A near real-time Evolution Strategy for Adaptive Color Quantization of image sequences

A.I. González, M. Graña, F.X. Albizuri, A. D'Anjou, F.J. Torrealdea

Dept. CCIA, Univ. Pais Vasco/EHU,Apdo. 649, 20080 San Sebastián, España.
e-mail: ccpgrrom@si.ehu.es

Abstract:
Color Quantization of still images can be easily stated as a Clustering problem. Color Quantization of sequences of images becomes a Non-stationary Clustering problem. In this paper we propose a very simple and effective Evolutive Strategy to perform the adaptive computation of the color representatives for each image in the sequence. Salient features of the Evolutive Strategy proposed here are: individuals correspond to individual cluster centers, to approach real-time response we impose one-generation adaptation for each image, mutation operators are guided by the actual covariance matrices of the clusters. Experimental results on a sequence of indoor images are presented.

1 Introduction.

Color Quantization (Heckbert 1980; Orchard, Bouman 1991; Lin, Chang 1995; Uchiyama, Arbib 1994), is an instance of the more general technique of Vector Quantization (VQ) (Gersho, Gray 1992) in the space of colors. Color Quantization has application in visualization, color image segmentation, data compression and image retrieval (Kankanhalli, Mehtre, Wu 1996) . The number of color representatives searched is tightly related to the application. In visualization and compression applications the typical size of the color palette (codebook, color representatives) is 256, whereas for segmentation and retrieval tasks the size of the color palette is smaller. We have chosen 16 as a typical number of color representatives for these latter kind of applications. At the present stage of our works, we do not deal with the problem of finding the *natural* number of colors, which is a much more involved problem. We think that it does not make sense to try to find out adaptively the natural number of clusters, without having tested the ability of the Evolution approach to perform adaptively the clustering into a fixed number of clusters. Besides, the definition of a numerical measure for the natural clustering problem is highly dependent of the application, and the subject of strong discussion inside the Clustering community.

Color Quantization of images within a sequence contains the essence of the paradigm of real time Non-stationary Clustering. Although sequences of images (video) lead naturally to the consideration of time varying Clustering/VQ problems, the usual approaches to the computation of codebooks for both Color Quantization and Vector Quantization of image sequences consider time invariant distributions of colors (Chen, Chien 1995) or image blocks (Chen, Chen 1995), and apply conventional Clustering methods. Some heuristic efforts (Gong, Zen, Ohsawa, Sakauchi 1992, Chen, Chen, Zhang 1994) have been reported that try to cope with the time varying characteristics inherent to image sequences. Our approach is to assume the problem as a Non-stationary Clustering problem that may be solved by the application of Adaptive VQ algorithms: the Color Quantization of the image sequences become the Adaptive Color Quantization problem. We propose an Evolution Strategy as an Adaptive Color Quantization algorithm. We demonstrate the effectiveness of our approach through the color quantization of a sequence of images that shows a smooth but clear variation over time of the distribution of colors.

Evolution Strategies (Back, Schwefel 1993,1996; Michalewicz 1996) have been developed mainly by Rechenberg and Schwefel since the sixties. They belong to the broad class of algorithms inspired by natural selection. The most widely accepted features of Evolution Strategies (ES) are the following:

(1) The individuals are vector real-valued,
(2) The main source of genetic variability is the mutation operator, although recombination operators are defined and applied by some authors.
(3) The individuals contain local information for mutation so that adaptive strategies can be formulated to self-regulate the mutation operator.

However, a lot of hybrid algorithms have been defined (Michalewicz 1996), so that it is generally difficult to assign a definitive "label" to a particular algorithm. The features that we identify as specific to the Evolution Strategy proposed in this paper are:

(1) The Evolution Strategy is intended to deal with a time varying environment. Therefore the fitness function will be time dependent.
(2) Individuals are defined as components of the solution instead of representing complete solutions. This implies the existence of a global fitness function for the entire population, on top of the individual fitness functions. The ES is expected to produce as a cooperative solution the best population to solve the problem. In this respect, our approach resembles largely the so-called Michigan approach to the design of Classifier Systems.
(3) Due to the problem representation chosen, the selection looks for entire populations, so that there is a strong interaction between parents and children in the evaluation of the selection operator.
(4) Mutation is the only operator that introduces evolution-like variability. The mutation guiding parameters, which in the ES literature are referred as self-adaptation parameters, are deduced from the interaction with the environment. In our final proposition mutation is performed deterministically to approach real time response. However, the Evolution Strategy that we will propose below remains an stochastic algorithm whose source of stochasticity is the input data.

The assumption of a time varying environment must not be confused with the case of a noisy environment (Fitzpatrick, Grefenstette 1988; Aizawa, Wah 1994). From our point of view the latter is a particular case of the former. The uncertainty associated to the environment is perceived by the algorithm through the variability of the fitness function. In the case of noisy environment as formulated in (Fitzpatrick, Grefenstette 1988), each evaluation of the fitness function involves a sampling procedure. This sampling procedure assumes a stationary random process as the source of the fitness function values. On the other hand, our approach assumes that the fitness function value will vary due to the inherent Non-stationarity of the environment. The fitness is measured upon a sample of the process. This sample is considered as representative of the environment for a limited amount of time. While the data sample remains the same, the fitness is a deterministic function. As far as the environment remains stationary, succesive data samples will posses the same statistical characteristics and the fitness function will continue to be (almost) the same. Unpredictable changes in the environment produce significant changes on the statistics of the data sample and, therefore, changes in the landscape of the fitness function. The Evolution Strategy tries to adapt as fast and smoothly as possible to the environment changes in an unsupervised way.

We want to obtain fast and good responses that can be used in a real time framework. To approach as much as possible to real time responses, we have imposed to our Evolution Strategy two restrictions:

(1) The adaptation must be performed in one generation
(2) The computations must be based on subsamples of the data.

These computational conditions are in conflict with the conventional view of Evolution Strategies (and similar algorithms) as random strategies for global optimization. We look for fast algorithms that give good, although suboptimal, solutions. We realize that this point of view is well outside the orthodox thinking of the Evolutionary Computation community, but we stick to our labeling of the algorithm presented below as an Evolution Strategy because this is the closest category of stochastic algorithms from structural and behavioral point of views.

The paper is organized as follows. Section 2 introduces the adaptive approach to Non-stationary Clustering and Adaptive Color Quantization. Section 3 presents the Evolution Strategy proposed. Section 4 presents the experimental results, and section 5 gives our conclusions and lines for further work.

2 Non-stationary Clustering and Adaptive Color Quantization.

Cluster Analysis and Vector Quantization are useful techniques in many engineering and scientific disciplines (Gersho, Gray 1992; Hartigan 1975; Diday, Simon 1980; Duda, Hart 1973; Jain, Dubes 1988; Fukunaga 1990). In their most usual formulation it is assumed that the data is a sample of a Stationary Stochastic Process, whose statistical characteristics will not change in time. Non-stationary Clustering and Adaptive Vector Quantization assume a Non-stationary Stochastic Process that is sampled at diverse time instants. An important remark: no knowledge of the time dependencies is assumed. If a model is known (or assumed), a predictive approach (Gersho, Gray 1992) would reduce the problem to a stationary one. We have found in the literature references to dynamic approaches to Clustering (Chaudhuri 1994; Garcia et alt 1995). These approaches are incremental or progressive strategies for the search of the optimal clustering of stationary data, and therefore not related to our present work.

Let us start with a working definition for the Clustering problem in the stationary case:

> Given a set of vectors $\aleph = \{\mathbf{x}_1,..,\mathbf{x}_n\}$. Obtain a partition of them into a set of disjoint clusters $\{\aleph_1,..,\aleph_c\}$ that minimizes a criterion function C.

The definition of the criterion function involves the definition of a dissimilarity measure. The most used for practical purposes is the Euclidean distance, and the most common criterion function is the within-cluster scattering

$$S_W = \sum_{i=1}^{c} \sum_{\mathbf{x}\in\aleph_i} p(\mathbf{x}) \left\| \mathbf{x} - \mathbf{m}_i \right\|^2 \quad \text{where} \quad \mathbf{m}_i = \left(\sum_{\mathbf{x}\in\aleph_i} p(\mathbf{x}) \right)^{-1} \left(\sum_{\mathbf{x}\in\aleph_i} p(\mathbf{x})\mathbf{x} \right) \qquad (1)$$

(usually $p(\mathbf{x})=n^{-1}$ is assumed). The Vector Quantization design problem is tightly related to the Stationary Clustering problem, a working definition can read as follows

> Search for a set of representatives $\mathbf{Y} = \{\mathbf{y}_1,..,\mathbf{y}_c\}$ that minimize the quantization error (distortion) function E that measures the error of the substitution of the vectors in the set $\aleph = \{\mathbf{x}_1,..,\mathbf{x}_n\}$ by their nearest neighbor representatives (codification) according to a similarity measure d .

$$E=\sum_{j=1}^{n}\sum_{i=1}^{c} d\left(\mathbf{x}_j-\mathbf{y}_i\right)\delta_{ij};\quad \delta_{ij}=\begin{cases}1 & \mathrm{i}=\underset{k=1,..,c}{\operatorname{argmin}}\left\{d\left(\mathbf{x}_j-\mathbf{y}_k\right)\right\}\\ 0 & \text{otherwise}\end{cases}. \tag{2}$$

When the similarity measure is the Euclidean distance, the quantization error function is the Squared Error, then both the set of optimal representatives $\mathbf{Y}=\{\mathbf{y}_1,..,\mathbf{y}_c\}$ and the centroids of the optimal clustering partition coincide. Then, Clustering and Vector Quantization are the same problem. The optimal partition of the data sample is defined by the nearest representative:

$$\mathbf{x}_j \in \aleph_i \Leftrightarrow \mathrm{i}=\underset{k=1,..,c}{\operatorname{argmin}}\left\{\left\|\mathbf{x}_j-\mathbf{y}_k\right\|^2\right\}. \tag{3}$$

Note that the above definitions do not apply if the number of clusters (codebook size) is not known beforehand. As said in the introduction, the determination of the natural clustering is a much more involved problem.

The formulation of the Non-Stationary Clustering problem must start with the explicit assumption of a time varying population described by an stochastic process $\{\mathbf{X}_t \;\; t=0,1,..\}$ (note that we have jumped into the discrete time case). A working definition of the Non-Stationary Clustering problem could read as follows:

> Given a sequence of samples $\aleph(t)=\{\mathbf{x}_1(t),..,\mathbf{x}_n(t)\};\; t=0,1,2,...$ Obtain a corresponding sequence of partitions, each over the sample data at each time instant, $\mathcal{P}(\aleph(t))=\{\aleph_1(t),..,\aleph_c(t)\}\;\; t=0,1,2,...$ This sequence of partitions minimizes an accumulative criterion function $C=\sum_{t\geq 0} C(t)$.

The similar Adaptive Vector Quantization design problem can be stated as

> Search for a sequence of sets of representatives $\mathbf{Y}(t)=\{\mathbf{y}_1(t),..,\mathbf{y}_c(t)\}\;\; t=0,1,2,..$ that minimizes the accumulative quantization error (distortion) function $E=\sum_{t\geq 0} E(t)$.

Again, we consider the squared Euclidean distance as the similarity/dissimilarity measure and, therefore, the within cluster variance and squared error as clustering criterion and quantization error function, respectively. Non-Stationary Clustering and Adaptive Vector Quantization are equivalent problems and the optimal sequence of partitions is given by the nearest (Euclidean) optimal cluster representative:

$$\mathbf{x}_j(t) \in \aleph_i(t) \Leftrightarrow \mathrm{i}=\underset{k=1,..,c}{\operatorname{argmin}}\left\{\left\|\mathbf{x}_j(t)-\mathbf{y}_k(t)\right\|^2\right\}. \tag{4}$$

In (Gersho, Gray 1992) predictive approaches to Vector Quantization are discussed as multidimensional extensions of the scalar predictive approach, widely used in one dimensional signal processing and compression. These approaches are based in the formulation of a model of the time dependencies of the data. Several adaptive approaches have been also proposed (Gersho, Gray 1992, Fowler 1996) to deal with non stationary data. The most popular ones are the algorithms that perform the substitution of the cluster representatives (codevectors) based on some heuristically determined parameters (Chen et alt. 1994). There have been some suggestions that stochastic algorithms such as the LMS stochastic minimization, related to neural network learning algorithms, could be of use for the adaptive computation of the time varying means. The Evolution Strategy proposed in this paper falls in the category of stochastic algorithms that compute adaptively the cluster means. As an illustration of the most general non-stationary case, we propose the Color Quantization of image sequences that show a smooth but unpredictable color variation. Figure 1 shows the distributions of pixels in the RGB unit cube for the images in the sequence used in the experiments reported below. It can be appreciated that the color distribution spreads and shrinks in unpredictable ways. This representation gives a straight illustration of the strong Non-Stationary nature of the data we are handling. The data samples $\aleph(t) = \{\mathbf{x}_1(t),..,\mathbf{x}_n(t)\};\ t = 0,1,2,..$ used in the experiments are randomly extracted from the data shown in figure 1.

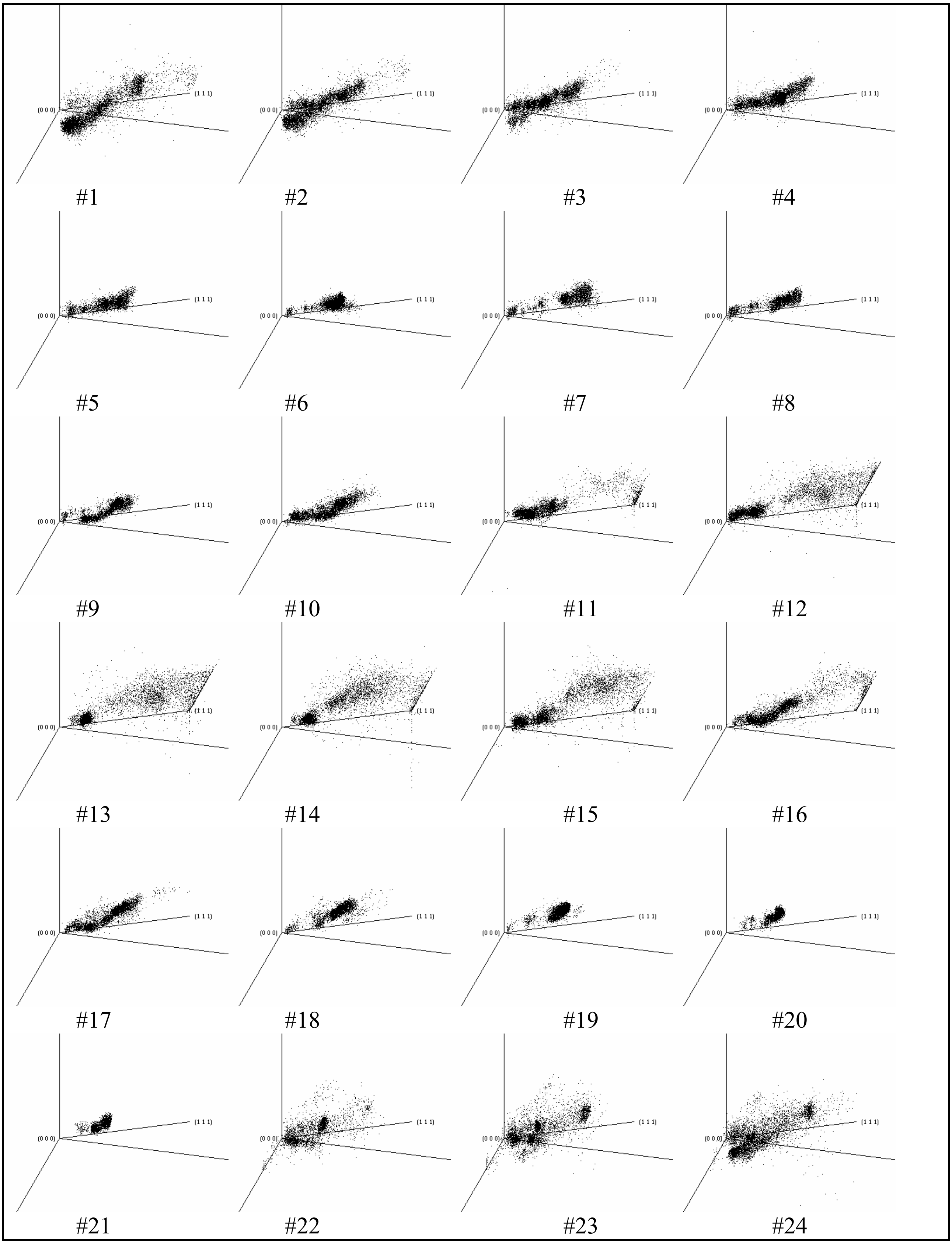

#1
#2
#3
#4
#5
#6
#7
#8
#9
#10
#11
#12
#13
#14
#15
#16
#17
#18
#19
#20
#21
#22
#23
#24

Figure 1. Distribution of pixel color for the images in experimental sequence.

The Adaptive VQ problem under the Euclidean distance assumption can be stated as the following stochastic minimization problem:

$$\min_{\{\mathbf{Y}(t)\}} \sum_{t\geq 0} \sum_{j=1}^{n} \sum_{i=1}^{c} \left\| \mathbf{x}_j(t) - \mathbf{y}_i(t) \right\|^2 \delta_{ij}(t); \quad \delta_{ij}(t) = \begin{cases} 1 & i = \operatorname{argmin} \left\{ \left\| \mathbf{x}_j(t) - \mathbf{y}_k(t) \right\|^2 k = 1,..,c \right\} \\ 0 & \text{otherwise} \end{cases}. \quad (5)$$

The proposition of adaptive algorithms to solve this stochastic minimization problem is based in the assumption of the following two conditions:

(1) The independent minimization of the error function at each time step produces the minimization of the accumulative error function
(2) The statistical characteristics of the underlying physical process have a smooth variation.

The first condition implies that the problem can be decomposed into a sequence of isolated problems. Therefore, bad solutions in a given time instant do not degrade the overall response of the adaptive algorithm along time. Besides, if we can compute an optimal solution for the stationary case, we can obtain an optimal solution of the non-stationary case, through the computation of the optimal solutions at each time instant. To obtain the benchmark sequence of color quantizers over the experimental sequence shown in figure 1, we compute a near optimal color quantizer for each image applying the Minimum Variance Heckbert Algorithm over it.

The second condition implies that adaptive algorithms can be formulated as local minization procedures. The solution computed for the previous time step can be assumed as a good initial condition for the next time step. Therefore, the local minimization performed by the adaptive algorithm can produce near optimal results. We will discuss at the end of the next section the adaptive application of the Evolution Strategy proposed in this paper.

3 An Evolution Strategy and its adaptive application to Adaptive Color Quantization of image sequences

A representative sample of the works found in the literature dealing with clustering problems via Evolutionary Computation is (Alippi, Cucchiara 1992; Andrey, Tarroux 1994; Babu, Murty 1994; Bezdek, Hathaway 1994; Bezdek et alt 1994; Bhuyan et alt 1991; Blekas, Stafylopatis 1996; Buckles et alt. 1994; Jones, Beltrano 1990; Ketaff, Asselin 1994; Lucasius et alt 1993; Luchian et alt 1994; Moraczewski et alt 1995). The common approach of all these works is the mapping of complete clustering solutions to population individuals. The fitness function is the ad-hoc clustering criterion function. The authors propose a wide variety of representation of clustering solutions as population individuals, ranging from the set of cluster representatives to the membership (hard or fuzzy) matrices of the clusters. Evolution operators, recombination and mutation, are defined suitably to be closed operators on the representation chosen.

Our conclusion from the literature review, is that most of the Evolutionary approaches suggested for clustering could not be applied to the non-stationary case in a stringent time frame. They can not guarantee a reasonable response in a reasonable time. Most of the approaches found in the literature have a big uncertainty about the proper setting of the algorithm parameters (population size, mutation and crossover rate, the appropriate operators,...). Assuming that the previous criticisms could be properly answered, the computational complexity of each generation is usually very big , so that even in the case that the evolutionary approach is used with a computational limit imposed, this limit will be necessarily very high for practical applications of the kind we are interested in. We have honestly tried to address the problem in a way that is both computationally effective and gives good solutions, assuming its suboptimality.

A widely accepted (Back, Schwefel 1996) pseudocode representation of the general structure of the algorithm of Evolution Strategies is given in figure 2. Note that the generation number in this figure is a time parameter **t** that we have typed in bold to distinguish it clearly from the time parameter t used in section 2. The adaptive application of the Evolution Strategy will impose the statement of a mapping between these two time parameters; which represent, respectively, the internal (computational) and external (environmental) times. The Evolution Strategy proposed is heavily influenced by the use of Euclidean distance, the consideration of other clustering measures will imply that some algorithm elements must be redefined. We will start describing in detail the elements of the Evolution Strategy, and then we will discuss its application as an Adaptive VQ algorithm.

```
t:= 0
initialize P(t)
evaluate P(t)
while not terminate do
    P'(t):= recombine P(t)
    P"(t):= mutate P'(t)
    evaluate P"(t)
    P(t+1):= select (P"(t) U Q)
    t:= t+1
end while
```

Figure 2. General structure of an Evolution Strategy.

3.1 Problem codification: The individuals and the population. Local and global fitness functions.

We make each individual to correspond to a single cluster center. A single solution to the Clustering/VQ problem is mapped into the entire population. The population at generation **t** is given by

$$P(\boldsymbol{t}) = \{\mathbf{y}_i(\boldsymbol{t}); i = 1..c\}. \tag{6}$$

The population size c corresponds to the number of clusters searched in the data. We have not included mutation parameters in the definition of the individuals, because we will use for this role the covariance matrices computed over the sample data.

The local fitness of each individual is its local quantization error relative to the sample considered in this generation.

$$F_i(t) = \sum_{j=1}^{n} \left\| \mathbf{x}_j(t) - \mathbf{y}_i(t) \right\|^2 \delta_{ij}(t). \tag{7}$$

The sample data $\aleph(t) = \{\mathbf{x}_1(t), .., \mathbf{x}_n(t)\}$ used to compute this fitness is determined by the correspondence between internal and external time parameters. This correspondence is specified when describing the adaptive application of the Evolution Strategy (see section 3.4). As the individuals do not specify clustering solutions, we must consider a fitness function for the population as a whole. This population fitness corresponds to the objective function to be minimized, because it is the population as a whole which specifies the clustering solution, and can be evaluated as

$$F(t) = \sum_{i=1}^{c} F_i(t). \tag{8}$$

If the individual fitness could be considered as non interacting functions, their separate optimization would trivially produce the optimization of the population fitness. However, in our case we have a clear interaction between individual fitness functions. Our population fitness corresponds to the within cluster scatter S_W of the clustering specified by the population. The well known equation relating the within cluster and between cluster scattering (Duda, Hart 1973)

$$S = S_W + S_B \tag{9}$$

can be written in the context of Non-stationary Clustering as:

$$S(t) = \sum_{j=1}^{n} \left\| \mathbf{x}_j(t) - \overline{\mathbf{y}}(t) \right\|^2 = \sum_{i=1}^{c} F_i(t) + \sum_{i=1}^{c} \left\| \mathbf{y}_i(t) - \overline{\mathbf{y}}(t) \right\|^2 \tag{10}$$

where $S(t)$ remains constant as far as the same data sample is considered, and $\overline{\mathbf{y}}(t)$ denotes the centroid of the entire data sample $\aleph(t)$ considered at time $\boldsymbol{t}$. What we expect of the Evolution Strategy is that it will implicitly react through the above equation balancing the minimization of the population fitness, from the local optimization of individual cluster representatives, and the maximization of the between cluster scattering.

3.2 The mutation operator.

The recombination operators found in the literature of Evolution Strategies do not look as appropriate sources for new cluster representatives, therefore we have not defined any recombination operator. Evolutive changes are introduced exclusively by the mutation operator. As is customary in Evolution Strategies, our mutation operator is a random perturbation that follows a zero-mean normal distribution. The design questions relevant to the definition of the mutation operator are :

(1) Which individuals will be mutated? The set of mutation parents is composed of the individuals whose local fitness is greater than the mean of the local fitness in its generation. Formally, this set is given by:

$$\phi(t) = \{ i | F_i(t) \geq \bar{F}(t) \} \text{ where } \bar{F}(t) = \frac{1}{c} \sum_{i=1}^{c} F_i(t). \quad (11)$$

(2) How many mutations will be allowed? We have decided to approach as much as possible to a fixed number of mutations *m*, so that the number of mutations per individual $m_i(t)$ will depend on the size of $\phi(t)$,

$$m_i(t) = \left\lceil \frac{m}{|\phi(t)|} \right\rceil. \quad (12)$$

(3) What information will be used to compute mutations?. We can use the local covariance matrices of the sample partition associated with each individual, so that the mutation operator is naturally adapted to each individual. The expression of the local covariance matrices is

$$\hat{\Sigma}_i(t) = (n-1)^{-1} \sum_{j=1}^{n} (\mathbf{x}_j(t) - \mathbf{y}_i(t))(\mathbf{x}_j(t) - \mathbf{y}_i(t))^t \delta_{ij}(t). \quad (13)$$

(3 cont.) We have tested a deterministic approximation to the mutation operator in order to avoid the variability introduced by the random generation of perturbations. Mutations are computed along the axes defined by the eigenvectors of the estimated local covariance matrix. The number of mutations along each eigenaxis is proportional to the relative magnitude of its eigenvalue. Let $\Lambda_i = \mathrm{diag}(\lambda_{ij}, j = 1..3)$ and $\Phi_i = [\mathbf{e}_{ij}, j = 1..3]$ denote, respectively, the eigenvalue and eigenvector matrices of $\hat{\Sigma}_i(t)$. Then the set of mutations generated along the axis defined by eigenvector $\mathbf{e}_{ij}$ is:

$$P''_{ij}(t) = \{ \mathbf{y}_i \pm \alpha_k \lambda_{ij} \mathbf{e}_{ij} | k = 1..m_{ij}(t), i \in \phi(t) \} \quad (14)$$

$$m_{ij}(t) = \mathrm{round}\left(\frac{m_i(t)\lambda_{ij}}{2\sum_{l=1}^{3} \lambda_{il}} \right), \ \alpha_k = \frac{1.96k}{m_{ij}(t)}. \quad (15)$$

The set of individuals generated by the mutation operator is

$$P''(t) = \bigcup_{i,j} P''_{ij}(t). \quad (16)$$

3.3 The selection operator.

The last operator to be defined is the selection operator, which determines the individuals of the population for the next generation. In the definition of this operator we have followed the so called (μ+λ)-strategy. We pool together parents and children, so that $Q = P(t)$. Selection can not be based on the original individual fitness functions $F_i(t)$ because they do not have information about the interaction effects introduced by the mutation generated individuals. The optimal approach to the implementation of the selector operator consists in computing the fitness of all the possible populations of size c extracted from $P''(t)\cup P(t)$. That means to compute $\binom{|P''(\mathbf{t})\cup P(\mathbf{t})|}{c}$ population fitness functions. This computational burden largely questions the feasibility of applying this approach in any real time application. Therefore, we have tested two alternative selection operators of lesser complexity. We will describe then in the order of decreasing complexity and optimality.

3.3.1 The Selection Operator 1.

One way to reduce the combinatorial growing of the complexity of the selection operator is to try to explore the solutions in order, performing a greedy search. The selection procedure results in a complexity that grows quadratically with (c+λ). That means that it requires the computation of $(c+\lambda)^2$ population fitness functions. The procedure tries to select the cluster representatives in order of decreasing distortion. Given a set of currently selected cluster centers $\left\{\mathbf{y}_{i_1},\ldots,\mathbf{y}_{i_k}\right\}\subset P''(t)\cup P(t)$, with k<c, we select the next cluster center as the one that added to the previous selected ones produces the smaller distortion: the minimum of the (c+λ–k) distortions that can be computed based on the sub-populations that can be formed adding one cluster representative (individual) to the k already selected.

More formally, this selection operator can be described as follows:

$$P(t+1)=\mathbf{select}_{op1}\left(P''(t)\cup P(t)\right)=P^*(t)\subset P''(t)\cup P(t) \tag{17}$$

where the set $P^*(t)=\left\{\mathbf{y}_{i_k};k=1..c\right\}$ is constructed iteratively by selecting the indices applying the following recursive expression (note that we drop the time index for simplicity):

$$i_k=\underset{\substack{i=1,\ldots,c+\lambda\\ i\notin I_{k-1}}}{argmin}\left\{\sum_{l\in I_{k-1}\cup\{i\}}\sum_{j=1}^{n}\left\|\mathbf{x}_j-\mathbf{y}_l\right\|^2\delta_{jl}^{I_{k-1}\cup\{i\}}\right\} \tag{18}$$

where $I_k=I_{k-1}\cup\left\{i_k\right\}$ and $I_0=\varnothing$ are the sets of indices considered at each step of the iteration, and the membership function is dependent on this set:

$$\delta_{jl}^{I}=\begin{cases}1 & l=\underset{i\in I}{argmin}\left\{\left\|\mathbf{x}_j-\mathbf{y}_i\right\|^2\right\}\\ 0 & otherwise\end{cases}. \tag{19}$$

3.3.2 The Selection Operator 2.

Given the time constraints imposed by our intended application, we looked for a faster, although suboptimal, approach. We pool together the parents and the individuals generated by mutation. Let us denote as $F^S(t)$ the fitness of the population $P''(t)\cup P(t)$. A way to measure the importance of a given cluster representative is to compute the effect of removing it from the set of cluster representatives. That is, we compute $F_i^S(t)$ as the fitness of the population $P''(t) \cup P(t) - \{\mathbf{y}_i\}$ for each $\mathbf{y}_i \in P''(t) \cup P(t)$. The significance of the individual would be measured by $F_i^S(t) - F^S(t)$. As $F_i^S(t) \geq F^S(t)$ for all the individuals, it suffices to compute $F_i^S(t)$ to measure the *significance* of the individual. Notice that for empty cluster representatives, which can be the case of some mutation generated individuals, their significance is null $F_i^S(t) = F^S(t)$, so that they will be discarded automatically. It is trivial to verify that no empty cluster will be selected using this fitness function, unless there is someone in the original population and all the mutations generate empty cluster representatives.

For notation simplicity, let $\lambda = |P''(t)|$ be the number of individuals effectively generated by mutation. A formal definition of the individual fitness function used by the selection operator is as follows:

$$F_i^S(t) = \sum_{k=1; i \neq k}^{c+\lambda} \sum_{j=1}^{n} \left\| \mathbf{x}_j(t) - \mathbf{y}_k(t) \right\|^2 \delta_{kj}^s(t) \tag{20}$$

$$\delta_{kj}^s(t) = \begin{cases} 1 & k = \underset{l=1,..,c+\lambda; l \neq i}{\operatorname{argmin}} \left\{ \left\| \mathbf{x}_j(t) - \mathbf{y}_l(t) \right\|^2 \right\} \\ 0 & \text{otherwise} \end{cases} \tag{21}$$

The selection operator selects the c best individuals according to this fitness function:

$$\begin{aligned} P(t+1) &= \mathbf{select}_{op2}\left(P''(t) \cup Q\right) = \left\{ \mathbf{y}_i \in P^*(t); i = 1..c \right\} \\ P^*(t) &= \left\{ \mathbf{y}_{i_1}, .., \mathbf{y}_{i_{c+\lambda}} \,\middle|\, i_j < i_k \Rightarrow F_j^S(t) > F_k^S(t) \right\} \end{aligned} \tag{22}$$

The computation requirements of this selection operator are linear in the number of cluster representatives, and it can be easily speed up using the simple programming trick of precomputing the two nearest cluster representatives. However, this selection operator is clearly suboptimal. The experimental works try to asses the trade-off of its suboptimality versus its computational efficiency.

3.4 The adaptive application of the Evolution Strategy.

The adaptive application of Evolution Strategies involves the mapping of the two discrete time axes: the data time parameter t and the generation **t** of the Evolution Strategy. The most conventional approach would allow for several generations between data samples, so that :

$$t = \left\lfloor \frac{\boldsymbol{t}}{\tau} \right\rfloor + 1 \tag{23}$$

where τ is the number of evolution generations allowed between input data samples. In the context of Color Quantization of image sequences, τ is the number of generations computed between presentations of image samples. The initial condition corresponds to the initial color representatives provided for the first image, and the adaptation starts upon the sample from the second image. A distinctive feature of our experiments below is that we impose a one generation framework, that is $\tau=1$.

4 Experimental results.

The experiments reported in this section were designed with two goals in mind: first to demonstrate that the proposed Evolution Strategy could be considered an adaptive algorithm, second to evaluate the sensitivity of the Evolution Strategy to some of its elements and parameters. The experiments have been performed over an image sequence that shows a smooth but unpredictable variation of the color distribution. The extent of the non-stationarity of the sequence can be appreciated in figure 1. The images come from a panning of the laboratory taken with an electronic camera. Original images have an spatial resolution of 480x640 pixels. Each image overlaps about 50% of the scene with the next image. This overlapping was intended to provide the required smooth (although unpredictable) variation of the color distributions. In a video sequence (25 to 30 images per second), these images would come from a time sub-sampling. Depending on the camera motion, the time between the shots represented in figure 1 would range from 0.1 up to 1 second. These are the time constraints that we have in mind in the design and experimentation with our Evolution Strategy.

As a benchmark non adaptive algorithm we have used a variation of the algorithm proposed by (Heckbert 1980) as implemented in MATLAB, we call this algorithm Minimum Variance Heckbert algorithm. This algorithm recursively partitions the RGB unit cube along the axis of maximum variance. The partition is performed by a plane orthogonal to the color axis chosen so as to minimize the sum of the residual variances. A time efficient method to compute the residual variances was presented in (Wu 1991). This method is implemented in many standard libraries. However, the algorithm involves the pre-computation of all the potential variances. Its time and space complexity grows exponentially with the dimension of the space, and the number of values of the discretization of each space axis. Color representatives are computed as the center of mass of the resulting partition cubes. This algorithm has been applied to the entire images in the sequence in two ways. Figure 3 shows the distortion results of the Color Quantization of the experimental sequence to 16 based on both applications of the Heckbert algorithm. The curve denoted *Time Varying Min Var* is produced assuming the non-stationary nature of the data and applying the algorithm to each image independently. The curve denoted *Time Invariant Min Var* come from the assumption of stationarity of the data: the color representatives obtained for the first image are used for the Color Quantization of the remaining images in the sequence. The gap between those curves gives an indication of the non stationarity of the data. Also this gap defines the response space left for truly adaptive algorithms. To accept an algorithm as an adaptive solution its response could not be worse than the *Time Invariant Min Var* curve.

In the application of the Evolution Strategy described in the previous section to Adaptive Color Quantization, the data samples at each time instant were sets of pixels picked randomly from the image. This image sub-sampling was aimed to approach as much as possible the real time constraints. The algorithm was applied in the one-generation time schedule, starting on the sample of the second image, and using as initial population $P(1)$ the Heckbert palette of the first image. The populations $P(t)$ are the color palettes used to perform the Color Quantization. The results of the Evolution Strategy are always shown together with the benchmark curves of figure 3, in order to show its adaptive behavior. The historical sequence of our experiments started with Selection Operator 2, so in the following it is the used selection operator unless stated otherwise.

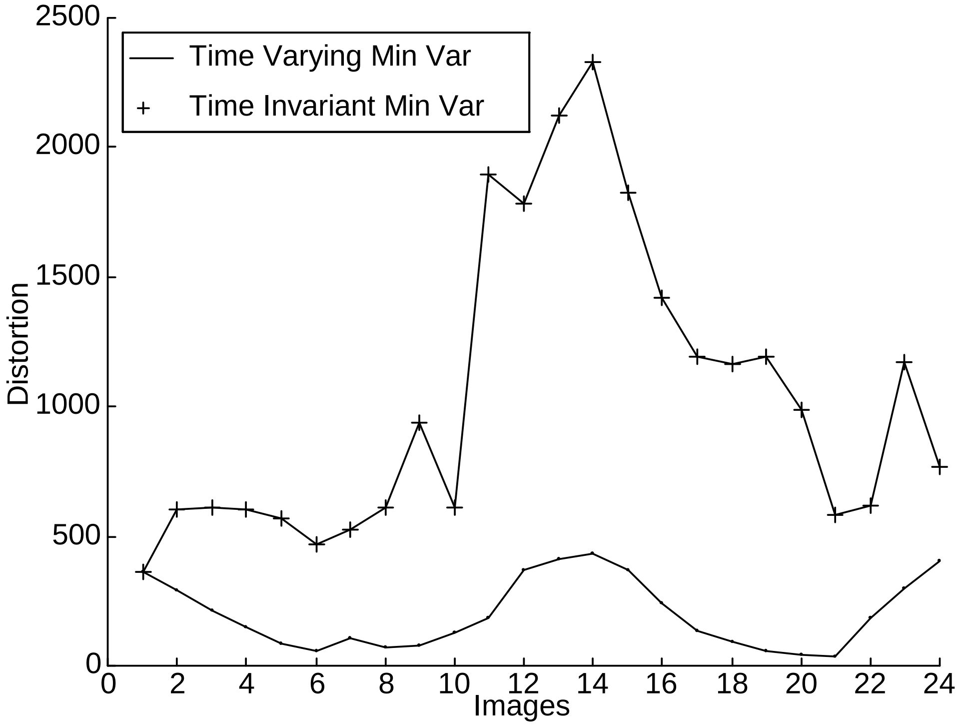


Figure 3. Benchmark distortion values obtained with the application of the Matlab implementation of the Heckbert algorithm to compute the color quantizers of 16 colors of the images in the experimental sequence.

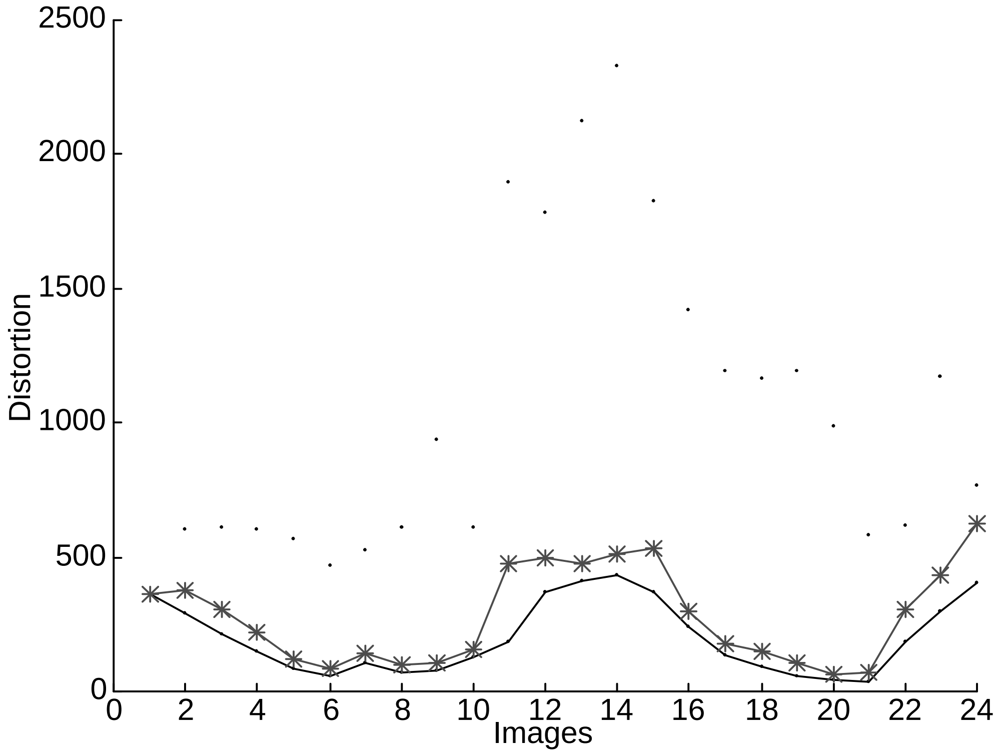


Figure 4. Results of the application of the Evolution Strategy with the deterministic mutation operator upon samples of n=1600 pixels with c=16, m=16.

The first experiment tried to evaluate the performance of the Evolution Strategy using the deterministic mutation operator. We have reported elsewhere (Gonzalez et alt. 1998) the results obtained with the Monte Carlo simulation of the random perturbations that realize the mutation operator. These experiments showed that the random nature introduced a high variance of the results when a one generation adaptation schedule was imposed. The deterministic mutation operator was proposed to avoid this variability, but the question was if the Evoltion Strategy remained an adaptive algorithm. Figure 4 shows the results of the application of the Evolution Strategy with the deterministic mutation operator to the color quantization of the experimental image sequence to 16 colors. In this figure the distortion results shown are those of the Color Quantization of the entire images, by the color quantizers computed using image samples of 1600 pixels. The inspection of figure 4 confirms that the deterministic mutation operator gives a good approximation to the mean behavior of the random mutations simulated via Monte Carlo Methods (Gonzalez et alt. 1998) and that the algorithm remains adaptive in the sense of performing better than the Time Invariant Min Var curve.

We have performed an exploration of the sensitivity of the deterministic Evolution Strategy to the ratio of sample size to the number of colors. Obviously, sample size influences the computational requirements, so that small samples are preferred. The tradeoff is the degradation in the response obtained due to the loss of information. Figure 5 shows the distortion of the Color Quantization of the entire images with palettes of 16 colors computed by the Evolution Strategy varying the size of the sample from 400 to 25600 pixels, which means a variation of the ratio sample:codebook from 6:1 up to 1600:1.

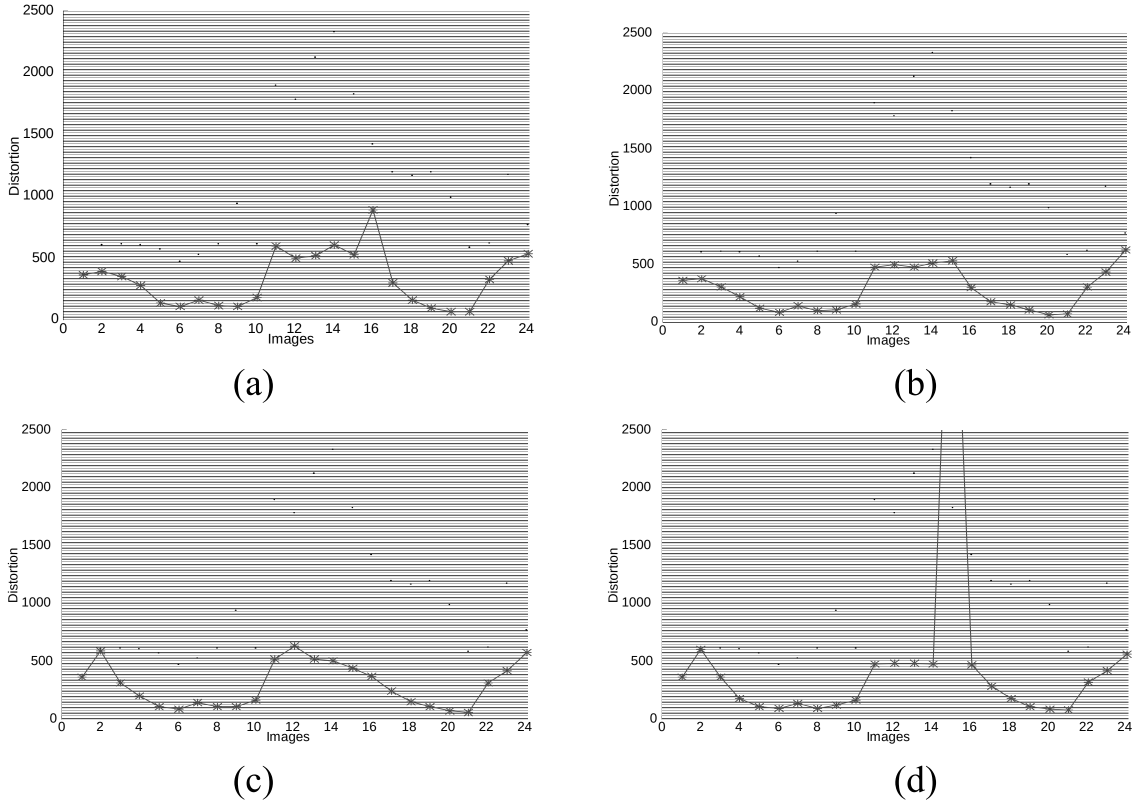
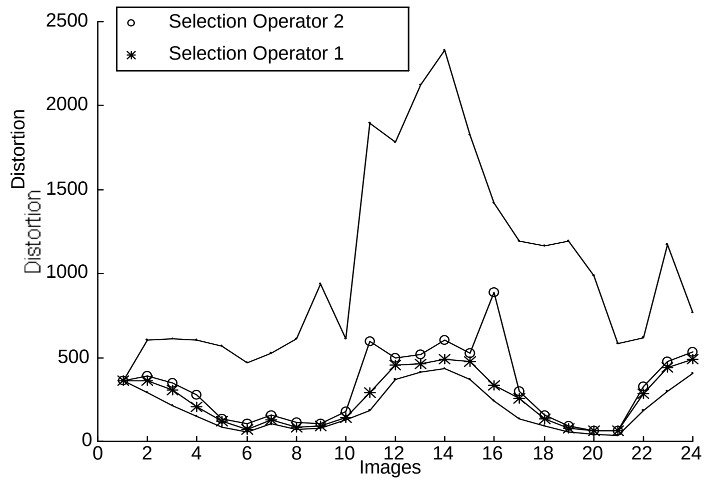
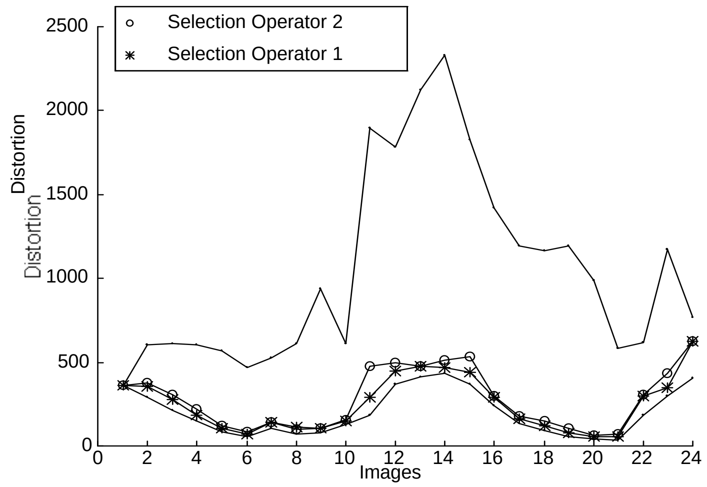


Figure 5. Sensitivity of the Evolution Strategy with deterministic mutation operator to the size of the sample, c=16, m=16. (a) 400, (b) 1600, (c) 6400, (d) 25600 pixels.

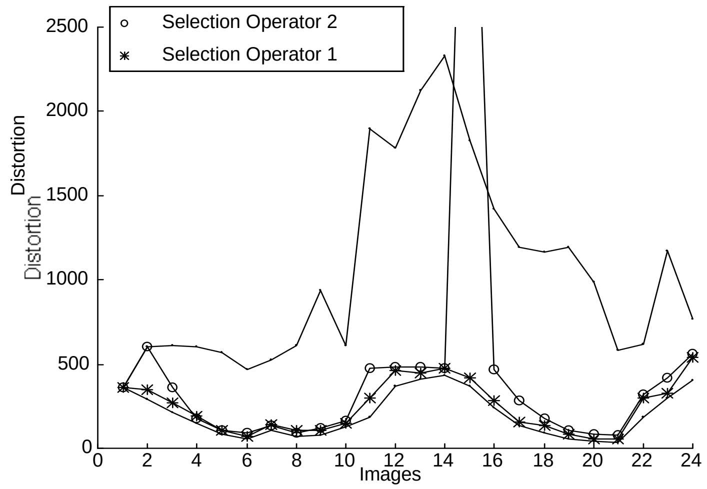
It can be appreciated in figure 5 that increasing the sample size improves the response of the Evolution Strategy, approaching that of the *Time Varying Min Var* algorithm. An optimal tradeoff between efficiency and computation requirements could be identified with a sample size of 1600 pixels (a sample:codebook ratio of 100:1). There is, however, an strange effect for the biggest sample size. The Evolution Strategy gives an anomalous response for image #15 and recovers its adaptive behavior afterwards. We have hipothesized that this unexpected degradation of the response may be related to the suboptimal definition of the Selection Operator 2 applied up to now. The significance measure computed decreases its variability as the number of colors and the sample size increase. That means that the Selection Operator 2 ability to discriminate good individuals decreases accordingly. This sensitivity could explain the anomaly in figure 5d. To test this hipothesis we have formulated the Selection Operator 1. We propose it as a complexity intermediate solution between the linear but suboptimal Selection Operator 2 and the infeasible optimal selection operator. The idea behind the next experiment is that if the Evolution Strategy with Selection Operator 1 recovers the anomalies, the responsability for them would no lie in the deterministic mutation operator, but in the suboptimal choice performed by the Selection Operator 2. This would allow the safe proposition of the Evolution Strategy with the deterministic mutation operator as a reduced complexity and variance Adaptive Color Quantization algorithm.



(a)



(b)



(c)

Figure 6. Comparative results of the application of the deterministic Evolution Strategy with Selection Operator 1 and 2. c*=16,* m*=16. Samples of size 400 (a) , 1600(b) and 25600 (c)*

Figure 6 shows the results of the final experiment that compare the response that both selection operators give in the application of the Evolution Strategy (with the deterministic mutation operator) in the case of 16 colors. There is a general improvement of the response in all sample sizes tested. The Selection Operator 1 improves in the case of bad sample size selection, the strange effect detected for sample size 25600 disappears, and it performs better in the case of very small (400) samples.

These results have two meanings. The first one is the expected conclussion that Selection Operator 2 improves the convergence of the Evolution Strategy, a natural result. The second, is the confirmation that the deterministic mutation operator can be applied without introducing serious degradations of the algorithm response. The anomalous effects are due to the selection operator used.

However, the quadratic growth of his complexity is a serious impediment for its practical application. Selection Operator 1 is more sensitive to the size of the sample and the number of mutations, due to his inherent suboptimality. However, it is very efficient computationally, and can be of practical use for real time applications, if properly tuned. On top of that, we remind the reader that Color Quantization is an instance of the general Clustering problem, where much bigger problems can be posed.

5 Conclusions and further work.

Most of the works done on Clustering deal with stationary data, assumed to come from a time invariant population. In this paper we deal with non-stationary data that comes from a population whose characteristics change with time. We have given working definitions of Non-stationary Clustering and Adaptive Vector Quantization. We have found that the problem of Color Quantization of image sequences is an instance of Adaptive Vector Quantization, that we have termed Adaptive Color Quantization. We have designed a Color Quantization experiment that shows the characteristics that are specific of Non-stationary Clustering problems: a sequence of population distributions that show an unpredictable smooth (bounded) change.

We propose an Evolution Strategy for the adaptive computation of the cluster representatives at each time, and we have applied it to the computation of the color representatives for the Color Quantization of the experimental image sequence. The design of this Evolution Strategy has a main computational constraint: it must approach real time performance as much as possible, with the lowest variance induced by the algorithm itself. This lead us to formulate a deterministic version of the mutation operator, something very unusual in the Evolution Computation literature. However, the algorithm remains an stochastic algorithm whose source of randomness lies in the data points themselves. We also enforce a one-pass adaptation schedule of the application of the Evolution Strategy, that means that only one generation is computed from each real world time instant. Each time instant a sample of the data was considered. For Adaptive Color Quantization, we take a small sample of the image pixels, compute one generation of the Evolution Strategy and use the resulting population to color quantize the entire image. The optimal selection strategy is infeasible, due to its large computational cost. This has forced us to propose a greedy and a suboptimal selection operators. We tested first the suboptimal selection operator because of its linear complexity. The Evolution Strategy with the suboptimal selection operator and the deterministic mutation operator performed adaptively for almost all the cases. An anomaly appear for a large sample case. We tested the greedy selection operator combined with the deterministic mutation operator. The Evolution Strategy improved its response. Therefore, the suboptimality of the selection operator does influence more than we expected the response of the algorithm. Also, the deterministic mutation operator does not introduce biases that could produce degradations of the algorithm.

The focus for the future work will be on the search for alternative definitions of the selection operator with two goals in mind: (1) it must be computationally efficient, and (2) it must be as close as possible to the optimal selection. Although we are interested in formal convergence results, we will pursue the experimental work. The formal study of the convergence of this algorithm, as that of any stochastic algorithm, is far from trivial. Actual convergence studies consider always that the input is a sequence of independent and identically distributed random variables, which is the simplest case of a stationary process. Although these convergence analysis methods could be applied to the Evolution Strategy proposed in this paper, we believe that the proper analysis of the issues presented here must involve the analysis of the convergence of the stochastic algorithms whose input is a non-stationary process.

Our work shows also the general feasibility of the so called Michigan approach, which can be applied to a wide variety of problems, besides the original classifier systems. There is, however an unavoidable tradeoff of complexity. The Michigan approach simplifies the individuals, the global population fitness functions introduces complexity in the definition of the selection operator. The Pittsburg approach maps the whole problem into each individual, but the independence of fitness functions makes the definition of the selection operator trivial. Further work must be addressed to explore this tradeoff in a diversity of problems.

Acknowledgements:
This work is being supported by a research grant from the Dpto de Economía of the Excma. Diputación de Guipuzcoa, and a predoctoral grant and projects UE96/9, IT97/2 of the Dept. Educación, Univ. e Inv. of the Gobierno Vasco, and a project TAP98-0294-C02-02 of the CICYT.

References.

Aizawa A.N., Wah B.W. (1994) "Scheduling of genetic algorithms in a Noisy Environment". *Evolutionary Computation* 2(2), pp.97-122.

Alippi C., Cucchiara R. (1992) "Cluster partitioning in image analysis classification: a genetic algorithm approach". Proc IEEE COMPEURO, The Hague, pp.139-144.

Andrey P., Tarroux P. (1994) "Unsupervised image segmentation using a distributed genetic algorithm". *Pattern Recognition* 27(5), pp.659-673.

Babu G.P., Murty N.M. (1994) "Clustering with evolution strategies". *Pattern Recognition* 27(2), pp321-329.

Back T., Schwefel H.P. (1993) "An overview of Evolution Algorithms for parameter optimization". *Evolutionary Computation* 1, pp.1-24.

Back T., Schwefel H.P. (1996) "Evolutionary computation: an overview". *IEEE ICEC'96,* pp.20-29.

Beyer H.G. (1993) "Toward a theory of Evolution Strategies: Some asymptotical results from the $(1,^{+}\lambda)$ theory". *Evolutionary Computation* 1(2), pp.165-188.

Beyer H.G. (1994) "Toward a theory of Evolution Strategies: The (μ,λ) theory". *Evolutionary Computation* 2(4), pp.381-407.

Beyer H.G. (1995a) "Toward a theory of Evolution Strategies: On the benefits of sex-the $(\mu/\mu,\lambda)$ theory". *Evolutionary Computation* 3(1), pp.81-111.

Beyer H.G. (1995b) "Toward a theory of Evolution Strategies: Self-Adaptation". *Evolutionary Computation* 3(3), pp.311-347.

Bezdek J.C., Boggavaparu S, Hall L.O., Bensaid A. (1994) " Genetic Algorithm guided Clustering". *Proc 1st IEEE COnf. Evol. Comp.,* pp.34-39.

Bezdek J.C., Hathaway R.J. (1994) "Optimization of fuzzy clustering criteria using genetic algorithms". *Proc 1st IEEE Conf. Evol. Comp.,* pp589-594 . IEEE press.

Bhanu B., Lee S. (1994) "Genetic Learning for adaptive image segmentation". *Kluwer Acad. Pub.*

Bhuyan J.N., Raghavan V.V., Elayavalli V.K. (1991) "Genetic algorithm for clustering with an ordered representation". *Proc. 4th Int. Cong. Gen. Alg.,* pp.408-415.

Blekas K., Stafylopatis A. (1996) "Real coded genetic optimization of fuzzy clustering". EUFIT-96, Aachen, Germany

Buckles B.P., Petry F.E., Prabhu D., George R., Srikanth R. (1994) "Fuzzy clustering with genetic search". *Proc. 1st IEEE Conf. Evol. Comp.,* pp.46-50.

Chaudhuri B.B. (1994) "Dynamic clustering for time incremental data". *Paternn Recog. Let.* 15(1), pp27-34.

Chen H.H., Chen Y.S., Hsu W.H. (1995) "Low rate sequence image coding via vector quantization". *Signal Processing* 26, pp.265-283.

Chen O.T., Chen B.J., Zhang Z. (1994) "An adaptive vector quantization based on the gold-washing method for image compression". *IEEE trans cirsuits & systems for video techn.* 4(2), pp.143-156.

Chen R.J., Chien B.C. (1995) "Three dimensional morphological pyramid and its application to color image sequence coding".. *Signal Processing* 44, pp.163-180.

Diday E., Simon J.C. (1980) "Clustering Analysis". In K.S. Fu (ed) *Digital Pattern Recognition,* pp.47-94. Springer Verlag .

Duda R.D., Hart P.E. (1973) "Pattern Classification and Scene Analysis". Wiley .

Fitzpatrick J.M., Grefenstette J.J. (1988) "Genetic algorithms in Noisy Environments". *Machine Learning* 3, pp.101-120.

Fowler J.E. (1996) "A survey of adaptive vector quantization-Part I: A unifying structure". submitted to *IEEE transactions on Communications*.

Fukunaga K. (1990) "Statistical Pattern Recognition". *Academic Press* .

Garcia J.A., Fdez-Valdivia J., Cortijo F.J., Molina R. (1995) "A dynamic approach for clustering data". *Signal Processing* 44(2), pp.181-196.

Gersho A., Gray R.M. (1992) "Vector Quantization and signal compression". *Kluwer Acad. Pub.*

Goldberg D.F. (1989) "Genetic Algorithms in Search, Optimization & Machine Learning". *Addison Wesley.*

Gong Y., Zen H., Ohsawa Y., Sakauchi M. (1992) "A color video image quantization method with stable and efficient color selection capability". *Int. Conf. Pattern Recognition* vol 3, pp.33-36.

Gonzalez A.I., Graña M., Lozano J.A., Larrañaga P. (1998) "Experimental results of a Michigan-like Evolution Strategy for Non-Stationary Clustering". in G.D. Smith, N.C. Steele, R.F. Albrecht (eds) Artificial Neural Nets and Genetic Algorithms,. Springer Verlag, pp.555-559.

Hartigan J. (1975) "Clustering Algorithms". *Wiley* .

Heckbert P. (1980) "Color image quantization for frame-buffer display". *Computer Graphics* 16(3) ,pp.297-307 .

Holland J.H. (1975) "Adaptation in Natural and Artificial Systems". *The Univ. of Michigan Press* .

Jain A.K., Dubes R.C. (1988) "Algorithms for clustering data". *Prentice Hall* .

Jones D.R., Beltrano M.A. (1990) "Clustering with genetic algorithms". Operating Sciences Dept., Generla Motors Res. Lab., Warren, Mich..

Kankanhalli M.S., Mehtre B.M., Wu J.K. (1996) "Cluster based color matching for image retrieval". *Pattern Recognition* 29(4), pp.701-708.

Ketaff F.Z., Asselin de Beauville J.P. (1996) "Genetic and fuzzy based clustering". Proc. 5th Conf. Int. Fed. Class. Soc. (Kobe,Japan) , pp100-103.

Lin T.S., Chang L.W. (1995) "Fast color image quantization with error diffusion and morphological operations". *Signal Processing* 43, pp.293-303.

Lucasius C.B., Dane A.D., Kateman G. (1993) "On k-medoid clustering of large data sets with the aid of a genetic algorithm: background, feasibility and comparison". *Analytica Chimica Acta* 282, pp.647-669.

Luchian S., Luchian H., Petriuc M. (1994) "Evolutionary Automated Classification". *Proc 1st IEEE Conf. Evol. Comp.,* pp.585-588.

Michalewicz Z. (1996) "Evolutionary computation: practical issues". *IEEE ICEC'96.,* pp.30-39 .

Moraczewski I.R., Borkowski W., Kierzek A. (1995) " Clustering geobotanical data with the use of a genetic algorithm". *COENOSES* 10(1), pp.17-28.

Orchard M.T., Bouman C.A. (1991) "Color quantization of images". *IEEE trans. Signal Processing* 39(12), pp.2677-2690.

Ostermeier A., Gawelczyk A., Hansen N. (1995) "A derandomized approach to Self-Adaptation of Evolution Strategies". *Evolutionary Computation* 2(4), pp.369-380.

Uchiyama T., Arbib M.A. (1994) "Color image segmentation using competitive learning". *IEEE . Patt. Anal. and Mach. Int.* 16(12), pp.1197-1206.

Wu X. (1991) "Efficient Statistical Computations for Optimal Color Quantization". In J. Arvo (ed) Graphics Gems II *Academic Press Professional,* pp.126-133.

Yin G., Rudolph G., Schwefel H.P. (1995) "Analyzing the (1,λ) Evolution Strategy via stochastic approximation methods". *Evolutionary Computation* 3(4), pp.473-490.