EuroVis Workshop on Visual Analytics (2017)
M. Sedlmair and C. Tominski (Editors)

Visual Analytics for Multitemporal Aerial Image Georeferencing

Albert Amor-Amorós[1], Paolo Federico[1], Silvia Miksch[1], Sebastian Zambanini[2], Simon Brenner[2], and Robert Sablatnig[2]

[1]Institute of Software Technology & Interactive Systems, TU Wien, Vienna, Austria
[2]Computer Vision Lab, Institute of Computer Aided Automation, TU Wien, Vienna, Austria

Abstract

Georeferencing of multitemporal aerial imagery is a time-consuming and challenging task that typically requires a high degree of human intervention, and which appears in application domains of critical importance, like unexploded ordnance detection. In order to make a semi-automatic scenario possible, we introduce a Visual Analytics approach for multitemporal aerial image georeferencing designed in close collaboration with real-world analysts that face the problem on a daily basis, and implemented by combining computer vision and interactive visual exploration methods. We report on informal validation findings resulting from the integration of our solution into our users' GIS platform of choice, which positively illustrate its effectiveness and time-saving potential.

Categories and Subject Descriptors (according to ACM CCS): Human–centered computing: Visualization – Visualization application domains – Visual Analytics; Computing methodologies: Computer vision problems – Matching.

1. Introduction

Image georeferencing, i.e., the association of geographical coordinates to the different points of an image, is a fundamental problem in several applications involving the analysis of historical aerial imagery, such as unexploded ordnance detection. In this context, analysts have to identify and annotate a sufficient number of correspondence points of each image on a supporting georeferenced image of the same area. This task, known as image registration, can be very taxing in scenarios involving a large number of images. While existing computer vision methods can be leveraged to partly automate this time-consuming operation, the critical nature of many of its applications makes a human-in-the-loop scenario highly desirable. In this work, we address this challenge by introducing a Visual Analytics approach to multitemporal aerial image georeferencing. In particular, we claim the following contributions:

- providing a Visual Analytics solution to multitemporal aerial image georeferencing intertwining computer vision and interactive visual exploration methods in a coherent workflow;
- integrating our solution into the users' GIS platform, and reporting on its effectiveness and user satisfaction.

2. State of the art

Automatic image registration methods enable the geometric alignment of images through the identification of similarities in local or global image regions. It is a key step in various remote sensing applications, such as image fusion, change detection and georeferencing [LMNE11]. Existing image registration methods can be divided into area-based and feature-based methods [ZF03]. The former aim at optimizing the transformation parameters based on a global similarity metric, whereas the latter detect and match salient control points and derive the transformation parameters from spatially verified correspondences. Due to their computational complexity, area-based methods are limited to registration problems with a restricted search space of transformation parameters, e.g. the fine registration of roughly aligned image pairs [LLH*14]. In contrast, the local nature of feature-based methods allows to use the most prominent features between the images for registration, consequently enabling wider transformation ranges and a higher robustness against dissimilar image parts [MZZ*15]. The efficiency of feature-based methods is grounded on powerful local image descriptors that provide a reasonable trade-off between distinctiveness and insensitivity to various imaging conditions [Gos12]. Additionally, spatial verification techniques [RFP08] are exploited to filter out falsely matched features that would otherwise critically disturb the final estimation of transformation parameters.

While existing automatic methods can be used for building high-performance aerial image registration systems, the critical nature of many application domains makes a purely automated workflow infeasible, requiring human supervision to validate the process outcomes and intermediate results. Visual Analytics [KAF*08] is an emergent discipline dealing with problems that have to be tackled using both the exceptional computational power of modern computers and the superior human perceptual and cognitive capabilities. The interactive visual exploration of time-oriented [AMST11] and spatio-temporal [AA06] data has been largely studied. However, to the best of our knowledge no Visual Analytics approach for aerial image georeferencing has been described in the literature.

© 2017 The Author(s)
Eurographics Proceedings © 2017 The Eurographics Association.

DOI: 10.2312/eurova.20171120

delivered by EUROGRAPHICS DIGITAL LIBRARY
www.eg.org diglib.eg.org

3. A Visual Analytics approach

In our target scenario, an analyst needs to georeference tens of images representing monochrome aerial photographs of a given region around an area of interest, taken during the 1930s and 1940s decades. In addition, metadata about the images, such as their date, approximate midpoint coordinates, and spatial resolution, are also available. The georeferencing task involves finding the transformation that best matches each particular image into modern imagery associated to a given geographic reference system. The calculation of the appropriate transformation can be performed using commercial GIS tools, given that a number of control points that connect the images are provided. Therefore, finding a sufficient number of control points for each image is the main analytical challenge faced by the analyst, and it is usually executed by repeated identification and selection of corresponding points in a side-by-side or layered arrangement of images. Even though analysts performing this task typically have great expertise, e.g., a strong sense of spatial orientation, and high competence for annotation, layer management, and pan and zoom navigation, it is generally a very taxing procedure requiring a great amount of time and human effort.

Our approach relies on the use of automatic methods for finding control points among images of the historical set (historical-to-historical registration), followed by interactive user-driven registration of one of the historical images on a modern georeferenced satellite image (historical-to-modern registration) to achieve automatic georeferencing of the complete historical set, on the basis of appropriate geometric transformations computed for each image. The rationale behind this strategy is the more challenging nature of modern-to-historical aerial image registration, caused by the greater amount of change that typically exists between images separated by a considerable time lapse. Accordingly, the workflow of the analyst consists in a sequence of two steps: 1) specifying a number of historical images, that will be registered to each other automatically, and 2) interactively registering one of those images (the so-called master image) to a modern satellite image of the same area. The second step requires the analyst to provide a sufficient number of control points in terms of their locations in both images. To accomplish this task, the analyst must rely on standard GIS layer management and navigation (i.e., pan and zoom) capabilities for individually locating and clicking points of both images that are considered to match. This solution does not provide fully-automated georeferencing, yet it exempts the analyst from the burden of interactively performing the registration of most images, eliminating a high amount of cognitive load.

3.1. Automatic aerial image registration

For the historical image pairs to be registered, translation and rotation differences are unknown *a priori*, and image changes can be partially strong due to different variation effects, such as destruction and reconstruction after bombing, changing weather conditions, cloud coverage, or image noise. This makes area-based techniques infeasible, and we consequently build upon feature-based registration for this task. We use SIFT [Low04] to extract local image descriptors in both images. SIFT allows to perform an automatic scale and orientation selection of detected control points which makes the descriptors invariant to changes in spatial resolution and rotation of the images. However, we found automatic control point detection too unreliable on the given image data, and therefore perform a dense feature extraction on a regular spatial grid. Automatic scale selection of control points is avoided by normalizing the two images with respect to their spatial resolution estimated from the aircraft altitude and camera's focal length.

The interval of the regular spatial grid as well as the patch size of the SIFT descriptors are both empirically set to 32 meters. After feature extraction, control points in the first and second image are matched based on a L1–distance nearest neighbors search, and a subsequent ratio test: a match is accepted only if the ratio between the first nearest neighbor and second nearest neighbor is less than 0.8. This ratio test helps to filter out matches that are too ambiguous and thus prone to confusion. A second filtering step is performed by RANSAC [FB81], where the subset of matches with highest support for the projective transformation between the two images is identified. The outcome of the registration process is a set of corresponding points between given image pairs.

3.2. Data model

From a data modelling perspective, the registration process endows the original image set with two more facets additional to their time-oriented nature. On the one hand, it yields the coordinates of a number of points that constitute matches between image pairs; in other words, it generates relational data describing how images are related to each other. On the other hand, given a set of corresponding points, a geometric transformation can be found that minimizes the overall mismatch between those points when both images are displayed together; such transformations endow the images with a spatial facet, enabling their positioning on a common space. On the basis of these considerations, we model the registration outcome using a graph in which images constitute the nodes and corresponding points define edges between them. Every node can be mapped to a time instant that defines its temporal facet, and to a geometric shape with a specific position, orientation, and scale, defining its spatial facet. Similarly, every edge has associated two pairs of spatial coordinates defining the position of the point they represent with respect to the images (i.e., nodes) at both ends.

3.3. Interactive visualization

We have introduced interactive visualization methods to appropriately support the analyst in supervising the automatic registration outcomes, and identifying the best master image candidate. The visualization must provide an overview that permits assessing the degree to which previously disconnected images match together into a coherent whole, in addition to the necessary interaction means to investigate matches between any given pair individually. Relevant factors that inform the selection of the master image are the location and distribution of the correspondence matches, the quality of pairwise registration in terms of the number of control points and their associated mismatch errors, or the centrality of the image in the registration graph, in order to minimize error propagation. In the following, we provide our design rationale, together with a short note on the implementation of the system.

Visual encoding For the visual interface, we use coordinated views in order to highlight different facets of the data. More specifically, we rely on a two-view design with the third facet incorporated into both of these views. Our design consists of a timeline

© 2017 The Author(s)
Eurographics Proceedings © 2017 The Eurographics Association.

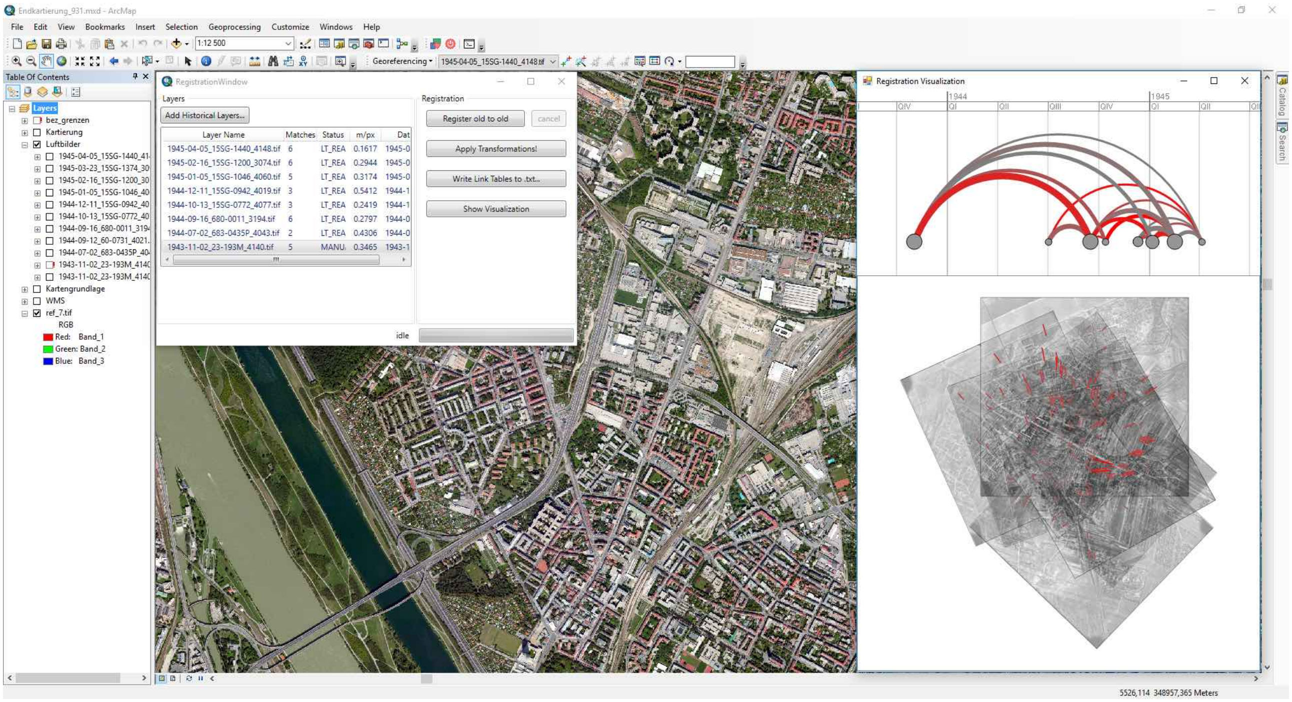

Figure 1: *A screenshot of the Visual Analytics environment for historical aerial image georeferencing. Two windows on top of the main GIS interface enable the analyst to control the image registration process (left), and interactively explore its outcome through a visual interface with two coordinated temporal and spatial views enhanced with the relational facet given by the registration (right).*

and a map, arranged in a vertical stack on which images are positioned according to their temporal and spatial facets respectively, and a dual representation for the relational facet with correspondence edges displayed simultaneously on both views (Figure 1).

In the **map** view, images are displayed in accordance to the position, scale, rotation, and shape defined by the geometric transformation computed on the basis of the correspondence points. Images generally have a considerable degree of overlap, and form a superposition of semi-transparent layers, with most recent images shown on top. Such an arrangement provides an initial overview on the spatial extent spanned by the images. On top of the images, we display the individual correspondence matches in order to highlight their spatial distribution and density. While such correspondences represent single points in each of the images, they must be represented as lines once the images have been transformed, since the transformation will not, in general, result in a perfect alignment for the points. Accordingly, the further apart the points, the longer these lines, and therefore the larger the errors associated to the transformation. We additionally encode the error associated to each of the lines using a single-hue red sequential color scale: longer lines representing a higher error are colored with a more saturated shade of red. Finally, the number of detected correspondences varies strongly between different image pairs, reaching up to thousands in some cases. Such a high amount of correspondences would not only lead to an undesirably cluttered visualization, but would also make a manual refinement of the control points by the analyst unpractical. Therefore, we limit the number of displayed correspondences by a greedy heuristic striving for an even spatial distribution across the images, and minimal distances of the corresponding points in the feature space.

In the **timeline** view, circles representing individual images are positioned in a horizontal arrangement according to their respective timing, providing an overview on the temporal extent spanned by the set of images. In addition, aggregated registration correspondences for each of the image pairs are represented as arcs between nodes, enhancing the timeline into an arc diagram [Wat02]. Node size encodes the number of total correspondences shared with any other image (i.e., its weighted degree), while arc thickness encodes the amount of individual matches represented by timeline arc (thicker arcs mean a higher number of matches). Additionally, as an indicator for pairwise registration quality, the average error associated to each set of correspondences is encoded using arc color (higher red saturation for higher average mismatch), resulting in the emphasizing of those pairs with lower registration quality, granting the user the possibility to investigate, and possibly exclude, registration links of unsatisfactory quality.

Interactive exploration The aforementioned visual encoding results in a data representation that provides an initial overview on the set of images and the outcome of their automatic registration, in terms of aspects such as the temporal and spatial extent of the images, or the spatial distribution and degree of mismatch for both individual and aggregated correspondences. However, such an overview does not convey more detailed pieces of information such as how images on the timeline relate to images on the map, or which set of individual matches on the map correspond to a given arc on the timeline. This appears whenever views representing different

© 2017 The Author(s)
Eurographics Proceedings © 2017 The Eurographics Association.

facets of data are displayed together, and can be effectively solved by introducing interaction-driven coordination. In the following, we elaborate on three actions available to analysts for interacting with the visual representation and exploring the results in detail:

- **Highlight.** Temporal and spatial views can be linked interactively by means of a visual query: whenever a node in the timeline is hovered with the mouse, the area on the map spanned by the corresponding image is highlighted with a red outline. Long hovering a timeline element results in the appearance of a tooltip with specific details like relational metrics (e.g., node degree, arc weight) or average registration error.
- **Focus.** Particular entities of interest can be focused upon by clicking on the corresponding element in the timeline. In the focus mode, the focus image is brought to the front and made opaque, and any elements that are not directly related to it are hidden. A small cross indicates the end of the line corresponding to the image in focus. Furthermore, specific pairs can be inspected in detail using a combination of hovering on the corresponding arc, which results in the hiding of non-relevant elements, and followed by scrolling to switch the top-most image.
- **Iterate.** The position of the focus can be iterated easily using mouse scrolling on any empty area, which results in the sequential displacement of the focus, either from oldest to newest (for up-scrolling), or from newest to oldest (for down-scrolling). By quickly iterating the position of the focus, the analyst can explore the registration graph locally in a clutter-free context.

3.4. Implementation

The system has been implemented as extension of ArcMap [Arc], a GIS environment that constitutes the fundamental tool for our analysts to execute their daily workflows. This architecture provides the analysts with a compatible solution with their existing environment and workflows: aerial images present in ArcMap can be selected for automatic registration; then, the selected master image can be manually registered onto up-to-date georeferenced satellite imagery; finally the rest of the selected project images are registered automatically using a selection of correspondences, which can then be refined by the analyst if necessary. We used OpenCV [Bra08], an extensive open-source computer vision library, for implementing the automatic registration, and TimeBench [RLA*13], a library for time-oriented data based on prefuse [HCL05], for the interactive visualization.

4. Validation

This work is the result of a collaboration with Luftbilddatenbank Dr. Carls GmbH, a company dealing with historical aerial image collection, localization, and analysis, for the assessment of unexploded ordnance risks from World War II in Central Europe. We identified the most important design requirements through the detailed characterization of the problem in terms of the Nested Workflow Model [FAAM16], an extension to the Nested Model methodology [Mun09] aimed at the analysis of workflows at multiple abstraction levels. In accordance to the human-centered approach to design and development [SJWM05], we continuously involved users (i.e., our analysts) along the different phases of design and validation. Threats to validity at the domain characterization level were addressed by a pre-design grounded evaluation [IZCC08], performed by means of contextual inquiries [HJ93] and *in situ* semi-structured interviews [Woo97]. We refined our findings in multiple iterations, in which users were asked to check the outcomes of requirement analysis and problem modeling in terms of correctness and completeness, in order to validate our design and implementation. Initially, we jointly discussed the first design sketches by means of several mock-ups and textual descriptions. Next, we introduced a first demonstrator that illustrated the complementary role of computer vision and interactive visualization in a controlled environment. Finally, we deployed the software to the users' work environment by integrating it into their GIS platform of choice, and asked them to use it with real data and report their usage experience.

Senior analysts from two different branches of the company used the tool to replicate the georeferencing phase of a typical project. Through this informal evaluation, we collected interesting feedback regarding minor usability issues, as well as anecdotal evidence about the tool's effectiveness and user satisfaction. As for usability, users reported problems about initial data loading and parameter setting, which were easily fixed. They appreciated the arc-diagram/timeline view, which combines the relational information with relevant time-oriented information. According to their comments, our Visual Analytics solution reduces the amount of effort required to accomplish the georeferencing task, and enables them to achieve the necessary degree of accuracy in a shorter period of time. For them, this represents a significant improvement, considering that the interactive registration of all the images involved in a typical project can take up to several hours for a senior analyst. Additional indication that the domain problem has been properly addressed and the target users can actually benefit from our Visual Analytics approach comes from the fact that the company reported plans for incorporating the tool into their production environment, where all analysts would adopt it. Nevertheless, while these preliminary findings indicate positive results, further evaluation by formal user studies is still required for a rigorous confirmation.

5. Conclusion

We presented a Visual Analytics approach for multitemporal aerial image georeferencing combining computer vision and interactive visual exploration methods. Our georeferencing strategy leverages automatic methods for performing historical-to-historical image registration, and relies on the user to interactively perform a single historical-to-modern image registration, in order to achieve the georeferencing of the complete image set. We are working towards the integration of object-detection capabilities, which will enable the automatic identification and highlighting of particular artifacts of interest (e.g., craters). Moreover, we aim at extending the approach with automatic historical-to-modern registration in order to further lower the amount of effort and time required. Nevertheless, we believe critical applications requiring human supervision will still benefit from the advantages of this approach even in a fully-automated scenario.

Acknowledgements

The authors wish to thank Luftbilddatenbank Dr. Carls GmbH, as well as Tim Lammarsch. This work was partially supported by the Bridge 1 Programme of the Austrian Research Promotion Agency (FFG), grant number 850695, project DeVisOR.

© 2017 The Author(s)
Eurographics Proceedings © 2017 The Eurographics Association.

References

[AA06] ANDRIENKO N., ANDRIENKO G.: *Exploratory Analysis of Spatial and Temporal Data: A Systematic Approach*. Springer-Verlag Berlin Heidelberg, Secaucus, NJ, USA, 2006. doi:10.1007/3-540-31190-4. 1

[AMST11] AIGNER W., MIKSCH S., SCHUMANN H., TOMINSKI C.: *Visualization of Time-Oriented Data*, 1st ed. Springer Publishing Company, Inc., 2011. doi:10.1007/978-0-85729-079-3. 1

[Arc] ArcMap. http://desktop.arcgis.com/en/arcmap/. [Online; accessed 24-Apr-2017]. 4

[Bra08] BRADSKI G.: *Learning OpenCV: computer vision with the OpenCV library*. O'Reilly, Sebastopol, CA, 2008. 4

[FAAM16] FEDERICO P., AMOR-AMORÓS A., MIKSCH S.: A nested workflow model for visual analytics design and validation. In *Proceedings of the Sixth Workshop on Beyond Time and Errors on Novel Evaluation Methods for Visualization* (New York, NY, USA, 2016), BELIV '16, ACM, pp. 104–111. doi:10.1145/2993901.2993915. 4

[FB81] FISCHLER M. A., BOLLES R. C.: Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM 24*, 6 (1981), 381–395. doi:10.1145/358669.358692. 2

[Gos12] GOSHTASBY A. A.: *Image Descriptors*. Springer London, London, 2012, pp. 219–246. doi:10.1007/978-1-4471-2458-0_5. 1

[HCL05] HEER J., CARD S. K., LANDAY J. A.: prefuse: A toolkit for interactive information visualization. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (New York, NY, USA, 2005), CHI '05, ACM, pp. 421–430. doi:10.1145/1054972.1055031. 4

[HJ93] HOLTZBLATT K., JONES S.: Contextual inquiry: A participatory technique for system design. In *Participatory design: Principles and practices*, Schuler D., Namioka A., (Eds.). Lawrence Erlbaum Associates, Hillsdale, NJ, USA, 1993, pp. 177–210. 4

[IZCC08] ISENBERG P., ZUK T., COLLINS C., CARPENDALE S.: Grounded evaluation of information visualizations. In *Proceedings of the 2008 Workshop on BEyond Time and Errors: Novel evaLuation Methods for Information Visualization* (New York, NY, USA, 2008), BELIV '08, ACM, pp. 6:1–6:8. doi:10.1145/1377966.1377974. 4

[KAF*08] KEIM D., ANDRIENKO G., FEKETE J.-D., GÖRG C., KOHLHAMMER J., MELANÇON G.: Visual analytics: Definition, process, and challenges. In *Information Visualization: Human-Centered Issues and Perspectives*, Kerren A., Stasko J. T., Fekete J.-D., North C., (Eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 2008, pp. 154–175. doi:10.1007/978-3-540-70956-5_7. 1

[LLH*14] LIANG J., LIU X., HUANG K., LI X., WANG D., WANG X.: Automatic registration of multisensor images using an integrated spatial and mutual information (smi) metric. *IEEE Transactions on Geoscience and Remote Sensing 52*, 1 (2014), 603–615. doi:10.1109/TGRS.2013.2242895. 1

[LMNE11] LE MOIGNE J., NETANYAHU N. S., EASTMAN R. D.: *Image Registration for Remote Sensing*. Cambridge University Press, 2011. 1

[Low04] LOWE D.: Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision 60*, 2 (2004), 91–110. doi:10.1023/B:VISI.0000029664.99615.94. 2

[Mun09] MUNZNER T.: A nested model for visualization design and validation. *IEEE Transactions on Visualization and Computer Graphics 15*, 6 (Nov 2009), 921–928. doi:10.1109/TVCG.2009.111. 4

[MZZ*15] MA J., ZHOU H., ZHAO J., GAO Y., JIANG J., TIAN J.: Robust feature matching for remote sensing image registration via locally linear transforming. *IEEE Transactions on Geoscience and Remote Sensing 53*, 12 (2015), 6469–6481. doi:10.1109/TGRS.2015.2441954. 1

[RFP08] RAGURAM R., FRAHM J.-M., POLLEFEYS M.: A comparative analysis of ransac techniques leading to adaptive real-time random sample consensus. In *European Conference on Computer Vision* (2008), Springer, pp. 500–513. doi:10.1007/978-3-540-88688-4_37. 1

[RLA*13] RIND A., LAMMARSCH T., AIGNER W., ALSALLAKH B., MIKSCH S.: TimeBench: A data model and software library for visual analytics of time-oriented data. *IEEE Transactions on Visualization and Computer Graphics 19*, 12 (Dec 2013), 2247–2256. doi:10.1109/TVCG.2013.206. 4

[SJWM05] STONE D., JARRETT C., WOODROFFE M., MINOCHA S.: *User interface design and evaluation*. Morgan Kaufmann, 2005. 4

[Wat02] WATTENBERG M.: Arc diagrams: visualizing structure in strings. In *IEEE Symposium on Information Visualization* (2002), pp. 110–116. doi:10.1109/INFVIS.2002.1173155. 3

[Woo97] WOOD L. E.: Semi-structured interviewing for user-centered design. *interactions 4*, 2 (Mar. 1997), 48–61. doi:10.1145/245129.245134. 4

[ZF03] ZITOVA B., FLUSSER J.: Image registration methods: a survey. *Image and Vision Computing 21*, 11 (2003), 977–1000. doi:10.1016/S0262-8856(03)00137-9. 1

© 2017 The Author(s)
Eurographics Proceedings © 2017 The Eurographics Association.