Dear A. Bistacchi,

We thank you for your time and effort reviewing the submitted manuscript, and are pleased that you appreciated our results. We have incorporated the majority of your suggestions into the revised manuscript, as detailed in the following pages. Please note that to facilitate the evaluation of our revision, the page and line numbers of the reviewer's comments refer to the originally submitted manuscript while page and line numbers of our responses refer to our revised manuscript.

Regards,

Samuel T. Thiele, Lachlan Grose, Anindita Samsu, Steven Micklethwaite, Stefan A. Vollgger & Alexander Cruden

Response to review by A. Bistacchi

General Comments:

The most important point, in my opinion, is that the "manual" and semiautomatic interpretations yield different results (e.g. Fig 2), hence a proper discussion in terms of false positives and undetected lineaments, with a matrix showing the results, must be included. This will allow the reader to form a better idea on the value of the proposed algorithm.

Clarify: The reviewer is correct that the manual and semi-automatic interpretations yield different results. This is largely due to the inherent variability of the interpretive decision processes applied when using these types of datasets – differing results would also be expected on manually digitising the same dataset multiple times. As discussed at lines 7.28-7.34, our method improves reproducibility by rigorously and consistently "interpolating" between points of interpretation, however results still depend on these interpretive decisions, hence the "natural" variation.

Regarding false-positives and undetected lineaments, our method does not do any "lineament detection" as it relies on the user to pick each lineament (by inserting the control points). Thus, the approach is subject to identical "false-positives" and "undetected-lineaments" as manual interpretations, something that we feel is outside the scope of this paper.

Situations where the semi-automatically generated trace does not match the user's expectations (i.e. they expected the algorithm to follow a different lineament) could be considered a false positive, though the inclusion of additional constraints (intermediate "waypoints") can be used to easily modify the results to meet expectations. This is discussed in the text at lines 8.26-8.30, with a specific example provided in Fig. 6.

Specific Comments:

31-32: paragraph not necessary

Agree: The paragraph-break has been removed. [Line 1.31]

33: "virtual" -> consider "digital" more used nowadays

Agree: "virtual" has been replaced with "digital" throughout the text.

33: consider using more proper references, both older ones that introduced the digital outcrop concept and newer ones with detailed studies

Agree: The reference has been updated as suggested, and now includes references to McCaffrey et al. (2005), Pringle et al. (2006) and Vollgger and Cruden (2016). [Line 1.30]

34: this could be not the best reference for photogrammetric workflow

Clarify: Bemis et. al. provides a useful and widely acknowledged review of the photogrammetric workflow as applied to structural geology and seismology. Hence, we believe that it is an appropriate reference in this context. To broaden the scope somewhat we have included an additional reference to Smith et al. (2015). [Line 1.31]

36: 2.5D should be better defined

Agree: The sentence has been rephrased as follows:

"While this method works well in topography containing slopes <45°, it is inherently limited to 2.5 dimensions (2.5D) as elevation data is gridded in the X-Y plane, causing problems when features crosscut steep or overhanging topography (Pavlis and Mason, 2017)" [Line 2.33-2.35]

23: add reference to Fig. 1a

Agree: The text has been modified such that each part of the figure is now explicitly referenced:

"In the first step, data points (points in a point cloud or pixels in an image; Fig. 1a) are linked with their nearest neighbours, using a spherical search radius slightly larger than the dataset resolution, to produce a neighbourhood network (Fig. 1b). The costs of moving along links in this network (hereafter referred to as "edges") are then calculated, using a cost function designed to promote movement along structural or lithological traces and inhibit movement across them (Fig. 1c)." [Lines 3.20-3.24]

24: this is Fig. 1b

Agree: This has been corrected (see above).

27-32: some equations might help the reader here

Clarify: Dijkstra's algorithm is a recursive path-growing procedure, and so cannot be simply expressed as an equation. Hence, we suggest that attempting to do so would only confuse the reader. Furthermore, detailed descriptions of the method and its implementations are readily available, including in the cited Dijkstra et al., 1959.

37: even if you refer to the appendix, please list and briefly describe the cost functions here - this is a key point

Agree: The cost functions used in this case study are now listed and described in the text:

"… simple cost functions such as point/pixel brightness or local colour gradient work well on most geological datasets; the examples presented below all map a single scalar attribute in the dataset directly to cost (point/pixel brightness for Study 1 and 2, topographic slope for Study 3 and bathymetric depth and vertical gravity gradient for Study 4). We have designed and implemented these and several other simple cost functions that give reasonable results for different structure and data types. Specific equations for these are included as Appendix 1." [Lines 3.34-4.2]

For clarity and brevity we have kept the specific equations in the appendix.

4-14: have you used third party libraries or have you written all the code? this must be clearly stated here.

Agree: The Compass plugin uses no third-party libraries other than CloudCompare and its dependencies. The GeoTrace implementation uses numpy data structures and scikit-image functionality. The text has been amended to reflect by adding the following:

"Our CloudCompare plugin (Compass) is written in C++ and works for point cloud data, while the QGIS version (GeoTrace) is implemented as a python plugin, using numpy (van der Walt et al., 2011) and scikit-image (van der Walt, 2014) to apply our method to raster data" [Lines 4.6-4.8]

24: specify which software you use and how many photos in model

Agree: The text has been updated as follows: "*… applying a Structure from Motion-Multi-View Stereo (SfM-MVS) workflow, as implemented in Agisoft Photoscan v1.2.6 to 297 digital photographs …*" [Lines 4.22-4.23]

29: Melbourne, Australia,

Agree: The text has been updated to specify that Melbourne is in Australia:

"*The Cape Woolamai sea stacks, located approximately 115 km southeast of Melbourne (Australia) on Phillip Island, have formed by erosion of …*" [Lines 4.32-4.33]

35-36: specify software, camera resolution, focal length etc. (as above)

Agree: The text has been updated to include this information:

"*For this study, a DJI Inspire 1 multi-rotor UAV and 20 mm fix-focal Zenmuse X3 camera were used to capture 274 aerial photographs, which were subsequently processed using Agisoft Photoscan.*" [Lines 5.1-5.2]

37: 2 million points is not so much with modern photogrammetry software. have you filtered the dataset? in any case please explain.

Agree: The 2 million points represents a small subset of a much larger survey, hence the relatively small number of points. To clarify the text has been modified as follows:

"A 45 × 40 × 25 m section of the resulting model containing a single sea-stack was then extracted, containing ~2 million points and representing an average ground sampling distance of ~2.5 cm/pixel." [Lines 5.2-5.4]

4: LIDAR-derived

Agree: As suggested, a hyphen has been added to the manuscript. [Line 5.8]

29: reference for Sobel filter

Agree: An appropriate citation to Sobel (1990) has been included [Line 6.1]

31-32: explain closest-point difference - this is a key point

Agree: The manuscript has been amended to describe the closest-point difference calculation: *"Closest-point differences, calculated by subsampling closely-spaced points from each assisted trace and computing the shortest-distance between these and a manually interpreted trace, show that…"* [Line 6.12 – 6.14]

9: explain why you get different orientation estimates. probably a discussion in terms of false positives and undetected lineaments will be very interesting.

Agree: The differing number of orientation estimates reflect the different number of traces digitised manually (146) and with the computer assisted method (114). The difference in the resulting stereoplots (Fig. 3.) probably results from the larger number of points in the assisted traces (which sample every point along each trace) as compared to manually digitised traces: Manually digitised traces only contain points picked by the user during digitisation, and hence will be more affected by outliers. The text has been updated to reflect this: "*Although far from conclusive, this indicates that the computer-assisted approach improves the consistency and precision of the orientation estimates, likely due to the larger number of points it samples. The assisted method samples every point along each trace, while the manual method only includes the polyline vertices created during digitisation, making the best-fit plane more susceptible to errors caused by outliers.*" [Lines 6.30-6.34]

15-21: this case study is not described in details as the other ones. if it is not important, consider deleting it, otherwise add a description as detailed as or the others.

Clarify: The Greendale Fault case study demonstrates an application of our method to LiDAR data, which is commonly used throughout the geosciences. Hence we feel that it is worth retaining even if it is not discussed in detail. Furthermore, a detailed description of this dataset is not necessary as: (1) the results are quite similar to those from the Bingie Bingie case study, and hence can be communicated succinctly, and; (2) a full interpretation and description of this dataset has been previously published by Duffy et al., 2012, as mentioned in the manuscript.

Dear T. Scheiber,

We thank you for your time and effort reviewing the submitted manuscript and for your constructive suggestions. We have incorporated the majority of these into the revised manuscript, as detailed in the following pages. Please note that to facilitate the evaluation of our revision, the page and line numbers of the reviewer's comments refer to the originally submitted manuscript while page and line numbers of our responses refer to our revised manuscript.

Regards,

Samuel T. Thiele, Lachlan Grose, Anindita Samsu, Steven Micklethwaite, Stefan A. Vollgger & Alexander Cruden

General Comments

(I) Missing description of the manual extraction method

While the semi-automatic method is well-documented and is proven to be scaledependent, the manual method is not explained at all. There is the need for a description on how the manual mapping was performed, since it is widely known that results obtained by manual extraction are strongly dependent on various factors (cf. e.g., Scheiber et al., 2015). Especially the scale during mapping and the fractal dimension of the dataset is of great importance for this study.

Agree: The reviewer is correct that our description of the manual digitisation method used is inadequate. This has been rectified by adding two paragraphs describing the manual methods to the end of Section 4. These are as follows:

"For each of these case studies, we also assess the similarity of our assisted results to manually derived interpretations. The aim of these comparisons is not to rigorously validate the computer-assisted method as: (1) given enough control-points our method could match any interpretation, and; (2) manual interpretation of structures from remotely sensed data is notoriously subjective, and so differing results can be expected from different operators (Bond et al., 2007) and scales (Scheiber et al., 2015). Instead, we seek to demonstrate the applicability and versatility of the assisted approach, and that similar results to a manual interpretation can be produced for less time and effort.

For each interpretation, the operator was instructed to digitize every structural feature within the dataset. To ensure this was an achievable task, the extent of the dataset used in each case study is small compared to its resolution. No attempt was made to ensure that the same number of features was extracted from each dataset, as this would affect timing measurements. Digitization was performed at or close-to the dataset resolution, so that the manual interpretations contained similar detail to the assisted approach (which always follows traces at the resolution of the dataset), although the operator was able to freely zoom in and out to inspect the data at multiple scales. The same operator performed the manual and assisted interpretations for Case Studies 1 and 2, while different operators generated each of the assisted and manual interpretations for Case Study 3 and 4. Manual interpretation in QGIS (Case Study 1, 3 and 4) was performed by digitising polyline features to a shapefile using the "Add feature" tool, while in CloudCompare (Case Study 2) the "Trace-polyline by point picking" tool was used." [Line 5.16-5.32]

The larger the scale, the more details will be recognized by the mapper and, as a consequence, the more turns/curves will a structural trace have. Larger scales would thus probably lead to more similar results of the manual extraction method compared to the new semi-automatic method. In this respect, a comparison of the data by closest-point difference calculation is problematic, unless it is clearly specified under which conditions (especially the fixed scale) the manual mapping was conducted.

Clarify: Each of case studies chosen in this paper represent small subsets of much larger datasets, and hence the interpretations were performed over a limited range of scales at or close-to the dataset resolution. The assisted-method always "follows" traces at the largest possible scale (i.e. at the data resolution), and hence will be scale-independent for individual traces (though the user-decision on what comprises a single-structure/trace will be scale dependent). Thus, we suggest that the comparison between the two (automatic vs assisted) is reasonable, especially given the limited and general conclusions we draw from these.

This has been clarified on lines 5.24– 5.28 (see response to previous point).

Finally, we are unsure why the reviewer considers that larger-scales would lead to smaller differences between the assisted and automated approaches – obvious errors in the digitisation with the assisted approach are corrected by the user during digitisation (as discussed at lines 8.26 –8.30), reducing the incidence of large errors at small scales.

(II) Comparison of data

A large part of the paper, especially the result section deals with the comparison of manually-extracted data and data derived from the new semi-automatic approach. The data are compared on the basis of digitization time, number of traces, mouse-clicks (Tab. 1) and by pixel distances between constructed lines (closest-point difference, cf. Figs.2, 4,5). In the case studies the authors state that the results obtained by the two different methods are "visually similar" , "broadly consistent", "very similar" and show "similar accuracies". These phrases are way too qualitative - what is "very similar", and what would "less similar" be then?

Clarify: The phrases are indeed qualitative, though the authors suggest that is reasonable given: (1) the inherent variability of manual data interpretation (as correctly pointed out by the reviewer, and described in lines 5.18-5.20), and; (2) our aims from the comparison are only to (qualitatively) demonstrate that results similar to a manual interpretation can be produced with the assisted-method (rather than a robust validation; now stated at lines 5.20-5.22)

Furthermore, we include both quantitative histograms of the closest-point distance between the two interpretations for each Case Study (Figs. 2, 4, and 5) and the percentage of the traces that fall within threshold distances of 2 and 5 pixels. These comparisons are entirely quantitative and are provided throughout the text (lines 6.14-6.15, 7.4-7.5, 7.9 and 7.16-7.17) as justification for the qualitative statements the reviewer has identified.

Having a closer look, however, the results of the two methods appear "not similar" to me: traces in one interpretation are longer than in the other (i.e., they have different start and end points), there are traces drawn in one interpretation and are missing in the other, and there are traces which consist of two segments in one interpretation and appear as one trace in the other, and vice versa (see Figs. 2-5). It is not clear how the authors handled these mismatches, especially when applying a closest-point difference calculation (a description of this comparison technique is missing).

Clarify: The reviewer is correct that the number and length of traces in each interpretation differ. This is an expression of the natural variability of the interpretation processes and, as now clearly stated at line 5.25, no attempt was made to reduce this (by, for example, enforcing traces to have the same start/end points) as doing so would significantly affect the timing comparisons.

Such variations in the dataset were not corrected for in the closest-point difference calculations as this technique does not require exactly analogous traces (as now described on lines 6.12-6.14). This is the reason the histograms in Figs. 2, 4 and 5 have such long tails. In Case Study 1, these were clipped at a large distance to remove points from traces that did not have equivalents in both datasets, as described in the captions of Fig. 2.

The only sound comparison presented in the manuscript would be the one shown in Fig. 5f, where start- and endpoints of a previous manual interpretation are used to recalculate the structural traces using the new tool. However, in this example the authors chose to use a different (updated) base map [Sandwell and Smith (2009) in Fig. 5c versus Sandwell (2014) in Fig. 5d; p. 6, lines 27-31]. This complicates a comparison, because the resulting differences stem from both the different basemaps and the different extraction methods, and the influence of each of these variables cannot be quantified. A proper comparison of manually- and semi-automatically produced datasets needs to fulfill the following requirements: (1) Different methods have to be tested on similar basemaps. (2) A thorough description of the fully manual method, especially by indicating the scale used while mapping is necessary. (3) The exactly same start and end point have to be used for each constructed trace. This is because the start- and end-point of the semi-automatic approach are defined manually as well! Thus it makes only sense to compare the actual traces (having the same start- to an endpoints) to figure out the differences of the fully-manual and your new method. As a consequence, the number of traces of the two opposing interpretations has to be the same. These requirements are not fulfilled in the presented case studies and should be considered when revising the manuscript.

Clarify: The method the reviewer suggests for comparing the assisted and manual approaches using fixed start/end points for each trace would produce interesting results, however this approach would make the timing/mouse-click measurements meaningless, as mentioned at line 5.25. Furthermore, one of the key advantages of our computer-assisted approach is that a user can iteratively modify/correct results in real-time (cf. lines 8.26 –8.30), making the "accuracy" of the results essentially dependent on the effort spent by the user on quality control and making such a rigorous validation largely arbitrary. Hence we suggest that such a comparison is outside the scope of this paper, which is focused on presenting our new method and demonstrating ways it can be applied.

We make this clear by stating "*The aim of these comparisons is not to rigorously validate the computer-assisted method as: (1) given enough control-points our method could match any interpretation, and; (2) manual interpretation of structures from remotely sensed data is notoriously subjective, and so differing results can be expected from different operators (Bond et al., 2007) and scales (Scheiber et al., 2015). Instead, we seek to demonstrate the applicability and versatility of the assisted approach, and that similar results to a manual interpretation can be produced for less time and effort.*" [Lines 5.17-5.22]

The purpose of Fig 5c. is to show that our method can be used to translate interpretations between datasets (as discussed at lines 7.13, 7.35-8.4), hence the use of different datasets.

Digitization time and number mouse clicks will vary significantly between users, as correctly stated by the authors (p. 7, 30/31). So if a comparison based on these numbers is presented, it has to be clearly stated that the exercises were conducted by one and the same user.

Agree: The text now clearly states that "*The same operator performed the manual and assisted interpretations for Case Studies 1 and 2, while different operators generated each of the assisted and manual interpretations for Case Study 3 and 4.*" [Lines 5.28-5.30].

(III) Practical application of the GeoTrace plugin

When installing the plugin (MS Windows), it would be much more user-friendly, if the installation would run automatically. Up to now users need to install and run several codes and files in python in

order to make the plugin work (see detailed instructions for MS Windows users on https://github.com/lachlangrose/GeoTrace). There must be an easier, more user-friendly solution.

Clarify: The difficulty installing the GeoTrace plugin on Windows results largely from software dependencies of the scikit-image package used, for performance reasons, to perform the shortest-path calculations. We are currently working to create an installation script/file, however unfortunately the security features of windows and the need for admin privileges make this no easy task. A feature-request has been posted on the GitHub page to record progress.

I have tried to run the method on a .dem file and an ovr file. While it worked out well with the .dem file, I got error messages and couldn't use the plugin with the .ovr file.

Agree: .ovr files are a "pyramid file", containing pre-computed data tiles at different scales for rapid visualisation. This causes issues in the plugin, as it requires easy access to data at full-resolution, hence the error messages. We have modified the plugin so that it throws a "friendly" error message when .ovr files are used. QGIS can easily be used to convert .ovr files to a compatible data format.

"Fit planes", "Stereonet" and "Rose": I tested the tool by tracing straight and sharp bedrock lineaments, which correspond to subvertical fault and fracture zones. In case the "fit planes" box is activated, the columns called DIP and DIP_DIR in the attribute table are filled with calculated values for each trace. The values in these columns, however, appear to be incorrect: In the column DIP_DIR occur numbers which do not represent dip direction (they even include negative values), but these values appear to be arbitrary. For the dip (column DIP), I got values between 0 and 10 degrees, which is obviously wrong as well. This issue needs to be fixed and the methods used for the calculation of dip and dip direction need to be explained in the manuscript (or in the appendix).

Clarify: We have tested the "fit-planes" functionality on several datasets, with reasonable results. We suggest that the "arbitrary" orientations and shallow dip values might result because the best-fit-plane is sub-parallel to topography (i.e. variation in the structure orientation is resulting in an incorrect best-fit-plane). Poor-quality fits can also be obtained from highly co-linear traces. If this issue persists we would appreciate it if the reviewer would add a bug-report to the GitHub page that includes further information on the situation in which the issue arose.

To help clarify the quality of the orientation estimates, the plugin has been modified to calculate the "planarity" metric described by Thiele et al., 2015, as well as a crude classification of fit plane quality based on this. A description of this, and the plane-fitting method used (Eigen analysis), has been added to the plugin documentation. As plane-fitting is outside the scope of this publication (which focusses on the least-cost-path method for trace rapid trace digitization) we have not included this in the manuscript.

Agree: The plugin no longer produces negative (i.e. anti-clockwise) dip-direction bearings.

The tabs called "Stereoplot" and "Rose" allows for directly plotting the structural data. To make the tool more user-friendly, I suggest to predefine and fix the fields "Direction" and "Dip" to the columns DIP_DIR and DIP in both tabs. In both these tabs, it is not clear what the checkbox "Dip Direction" means and does.

Agree: The requested features have been included in the plugin. The stereonet tab has been modified to make it clear that by checking "Dip-Direction" orientations are expected in the format Dip/Dip-Direction rather than Strike/Dip.

For the stereonet it needs to be indicated if the plot corresponds to lower/upper hemisphere and equal area/angle. And for the "Rose" tab, it has to be stated if the plot is number/area-weighted. In the manuscript, as it is now, no stereoplots and no rose diagrams produced by the GeoTrace tool are shown.

Agree: The requested information is now included on the stereonet/rose-diagram plots.

The authors need to provide orientation data plotted in stereoplots and rose diagrams (produced by GeoTrace) for at least one case study to show the functionality and the full output spectrum of their tool.

Clarify: The purpose of this publication is to present the least-cost-path method for trace digitisation. The functionality for drawing stereoplots/rose-diagrams, included for convenience in the GeoTrace plugin, is not scientifically novel and is outside the scope of this paper.

Specific comments

Abstract

7-9: The first sentence of the abstract is far-fetched, especially when regarding the fact that mapping at this time concentrated on lithologies and lithological boundaries, not specifically structures. I therefore suggest to delete this sentence.

Agree: The sentence has been removed. [Line 1.7]

12: "...extract..." There is the need to add a sentence describing how your new extraction method practically works - you manually add a start and end point between which a connecting line is calculated, which can be manually tuned by adding additional points in between.

Agree: Line 1.9 has been modified to reflect this, and now reads: "*Here we adapt a least-cost-path solver and specially tailored cost-functions to rapidly interpolate structural features between manually defined control points in point cloud and raster datasets.*" [Line 1.9-1.10]

21: "The approach improves the objectivity and consistency..." Since start and end points are set manually, I don't really see how this method then improves objectivity 21: change "expert" to "user". This accounts for all places where "expert" is used in the manuscript. Not every user can be considered an expert.

Clarify: The method improves objectivity in that a significant portion of the trace length is directly guided by the underlying dataset (e.g. Fig. 7), and so will consistently be reproduced given the same input/control points. The reviewer is correct, however, that as the user chooses the start and end-points (and can add intermediate points to "force" the trace to do almost anything), so the approach is certainly not completely objective (as discussed in detail at lines 7.25-7.34).

To avoid implying objectivity without the associated caveats included in the discussion, the sentence has been modified to:

"*The approach improves the consistency of the interpretation process while retaining expert-guidance, and achieves significant improvements (35-65%) in digitisation time compared to traditional methods*" [Line 1.19].

Also note that we retain the term "expert-guidance" as: (1) it is used elsewhere in the literature (e.g. Vasuki et al., 2014), and; (2) we consider that the "user" knows best what they want to extract from the dataset, making them an "expert" by comparison with our humble computer program.

23-24: "it... can quantify the agreement between datasets and interpretation." Unclear. What interpretation (manual or semi-automatic)? How is this practically done? - In the GeoTrace plugin, I couldn't find any tool to compare data - ?

Clarify: This statement relates to the method we present rather than its specific implementations (GeoTrace and Compass), and is a summary of the discussion points presented in lines 8.5 – 8.14. The authors are of the opinion that the sentence is clearly a general statement about the approach rather than a specific reference to either of the plugins, and hence it has been retained.

Introduction

33: add e.g., to the references; refer only once to Bemis et al., 2014

Agree: The references at line 1.29 have been updated as per the comment of Reviewer #1, and no longer refer to Bemis et al., 2014.

35 and following: rephrase this sentence

Agree: The sentence has been rephrased as follows:

"*These techniques, combined with inexpensive and easy-to-use UAV technology, now make it feasible to acquire topographic data at mm-to-cm resolution over areas of several square kilometres (e.g., Vollgger and Cruden, 2016; Cruden et al., 2016), providing for the first time an objective method for rapidly collecting detailed 3D information on geological structures.*" [Line 1.31 – 1.34]

1: something is missing between easy and to

Agree: The missing word has been added (see above).

11: remove "and contacts". Contacts are geological structures

Agree: "and contacts" has been removed as suggested. [Line 2.9]

13: in brackets: name only the sort of data you used in your case studies.

Agree: The listed datasets now explicitly relate to data types presented in this study:

" … 2D gridded datasets (imagery, geophysical rasters etc.) and dense 3D point clouds (digital outcrop models)." [Line 2.11]

Existing Methods

Add a short description of the fully manual method here as well and refer to its drawbacks regarding objectivity. Refer to Scheiber et al. (2015). Scheiber, T., O. Fredin, G. Viola, A. Jarna, D. Gasser, and R. Łapi´nska-Viola (2015):Manual extraction of bedrock lineaments from high-resolution LiDAR data: methodological bias and human perception, GFF, 137, 362-372, doi: 10.1080/11035897.2015.1085434.

Agree: Please refer to our response to General Comment I

16: write more specific: "outcrop structures" instead of "outcrops"

Agree: The sentence has been amended as suggested. [Line 2.13]

18/19: refer to more recent papers as well

Agree: Additional references to more recent papers (Holden et al., 2012 and Masoud and Koike, 2017) have been included as suggested [Line 2.17]

33 onwards: explain the dimensions 3D versus 2.5D

Agree: The sentence has been rephrased as follows:

"While this method works well in topography containing slopes <45°, it is inherently limited to 2.5 dimensions (2.5D) as elevation data is gridded in the X-Y plane, causing problems when features crosscut steep or overhanging topography (Pavlis and Mason, 2017)" [Line 2.34-2.35]

35/36: what do you consider a simple topography? Be more precise.

Agree: "simple topography" has been replaced with "topography containing slopes <45°". [Line 2.34]

2: refrain from using phrases such as "such as described above". Follow this suggestion throughout the manuscript.

Agree: The sentence has been reworded to: "*Unfortunately, the unstructured nature of 3D point data mean that methods for trace detection in raster data, including those previously described, cannot be easily applied.*" [Line 2.37-3.38]

4: use the more general term "fractures" instead of "joints"

Agree: "joints" have been replaced with "fractures" as suggested. [Line 6.2]

Method

14: "DEM based plane fitting and orientation analysis" didn't work (see general comment III)

Clarify: General comment III is dealt with earlier in this response.

16: To demonstrate the capability of our new computer-assisted trace detection approach,

Agree: The sentence has been rephrased as suggested to avoid using "described above". [Line 4.18]

18: "established manual methods" - A thorough description of the manual method is of great importance as well: Who mapped? At which scale was the mapping performed (see general comment I)?

Agree: Please refer to our response to General Comment I

21: "plutonic diorites and tonalites" - either "plutonic rocks" or "diorites and dolerites" instead

Agree: The sentence has been rephrased as suggested *"… Cretaceous to Paleogene dykes intruding Devonian diorites and tonalities…"* [Line 4.23]

28: use "fractures" instead of "joints" throughout

Disagree: The term "joint" is widely used in the structural geology literature, and conveys useful information (e.g. small displacement, Mode I opening) as opposed to the more generic "fracture". Hence in this specific context (i.e. the small displacement, Mode I fractures observed in the Bingie Bingie and Cape Woolamai datasets) we use the term "joint".

33: "Several sets of systematic and non-systematic joints..." if a joint/fracture is nonsystematic, it doesn't belong to any set.

Agree: "several *sets of joints*" has been replaced with "several *generations of joints*" to avoid this issue. [Line 4.36]

34: "cooling of the intrusion, subsequent deformation and recent unloading." Are these guesses? If not add references, otherwise remove it. What does "recent" mean in geological time scale? Be more precise.

Clarify: These are interpretations based on field-observations and context (and the fact that there are few other processes that generate joint-sets), hence the use of the term "likely". The sentence has been retained (rather than removed as suggested) as the authors feel it provides useful context to the case-study. The timing of unloading is largely poorly constrained (though probably neotectonic), and hence "recent" has been removed.

1: "...accurate orientation measurement" - of what?

Agree: The sentence has been amended to specify that joint-orientation is being measured: "… *allows for accurate joint-orientation measurement …*" [Line 5.5]

Results

14: "manual workflows" - the manual extraction method needs to be described for each case study.

Agree: See response to General Comment I

18: "...compare and contrast the results of both manual and assisted interpretations..." The result sections dealing with comparison of the data have to be rephrased in line with newly obtained results (see general comment II). And the closest-point difference calculation needs to be explained in the method section.

Agree: We respond to General Comment II earlier in this response. The manuscript has been amended to describe the closest-point difference calculation:
"Closest-point differences, calculated by subsampling closely-spaced points from each assisted trace and computing the shortest-distance between these and a manually interpreted trace, show that…" [Line 6.12 – 5.14]

29: remove "previously mentioned"

Agree: The words were unnecessary and have been removed.

31-33: "The results are visually similar..." - see general comment II

Clarify: This point is addressed in the response to General Comment II.

1: "As in the previous example,..." Name it.

Agree: The example has now been explicitly named: *"As in the Bingie Bingie example, …"* [Line 6.19]

3: What does this difference in numbers of traces reflect? (see also comment to table 1 and general comment I and II)

Clarify: The difference in the number of traces interpreted reflects inherent variability of manual interpretation. This point is discussed in detail in our response to General Comment I and III.

8: "as a post processing step..." shorter: "after processing"

Agree: Shorter is better!

15: "fault scarps" In the LiDAR DEM (Fig. 4a) the interpreted fault scarps (Figs. 4b and c) are not visible at all. The slope map which was used for doing the interpretation has to be provided in Fig. 4.

Agree: The "slope-map" has been included in Fig. 4 as suggested.

27: Replace "described in" by "interpreted by"

Agree: "described" has been changed to "interpreted" [Line 7.11]

33 onwards: Here it is unclear whether the errors refer to the 21% pixels located >5 pixels away. "the computed shortest-path ... would "detour" through..." Does this mean that you reduced these errors by adding control points to "guide" the trace? Clarify

Agree: The quoted error values refer to unmodified traces. This has been clarified by adding "would" to the sentence: *"A small number of additional control points along these traces would resolve this issue…"* [Line 7.19]

36 "correct fracture zone" A fracture zone cannot be correct or incorrect. Change to for example "desired structural trace"

Agree: "Correct" has been replaced with "desired" as suggested. [Line 7.20]

Discussion

3: DEM: use plural: DEMs

Agree: The plural has been used. [Line 7.24]

6: "operates in co-operation" Rephrase

Agree: The sentences has been rephrased as follows to avoid the repetition:

"*… works in co-operation …*" [Line 7.27]

6/7: change to "user guidance"

Disagree: The authors suggest that "expert-guidance" is an appropriate term here (see response to the previous comment regarding "expert-guidance").

8: if there is a firstly, there has to be somewhere a secondly as well

Agree: Secondly has been added to the following paragraph ("*Secondly, similarly to*") [Line 7.35]

12: "improved consistency" Yes there is improved consistency, but it doesn't necessarily mean that it increases precision and that it is closer to reality. This remains a bit abstract unless a reference to the true (nature) pattern of the fractures is given.

Agree: Hence the use of "precision" rather than "accuracy" in line 7.32.

30 "...manually interpret datasets using GeoTrace or Compass..." this is confusing. GeoTrace and Compass are semi-automatic tools. I suggest to use "manual" for the fully-manual extraction method and "semi-automatic" or "computer-assisted" for your new extraction tool consistently throughout the manuscript.

Agree: The use of "manually" here was incorrect and unnecessary, and so has been removed. The sentence now reads: "*The time it takes for users to interpret datasets using GeoTrace or Compass will vary*" [Line 8.14]

34/35 cf. general comment II

Clarify: Please see our response to General Comment II.

6: add "fully" to "automated"

Agree: "automated" has been replaced with "fully-automated". [Line 8.29]

Tables and Figures

Table 1:
Why do you get different numbers of traces for the manual and the assisted method? Is this due to different operators for each method? Or does this reflect different perception states of the same operator? What does the comparison in numbers tell us here? Cf. general comment II.

Clarify: The difference in the number of traces interpreted reflects inherent variability of manual interpretation. This point is discussed in detail in our response to General Comment I and III.

Figures 2-5 should be considered to get redrawn/rearranged according to the general comment II and the comments below.

Figure 2: North arrow is missing in both (a) and (b). Here you could add stereoplots and rosediagrams produced by GeoTrace (cf. general comment III).

Agree: A north arrow has been added to Fig. 2a and 2b as suggested.

Figure 3: North arrow is missing in both (a) and (b). What is the reason for the more consistent orientation estimates using the computer-assisted approach (Fig. 3f)? Is this an artefact of the computer-assisted approach? Is the larger spread in orientations in the manual-extracted dataset more realistic? What is closer to reality? See also comment to page 7, line 12. Discuss. caption: get rid of the "similar results" (see general comment II).

Clarify: Figure 3a includes a lat/long grid, and hence a north arrow is unnecessary and will only clutter an already busy figure.
Agree: A north arrow has been added to 3b and 3d for consistency with 3c.

Figure 4: North arrow is missing. Scale bar: show it from 0 to 1000m only, don't go to -250. What are the straight lines across the DEM? Roads or tractor tracks? Indicate. Show the map you used for extraction of structural traces. In the un-interpreted hillshaded DEM (Fig. 4a) I would not dare drawing any of the interpreted traces. Show the slope map, which was used to extract the traces.

Clarify: As above, a north arrow is unnecessary as the figure includes a lat/long grid.
Agree: The scale bar has been fixed as suggested. We are unsure what causes all of the "straight lines" across the DEM, though they could quite possibly be farm-tracks or patterns resulting from tractors. Regardless, the lines are not relevant to this publication.

Figure 5:
Legend for 5a/b (bathymetry?) and for 5c/d (vertical gravity gradient) is missing.

Agree: The missing legends have been added to the figure.

The original/uninterpreted data (basemap) needs to be shown for 5a/b. You did the mapping probably at larger scales? If yes, provide zoom-in maps showing exemplary oceanic fracture zone and their interpretation. Indicate location of Fig. 6 in Fig. 5c or 5d.

Agree: The original/uninterpreted bathymetry has been shown suggested. The location of Fig. 6 has been included.

Caption: remove the last sentence (see general comment II)

Clarify: This point is discussed in detail in our response to General Comment II.

Add scale and north arrow.

Disagree: As in the previous figures, Fig. 5 a-d contain lat/long grids, hence a north arrow is unnecessary and distracting. Also over this extent, a scale-bar would be misrepresentative due to map distortions, hence the use of a lat/long grid.

Technical corrections:

1: "sea-stacks" - in other places written without hyphen. Be consistent.

Agree: "Sea-stacks" (with a hyphen) is now used consistently throughout the manuscript.

Rapid, semi-automatic fracture and contact mapping for point clouds, images and geophysical data

Samuel T. Thiele[1], Lachlan Grose[1], Anindita Samsu[1], Steven Micklethwaite[1], Stefan A. Vollgger[1], Alexander R. Cruden[1]

[1]School of Earth, Atmosphere and Environment, Monash University, Melbourne, 3800, Australia

Correspondence to: Samuel T. Thiele (sam.thiele@monash.edu)

Abstract. ~~Two centuries ago William Smith produced the first geological map of England and Wales, an achievement that underlined the importance of mapping geological contacts and structures as perhaps the most fundamental skill set in earth science.~~ The advent of large digital datasets from unmanned aerial vehicle (UAV) and satellite platforms now challenges our ability to extract information across multiple scales in a timely manner, often meaning that the full value of the data is not realised. Here we adapt a least-cost-path solver and specially tailored cost-functions to rapidly ~~extract and measure~~interpolate structural features ~~from~~between manually defined control points in point cloud and raster datasets. We implement the method in the geographic information system QGIS and the point cloud and mesh processing software CloudCompare. Using these implementations, the method can be applied to a variety of three-dimensional (3D) and two-dimensional (2D) datasets including high-resolution aerial imagery, ~~virtual~~digital outcrop models, digital elevation models (DEMs) and geophysical grids.

We demonstrate the algorithm with four diverse applications, where we extract: (1) joint and contact patterns in high-resolution orthophotographs; (2) fracture patterns in a dense 3D point cloud; (3) earthquake surface ruptures of the Greendale Fault associated with the M_w7.1 Darfield earthquake (New Zealand) from high-resolution light detection and ranging (LiDAR) data, and; (4) oceanic fracture zones from bathymetric data of the North Atlantic. The approach improves the ~~objectivity and~~ consistency of the interpretation process while retaining expert-guidance, and achieves significant improvements (35-65%) in digitisation time compared to traditional methods. Furthermore, it opens up new possibilities for data synthesis and can quantify the agreement between datasets and an interpretation.

1 Introduction

Remote sensing datasets are commonly used in the earth sciences to interpret the morphology, location, timing and orientation of geological features. These data types, now routinely delivered by satellite, aerial and UAV platforms, have advanced to the point where they are widely available at high-resolution, and in some instances frequently updated. This proliferation of data has led to a situation where it is now no longer practical to use manual methods to extract geological information, meaning that the full geological value of high-quality datasets is often not extracted.

For example, high and ultra-high resolution (cm to mm) photorealistic reconstructions of geological outcrops ("~~virtual~~digital outcrop models") are becoming widely available (~~Bemis~~McCaffrey et al., ~~2014; De Paor~~2005; Pringle et al., 2006; Vollgger and Cruden, 2016), typically acquired using either laser scanning technology (cf. Buckley et al., 2008) or photogrammetric workflows (cf. Bemis et al., 2014~~). It is~~; Smith et al., 2015). These techniques, combined with inexpensive and easy-to-use UAV technology, now make it feasible to ~~use these~~

techniques to captureacquire topographic data at mm-to-cm resolution over areas of several square kilometres at mm-to-cm resolution using off-the-shelf and easy to UAV technology (e.g., Vollgger and Cruden, 2016; Cruden et al., 2016), providing for the first time an objective method for rapidly collecting detailed 3D information on geological structures.

There has recently been significant effort to develop automatic or computer-assisted methods for digitising structural data, in particular from orthorectified photographs or image sequences (Seers and Hodgetts, 2016; Vasuki et al., 2014; Jones et al., 2009). Achieving satisfactory automated digitisation is challenging for the mapping of geological structures due to intrinsic variables such as geometry, soft-linkage and segmentation over multiple scales, as well as extrinsic variables such as natural variations in colour, shadows, glare, and/or incomplete geological exposure. Due to this complexity, fully automatic methods often require significant manual adjustment and vetting to remove false positives while retaining real geological features (Vasuki et al., 2014; Seers and Hodgetts, 2016).

In this paper, we first review existing approaches to the mapping of geological structures and contacts from digital data, and then describe a novel least-cost path method that can "follow" structure traces and lithological contacts between user-defined control points in both 2D gridded datasets (photographsimagery, geophysical imageryrasters etc.) and dense 3D point clouds (virtualdigital outcrop models). We then describe its implementation in two widely used software packages (QGIS and CloudCompare), and introduce four applications demonstrating the efficacy of the method for mapping outcropsoutcrop structures, earthquake surface ruptures and oceanic fracture zones.

2 Existing methods

Many automated methods have been developed to extract linear features in the geosciences (e.g., Jinfei and Howarth, 1990; Tzong-Dar and Lee, 2007; Jinfei and Howarth, 1990Holden et al., 2012; Masoud and Koike, 2017). These use computer vision algorithms for edge and lineament detection and, while often successful in ideal situations, require substantial fine-tuning to achieve optimal performance on real-world data. They also have a tendency to detect many false positives related to non-geological features such as shadows, roads or vegetation (Vasuki et al., 2014). Hence, even fully automated methods currently require significant manual effort to remove non-geological features while ensuring features of interest are correctly detected.

To circumvent these difficulties, several methods have been developed which remain user-driven but also use computational power to optimise the interpretation process and improve objectivity and consistency. Vasuki et al. (2014), for example, use an edge detection algorithm (phase congruency; cf. Kovesi, 1999) on orthophotographs to optimise manually defined fracture traces and contacts. This allows the user to quickly define the approximate locations of interesting features and then automatically refine them, speeding up the digitisation process significantly while avoiding problems associated with false positives. Similar computer-assisted approaches have also been applied to improve the interpretation of faults in regional magnetic surveys (Holden et al., 2016) and oceanic fracture zones in global gravity datasets (Wessel et al., 2015).

In many situations, the 3D orientations of detected features are of interest. This is typically calculated using a digital elevation model to add height information to each trace and then computing a best-fit plane (e.g., Dering et al., 2016; Jaboyedoff et al., 2009; Banerjee and Mitra, 2005). While this method works well in simple

topography~~,~~ containing slopes <45°, it is inherently limited to 2.5 dimensions (2.5D~~),~~) as elevation data is gridded in the X-Y plane, causing problems when features crosscut steep or overhanging topography (Pavlis and Mason, 2017). For this reason, direct analysis of 3D point cloud data is preferable over methods that are limited to 2.5D. Unfortunately, the unstructured nature of 3D point data mean that methods for trace detection in raster data, ~~such as~~including those previously described ~~above~~, cannot be easily applied.

A number of automatic methods for analysing point-cloud data have been proposed. These use clustering or plane-fitting algorithms to automatically segment and extract ~~joint~~fracture or bedding faces (*facets*) exposed on the surface of the outcrop, with reasonable success (e.g., Dewez et al., 2016; Lato and Vöge, 2012; García-Sellés et al., 2011). However, structural surfaces are not always directly exposed, and instead intersect the outcrop surface to form linear features referred to as *structural traces*. These traces cannot be detected using facet-based techniques, and require a different approach.

Seers and Hodgetts (2016) demonstrate one such approach, automatically extracting 3D structural traces by applying image-based edge detection techniques (phase congruency) to a set of images and then projecting the identified traces into 3D, using depth information derived from photogrammetric reconstructions or associated laser scan data. This approach uses multiple images to overcome issues associated with out-of-plane geometry, however as with other fully automated methods, a variety of parameters and thresholds require careful calibration and the results must be manually vetted to remove false positives.

3 Method: A least-cost path approach to digital mapping

3.1. Theory

The approach presented here couples algorithms for solving least-cost path problems with both general and use-case specific cost functions to capture structural features in both point-cloud and raster datasets. Least-cost path algorithms have previously been used to detect linear features in a variety of image data and have proven robust even when signal-to-noise ratios are very low (Sun and Pallottino, 2003; Vincent, 1998; Buckley and Yang, 1997).

Conceptually the algorithm can be divided into two steps, although for performance reasons our implementation performs these simultaneously. In the first step, data points (points in a point cloud or pixels in an image; Fig. 1a) are linked with their nearest neighbours, using a spherical search radius slightly larger than the dataset resolution, to produce a neighbourhood network (Fig. ~~1a~~1b). The costs of moving along links in this network (hereafter referred to as "edges") are then calculated, using a cost function designed to promote movement along structural or lithological traces and inhibit movement across them (Fig. ~~1b~~1c).

In the second step, an optimised version of Dijkstra's algorithm (Dijkstra, 1959) is used to derive the least-cost path between user-defined control points, providing the estimated trace (Fig. 1c). Djikstra's algorithm, in essence, progressively "grows" least-cost paths from the start point until the end is found. We optimise this by requiring paths to move closer to the end at each step, eliminating tortuous geometries that tend not to be geologically feasible. Once a trace has been estimated, manual adjustments can be easily applied by adding intermediate waypoints and recalculating the relevant least-cost paths.

The critical component in this approach is the cost function. A well-designed cost function produces low values for edges following structure or contact traces, and high values for edges outside or crosscutting traces. Our optimised implementation of Dijkstra's algorithm then follows edges with the lowest cost-values in order to map

out the feature of interest. ~~We have designed and implemented five simple cost functions that give reasonable results for different structure and data types (Appendix 1).~~ Conveniently, simple cost functions such as point/pixel brightness or local colour gradient work well on most geological datasets; the examples presented below all map a single scalar attribute in the dataset directly to cost (point/pixel brightness for Study 1 and 2, topographic slope for Study 3 and bathymetric depth and vertical gravity gradient for Study 4). We have designed and implemented these and several other simple cost functions that give reasonable results for different structure and data types. Specific equations for these are included as Appendix 1.

3.2. Implementation

The above methodology has been implemented as plugins for Cloud Compare (Girardeau-Montaut, 2015) and QGIS (QGis, 2011), both of which are cross-platform, open-source and widely used software packages for geospatial analysis. Our CloudCompare plugin (*Compass*) is written in C++ and works ~~on~~for point ~~clouds~~cloud data, while the QGIS ~~implementation~~version (*GeoTrace*) ~~works on~~is implemented as a python plugin, using numpy (van der Walt et al., 2011) and scikit-image (van der Walt, 2014) to apply our method to raster data.

Compass is bundled with the default CloudCompare distribution (since version 2.9), and the source code is freely available at https://github.com/CloudCompare/CloudCompare. Similarly, *GeoTrace* can be found on the QGIS plugin repository (https://plugins.qgis.org/plugins/), and the source code downloaded from https://github.com/lachlangrose/GeoTrace. Complete documentation for the plugins is found at the CloudCompare wiki (http://www.cloudcompare.org/doc/wiki/) and on the *GeoTrace* GitHub page.

In addition to our method for rapidly extracting structural traces, a variety of other functionality has been implemented, including tools for measuring surface orientations, lineations and true thicknesses in the Cloud Compare plugin, and DEM based plane fitting and orientation analysis in the QGIS plugin.

4 Case Studies

To demonstrate the capability of ~~the~~our computer-assisted trace detection approach ~~described above~~, we present the results of four case studies. These studies highlight the versatility of our method and its increased efficiency as compared to established manual methods.

The first case study involves the interpretation of joint sets in two 10 × 10 m areas from a ~1 cm resolution orthophotograph of a wavecut platform at Bingie Bingie Point, New South Wales, Australia. The outcrop contains several Cretaceous to Paleogene dykes intruding Devonian ~~plutonic~~ diorites and tonalities and crosscut by a series of complex joint sets. The orthophotograph was generated by applying a Structure from Motion-Multi-View Stereo (SfM-MVS) workflow ~~(Cruden et al., 2016) to~~, as implemented in Agisoft Photoscan v1.2.6, to 297 digital photographs captured from a DJI S800 Evo multi-rotor UAV fitted with a 24.3-megapixel Sony Nex-7 camera and 16mm F2.8 prime lens~~.~~ (Cruden et al., 2016). The two 10 × 10 m areas (Fig. 2a, b) were selected from the survey as they contain well exposed dykes and joint sets as well as common confounding effects such as shadows and puddles. For demonstration purposes the selected areas are relatively small, but the workflow is equally applicable to much larger outcrops.

Our second case study focuses on the extraction of 3D joint traces and orientations, which are interpreted directly on a dense 3D point cloud. The Cape Woolamai sea~~–~~stacks, located approximately 115 km southeast of Melbourne

(Australia) on Phillip Island, have formed by erosion of the coarse-to medium-grained Cape Woolamai granite, which intruded Silurian to Lower Devonian meta-turbidites during or slightly after the mid-Devonian Tabberabberan Orogeny (a widespread episode of deformation and plutonism across Victoria; Gray, 1997; Richards and Singleton, 1981). Several ~~sets~~generations of systematic and non-systematic joints crosscut this granite, likely related to the cooling of the intrusion, subsequent deformation and ~~recent~~ unloading.

For this study, a DJI Inspire 1 multi-rotor UAV and ~~a~~20 mm fix-focal Zenmuse X3 camera were used to capture 274 aerial photographs, which were subsequently processed using ~~a SfM-MVS workflow. The~~Agisoft Photoscan v1.2.6. A 45 × 40 × 25 m section of the resulting ~~3D~~ model ~~is 45 × 40 × 25 m in size and comprises~~containing a single sea-stack was then extracted, containing ~2 million points~~, with a~~ and representing an average ground sampling distance of ~2.5 cm/pixel. The topographic complexity of this ~~outcrop~~sea-stack allows for accurate joint-orientation measurement, but makes interpretation from 2.5D datasets (orthophotograph + DEM) impractical (Fig. 3a, b).

For the third case study, surface ruptures that formed along the Greendale Fault after the 2010 M_w7.1 Darfield earthquake are extracted from a 1 m resolution LiDAR~~-~~-derived DEM. The data was collected a few days after the earthquake and was used, along with a variety of other data, to measure the surface displacement resulting from the earthquake and to interpret the kinematics of the Greendale Fault (Duffy et al., 2012).

Finally, for our last case study we interpret oceanic fracture zones in the North Atlantic from 30 arc-second bathymetry (Weatherall et al., 2015) and vertical gravity gradient data (Sandwell et al., 2014). From their inception at mid-ocean ridges, fracture zones can be used to constrain plate motion vectors and are widely used in tectonic reconstruction (Williams et al., 2016; Sandwell et al., 2014). Both these datasets provide an opportunity to test our method on global-scale geophysical data.

For each of these case studies, we also assess the similarity of our assisted results to manually derived interpretations. The aim of these comparisons is not to rigorously validate the computer-assisted method as: (1) given enough control-points our method could match any interpretation, and; (2) manual interpretation of structures from remotely sensed data is notoriously subjective, and so differing results can be expected from different operators and scales (Bond et al., 2007; Scheiber et al., 2015). Instead, we seek to demonstrate the applicability and versatility of the assisted approach, and that similar results to a manual interpretation can be produced for less time and effort.

For each interpretation, the operator was instructed to digitize every structural feature within the dataset. To ensure this was an achievable task, the extent of the dataset used in each case study is small compared to its resolution. No attempt was made to ensure that the same number of features was extracted from each dataset, as this would affect timing measurements. Digitization was performed at or close-to the dataset resolution, so that the manual interpretations contained similar detail to the assisted approach (which always follows traces at the resolution of the dataset), although the operator was able to freely zoom in and out to inspect the data at multiple scales. The same operator performed the manual and assisted interpretations for Case Studies 1 and 2, while different operators generated each of the assisted and manual interpretations for Case Study 3 and 4. Manual interpretation in QGIS (Case Study 1, 3 and 4) was performed by digitising polyline features to a shapefile using the "Add feature" tool, while in CloudCompare (Case Study 2) the "Trace-polyline by point picking" tool was used.

5 Results

~~In each of the case studies described above the data have been interpreted twice, once using our computer-assisted method and once using manual workflows in QGIS (Study 1, 3 and 4) or CloudCompare (Study 2).~~ The time required to extract comparable amounts of structural data was significantly reduced using the computer-assisted method (Table 1). This efficiency increase was especially pronounced (61%) for the point cloud example, as manual methods for digitising linear features on 3D point clouds are particularly time consuming.

The following four subsections compare and contrast the results of both manual and assisted interpretations in more detail.

5.1. Bingie Bingie Point

Both areas (Fig. 2a, b) of the Bingie Bingie Point orthophotographs contain joints over a range of scales and in a variety of host rocks, as well as features that make automated interpretation challenging such as water, shadows and debris-filled joints. Fracture and contact traces were digitised manually in QGIS (Fig. 2c, d), and with the *GeoTrace* implementation of our assisted method (Fig. 2e, f). For the assisted interpretation, different cost functions were used to pick the fractures and the dyke contacts. Fractures in the orthophotographs are clearly darker than their surroundings, so a greyscale version of the orthophotograph (easily calculated using *GeoTrace*) was used to define the shortest-path cost function during fracture digitisation. Dyke contacts were mapped using a cost function derived from the inverse of the local brightness gradient (high gradient = low cost). This was achieved by applying a Sobel filter (essentially a local gradient operator; Sobel 1990) to the ~~previously mentioned~~ greyscale image, using scikit-image functionality (van der Walt et al., 2014) integrated into *GeoTrace*.

The results are visually similar to the manually derived reference interpretation (Fig. 2c-f). Closest-point differences ~~between the manual and assisted interpretations~~, calculated by subsampling closely-spaced points from each assisted trace and computing the shortest-distance between these and a manually interpreted trace, show that the majority of traces (78% in Area 1 and 70% in Area 2) match to within 2 pixels (≈2 cm), smaller than the ambiguity of the dataset.

5.2. Cape Woolamai

Joints in the Cape Woolamai ~~virtual~~digital outcrop model were interpreted in 3D using CloudCompare, first with the manual "draw polyline" tool and then using the *Compass* implementation of our method. The complex topography of the sea-stacks makes 2.5D analysis inappropriate (Fig. 3a, b). As in the ~~previous~~Bingie Bingie example, cost was defined by point brightness as fractures are defined by their darker colour.

In total, 146 joint traces were interpreted manually over ~3 hours, while 114 joint traces were digitised using the *Compass* plugin in less than an hour (Table 1). Joint orientations were estimated by calculating the least-squares plane-of-best-fit for each trace. The ratio between the second and third eigenvectors of each trace was then used to reject arbitrary planes resulting from sub-linear traces, using a planarity threshold of 0.75 (cf. Thiele et al., 2015 for a more detailed description of this method). *Compass* does this in real-time during the digitisation processes, while orientation estimates from the manually digitised dataset were calculated ~~as a post~~ after processing ~~step~~. The manual and computer-assisted methods resulted in 133 and 91 orientation estimates respectively.

Both sets of interpreted traces and associated orientation estimates appear to be broadly consistent for each method (Fig. 3c-f). Significantly, orientation estimates from the computer assisted method form more-pronounced clusters than equivalents estimated using the manually digitised traces. Although far from conclusive, this indicates that the computer-assisted approach improves the consistency and precision of the orientation estimates, likely due to the larger number of points it samples. The assisted method samples every point along each trace, while the manual method only includes the polyline vertices created during digitisation, making the best-fit plane more susceptible to errors caused by outliers.

5.3. Greendale Fault

Surface ruptures of the Greendale Fault form a series of en échelon fault-scarps visible in the LiDAR dataset (Fig. 4a). Our shortest-path method can be used to pick the fault scarps using a cost function where slope maps inversely with cost. This was achieved by calculating a slope raster using the QGIS DEM (Terrain models) tool and inverting it using *GeoTrace*.

As in the previous examples the assisted interpretation achieved very similar results to a manual interpretation, in about half the time. Closest-point difference calculations between the manual and assisted traces also show that the two sets of interpretation are consistently within ~1–2 pixels (~2 m).

5.4. Oceanic Fracture Zones

Oceanic fracture zones in the North Atlantic were digitised in *GeoTrace* using bathymetric depth to define trace cost. Comparison with an interpretation that was digitised manually shows similar accuracies to the previous case studies, with the majority of traces within ~2 pixels, as well as an improvement of 36% in per-trace digitisation time (Fig. 5).

Additionally, we used the start and end points of oceanic fracture zones described ininterpreted by Matthews et al. (2011), which are based on a 2009 gravity gradient compilation (Sandwell and Smith, 2009), to constrain an otherwise unguided *GeoTrace* interpretation of an updated vertical gravity gradient dataset (Sandwell et al., 2014). This was achieved using the vertical gravity gradient directly as the cost function, such that traces follow areas of low vertical gradient, and then solving the shortest-path between the Matthews et al. (2011) start and end points.

The results (Fig. 5d5e) again highlight the tool's general accuracy, with 65% of traces falling within 2 pixels of the Matthews et al. (2011) interpretation and 79% within 5 pixels. Most errors occurred in areas of closely spaced fracture zones, where the computed shortest-path for many fracture zones would "detour" through adjacent low-cost features (Fig. 6). A small number of additional control points along these traces would resolve this issue (Fig. 6c) by forcing the computed path to stay in the local cost-minima (the correctdesired fracture zone), rather than taking advantage of larger adjacent minima.

6. Discussion

The four case studies presented above highlight applications of the least-cost-path method to the interpretation of high-resolution aerial orthophotographs, 3D point clouds, LiDAR DEMDEMs and bathymetric data - all datasets commonly used in the earth sciences to interpret and characterise geological features. We discuss here three aspects of our least-cost-path approach: it improves objectivity and reproducibility, allows automatic refinement

if better data becomes available, and unlike fully-automated ~~workflows, operates~~methods, works in co-operation with expert guidance.

Firstly, the approach is more objective than manual digitisation. Although not as objective as fully-automated methods (the location of the trace start and end are interpreted), most of the length of each trace is determined algorithmically, and hence will consistently locate in the same spot. Indeed, as demonstrated in Figure 7, the calculated shortest path varies only slightly when control points are interpreted at different locations. Results from Study 2 indicate that this improved consistency might increase the precision of the derived orientation estimates (Fig. 3e, f). Furthermore, each control point can easily be stored, providing a record of the locations at which interpretive decisions were made.

~~Similarly~~Secondly, similarly to the method outlined by Wessel et al. (2015) for extracting oceanic fracture zones, these control points can also be reused to generate an updated interpretation if higher-resolution or more accurate information become available. This possibility is demonstrated in Study 4, where published oceanic fracture zones were reconstructed automatically using an updated underlying dataset and the start and end points of a previous interpretation (Matthews et al., 2011). Although some quality control is required after such an operation, the digitisation process no longer needs to be completely repeated, and interpretations can be rapidly updated as datasets evolve.

Where multiple datasets are available, the similarity and total cost of paths reconstructed using different datasets can be used to quantitatively assess the degree to which different datasets support an interpretation. It is common in the geosciences to bring interpretations from multiple types of data into a single synthesis (e.g., Seton et al., 2016; Blaikie et al., 2017), especially when using geophysical datasets such as gravity and magnetics. Limiting factors during such data synthesis include both the time required and highly subjective nature of multi-data type interpretations, so a method for rapidly quantifying the extent to which different datasets support an interpretation serves as an important addition. Similarly, sensitivity analyses could be performed by randomly moving control points and measuring the response of the traces to quantify the robustness of the interpretation to uncertainty (similar to Fig. 7).

The time it takes for users to ~~manually~~ interpret datasets using *GeoTrace* or *Compass* will vary significantly between users, and the purpose of this study was not to comprehensively measure the efficiency of our approach. Nevertheless, in each of the case studies, our initial assessment indicates that computer-assisted interpretation required ~35-66% less user effort, as measured by both average time and mouse-clicks per structure trace, when compared to manual methods (Table 1). The resulting traces also appear to be comparable to manual traces in each case (~±2 pixels), demonstrating that our method can be used to achieve equivalent results.

The *Compass* implementation of the technique produces especially impressive results, reducing interpretation time in the Cape Woolamai example by 61%. This is pertinent given the rapid growth in both size and availability of high-resolution point cloud data and the limited range of available tools for extracting structural data from them. Significantly, the implementation of our least-cost-path method in *Compass* requires only local information, such that calculation time scales with trace length and not dataset size. This means the tool can be used to interpret arbitrarily large point clouds.

Finally, the computer-assisted philosophy behind our method keeps the expert in control of the entire digitisation process, allowing data vetting and correction during digitisation. The approach ensures the expert becomes familiar with the particular intricacies of each dataset, a key part of further data analysis and something not

possible using fully-automated methods yet essential for the creative process of understanding and interpreting spatial information.

7. Conclusions

We have described a least-cost-path based method for the computer-assisted digitisation of structural traces in point cloud, image and raster datasets. The method enhances an expert's ability to extract geological information from the wide range of high-resolution data available to geoscientists while reducing the required time and effort. In summary, the method:

- Allows expert-guided interpretation in a way that seamlessly utilises computing power to significantly optimise the interpretation process and improve objectivity and consistency.
- Can be applied to both raster and point-cloud datasets. This is particularly significant in situations where complex topography prevents a more conventional 2.5D raster based workflow.
- Requires only local knowledge of a dataset, so that the total dataset size does not affect performance; thereby allowing computer-assisted interpretation of exceedingly large datasets.
- Is implemented as two freely available and open-source plugins for the widely used CloudCompare and QGIS software packages.

Data Availability

Datasets used for the Bingie Bingie Point and Cape Woolamai case studies are freely available from https://doi.org/10.4225/03/5981b31091af9. The bathymetric and vertical gravity gradient datasets used for the oceanic fracture zone example can be downloaded from the University of California San Diego at http://topex.ucsd.edu/grav_outreach/, while the Greendale Fault LiDAR dataset is available on request from the authors of Duffy et al. (2012).

Author Contributions

ST and LG developed the methodology described in this study. ST implemented it in CloudCompare and LG implemented it in QGIS. All of the authors contributed to the case studies and helped prepare the manuscript. The authors declare that they have no conflict of interest.

Acknowledgements

The authors would like to gratefully acknowledge Daniel Girardeau-Montaut and other CloudCompare developers for creating a fantastic software package and for their assistance creating the *Compass* plugin. ST was supported by a Westpac Future Leaders Scholarship and Australian Postgraduate Award. LG was supported by an Australian Postgraduate Award. AS was supported by a Monash University Faculty of Science Dean's International Postgraduate Research Scholarship and an American Association of Petroleum Geologists Grants-in-Aid award. Finally, we acknowledge A. Bistacchi and T. Scheiber for their insightful and constructive reviews.

References

Banerjee, S., and Mitra, S.: Fold–thrust styles in the Absaroka thrust sheet, Caribou National Forest area, Idaho–Wyoming thrust belt, Journal of Structural Geology, 27, 51-65, 10.1016/j.jsg.2004.07.004, 2005.

Bemis, S. P., Micklethwaite, S., Turner, D., James, M. R., Akciz, S., Thiele, S. T., and Bangash, H. A.: Ground-based and UAV-Based photogrammetry: A multi-scale, high-resolution mapping tool for structural geology and paleoseismology, Journal of Structural Geology, 69, 163-178, 10.1016/j.jsg.2014.10.007, 2014.

Blaikie, T. N., Betts, P. G., Armit, R. J., and Ailleres, L.: The ca. 1740–1710Ma Leichhardt Event: Inversion of a continental rift and revision of the tectonic evolution of the North Australian Craton, Precambrian Research, 292, 75-92, http://dx.doi.org/10.1016/j.precamres.2017.02.003, 2017.

Bond, C.E., Gibbs, A.D., Shipton, Z.K. and Jones, S.: What do you think this is? "Conceptual uncertainty" in geoscience interpretation. GSA today, 17, 4-10, 10.1130/gsat01711a.1, 2007.

Buckley, M., and Yang, J.: Regularised shortest-path extraction, Pattern Recognition Letters, 18, 621-629, 10.1016/s0167-8655(97)00076-7, 1997.

Buckley, S. J., Howell, J. A., Enge, H. D., and Kurz, T. H.: Terrestrial laser scanning in geology: data acquisition, processing and accuracy considerations, Journal of the Geological Society, 165, 625-638, 10.1144/0016-76492007-100, 2008.

Cruden, A., Vollgger, S., Dering, G., and Micklethwaite, S.: High Spatial Resolution Mapping of Dykes Using Unmanned Aerial Vehicle (UAV) Photogrammetry: New Insights On Emplacement Processes, Acta Geologica Sinica (English Edition), 90, 52-53, 10.1111/1755-6724.12883, 2016.

~~De Paor, D. G.: Virtual Rocks, GSA Today, 4-11, 10.1130/gsatg257a.1, 2016.~~

Dering, G., Micklethwaite, S., Barnes, S. J., Fiorentini, M., Cruden, A., and Tohver, E.: An Elevated Perspective: Dyke-Related Fracture Networks Analysed with Uav Photogrammetry, Acta Geologica Sinica - English Edition, 90, 54-55, 10.1111/1755-6724.12884, 2016.

Dewez, T. J. B., Girardeau-Montaut, D., Allanic, C., and Rohmer, J.: Facets : A CloudCompare plugin to extract geological planes from unstructured 3D point clouds ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, XLI-B5, 799-804, 10.5194/isprsarchives-xli-b5-799-2016, 2016.

Dijkstra, E. W.: A note on two problems in connexion with graphs, Numerische Mathematik, 1, 269-271, 10.1007/bf01386390, 1959.

Duffy, B., Quigley, M., Barrell, D. J. A., Van Dissen, R., Stahl, T., Leprince, S., McInnes, C., and Bilderback, E.: Fault kinematics and surface deformation across a releasing bend during the 2010 MW 7.1 Darfield, New Zealand, earthquake revealed by differential LiDAR and cadastral surveying, Geological Society of America Bulletin, 125, 420-431, 10.1130/b30753.1, 2012.

García-Sellés, D., Falivene, O., Arbués, P., Gratacos, O., Tavani, S., and Muñoz, J. A.: Supervised identification and reconstruction of near-planar geological surfaces from terrestrial laser scanning, Computers & Geosciences, 37, 1584-1594, 10.1016/j.cageo.2011.03.007, 2011.

Gray, D. R.: Tectonics of the southeastern Australian Lachlan Fold Belt: structural and thermal aspects, Geological Society, London, Special Publications, 121, 149-177, 10.1144/gsl.sp.1997.121.01.07, 1997.

Holden, E.-J., Wong, J.C., Kovesi, P., Wedge, D., Dentith, M., Bagas, L.: Identifying structural complexity in aeromagnetic data: An image analysis approach to greenfields gold exploration, Ore Geology Reviews, 46, 47-59, 10.1016/j.oregeorev.2011.11.002, 2012.

Holden, E.-J., Wong, J. C., Wedge, D., Martis, M., Lindsay, M., and Gessner, K.: Improving assessment of geological structure interpretation of magnetic data: An advanced data analytics approach, Computers & Geosciences, 87, 101-111, 10.1016/j.cageo.2015.11.010, 2016.

Jaboyedoff, M., Couture, R., and Locat, P.: Structural analysis of Turtle Mountain (Alberta) using digital elevation model: Toward a progressive failure, Geomorphology, 103, 5-16, 10.1016/j.geomorph.2008.04.012, 2009.

Jinfei, W., and Howarth, P. J.: Use of the Hough transform in automated lineament, IEEE Transactions on Geoscience and Remote Sensing, 28, 561-567, 10.1109/tgrs.1990.572949, 1990.

Jones, R. R., McCaffrey, K. J. W., Clegg, P., Wilson, R. W., Holliman, N. S., Holdsworth, R. E., Imber, J., and Waggott, S.: Integration of regional to outcrop digital data: 3D visualisation of multi-scale geological models, Computers & Geosciences, 35, 4-18, http://dx.doi.org/10.1016/j.cageo.2007.09.007, 2009.

Kovesi, P.: Image features from phase congruency, Videre: Journal of computer vision research, 1.3, 1-26, 1999.

Lato, M. J., and Vöge, M.: Automated mapping of rock discontinuities in 3D lidar and photogrammetry models, International Journal of Rock Mechanics and Mining Sciences, 54, 150-158, 10.1016/j.ijrmms.2012.06.003, 2012.

Masoud, A., and Koike, K.: Applicability of computer-aided comprehensive tool (LINDA: LINeament Detection and Analysis) and shaded digital elevation model for characterizing and interpreting morphotectonic features from lineaments. Computers & Geosciences. 106, 89-100, 10.1016/j.cageo.2017.06.006, 2017.

Matthews, K. J., Müller, R. D., Wessel, P., and Whittaker, J. M.: The tectonic fabric of the ocean basins, Journal of Geophysical Research, 116, 10.1029/2011jb008413, 2011.

Pavlis, T. L., and Mason, K. A.: The New World of 3D Geologic Mapping, GSA Today, 10.1130/gsatg313a.1, 2017.

McCaffrey, K.J.W., Jones, R.R., Holdsworth, R.E., Wilson, R.W., Clegg, P., Imber, J., Holliman N., and Trinks I: Unlocking the spatial dimension: digital technologies and the future of geoscience fieldwork, Journal of the Geological Society, 162, 927-938, 10.1144/0016-764905-017, 2005

Pringle, J.K., Howell, J.A., Hodgetts, D., Westerman, A.R. and Hodgson, D.M.: Virtual outcrop models of petroleum reservoir analogues: a review of the current state-of-the-art, First break, 24, 33-42, 10.3997/1365-2397.2006005, 2006

Richards, J. R., and Singleton, O. P.: Palaeozoic Victoria, Australia: Igneous rocks, ages and their interpretation, Journal of the Geological Society of Australia, 28, 395-421, 10.1080/00167618108729178, 1981.

Sandwell, D. T., and Smith, W. H. F.: Global marine gravity from retracked Geosat and ERS-1 altimetry: Ridge segmentation versus spreading rate, Journal of Geophysical Research, 114, 10.1029/2008jb006008, 2009.

Sandwell, D. T., Muller, R. D., Smith, W. H. F., Garcia, E., and Francis, R.: New global marine gravity model from CryoSat-2 and Jason-1 reveals buried tectonic structure, Science, 346, 65-67, 10.1126/science.1258213, 2014.

Scheiber, T., Fredin, O., Viola, G., Jarna, A., Gasser, D., Lapinska-Viola, R.: Manual extraction of bedrock lineaments from high-resolution LiDAR data: methodological bias and human perception, GFF, 137, 362-372, 10.1080/11035897.2015.1085434, 2015.

Seers, T. D., and Hodgetts, D.: Extraction of three-dimensional fracture trace maps from calibrated image sequences, Geosphere, 12, 1323-1340, 2016.

Seton, M., Mortimer, N., Williams, S., Quilty, P., Gans, P., Meffre, S., Micklethwaite, S., Zahirovic, S., Moore, J., and Matthews, K. J.: Melanesian back-arc basin and arc development: Constraints from the eastern Coral Sea, Gondwana Research, 39, 77-95, https://doi.org/10.1016/j.gr.2016.06.011, 2016.

Smith, M. W., Carrivick, J. L., and Quincey, D. J.: Structure from motion photogrammetry in physical geography. Progress in Physical Geography, 40, 247-275, 10.1177/0309133315615805, 2016.

Sobel, I.: "An isotropic 3× 3 image gradient operator." Machine vision for three-dimensional scenes: 376-379, 10.13140/RG.2.1.1912.4965, 1990

Sun, C., and Pallottino, S.: Circular shortest path in images, Pattern Recognition, 36, 709-719, 10.1016/s0031-3203(02)00085-7, 2003.

Thiele, S. T., Micklethwaite, S., Bourke, P., Verrall, M., and Kovesi, P.: Insights into the mechanics of en-échelon sigmoidal vein formation using ultra-high resolution photogrammetry and computed tomography, Journal of Structural Geology, 77, 27-44, 10.1016/j.jsg.2015.05.006, 2015.

Tzong-Dar, W., and Lee, M. T.: Geological lineament and shoreline detection in SAR images, 2007 IEEE International Geoscience and Remote Sensing Symposium, 2007.

van der Walt, S., Colbert, S. C., and Varoquaux, G.: The NumPy Array: A Structure for Efficient Numerical Computation, Computing in Science & Engineering, 13, 22-30, 10.1109/MCSE.2011.37, 2011

van der Walt, S., Schönberger, J. L., Nunez-Iglesias, J., Boulogne, F., Warner, J. D., Yager, N., Gouillart, E., and Yu, T.: scikit-image: image processing in Python, PeerJ, 2, e453, 10.7717/peerj.453, 2014.

Vasuki, Y., Holden, E.-J., Kovesi, P., and Micklethwaite, S.: Semi-automatic mapping of geological Structures using UAV-based photogrammetric data: An image analysis approach, Computers & Geosciences, 69, 22-32, 10.1016/j.cageo.2014.04.012, 2014.

Vincent, L.: Minimal path algorithms for the robust detection of linear features in gray images, Computational Imaging and Vision, 12, 331-338, 1998.

Vollgger, S. A., and Cruden, A. R.: Mapping folds and fractures in basement and cover rocks using UAV photogrammetry, Cape Liptrap and Cape Paterson, Victoria, Australia, Journal of Structural Geology, 85, 168-187, https://doi.org/10.1016/j.jsg.2016.02.012, 2016.

Weatherall, P., Marks, K. M., Jakobsson, M., Schmitt, T., Tani, S., Arndt, J. E., Rovere, M., Chayes, D., Ferrini, V., and Wigley, R.: A new digital bathymetric model of the world's oceans, Earth and Space Science, 2, 331-345, 10.1002/2015ea000107, 2015.

Wessel, P., Matthews, K. J., Müller, R. D., Mazzoni, A., Whittaker, J. M., Myhill, R., and Chandler, M. T.: Semiautomatic fracture zone tracking, Geochemistry, Geophysics, Geosystems, 16, 2462-2472, 10.1002/2015gc005853, 2015.

Williams, S. E., Flament, N., and Müller, R. D.: Alignment between seafloor spreading directions and absolute plate motions through time, Geophysical Research Letters, 43, 1472-1480, 10.1002/2015gl067155, 2016.

Appendix 1

We outline five simple cost functions that give reasonable results for different structure types. Each function is designed to give values between 0 and 1, allowing combinations of functions to be used (by summation), and to work on both unstructured datasets (i.e. point clouds) and structured datasets (images). Hence, we do not present any cost functions that rely on commonly used image processing techniques such as edge enhancement, although these could be easily incorporated for raster datasets. These functions are implemented directly in the *Compass* plugin, while simple QGIS functionality can be used to apply them to raster data for use with *GeoTrace.*

Colour Brightness

The brightness of an edge's end colour (e_{RGB}) can be mapped directly to edge cost (the brightness of an edges start colour will be incorporated into the previous edge in the path). Despite its simplicity, this function (Eq. 1) is surprisingly effective at picking fracture traces, which are typically darker than their surroundings due to shadowing. Similarly, bright traces such as thin quartz or calcite veins can be identified using the opposite of this cost function (Eq. 2). Note that the division by 3 ensures that the function maps to the 0 – 1 range (assuming red, green and blue values also range from 0 to 1).

$$\text{cost} = \frac{\text{e\_R} + \text{e\_G} + \text{e\_B}}{3} \qquad (Eq.\ 1)$$

$$\text{cost} = 1 - \frac{\text{e\_R} + \text{e\_G} + \text{e\_B}}{3} \qquad (Eq.\ 2)$$

Colour Similarity

A similar cost function, based on colour similarity rather than brightness alone, is useful in more generic situations where traces have a distinctive colour but are not necessarily darker or lighter than their surroundings. This function (Eq. 3) considers an edge to be low cost if: (1) the start and end colours are similar, and; (2) the start and end colours are similar to the colour of the start and end of the trace (B_{RGB} and E_{RGB}), minimizing along-path gradient and maximizing similarity with the trace start and end points. This function works well when traces have a specific colour, such as for cemented joints, though it is comparatively slow compared to the brightness-based functions described above due to the large number of square roots. Similarly to the previous equations, the factors of $\sqrt{3}$ ensure that the function maps to the 0 – 1 range.

$$\text{cost} = \frac{1}{2} \times \frac{\left|s_{RGB} - e_{RGB}\right|}{\sqrt{3}} + \frac{1}{2} \times \frac{\left|s_{RGB} - B_{RGB}\right| + \left|s_{RGB} - E_{RGB}\right| + \left|e_{RGB} - B_{RGB}\right| + \left|e_{RGB} - E_{RGB}\right|}{4\sqrt{3}} \qquad (Eq.\ 3)$$

Gradient

The previous cost functions are useful for identifying discrete structural traces such as faults, joints or thin veins, but will not be sensitive to lithological contacts. Lithological contacts are typically defined by changes in colour, and hence we base a cost function around colour gradient to identify them. This function (Eq. 4) evaluates the

gradient *G[N]* of the magnitude of the colour vectors across the start and end neighbourhoods N_{start} and N_{end}. To calculate the gradient for point cloud data, we use a simple method that calculates the average distance-weighted point-to-point gradient for each neighbourhood. More complex methods would highlight contacts better, but at a computational cost. For raster data, we implement a Sobel filter to achieve equivalent results.

An upper limit (*l*) is applied to the gradient in order to maintain a cost value between 0 and 1. A reasonable value for this limit can be approximated by dividing the maximum change in colour magnitude ($\sqrt{3}$) by the average distance between data points. This cost function can also be improved by log-transforming it to increase the importance of gradients resulting from more subtle features.

$$\text{cost} = 1 - \frac{\min\left(\mathrm{G}\left[N_{\text{start}}\right] + G\left[N_{end}\right], l\right)}{\mathrm{l}} \qquad (Eq.\ 4)$$

Curvature

In some situations, resolution is high enough that structural traces, fractures in particular, appear as topographic ridges or valleys. Hence, we include a final cost function (Eq. 5) which considers points with a high curvature as low cost, allowing paths to 'follow' ridges and valleys, where C[N] calculates the mean curvature of a point or pixel neighbourhood N, and *l* is an arbitrarily large upper limit (that allows the log-curvature to scale from 0 to 1). Note that calculating the mean curvature of a neighbourhood is computationally expensive, so this cost function performs significantly slower than the previously described ones unless curvature is pre-computed.

$$\text{cost} = 1 - \frac{\min\left(\log\left(\mathrm{C}\left[N_{\text{end}}\right]\right), l\right)}{\mathrm{l}} \qquad (Eq.\ 5)$$

Table 1. Manual vs computer-assisted digitisation for the different study areas. Percentage improvements are calculated by comparing the average time and mouse clicks per digitised trace. Each case study shows a clear reduction in digitisation time, especially for the 3D datasets where manual interpretation can be especially tedious.

Method	Digitisation time (h:min)	Number of traces	Improvement %	Number of mouse clicks	Improvement %
Study 1: Bingie Bingie					
Area 1					
Manual	0:54	270	-	2253	-
Assisted	0:37	283	35%	917	61%
Area 2					
Manual	0:57	338	-	2509	-
Assisted	0:35	383	46%	1122	61%
Study 2: Cape Woolamai					
Manual	3:04	146	-	6026	-
Assisted	0:56	114	61%	1703	64%
Study 3: Greendale Fault					
Manual	0:18	74	-	1039	-
Assisted	0:07	93	51%	282	66%
Study 4: Oceanic Fracture Zones					
Manual	1:17	432	-	5731	-
Assisted	0:35	310	36%	1265	69%

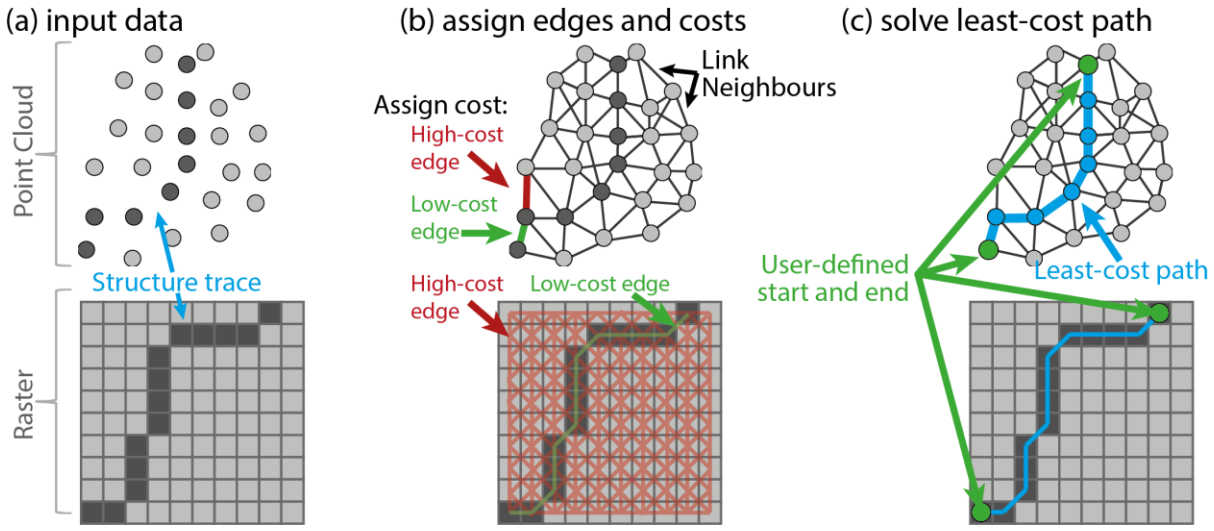

Figure 1. Schematic representation of the least-cost path approach to trace detection for point cloud (top) and raster (bottom) data. Points/pixels on the structural trace have a lower brightness in this example (a), so a brightness-based cost function will result in low-cost edges between adjacent points/pixels that both fall on the structure trace (b). A least-cost path calculation (c) then provides an estimate of the structure trace.

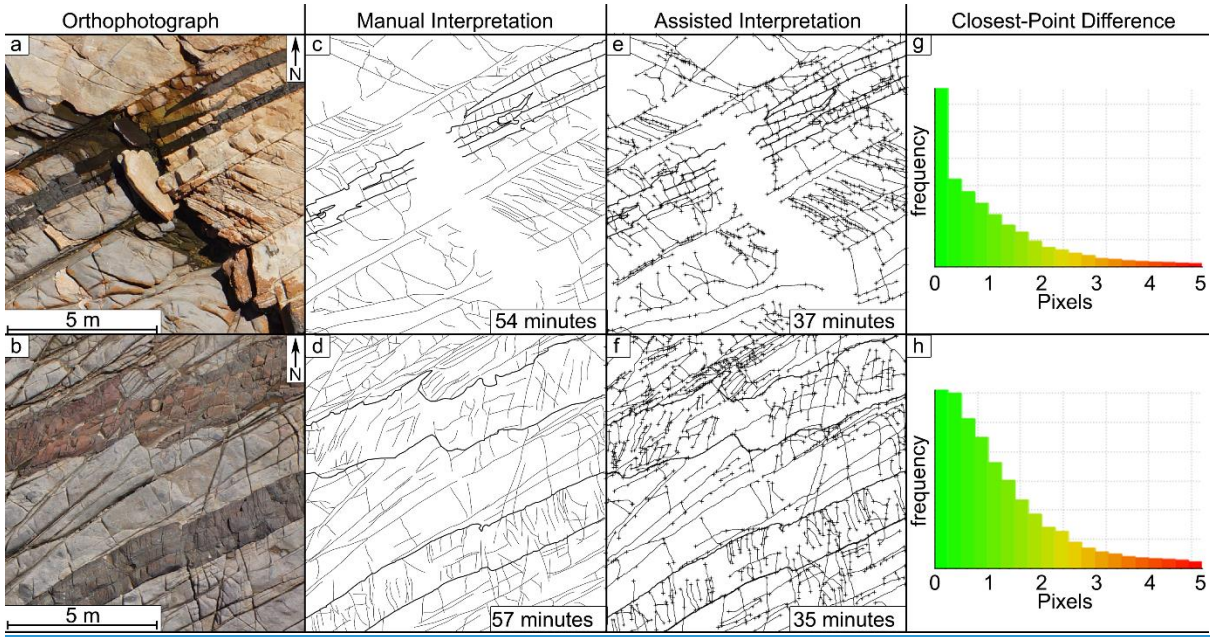

Figure 2. The two 10 × 10 m orthophotographs (a, b) interpreted in Study 1. Fracture traces were digitised manually (c, d) and with our assisted method (e, f). Closest-point distances between the assisted and manual interpretations are also shown (g, h). Note the tails of these distributions have been clipped to 5 cm, as some assisted traces did not have manual equivalents, and hence gave incorrectly large closest-point differences. Small crosses in (e) and (f) represent the control points that were digitised by the user to constrain the shortest path algorithm.

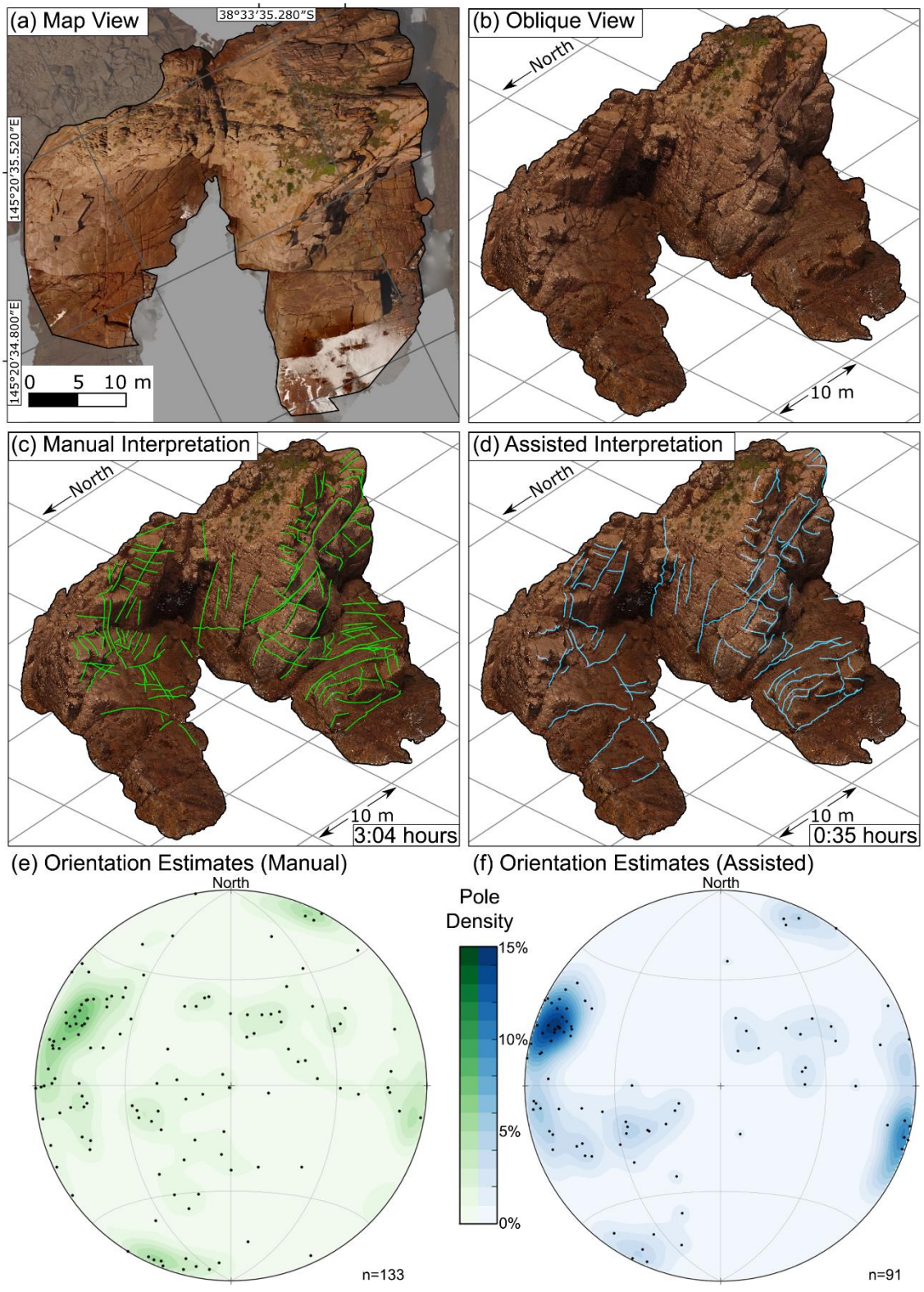

Figure 3. Orthophotograph of the Cape Woolamai sea–stacks (a) and oblique view of the equivalent dense point cloud (b). A 2.5D analysis conducted using the orthophotograph would significantly under sample the moderately to shallowly dipping joint sets which are clearly visible on sub-vertical exposures in (b). Hence fractures were digitised in 3D, both manually (c) and using the computer assisted approach (d). Equal-area lower hemisphere stereographic projections of poles to joint orientations estimated from each of these interpretations (e-f) show that both methods produce similar results. Poles from the computer-assisted dataset cluster more tightly (maximum density = 14.7%) than the manually interpreted dataset (maximum density = 8.1%), indicating that the computer-assisted approach results in more consistent orientation estimates.

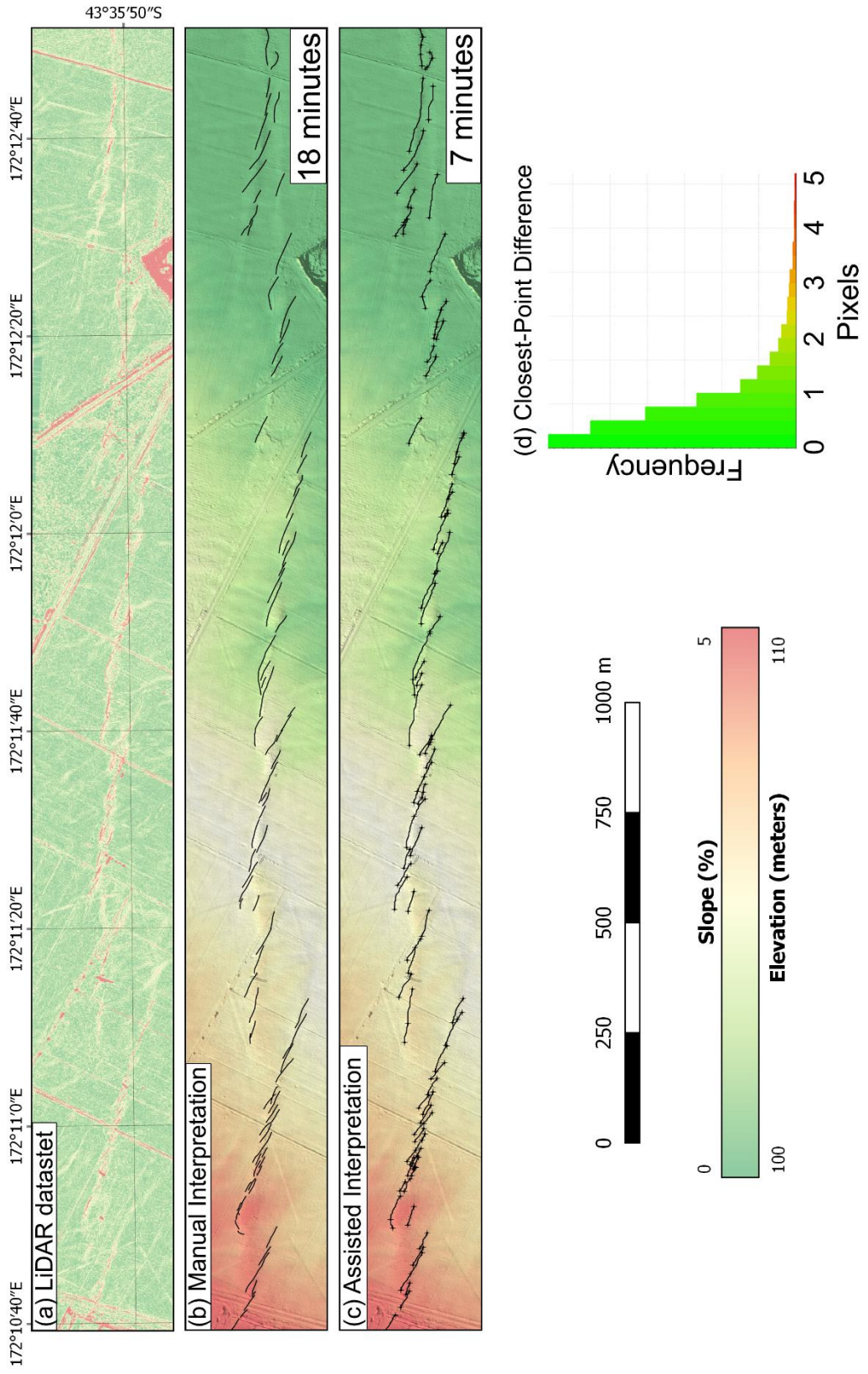

Figure 4. Slope-map derived from the Greendale **LiDAR dataset** ~~illuminated from the NW~~ **showing surface ruptures of a section of the Greendale Fault, New Zealand, collected shortly after the M_w7.1 Darfield earthquake (a). Traces interpreted manually (b) and using the *GeoTrace* implementation of our least-cost-path method (c) are essentially**

equivalent (d). Control points for the assisted interpretation are shown as small crosses. Background for (b) and (c) shows the elevation.

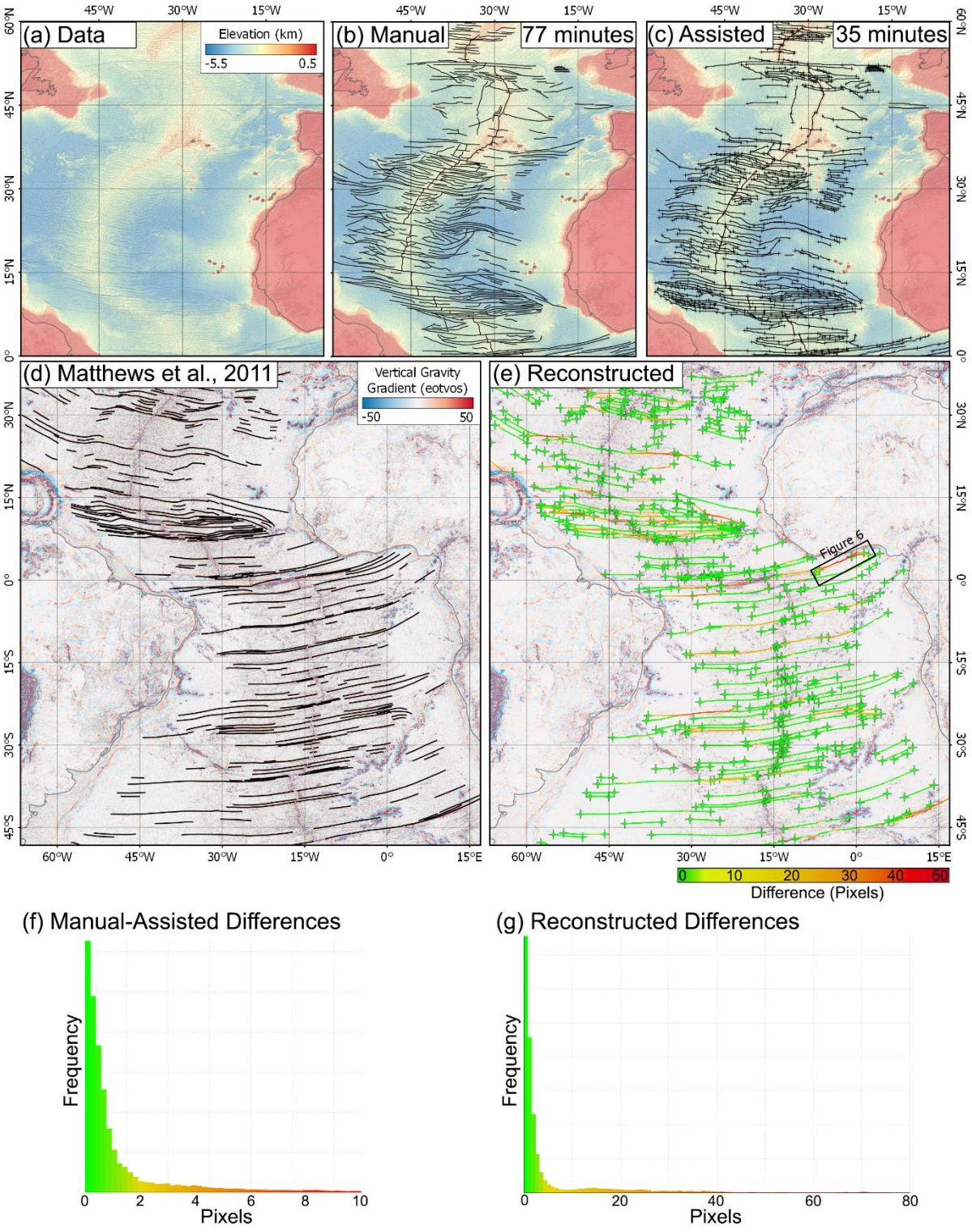

Figure 5. ~~Manual (a) and assisted (b) interpretations of~~Bathymetry showing oceanic fracture zones in the north Atlantic~~.~~ (a) and associated manual (b) and assisted (c) interpretations. Fracture zones interpreted manually (~~c~~d) from vertical gravity gradient by Matthews et al. (2011) and reconstructed in *GeoTrace* using the start and end points only (~~d~~e) are also shown. ~~Red and blue colours in (c) and (d) show areas of high and low vertical gravity gradient, respectively.~~ As in the previous case studies, most equivalent manual and assisted traces fall within 2 pixels (~~e~~f), though differences of up to 80 pixels occur in the reconstructed dataset (~~d~~).g). The location of Fig. 6 is shown in (e) for reference.

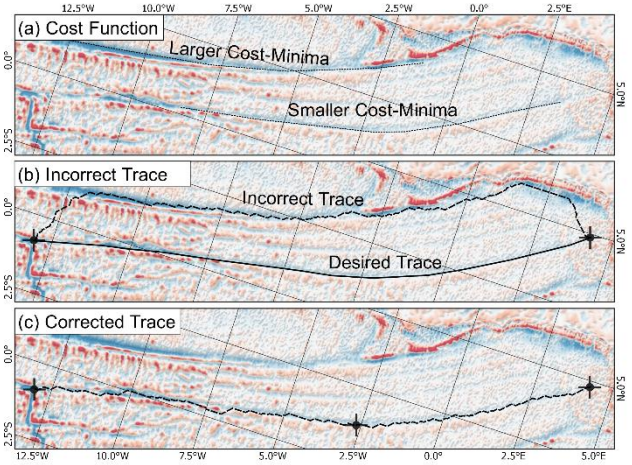

Figure 6. Example of a larger cost minima (a) causing the incorrect reconstruction (b) of an oceanic fracture zone. In this case the trace can be corrected by adding a single additional control point midway along the fracture zone (c).

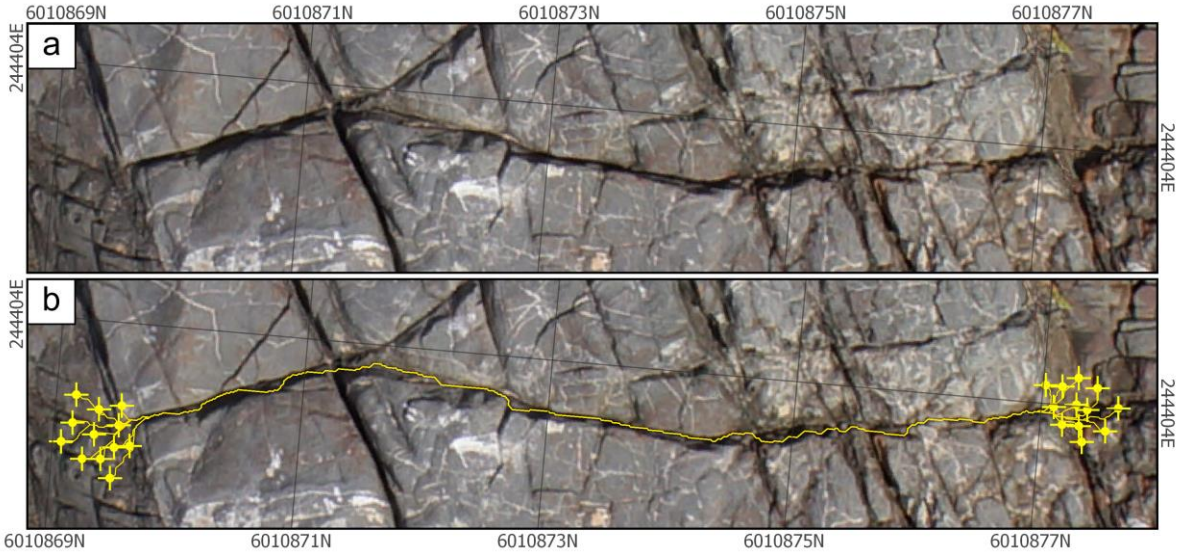

Figure 7. Fracture (a) from the Bingie Bingie Point dataset showing that the majority of the resulting trace (b) consistently follows the same path despite variation of the location of the control points.