# An Entropy Minimization Approach for Designing W-Operators

Daniel André Vaquero[1], Junior Barrera[2] and Roberto Hirata Jr.[2]

[1] *University of California, Santa Barbara (UCSB), USA*
daniel@cs.ucsb.edu
[2] *Universidade de São Paulo (USP), Brazil*
{jb,hirata}@ime.usp.br

## 1. Introduction

The task of solving an image processing problem is often very complex. It depends on prior knowledge of the problem domain and on the knowledge, experience and intuition of an image processing specialist. Motivated by this complexity, some research groups have been working on the development of automatic techniques which try to mimic the image processing expert. Previous works in this area include [1] and [2].

The goal of these techniques is to have a system which receives a collection of image pairs that represent the desired input and output of the image processing operation, and, based on these example images, designs an image operator with behavior similar to the illustrated in the examples. The existence of such a system would also make the design process accessible to professionals who are not familiar with the underlying technical details of the image processing field.

## 2. Learning W-Operators

In this work, we constrain the search for image operators to a special class of transformations, the so called W-operators. They are translation invariant and locally defined by a window W [3].

Learning a W-operator from a collection of input-output examples consists of three sequential steps:

- **Collection of examples**: a table that relates the input patterns to the number of times each output value appeared in the examples associated with each specific input pattern is computed. This is done by sliding the window through the example images;
- **Conditional probability distribution estimation**: once the examples are collected, the conditional probability distribution of the output values (denoted by $Y$) given the input patterns (denoted by $\mathbf{x}$) is estimated by dividing, for each input pattern, the number of times that it appears associated with a specific output value by the total number of times that it appears in the examples;
- **Optimal decision**: the operator is optimally designed with the objective of minimizing the error according to the estimated probability distribution. Some error measures typically used in this step are the Mean Absolute Error (MAE) and the Mean Squared Error (MSE). Then, the designed operator can be further applied to other images.

## 3. The Multiresolution Approach

When creating a system to design image operators from input-output pairs, a key component is the learning algorithm employed to estimate the image transformation from the examples. Often the number of possible patterns which can be observed by W is very large, and there are not enough examples among the input-output image pairs to achieve a good estimation of the operator. It is interesting to note that the W-operator estimation problem is a particular case of the supervised learning [4] problem.

The multiresolution approach [5] is based on the idea of observing patterns at a lower resolution when there are not enough examples at a higher resolution to provide a good estimate of the function's output value. The resolution reduction process is determined by a **window pyramid** that specifies the resolution mappings applied. In Figure 1 an example of a window pyramid with three levels is shown. The resolution mappings are usually subsampling operators, but mathematically any mapping between pattern spaces could be used.

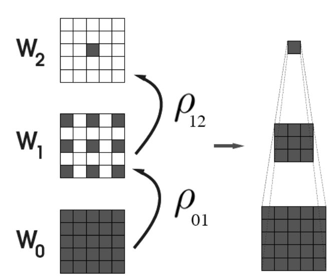


*Figure 1.* A window pyramid. The dark points represent the pixels selected by the resolution reduction operations $\rho_{01}$ and $\rho_{12}$. By eliminating the white points, the structure resembles a pyramid, as can be seen in the right hand side.

## 4. Joint Distribution Estimation

The multiresolution strategy estimates the conditional probability distribution $p(Y \mid \mathbf{X})$ using the resolution mappings given by a *window pyramid* to guide the estimation for non-observed input patterns. We developed a novel algorithm that, given a collection of input-output examples and a window pyramid, estimates the conditional distribution $p(Y \mid \mathbf{X})$ but also estimates $p(\mathbf{X})$, the probability of occurence of each input pattern. This way, we have a representation of the joint distribution $p(\mathbf{X}, Y)$ by means of $p(\mathbf{X})$ and $p(Y \mid \mathbf{X})$, and the following interpretation holds: a window pyramid can be seen as a model which induces a joint probability distribution given a collection of input-output examples.

## 5. Pyramid Choice

In the multiresolution approach, the quality of the designed operator directly depends on the chosen window pyramid. Hence, an interesting question is: how to choose the resolution mappings? The number of possibilities is huge, and the choice also depends on the specific image processing problem at hand.

To approach this problem, we worked under the assumption that for each possible pattern $\mathbf{x}$ observed by W there is one output value $y$ such that $P(Y = y \mid \mathbf{x})$ is much greater than $P(Y = k \mid \mathbf{x})$, $k \neq y$. This is reasonable for problems that have a good solution, as it means that only one output value is associated with each input pattern with high probability. The assumption is equivalent to saying that the conditional entropy of the distribution $p(Y|\mathbf{X})$ has low conditional entropy, a measure from Information Theory [6].

We then formulate the problem of choosing a pyramid as the problem of finding the pyramid that induces conditional probability distributions with lowest entropy. We use the algorithm mentioned in Section 4 to estimate the joint distribution, and the conditional entropy $H(Y \mid \mathbf{X})$ is then computed from $p(\mathbf{X})$ and $p(Y \mid \mathbf{X})$. A technique to aid in the choice of the pyramid would then choose the pyramid that induces the joint distribution with lowest conditional entropy.

## 6. Experimental Results

We used the MNIST database [7] to train and test classifiers of handwritten digits based on W-operators designed using our technique. The designed operators are mappings from $\{0, 1\}^W \rightarrow \{0, \ldots, 9\}$, since the input images are binary. Given a W-operator and a binary image containing a single digit, the operator is applied to the image and then a voting scheme classifies the digit as the pixel value that mostly appears in the output image.

We used our technique for a set of fourteen window pyramids based on subsampling resolution mappings, with various base window shapes and sizes. Comparing the ordering of the pyramids with respect to entropy values and the ordering regarding the error obtained over the test set, the pyramid of minimum error matched the pyramid of minimum entropy. The accuracy of the best classifier was 96.83%, obtained with a pyramid built from a variant of the quincunx sampling scheme [8].

The result is comparable to the accuracies obtained by some well-known pattern recognition algorithms, as listed in [7]. This suggests that our method may be worth considering as a pattern recognition tool when solving classification problems. Another experiment also shows that the accuracy of the W-operator increases with the number of training examples, similarly to good learning algorithms.

## Acknowledgements

Daniel André Vaquero is grateful to FAPESP (grant number 03/10066-7) for the financial support during his Master's studies at USP.

## References

[1] J. Barrera, E. R. Dougherty, and N. S. Tomita, *Automatic Programming of Binary Morphological Machines by Design of Statistically Optimal Operators in the Context of Computational Learning Theory*, Electronic Imaging **6** (1997), no. 1, 54-67.

[2] R. Hirata Jr., E. R. Dougherty, and J. Barrera, *Aperture Filters*, Signal Processing **80** (2000), no. 4, 697-721.

[3] J. Barrera and G. P. Salas, *Set Operations on Closed Intervals and Their Applications to the Automatic Programming of Morphological Machines*, Electronic Imaging **5** (1996), no. 3, 335-352.

[4] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*, John Wiley and Sons, 2001.

[5] E. R. Dougherty, J. Barrera, G. Mozelle, S. Kim, and M. Brun, *Multiresolution Analysis for Optimal Binary Filters*, Journal of Mathematical Imaging and Vision (2001), no. 14, 53-72.

[6] C. E. Shannon, *A Mathematical Theory of Communication*, Bell System Technical Journal **27** (1948), 379-423, 623-656.

[7] Y. LeCun and C. Cortes, *The MNIST Database of Handwritten Digits.* URL: <http://yann.lecun.com/exdb/mnist/>.

[8] H. J. A. M. Heijmans and J. Goutsias, *Morphological Pyramids and Wavelets Based on the Quincunx Lattice*, ISMM 2000, International Symposium on Mathematical Morphology and its Applications to Image and Signal Processing V (2000).