On combining Learning Vector Quantization and the Bayesian classifiers for natural textured images

María Guijarro
Dept. Ingeniería del Software e Inteligencia Artificial
Facultad Informática
Universidad Complutense
28040 Madrid
mariaguijarro@gmail.com

Raquel Abreu
Dpto. Informática y Automática
ETS Ingeniería Informática
Univ. Nacional Educación a Distancia
28040 Madrid
raquel.abreuhernando@telefonica.es

Gonzalo Pajares
Dept. Ingeniería del Software e Inteligencia Artificial
Facultad Informática
Universidad Complutense
28040 Madrid
pajares@dacya.ucm.es

Abstract

One objective for classifying textures in natural images is to achieve the best performance possible. Unsupervised techniques are suitable when no prior knowledge about the image content is available. The main drawback of unsupervised approaches is its worst performance as compared against supervised ones. We propose a new unsupervised hybrid approach based on two well-tested classifiers: Vector Quantization (VQ) and Bayesian (BY). The VQ unsupervised method establishes an initial partition which is validated and improved through the supervised BY. A comparative analysis is carried out against classical classifiers, verifying its performance.

1. Introduction

Nowadays the increasing technology of aerial images is demanding solutions for different image-based applications. The natural texture classification is one of such applications due to the high spatial resolutions achieved in the images. The areas where textures are suitable include agricultural crop ordination, forest or urban identifications and damages evaluation in catastrophes or dynamic path planning during rescue missions or intervention services also in catastrophes (fires, floods).

Different classical techniques have been studied for image texture classification, namely: Bayesian, K-Nearest, Neural Networks, Vector Quantization [1, 5, 6, 7, 10]. These classifiers are supervised methods, which perform appropriately if a correct partition is used for training them. This implies that the samples have been correctly assigned as belonging to each cluster. Unfortunately, the variety of textures in aerial images could become high or even if unpredictable. Hence, unsupervised automatic classification approaches should be suitable in order to establish the best cluster partition. The Vector Quantization approach (VQ) is one of the possible unsupervised approaches, the main drawback of VQ is its strong dependency from the threshold used for the partition and the order in which the samples are processed.

We propose the use of the unsupervised VQ classifier supervised by the Bayesian (BY) classifier. Hence, the cluster partition supplied by the VQ is verified by the BY. This is a hybrid approach which making the main finding of this work.

There are pixel-based and region-based approaches. A pixel-based approach tries to classify each pixel as belonging to one of the classes. The region-based identifies patterns of textures within the image and describes each pattern by applying filtering (laws masks, Gabor filters, Wavelets, etc.), it is assumed that each texture displays different levels of energy allowing its identification at different scales [12], [11]. The behaviour of features has been also studied in texture classifications, where the set of features describes each pattern [3, 12]. This is out of the scope of this paper.

The aerial images used in our experiments do not display texture patterns. This implies that textured regions cannot be identified by applying region-based. In this work we use a pixel-based approach under red-green-blue (RGB) colour representation because it performs favourably against other colour mappings as reported in [11]. Hence, the three RGB spectral values are the features used in our method.

The paper is organized as follows. Section 2 describes the design of the automatic hybrid classifier, where the VQ and BY methods are briefly described. Section 3 shows experimental and comparative results. Finally, in the section 4 the conclusions are presented.

2. Automatic Unsupervised Classifier

Our systems works in two stages: training and classification. The training phase tries to obtain the best partition for the available training patterns. The classification phase uses the results obtained during the training and classifies the new incoming patterns.

Figure 1 displays the training system architecture with two main modules (VQ and BY).

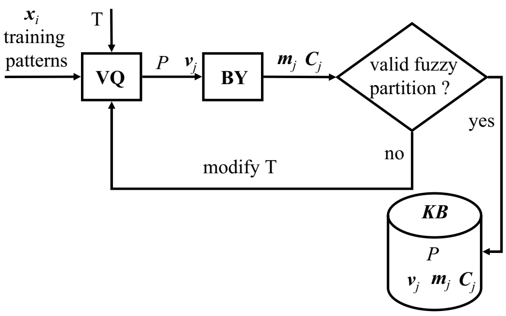

Figure 1. Combination of Classifiers

The working process is described below,

1) The VQ module receives the training patterns $\boldsymbol{x}_i$ and a threshold T and provides a partition (clusters) P. This means that the BY module receives the number of clusters (c), their centres ($\boldsymbol{v}_j$) and the training samples $\boldsymbol{x}_i^j$ belonging to each cluster w_j.

2) The BY module estimates the cluster centres ($\boldsymbol{m}_j$) and covariance matrices C_j. This allows us to verify the partition's validation.

3) If P is a valid partition it is stored in the knowledge base (*KB*) so that the cluster centres are available during the classification phase after the training one.

4) If P is not validated, the threshold T is modified and a new partition, with the initial set of training samples available, is intended until the validation is achieved. As before the new partition is generated by the VQ module.

5) During the classification phase the system recovers, from the *KB*, the cluster centres $\boldsymbol{v}_j$, $\boldsymbol{m}_j$ and C_j obtained during the training phase. Each new pattern is classified as belonging to the available clusters by applying the Bayes approach. Hence, this pattern will belong to each cluster with a different probability value. The final decision is made based on the maximum probability value.

2.1. Vector Quantization

The VQ approach [2,4] starts with the observation of a set X of n training patterns, i.e $X = \{\boldsymbol{x}_1, \boldsymbol{x}_2, ..., \boldsymbol{x}_n\} \in \Re^3$. The samples are three-dimensional because the features are the three spectral RGB values. The VQ process is as follows:

1) Define the threshold T value and select a metric distance (euclidean).

2) For each training pattern $\boldsymbol{x}_i$ compute its distance to each cluster centre $\boldsymbol{v}_j$, $d_{ij}(\boldsymbol{x}_i, \boldsymbol{v}_j)$. If $\boldsymbol{x}_i$ is the first pattern it is the first centre, i.e. $\boldsymbol{v}_1 = \boldsymbol{x}_i$.

3) Compute $d_{ik}(\boldsymbol{x}_i, \boldsymbol{v}_k) < d_{ij}(\boldsymbol{x}_i, \boldsymbol{v}_j)\ \ \forall\, j \neq k$. If $d_{ik}(\boldsymbol{x}_i,\boldsymbol{v}_k) < T$ then assign $\boldsymbol{x}_i$ to the cluster associated to the centre $\boldsymbol{v}_k$ which gives the minimum distance and update this centre $\boldsymbol{v}_k$ by averaging the samples belonging to this cluster.

4) If $d_{ik}(\boldsymbol{x}_i, \boldsymbol{v}_k) \geq T$ then a new cluster is created and its new cluster centre is exactly the sample, i.e. $\boldsymbol{v}_k = \boldsymbol{x}_i$.

After this process a partition P is obtained with c clusters, each one with its cluster centre $\boldsymbol{v}_j$ where $j = 1,2,\ldots,c$ and a subset $X^j \subset X$ of n_j pattern samples belonging to each cluster w_j.

2.2. Bayesian approach

The BY module receives the partition, i.e. the samples distributed into the clusters.

Given a pattern vector $\boldsymbol{x}$, under the Bayesian framework, the main problems to be solved are the estimations of a set of class-conditional probability density functions $p(\boldsymbol{x} \mid w_j)$ and the *a priori* probabilities $P(w_j)$ for each class w_j [4]. The probability density function $p(\boldsymbol{x})$ can be modelled as a mixture density distribution consisting of c density components associated to the c clusters,

$$p(\boldsymbol{x}) = \sum_{j=1}^{c} p\left(\boldsymbol{x} \mid w_j\right) P(w_j) \tag{1}$$

The method is summarized as follows,

1) For each class w_j compute its associated mean vector $\boldsymbol{m}_j$

2) With the n_j samples $\boldsymbol{x}_i^j$ belonging to the class w_j compute de covariance matrix as follows,

$$C_j = \frac{1}{n_j - 1} \sum_{j=1}^{n_j} \left(\boldsymbol{x}_i^j - \boldsymbol{m}_j\right)\left(\boldsymbol{x}_i^j - \boldsymbol{m}_j\right)^t \tag{2}$$

where t denotes transpose.

3) Once the parameters $\boldsymbol{m}_j$ and C_j are estimated, the class conditional probability density function is given by,

$$p\left(\boldsymbol{x} \mid w_j\right) = \frac{1}{\left(2\pi\right)^{\frac{n}{2}} \left|C_j\right|^{\frac{1}{2}}} exp\left\{-\frac{1}{2}\left(\boldsymbol{x} - \boldsymbol{m}_j\right)^t C_j \left(\boldsymbol{x} - \boldsymbol{m}_j\right)\right\} \tag{3}$$

2.3. Partition validation

The next step consists in the partition validation. This is carried out by computing the divergence between two clusters w_i and w_j through the equation (4) according to Jensen [8].

$$\begin{aligned} Diverg_{ij} &= \frac{1}{2} Tr\left[\left(C_i - C_j\right)\left(C_j^{-1} - C_i^{-1}\right)\right] \\ &+ \frac{1}{2} Tr\left[\left(\boldsymbol{m}_i - \boldsymbol{m}_j\right)^t \left(C_i^{-1} + C_j^{-1}\right)\left(\boldsymbol{m}_i - \boldsymbol{m}_j\right)\right] \end{aligned} \tag{4}$$

where Tr is the trace for the matrix; $\boldsymbol{m}_i$, C_i and $\boldsymbol{m}_j$ C_j are the mean and covariance matrices for clusters w_i and w_j respectively; t denotes transpose.

The divergence is a measure of separability between two clusters, so that the higher the divergence, the higher the cluster separation.

In order to determine the validity of the partition, we compute the divergence between each two clusters and then average the values of the divergences.

The maximum value of the averaged divergence for different values of c determines the best partition, i.e. the best number of clusters for the set of training samples available. Values of the divergence greater than 80 are acceptable.

Following the scheme in the figure 1 and based on the averaged divergence value, if the partition is rejected then the threshold T must be modified in order to try a new better partition P; otherwise if the current partition is accepted, it is stored in the KB.

Different partitions are intended by supplying the samples in different orders of processing. The order is randomly established.

In our experiments, we have verified that the divergence values range between 0 and 130. Hence we map linearly these divergence values so that they range in [0,1] as $D_{ij} = D_{ij} / 130$.

Moreover, we are using RGB values as features ranging in [0,255]. This implies that the clusters (samples) are mapped into the 3-dimensinal space, where each axis varies between 0 and 255. Therefore, the maximum Euclidean distance d_E in this space is given by the two opposite points (0,0,0) and (255,255,255) resulting the following value $d_E \approx 442$. Hence, this will be the maximum threshold value T. Its minimum value is bounded by cero. We define the following quantum magnitude $Q = d_E / 40$. The modification process for T is expressed by (6). We have specified the iteration as k

$$T(k) = T(k-1) + Q\left(1 - D\right) \tag{5}$$

where D is the averaged divergence value ranging in [0,+1]. Initially we set $T(0) = 0$ and $D = 0$, i.e. $T(1) = Q$.

2.4. Decision process

During this on-line process new images and consequently new pattern samples are to be processed by the system. A decision must be made about them. With such purpose, we recover the class conditional probability density functions stored in KB and estimated through the equation (3) during the off-line process.

Now, given a new sample $\boldsymbol{x}_s$, the problem is to decide what class it belongs to. This is carried out by applying the Bayes rule, in order to obtain the posterior probabilities $P\left(w_j \mid \boldsymbol{x}_s\right)$. A posterior probability determines the membership degree of

$\boldsymbol{x}_s$ to the class w_j once the sample has been observed and becomes available. We compute the a priori probability following the Bayesian framework [4], as follows,

$$P\left(w_j \mid \boldsymbol{x}_s\right) = \frac{p\left(\boldsymbol{x}_s \mid w_j\right)P\left(w_j\right)}{p\left(\boldsymbol{x}_s\right)} \tag{6}$$

So, the decision is made as follows,

$$\boldsymbol{x}_s \in w_j \text{ if } P\left(w_j \mid \boldsymbol{x}_s\right) > P\left(w_h \mid \boldsymbol{x}_s\right) \quad \forall h \neq j \tag{7}$$

The equation (7) can be re-written avoiding $p(\boldsymbol{x}_s)$, as it appears in both members of the inequality,

$$\boldsymbol{x}_s \in w_j \text{ if } p\left(\boldsymbol{x}_s \mid w_j\right)P(w_j) > p\left(\boldsymbol{x}_s \mid w_h\right)P(w_h) \quad \forall h \neq j \tag{8}$$

We still require computing the prior probabilities $P(w_j)$ and $P(w_h)$ involved in (8) because $p(\boldsymbol{x}_s \mid w_j)$ and $p(\boldsymbol{x}_s \mid w_h)$ can be obtained through the equation (3).

The computation of the a priori probabilities is carried out by exploiting the information provided previously by the VQ approach. Indeed, given $\boldsymbol{x}_s$ we compute the Euclidean distance $d_{sk}(\boldsymbol{x}_s,\boldsymbol{v}_j)$ from $\boldsymbol{x}_s$ to each cluster centre $\boldsymbol{v}_j \in w_j$, where the centres have been provided by the VQ module. The corresponding a priori probability is computed by applying the logistic function [4] as follows,

$$P(w_j) = \left(1 + e^{-Ad_{sk}(\boldsymbol{x}_s,\boldsymbol{v}_j)}\right)^{-1} \tag{9}$$

The constant A is included to avoid severe bias, it is set to 10^{-1} in this paper after trial and error experimentation.

Now, the final decision can be made through the equation (8).

As one can see the decision is made by combining two classifiers: VQ and BY. This is a suitable practice for solving classification problems [9].

3. Comparative and performance evaluation

We have used a set of 26 digital aerial images acquired during May in 2006 from the Abadia region located at Lugo (Spain). They are multi-spectral images with 512*x*512 pixels in size. The images are taken at different days from an area with several natural textures. The initial training patterns are extracted from 10 images of the full set. The remainder 16 images are used for testing and four sets, S0, S1 S2 and S3 of four images each one, are processed during the test according to the strategy described below. The images assigned to each set are randomly selected from the 26 images available.

3.1. Design of a test strategy

In order to assess the validity and performance of the proposed hybrid VQ and BY method we obtain two initial partitions, namely: P_{VB} and P_{MI}. P_{VB} is the automatic validated partition obtained by the unsupervised procedure described in this paper, Figure 1. P_{MI} is a manual partition obtained as described below.

Each partition is used as the initial training set of samples by the following three supervised classical clustering procedures [4]: *a*) VQ; *b*) BY and *c*) the Self-Organizing feature maps (SO).

This is intended in order to verify the performance of the proposed fusion approach, hereinafter VB, against the single methods VQ, BY and SO.

The test is carried out according to the following steps:

1) STEP 0 (initial partitions): for each image (from the 10 available) we perform a down sampling by a factor of 16, obtaining 10*x*32*x*32 training samples, i.e. n = 10240.

2) We apply our VB approach starting with the threshold $T = Q$ until the partition is validated. At this time we have available the final P_{VB} partition i.e. the number of clusters c and their corresponding cluster centres are known. Now, we manually select n training patterns from the same set of the 10 images and build c clusters driven by the previous cluster centers. Each manually selected training sample is compared with each cluster center and it is assigned to the cluster which gives the minimum feature (spectral)

distance between the sample and the corresponding cluster center. So, we obtain the manual initial partition P_{MI}.

The P_{VB} and P_{MI} partitions are used for classifying the new pattern samples during the next steps.

3) STEP 1: given the images in the sets S0 and S1, classify each pixel as belonging to a cluster according to the VB, BY, VQ and SO methods.

4) Compute the percentage of successes according to the ground truth defined for each class and for each image. The classified pattern samples from S1 are added to the previous training samples and a new training process is carried out according to each method. The set S0 is used as a sample set in order to verify the performance of the training process as the learning increases.

5) Perform the same process for STEPs 2 and 3 but using the sets S2 and S3 respectively instead of S1. Note that S0 is also processed as before.

The number of training samples added at each STEP is 4*x*512*x*512 because this is the number of pixels classified during the STEPs 1 to 3 belonging to the sets S1, S2 and S3.

3.2. Analysis of results

Figure 2 displays different threshold values against the averaged *divergence*, attached is displayed the number of clusters obtained.
The best mean *Divergence* score value is obtained for the threshold *T* set to 70.83, obtaining four clusters. Larger threshold values result in two clusters and with values greater than 160 a unique cluster is obtained (they are not displayed).

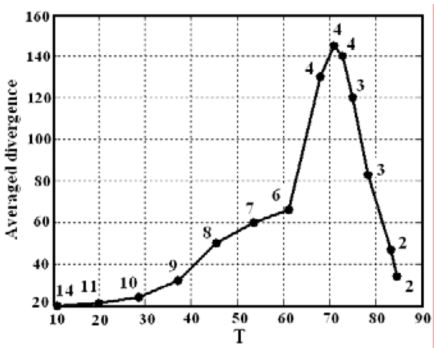

Figure 2. Threshold values against averaged divergence scores (attached is the number of cluster centres)

Figure 3 displays: (*a*) a representative original image to be classified belonging to the set S0; (*b*) the inter-clusters correspondence between the original colours and the labels assigned to each cluster; (*c*) the labelled image after the classification with our VB approach and (*d*) the ground truth for the cluster number two.

Each ground truth is built by applying the BY and then modifying the results manually according to the human expert criterion.

The colour for each cluster, Figure 3(*b*), matches with the natural colour assigned to the corresponding cluster centre. The labels are artificial colours derived from a colour map identifying each cluster.

The correspondence between labels and the different textures is as follows: 1) yellow with forest vegetation; 2) blue with bare soil; 3) green with agricultural crop vegetation and 4) red with buildings and man made structures.

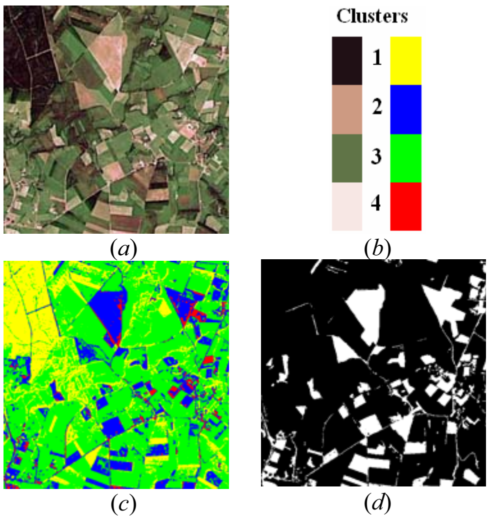

Figure 3. (*a*) original image; (*b*) colours and labels; (*c*) labelled textures; (*d*) ground truth for the cluster number two

Table 1 shows the percentage of successes in terms of correct classifications obtained for the different methods and for each initial partition. For each STEP we show the both sets of testing images processed SP0 and SP1/SP2/SP3. These percentages are computed taking into account the correct classifications for the four clusters according to the corresponding ground truth.

Figure 1 Percentage of successes obtained for the methods analysed for each partiton at the three STEPs.

%		STEP 1		STEP 2		STEP 3	
Partition	Methods	SP0	SP1	SP0	SP2	SP0	SP3
P_{VB}	VB	73.3	76.3	82.1	80.3	90.0	91.3
	BY	72.1	75.8	80.5	78.4	90.3	89.9
	VQ	71.3	70.1	75.1	77.3	79.5	83.3
	SO	72.7	72.0	77.2	79.4	84.3	87.4
P_{MI}	VB	75.6	77.2	83.1	81.1	93.7	91.7
	BY	75.2	76.9	81.6	80.4	92.5	90.9
	VQ	73.0	71.1	75.9	76.2	82.7	84.1
	SO	74.8	73.1	78.9	78.4	83.1	86.5

From results in table 1, one can see that the performance obtained with the initial partition P_{VB} is comparable to that obtained when P_{MI} is used for all methods and the three STEPs. This means that the proposed unsupervised approach performs in a similar fashion than the supervised ones.

The worst results are obtained with VQ when it is used isolated, but its combination with BY gives acceptable results. This behaviour was expected because supervised approaches perform better than unsupervised ones.

We can infer that the performance of the different methods improves as the number of training samples increases, i.e. high learning rates obtain better performances than low ones as expected in any learning process.

4. Conclusion

We propose a new unsupervised hybrid and automatic making decision process for classifying natural textures. The proposed method combines the unsupervised VQ and the supervised BY strategies achieving results which are comparable to those obtained by the supervised ones.

The performance of our proposed approach is analysed with different classical methods, including the methods which are combined, verifying that it performs favourably in the set of aerial images tested.

The method is applicable to other textured images even if using different attributes. The unique adaptation for new attributes comes from the computation of the attribute's values.

In the future, additional experiments should be required in order to deal with illumination variability. This is because the images are normally acquired during different days and obviously under different illumination conditions.

References

[1] Chan, J.W.C., Laporte, N. and Defries, R.S., Texture Classification of logged forest in tropical Africa using machine-learning algorithms. Journal Remote Sensing. 24(6), 1401-1407, 2003.

[2] Cherkassky, V., Mulier, F., Learning from data: concepts, theory and methods, Wiley, New York, 1998.

[3] Drimbarean, P.F. and Whelan, P.F., Experiments in colour texture analysis. Pattern Recognition Letters, 22, 1161-1167, 2001

[4] Duda, R.O., Hart, P.E. and Stork, D.G., Pattern Classification, Jhon Willey and Sons, New York, 2001.

[5] Frate, F. del, Pacifici, F., Schiavon, G. and Solimini, C., Use of Neural Networks for Automatic Classification from High-Resolution Images. IEEE Trans. Geoscience and Remote Sensing. 45(4), 800-809, 2007.

[6] Giacinto, G., Roli, F. and Bruzzone, L., Combination of neural and statistical algorithms for supervised classification of remote-sensing. Pattern Recognition Letters. 21(5), 385–397, 2000

[7] Hanmandlu, M., Madasu, V.K. and Vasikarla, S., A Fuzzy Approach to Texture Segmentation. Proc. IEEE Int. Conf.

Information Technology: Coding and Computing (ITCC04), The Orleans, Las Vegas, Nevada, 636-642, 2004.

[8] Jensen J.R., Introductory Digital Image Processing, Prentice-Hall, Englewood Cliffs, NJ, 1982.

[9] Kittler, J., Hatef, M., Duin R.P.W. and Matas, J., On Combining Classifiers IEEE Trans. on Pattern Analysis and Machine Intelligence, 20(3), 226-239, 1998.

[10] Lam, E. P., Wavelet-based texture image classification using vector quantization. In: Astola, J.T., Egiazarian, K.O and Dougherty, E.R. (eds.): Proc. SPIE, Image Processing: Algorithms and Systems V, vol. 6497, 687-696, 2007.

[11] Maillard, P., Comparing Texture Methods through Classification. Photogrammetric Engineering and Remote Sensing, 69(4) (2003) 357-367, 2003.

[12] Puig, D. and García, M.A.: Automatic Texture Feature Selection for Image Pixel Classification. Pattern Recognition, 39(11) 1996-2009, 2006.