Symposium - 2nd Nordic Symposium on Digital Pathology

RandomSpot: A web-based tool for systematic random sampling of virtual slides

Alexander I. Wright¹, Heike I. Grabsch¹,², Darren E. Treanor¹,³

¹Section of Pathology and Tumour Biology, Leeds Institute of Cancer and Pathology, University of Leeds, Leeds, England, UK
²Department of Pathology, Maastricht University Medical Center, Maastricht, Netherlands
³Department of Histopathology, St. James's University Hospital, Leeds Teaching Hospitals NHS Trust, Leeds, England, UK

E-mail: *Alexander I. Wright - a.wright@leeds.ac.uk
*Corresponding author

Received: 24 November 14 Accepted: 25 November 14 Published: 24 February 15

Abstract

This paper describes work presented at the Nordic Symposium on Digital Pathology 2014, Linköping, Sweden. Systematic random sampling (SRS) is a stereological tool, which provides a framework to quickly build an accurate estimation of the distribution of objects or classes within an image, whilst minimizing the number of observations required. RandomSpot is a web-based tool for SRS in stereology, which systematically places equidistant points within a given region of interest on a virtual slide. Each point can then be visually inspected by a pathologist in order to generate an unbiased sample of the distribution of classes within the tissue. Further measurements can then be derived from the distribution, such as the ratio of tumor to stroma. RandomSpot replicates the fundamental principle of traditional light microscope grid-shaped graticules, with the added benefits associated with virtual slides, such as facilitated collaboration and automated navigation between points. Once the sample points have been added to the region(s) of interest, users can download the annotations and view them locally using their virtual slide viewing software. Since its introduction, RandomSpot has been used extensively for international collaborative projects, clinical trials and independent research projects. So far, the system has been used to generate over 21,000 sample sets, and has been used to generate data for use in multiple publications, identifying significant new prognostic markers in colorectal, upper gastro-intestinal and breast cancer. Data generated using RandomSpot also has significant value for training image analysis algorithms using sample point coordinates and pathologist classifications.

Key words: Random sampling, randomspot, systematic, web system

INTRODUCTION

This paper describes work presented at the Nordic Symposium on Digital Pathology 2014, Linköping, Sweden. The manual quantification of tissue types within regions of interest (ROI) on pathological slides is laborious and prone to interobserver and intraobserver variation. Current quantification methods (scoring) rely on experienced pathologists to make informed estimations of the proportions of tissue, either on a whole slide or within a given ROI. These estimations are subjective and require broad category bins in order to maintain inter and intra scorer consistency (typically three to four bins). Virtual slide viewing software solutions often provide inbuilt tools for drawing boundaries around tissue types, or counting cells with mouse clicks. Although...
applying these techniques to gigapixel resolution images is possible, due to the sheer number of cells per image, the application of these techniques is not practical. Random sampling provides a feasible alternative to manual whole slide quantification but is prone to biases when choosing areas for sampling. Systematic random sampling (SRS) allows accurate, unbiased estimation of the proportion of classes within a given ROI, when using the appropriate number of measurements. Traditionally, SRS involves placing a fixed grid (usually on an optical graticule) at a random seed point on a slide and counting objects under the points on the grid.[1] The efficacy of the SRS technique fundamentally relies on selecting an appropriate density of the sampling points within the grid. The density of the grid should be based on the estimation of the relative proportions of tissue within the area being analyzed (low vs. high frequency), as well as the distribution of those proportions (sparsely vs. densely distributed).[2]

With an appropriate number of samples, the true frequency of objects in the whole tissue can be estimated effectively from the sampled frequency using the grid,[3,4] which ensures higher reproducibility and consistency in scoring. However, SRS is still laborious, requiring the user to make hundreds of measurements using a conventional microscope and an optical graticule. Errors are easily made, and it is difficult to pause and recommence work.

**METHODS**

RandomSpot is a web-based system that enables pathologists to apply SRS to their own tissue samples, using digital slides. Digital slides may be annotated using specialist viewing software, which provides drawing tools for measuring distances and areas, or delineating boundaries of tissue in order to highlight ROIs, such as epithelial cancer cells.

The system uses HTML5, JQuery and PHP, with a MySQL database, and MATLAB compiled executable programs used for background data processing. As a result, the system is both platform and browser independent. RandomSpot is primarily modeled on current pathologist workflow, which has been simplified to identify two separate tasks (use cases):

**Use case one:** Creating an equidistant set of spatial co-ordinates, within a given region/several ROI (for SRS)

**Use case two:** Collecting pathologist-labeled SRS points (after SRS points have been scored) as expert classified coordinates.

Use case one utilizes HTML5 and JQuery to create a simple, user-friendly interface for uploading XML ROI. ROI can be rectangular, elliptical or polygonal [Figure 1]. Once uploaded, the RandomSpot algorithm places equidistant, systematic, randomly distributed spots within the ROIs.

The user can create their own ROIs by drawing around desired areas in their virtual slide viewing software, and export the delineations as annotation(s), stored in one or many files, in XML format. Users can then upload a single XML file containing one or more ROI, or multiple XML files contained within a zip archive, for batch processing. Once uploaded, the number of spots within the ROI can then be set, and the tolerance with which that number may be increased or decreased by, as a percentage. It should be noted that prior to use, calculations must be made in order to obtain the appropriate sample number for the given area of the ROI provided. The number of spots required will rely on the expected frequency of the tissue type being measured. By default, the system sets 300 as the target number of spots, as this is optimal for minimizing the coefficient of variation for a target frequency of 50% (i.e. a normal two class system).

The user may also customize the viewing settings for each spot, setting the default zoom level (at which each spot is viewed), the annotation type and the size of the annotation.

Spots are created using a MATLAB compiled executable program, which initially generates a grid with an arbitrary number of spots, with a random seed to initiate the first spot.[5] Spots are spaced equidistantly using a hexagonal mesh, which is iteratively increased or decreased in size until the number of spots within the ROI matches the number chosen by the user, within the percentage tolerance specified. Figure 2 shows the ROI illustrated in Figure 3, overlaying a hexagonal mesh grid. The number of hexagons within the ROI (colored green) is 373, which

**Figure 1:** Examples of hand drawn annotation types, drawn on the same digital slide-rectangular, elliptical and polygonal

**Figure 2:** (a) Example of the hexagonal grid applied to the given regions of interest, (b) and a graph of the iterations required in order to create the correct sized grid
falls within the tolerance limits of the target number of spots set in Figure 4 (400 ± 15%). Decreasing the tolerance level results in more iterations being required until the desired mesh is created.

If processing individual XML files, the system will return the annotations as a single xml file ready for immediate download. When processing a batch of XML files, submitted using a zip archive, processing is done as a background process, and a zip file containing the resulting XML files will be e-mailed to the user upon completion. Once the user obtains the XML annotations containing the spots, they can then open the annotations in their virtual slide viewing software. From there, they are able to classify each spot individually, in order to get an unbiased sample of the proportions of tissue within the ROI.

Use case two focuses on retaining the value of the hand annotated data, by encouraging users to submit their completed XML files after they have used the data for their own research. As with creating spots, XML files can be uploaded individually or as a zip archive. These files are processed, and each spot is added to the RandomSpot database – RandomSpotDB. These hand labeled locations are processed and stored as text classifications, paired with the URL strings containing x and y co-ordinates of the spot that has been scored, as well as the location of the virtual slide which the spot relates to. In addition to uploading XML files, users are encouraged to submit a scoring key, which matches the shortcut keys they used whilst scoring, to the actual semantic text classifications that are used when analyzing the data. Also, to add further value to RandomSpotDB, each time a user submits their XML files, they are asked which type of tissue it is that they have scored. Having an extensive set of human scored images, which is searchable by tissue type is extremely useful for training computer vision algorithms. Currently, a separate web-based system is being developed in order to allow researchers access to this database.

A simple evaluation of the performance of the RandomSpot algorithm was undertaken in order to identify processing limitations that pathologists may encounter. Performance of the algorithm was assessed on three types of annotations: Rectangular, elliptical and polygonal. Each annotation was drawn with the same area (85,359,118 µ²) in order to preserve maximum comparability of performance between types. For each annotation type, two separate sets of processing were carried out, manipulating the number of target spots (nT) and the percentage tolerance (pT) as the independent variables respectively. The effects of both variables on the time taken to process were observed separately. For evaluating the effect of nT, the algorithm was run using values from 1 to 1000 in intervals of 100, with a fixed pT of (the default value) 15%. The algorithm was repeated for each nT observed (n = 20) in order to account for the random initiation of the hexagonal grid. The effect of pT was evaluated using values from 1 to 5 in intervals of 5, with a fixed pT of 300. Again, the algorithm was repeated for each pT observed (n = 20).

RESULTS

RandomSpot has been used extensively at Leeds for the past 6 years[6] and is continually being improved iteratively, in response to user feedback. Since the introduction of this sampling system, research at Leeds has shown that the proportion of tumor to stroma can predict the response to therapy in colorectal cancer patients,[7] and that the level of tumor cell density is a prognostic indicator of response to preoperative therapy.[8] It has also been used in studies that have successfully
identified the prognostic significance of the ratio of tumor to stroma in breast[9] and upper gastro-intestinal[10] cancers. Currently, the system has been adopted by over 40 active users, which have generated over 21,000 sets of spots, with an estimated total of 6.3 million classifications (expecting 300 spots per ROI, and only one ROI per set). These sets of data are being used extensively by researchers, research students and clinical trials.[11‑17] After publication, the clinically valuable data sets have been reused and collected in RandomSpotDB, which currently contains over 200,000 human labeled spot locations.

Figure 5 illustrates that nT has a direct effect on the time taken to process the ROI for all three types, which was confirmed by a Spearman’s rank order correlation coefficient test (i.e. Spearman’s rho). Tests revealed a statistically significant positive correlation between polygon (r = 0.9911, P < 0.001), square (r = 0.9945, P < 0.001) and circle (r = 0.9945, P < 0.001) ROIs, and nT. Effects of pT were found to be minimal above 15%, and Spearman’s correlation tests revealed a moderate negative correlation between polygon (r = −0.6829, P < 0.001) and pT, a slight positive correlation between square ROI (r = 0.1505, P = 0.0256) and pT, and a slight negative correlation between circle (r = −0.2733, P < 0.001) ROIs and pT.

DISCUSSION

The RandomSpot system provides a reproducible, objective and quantitative method of estimating tissue proportions within various types of ROIs on virtual slides. Digitally creating and storing randomly generated digitized grids for sampling of slides facilitates the inclusion of multiple grids per virtual slide, as well as allowing for grids of various shapes and sizes. The annotations can be reused by experimenters for a dual review or collaboration with remote sites, and the data can easily be exported into statistical software. The system itself has been designed for ease of use and minimal disruption to pathologist workflow. A simple performance evaluation was undertaken, which indicated that processing ROIs with target numbers of spots > 1000 will quickly become computationally prohibitive.

Applying SRS to digital slides offers the pathologist the benefits associated with digital pathology, such as saving and recommencing work, automatic navigation between points and facilitated collaboration. In using the RandomSpot system, the pathologist can generate quantifiable data, which accurately represents the frequency of tissue classes within a given ROI (compared to subjective visual estimations). However, manually inspecting each spot is still time consuming task, with each spot taking an experienced pathologist an estimated 5 s (25 min/300 spot case). Also, manually inspecting and classifying each spot still suffers from subjectivity, with mean pathologist agreement of 89%. Automation of the spot counting task is highly desirable. The RandomSpotDB has huge potential for recycling clinical data that is no longer needed by investigators, and subsequently the reuse of spot counting data sets will facilitate the development and validation of computer vision algorithms.[18]

RandomSpot is available at http://www.virtualpathology.leeds.ac.uk/RandomSpot

REFERENCES

1. Hyman BT, Gomez-Isla T, Frizzary MC. Stereology: A practical primer for neuropathology. J Neuropathol Exp Neurol 1998;57:305‑10.
2. Gundersen HJ, Bagger P, Bendtsen TF, Evans SM, Korbo L, Marcussen N, et al. The new stereological tools: Disector, fractionator, nucleator and point sampled intercepts and their use in pathological research and diagnosis. APMIS 1988;96:687‑81.
3. Mayhew TM, Gundersen HJ. If you assume, you can make an ass out of u and me': A decade of the disector for stereological counting of particles in 3D space. J Anat 1996;188:1‑15.
4. Cruz‑Orive LM, Weibel ER. Recent stereological methods for cell biology: A brief survey. Am J Physiol 1990;258:L148‑56.
5. Hill D. Practical Example: Placing a sampling grid in a polygon. MATLAB Central, 2008. Available from: http://www.mathworks.com/videos/2008/03/11/practical-example-placing-a-sampling-grid-in-a-polygon/ [accessed 2014 Oct 29].
6. Treanor D, Dattani M, Quirke P, Grabsch H. Systematic random sampling with virtual slides: A new software tool for tissue research. Abstr J Pathol 2008;216:1:543.
7. West NP, Dattani M, McShane P, Hutchins G, Grabsch J, Mueller W, et al. The proportion of tumour cells is an independent predictor for survival in colorectal cancer patients. Br J Cancer 2010;102:1519‑23.
8. West N, Grabsch H, Treanor D, Sebag‑Montefiore D, Thorpe H, Jayne D, et al. Quantitative assessment of tumor cell density in rectal cancer following three different preoperative therapies compared to surgery alone. Clin Oncol 2010;28:15s.
9. Downey CL, Simpkins SA, White J, Holliday DL, Jones JL, Jordan LB, et al. The prognostic significance of tumour-stroma ratio in oestrogen receptor-positive breast cancer. Br J Cancer 2014;110:1744‑7.
10. Wu Y, Grabsch H, Ivanova T, Tan IB, Murray J, Ooi CH, et al. Comprehensive genomic meta-analysis identifies intra-tumoural stroma as a predictor of survival in patients with gastric cancer. Gut 2013;62:1100‑11.
11. Elmoursi M, Treanor D, Simpson NA. Novel insights from using stereology
based volume estimation of syncytial nuclear aggregates in diabetic placenta. Arch Dis Child Fetal Neonatal Ed 2014;99:A158.

12. Elmoursi M, Stahlshmidt J, Treanor D, Simpson NA. Syncytial nuclear aggregates and villous capillary volume in IUGR placentas: A Stereology-based study on virtual slides. Placenta 2014;35:A11.

13. Hale MD, Nankivell MG, Mueller W, West NP, Stenning SP, Wright AI, et al. The relationship between tumour cell density in the pre-treatment biopsy and survival after chemotherapy in OE02 trial oesophageal cancer patients. J Clin Oncol 2014;32:49.

14. Earle S, Aoyama T, Wright AI, Treanor DE, Miyagi Y, Hewitt LC, et al. Prognostic and predictive value of tumor-infiltrating immune cells in Japanese patients with stage II/III gastric cancer. J Clin Oncol 2014;32:46.

15. Pettinger KJ, Ward L, Wright A, Treanor D, Grabsch HJ. Immune cell density is associated with tumour cell density, histological subtype, depth of invasion and outcome in gastric cancer. In Joint Meeting of the Pathological Society of Great Britain and Ireland and the Dutch Pathological Society-Winter Meeting; 2013. p. S21.

16. Toh E, Brown P, Botterill I, Quirke P. The proportion of tumour stroma as a prognostic factor in predicting lymph node metastasis in pT1 colorectal cancers. In American Society of Colon and Rectum Surgeons Annual Meeting; 2014. p. P198.

17. Toh E, Brown P, Botterill I, Quirke P. The proportion of tumour stroma as a prognostic factor in predicting lymph node metastasis in pT1 colorectal cancers. In The Association of Surgeons of Great Britain and Ireland International Congress; 2014. p. 0549.

18. Wright A, Magee D, Quirke P, Treanor DE. Towards automatic patient selection for chemotherapy in colorectal cancer trials. In Proc SPIE 9041. Medical Imaging 2014: Digital Pathology 2014. p. 90410A.