Integrating Copies Obtained from Old and New Preservation Efforts

Yoram Zarai
School of Electrical Engineering, Tel Aviv University
Tel Aviv, Israel
Email: yoramzar@mail.tau.ac.il

Tamar Lavee Nachum Dershowitz Lior Wolf
School of Computer Science, Tel Aviv University
Tel Aviv, Israel
Email: tamarla1@mail.tau.ac.il; {nachum,wolf}@cs.tau.ac.il

***Abstract*—The Dead Sea Scrolls were discovered in the Qumran area and elsewhere in the Judean desert beginning in 1947 and were photographed in infrared in the 1950s. Recently, the Israel Antiquities Authority embarked on an ambitious project to digitize all the fragments using multi-spectral cameras. We describe a method that utilizes information from both of these image sets: the highly detailed multispectral images and the older infrared images, which preserve the state of the fragments as it was shortly after discovery. We use a two-step registration procedure to align the image sets. First, a coarse global transformation is applied to the whole image of the new set, producing a rough alignment, followed by a fine, local wrapping based on interest point matching. The aligned images can be used to improve image binarization and to identify and repair fragments that have degraded further over the years. Additionally, the fine alignment parameters can be used for coarse attribute classification, such as the period when written.**

I. Introduction

The Dead Sea Scrolls (DSS) were written between the third century BCE and the first CE, and were discovered in the Qumran area and elsewhere in the Judean desert beginning in 1947. These fragmentary documents have taught the world a great deal about Jewish history in the late Second Temple period and about sectarian Judaism, and they have provided much background knowledge relevant to the birth of Christianity. In addition, the Dead Sea Scrolls have enhanced our understanding of the textual transmission of the Bible. Following their discovery, the scrolls were photographed during the 1950s, using infrared (IR) film and Kodak filters [1].

Removal from the stable environment in the Judean Desert caves has placed the scrolls at great risk of deterioration. Furthermore, in the early years, scholars mishandled the scrolls unwittingly, causing irreversible damage, as they attempted to piece fragments together. Recently, the Israel Antiquities Authority, in collaboration with Google, Inc., embarked on an ambitious project to digitize all the fragments using multi-spectral cameras, providing high-resolution color and IR images, and to make them, as well as the old IR images, available to everyone on the Internet [2], [3].

When examining the two image sets, one can see a number of conspicuous differences: naturally, the new images are of better quality and higher contrast, but in many cases the fragment in the new image is damaged, while the older image preserves the state of the fragments before they degraded further. Many images are of assemblages of several tiny fragments, and in some cases the layout was changed before the new image was taken. Fig. 1 shows a typical example of the differences between the images: part of the bottom edge of the fragment has eroded, causing letters to disappear, and a piece at the top is displaced in the new image.

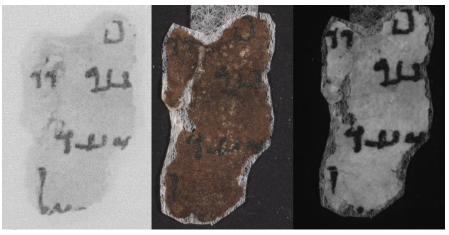

Figure 1. Old infrared image (left), new color image (center) and new high-contrast infrared image (right).

Our goal is to combine the advantages of both sets of images. We use a two-step registration procedure to align the two image sets in a coarse-to-fine approach, and we show how registration may contribute to more accurate results for various processing and analyzing tasks.

II. Image Registration

Image *registration* is the process of aligning two (or more) images of the same scene, taken at different times, positions and/or capturing devices. The task is to map each pixel from one image to its corresponding pixel in the other. Registration is widely used in many image analysis applications, aiming to increase the overall information gained from the multiple images. A comprehensive survey of image registration methods may be found in [4] (and reference [26] therein).

Document registration is needed in many applications, including office automation, document databases and digital libraries. Template matching [5], line information and cell structure utilization [6] and projective geometry [7] are just a few among many methods proposed in recent years. The special case of historic document analysis, containing handwritten text, is gaining considerable momentum as more and more collections are being digitized and made available on the Internet and in libraries. In addition to the distortions that must be handled by document registration methods, historical documents present additional challenges, including poor image quality, tears, bleed-through and in some cases local small displacements of small sub-fragments within the image. A

method to automatically select interest points for historical document registration, ignoring bleed-through areas, was presented in [8]. However, the algorithm assumes specific stroke patterns when selecting interest points, which is not necessarily valid in our case.

The types of distortion between the two DSS image sets (old and new) that are encountered vary from global misalignment (e.g. rotation, scaling) to local deformations sustained differently by different parts of the document. Thus, a registration algorithm applied to the DSS's old and new images must be able to address all these types of distortion. Efforts to address the problem of alignment of old and new DSS fragments have been performed by the team acquiring the new images [Gregory Bearman, personal communication, 2013].

We take our inspiration from the work of Jagalur et al. [9], describing a registration algorithm comprising two main steps: a coarse similarity transformation step (a subset of affine transformation including translation, rotation and scaling, excluding shearing) is first applied globally to one of the images (new image) to bring it into rough alignment with the other (old image). Then, a fine local warping step is performed to handle local misalignments. The fine step uses an entropy measure to select interest points (landmarks) in the reference image, finds matching interest points in the target image based on mutual information (MI) and performs inverse warping by bi-cubic interpolation to finely align the new image.

A. Algorithm

As the algorithm in [9] was targeted to mice brain images, it required modification to be suited for registration of historical, hand-written texts. We present our registration algorithm, highlighting the changes relative to [9], which will be referred to as the baseline algorithm.

Given the old (reference) image and the new (target) image, the algorithm first performs a similarity transformation on the target image to try and roughly align it to the reference image, as in [9]. The main objective of this step is to reduce fine-step complexity. We perform the search for the best transformation parameters over a scaled-down version of the images, where the scaling is coarsely determined by the image sizes. The best transformation parameters (vertical and horizontal scales, rotation and vertical and horizontal translations) are searched for sequentially. We tested different metrics to asses the transformation performance during the search, including sum-of-square-distance (SSD), earth mover distance (EMD) [10], quadratic-chi (QC) [11], normalized-cross-correlation (NCC) and mutual-information (MI). We found that the results using the MI metric outperformed all the others.

To present the different metrics' performance, we introduced scale, phase and translation errors to a registered target image and used it, with the corresponding reference image, to test the similarity transformation step. We used a registered target image to emulate as closely as possible the actual registration of the two image sources. Table I presents the transformation parameter's l_1 errors (averaged over a few images), where scale and translation errors are normalized to image size.

TABLE I
SIMILARITY TRANSFORMATION PARAMETER ERRORS.

	Scale Mean Error	Rotation Mean Error (rad)	Translation Mean Error
MI	0.020	0.011	0.015
SSD	0.114	0.072	0.211
EMD	0.217	0.070	0.167
QC	0.212	0.084	0.233
NCC	0.181	0.087	0.110

B. Probabilistic Graphical Model

After the similarity transformation, we use, differently than [9], a probabilistic graphical model (PGM) to finely align local image patches.

We first identity interest points in the reference image. For the most part, each reference image contains one fragment with some background, where the text region occupies most (but not all) of the fragment area. Since the fragments, in general, are distorted, we would like to search for interest points over the text region only and avoid other areas in the image (in contrast to the search performed by the baseline algorithm). This will reduce unnecessary warping distortions performed by the registration's fine step. To do so, we use Otsu's method [12] to find a global threshold, mask the reference image and define the search region as the mask's convex hull. Interest points are then defined as patches having a large amount of information over the search region. To ease the computation burden, the search is performed over a pre-defined grid. We use the brightness histogram's entropy as a measure of the patch's information. The patch size is coarsely determined by the reference image size (with patch sizes ranging from 50×50 pixels for small images to 100×100 pixels for larger ones).

Fig. 2 illustrates the search region in a reference fragment. The red *'s are the pre-defined search grid, each represents the upper-left corner location of a patch. We use a patch size of 100×100 with grid spacing of 50 pixels in each dimension. The blue +'s are the selected interest points, where we allow up to 25% patch overlap.

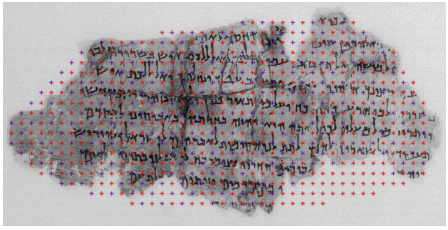

Figure 2. Search grid (red *) and selected interest points (blue +).

Having detected interest points in the reference image, the corresponding interest points in the target image need to be determined. Since we assume that the similarity transformation coarsely aligned the two images, the location of each corresponding interest point in the target image should reside in a local area surrounding the reference's interest point location. In general, there is some correlation between neighboring interest

points in the reference image, which should be exploited. This is formalized using PGM as described next.

Let l^m be the pixel location (row and column) of interest point m in the reference image (which is already identified as indicated by the + set in Fig. 2) and by d^m the displacement (row and column) of the corresponding interest point pixel location in the target image. Then d^m is a (2D) random variable with realization over a local search space defined by a $P\times P$ square region around l^m (meaning $d^m \in [-P/2, P/2]\times [-P/2, P/2]$). Thus, the corresponding target's interest point location is $l^m + d^m$. Our goal is to find the "best" d^m value for all m's (all interest points). We use MI to capture the fidelity of each d^m realization by computing the MI between the reference patch, located at l^m, and the target patch, located at $l^m + d^m$, for each d^m realization. The overall MI measure ($P\times P$ matrix) is then defined as $\theta_m(d^m)$. This is done for each interest point.

To formulate the underlying correlation between neighboring interest points in the reference image, we define $\phi_k(d^m, d^n)$ to be the correlation measure (potential) between adjacent interest points m and n having displacement d^m and d^n respectively. The subscript k, defining different correlation potentials, identifies the neighborhood space index, where $k = 1$ corresponds to interest points (within the reference image) having distance of $\|l^m - l^n\| = g$, $k = 2$ to interest point distance of $\sqrt{2}g$, $k = 3$ to $2g$, etc., where g is the search grid spacing in each dimension. Given two neighboring interest points m and n, the value of ϕ_k is computed by:

$$\phi_k(d^m, d^n) = C(k) \cdot (L - \|d^m - d^n\|)$$

where $L \equiv max\{\|d^m - d^n\|\}$ over all d^m and d^n realizations, $C(k)$ is a constant per k and $C(k) > C(k+1)$ for all k. This suggests that the correlation assumption weakens as the two interest points become farther apart.

Given the displacement fidelity measure, $\theta_m(d^m)$, and the different neighbor correlation potentials, $\phi_k(d^m, d^n)$, the integrated (over all interest points m) displacement realization log-probability is (based on PGM):

$$Log(P(\{d^m\})) = \sum_m \theta_m(d^m) + \sum_{k,m,n} \phi_k(d^m, d^n) - \log Z$$

where Z is the partition function ensuring that the probabilities sum to 1 and $n \in A^m(k)$, where $A^m(k)$ defines the interest point m's neighbor space corresponding to the neighbor index k. Finding the optimal displacement for all interest points is then equivalent to finding the maximum a-posteriori (MAP) argument of the integrated log-probability function.

C. Inference

To find the optimum displacement for each interest point, we perform inference based on convex belief propagation (BP), as presented in [13]. The method uses an approximate convex function to solve the MAP assignment (as exact inference is NP-hard), guaranteeing the convergence of the BP algorithm. Detailed information can be found in [13].

A factor graph is constructed with variable nodes representing $\theta_m(d^m)$ and a factor node for each $\phi_k(d^m, d^n)$. An edge connects a variable node to its corresponding factor node. We use the distributed convex-BP implementation presented in [14]. It allows parallel inference on p processors by partitioning the graph into p subgraphs and enforcing consistency between the subgraphs' shared nodes (inter-subgraph edges).

The solution of the distributed convex-BP is the set of optimal interest points' displacements. Thus, the corresponding optimal location of the reference's interest point m in the target image is $l^m + \widehat{d^m}$, where $\widehat{d^m}$ is the optimal displacement value.

Given the set of interest points in the reference image and the corresponding interest points in the target image, Delaunay triangulation is performed using the interest points as vertices, followed by inverse warping, as in [9]. A bi-cubic interpolation is used to map each pixel in the target image to its new location based on the interest points (vertices) of the surrounding triangle in the reference image.

D. Results

The proposed algorithm was tested on many of the DSS old and new image sets. We used two neighborhood space indexes for the correlation potential: $\phi_1(d^m, d^n)$ describing the potential values for two interest points having a Euclidean distance of g and $\phi_2(d^m, d^n)$ having a distance of $\sqrt{2}g$, where g is the search-grid spacing in each dimension.

To present the registration results, we superimpose the two images within a color image where the red channel contains the reference image and the green channel the target image (either before or after registration). Thus a perfectly aligned text will appear black. The arrows point to sample regions showing marked improvement.

Fig. 3(a) presents the registration result on a Herodian fragment. The figure contains the superimposed image before registration (left side), after registration using the baseline algorithm [9] (center) and after registration using the proposed algorithm (right size). Note the performance differences at the lower-left edge of the fragment between the two algorithms. Fig. 3(b) is a Hasmonean fragment; note the performance differences at the right edge of the fragment between the two algorithms. Fig. 3(c) is another Herodian fragment; note the performance differences on the middle and bottom sides between the two.

III. Applications

We show now how the registration of images from the two sets enables us to further process and analyze the DSS images.

A. Binarization

Binarization is an essential stage of every document image-processing system. For images of historical manuscripts, this is a very challenging task, as degradation and different kinds of damage lead to variability both in the background and the foreground. Many local methods have been suggested to binarize non-uniform document images ([15], [16]). Other

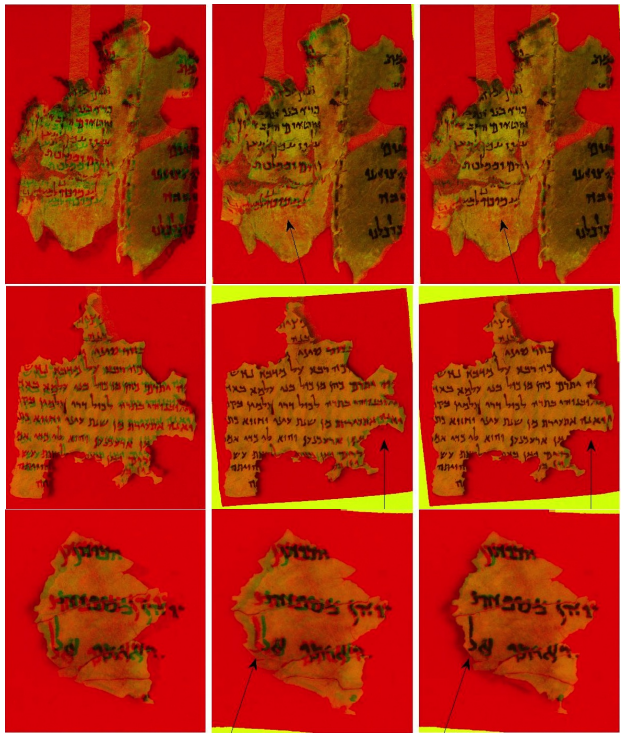

Figure 3. Registration input (left), output using baseline (center) and output using proposed algorithm (right). The arrows point to sample regions showing marked improvement.

methods have been proposed to solve typical problems of historical documents ([17], [18]).

We investigated the benefits of multispectral imaging, as is available for the DSS. We show how the different wavelengths are combined to achieve a binary image, and then use the registration results and integrate the old image as an additional reference. We show how using the old image may produce a binary image that contains information lost over the years.

Each fragment of the DSS is being photographed in 12 different wavelengths, between 445nm and 924nm. Our assumption was that the different wavelength images together contain more information than any one separately. As a baseline, we use the results of Sauvola's method [15] on the single highest wavelength (924nm).

Given k different wavelengths, we classify each pixel as foreground or background in the following way: we train an SVM classifier using an equal number of foreground and background pixels, sampled from the baseline binary image. The features of each pixel are its grayscale values in the k different images. Then, the achieved SVM model is used to classify every pixel in the image according to its k features. The obtained image is our binary result. After registering the old and new images, we can use the old image as an additional reference, as it may contain details that have subsequently degraded. Therefore, we repeat the procedure using the 12 wavelengths plus the old image.

To evaluate the output obtained from different combinations, we used a manually-made ground-truth binary image, and calculated three measures: S_{tot}, the fraction of pixels that were correctly classified as foreground or background; S_{fg}, the success rate on the foreground area, namely, the fraction of the foreground that was correctly classified; S_{bg}, the success rate on the background area, namely, the fraction of the background that was correctly classified. Table II shows an example of the scores obtained for different combinations of reference images of a typical fragment.

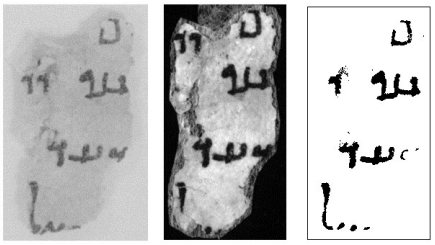

Figure 4. Old image (left), new image of wavelength 924nm (center) and binary image sampled from old and new references (right).

TABLE II
BINARIZATION SCORES OF DIFFERENT COMBINATIONS OF REFERENCES.

	S_{tot}	S_{fg}	S_{bg}
New image 924nm	0.925	0.999	0.921
Old image	0.873	0.992	0.874
New image 924nm & old image	0.928	0.999	0.923
All new wavelengths	0.929	0.999	0.925
All new wavelengths & old image	0.927	0.999	0.923

The combination of the old and the new images seems to achieve better accuracy than any single one of them, specifically with regard to background identification, meaning that it helps avoid noise caused by ruptures, stains or parchment texture. This is very important for the next stages of processing, such as character recognition, where we want to minimize the number of errors. Investigation of the contribution of the different reference images and obtaining an optimal combination is left for future research.

Fig. 4 shows another example where using the old image as reference contributes to the accuracy of the binary result. The fragment in the new image has deteriorated so that letters at the bottom disappeared, but in the old image they are still apparent. The binary image on the right combines the high contrast of the new image with the information that was lost at the bottom of the fragment.

B. Damage Detection

Some of the fragments were significantly damaged over the years, so that it is almost impossible to align the two images with one another. In such cases, the registration results cannot be used, but they do give an indication that the fragment's condition is very poor. In severe cases, we found that it is preferable to use the old image for the next processing stages, although its quality is lower.

However, in many other fragments, a milder deformation occurred, which influenced mostly the edges of the fragments. In some cases, as seen in Fig. 4, parts of letters that clearly show in the old image have disappeared by time the new

images were taken. Another common difference between the image sets is displacement of small pieces that are part of a composite, presumably caused during conservation efforts. This is demonstrated in the upper part of the fragment in Fig. 4. The two registered images can be used to identify such cases by comparing the margins of the old and the new photo after they are aligned. This can serve as a starting point for a "damage repair" algorithm that will use the old image to fix the new one, and vice versa.

C. Classification

We considered the three DSS periods, Herodian (Her), Hasmonean (Has) and Hellenistic-Roman (Hel), and evaluated the discrimination capabilities of the following measures: (1) Interest-point displacement norm, $R(s) = E\{\|d^m\|\}$, where the average is over all interest points in a fragment. There is one measure per registration image set s (old and new images). (2) Jaccard index, $J(s)$, between the interest points' convex hull in the reference image and the corresponding convex hull in the target image (one measure per set). (3) Mutual information, $M(s)$, of the registration's coarse step (one measure per set). Analysis of variance (ANOVA) was performed with each of the above measures, together with a multiple comparison test to evaluate discrimination between pairs of periods. The results are provided in Table III. Each row corresponds to a different measure and gives the p-value (under the null hypothesis of equal means) and the 95% confidence interval (CI) range [low mid high] of the estimated means difference for each period pair.

TABLE III
ANOVA TEST RESULTS FOR PERIOD DISCRIMINATION.

	p-value	CI(Her,Has)	CI(Her,Hel)	CI(Has,Hel)
R(s)	0.0175	[-6.00 -1.65 +2.70]	[-11.54 -6.20 -0.86]	[-9.03 -4.55 -0.06]
J(s)	0.0024	[-0.02 +0.07 +0.16]	[-0.16 -0.06 +0.05]	[-0.22 -0.13 -0.04]
M(s)	0.0111	[-0.15 -0.03 +0.08]	[-0.31 -0.17 -0.03]	[-0.25 -0.13 -0.01]

These results suggest a rejection of the null hypothesis at the 5% significance level for all measures. In addition, both the displacement norm and mutual information can discriminate between the Hellenistic-Roman period and each of the other two, whereas the Jaccard index can only discriminate between the Hellenistic-Roman and Hasmonean periods.

IV. Conclusion

We have demonstrated a registration method that was developed to address the characteristics of images of historical document. We used this method to align together image sets that were taken at two different points in time. The images show the changes that occurred in a fragment's state, and the alignment between them allows for the combining of information that is preserved in each of the two image sets. The parameters used in different stages of the registration may also highlight the features of the different images, and the differences between the image sets. As part of future research, we plan to further investigate the ability of the two image sets to contribute to DSS preservation efforts and to enhance the results of further image processing steps, like character recognition and fragment classification. The registration method presented here is also applicable to the many other collections that have been imaged multiple times over the years.

Acknowledgment

We would like to thank the Israel Antiquities Authority, and especially Pnina Shor and Orit Kuslansky, for their collaboration and help with the Dead Sea Scroll images. Research was supported in part by a grant from Google, Inc. We thank M. Jagalur, C. Pal, E. Learned-Miller, R. T. Zoeller, and D. Kulp for sharing the code of their alignment software with us, as well as Gregory Bearman for fruitful discussions.

References

[1] G. Bearman, "Imaging the Dead Sea Scrolls for conservation purposes," *SPIE Newsroom*, Dec. 29 2008.
[2] "The Leon Levy Dead Sea Scrolls digital library website," http://www.deadseascrolls.org.il/home.
[3] S. Tanner and G. Bearman, "Digitizing the Dead Sea Scrolls," in *Proc. Archiving Conf., IS&T*, 2008, pp. 119–123.
[4] B. Zitová and J. Flusser, "Image registration methods: A survey," *Image Vis. Comput.*, vol. 21, pp. 977–1000, 2003.
[5] H. Peng, F. Long, and Z. Chi, "Document image recognition based on template matching of component block projections," *IEEE Trans. Pattern Analysis and Machine Intell.*, vol. 25, no. 9, pp. 1188–1192, 2003.
[6] K.-C. Fan and M.-L. Chang, "From document identification using line structure based features," in *Proc. 14th International Conference on Pattern Recognition*, vol. 2, 1998, pp. 1098–1100.
[7] R. Safari, N. Narasimhamurthi, M. Shridhar, and M. Ahmadi, "Document registration using projective geometry," *IEEE Trans. on Image Processing*, vol. 6, no. 9, pp. 1337–1341, 1997.
[8] J. Wang, M. S. Brown, and C. L. Tan, "Automatic corresponding control points selection for historical document image registration," in *Int. Conf. on Document Analysis and Recognition*, 2009, pp. 1176–1180.
[9] M. Jagalur, C. Pal, E. Learned-Miller, R. T. Zoeller, and D. Kulp, "Analyzing in situ gene expression in the mouse brain with image registration, feature extraction and block clustering," *BMC Bioinformatic*, vol. 21, pp. 1–21, 2007.
[10] C. T. Yossi Rubner and L. Guibas, "The earth mover's distance as a metric for image retrieval," *Intl. J. Computer Vision*, vol. 40, no. 2, pp. 99–121, 2000.
[11] O. Pele and M. Werman, "The quadratic-chi histogram distance family," in *Proc. ECCV*, 2010, pp. 749–762.
[12] N. Otsu, "A threshold selection method from gray-level histograms," *IEEE Transactions on Systems, Man., and Cybernetics*, vol. 9, no. 1, pp. 62–66, 1979.
[13] T. Hazan and A. Shashua, "Norm-product belief propagation: Primal-dual message-passing for approximate inference," Arxiv preprint arXiv:0903.3127, 2009.
[14] A. Schwing, T. Hazan, M. Pollefeys, and R. Urtasun, "Distributed message passing for large scale graphical models," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2011.
[15] J. Sauvola and M. Pietikäinen, "Adaptive document image binarization," *Pattern Recognition*, vol. 33, no. 2, pp. 225–236, 2000.
[16] Y. Yang and H. Yan, "An adaptive logical method for binarization of degraded document images," *Pattern Recognition*, vol. 33, no. 5, pp. 787–807, 2000.
[17] B. Su, S. Lu, and C. L. Tan, "Binarization of historical document images using the local maximum and minimum," in *Proc. 9th IAPR Intl. Workshop on Document Analysis Systems*, 2010, pp. 159–166.
[18] I. Bar-Yosef, I. Beckman, K. Kedem, and I. Dinstein, "Binarization, character extraction, and writer identification of historical Hebrew calligraphy documents," *International Journal on Document Analysis and Recognition*, vol. 9, no. 2, pp. 89–99, 2007.