A sensitivity analysis of the Self Organizing Map as an Adaptive One-pass Non-stationary Clustering algorithm: the case of Color Quantization of image sequences

A. I. Gonzalez[1], M. Graña[1], A. D´Anjou,[1] F.X. Albizuri[1], M. Cottrell[2]

[1]Dept. CCIA, UPV/EHU+ Apartado 649, 20080 San Sebastián
e-mail: ccpgrrom@si.ehu.es
[2]Institute SAMOS, Univ. Paris I

Abstract: In this paper we study the sensitivity of the Self Organizing Map to several parameters in the context of the one-pass adaptive computation of cluster representatives over non-stationary data. The paradigm of Non-stationary Clustering is represented by the problem of Color Quantization of image sequences.

0 Introduction

Cluster analysis and Vector Quantization have applications in signal processing, pattern recognition, machine learning and data analysis [1,2,3,4,5,6]. A vast number of approaches have been proposed to solve these problems, among them Competitive Neural Networks [7,8,9]. Conventional formulations assume that the underlying stochastic process is stationary and that a given set of sample vectors properly characterizes this process. Non-stationary processes are dealed with applying a predictive approach to reduce the non-stationary problems to the stationary framework [1]. This paper continues a line of work [10, 19] that consist in the exploration of the efficiency of competitive neural networks as one-pass adaptive algorithms for the computation of clustering representatives in the non-stationary case whithout knowledge of a time dependence model. This paper and [19] focus on the Self Organizing Map (SOM) [9]. The one-pass adaptation framework is not very common in the neural networks literature, in fact the only related recent reference that we have found is [18]. This restriction imposes very strong computational limitations. The effective scheduled sequences of the learning parameters applied to meet the fast adaptation requirement fall far from the theoretical conditions for convergence. A sensitivity analysis is justified in order to asses the behaviour of the SOM under a wide range of conditions and parameter values.

We have found that Color Quantization of image sequences is a privileged instance of the Non-stationary clustering problem. Nevertheless Color Quantization has a strong appealing by itself for his practical applications in visualisation [11,12,13], color image segmentation [14], data compression [15] and image retrieval [16]. In the context of Color Quantization of image sequences, one-pass adaptation is naturally enforced by the real time constraints of the processing of each image whithin the sequence.

Section 1 gives a review of the application of the SOM to the one-pass adaptive computation of cluster representatives in the general Non-stationary Clustering problem Section 2 discusses the experimental results obtained. Finally, section 3 gives some conclusions and further work

+ This work is being supported by the Dpto. de Economía of the Excma. Diputación de Guipuzcoa and the Dept. Educación, Univ. e Inv. of the Gobierno Vasco, under projects PI94-78, UE96/9, and a predoctoral grant for A.I. Gonzalez

1 Adaptive application of SOM to Non-stationary Clustering

Conventional formulations of Cluster analysis assume that the data $\aleph = \{\mathbf{x}_1,..,\mathbf{x}_n\}$ is a sample of an stationary stochastic process, whose statistical characteristics will not change in time. Our approach to Non-stationary Clustering assume that the data come from a non-stationary stochastic process that is sampled at diverse time instants. That is, the population can be modelled by a discrete time stochastic process $\{\mathbf{X}_t \;\; t = 0,1,..\}$ of unknown joint probability distribution. We do not assume any knowledge of the time dependencies that could allow a predictive approach [1].

A working definition of the time varying Clustering problem could read as follows: Given a sequence of samples $\aleph(t) = \{\mathbf{x}_1(t),..,\mathbf{x}_n(t)\}$ of the population obtain a corresponding sequence of partitions of each sample that consists of a sequence of sets of disjoint clusters $P(\aleph(t)) = \{\aleph_1(t),..,\aleph_c(t)\}$. This sequence of partitions minimizes a criterium function $C = \sum_{t\geq 0} C(t)$. In the general statement of the problem the difinition of the criterium function is based on the definition of an appropriate distance. We follow the conventional approach of using the Euclidean distance. We consider a sequence of representatives $\mathbf{Y}(t) = \{\mathbf{y}_1(t),..,\mathbf{y}_c(t)\}$ such that the desired partitions are defined by the nearest (Euclidean) representative.

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

The criterium function that we will consider at each time step is, therefore, the so-called distortion (or whithin cluster variance)

$$C(t) = \sum_{\mathrm{j}=1}^{\mathrm{n}} \sum_{i=1}^{c} \left\| \mathbf{x}_j(t) - \mathbf{y}_i(t) \right\|^2 \delta_i\left(\mathbf{x}_j(t), \mathbf{Y}(t)\right); \;\; \delta_i(\mathbf{x}, \mathbf{Y}) = \begin{cases} 1 & \mathrm{i} = \underset{k=1,..,c}{\operatorname{argmin}} \left\{ \left\| \mathbf{x} - \mathbf{y}_k \right\|^2 \right\} \\ 0 & \text{otherwise} \end{cases} \qquad (1)$$

The schema of the adaptive computation of the cluster representatives along time can be stated as follows: At time t take as initial cluster representatives the ones already computed from the sample of the process at time t-1. Use the sample vectors $\aleph(t) = \{\mathbf{x}_1(t),..,\mathbf{x}_n(t)\}$ to perform an adaptive computation leading to the new estimates of the cluster representatives. The adaptive computation proposed in this paper is the Self Organizing Map (SOM).

For notational simplicity, let us denote $\aleph = \{\mathbf{x}_1,..,\mathbf{x}_n\}$ the sample of the process at a given time instant. The SOMis a particular case of the general Competitive Neural Network algorithm: :

$$\mathbf{y}_i(\tau+1) = \mathbf{y}_i(\tau) + \alpha_i(\tau)\, \vartheta_i(\mathbf{x}(\tau), \mathbf{Y}(\tau)) \left[\mathbf{x}(\tau) - \mathbf{y}_i(\tau)\right] \;\; ; \;\; \mathbf{x}(\tau) \in \aleph; 1 \leq i \leq c \qquad (2)$$

where τ is the order of presentation of the sample vectors. One-pass adaptation means that each sample vector will only appear once in the sequence of presentations implied by (2). In their general statement, Competitive Neural Networks are designed to perform stochastic gradient minimisation of a distortion-like function. In order to guarantee theoretical convergence, the (local) learning rate $\alpha_i(\tau)$ must cope with the conditions:

$$\lim_{\tau \to \infty} \alpha(\tau) = 0, \; \sum_{\tau=0}^{\infty} \alpha(\tau) = \infty, \text{ and } \sum_{\tau=0}^{\infty} \alpha^2(\tau) < \infty \qquad (3)$$

However, these conditions imply very lengthy adaptation processes, which do not agree with the "one pass" adaptation constraint. In the experiments, the learning rate follows the expression :

$$\alpha_i(\tau) = 0.1\left(1 - \frac{\sum_{k=1}^{\tau}\delta_i(\mathbf{x}(k),\mathbf{Y}(k))}{n}\right) \qquad (4)$$

were $\delta_i(\mathbf{x},\mathbf{Y})$ follows the definition given in (1). This expression implies that the learning rate decreases proportionally to the number of times that a codevector "wins" the competition. The adaptation induced by the neighbouring function does not alter the local learning rate. The expression (4) also implies that a local learning rate only reaches the zero value if the codevector "wins" for all the sample vectors. This expression of the learning rate is the best we have found that fits in the one-pass adaptation framework. Obviously the sequences of the learning rate parameters given by (4) do not comply with the conditions (3) that ensure the theoretical convergence of the stochastic gradient minimization algorithm.

The function $\vartheta_i(\mathbf{x},\mathbf{Y})$ is the so-called *neighbouring function*. According to its definition, the shape of the minimized distortion function and, therefore, the qualitative properties of the learning rule equilibria can be very different. In the case of the SOM the neighbouring function is defined over the space of the neuron (cluster) indices. In our works we have assumed a 1D topology of the cluster indices. The neighbourhoods considered decay exponentially following the expression:

$$\vartheta_i(\mathbf{x}(\tau),\mathbf{Y}(\tau)) = \begin{cases} 1 & |w - i| \le \left\lceil (\upsilon_0 + 1)\,exp\left(\upsilon^{(0)}\tau\, log\left(1/(\upsilon_0+1)\right)/n\right)\right\rceil \\ 0 & otherwise \end{cases} ;1 \le i \le c \qquad (5)$$

$$w = \text{argmin}\left\{\|\mathbf{x}(\tau) - \mathbf{y}_k(\tau)\|^2 k = 1,..,c\right\}$$

The size of the sample considered at each time instant is n. The initial neighbourhood radius is υ_0. The expression ensures that the neighbouring function reduces to the simple competitive case (null neighbourhood) after the presentation of the first $1/\upsilon^{(0)}$ vectors of the sample. Along the experiments we have called $\upsilon^{(0)}$ the neighbourhood reduction rate. .

2 Experimental results on the Color Quantization of an image sequence

The sequence of images used for the computational experiments is a panning of the laboratory taken with an electronic camera. Original images have an spatial resolution of 480x640 pixels. Each two consecutive images overlap 50% of the scene. Figure 1 shows the distribution of the pixels in the RGB unit cube for each image in the sequence and gives an straight impression of the non-stationary nature of the data we are dealing with.

As a benchmark non adaptive algorithm we have used a variation of the algorithm proposed by Heckbert [11] as implemented in MATLAB. This algorithm has been applied to the entire images in the sequence under stationary and non-stationary assumptions. Figure 2 shows the distortion results of the Color Quantization of the experimental sequence to 16 and 256 colors based on both applications of the Heckbert algorithm. The curve $\{C^{TV}(t); t = 1,..,24\}$, named *Time Varying Min Var* in the figure, is produced assuming the non-stationary nature of the data and applying the algorithm to each image independently. The curve $\{C^{TI}(t); t = 1,..,24\}$, named *Time Invariant Min Var* in the figure, comes 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 another indication of the non stationarity of the data. Also this gap defines the response space left for truly adaptive algorithms. All the figures giving distortion results for the experimental sequence will include these two curves as a reference frame.

The adaptive application of the SOM assumes that the adaptation process starts with the second image, taking as initial cluster representatives the assumed color representatives for the first image. In the two first experiments the initial codebook was the Heckbert palette for the first image. The adaptation is performed over a random sample of the pixels of each image. In the experiments that foloow, we have tried to explore the sensitivity of the SOM to the following parameters: number of clusters (codebook size), size of the sample taken from each image, neighbouring function parameters: neighbourhood initial size and reduction rate, and, finally, the initial color representatives of the whole process (the assumed color representatives of the first image). The scheduling of the learning rate remains the same through all the experiments.

The first experiment tries to evaluate the sensitivity of the SOM to the sample size and the number of cluster representatives (codebook size) searched. Two codebook sizes have been considered 16 and 256 colors. The neighbouring function parameters were reasonably set to: $\upsilon_0 = 1$ and $\upsilon^{(0)} = 4$ for 16 colors, and $\upsilon_0 = 8$ and $\upsilon^{(0)} = 4$ for 256 colors. Figure 3 shows the results of the SOM for several sample sizes. These results consist of the sequence of distortions over the image sequence of the Color Quantization using the color representatives computed adaptively by the SOM over the image samples. The first general conclussion that can be drawn from this figure is that the SOM performs adaptively under a wide variety of conditions, but that it is clearly sensitive to the sample size. A closer inspection of the figure leads to the conclussion that the SOM is highly sensitive to the number of color representatives (clusters) searched. The sample sizes 100 for 16 colors and 1600 for 256 have the same ratio of sample size to codebook size (roughly 6:1). However, the response of the SOM in either case is qualitatively very different, it is clearly worse in the 256 colors codebook case. In the case of the the 16 color codebook, as the sample size grows, the distortion curves overlap very fast in near optimal results. In the case of 256 colors this convergence to near optimal results (as the sample size grows) is very smooth. The influence of the sample size seems to be stronger in the 256 colors codebook case. Finally, if we consider the highest sample:codebook ratio that appears in both figures (100:1), we note that the response in the 16 colors codebook case is qualitatively better than in the 256 colors codebook case. Our main conclussion from this first experiment is that the codebook size is the prime factor in the performance of the SOM. Once the codebook size is fixed, the size of the sample used for the one-pass adaptation can be a very sensitive performance factor.

The second experiment was intended to explore the sensitivity of the SOM to the neighbouring function parameters: the initial neighbourhood υ_0 and the neighbourhood reduction rate $\upsilon^{(0)}$. Not all the combinations of codebook and sample size tested in fig 3 are retried in this experiment. The measure of the behaviour of the color quantizers computed by the SOM is the accumulated distortion along the entire image sequence. This measure was computed from the samples instead of the entire images (the magnitudes of the errors can not be compared between surfaces). This simplification is justified because we are interested in the qualitative shape of the response surface, and because we have observed that the distortion of the color quantization of the entire image is proportional to that of the of the sample. The values of the neighbouring reduction factor tested were {1,2,3,4,5,8} and {1.25,1.3,1.5,2,3,4} in the case of 16 and 256 colors, respectively. The initial neighbourhoods considered were {1,2,3,4,5,8} and {2,8,16,32,64,128} in the case of 16 and 256 colors, respectively. Figures 4 and 5 show the results, and table 1 summarizes the experimental design. Shown in the figures are both the response surfaces (figs 4a,4c,4e,5a,5c) and the projections on the experiment axes (figs 4b,4d,4f,5b,5d).

Sample size

codebook		100	400	1600	4096	6400	12800	25600
16	surface	fig4a	fig4c	fig4e				
	project.	fig4b	fig 4d	fig 4f	--	--	---	--
256	surface			fig 5a				fig5c
	project.	--	--	fig 5b	--	--	--	fig 5d

Table 1. Summary of the neighbouring function sensitivity experiment results

The study of figures 4 and 5 confirm the previous assertion of the importance of codebook and sample size. The sensitivity of the SOM to the setting of the neighbouring function parameters varies strongly with them. In the case of the smaller sample:codebook ratio (6:1) (figs 4a, 5a) the response surface has a counter intuitive shape. It appears that for this ratio the best results are obtained with the smaller initial neighbourhoods. This result may be due to fluctuations produced during the reordering phase of the SOM by the combined effect of the sparse distribution of the small sample and the relatively big initial neighbourhood. For a more sensible ratio (100:1), whose results are shown in figs 4e and 5c, the response surface has a more natural shape giving the best results for the largest initial neighbourhood. The comparison of figs 4c and 4e confirms the quick convergence of the SOM to the optimal behaviour as the sample:codebook ratio grows, in the case of 16 colors. The examination in both figures 4 and 5 of the projections of the surfaces reveals a very clear trend for the neighbourhood reduction rate. In general, a reduction factor such that the neighbourhood disappears after presentation of one quarter of the sample gives the best results in all the cases. After codebook and sample size, the neighbouring reduction rate seems to be the next significant performance factor. With all the other performance factors set to appropriate values, the optimal values of the initial neighbourhood are the largest ones.

The last experiment conducted was the exploration of the sensitivity to the initial codebooks. As said before, the previous experiments were conducted starting the adaptive process in the second image of the sequence, assuming the initial codebook to be the Heckbert palette (Matlab) for the first image. In figures 6 and 7 it is shown the response of the SOM to other settings of the initial codebook: a threshold based selection of the sample of image #1 (Thresh), random points in the RGB cube (RGBbox) and a random selection of the sample of image #1 (Sample). For 16 colors the SOM parameters were: sample size 1600, $\upsilon_0 = 1$, and $\upsilon^{(0)} = 4$. For 256 colors sample size was 25600, $\upsilon_0 = 128$, and $\upsilon^{(0)} = 4$. Figures 6a,7a,7c show the distortion along the image sequence of experimental images together with the benchmark results. Let us denote $\left\{C^{SOM}(t); t = 1,..,24\right\}$ the sequence of distortion values obtained from the color quantizers computed by the SOM starting from a given initial codebook. Figures 6b,7b,7d show these sequences relative to the error committed when assuming stationarity, that is for each initial condition we plot:

$$\left\{\left(C^{SOM}(t) - C^{TV}(t)\right) \Big/ \left(C^{TI}(t) - C^{TV}(t)\right); t = 1,..,24\right\}$$

Figure 6 shows that the SOM is quite insensitive to initial conditions for small codebooks. However figs 7a,b show a rather high sensitivity to the initial codebook. The obvious hypothesis for this degradation is that our one-pass implementation of the SOM can not perform properly the self-organization phase when the codebook size is relatively large. To test this idea, we have applied a simple ordering by components to the codebooks before starting the adaptation with the SOM. The results are shown in figs 7c,d. Given a good ordering of the initial cluster representatives, the SOM becomes insensitive to initial conditions regardless of codebook size. We can conclude that the our one-pass SOM is capable of performing fast self-organization in the case of small codebooks, but as the size of the codebook grows it becomes very sensitive to the bad ordering of the initial cluster representatives. The strong influence of the network size (the number of clusters) extends to the ability of our one-pass SOM to recover from bad initial topological orderings of the neurons that incorporate the cluster representatives.

4 Conclusions and further work

This work has explored the sensitivity of the Self Organizing Map as a one-pass adaptive algorithm for the computation of cluster representatives in the framework of Non stationary Clustering problems. From an experimental point of view, we have found that the paradigm of Non-stationary Clustering is summarized in the problem of Color Quantization of image sequences. The experiments show that the SOM is a very robust algorithm for the one-pass adaptive computation of cluster representatives in the non-stationary case.The detailed sensitivity experiments reported here are motivated by the fact that the SOM has given the best results so far for this task. In the sensitivity experiments, we have found that the SOM is highly sensitive to the number of clusters searched, that is, to the size of the network to be adapted. The number of clusters searched impose restrictions on the size of the sample used. These two problem parameters condition the response of the SOM to changes in the neighbouring function parameters. Finally, when the SOM is quite insensitive to initial conditions, for small codebook sizes. For larger codebooks, our one-pass SOM is sensitive to the topological ordering of the initial cluster representatives. An extensive and comprehensive report on the application of several neural network and evolutionary approaches to one-pass adaptive Color Quantization of image sequences is on the way.

References

[1] A. Gersho, R.M. Gray "Vector Quantization and signal compression" Kluwer 1992
[2] J. Hartigan "Clustering Algorithms" Wiley 1975
[3] E. Diday, J.C. Simon, "Clustering Analysis" In K.S. Fu (ed) Digital Pattern Recognition pp.47-94 Springer Verlag 1980
[4] R.D. Duda, P.E. Hart "Pattern Classification and Scene Analysis", Wiley 1973
[5] A. K. Jain, R.C. Dubes "Algorithms for Clustering data" Prentice Hall 1988
[6] K. Fukunaga "Statistical Pattern Recognition" Academic Press 1990
[7] J. Mao, A.K. Jain "A Self-Organizing network for hyperellipsoidal Clustering" IEEE trans. Neural Networks 7(1) pp.16-29 1996
[8] Ahalt S.C., A.K. Krishnamurthy, P. Chen, D.E. Melton "Competitive Learning Algorithms for Vector Quantization Neural Networks" 3 pp.277-290 1990
[9] T. Kohonen "Self Organization and Associative memory" Springer Verlag 1989
[10] A.I. Gonzalez, M. Graña, A. d'Anjou, M. Cottrell "On the application of Competitive Neural Networks to Time-varying Clustering problems" Spatiotemporal Models in Biological and Artificial Systems F.L Silva, J. Principe, L.B. Almeida (eds) IOS press(1996) pp.49-55
[11] P. Heckbert "Color image quantization for frame-buffer display" Computer Graphics 16(3) pp.297-307 1980
[12] M.T. Orchard, C.A. Bouman "Color quantization of images" IEEE trans. Signal Processing 39(12) pp.2677-2690 1991
[13] T.S. Lin, L.W. Chang "Fast color image quantization with error diffusion and morphological operations" Signal Processing 43 pp.293-303 1995
[14] T. Uchiyama, M.A. Arbib "Color image segmentation using competitive learning" IEEE trans. Pattern Analysis and Machine Intelligence 16(12) pp.1197-1206 1994
[15] Y. Gong, H. Zen, Y. Ohsawa, M. Sakauchi "A color video image quantization method with stable and efficient color selection capability" Int. Conf. Pattern Recognition 1992 vol3 pp.33-36
[16] M.S. Kankanhalli, B.M. Mehtre, J.K. Wu "Cluster based color matching for image retrieval" Pattern Recognition 29(4) pp.701-708 1996
[17] O.T. Chen, B.J. Chen, Z. Zhang "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 1994
[18] C. Chan, M. Vetterli, "Lossy Compression of Individual Signals Based on String Matching and One Pass Codebook Design" ICASSP'95, Detroit, MI, 1995
[19] A.I. Gonzalez, M. Graña, A. d'Anjou, F.X. Albizuri, M. Cottrell " Self Organizing Map for Adaptive Non-stationary Clustering: some experimental results on Color Quantization of image sequences" submitted to ESSAN97

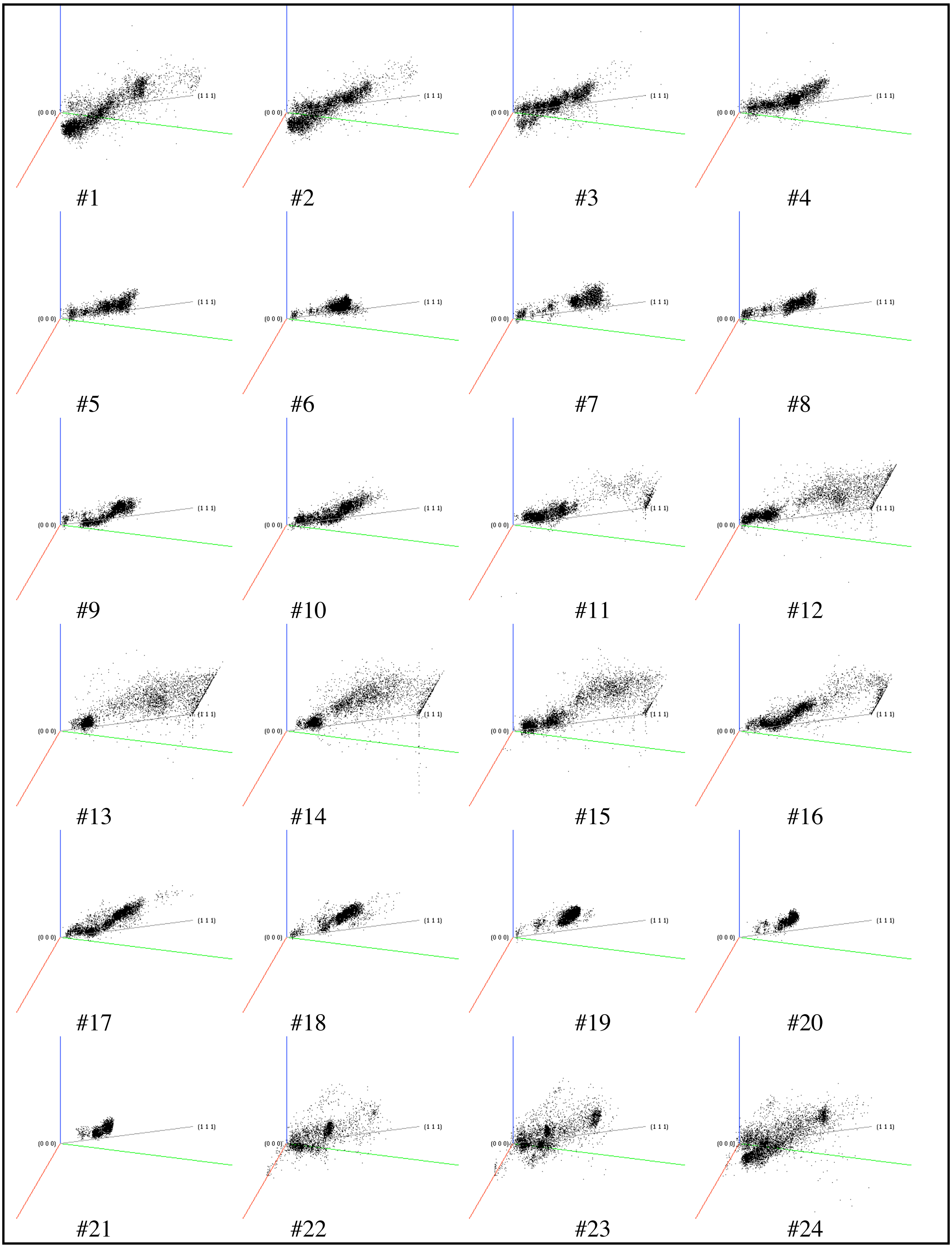
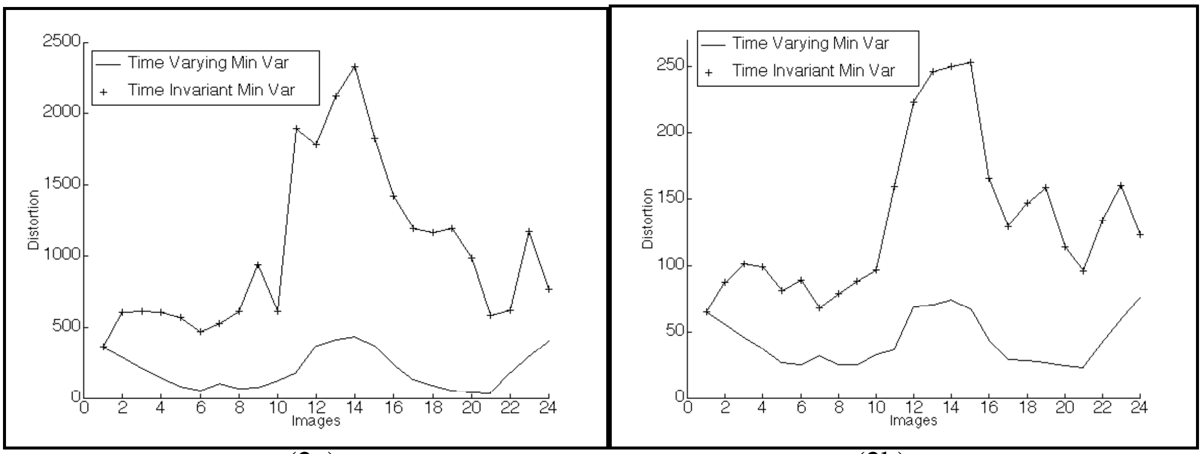
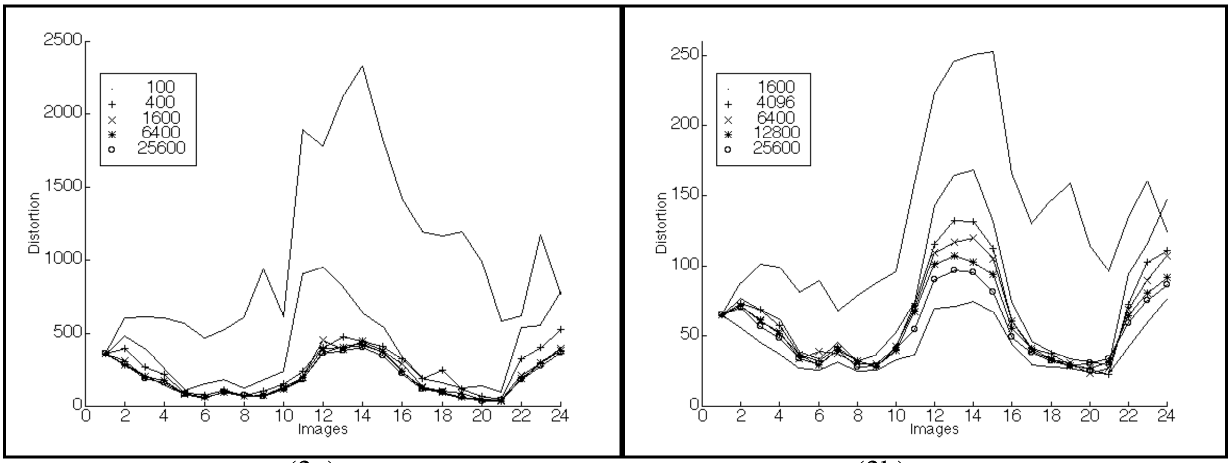


Figure 1 Distribution of pixel color for the images in experimental sequence



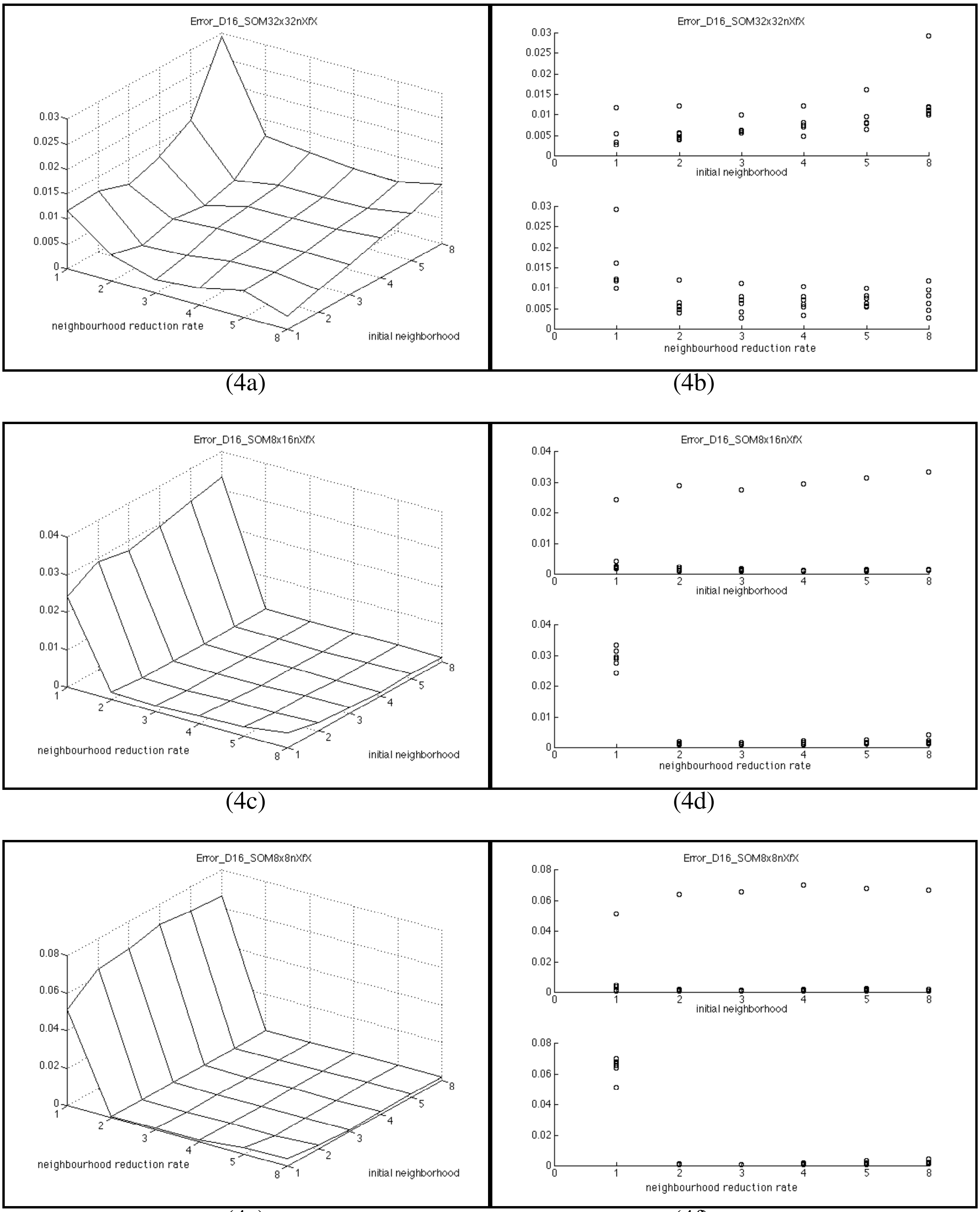
(2a) (2b)

Figure 2. Benchmark distortion values obtained with the application of the Matlab implementation of the Heckbert algorithm to compute the color quantizers of 16 (2a) and 256 (2b) colors of the images in the experimental sequence.



(3a) (3b)

Figure 3 Distortion results obtained with the adaptive application of SOM over samples of diverse sizes to compute the color quantizers of (3a) 16 (with $\upsilon_0 = 1$ and $\upsilon^{(0)} = 4$) and (3b) 256 colors (with $\upsilon_0 = 8$ and $\upsilon^{(0)} = 4$)



(4a) (4b)

(4c) (4d)

(4e) (4f)

Figure 4- Sensitivity to the neighbouring function parameters υ_0 and $\upsilon^{(0)}$ of the SOM applied to compute the color quantizers of 16 colors. (see table 1), measured by the accumulated distortion along the experimental sequence

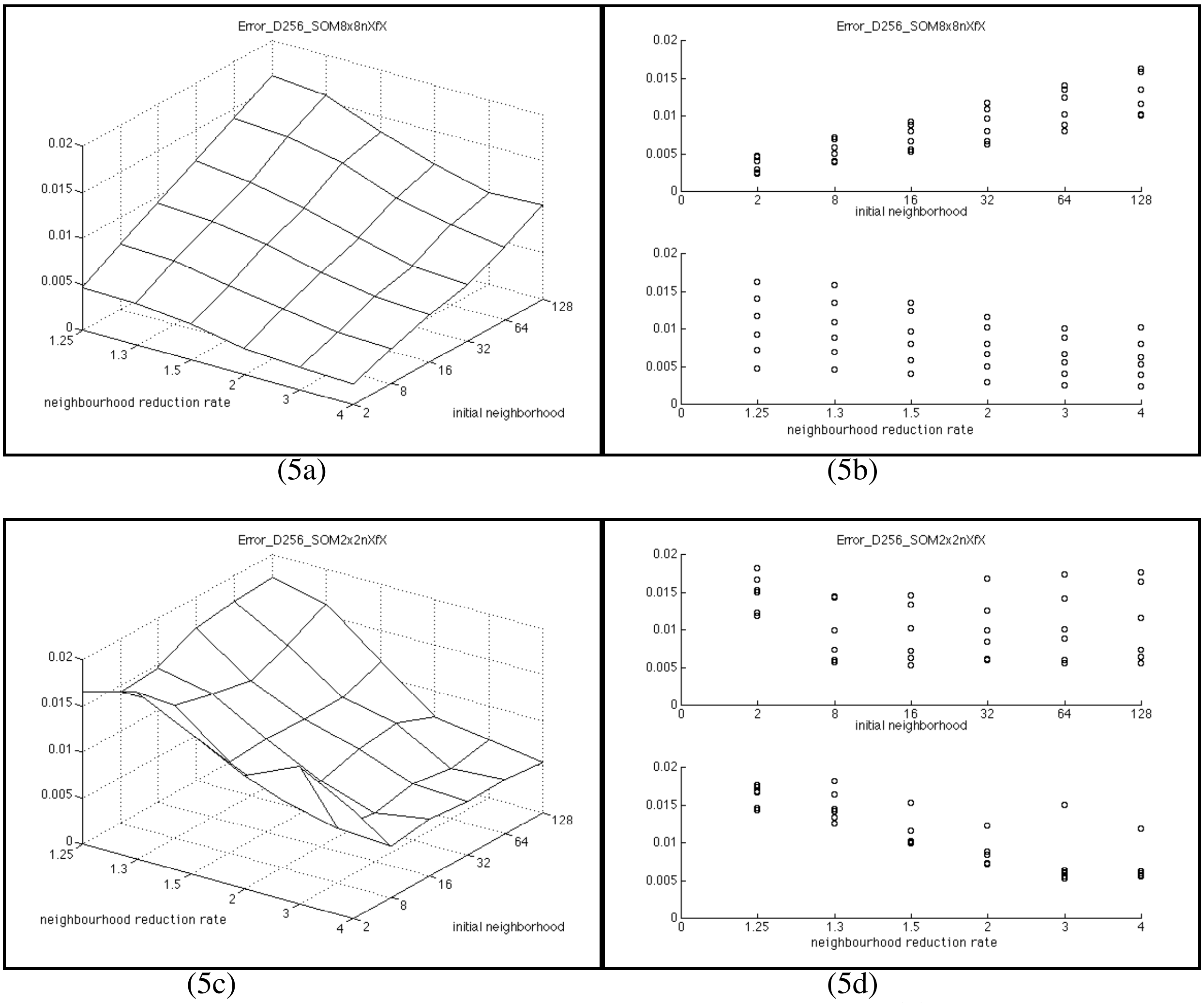
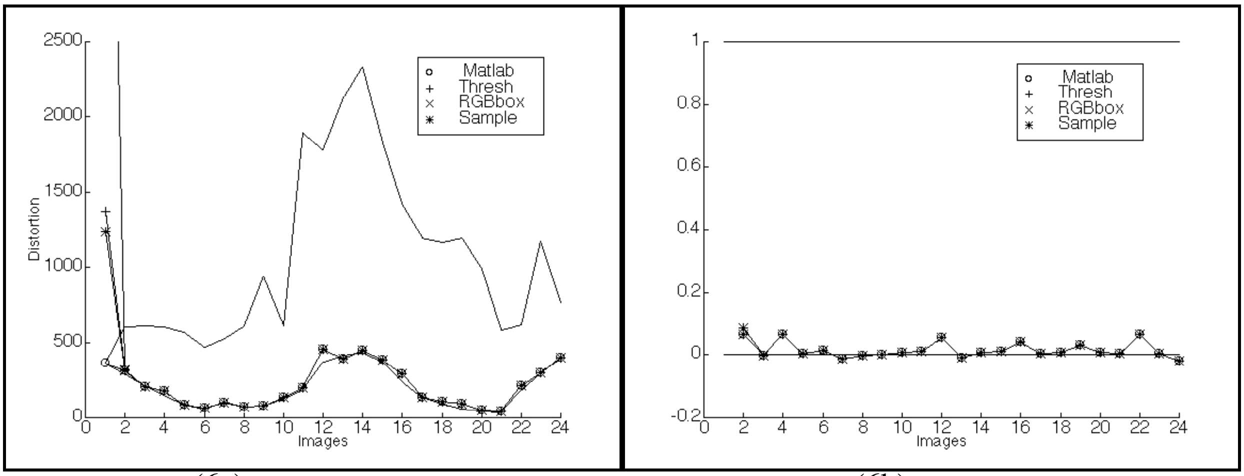
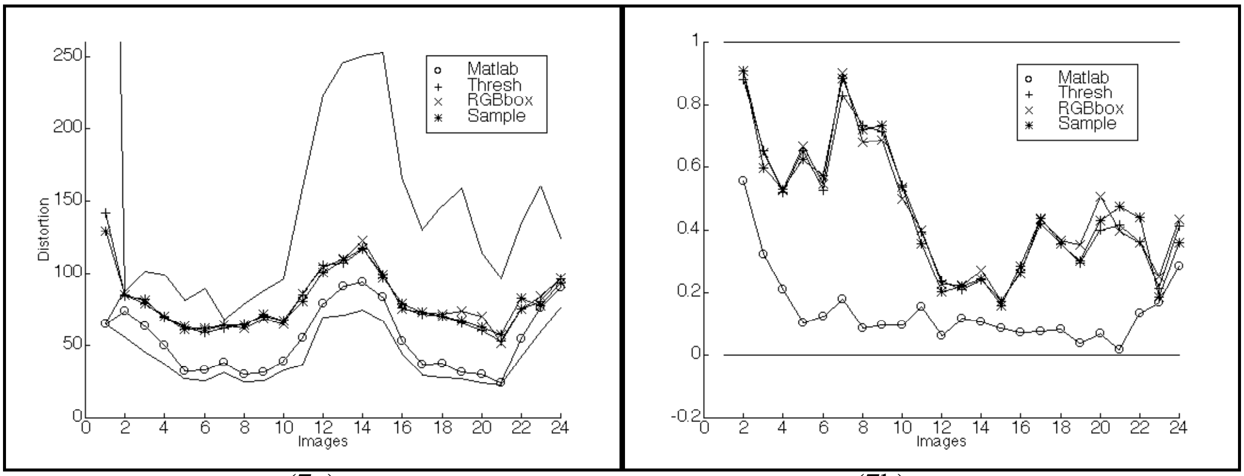


Figure 5- Sensitivity to the neighbouring function parameters υ_0 and $\upsilon^{(0)}$ of the SOM applied to compute the color quantizers of 256 colors. (see table 1), measured by the accumulated distortion along the experimental sequence

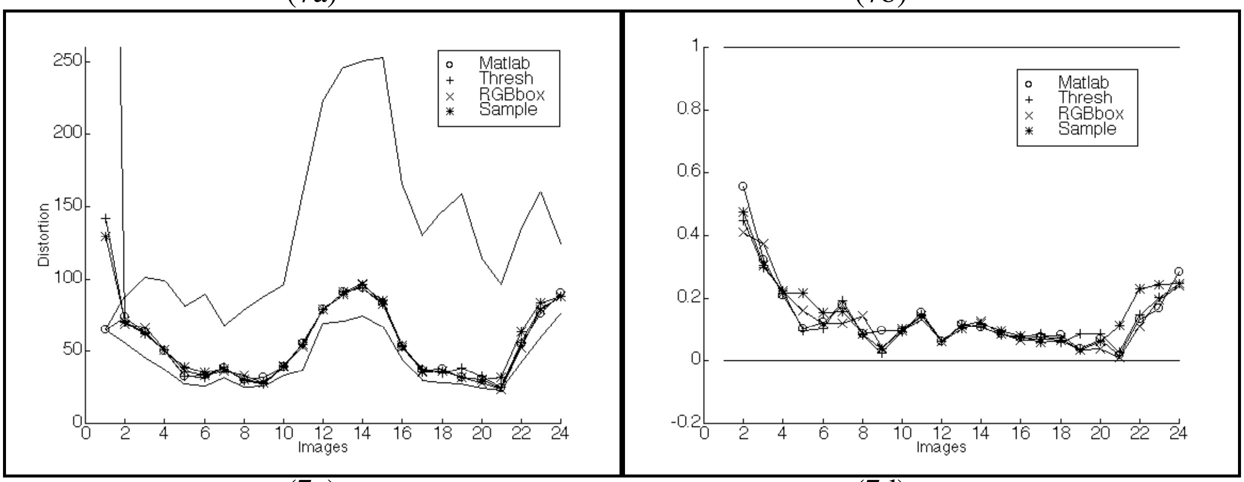


(6a) (6b)

Figure 6 Distortion of the color quantizers of 16 colors computed by the adaptive application of the SOM starting from several intial cluster representatives (sensitivity to initial conditions) 6a absolute values, 6b normalized relative to the stationarity assumption.



(7a) (7b)



(7c) (7d)

Figure 7 Distortion of the color quantizers of 256 colors computed by the adaptive application of the SOM starting from several intial cluster representatives (sensitivity to initial conditions) 7a,b unprocessed initial cluster representatives, 7c,d the same initial cluster representatives ordered before starting the adaptation process