W-operator Window Design for Color Texture Classification

David C. Martins-Jr.[1]*, Roberto M. Cesar-Jr.[1], Junior Barrera[1], Mario A. S. Liziér[2], Luiz G. Nonato[2]

[1]*Instituto de Matemática e Estatística–Universidade de São Paulo*
[2]*Instituto de Ciências Matemáticas e de Computação–Universidade de São Paulo*
{davidjr,cesar,jb}@ime.usp.br, {lizier,gnonato}@icmc.usp.br

Abstract

This work generalizes the technique described in [1] to image processing applications based on color. This method chooses a subset of variables W (i.e. pixels seen through a window) that maximizes the information observed in a set of training data by mean conditional entropy minimization. The task is formalized as a combinatorial optimization problem, where the search space is the powerset of the candidate variables and the measure to be minimized is the mean entropy of the estimated conditional probabilities. As a full exploration of the search space requires an enormous computational effort, an algorithm of the feature selection literature is applied. The introduced approach is well fundamented mathematically and experimental results with color texture recognition applications show that it is also adequate to treat problems with color images. Comparative performance assessment of this technique including an artificial neural network alternative (Multi-Layer Perceptron) approach is presented.

1. Introduction

The paper [1] discusses a technique based on information theory concepts to estimate a good W-operator to perform binary image transformations (e.g. noisy image filtering) and gray scale analysis (e.g. gray-level texture recognition). A W-operator is an image transformation that is locally defined inside a window W and translation invariant [2]. This means that it depends just on shapes of the input image seen through the window W and that the applied transformation rule is the same for all image pixels.

The main contribution of this research is the proposal of an extension of the W-operator concept to be applied to color texture classification. For this, we will consider it as a function whose domain is a set of vectors of 3 integer numbers, one for each channel (red, green, blue), and the output is an integer number (one of the considered classes). The fact that each pixel in color images contains a triple of integer values worsens even more the problem of lack of training data. Because of this, an approach for dealing with the lack of training data becomes even more required. Therefore, as in gray level applications, quantization is usually necessary for RGB applications. Besides, we propose a strategy to increase the number of training samples, generating new samples from rotations of the observed samples.

As the number of possible sub-windows is exponential in terms of the cardinality of W, we adopted some algorithms to explore this space in reasonable computational time. The adopted algorithm is the Sequential Floating Forward Search (SFFS) feature selection algorithm [3]. In this text, we briefly discuss how to apply this algorithm guided by the criterion function based on condition entropy to estimate a sub-window W^* that gives one of the best operators to perform classification over images with arbitrary number of channel levels and arbitrary number of classes.

* Financially supported by FAPESP and CNPq

2. W-operator design by feature selection with mean conditional entropy

Let X, Y be two random variables and P be its probability distribution. The conditional entropy of Y given X is defined as:

$$H(Y|X) = -\sum_{x\in X}\sum_{y\in Y} P(y|x) log P(y|x), \qquad (1)$$

with $0log0 = 0$. Similar definitions hold for random vectors $\mathbf{X}$.

The important property of entropy explored in this work is that when the probability mass of a distribution becomes more concentrated somewhere in its domain, the entropy decreases. This means that when a given feature vector defined in a window has a majoritary label $Y = y$ (i.e. it is classified almost always in a same class), its entropy of the conditional distribution should be low. Thus, the optimization algorithm consists in estimating the mean conditional entropy for the joint distribution estimated for each sub-window and choosing the one that minimizes this measure. The estimator of the mean conditional entropy idealized in this work is given by the following equation:

$$\hat{E}[H(Y|\mathbf{X}_{\mathcal{Z}})] = \sum_{\mathbf{x}_{\mathcal{Z}}:P(\mathbf{x}_{\mathcal{Z}})>\frac{\alpha}{t}} P(\mathbf{x}_{\mathcal{Z}}) \cdot \hat{H}(Y|\mathbf{x}_{\mathcal{Z}}) + \\ + \sum_{\mathbf{x}_{\mathcal{Z}}:P(\mathbf{x}_{\mathcal{Z}})\leq\frac{\alpha}{t}} P(\mathbf{x}_{\mathcal{Z}}) \cdot H(U(Y)),$$

where $\mathcal{Z}$ is a subset of the indexes $\mathcal{I} = \{1, 2, ..., |W|\}$, $\mathbf{x}_{\mathcal{Z}}$ is one of the possible instances of $\mathbf{X}_{\mathcal{Z}}$, U is the uniform probability distribution function, t is the number of training samples and α is a parameter that indicates if an instance is poorly observed or not. Instances that have probability less or equal than α/t are considered poorly observed and the entropies are set to maximum (entropy of uniform distribution). It is important to set $\alpha > 0$ in order to penalize the poorly observed instances and to avoid the curse of dimensionality where the error estimation becomes too high for larger feature space dimensions. It was adopted $\alpha = 1$ in the experiments shown here.

Thus, feature selection may be defined as an optimization problem where we search for $\mathcal{Z}^* \subseteq \mathcal{I}$ such that:

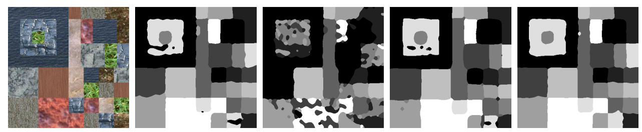
$$\mathcal{Z}^* : H(Y|\mathbf{X}_{\mathcal{Z}^*}) = min_{\mathcal{Z}\subseteq\mathcal{I}}\{\hat{E}[H(Y|\mathbf{X}_{\mathcal{Z}})]\}$$

The W-operator is represented by a table of conditional probabilities where each row is a possible instance $\mathbf{x}_{\mathcal{Z}^*}$ of $\mathbf{X}_{\mathcal{Z}^*}$, each column is a possible class $Y = y$ and each cell of this table is $P(Y|\mathbf{X}_{\mathcal{Z}^*} = \mathbf{x}_{\mathcal{Z}^*})$. This table is used as a Bayesian classifier where for each given instance, the chosen label $Y = y$ is the one with maximum conditional probability for the considered instance. In case of instances that have two or more labels of maximum probability, one of these labels is chosen according to their number of occurrences around the neighborhood of the pixel that is to be classified (the most frequent label is assigned to the referred pixel).

3. Experimental Results

Two experiments were performed, one involving generic textures and another involving ring classification of a tree trunk slice. For both experiments, we compared the method proposed here (RGB W-operator with window 5×5) with other three methods. The first uses only pixel to pixel color information (window 1×1), the second is the gray-level W-operator as described in [1] with window 5×5, and the last is a method based on Multi-Layer Perceptron (MLP) neural network. The quantization degree for each RGB channel was $k = 16$ for all compared methods. Besides, sucessive applications of the mode filter (a window-based classifier that translates a window over all pixels of the produced labeled image and assigns the most frequent label to its central pixel) with decreasing window dimensions have been applied to the results in order to clean small noises on the labeled images.

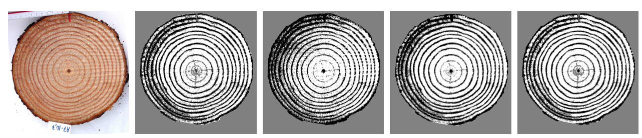
We have analyzed the MAE (Mean Absolute Error) obtained by application of the techniques mentioned above for the two experiments. For synthetic textures, nine textures were used for training and testing sets (sets obtained separately from the same source). A mosaic of the testing sets as illustrated in Figure 1-a was adopted as test image for results generation. Figures 1-b, 1-c, 1-d, and 1-e show the results by applying respectively, only color information pixel to pixel (MAE = 3.92%), gray-scale W-operator (MAE = 36.27%), MLP approach (MAE = 4.96%) and RGB W-operator (MAE = 2.32%).



(a) (b) (c) (d) (e)

Figure 1. Results for generic textures

For tree rings classification, we adopted the image of Figure 2-a as training (2.81% of the image) and testing image. Figures 2-b, 2-c, 2-d and 2-e show the results by applying respectively, only color information pixel to pixel (MAE = 7.18%), gray-scale W-operator (MAE = 7.63%), MLP approach (MAE = 9.25%) and RGB W-operator (MAE = 5.21%).



(a) (b) (c) (d) (e)

Figure 2. Results for tree rings classification

4. Conclusion

This work presented an extension of W-operator design by feature selection technique to color image classification. The new method has been very effective in both applications focused, presenting the best results when compared with other approaches, as MLP neural network.

5. References

[1] D. C. Martins-Jr, R. M. Cesar-Jr, and J. Barrera. W-operator window design by minimization of mean conditional entropy. *Pattern Analysis & Applications*, 2006.

[2] J. Barrera, R. Terada, R. Hirata-Jr., and N. S. T. Hirata. Automatic programming of morphological machines by pac learning. *Fundamenta Informaticae*, pages 229–258, 2000.

[3] P. Pudil, J. Novovicová, and J. Kittler. Floating search methods in feature selection. *Pattern Recognition Letters*, 15:1119–1125, 1994.