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

April 19, 1999

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 [2] and Li [3] are unreliable when the quality of the word image is degraded by noise. In this paper we investigate the effectivenes of simple median filter denoising for preprocessing noise degraded images prior to moment based classification using the moment matrix discriminants introduced by Hero etal [1].

1. The Word Spotting Problem

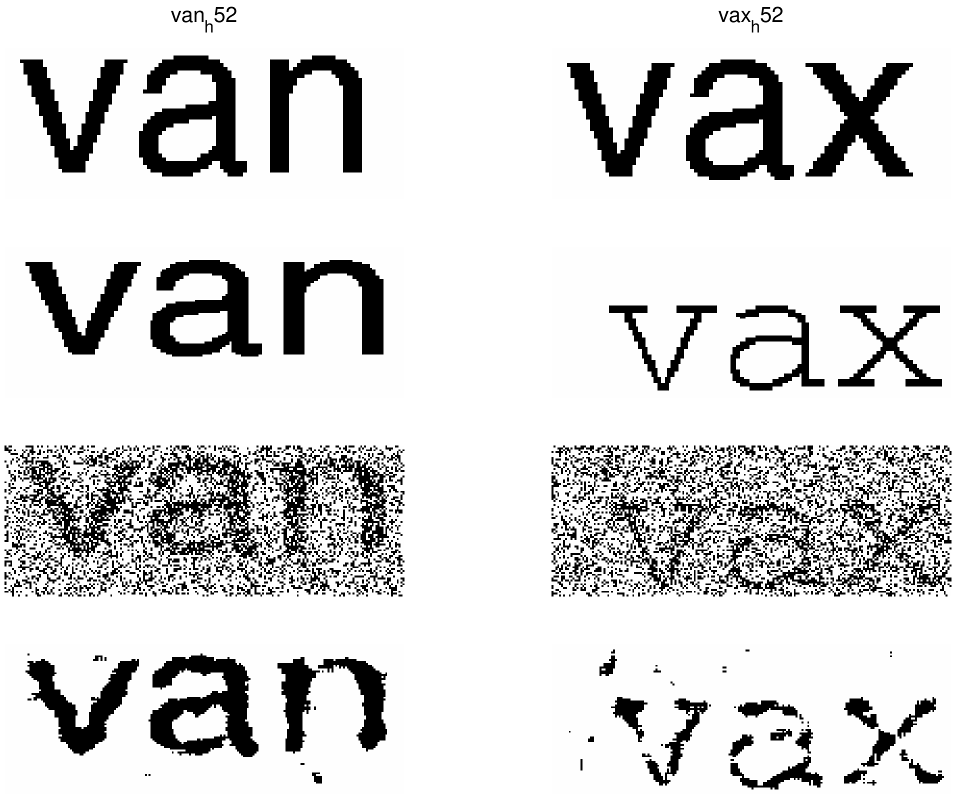

Figure 1. *The words "van" and "vax." Row 1: van and vax in Helvetica 52. Row 2: row 1 corrupted with noise. Row 3: row 2 denoised by median filtering with a* 7×7 *window.*

- $\mathbf{g} = ((g^p_{m,n}))_{m=1,n=1}^{M,N}$: observed $M \times N$ binary image
- $\{\mathbf{f}^p\}_{p=1}^P$: dictionary of binary template words

2. Moment Matrix

- $\mu(k,l)$: bivariate power moment of order (k,l)

$$\mu_{k,l} = \frac{\sum_{m=1}^{M}\sum_{n=1}^{N} m^k n^l f_{m,n}}{\sum_{m=1}^{M}\sum_{n=1}^{N} f_{m,n}}$$

- $\mathbf{M}(\mathbf{f})$: $(2L+1)\times(2L+1)$ dimensional moment matrix [1]:

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

- Properties of $\mathbf{M}(\mathbf{f})$:
  - symmetric non-negative definite
  - Hankel structured diagonal blocks give row and column marginals
  - Off-diagonal blocks give row and column mixed moments
  - $((\mathbf{M}(\mathbf{f})))_{(1,1)} = 1$

Equivalent form: let X and Y be row and column coordinates of "1",s

$$\mathbf{M}^{2L+1}_{X,Y} =$$

$$E \left[\begin{array}{cccccccc} 1 & \overline{X} & \ldots & \overline{X^L} & \overline{Y} & \ldots & \overline{Y^L} \\ \overline{X} & \overline{X^2} & \ldots & \overline{X^{L+1}} & \overline{XY} & \ldots & \overline{XY^L} \\ \vdots & \ddots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \overline{X^L} & \overline{X^{L+1}} & \ldots & \overline{X^{2L}} & \overline{X^LY} & \ldots & \overline{X^LY^L} \\ & & & & & & \\ & & & & \overline{Y^2} & \ldots & \overline{Y^{L+1}} \\ & (sym) & & & \vdots & \ddots & \vdots \\ & & & & \overline{Y^{L+1}} & \ldots & \overline{Y^{2L}} \end{array} \right].$$

3. Invariant Moment Matrix

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

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

4. Moment Matching

Unwhitened moment matching :

$\Rightarrow$ Find index q of template word $\mathbf{f}^q$ to minimize

$$\|\mathbf{M}(\mathbf{f}^q)-\mathbf{M}(\mathbf{g})\|_W = \sum_{i,j}((\mathbf{W}^{\frac{1}{2}}[\mathbf{M}(\mathbf{f}^q)-\mathbf{M}(\mathbf{g})][\mathbf{M}(\mathbf{f}^q)-\mathbf{M}(\mathbf{g})]^T\mathbf{W}^{\frac{T}{2}}))_{i,j}$$

where $\mathbf{W}$ is non-negative definite weighting matrix

Whitened moment matching :

- Define normalized image $\tilde{\mathbf{g}}$:

$$\tilde{\mathbf{g}} = \frac{\mathbf{g}}{\sum_{m=1}^{M}\sum_{n=1}^{N} g_{m,n}}$$

- Additive noise mixture model

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

- $\beta \in [0, 1]$
- $\tilde{\mathbf{f}}^p$ and $\tilde{\mathbf{f}}^o$ are normalized word-alone and noise-alone images.

$\Rightarrow$ non-invariant moment matrix $\mathbf{M}(\mathbf{g})$ has the decomposition

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

- Assume that $\mathrm{rank}\mathbf{M}(\mathbf{f}^p) \leq d$, $p = 1, \ldots, P$ (reduced rank approximation).
- Assume $\mathbf{M}(\mathbf{f}^o)$ is positive definite and known and let $\mathbf{C}$ be its Cholesky factor
- As $\mathbf{M}(\mathbf{f}^o) = \mathbf{C}^T\mathbf{C}$, we have decomposition for

$$\mathbf{M}^w(\mathbf{g}) \stackrel{\text{def}}{=} \mathbf{C}^{-T}\mathbf{M}(\mathbf{g})\mathbf{C}^{-1}$$

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

- $\mathbf{M}^w(\mathbf{f}^p) = \mathbf{C}^{-T}\mathbf{M}(\mathbf{f}^p)\mathbf{C}^{-1}$: the "whitened" word moment matrix
- $\mathbf{I}$ is $(2L+1) \times (2L+1)$ dimensional identity matrix.
- From the whitened mixture model (2) the word and noise subspaces can be identified from the SVD of $\mathbf{M}^w(\mathbf{g})$.

$$\mathbf{M}^w(\mathbf{g}) \qquad = \sum_{i=1}^{(2L+1)} \lambda_i \xi_i \xi_i^T$$

$$= \underbrace{\sum_{i=1}^{d} [\beta\lambda_i^p + (1-\beta)]\xi_i\xi_i^T}_{word\ subspace} + \underbrace{(1-\beta)\sum_{i=d+1}^{(2L+1)} [\beta\lambda_i^p + (1-\beta)]\xi_i\xi_i^T}_{noise\ subspace}$$

- The parameter β can be identified using $((\mathbf{C}^T\mathbf{M}^w(\mathbf{f}^p)\mathbf{C}))_{1,1} = 1$.

5. Applications

- Two words "van" and "vax" generated 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.

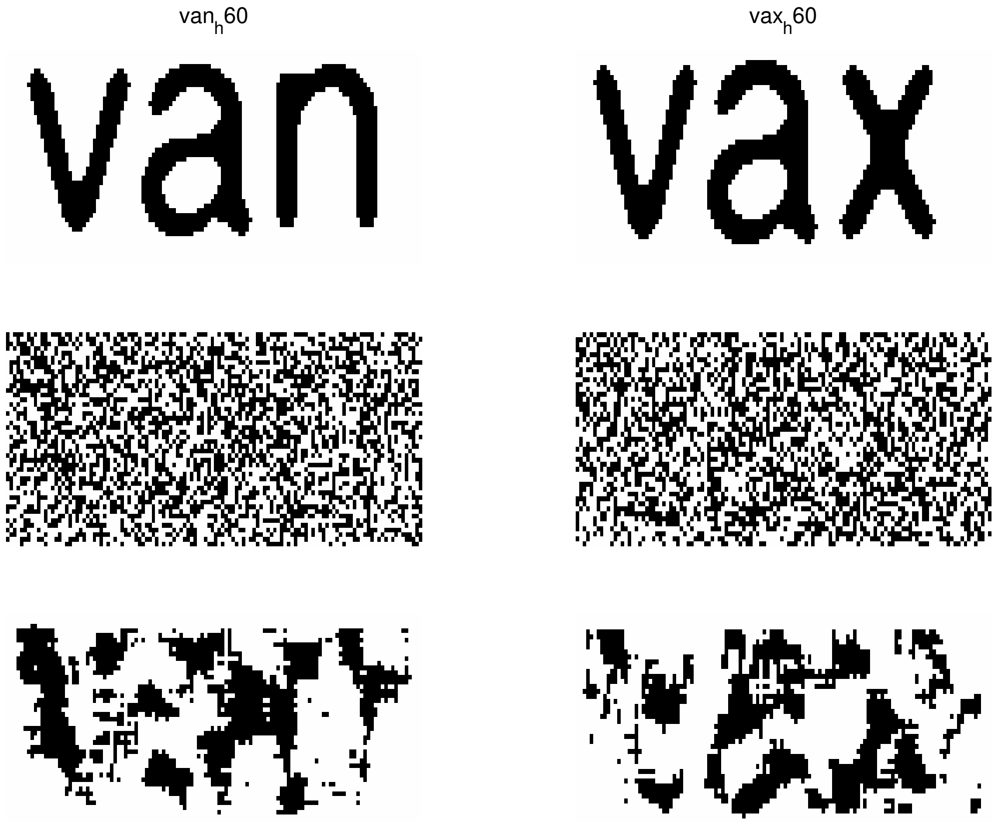

Figure 2. *Row 1: van and vax in Helvetica 60. Row 2: row 1 corrupted with noise. Noise corresponds to maximum noise level of 1 displayed on horizontal axis of Figures 4 and 7. Row 3: row 2 denoised by median filtering with a* 7×7 *window.*

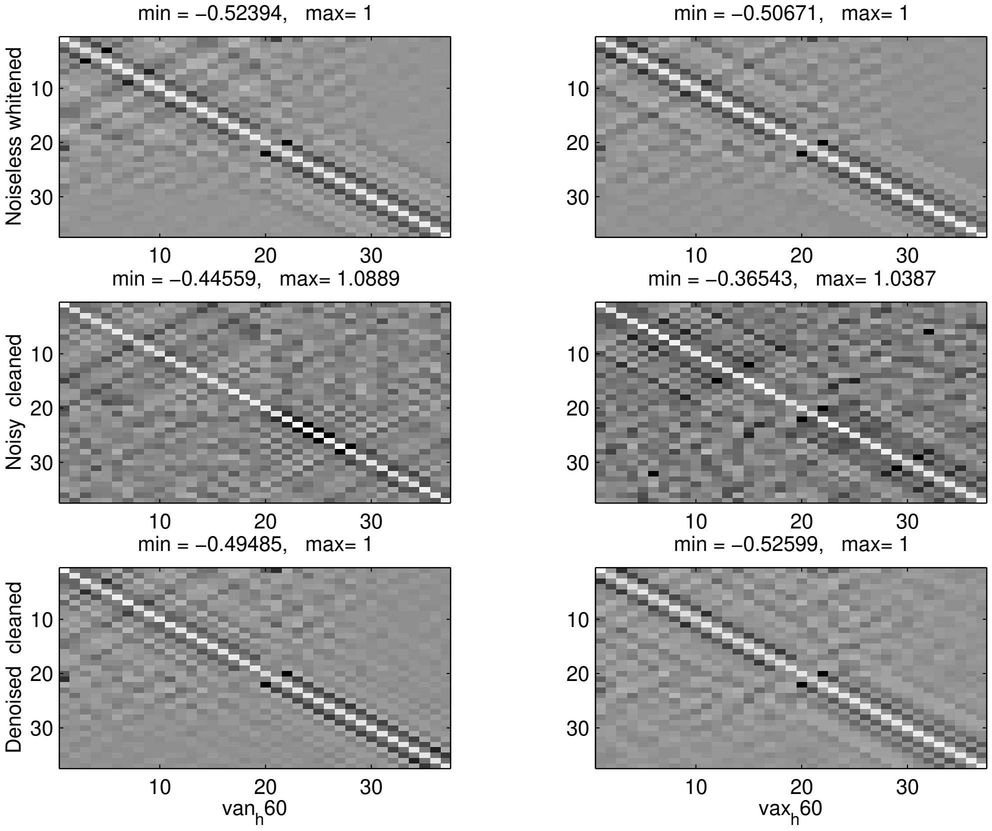

Figure 3. *Row 1: whitened moment matrices for the template words in row 1 of Figure 2. Row 2: SVD cleaned whitened moment matrices for noisy words in row 2 of Figure 2. Row 3: SVD cleaned whitened moment matrices for denoised words in row 3 of Figure 2.*

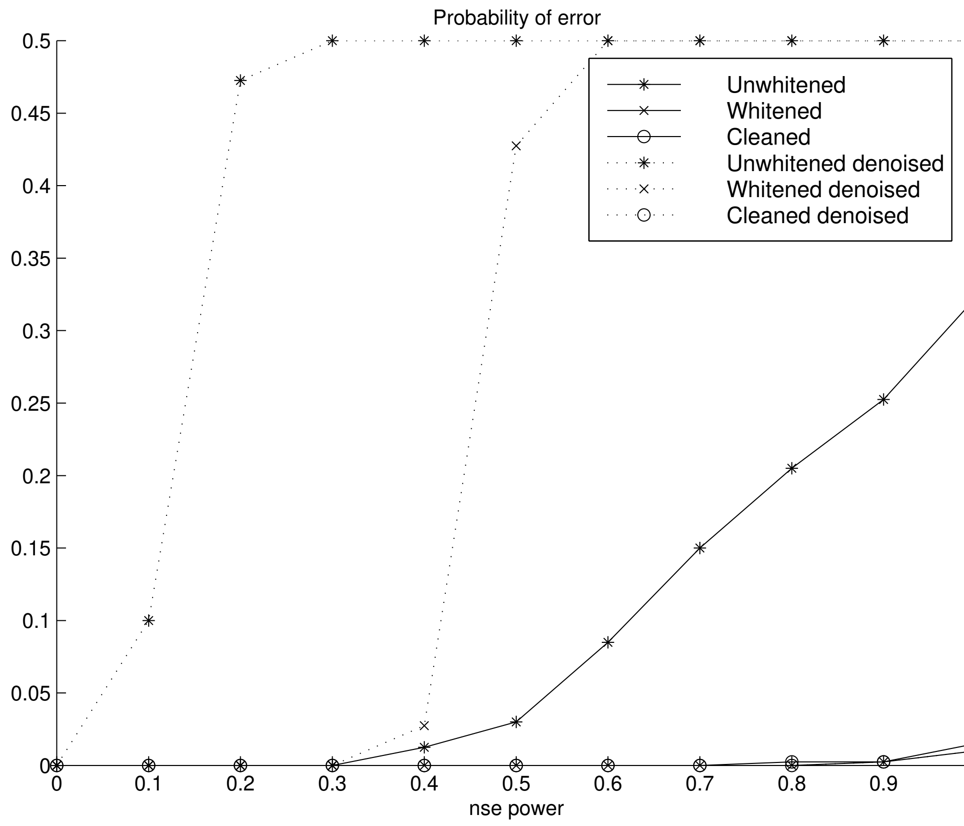

Figure 4. *Probability of error curves for noninvariant moment methods applied to words of known font, size and position with and without median filtering as a preprocessing step (200 trials for each noise level). Here $L = 18$ and a 2% energy threshold was used to select signal and noise subspaces in the SVD clean procedure.*

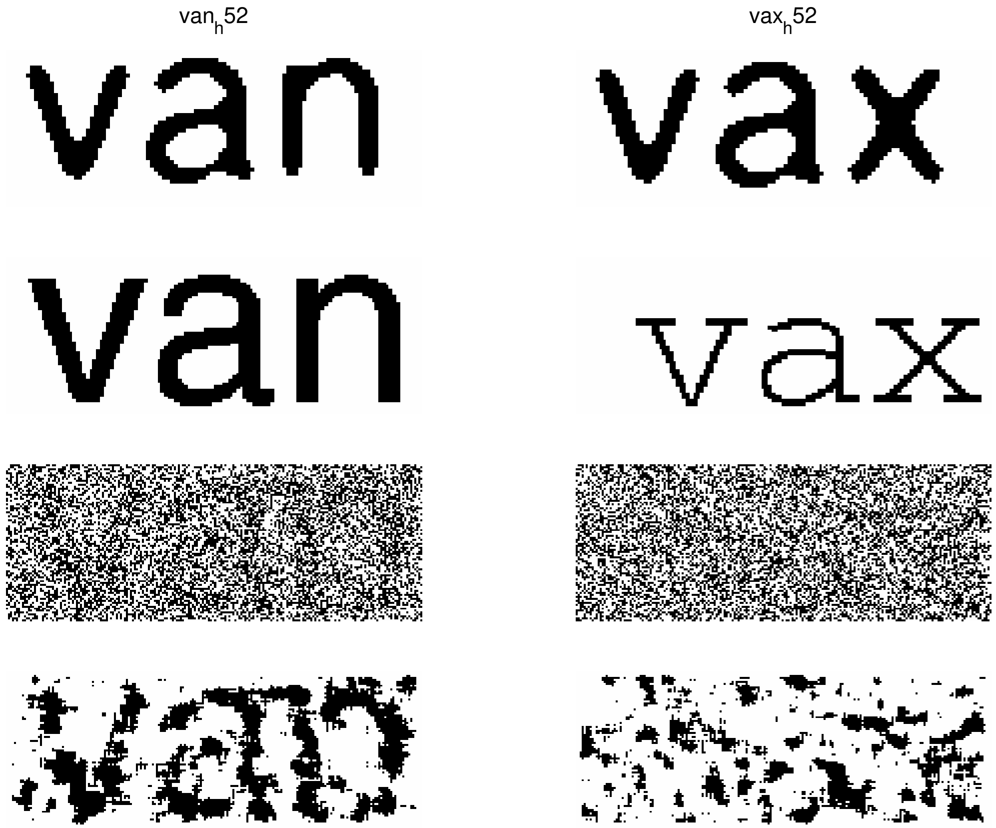

Figure 5. *Row 1: van and vax in Helvetica 52. Row 2: van in Helvetica 64 and vax in Courier 52 with random translation. Row 3: row 2 corrupted with noise. Noise level corresponds to maximum noise level of 1 displayed on horizontal axis of Figures 4 and 4. Row 4: row 3 denoised by median filtering with a 7 × 7 window.*

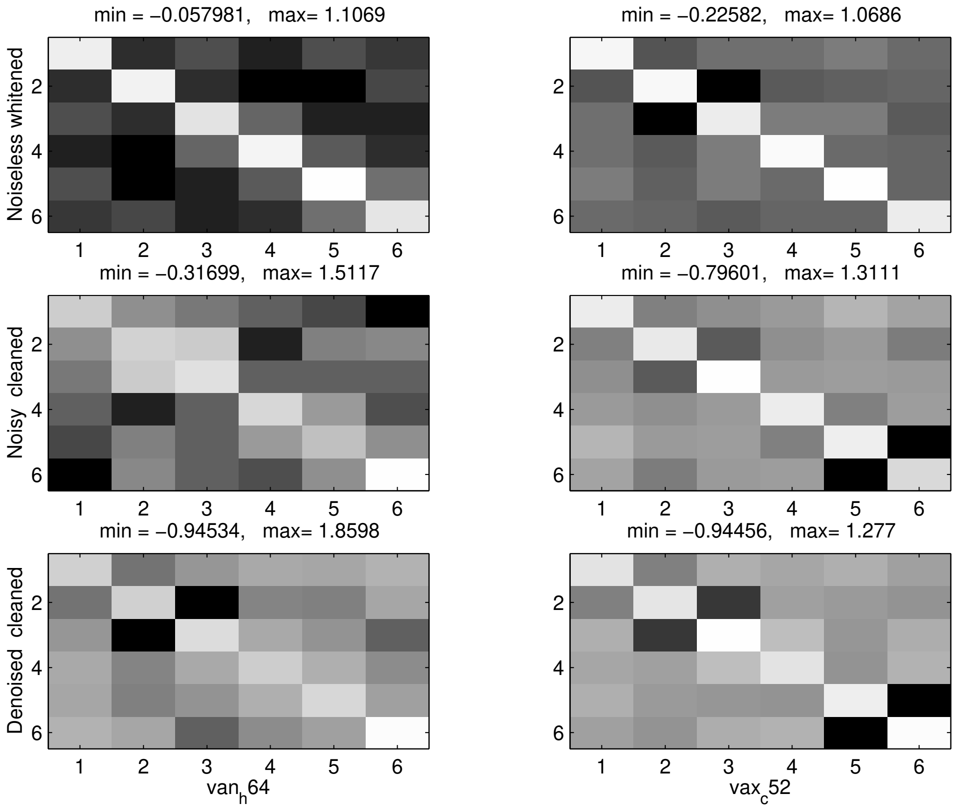

Figure 6. *Row 1: whitened invariant moment matrices for the template words in row 1 of Figure 5 ($L = 3$). Row 2: SVD cleaned whitened invariant moment matrices applied to noisy words in row 3 of Figure 5. Row 3: SVD cleaned whitened moment matrices applied to denoised words in row 4 of Figure 5.*

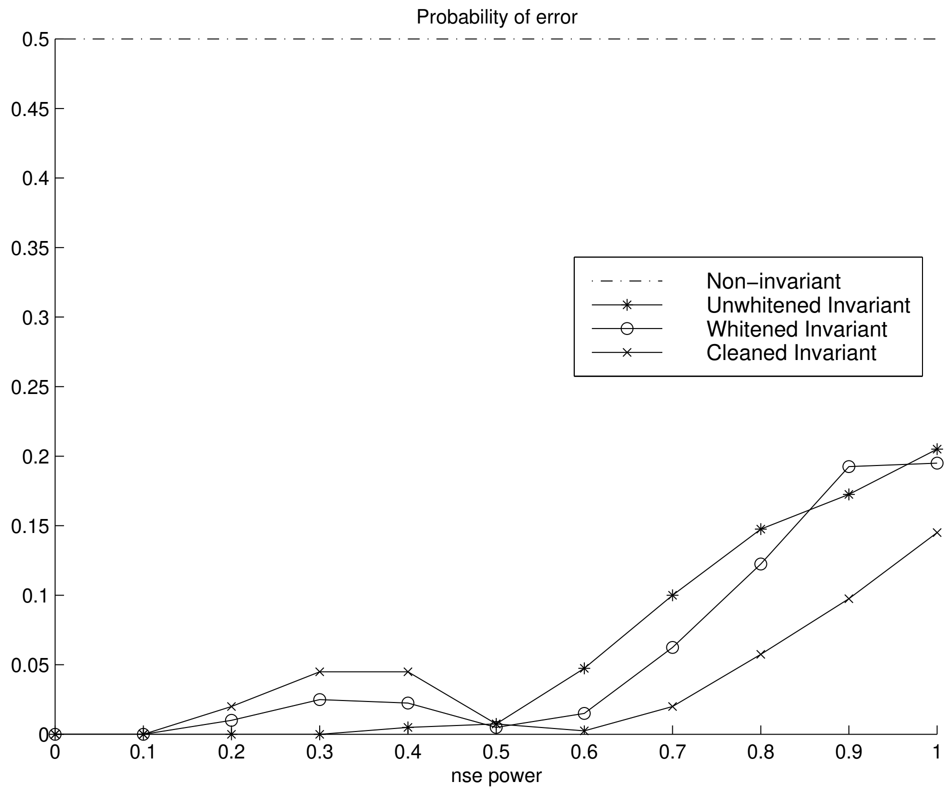

Figure 7. *P_e for invariant moment methods applied with median filtering as a preprocessing step. Here $L = 3$ and a 2% energy threshold was used. Non-invariant method has error probability of 0.5 for all noise levels with or without denoising.*

References

[1] 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.

[2] M. Hu, "Pattern recognition by moment invariants," *Proc. of the Institute of Radio Engineers (IRE)*, vol. 49, pp. 1428, 1961.

[3] Y. Li, "Reforming the theory of invariant moments for pattern recognition," *Pattern Recognition*, vol. 25, no. 7, pp. 723–730, 1992.