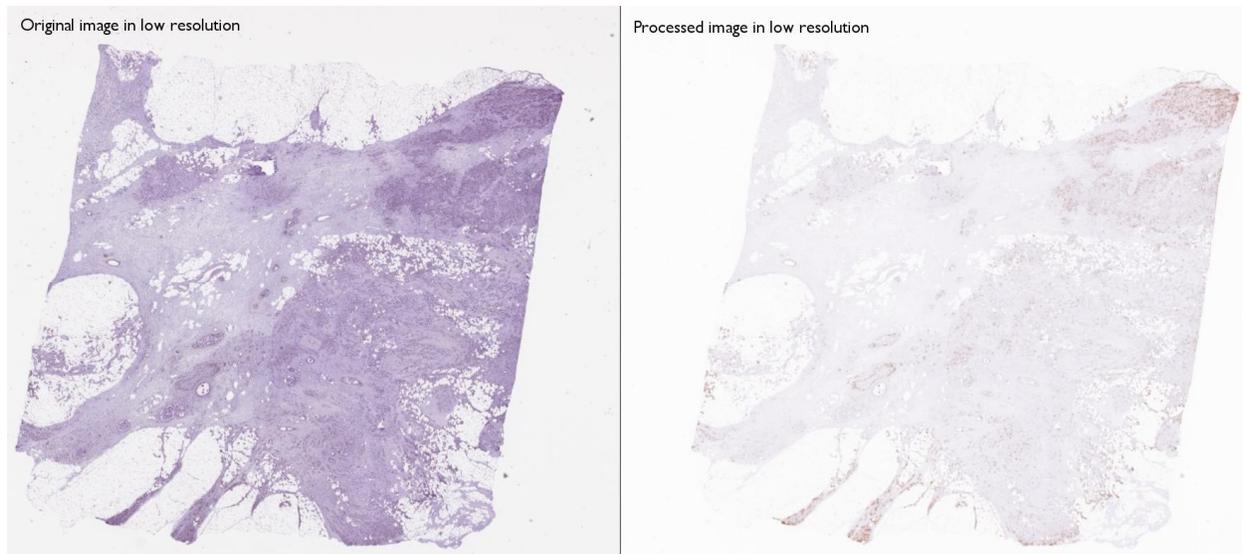


Figure 2. The original image (left) and the processed image (right) side-by-side in low magnification. In the processed image the positive nuclei (brown) become visible in order to enable detection of hot spots at a low magnification level.

3. EVALUATION

A dataset of eleven digital whole slide images from the Pathology Department of Linköping University Hospital, Sweden, were included in the study. The specimens were stained with the Ki-67 antibody MIB1 combined with the DAB cytochrome and then counterstained with Hematoxylin. The slides were scanned using an Aperio AT Turbo scanner (Vista, USA) at 20X magnification (0.5 microns/pixel). To evaluate the feasibility of the tool for clinical use, six pathologists with varying degree of experience were recruited from Stockholm South General Hospital and Linköping University Hospital, Sweden. For the establishment of the ground truth, an experienced pathologist compiled a dataset of candidate hot spots within the eleven whole slide images. The hot spots were divided up into 64 pairwise comparisons, for each pair, the preferred hot spot (having the highest positive tumor cell count) was determined at full magnification. Any pair without an evident difference was removed from the trial, which then finally consisted of 50 pairwise comparisons.

The task of the trial participants was to go through the 50 preselected pairs and select their preferred hot spot in each pair. 25 pairs were selected in a mode using the modified images and the remaining 25 with the original image at 1X magnification. The order of the pairs was randomized. After a four week washout period, the participants again selected their preferred hot spot for each pair, but in the opposite mode. The average detection rate, standard deviation as well as a two-tailed paired T-test was calculated. $P < 0.05$ was considered statistically significant.

After the second trial, an observer study was performed in order to assess the participant's opinion about the feasibility and benefits of the tool. The pathologists were given a free reign to pan and zoom within the tool and while expressing their opinions about in a semi-structured interview. The discussions were recorded and transcribed before being analyzed.

4. RESULTS

The trial test was conducted in two sessions which were held four weeks apart to include a washout period. The pathologists had an average detection rate of $69 \pm 4.6\%$ ($n=50$) using the original histology images. Using the Ki-67 hot spot detection tool, the pathologists were able to have an average detection rate of $75 \pm 4.1\%$ ($n=50$). A two-tailed paired T-test was conducted between detection rates of the original image and the processed image. The difference was not statistically significant, $p=0.272$ ($n=4$).

Five pathologists participated in the semi-structured interview focused around three different themes: comprehensibility of the image processing, system improvements and clinical routine usage.

4.1 Comprehensibility of the image processing

Three out of five pathologists could directly describe the effect of the image processing in low magnification while the other two had to be aided by a side-by-side comparison as in Figure 2 to see the effect. The effect was generally described as a decrease in intensity of the “blue” and a few mentioned that the contrast of the “brown” had been increased. None of the pathologists realized that the size of the nuclei had been increased.

Different comparisons with common artifacts were also made, the intensity decrease of the “blue” was compared to “something gone wrong in the lab” and the increased brown intensity was compared to “having a lot of anti-bodies”.

In some improvement suggestions it was questionable whether the pathologist had an accurate mental model of digital image processing. For instance, it was suggested that the nuclei should be in higher resolution at lower magnification and that when describing that the nuclei had been enlarged, a question was raised whether the cytoplasm also had been enlarged.

To summarize:

- The image processing could be described to some extent by most pathologists using simple words.
- There might be a risk of confusion between the image processing and artifacts from the laboratory processes.
- The mental model regarding digital image processing of the pathologists might not always be accurate.

4.2 System improvements

The pathologists did not agree at what level of magnification of the transition between the processed image and the original image should occur. Two pathologists wanted the processed image to stay on longer when zooming in while the other three wanted the original image earlier. It was suggested that a good rule of thumb for the transition would be to start using the original image when the nuclei are clearly separable, at a magnification level around 2X.

Two of the pathologists felt that it was harder for them to distinguish between tumor and non-tumor cells at low magnification as this decision includes information given by the blue channel which had been suppressed by the image processing. Another pathologist pointed out that while this technique was useful when DAB-staining was used, other cytochromes might give enough contrast at low magnification without the use of image processing.

To summarize:

- It was unclear at what magnification level the processed image should transition into the original image.
- The intensity decrease of the blue channel made it harder to distinguish tumor from non-tumor cells.
- The image processing might not be necessary when other cytochromes are used.

4.3 Clinical routine usage

When discussing how this tool could be used clinically, all the pathologists thought it was a good idea to use the low magnification level to display the output from the image processing since no important decisions usually were made at that level, most of them usually looked for hot spots at a magnification level around 5X. Two of the pathologists who wanted the transition between the processed image and the original image at a lower magnification, usually looked for the hot spots at a lower magnification than the others.

Possible sources of error were also discussed using the artifact given in Figure 3 as an example of when unwanted artifacts and possible hotspots could be confused.

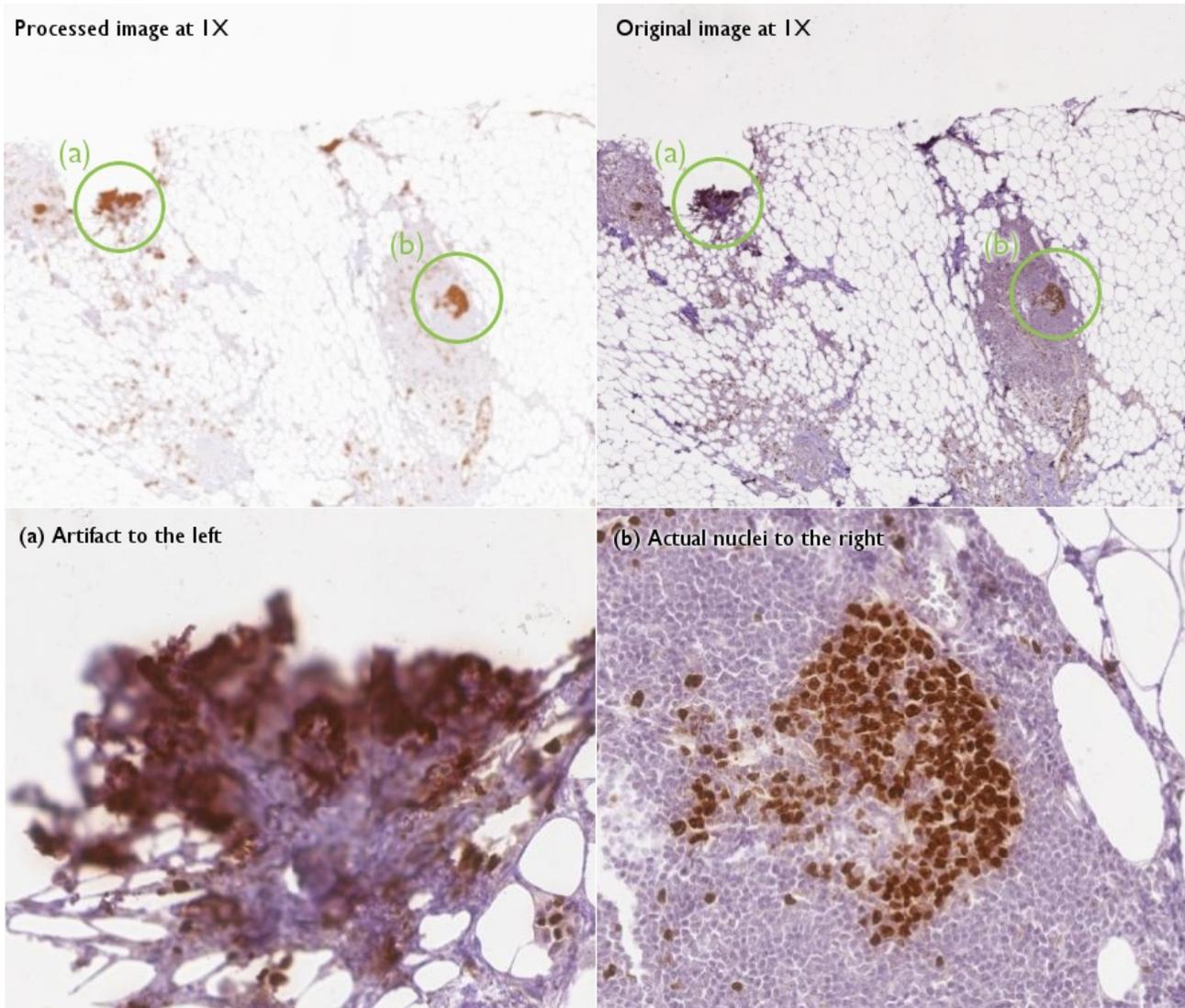


Figure 3. Upper-left image show two possible hot spots in the processed image, upper-right shows the same spots with the original image. Bottom-left shows the possible hot spot to the left in the processed image (a) at high magnification and is an artifact. Bottom-right shows the possible hot spot to the right (b).

While the processed image might make it harder to distinguish hot spots from artifacts, one pathologist pointed out that this problem still exists even without any image processing. The pathologists did not think that there was any risk of confusion between artifacts and hot spots since they have to zoom in to make the actual count of positive nuclei. During the trial, two pathologists discarded the possible hot spot to the right in the original image at 1X (Figure 3, upper-right) and when confronted with this in the interview they revealed that they thought it was an artifact.

The pathologists' attitudes towards using the tool in a clinical setting were generally positive. The processed image provided a better sense of the distribution of positive nuclei and it would probably become easier and faster to select hot spots using the tool. To increase the acceptability and to avoid complaints it was suggested to offer the possibility to turn-off the visualization. Another suggestion was to create a system that automatically detects the hot-spot and highlights the region using a pointer.

When asked if they would use the tool in their daily work, four out of five pathologists agreed.

To summarize:

- The processed image at low magnification did not interfere with their usual way of viewing the slides.
- Artifacts at low magnification were not seen as harmful since they had to zoom in to do the counting.
- The tool improved their sense of distribution of positive nuclei and would probably make hot spot selection faster and easier.
- It was suggested that an option to turn off the visualization was added to increase the acceptability and to avoid complaints.
- Four out of five pathologists was positive towards using the tool in their daily routine work.

5. DISCUSSION

In this study, we introduced a new tool able to aid the pathologists in detection candidate hot spots. From the two test trials conducted, we found that the improvement obtained was not statistically significant. In order to further investigate whether the tool has the potential to aid and improve hot-spot detection, we conducted a short observer study with a semi-structured interview with five participating pathologists.

The number of participants used in this study was not large enough determine significance because the anticipated effect size was smaller than expected, mainly because the pathologists performed better than expected using the original image. A possible reason for that could have been due to the fact that the viewing distance to the screen was not normalized. It was noted during the trial that the participants tended to get closer to the screen when viewing the original version of the images. This might have caused them to have a better view when trying to detect the spots in the original image than with the processed image.

The test was chosen to imitate a possible future interaction with the system, the decision to zoom in over a possible hot spot. Since hot spot selection is directly verified upon that interaction, the test reflects the probability of having to zoom out again and select another possible hot spot and should therefore be interpreted as a measure of efficiency rather than accuracy.

While the number of participants was not large enough to determine significance, it was sufficient to gain a rich understanding of how the tool was perceived.

Most of the pathologists were able to detect that something was different with the processed image at least when viewing it side-by-side with the original, but the fact that the processing was compared with laboratory artifacts and that the processing could not be accurately described, implies that the mental model of the pathologists with regards to the processing might not always be accurate. Having an appropriate mental model of the processes underlying a system is important in order to detect and understand errors. This is exemplified within the field of diagnostic pathology by the fact that pathologists are quite skilled at compensating for errors in staining quality partly because they have a deep understanding of the underlying laboratory processes.

We therefore think it is an interesting future research endeavor to see whether the negative effects of this mismatch in the pathologists' mental models could be overcome using the proposed interaction design for experiential learning. This experimental design could unfortunately not give any insight in long-term learning effects. In the test trials, the pathologists were only given a few minutes to familiarize themselves with using the digital tool before the trial was conducted.

The pathologists provided useful feedback which can be used to improve the tool. The image processing should improve the ability to differentiate between the tumor and non-tumor cells. With the current image processing, this was due to the fact that the intensity of the blue channel deliberately had been decreased in order to make the brown nuclei stand out. Turning up the blue intensity might require a more offensive processing of the brown channel in order to keep the popping out effect, which in turn might decrease the tolerance for staining variations.

The overall impression from the pathologists were positive, they felt that they gained a better understanding of the distribution and structure of the positive nuclei by making them visible without getting in the way of their usual habits. This is probably why they thought the tool would make hot spot selection faster and easier. Image processing tools which does not get in the way is quite unusual and we believe that the interactive visualization pattern to present some kind of overlay at low magnification and gradually zoom in to the original image could be adapted for other image processing applications as well.

In order to improve the current tool to achieve a better detection rate, we would aim at preserving the information given by the blue channel while still improving the visibility of the positive nuclei at the low magnification level. We would also perform a larger test study using a higher number of images and participants. Investigating the improvement in Ki-67 scoring time using this tool could be another interesting avenue of approach. Thus, we think that that our study does show that the tool has potential if the above proposed changes are made.

6. CONCLUSION

A novel hot spot detection tool is presented based on a method paradigm meeting demands from clinical routine work: use image processing to aid the pathologist without introducing explicit classifications. The short observer study with the pathologists using the tool revealed that the low magnification levels can be used for visualization of generated data, without interfering with the pathologists' usual way of examining the whole slide images. From the feedback of the pathologists, it can be concluded the proposed tool is feasible to be implemented in clinical routine. However, the tool needs further validation using a larger number of participants to verify its reproducibility and sensitivity.

ACKNOWLEDGEMENTS

This work has been supported by project grants from the Swedish Research Council (Dnr: 2011-4138) and from the Swedish Innovation Agency (Dnr: 2012-0112). We would like to thank Sten Thorstenson, senior consultant pathologist, Department of Pathology and Clinical Genetics, University Hospital, Linköping, Sweden for his consultation in this work. We would also like to thank the pathologists from Stockholm South General Hospital and Linköping University Hospital, Sweden that participated in our test trials.

REFERENCES

- [1] Dowsett, M., Nielsen, T.O., A'Hern, R., Bartlett, J., Coombes, R.C., Cuzick, J., Ellis, M., Henry, N.L., Hugh, J.C., et al., "Assessment of Ki67 in breast cancer: recommendations from the International Ki67 in Breast Cancer working group," *J. Natl. Cancer Inst.* 103(22), 1656–64 (2011).
- [2] Romero, Q., Bendahl, P.-O., Klintman, M., Loman, N., Ingvar, C., Rydén, L., Rose, C., Grabau, D., and Borgquist, S., "Ki67 proliferation in core biopsies versus surgical samples - a model for neo-adjuvant breast cancer studies.," *BMC cancer* 11(1), 341 (2011).
- [3] Tang, L.H., Gonen, M., Hedvat, C., Modlin, I.M., and Klimstra, D.S., "Objective quantification of the Ki67 proliferative index in neuroendocrine tumors of the gastroenteropancreatic system: a comparison of digital image analysis with manual methods," *Am. J. Surg. Pathol.* 36(12), 1761–1770 (2012).
- [4] Varga, Z., Diebold, J., Dommann-Scherrer, C., Frick, H., Kaup, D., Noske, A., Obermann, E., Ohlschlegel, C., Padberg, B., et al., "How reliable is Ki-67 immunohistochemistry in grade 2 breast carcinomas? A QA study of the Swiss Working Group of Breast- and Gynecopathologists.," *PloS one* 7(5), e37379 (2012).

- [5] Aleskandarany, M. A, Rakha, E. A, Macmillan, R.D., Powe, D.G., Ellis, I.O., and Green, A.R., "MIB1/Ki-67 labelling index can classify grade 2 breast cancer into two clinically distinct subgroups," *Breast Cancer. Res. Treat.* 127(3), 591–9 (2011).
- [6] Honma, N., Horii, R., Iwase, T., Saji, S., Younes, M., Ito, Y., and Akiyama, F., "Ki-67 evaluation at the hottest spot predicts clinical outcome of patients with hormone receptor-positive/HER2-negative breast cancer treated with adjuvant tamoxifen monotherapy," *Breast cancer* (2013).
- [7] Svensk förening för patologi, "KVASt dokument brösttumörer, utgåva 3.1.2." <http://www.svfp.se/node/214>.
- [8] Stephens, T. J., Cross S. S., Chetty R., [Standards and datasets for reporting cancers: Dataset for endocrine tumours of the gastrointestinal tract including pancreas (3rd edition)], The Royal College of Pathologists, 12 (2012).
- [9] Pantanowitz, L., Wiley, C.A., Demetris, A., Lesniak, A., Ahmed, I., Cable, W., Contis, L., and Parwani, A. V., "Experience with multimodality telepathology at the University of Pittsburgh Medical Center," *J. Pathol. Inform.* 3(1), 45 (2012).
- [10] Evans, A.J., Chetty, R., Clarke, B.A., Croul, S., Ghazarian, D.M., Kiehl, T.-R., Perez Ordonez, B., Ilaalagan, S., and Asa, S.L., "Primary frozen section diagnosis by robotic microscopy and virtual slide telepathology: the University Health Network experience," *Hum. Pathol.* 40(8), 1070–81 (2009).
- [11] Thorstenson, S., "Digital Pathology System," <http://laboratory-manager.advanceweb.com/Archives/Article-Archives/Digital-Pathology-System.aspx>.
- [12] Al-Janabi, S., Huisman, A., Nap, M., Clarijs, R., and van Diest, P.J., "Whole slide images as a platform for initial diagnostics in histopathology in a medium-sized routine laboratory," *J. Clin. Pathol.* 65(12), 1107–11 (2012).
- [13] Stathonikos, N., Veta, M., Huisman, A., and van Diest, P.J., "Going fully digital: Perspective of a Dutch academic pathology lab," *J. Pathol. Inform.*, (2013).
- [14] Riber-Hansen, R., Vainer, B., and Steiniche, T., "Digital image analysis: a review of reproducibility, stability and basic requirements for optimal results," *APMIS* 120(4), 276–289 (2012).
- [15] Tuominen, V.J., Ruotoistenmäki, S., Viitanen, A., Jumppanen, M., and Isola, J., "ImmunoRatio: a publicly available web application for quantitative image analysis of estrogen receptor (ER), progesterone receptor (PR), and Ki-67," *Breast cancer res.* 12(4), R56 (2010).
- [16] Niazi, M.K.K., Pennell, M., Elkins, C., Hemminger, J., Jin, M., Kirby, S., Kurt, H., Miller, B., Plocharczyk, E., et al., "Entropy based quantification of Ki-67 positive cell images and its evaluation by a reader study," *Proc SPIE* 8676, 86760I (2013).
- [17] Ventana, "Ventana Receives FDA Clearance for Ki-67 (30-9) Image Analysis and Digital Read Applications" <http://www.ventana.com/site/page?view=press-release-february28-2012>.
- [18] Roullier, V., Lézoray, O., Ta, V.-T., and Elmoataz, A., "Multi-resolution graph-based analysis of histopathological whole slide images: application to mitotic cell extraction and visualization," *Comput. Med. Imaging Graph.* 35(7-8), 603–15 (2011).
- [19] Bainbridget, L., "Ironies of Automation," *Automatica* 19(6), 775–779 (1983).

- [20] Dekker, S.W.A., and Woods, D.D., "MABA-MABA or Abracadabra ? Progress on Human – Automation Coordination," *Cogn. Technol. Work* 4, 240–244 (2002).
- [21] Ruifrok, A.C., and Johnston, D.A., "Quantification of histochemical staining by color deconvolution.," *Anal. Quant. Cytol. Histol. Int. Acad. Cytol. [and] Am. Soc. Cytol.* 23(4), 291–299 (2001).