RAT: A Referencing and Annotation Tool for Digitized Early Maps

Winfried Höhn
University of Luxembourg
winfried.hoehn@uni.lu

Christoph Schommer
University of Luxembourg
christoph.schommer@uni.lu

Early maps are valuable sources as they are a political and cultural mirror of their time. Available nowadays in a digital format, such digitized early maps represent a fundament for getting evidences and detecting novelties in the field of historic research. In particular, modern software systems become designed to support users in receiving beneficial answers to research questions, for example in identifying places and monitoring them over time or in matching and positioning place markers of early maps to modern maps. However, since early maps have been created manually with a high variance of used symbols (like villages, rivers, forests, et cetera), a single map can contain thousands of diverse place markers. There is a wide range of active research on annotating and exploring early maps. For example the YUMA Map Annotation Tool (Simon et al. 2011) is a manual annotation tool for maps and there is also a wide variety of automated tools with different focus (Budig and Dijk 2015; Höhn and Schommer 2016; Shaw and Bajcsy 2011).

The software system RAT is designed to support users in identifying symbolic place markers in digitized early maps, record their name and to link these place markers to modern maps. A more technical description of an earlyer version can be found in Höhn, Schmidt and Schöneberg 2013. RAT facilitates a georeferencing by suggesting the most likely modern places based on an estimated mapping. The number of estimated suggestions can be limited by additional filters, for example by applying a phonetic search (with Cologne phonetics) to places, which sound similar to names given on the map. This allows an identification of modern places, whose historic name has changed over time, but where its name still is close. The database of modern places, which is used for the phonetic matching, can also hold alternative names and all of them are considered for the phonetic matching.

RAT can semi-automatically detected the places markers contained in maps. This means that a sample place marker is firstly selected as template by dragging a rectangle containing it. In this case, RAT uses a template matching algorithm, based on the normalized cross-correlation, to detect place markers. If there are colored place markers in a map, a color segmentation methodology can be used to detect these markers. With respect to the template matching there is a threshold, which specifies how similar an image region has to be to be identified as place marker candidate. This threshold is calculated based on a few sample annotations that the user has to provide. The normalized cross-correlation is a template matching method which can find matches even when the brightness and contrast vary. These properties are crucial for the template matching in early maps to cope with uneven yellowing and fading of the map as also to work in case of differently colored regions. Color segmentation is performed by a mean

shift algorithm, which denotes a clustering algorithm that takes not only the similarity in color but also the positions of the colors in the image into account.

The estimation of geolocation as well as the phonetic search are potential filters to assist the user in identifying the correct place marker. The geolocation estimation is based on a projective transform of spatial coordinates and geographical coordinates calculated by a least squares method. Outliers may become filtered out.

RAT is currently in an advanced implementation phase. At present, we have performed some testing on early maps of the 16th to 18th century. For the biggest test map, containing 3809 place markers and manually annotating an area containing 47 of them, RAT detected 87.7 percent of all place markers correctly while only containing 1.8 percent wrong matches. The annotation features of RAT are not restricted to maps of the 16th to 18th century, but the variance of used symbols and distortions increase the older the maps are. As a result the automatic support given by RAT decreases.

Currently we are working on reducing the cost for manual annotations by taking into account similarities between the map areas in order to stimulate a learning, in order to provide better suggestions for existing places and their georeferencing on new maps. An additional idea is to use the color segmentation for detecting other symbols on the map, in particular forests (mostly shaded green), mountains (mostly brown), and rivers (mostly blue).

References

Budig, B. and Dijk, T. C. (2015) 'Active Learning for Classifying Template Matches in Historical Maps', *Discovery Science: 18th International Conference, DS 2015, Banff, AB, Canada, October 4-6, 2015. Proceedings*, ed. by N. Japkowicz and S. Matwin, Springer International Publishing, pp. 33–47.

Höhn, W., Schmidt, H.-G. and Schöneberg, H. (2013) 'Semiautomatic recognition and georeferencing of places in early maps', *Proceedings of the 13th ACM/IEEE-CS joint conference on Digital libraries*, JCDL '13, Indianapolis, Indiana, USA: ACM, pp. 335–338.

Höhn, W. and Schommer, C. (2016) 'Annotating and Georeferencing of Digitized Early Maps', *To appear in: Digital Humanities (DH)*, Digital Humanities.

Shaw, T. and Bajcsy, P. (2011) 'Automation of Digital Historical Map Analyses', *Proceedings of the IS&T/SPIE Electronic Imaging*, vol. 7869.

Simon, R. et al. (2011) 'Semantically Augmented Annotations in Digitized Map Collections', *Proceedings of the 11th ACM/IEEE-CS joint conference on Digital libraries*, JCDL '11, pp. 199–202.