Improving Binarization via Sparse Methods

Arie SHAUS [a], Barak SOBER [a], Eli TURKEL [a] and Eli PIASETZKY [b]
[a] *The Department of Applied Mathematics, Tel Aviv University, 69978 Tel Aviv, Israel*
[b] *The Sackler School of Physics and Astronomy, Tel Aviv University, 69978 Tel Aviv, Israel*
ashaus@post.tau.ac.il, baraksov@post.tau.ac.il, turkel@post.tau.ac.il, eip@tauphy.tau.ac.il.

Abstract. A binarization of challenging historical inscription is improved by means of sparse methods. The approximation is based on a binary dictionary learned by k-medians and k-medoids algorithms from a clear source. Some preliminary results show superiority to the existing binarization with respect to fine features such as strokes continuity, deviations from a straight line, edge noise and the presence of stains. The k-medians dictionary-learning scheme shows a robust behavior when initial patches database is reduced.

1. Introduction

Binarization of First Temple Period ostraca (ink written on clay sherds, see Fig 1a) inscriptions is a challenging problem even for the state-of-the-art algorithms in the field of Historical Document Analysis (Shaus & al., 2012a). Currently, Iron Age palaeographers prefer to use manually produced binary images, termed facsimiles (Shaus & al., 2010 and 2012b). Following our previous research, we aim at laying the ground for automatic production of such binary images, with the aid of already existing binarizations of unsatisfactory quality.

In the last decade, there has been a rapid development in the field of sparse coding methods, for various Image Processing tasks. We explore the possibility of using similar techniques in order to produce an improved inscription binarization.

Let $D \in \mathbb{R}^{n\times K}$ be an over-complete dictionary that contains K atoms $\{d_j\}_{j=1}^{K}$ for columns. A signal $y \in \mathbb{R}^n$ can be represented or approximated by a sparse linear combination of these atoms, $y \approx Dx$, $x \in \mathbb{R}^K$. The approximation is chosen in the sense that $\|y - Dx\|_p \le \varepsilon$ (p is commonly selected to be 1, 2 or ∞), with the sparsity of x minimized by the l_0 norm, counting the number of nonzero coefficients. In other words:

$$\min_x \|x\|_0 \text{ subject to } \|y - Dx\|_p \le \varepsilon \quad (1)$$

Since an exact determination of the sparsest representation is an NP-hard problem (Davis & al., 1997), there arises a need for reducing the size of the dictionary D. This can either be performed prior to solving the minimization problem in Eq. 1, or in parallel. The most prominent methods for dictionary training and quantization are k-means (Gersho & al., 1991), ML - Maximum Likelihood methods (Olshausen & al., 1996 and Lewicki & al., 1999), MOD - Method of Optimal Directions (Engan & al., 1999), MAP - Maximum A-Posteriori Probability (Kreutz-Delgado & al., 2000), UONB - Union of Orthonormal Bases (Lesage & al., 2000) and the highly popular k-SVD (Aharon & al., 2006).

In order to produce a binarization, we would like to use a dictionary containing black and white patches. In addition, the representation should avoid any combinations of such atoms in order to maintain the binary property of the resulting approximation. Only one atom d_j from the dictionary, with a corresponding coefficient of $x_j = 1$ should be used for each approximated patch, while $\forall i \ne j \quad x_i = 0$. Therefore, the problem is slightly changed: we set the l_0 norm of x exactly to 1, while we wish to find the best approximation of y. Though it is tempting to approximate the inscription image by itself (using it as a source for y), empirically its imperfect binarized images perform much better in this role.

Thus, using the above-mentioned formalism, our problem is composed of two steps:

1. Learn an over-complete **binary** dictionary $D \in \{0, 255\}^{n\times K}$, representing black and white patches.
2. For each patch y in an existing **imperfect binarization**, find $\min_x \|y - Dx\|_p$ subject to $\|x\|_0 = 1$.

2. Proposed Solution

The most demanding task is the first step of dictionary learning. It would seem that after obtaining a large database of patches, one would be able to plug-in any of the above mentioned off-the-shelf solutions. However, this is not the case. Almost all of these methods can be essentially interpreted as generalizations of the k-means algorithm. Thus, the constructed dictionary would almost certainly result in gray level, rather than black and white values. Moreover, most of the methods assume that d_j are part of a linear space (possibly coupled with an inner product), which is untrue in our case.

One possible solution may be applying "extensive" methods, avoiding the need for quantization altogether and using a large patches database as a dictionary. This may not always be feasible. A more elegant suggestion would be the utilization of the k-medians (Jain & al., 1981) or k-medoids (Kaufman & al., 1987) methods, which results in K atoms with appropriate values of 0 or 255 (the uncommon case of 127.5 can be assigned to either one of them).

The formal description of the algorithm, for each inscription, is:

1. Collect a large database of clean black and white patches. We use the above mentioned hand-made facsimiles as the primary source.
2. Learn a dictionary based on the database, using either k-medians or k-medoids method. Alternatively (if allowed computationally), use the whole database as a dictionary.
3. For each patch y in the existing imperfect binarization, find the most suitable replacement $d_j \in D$, chosen by the solution of $\min_{\|x\|_0=1} \|y - Dx\|_p$. If the patches y overlap, construct the binarization by prioritizing the patches with a better score.

3. Preliminary Experimental Results

Our experiments tested the soundness and performance of the technique with respect to different algorithm parameters and various ostraca inscriptions. The following parameters were kept constant: patch size = 11x11 pixels (in the initial database, patches are sampled on 3 pixels grid, with at most 73% overlap), dictionary size = 100 atoms (except for the extensive dictionary solution, where a typical number of atoms was 1000 up to 30000), number of repeated random initializations for k-medians and k-medoids = 100.

The first experiment tested the relationship between the best binary images available for our medium (Shaus & al., 2012a), and the improved binarization obtained by k-medians, k-medoids and extensive dictionary methods. The results for the ostracon of Arad #1 can be seen in Fig. 1, while the results for the ostracon of Arad #34 can be seen on Fig. 2.

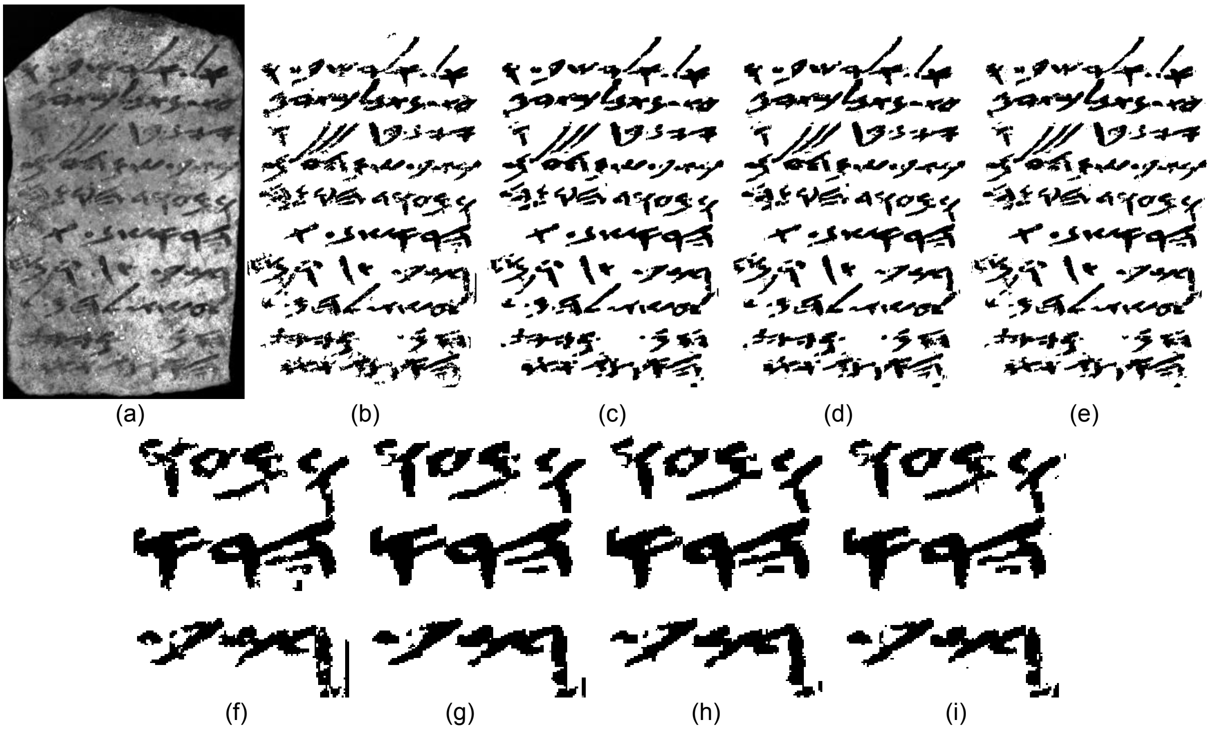

Figure 1. Arad #1 (a) ostracon image (b) best available binarization (c) k-medians result (d) k-medoids result (e) extensive dictionary result; zoom on right-center: (f) best available binarization (g) k-medians result (h) k-medoids result (i) extensive dictionary result

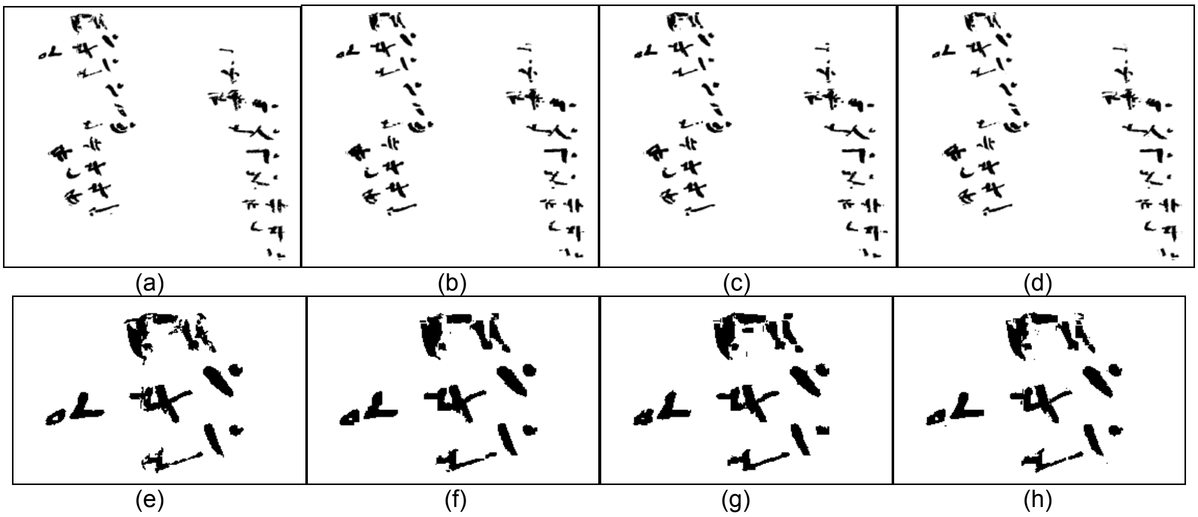

Figure 2. Arad #34 (a) best available binarization (b) k-medians result (c) k-medoids result (d) extensive dictionary result
Zoom on top-left: (e) best available binarization (f) k-medians result (g) k-medoids result (h) extensive dictionary result

The results show that the performance of the sparse models rivals that of the best binarization. In fact, when looking on a fine-grained details like strokes continuity, deviations from straight line, edge noise and the presence of stains, k-medians (on Fig. 1) and k-medoids (on Fig. 2) outcomes are superior to the available binarization, though not by a far margin. We note that despite its heavy computational burden, the extensive dictionary solution does not surpass the k-medians and k-medoids in both cases. It may be that the optimally fitting patches of the extensive dictionary result lack the robustness of the k-medians and the k-medoids solutions.

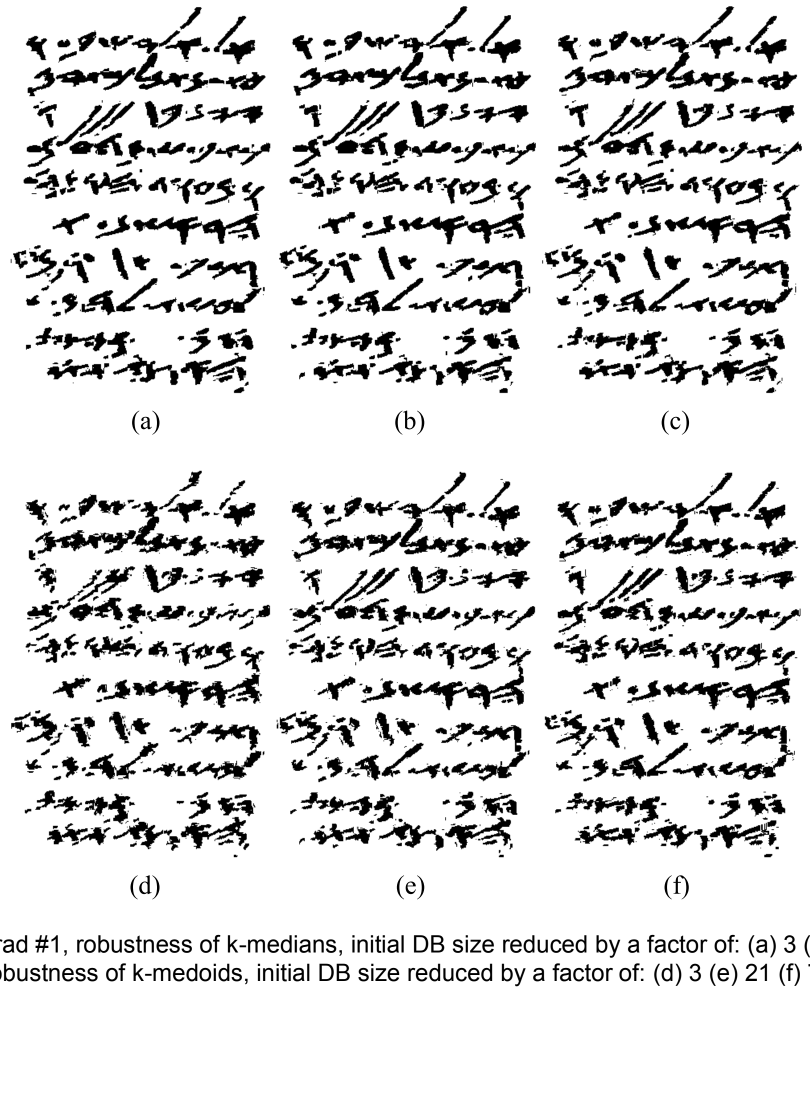

Figure 3. Arad #1, robustness of k-medians, initial DB size reduced by a factor of: (a) 3 (b) 21 (c) 75.
Robustness of k-medoids, initial DB size reduced by a factor of: (d) 3 (e) 21 (f) 75.

The robustness of the different methods was put to a test in the second experiment. The initial patches database was reduced by a factor of 3 by removing duplicate patches (in a non-robust scenario, this may bias the selection of the dictionary atoms). It was then further reduced by 9 and by 25 by changing the sampling ratio. The results of this test can be seen on Fig. 3.

The results show that the k-medians algorithm has an impressively robust behaviour, even under relatively strenuous initial database shrinkage. On the other hand, the performance of k-medoids is less robust and hard to predict. It may be that the medoids are prone to be altered upon changes in database (since medoids are database members) or in the random initialization.

4. Summary and Future directions

We presented a method to improve an already existing unsatisfactory binarization utilizing a sparse model. A database of black and white patches was created from a clean source. Existing dictionary learning methods were found to be unsuitable for our needs. Therefore, a dictionary was created via k-medians, k-medoids and extensive dictionary techniques. The results of k-medians and k-medoids were found to be sound, with fine-grained details superior to the available binarization, though not by a far margin. Further tests revealed that k-medians algorithm is more robust to initial database shrinkage than k-medoids.

This paper presents on-going research on binary sparse methods. Many future research directions are envisioned: different parameters tuning, learning a global dictionary suitable for various inscriptions; an adaptation of Maximum Likelihood dictionary learning methods; and shifting from 1-atom fit to non-linear combinations.

Acknowledgments

The research was partially 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 of Mr. 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. We would also like to acknowledge the valuable help of Dr. Leah Bar and Ms. Shira Faigenbaum. The algorithms of k-medians and k-medoids were implemented via the Pycluster toolkit.

References

Aharon, M., Elad, M. and Bruckstein, A. (2006). K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation, in *IEEE Transactions on Signal Processing*, Vol. 54, No. 11, pp. 4311-4322.

Davis, G., Mallat, S. and Avellaneda, M. (1997). Adaptive greedy approximations, *J. Construct. Approx.*, Vol. 13, pp. 57–98.

Engan, K., Rao, B. D., and Kreutz-Delgado, K. (1999). Frame design using focus with method of optimal directions (MOD), in *Norwegian Signal Process. Symp.*, pp. 65-69.

Gersho, A., and Gray, R. M. (1991). *Vector Quantization and Signal Compression*, Norwell, MA, Kluwer Academic.

Jain, A. K. and Dubes, R. C. (1981). *Algorithms for Clustering Data*, Prentice-Hall.

Kaufman, L. & Rousseeuw, P.J. (1987). Clustering by means of Medoids, in *Statistical Data Analysis Based on the L1-Norm and Related Methods*, Dodge Y. ed., North-Holland, Birkhäuser Basel, pp. 405–416.

Kreutz-Delgado, K. and Rao, B. D. (2000). FOCUSS-based dictionary learning algorithms, in *Wavelet Applications in Signal and Image Process*. VIII, pp. 4119-53.

Lesage, S., Gribonval, R., Bimbot, F. and Benaroya, L. (2005). Learning unions of orthonormal bases with thresholded singular value decomposition, in *IEEE Int. Conf. Acoust., Speech, Signal Process*, Vol. 5, pp. 293-296.

Lewicki, M. S. and Olshausen, B. A. (1999), A probabilistic framework for the adaptation and comparison of image codes, *J. Opt. Soc. Amer.* A 16, pp. 1587–1601.

Olshausen, B. A., and Field, D. J. (1996). Natural image statistics and efficient coding, *Network Comp. Neural Syst.*, Vol. 7, no. 2, pp. 333–339.

Shaus, A., Finkelstein, I., Piasetzky, E. (2010). Bypassing the eye of the beholder: Automated ostraca facsimile evaluation, *Maarav*, Vol. 17.1, pp. 7-20.

Shaus, A., Turkel, E., Piasetzky, E. (2012a). Binarization of First Temple Period inscriptions - performance of existing algorithms and a new registration based scheme. *Proc. of the 13th International Conference on Frontiers in Handwriting Recognition (ICFHR 2012)*, pp. 641-646.

Shaus, A., Turkel, E., Piasetzky, E. (2012b). Quality evaluation of facsimiles of Hebrew First Temple Period inscriptions, *Proc. of the 10th IAPR International Workshop on Document Analysis Systems (DAS 2012)*, pp 170-174.