FACSIMILE CREATION: REVIEW OF ALGORITHMIC APPROACHES

Arie Shaus, Barak Sober, Shira Faigenbaum-Golovin, Anat Mendel-Geberovich, Eli Piasetzky & Eli Turkel
Tel Aviv University

Our research team enjoyed the privilege of collaborating with Benjamin Sass over a period of several years. We are happy to dedicate this article to him and wish to express our gratitude for what has been both a prodigious and enjoyable experience. The purpose of our joint endeavor has been the introduction of modern techniques from computer science and physics to the realm of Iron Age epigraphy. One of the most important issues addressed during our cooperation was the topic of facsimile creation.

Facsimile creation is a necessary preliminary step in the process of deciphering and analyzing ancient inscriptions. Several manual facsimile construction techniques are currently in use: drawing upon collation of the artifact; outlining on transparent paper overlaid on a photograph of the inscription; and computer-aided depiction via software such as Adobe Photoshop, Adobe Illustrator, Gimp or Inkscape (see Summary section below for software web links). Despite their importance for the field of epigraphy, little attention has thus far been devoted to the methodology of facsimile creation (though the recent comprehensive treatment by Parker and Rollston 2016).

Recent decades have seen rapid development and consolidation of various computerized image processing algorithms. Among the most basic and popular tasks in this field is the creation of a black-and-white version of a given image, denoted as image binarization (see Fig.1a–b). Often, such a binarized image is used as a first step for further image processing missions, such as Optical Character Recognition (OCR), texts digitization and text analysis tasks. An algorithmic creation of binarizations can therefore be seen as another method of facsimile creation. Furthermore, a relatively new sub-domain of image processing, Historical Imaging and Processing (HIP), specializes in handling antique documents of different types, periods and origins. Accordingly, binarization algorithms stemming from HIP are even more suitable for archaeological purposes.

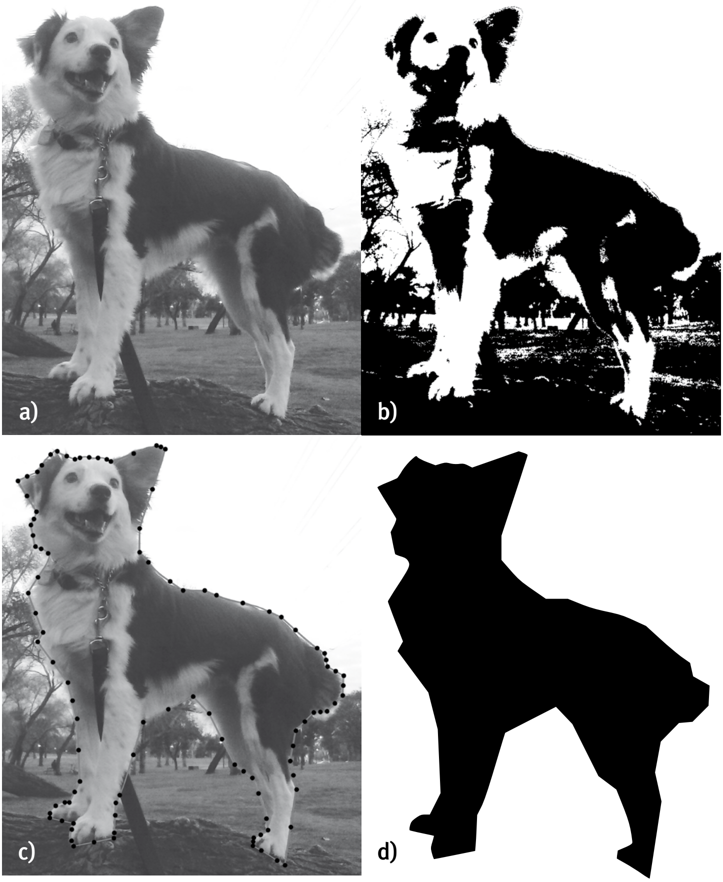

Figure 1: Examples of binarization of natural scene. (a) Original image; (b) automatic binarization; (c) CAGD contour selection; (d) CAGD-based binarization.

However, in some cases, even the most sophisticated technology might not be up to the task of automatically distinguishing between the text and its carrier medium. In such cases, binarization can be achieved via semi-automatic solutions, relating to the field of computer-aided geometrical design (CAGD). Such tools make use of limited information received from the user (e.g., rough guidelines), in order to construct a geometrical approximation of certain shapes (Fig.1c–d). In particular, any character, or even its composing strokes, can be viewed as geometric curves, and approximated accordingly (e.g., upon marking the start and end points by the user).

In what follows, we detail several computer-based methods that can be utilized by the epigraphic community at large, in order to produce automatic or semi-automatic facsimiles. Excluded from the discussion herein are diverse (e.g., multispectral) methods of obtaining the most favorable initial image (see Faigenbaum *et al.* 2012, with examples in Sober *et al.* 2014, Faigenbaum *et al.* 2014, 2015 and Faigenbaum-Golovin *et al.* 2015), which can naturally influence the quality of the facsimiles. We have also excluded computerized paleographical analysis, which can be based upon the facsimiles (e.g., Faigenbaum-Golovin *et al.* 2016).

COMPUTERIZED METHODS VIA AVAILABLE SOFTWARE

Computer image editing software relevant for facsimile production can roughly be divided into two types. The first type (which includes programs such as Adobe Photoshop and the open-source Gimp) deals with editing the images themselves, e.g., brightness and contrast manipulations, color corrections and noise removal. These programs provide sophisticated tools for foreground/background separation. For example, in the case of an inscription's image, ink traces can be separated from the background medium in order to create a facsimile. In Figs. 2a–b, a foreground selection of a character can be seen resulting in a facsimile of the character in Fig. 2c.

The second type of image editing software (such as Adobe Illustrator and its open-source alternative, Inkscape) deals with creating computer-aided drawings and illustrations. These programs provide tools for geometric manipulations of lines and curves, including connecting points, smoothing corners and controlling curvatures. For example, a curve representing a character's edge can be drawn and manipulated (Fig. 3a) in order to create a clean facsimile (Fig. 3b).

Such programs are intuitive and tend to be easy to use. They are semi-automatic (and semi-manual...), in the sense that they necessitate manual input from the user, and at the same time involve autonomous computerized algorithms in order to achieve eye-pleasing results.

AUTOMATIC BINARIZATION METHODS

In the discipline of image processing, and specifically its Historical Imaging and Processing sub-domain, numerous binarization methods were developed (see Shaus *et al.* 2012a with further literature). These methods allow for a relatively fast (typically a matter of seconds per inscription) automatic production of a black-and-white version of a given image. For our purposes, a facsimile can be created out of an inscription image with little or no human intervention.

In Figs. 4–6, we present the performance of several well-established computerized binarization techniques, operating on three different inscriptions, namely Hebrew ostraca Lachish No. 3 (Torczyner 1938) and Arad No. 1 (Aharoni 1981), as well as Arad inscription No. 34 containing Hieratic numerals. The binarization methods are either

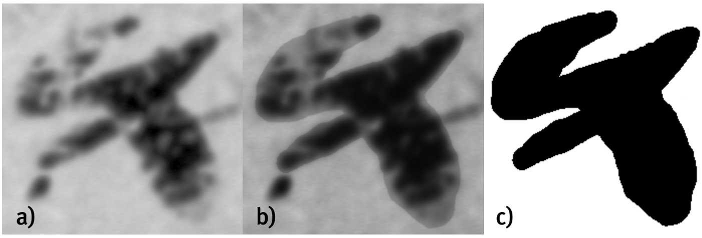

Figure 2: Example of foreground/background separation via selection tool in Adobe Photoshop.(a) the original character «waw»*, taken from Arad Ostracon No. 24 verso, line 3; (b) selection of a character; (c) the resulting facsimile.*

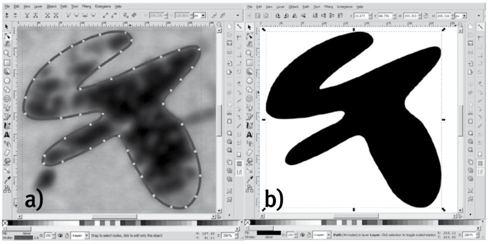

Figure 3: Example of facsimile creation via Inkscape of «waw» *taken from Arad Ostracon No. 24 verso, line 3. (a) drawing of curve representing the character's edge; (b) the resulting facsimile.*

general-purpose (Otsu 1975; Bernsen 1986; Niblack 1986), or specifically adapted for document analysis (White and Rohrer 1983; Sauvola and Pietikainen 2000; Gatos *et al.* 2004). In addition, results of two additional algorithms developed by our team, specifically tailored for ink inscription on clay (Shaus *et al.* 2012a, 2013), are also presented. Both techniques utilize an already existing, imperfect facsimile in order to produce an improved one.

The results show that the algorithms vary greatly in their performance, when applied to different ostraca. In general, it can be seen that the last two algorithms (Shaus *et al.* 2012a, 2013) outperform the others. It is worth mentioning that open source implementations of all of these algorithms can be found online; for more details see Table I below.

COMPUTER-AIDED STROKES RESTORATION

The results stemming from various automatic binarization methods may still be insufficient for epigraphic analysis purposes. This is especially relevant in cases of partly erased characters. Therefore, a semi-automatic procedure for restoration of incomplete handwritten character strokes was developed (Sober and Levin 2016).

The method is based on the representation of a character as a union of individual strokes that are treated independently and later recombined.

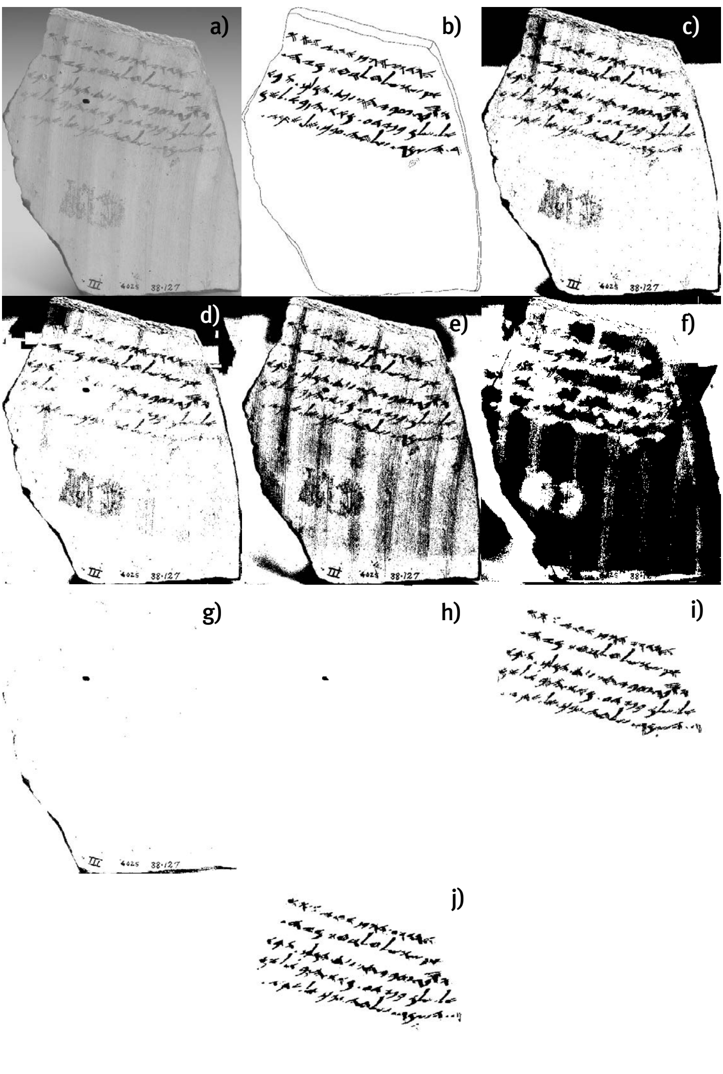

Figure 4: Binarizations of Lachish Ostracon No. 3 verso via various algorithms. (a) ostracon image (b) manual facsimile (c) Otsu (d) Bernsen (e) Niblack (f) White (g) Sauvola (h) Gatos (i) Shaus et al. 2012a (j) Shaus et al. 2013.

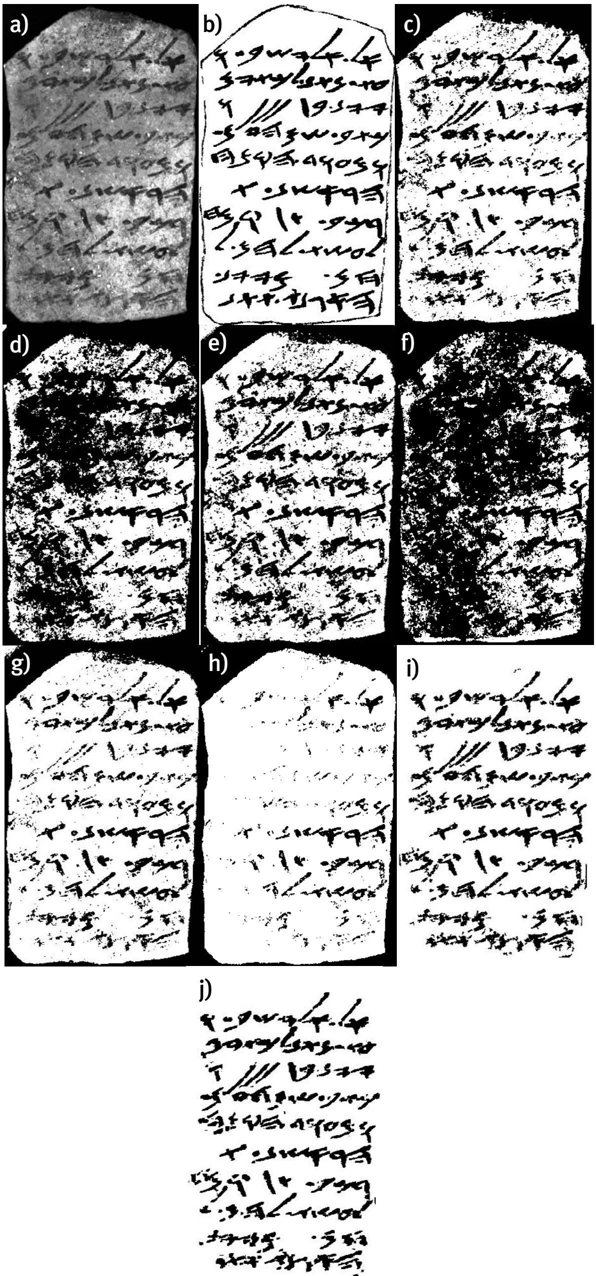

Figure 5: Binarizations of Arad Ostracon No. 1 via various algorithms. (a) ostracon image (b) manual facsimile (c) Otsu (d) Bernsen (e) Niblack (f) White (g) Sauvola (h) Gatos (i) Shaus et al. 2012a (j) Shaus et al. 2013.

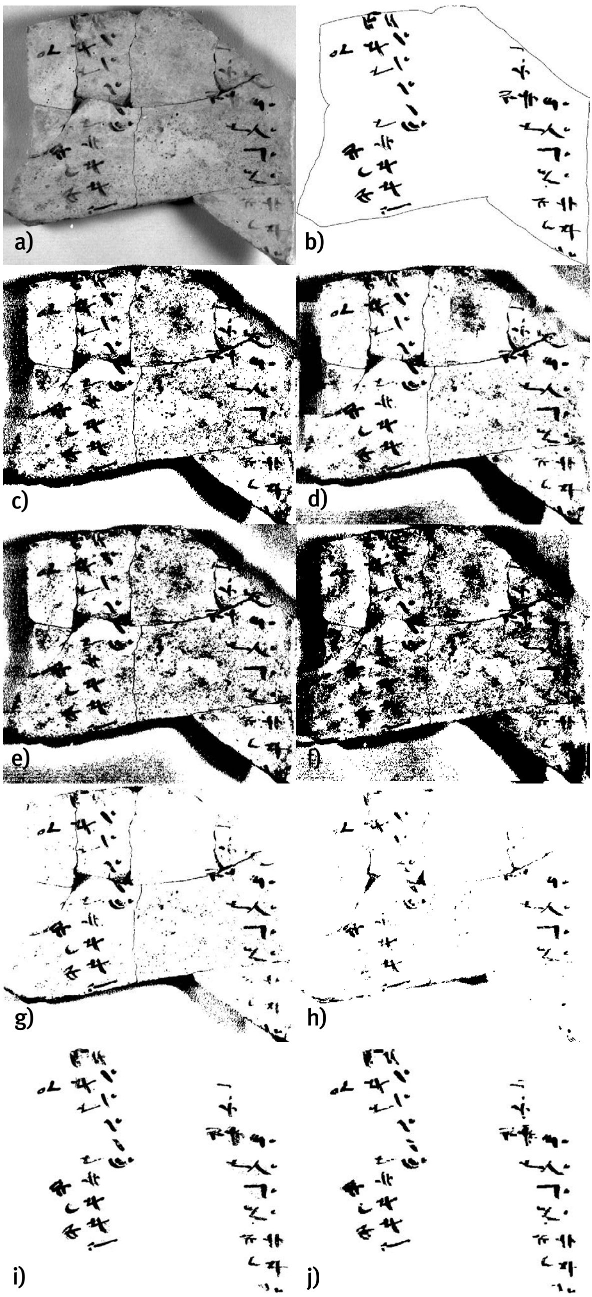

Figure 6: Binarizations of Arad inscription No. 34 via various algorithms. (a) ostracon image (b) manual facsimile (c) Otsu (d) Bernsen (e) Niblack (f) White (g) Sauvola (h) Gatos (i) Shaus et al. 2012a (j) Shaus et al. 2013.

The purpose of stroke restoration is to imitate a reed-pen's movement using several manually sampled key-points. An optimization of the pen's trajectory is then performed (e.g., Fig. 7 for a reconstruction of a Latin character "e"). The end product of the reconstruction is a binary image of the character, incorporating all its strokes. Based on reconstruction of individual characters, a full inscription's facsimile can be created.

In Figs. 8 and 9, we see an automatic reconstruction of *waw* and *yod* characters (Figs. 8d and 9e), respectively, guided by a user's selection of a few key-points (Figs. 8b, 9a and 9c). Additionally, in Fig. 10 a whole facsimile of the Ophel ostracon (Faigenbaum-Golovin *et al.* 2015), created on a character-by-character basis via the algorithm (Sober and Levin 2016), is presented.

QUALITY EVALUATION

A related, yet independent, topic is the evaluation of the quality of either the whole facsimile, or its constituting characters. By quality evaluation we mean an assessment of how well a given facsimile represents the original inscription. Occasionally, epigraphic publications include a facsimile superimposed over the inscription image, but this is performed manually, with no attempt at measuring the quality of the fit (e.g., Hunt *et al.* 2001; Barkay *et al.* 2003). In our view, a more accurate approach is desired.

Our team has developed a facsimile evaluation tool (Shaus *et al.* 2010 and 2012b, with further development in Schaus *et al.* 2016) that measures the adherence of the facsimile to the inscription's image. As a preliminary step, in order to compare a facsimile to an ostracon image, one needs to align them to one another. Our method keeps the computerized image of the artifact unchanged, while the facsimile under examination is slightly rotated/squeezed/stretched in order to achieve the best fit ("registration"). Thereafter, a simple calculation derives the quality of the fit, denoted as CMI ("Clayness Minus Inkness"; the terms were introduced by us) index. The larger the CMI, the better the quality of the fit.

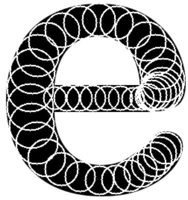

Figure 7: The Latin character "e" as unification of discs. The discs over the character were created using the stroke restoration algorithm.

It should be stressed that the CMI index depends on the inscription image. Camera position and angle (vis-à-vis the inscription) as well as illumination characteristics are critical factors affecting the image, and therefore the CMI index. However, our experiment demonstrated that the CMI *ranking order* of different facsimiles of the same artifact is maintained, even if the underlying inscription's image is substituted.

Figure 11 presents an example of registration between the facsimile and the ostracon image. As can be seen in Figs. 11d and 11e, the adherence of the facsimile to the ostracon image is improved upon registration. In Fig. 12, a comparison between different facsimiles of

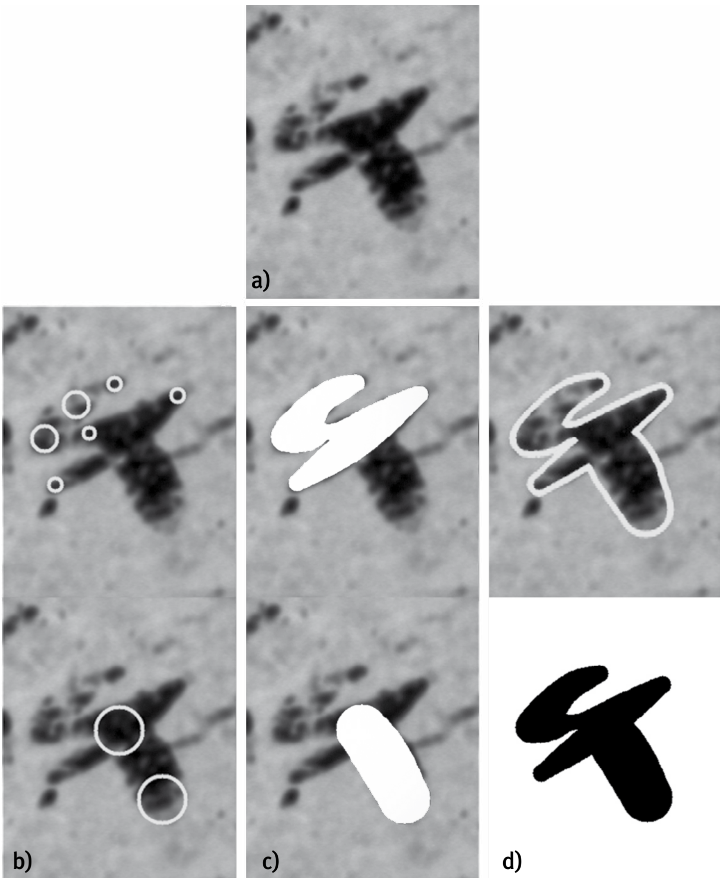

Figure 8: Reconstruction of a waw *from Arad Ostracon No. 24. (a) Image of the character to be reconstructed; (b) manually sampled key points (of top and bottom strokes, respectively); (c) the semi-automatic stroke restorations (of top and bottom strokes, respectively); (d) the reconstructed character (top: the contour of the reconstructed character overlaid on top of the original image; bottom: the binary image of the restored character).*

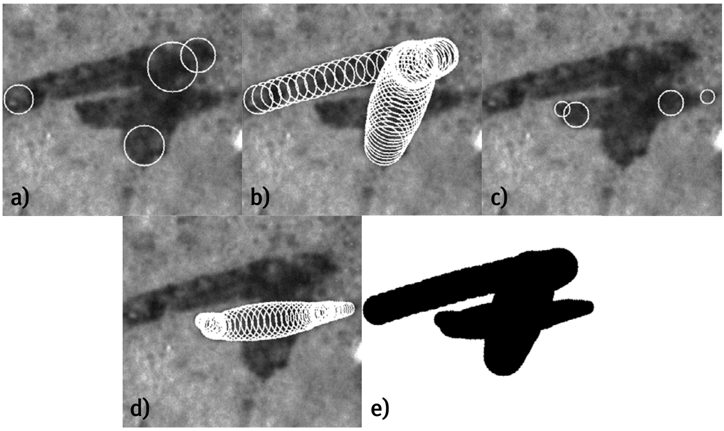

Figure 9: Reconstruction of a yod *from Arad Ostracon No.1, line 3. (a, c) the initially sampled points; (b, d) the reconstructed strokes; (e) the resulting facsimile.*

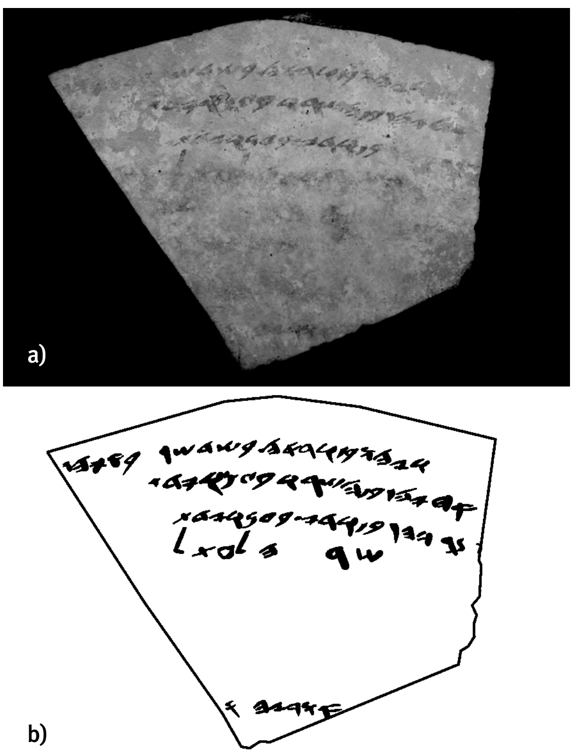

Figure 10: The Ophel Ostracon (Jerusalem). (a) the original grayscale image; (b) a facsimile created by utilizing the stroke restoration algorithm; the restoration was taken from (Faigenbaum-Golovin et al. *2015).*

Figure 11: Registration of Arad Ostracon No. 1 facsimile. (a) ostracon image; (b) manual facsimile; (c) manual facsimile after registration; (d) original facsimile overlaid; (e) registered facsimile overlaid.

the same inscription, created by different individuals, is performed. Both the detailed analysis, and the CMI score, demonstrate that facsimile C is the most accurate depiction of the inscription.

The quality of facsimile's constituting characters can also be evaluated. One option would be to use the same CMI machinery mentioned above. Another approach, especially suited for computer-generated binarizations, is an evaluation of the intrinsic properties of individual characters. Algorithmically-generated binarizations tend to produce undesired features such as "speckle noise," irregular character edges, as well as stains mistaken for character segments. Our algorithm (Faigenbaum *et al.* 2013) analyzes the binarized characters and evaluates them according to the presence or absence of such effects. This allows for an automatic comparison between the outcomes of different binarization algorithms. Figure 13 shows examples of nine binarizations of a single character, ranked according to their quality. Some of the

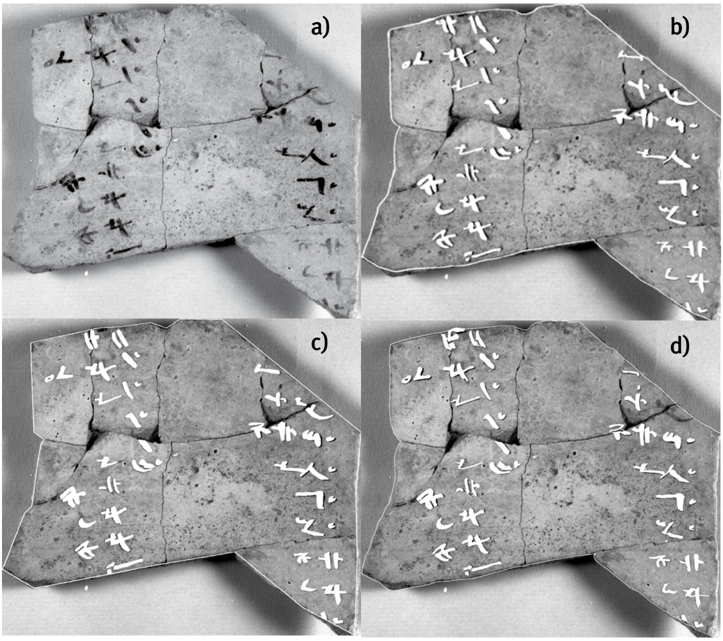

Figure 12: Comparison of Arad inscription No. 34 facsimiles: (a) ostracon image; (b) facsimile A overlaid, CMI= 71.1; (c) facsimile B overlaid, CMI= 82.6; (d) facsimile C overlaid, CMI= 84.0.

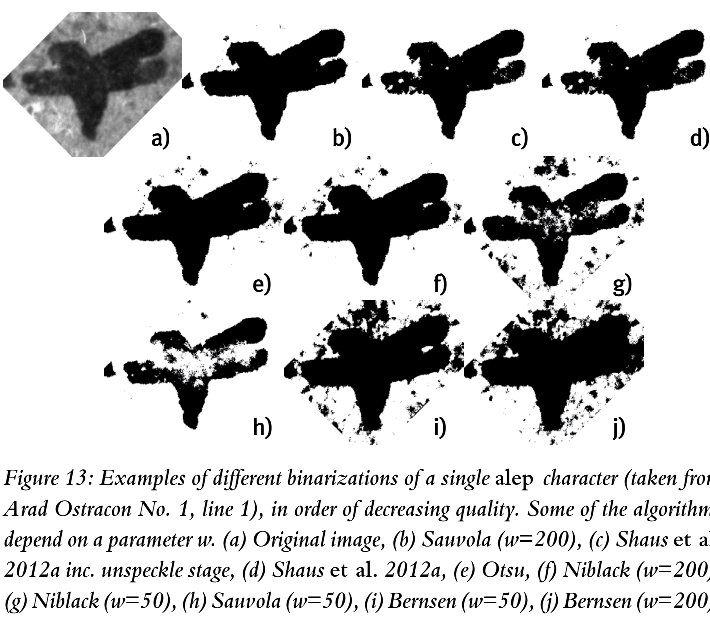

Figure 13: Examples of different binarizations of a single alep *character (taken from Arad Ostracon No. 1, line 1), in order of decreasing quality. Some of the algorithms depend on a parameter w. (a) Original image, (b) Sauvola (w=200), (c) Shaus* et al. *2012a inc. unspeckle stage, (d) Shaus* et al. *2012a, (e) Otsu, (f) Niblack (w=200), (g) Niblack (w=50), (h) Sauvola (w=50), (i) Bernsen (w=50), (j) Bernsen (w=200).*

Table I: List of Image Processing Software Available to the Aid of epigrapher

Software	Description and remarks	Weblink	Free & Open Source
Adobe Photoshop	Image editing	adobe.com/products/photoshop.html	No (commercial)
Gimp	Image editing	gimp.org	Yes
Adobe Illustrator	Creation of drawings	adobe.com/products/illustrator.html	No (commercial)
Inkscape	Creation of drawings	inkscape.org	Yes
Gamera	Document analysis toolkit, based on Python programming language. Implements various document binarization algorithms.	gamera.informatik.hsnr.de	Yes
Ostraca binarization	Implementation of (Shaus et al. 2012a), based on Python programming language.	www-nuclear.tau.ac.il/~eip/ostraca/Code/OstracaBinarizations2012.zip	Yes
Stroke restoration	Implementation of (Sober and Levin 2016), based on Python programming language.	www-nuclear.tau.ac.il/~eip/ostraca/Code/StrokeRestoration2016.zip	Yes
Facsimile evaluation	Implementation of (Shaus et al. 2010, Shaus et al. 2012b), based on Python programming language.	www-nuclear.tau.ac.il/~eip/ostraca/Code/FacsimileEvaluation2010.zip	Yes
Character evaluation	Implementation of (Faigenbaum et al. 2013), based on Python programming language.	www-nuclear.tau.ac.il/~eip/ostraca/Code/CharacterEvaluation2013.zip	Yes

binarization algorithms depend on a parameter *w*, which determines the size of window within which each algorithm step is performed.

SUMMARY

The article presents several approaches to facsimile creation, via modern computational tools. While some of the tools are familiar to the epigraphic community, it is our conviction, that other algorithmic approaches, presented above, can be quite beneficial for the creation of facsimiles and the quality evaluation processes. A cautious reader may prefer the simplicity of use of complete off-the-shelf image processing or drawing software, while a more adventurous researcher might prefer installing, experiencing and experimenting with stand-alone algorithms. In either case, as a service to the interested reader, we present a table with links to the recommended software.

ACKNOWLEDGMENTS

The research reported here received initial funding from the Israel Science Foundation—F.I.R.S.T. (Bikura) Individual Grant no. 644/08, as well as the Israel Science Foundation Grant no. 1457/13. The research was also funded by the European Research Council under the European Community's Seventh Framework Programme (FP7/2007-2013)/ERC grant agreement no. 229418, and by an Early Israel grant (New Horizons project), Tel Aviv University. This study was also supported by a generous donation from Jacques Chahine, made through the French Friends of Tel Aviv University. Arie Shaus is grateful to the Azrieli Foundation for the award of an Azrieli Fellowship. The kind assistance of Shirly Ben-Dor Evian, Sivan Einhorn, Noa Evron, Myrna Polak and Hatti Shaus is greatly appreciated. Ostraca and facsimile images are courtesy of the Institute of Archaeology, Tel Aviv University and of the Israel Antiquities Authority.

REFERENCES

Aharoni, Y. 1981. *Arad Inscriptions*. Israel Exploration Society.

Barkay, G., Lundberg, M.J., Vaughn, A.G., Zuckerman, B. and Zuckerman, K. 2003. The Challenges of Ketef Hinnom: Using Advanced Technologies to Reclaim the Earliest Biblical Texts and Their Context. *Near Eastern Archaeology* 66: 162–171.

Bernsen, J. 1986. Dynamic Thresholding of Grey-Level Images. *Proceedings of the Eighth International Conference on Pattern Recognition*, Paris: 1251–1255.

Faigenbaum, S., Shaus, A., Sober, B., Turkel, E. and Piasetzky E. 2013. Evaluating Glyph Binarizations Based on their Properties. *Proceedings of the 13th ACM Symposium on Document Engineering (DocEng 2013)*, Florence:127–130.

Faigenbaum, S., Sober, B., Shaus, A., Moinester, M., Piasetzky, E., Bearman, G., Cordonsky, M. and Finkelstein I. 2012. Multispectral Images of Ostraca: Acquisition and Analysis. *Journal of Archaeological Science* 39: 3581–3590.

Faigenbaum, S., Sober, B., Finkelstein, I., Moinester, M., Piasetzky, E., Shaus, A. and Cordonsky, M. 2014. Multispectral Imaging of Two Hieratic Inscriptions from Qubur el-Walaydah. *Ägypten und Levante* XXIV: 349–353.

Faigenbaum, S., Sober, B., Moinester, M., Piasetzky, E. and Bearman, G. 2015. Multispectral Imaging of Tel Malhata Ostraca. In: Beit-Arieh, I. and Freud, L., eds. *Tel Malhata: A Central City in the Biblical Negev, Volume I* (Monograph Series of the Institute of Archaeology of Tel Aviv University 32). Tel Aviv: 510–513.

Faigenbaum-Golovin, S., Rollston, C.A., Piasetzky, E., Sober, B. and Finkelstein, I. 2015. The Ophel (Jerusalem) Ostracon in Light of New Multispectral Images. *Semitica* 57: 113–137.

Faigenbaum-Golovin, S., Shaus, A., Sober, B., Levin, D., Na'aman, N., Sass, B., Turkel, E., Piasetzky, E., Finkelstein, I. *Proceedings*. Algorithmic Handwriting Analysis of Judah's Military Correspondence Sheds Light on Composition of Biblical Texts, Vol. 113, no. 17: 4664–4669.

Gatos, B., Pratikakis, I. and Perantonis, S.J. 2004. An Adaptive Binarization Technique for Low Quality Historical Documents, *Lecture Notes in Computer Science* 3163: 102–113.

Hunt, L., Lundberg, M.J. and Zuckerman, B. 2001. Eyewitness to the Past: Reclaiming Ancient Inscriptions with Modern Technologies through USC's West Semitic Research and InscriptiFact Projects. *Biblos* 50: 79–100.

Niblack, W. 1986. *An Introduction to Digital Image Processing*. Prentice-Hall: 115–116.

Otsu, N. 1975. A Threshold Selection Method from Gray-Level Histograms. *Automatica* 11:23–27.

Parker H.D.D. and Rollston C.A. 2016. The Epigraphic Digital Lab: Teaching Epigraphy in the Twenty-First Century C.E. *Near Eastern Archaeology* 79.1: 44–56.

Sauvola, J. and Pietikainen, M. 2000, Adaptive Document Image Binarization, *Pattern Recognition* 33: 225–236.

Shaus, A., Finkelstein, I. and Piasetzky, E. 2010. Bypassing the Eye of the Beholder: Automated Ostraca Facsimile Evaluation. *Maarav* 17: 7–20.

Shaus, A., Turkel, E. and Piasetzky, E. 2012a, Binarization of First Temple Period Inscriptions: Performance of Existing Algorithms and a New Registration Based Scheme, *Proceedings of the 13th International Conference on Frontiers in Handwriting Recognition (ICFHR 2012)*, Bari, Italy: 645–650.

Shaus, A., Turkel, E. and Piasetzky, E. 2012b. Quality Evaluation of Facsimiles of Hebrew First Temple Period Inscriptions. In: *Proceedings of the 10th IAPR International Workshop on Document Analysis Systems (DAS 2012)*, IEEE. Gold Coast, Australia: 170–174.

Shaus, A., Sober, B., Turkel, E. and Piasetzky, E. 2013, Improving Binarization Via Sparse Methods. *Proceedings of the 16th International Graphonomics Society Conference (IGS 2013)*, Nara, Japan:163–166.

Shaus, A., Sober, B., Turkel, E., and Piasetzky, E. 2016,

Beyond the Ground Truth: Alternative Quality Measures of Document Binarizations Proceedings of the 15th International Conference on Frontiers in Handwriting Recognition (ICFHR 2016), forthcoming.

Sober, B., Faigenbaum, S., Beit-Arieh, I., Finkelstein, I., Moinester, M., Piasetzky, E. and Shaus, A. 2014. Multispectral Imaging as a Tool to Enhance the Reading of Ostraca, *Palestinian Exploration Quarterly* 146: 185–197.

Sober, B. and Levin, D. 2016. Computer Aided Restoration of Handwritten Character Strokes. arXiv preprint arXiv:1602.07038.

Torczyner, H., 1938. *Lachish I: The Lachish Letters*. Oxford.

White, J.M. and Rohrer, G.D. 1983. Image Thresholding for Optical Character Recognition and Other Applications Requiring Character Image Extraction. *IBM Journal of Research and Development* 27: 400–411.