

# Noisy Word Recognition Using Denoising and Moment Matrix Discriminants

Mila Nikolova  
 Département TSI—ENST,  
 46 rue Barrault,  
 75634 Paris Cedex 13, France,  
 nikolova@tsi.enst.fr

Alfred Hero  
 Dept. of EECS, Univ. of Michigan,  
 1301 Beal Avenue,  
 Ann Arbor, MI 48109-2122, USA,  
 hero@eecs.umich.edu

## Abstract

*We consider the problem of recognition of a printed word belonging to a limited dictionary. The main difficulty comes from the fact that this word can be printed using different fonts, sizes, and positions on the page. Invariant moment methods for word recognition developed by Hu [5] and Li [6] are unreliable when the quality of the word image is degraded by noise. In this paper we investigate the effectiveness of simple median filter denoising for preprocessing noise degraded images prior to moment based classification using the moment matrix discriminants introduced by Hero et al [4].*

## 1. Introduction

Recognition of a bit-mapped digitized printed word is important in many applications, see e.g. the citations in [4]. Recognition should rely on a method which is able to capture the essential features of a word while being invariant to deformations such as font, position and scale. In the more general context of pattern recognition, several authors have proposed efficient methods to describe feature maps by using a set of higher-order moments [6]. In the context of omnifont word recognition, a moment-based approach has been adopted by Hero *et al* [4]. Specifically, a word is characterized by a set of higher-order moments of mixed orders. This description has been shown to be able to provide an important invariance with respect to scale, font and position. However, numerical calculation of a set of higher-order moments involves linear operators which are extremely ill-conditioned. As a consequence, the presence of even

a small amount of noise in an observed image of a word can make its moment description unreliable.

When word scale and position are known noise subspace processing of the matrix of spatial moments is a very effective way to obtain more reliable moment descriptions [4] which are robust to noise degradation. This is because the effect of binary bit-flip noise is approximately additive in the moment domain. When scale and position are unknown it is more appropriate to use a matrix of scale and translation invariant moments. However, in this case noise subspace processing is less effective since the noise is no longer approximately additive. As a consequence, the capability to discriminate among words of unknown position and scale is greatly reduced. In this paper, we focus on an invariant moment matrix extension of [4] with the addition of an image domain denoising step to preprocess the images prior to moment discrimination.

## 2. Classification via Moment Representation

Assume we are given a dictionary of template words consisting of bit-mapped binary images  $\mathbf{f}^p = \{f_{m,n}^p\}$ ,  $p = 1, \dots, P$  defined over a rectangular lattice  $\{(m, n), 1 \leq m \leq M, 1 \leq n \leq N\}$ ,  $\mathbf{f}_{m,n} \in \{0, 1\}$ . Any image  $\mathbf{f}$  can be represented by its set of bivariate power moments of mixed orders  $\{\mu(k, l)\}_{k,l=0}^{\infty}$ . The bivariate power moment of order  $(k, l)$  is defined as  $\mu_{k,l} = \sum_{m=1}^M \sum_{n=1}^N m^k n^l f_{m,n} / \sum_{m=1}^M \sum_{n=1}^N f_{m,n}$ . The  $(2L + 1) \times (2L + 1)$  dimensional moment matrix  $\mathbf{M}(\mathbf{f})$  is

$$\mathbf{M}(\mathbf{f}) = \begin{bmatrix} ((\mu_{i+j,0})_{i,j=0}^{L-1}) & ((\mu_{i,j})_{i=0,j=1}^{L-1,L-1}) \\ \text{sym} & ((\mu_{0,i+j})_{i,j=1}^{L-1}) \end{bmatrix}.$$

This matrix is symmetric non-negative definite [4] for all integers  $L \geq 0$ . Let  $\mathbf{g}$  be an observed bit-mapped word digitized from a document. As explained in [4] classification of  $\mathbf{g}$  can be performed by comparing the observed moment matrix  $\mathbf{M}(\mathbf{g})$  to the template moment matrices  $\mathbf{M}(\mathbf{f}^p)$ ,  $p = \dots, P$ . These moment matrices are not invariant to scale and translation.

A related symmetric non-negative definite moment matrix  $\tilde{\mathbf{M}}(\mathbf{f})$  which is invariant can be obtained by replacing the entries  $\mu_{k,l}$  in  $\mathbf{M}$  by

$$\tilde{\mu}_{k,l} = \frac{\sum_{m=1}^M \sum_{n=1}^N \left( \frac{m-\mu_{1,0}}{\mu_{2,0}-\mu_{1,0}^2} \right)^k \left( \frac{n-\mu_{0,1}}{\mu_{0,2}-\mu_{0,1}^2} \right)^l f_{n,m}}{\sum_{m=1}^M \sum_{n=1}^N f_{n,m}}.$$

**Unwhitened moment matching** This approach is based on searching for the index  $q$  of the template word  $\mathbf{f}^q$  which minimizes the weighted distance between invariant or non-invariant moment matrices:  $q = \arg \min_{p=1,\dots,P} \|\mathbf{M}(\mathbf{f}^p) - \mathbf{M}(\mathbf{g})\|_W$ , where for any square matrix  $\mathbf{A}$ :  $\|\mathbf{A}\|_W = \sum_{i,j} ((\mathbf{W}^{\frac{1}{2}} \mathbf{A} \mathbf{W}^{\frac{T}{2}}))_{i,j}$  stands for Frobenius norm with non-negative definite weighting matrix  $\mathbf{W}$ . It is necessary to use a weighting matrix  $\mathbf{W}$  to equalize the contribution of any repeated entries in the matrix  $\mathbf{A}$ , e.g. due to symmetry.

**Whitened moment matching** To reduce sensitivity to noise it was proposed by Hero *et al* [4] to whiten the moment matrix  $\mathbf{M}(\mathbf{g})$  and to identify and eliminate the noise subspace component via eigendecomposition. What results is a rank reduced moment matrix  $\hat{\mathbf{M}}(\mathbf{g})$  which is a better approximation to the noiseless moment matrix of the observed word. The method is formulated by assuming the additive mixture model for the normalized image  $\tilde{\mathbf{g}}$ , where  $\sum_{m=1}^M \sum_{n=1}^N \tilde{g}_{m,n} = 1$ ,

$$\tilde{\mathbf{g}} = \beta \tilde{\mathbf{f}}^p + (1 - \beta) \tilde{\mathbf{f}}^o,$$

where  $\beta \in [0, 1]$  and  $\tilde{\mathbf{f}}^p$  and  $\tilde{\mathbf{f}}^o$  are normalized word-alone and noise-alone images. Under this model the non-invariant moment matrix  $\mathbf{M}(\mathbf{g})$  has the decomposition

$$\mathbf{M}(\mathbf{g}) = \beta \mathbf{M}(\mathbf{f}^p) + (1 - \beta) \mathbf{M}(\mathbf{f}^o) \quad (1)$$

Assume that the moment matrix  $\mathbf{M}(\mathbf{f}^o)$  is positive definite and known, and let  $\mathbf{C}$  be its Cholesky factor. Since  $\mathbf{M}(\mathbf{f}^o) = \mathbf{C}^T \mathbf{C}$ , the model (1) implies that  $\mathbf{M}^w(\mathbf{g}) \stackrel{\text{def}}{=} \mathbf{C}^{-T} \mathbf{M}(\mathbf{g}) \mathbf{C}^{-1}$  obeys the equivalent diagonal mixture:

$$\mathbf{M}^w(\mathbf{g}) = \beta \mathbf{M}^w(\mathbf{f}^p) + (1 - \beta) \mathbf{I} \quad (2)$$

Where  $\mathbf{M}^w(\mathbf{f}^p) = \mathbf{C}^{-T} \mathbf{M}(\mathbf{f}^p) \mathbf{C}^{-1}$  is the “whitened” word moment matrix and  $\mathbf{I}$  is the  $(2L + 1) \times (2L + 1)$  dimensional identity matrix. From the whitened mixture model (2) the word and noise subspaces can be identified from the SVD of  $\mathbf{M}^w(\mathbf{g})$ . The parameter  $\beta$  can also be identified using the fact that, by construction,  $((\mathbf{C}^T \mathbf{M}^w(\mathbf{f}) \mathbf{C}))_{1,1} = 1$  for any image  $\mathbf{f}$ . Since the sequence of singular values of typical whitened word moment matrices decay rapidly towards zero, the SVD can be used to recover a rank reduced approximation to  $\mathbf{M}^w(\mathbf{f}^p)$  from  $\mathbf{M}^w(\mathbf{g})$ ; the resultant approximation is denoted  $\hat{\mathbf{M}}(\mathbf{f}^p)$  and is called the “SVD cleaned whitened moment matrix.” Classification of an observed word  $\mathbf{g}$  is then performed by finding the word in the dictionary whose whitened moment matrix is closest to this rank reduced approximation.

In practice the noise moment matrix  $\mathbf{M}(\mathbf{f}^o)$  is not known exactly. On the other hand, one may have a good model for the mean noise matrix, e.g. when the noise is assumed to be i.i.d. uniform (salt and pepper) over the image. In this case whitening and SVD cleaning is accomplished by using the Cholesky factor of the known mean noise moment matrix in place of the Cholesky factor of the actual unknown noise moment matrix.

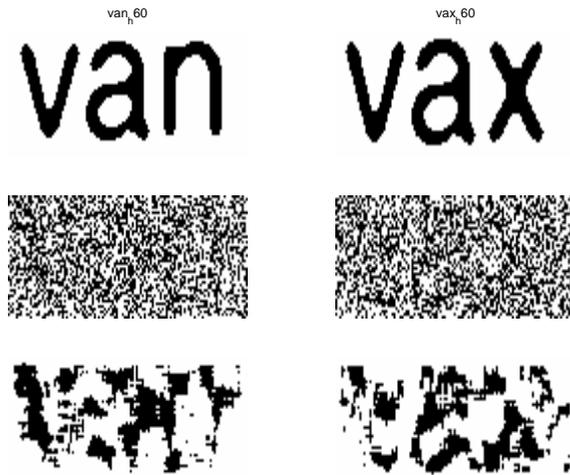
### 3. Image Preprocessing via Denoising

Optimal denoising of binary images is a difficult problem which can be addressed by means of binary Markov random fields [1]. Direct calculation of the resultant estimates requires solving a prohibitive combinatorial problem, while approximate methods are often unreliable [3]. Various suboptimal techniques for denoising have been developed which circumvent these difficulties. Two of the simplest denoising methods are based on lowpass filtering and rank order statistical filtering, e.g. the median filter, [2]. As contrasted with rank order statistical filtering, lowpass filtering has the disadvantage of producing non-binary gray scale images which introduces additional nuisance parameters (gray level) into the pattern matching problem. We investigate the use of median filtering for denoising in the next section.

### 4. Applications

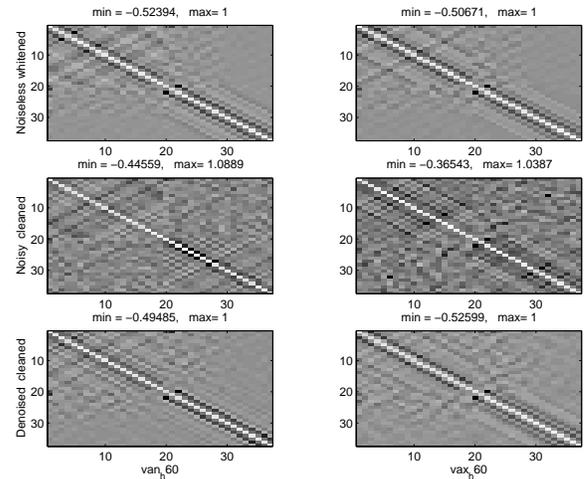
We generated two words “van” and “vax” in various postscript fonts, font sizes, and positions. Varying levels of spatially homogeneous salt and pepper noise were

added modulo-2 to the bitmaps of each word. Raw moments of various mixed orders were computed empirically and sample moments matrices were constructed using Matlab 5.0. Note that the number of pixels, or window size, for each word depends on the number of letters in the word, the presence of capitalization, punctuation, etc. To standardize the computation the bitmap coordinates for each word were scaled to a square of length 1 on a side.

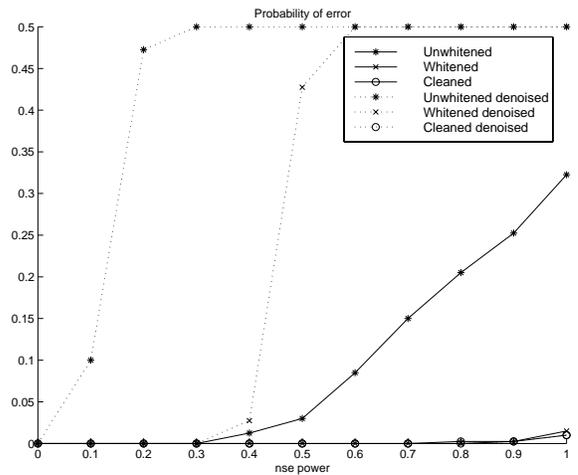


**Figure 1.** Row 1: *van* and *vax* in Helvetica 60. Row 2: row 1 corrupted with noise. Noise corresponds to maximum noise level of 1 displayed on horizontal axis of Figures 3 and 6. Row 3: row 2 denoised by median filtering with a  $7 \times 7$  window.

**Experiment 1:** We first considered denoising for the case of known font, scale and position of the two words (see Figure 1 for a representative example). The non-invariant whitening and cleaning methods of [4] were applied to the two template words and two noise corrupted words in Helvetica 64 font. Simple median filtering (with  $7 \times 7$  footprint) was applied to denoise the noise corrupted words prior to moment matrix construction. The same median filter was applied to the template words to reduce moment matching bias. The moment matrices for  $L = 18$  (each corresponding to over 150 different mixed moment discriminants) are shown in Figure 2 for the representative example of Figure 1. Note that the denoising has visually improved the match between the template moment matrices (row 1) and corresponding denoised moment matrices (row 3 appears to be a less noisy estimate of row 1 than row 2). The prob-

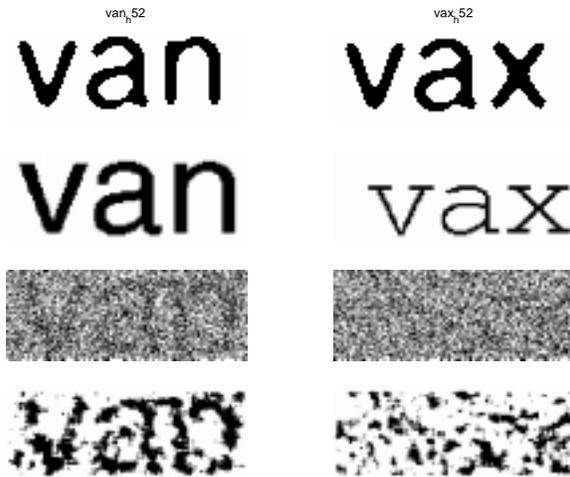


**Figure 2.** Row 1: whitened moment matrices for the template words in row 1 of Figure 1. Row 2: SVD cleaned whitened moment matrices for noisy words in row 2 of Figure 1. Row 3: SVD cleaned whitened moment matrices for denoised words in row 3 of Figure 1.



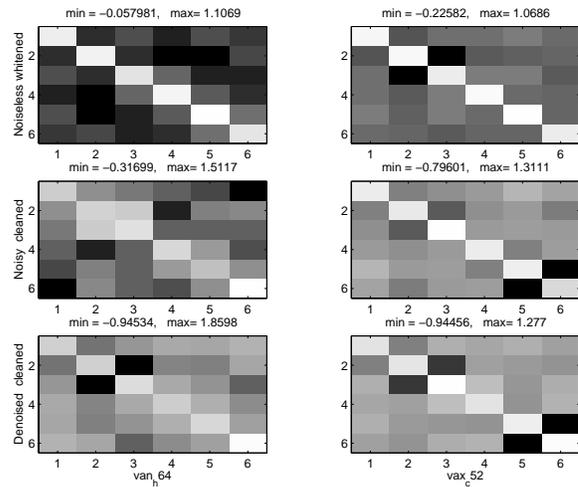
**Figure 3.** Probability of error curves for nonvariant moment methods applied to words of known font, size and position with and without median filtering as a preprocessing step (200 trials for each noise level). Here  $L = 18$  and a 2% energy threshold was used to select signal and noise subspaces in the SVD clean procedure.

ability of error shown in Figure 3 indicates the benefit of applying denoising prior to classification. Both with and without denoising the use of the whitened sample moment matrix leads to lower error than the raw moment matrix and the use of the SVD cleaning procedure is better yet. These results are consistent with those reported in [4]. Interestingly, at least over the range of noise power studied, the SVD cleaning procedure does not seem to benefit from median filter denoising. This is despite the visual improvement in the quality of the denoised moment matrices observed from Figure 2. This may be due to the fact that the residual noise in the denoised image no longer satisfies the i.i.d. uniform assumptions underlying the mixture model underlying the SVD clean procedure.



**Figure 4.** Row 1: *van* and *vax* in Helvetica 52. Row 2: *van* in Helvetica 64 and *vax* in Courier 52 with random translation. Row 3: row 2 corrupted with noise. Noise level corresponds to maximum noise level of 1 displayed on horizontal axis of Figures 3 and 3. Row 4: row 3 denoised by median filtering with a  $7 \times 7$  window.

**Experiment 2:** We next considered the case of unknown font, position and scale. Due to the mismatch between the template moment matrices and the observed moment matrices, for no values  $L$ ,  $1 \leq L \leq 12$ , were the non-invariant moment matrices used in experiment 1 successful in discriminating between the font-differentiated words (even with no noise added!). The moment matrices for  $L = 3$  are shown in Figure 5 for the representative example of Figure 4. Denoising (row



**Figure 5.** Row 1: whitened invariant moment matrices for the template words in row 1 of Figure 4 ( $L = 3$ ). Row 2: SVD cleaned whitened invariant moment matrices applied to noisy words in row 3 of Figure 4. Row 3: SVD cleaned whitened moment matrices applied to denoised words in row 4 of Figure 4.

3) has not appreciably improved the match between template and cleaned moment matrices as compared to the cleaned method applied directly to the noisy image (row 2). While matrix whitening generally appears to improve performance for high noise levels, the SVD cleaning procedure does not appear to reduce probability error. We think this is due to the fact that the invariant moment matrix no longer satisfies the additive mixture model due to the scale and position invariance transformations. In Figure 6 the total probability of error for the invariant moment matrix discriminants are shown with median filter denoising. The non-monotonicity of the curves as a function of noise power is due to the “dithering” effect: addition of noise can actually bring two words at different fonts closer together than without noise. The probability of error without median-filter denoising oscillated wildly and is not shown.

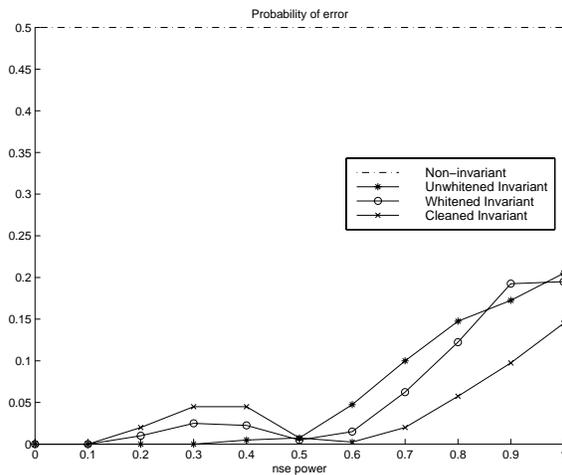
## 5. Conclusion

This paper provides motivation for applying denoising strategies prior to classification of noisy word images. Even with simple median filter denoising, significant reduction in classification performance was ob-

served for a binary classification of words at different fonts, scales, and positions. Based on this preliminary study we think that investigation of more sophisticated and accurate denoising procedures are justified. Examples of denoising methods which we will consider in the future are total variation denoising [8] and binary nonsmooth regularization [7]

## References

- [1] J. Besag, "Digital image processing: towards Bayesian image analysis," *Journal of Applied Statistics*, vol. 16, no. 3, pp. 395–407, 1989.
- [2] R. Bracewell, *Two-Dimensional Imaging*, Prentice Hall, Englewood Cliffs, 1995.
- [3] D. M. Grieg, B. T. Porteous, and A. H. Scheult, "Exact maximum a posteriori estimation for binary images," *J. Royal Statistical Society, Ser. B*, vol. 51, no. 2, pp. 271–279, 1989.
- [4] A. Hero, J. O'Neill, and W. Williams, "Moment matrices for recognition of spatial pattern in noisy images," in *IEEE Int. Conf. on Image Processing*, pp. 378–381, 1997.
- [5] M. Hu, "Pattern recognition by moment invariants," *Proc. of the Institute of Radio Engineers (IRE)*, vol. 49, pp. 1428, 1961.
- [6] Y. Li, "Reforming the theory of invariant moments for pattern recognition," *Pattern Recognition*, vol. 25, no. 7, pp. 723–730, 1992.
- [7] M. Nikolova, "Estimation of binary images by minimizing convex criteria," in *IEEE Int. Conf. on Image Processing*, Chicago, 1998.
- [8] L. Rudin, S. Osher, and C. Fatemi, "Nonlinear total variation based noise removal algorithms," *Physica D*, vol. 60, pp. 259–268, 1992.



**Figure 6.** Probability of error curves for invariant moment methods applied with median filtering (denoising) as a preprocessing step (200 trials for each noise level). Here  $L = 3$  and a 2% energy threshold was used to select signal and noise subspaces in the SVD clean procedure. Non-invariant method has error probability of 0.5 for all noise levels with or without denoising. Invariant methods without denoising were extremely sensitive to noise and are not shown.