Quality Evaluation of Facsimiles of Hebrew First Temple Period Inscriptions

Arie Shaus, Eli Turkel
The Department of Applied Mathematics
Tel Aviv University
Tel Aviv 69978, Israel
e-mails: ashaus@post.tau.ac.il, turkel@post.tau.ac.il

Eli Piasetzky
The Sackler School of Physics and Astronomy
Tel Aviv University
Tel Aviv 69978, Israel
e-mail: eip@tauphy.tau.ac.il

***Abstract*—The discipline of First Temple Period epigraphy (the study of writing) relies heavily on manually-drawn facsimiles (black and white images) of ancient inscriptions. This practice may unintentionally mix up documentation and interpretation. The article proposes a new method for evaluating the quality of the facsimile. It is based on a measure, comparing the image of the inscription to the registered facsimile. Some empirical results, supporting the methodology, are presented. The technique is also relevant to quality evaluation of other types of facsimiles and binarization in general.**

Facsimile, quality evaluation, registration, CMI, epigraphy, First Temple Period, Iron Age, ostracon.

I. INTRODUCTION

Most of the Hebrew texts of the Iron Age (First Temple period) Israel and Judah were written in ink on papyrus. However, these documents did not survive the journey down the millennia. The most abundant among the meaningful surviving texts were written in ink on ostraca (pieces of pottery). Important ostraca corpora were unearthed in Samaria [1], Lachish [2] and Arad [3].

The discipline of Iron Age epigraphy relies heavily on manually-drawn facsimiles (binary documents) of these ostraca inscriptions. However, facsimiles crafted by hand may unintentionally mix up documentation with interpretation. Surprisingly, despite their importance for the field of epigraphy, to the best of our knowledge no attention has thus far been devoted to facsimile quality evaluation (see however [4], intended for the general epigraphic audience). Some epigraphical publications (e.g. [5] and [6], though they do not deal with ostraca) superimpose the facsimile over the inscription image, but this is performed manually with no attempt at measuring the quality of the fit. On the other hand, document analysis approaches of binarization results quality such as [7-9], require the creation of a manual or semi-automatic ground truth (which can be potentially influenced by the human factor, see [10]). Candidate binarized images (facsimiles) are then graded in one way or another, according to the quality of their fit to the ground truth, with no reference to the inscription image.

As an alternative, we shall establish an effective facsimile quality measure, simple enough to be explained to non-scientists. The measure will be based upon registering the facsimile directly to an inscription image (kept constant). The performance of the measure will be tested in order to assess its reliability.

II. FACSIMILE EVALUATION

Given a gray-level $O(p)$ ostracon image, and the facsimile image $F(p)$, $p \in [1,m] \times [1,n]$, several image-fit functions can be defined (as will be explained later, given images of different sizes, a registration of the facsimile image to the ostracon image is required). Natural candidates for comparing different versions are the commonly used L_1 and L_2 norms. While the latter may entail nice analytic properties, it also has the tendency to heavily penalize large deviations, which might lead to non-robust behavior. Thus, we prefer the L_1 norm. Since the facsimile documents are binary we denote $I = \{p \mid F(p) = 0\}$ ("ink pixels") and $C = \{p \mid F(p) = 255\}$ ("clay pixels"), which will function as a partition of $O(p)$ induced by $F(p)$. We begin with the following measure which we wish to *minimize*:

$$\sum_{p \in I \cup C} |F(p) - O(p)|. \quad (1)$$

As the facsimile image is restricted to 0=ink and 255=clay values, it is easy to show the equivalence to *maximizing*

$$\sum_{p \in C} O(p) - \sum_{p \in I} O(p) - \sum_{p \in I \cup C} F(p). \quad (2)$$

First, it is expected that among the various facsimiles depicting a given inscription, the relative proportions of ink and clay pixels (as opposed to their location) would be almost constant. Thus, the rightmost sum can be neglected. Second, a possible problem with this measure is the dominance of the left component over the middle one, as the "ink" pixels (within the facsimile image) are relatively rare. A more "egalitarian" approach is to use averages (i.e. $\mu_D = \underset{p \in D}{Avg}\{O(p)\}$, where D is a domain within an image) instead of sums, thus biasing the measure towards the ink pixels:

978-0-7695-4661-2/12 $26.00 © 2012 IEEE
DOI 10.1109/DAS.2012.70

$$CMI(F,O) = \mu_C - \mu_I \,. \qquad (3)$$

Within the last formula, μ_C is denoted by "clayness", while μ_I is denoted by "inkness". The overall measure is thus abbreviated as CMI ("clayness minus inkness").

The CMI index exhibits a connection to the Otsu [11] binarization measure, which is equivalent to:

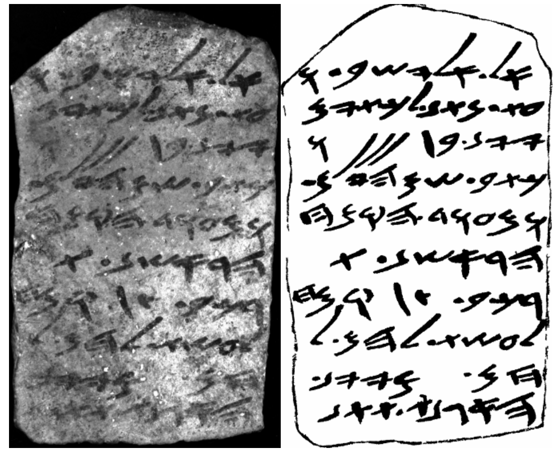
$$\omega_0 \omega_1 (\mu_1 - \mu_0)^2 \,, \qquad (4)$$

where μ_0 and μ_1 are averages of the two pixel "populations", and ω_0, $\omega_1 = 1 - \omega_0$ are their appropriate proportions. Since, $\omega_0 \omega_1$ reaches a maximum when $\omega_0 = \omega_1 = 0.5$, Otsu's criteria may be viewed as the square of the CMI measure, biased towards the histogram median. On the other hand, it should be noted that the underlying problems are quite different: while Otsu deals with unknown pixel population separation by histogram thresholding, our mission is to evaluate existing pixel populations induced by another (facsimile) image.

The main difficulty of comparing two documents is that the manually-crafted facsimile may depict the ostracon from a slightly different angle, or to be somewhat rotated with respect to the ostracon image. For that reason, there arises the need for a registration between the facsimile image and the ostracon image, resulting in the facsimile registered image fitting the dimensions of the ostracon. For registration purposes, we use the same CMI target function. We design a registration that allows for rotations, with subsequent height/width adjustments in order to impose the dimensions of the ostracon image on the facsimile. Therefore, the optimization is only performed with respect to one parameter (the angle θ, sampled with 0.1 degrees resolution):

$$\theta_{\max} = \arg\max_{\theta} CMI(R_\theta(F), O) \,. \qquad (5)$$

The final CMI is chosen to be $CMI(R_{\theta_{\max}}(F), O)$. An example of an ostracon image, a facsimile image, their initial fit and their CMI-based registration can be seen at Figs. 1 and 2 (depicting Arad ostracon No. 1, see [3]).



(a) (b)

Figure 1. Example of (a) an ostracon image, (b) a facsimile image

(a) (b)

Figure 2. Example of (a) initial facsimile-ostracon fit, (b) CMI-based registration

Using the two images of Fig. 2, it is quite easy to explain the mechanics of the CMI measure even to a person lacking an image processing background:

1. The average pixel value "below" the facsimile ("inkness") should be as small (dark) as possible.
2. The average pixel value "not obstructed" by the facsimile ("clayness") ought to be as large (bright) as possible.

Merging the "inkness" and the "clayness" is desirable. The simplest formula combining the two is the CMI ("clayness minus inkness") measure, given in (3).

III. Experimental Results

A. Methodology Verification I

A comparison of several facsimiles of the same Arad No. 34 ostracon (containing hieratic, i.e. Egyptian, numerals; see [3] and Fig. 3), created by different individuals, was performed.

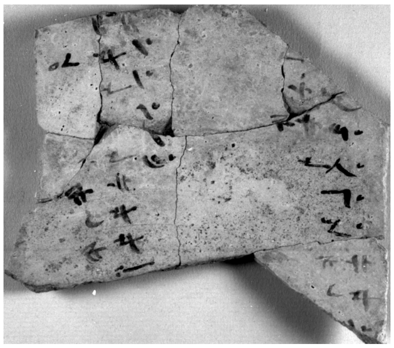


Figure 3. Arad Ostracon No. 34.

Two of the facsimiles were drawn by epigraphers and one by an artist. In order to avoid identifying these scholars, they are denoted below as A, B and C. The results of the CMI-based registration and evaluation can be seen in Figs. 4-6. The performed analysis received experts' approval, confirming the soundness of the approach and of the results.

Fig. 4 shows the ostracon image compared with facsimile A. The registration of the facsimile is excellent (attesting to the effectiveness of the CMI measure). The overall fit of the overlaid facsimile A is good. Nevertheless, the facsimile characters are not always correlated with the ostracon image characters (see for example Fig. 4 in the lower left), which results in typical "shadows" (un-obstructed ink). The strokes are not always long enough (e.g., Fig. 4, lower right). The strokes themselves are somewhat wide. The resulting CMI score is 71.1.

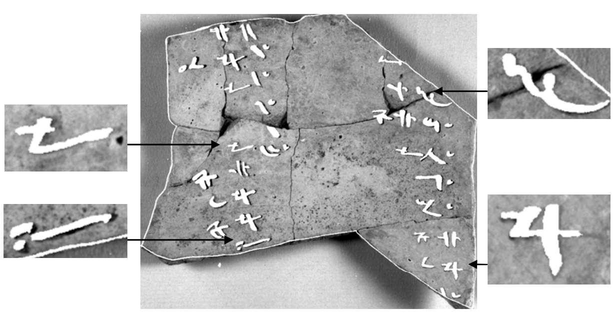


Figure 4. Overlaid Facsimile A, CMI = 71.1

The overlaid facsimile B, seen on Fig. 5, again has good registration. This time, the fit is also good and the facsimile characters seem to be in better correlation with the ostracon image. On the other hand, the character strokes are sometimes a bit too wide (e.g., Fig. 5, upper left) and the overlap is not always perfect. Also notice cases where the strokes are not long enough (e.g., Fig. 5, upper left and lower right). Overall, the facsimile is of better quality and the CMI measure, 82.6, is understandably higher.

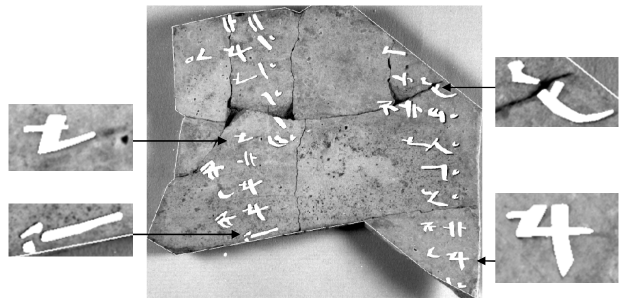


Figure 5. Overlaid Facsimile B, CMI = 82.6

In the case of facsimile C, Fig. 6, the registration is outstanding. The characters are narrow and "crisp"; they seem to be in excellent agreement with the ostracon image. The CMI score, 84.0, is justifiably the highest among the three facsimiles. This is despite one possibly missing character, taken for a scratch or stain (Fig. 6, upper right); owing to the fact that empirically, the CMI measure prefers mistaking ink for clay than vice versa (i.e. it is "conservative" with respect to "character-invention", but will not heavily penalize for dropping dubious character).

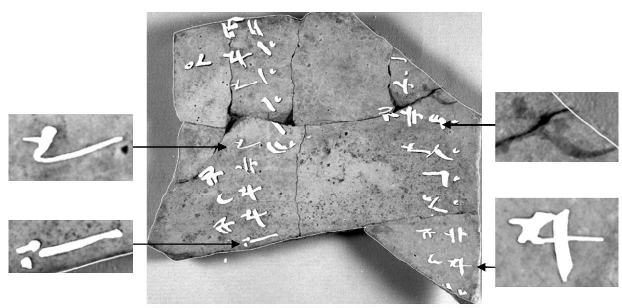


Figure 6. Overlaid Facsimile C, CMI = 84.0

In conclusion, the procedure correctly indicates that facsimile C is the best of the three. The superb registration, also based on a CMI index, is also a good indicator of the soundness of this measure.

B. *Methodology Verification II*

It follows from the definition (3), that the CMI index depends on the ostracon image. Hence, it is reasonable to assume that camera position and angle (vis-à-vis the ostraca), as well as illumination characteristics, are significant factors in obtaining the image and so may change not only the CMI scores, but also their relative rankings. In order to empirically test the degree of invariance of the CMI measurements and their rankings to ostracon image change, we used yet another image of the same ostracon, which can be seen in Fig. 7.

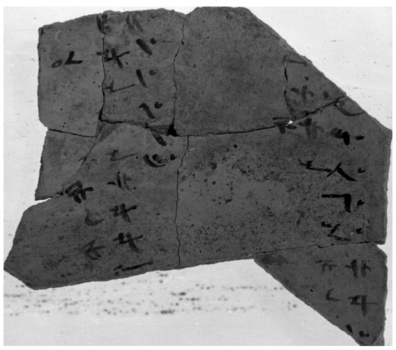


Figure 7. Another image of Arad Ostracon No. 34.

Comparing Figs. 3 and 7, it is obvious that the latter image is markedly different from the former. It is viewed from a different angle, it is slightly rotated, the background is brighter and lacks shadows, and the ostracon itself is darker. We repeated the previous methodology verification stage, the protocol included the usage of the unchanged A, B and C facsimiles and an application of the same CMI registration and quality estimation apparatus. Table I summarizes the results of the first and the second methodology verifications.

TABLE I. RESULTS FOR TWO METHODOLOGY VERIFICATIONS

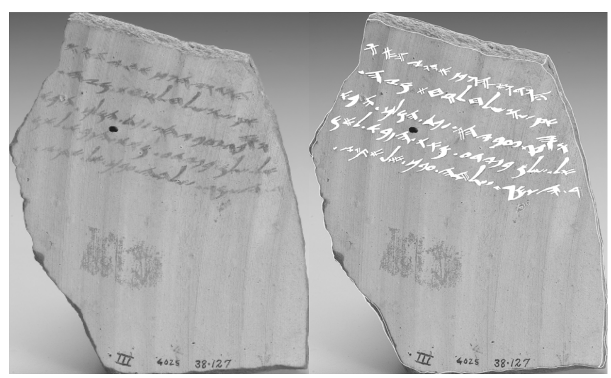
Facsimile	CMI score using Image #1	CMI score using Image #2
A	71.1	64.5
B	82.6	71.6
C	84.0	75.1

The change in the magnitude of the results is hardly surprising, as the image has a different grayscale level distribution. What is important is that the order of the CMI scores is maintained despite the completely different ostracon images. The A score is lower than B, while the C score is higher than both A and B. Therefore, despite using substantially different ostracon images, the relative results of the facsimile evaluation remain effectively the same. This current empirical validation shows, that the facsimile rankings are fairly invariant even under drastic ostracon image alterations.

IV. DISADVANTAGES

Several shortcomings in the method and its verification ought to be mentioned:

- In any given quality assessment metric, some cases can lead to misleading results. The CMI index is no exception. As an illustration, assume an extremely faint character, with gray levels comparable to typical clay gray levels. In such a case, omitting the character from the facsimile might be preferable from the CMI index point of view. A compromise could be to draw only a silhouette of such a faint character (in fact, this is an accepted epigraphical practice). Another example is that of a dark stain. From the CMI index perspective it may be better to record it on the facsimile as if it were a character. As already stated, the CMI score is “conservative” with respect to “character-invention”, and is not expected to benefit substantially from the addition of a letter.
- The CMI-based evaluation depends on registration of the facsimile to the ostracon image. We use a registration of a very simple type, which empirically works for our purpose. More sophisticated registrations can be considered (see [12] for a survey on the subject). Registering on a per-character basis, for instance, may lead to another quality measure and allow for low scale correction of the drawing. Such a method of registration may also compensate for nonlinear camera distortions.
- The results presented here were obtained from a limited number of test cases. In addition to these, we successfully experimented with several other ostraca (e.g. Fig. 8, Lachish ostracon No. 3) and tested the technique on different scales (1/4 and 1/8). Additional research is expected to strengthen the confidence in this methodology.



(a) (b)

Figure 8. Another example of (a) ostracon image, (b) a fit to a high-quality facsimile

V. Conclusions And Future Directions

We presented a facsimile (or binarization) quality measure (CMI), based upon registering the facsimile directly to an inscription image. The technique was tested on three facsimiles and two different images of the same ostracon (Arad No. 34). The CMI grades received for the facsimiles reflect their relative merits. Based on the CMI scores, the ranking of the facsimiles are minimally influenced by the ostracon image. It can therefore be concluded that the proposed technique is sound and can be used to evaluate the accuracy of a facsimile in relation to the original ostracon. Our method can also be used as a facsimile-drawing aid for epigraphers, allowing for iterative improvements of the facsimile drawing.

Though we only presented an evaluation of the technique on ostraca, other types of inscriptions, such as papyri, parchment, stone, seal impressions etc., along with their facsimiles, can also be handled, with minor adaptations.

It is possible to extend the methodology for the use of a computer-based binarization (automatic facsimile production). In other words, although the facsimiles cannot be regarded as a perfect representation of the ostracon, they may be used as a "first draft" for a new facsimile, produced by a computer. This will be dealt with in future publication.

Acknowledgment

This study was supported by the F.I.R.S.T. (Bikura) Individual Grant no. 644/08 and by the European Research Council under the European Community's Seventh Framework Program (FP7/2007-2013) / ERC grant agreement no. 229418.

Arie Shaus is grateful to the Azrieli Foundation for the award of an Azrieli Fellowship. We would also like to acknowledge the valuable help of Ms. Yael Barschak of the Israel Antiquities Authority, Prof. Christopher Rollston of the Emmanuel School of Religion, Prof. Bruce E. Zuckerman of the University of Southern California, and Ms. Judith Dekel, Ms. Shira Faigenbaum, Prof. Israel Finkelstein, Ms. Myrna Pollak, Prof. Benjamin Sass, Mr. Pavel Shargo and Mr. Barak Sober, all of Tel Aviv University.

Ostracon images: courtesy of the Institute of Archaeology, Tel Aviv University; and of the Israel Antiquities Authority. Facsimiles: courtesy of Ms. Judith Dekel; of the Israel Exploration Society (see [3]); and of Prof. C. A. Rollston (see [13]).

References

[1] G. A. Reisner, C. S. Fisher, and D. G. Lyon, Harvard Excavations at Samaria. Cambridge, 1924.

[2] H. Torczyner et al. Lachish I: The Lachish Letters. London, 1938.

[3] Y. Aharoni, Arad Inscriptions. Israel Exploration Society, 1981.

[4] A. Shaus, I. Finkelstein, E. Piasetzky, "Bypassing the eye of the beholder: automated ostraca facsimile evaluation," Maarav, in press.

[5] L. Hunt, M. J. Lundberg, B. Zuckerman, "Eyewitness to the past: reclaiming ancient inscriptions with modern technologies through USC's West Semitic Research and InscriptiFact projects," Biblos 50, 1, 2001, pp. 79-100.

[6] G. Barkay, M. J. Lundberg, A. G. Vaughn, B. Zuckerman, K. Zuckerman, "The challenges of Ketef Hinnom: using advanced technologies to reclaim the earliest biblical texts and their context," Near Eeastern Archaeology, 66:4, 2003, pp. 162-171.

[7] K. Ntirogiannis, B. Gatos and I. Pratikakis, "An Objective Evaluation Methodology for Document Image Binarization Techniques," 8th International Workshop on Document Analysis Systems (DAS), 2008, pp. 217-224.

[8] B. Gatos, K. Ntirogiannis, I. Pratikakis, "DIBCO 2009: document image binarization contest," International Journal on Document Analysis and Recognition (IJDAR), special issue on performance evaluation, Vol.14, No. 1, 2011, pp. 35-44.

[9] I. Pratikakis, B. Gatos, K. Ntirogiannis, "H-DIBCO 2010 – Handwritten Document Image Binarization Competition," 12th International Conference on Frontiers in Handwriting Recognition (ICFHR), 2010, pp. 727-732.

[10] E. H. Barney Smith. "An analysis of binarization ground truthing," Proceedings of the 9th IAPR International Workshop on Document Analysis Systems (DAS), 2010, pp. 27-33.

[11] N. Otsu, "A threshold selection method from gray-level histograms," IEEE Trans. Systems Man Cybernet. Vol. 9 (1), 1979, pp. 62–66.

[12] B. Zitová, J. Flusser, "Image registration methods: a survey," Image and Vision Computing 21, 2003, pp. 977–1000.

[13] C. A. Rollston, "Scribal education in ancient Israel: the Old Hebrew epigraphic evidence," BASOR 344, 2006, pp. 47–74.