Ecole Polytechnique, Promotion X 98

Rapport de stage d'option scientifique

Benoit Labrousse **Jérôme Le Ny**

A Model For Video Sequences

Option: Image et Signal

Département de Mathématiques Appliquées

Directeur de l'option: M. Stéphane Mallat

Organisme d'accueil: University of Central Florida
Department of Computer Science. Orlando, FL USA

Directeur de Stage: Dr. Erol Gelenbe

Dates du stage: 6 avril - 8 juillet 2001

Abstract

In this paper, we examine how we can use probabilities to model video sequences. First we develop a spatio-temporal model, showing the relationship between the gray-level of a pixel and its speed. We characterize the state of a 8x8 pixels-block by the distribution of its successive speed vectors. We then suppose that for a given block, the evolution of such states is governed by a semi-Markovian process. When in a particular state, the speed is assumed to follow a wide-stationary distribution of which we compute both the mean and the auto-covariance. The practical interest of this model stems from the theoretical evaluation of the quality of a compression scheme and the comparison between different schemes.

Résumé

Dans ce rapport, nous examinons comment l'usage des probabilités permet de modéliser le comportement de séquences vidéo. Dans un premier temps nous développons le modèle spatio-temporel, en montrant quelles sont les relations entre les valeurs des niveaux de gris d'un pixel et sa vitesse. Puis nous caractérisons l'état d'un bloc de 8x8 pixels par la distribution de ses vecteur vitesse successifs. Nous supposons ensuite que pour un bloc donné, l'évolutoin de cet état est gouvernée par un processus semi-Markovien. Quand un bloc est dans un état particulier, sa vitesse évolue selon une distribution de probabilité stationnaire dont nous évaluons la moyenne et l'autocovariance. L'intérêt pratique d'un tel modèle est d'évaluer de manière théorique la qualité d'un schéma de compression et éventuellement de comparer différents algorithmes.

Contents

Introduction

Video compression is one of the most rapidly developping field in signal processing because the transmission of video with a good quality through the different networks is one major challenge for the near future. Thus, the number of different compression schemes is also growing very rapidly, involving the last techniques and theories available, like wavelets or neural networks.

Considering all those different works, it could be interesting to characterize the efficiency of a compression scheme through a set of general parameters, and not only through the computation based on a few particular examples. This is the first motivation for the development of a model describing a video sequence in general: the aim is to determine some parameters describing a film, on which we will generally be able to compute the effect of the compression scheme. Thus ideally it would permit to compare theoretically different compression algorithms, and to estimate the domains of efficiency of each algorithm.

Very few work has been done in modelling the data structure represented by a film. Some authors ([1],[2]) began to work in this direction in the 60's and 70's, but later there is no continuation of such studies. Due to the complexity and the variety of video sequences, we had necessarily to make some simplifying assumptions in this work. Here we will consider most of the time sequences with only few scene changes or no scene changes, with a quite simple behavior. A good example for such a sequence could be a video conference.

In this project, we discuss a possible model for the temporal evolution of the pixel colors through successive frames. It is obvious that there exist a temporal and a spatial dependency between pixels of the same image and of successive frames, and this led to the principle of the MPEG algorithm. We try to incorporate these dependancies to a stochastic model, where we consider a film, the color of a certain pixel, the motion of a point through successive frames, as random processes. The first part recalls some important ideas about stochastic processes that we will use later. Then we present the characteristics of our model, which gives to the motion process through successive frames the most important role. Finally, we apply these ideas to the computation of the theoretical quality of some simple compression schemes, and compare it to the real results with some sample sequences.

Chapter 1

Stochastic Processes

In the following, we will consider either discrete time or continuous time processes, i.e. sequences of random variables $\{X(t)\}_{\{t\in\mathbb{Z}\}}$ or $\{X(t)\}_{\{t\in\mathbb{R}\}}$ defined on a probability space $(\Omega, \mathcal{A}, P)$. A process is associated with a probability distribution $P(X(t_1), \ldots, X(t_k))$ for every subset of k random variables. Although we will usually consider discrete time processes, corresponding to the succesive frames of a video sequences, the use of continuous time will be more convenient in some explanations.

For example, a process $\{X(t)\}_{\{t\in\mathbb{Z}\}}$ can represent the value of a pixel at a fixed position in sucessive images of a sequence, considering the different discrete values among grey levels. Thus we will consider video sequences as realizations of such a process.

In order to have usable parameters to describe video, we first need to introduce some stochastic tools which will serve later.

1.1 Ergodic markov chains

Let us first recall the ergodic theorem for markov chains, which will be useful in the following. For the proof of this theorem and more detailed descriptions of markov chains, one can refer to [3].

If E is the state space, that we suppose to be finite, we say that $\{X(t)\}_{\{t\in\mathbb{N}\}}$ is a discrete time markov chain if it has the following property:

$$P(X(t+1) = x | X(0) = x_0, \ldots, X(t) = x_t) = P(X(t+1) = x | X(t) = x_t)$$
$$\forall t \in \mathbb{N} \quad \text{and} \quad x_0, \ldots, x_t, x \in E.$$

An homogeneous Markov chain is caracterized by a transition matrix Q and an initial distribution π_0:

$$Q(x_t, y) = P(X(t+1) = y | X(t) = x_t)$$

$$\pi_0(x) = P(X(0) = x)$$

Thus, $\forall t \in \mathbb{N}$, we have:

$$P(X(0) = x_0, X(1) = x_1, \dots, X(t) = x_t) = \pi_0(x_0)Q(x_0, x_1) \dots Q(x_{t-1}, x_t)$$

If we define $T_y = \min(t \geq 1, X(t) = y)$ and $P_x = P(.|X(0) = x)$, we will say that a markovian matrix Q is irreducible if for all $x \neq y$ $P_x(T_y < \infty) \neq 0$. That is to say, for every initial state x, the chain has a non zero probability to visit y.

Moreover, we know that if E is finite, if Q is an irreducible matrix, it is also recurrent, i.e. for every initial state x, the probability for the chain to visit any state an infinite number of times is equal to 1.

We can now give the ergodic theorem; here we will note π_t the distribution of the considered chain at time t, with π_0 the initial distribution.

Theorem 1.1 (Ergodic theorem)
Let Q be an irreducible, recurrent matrix on the state space E. Then for all x,y, the frequency of the visits to state y of the chain with transition matrix Q and initial state x has a limit which is independant of x:

$$\frac{1}{t}\sum_{s=1}^{t} 1_{X_s=y} \xrightarrow[t\to\infty]{} \pi(y) \qquad P_x \;\; a.s. \tag{1.1}$$

The limits $\pi(y)$ $(y \in E)$ represent the unique stationnary probability on E, i.e:

$$\sum_y \pi(y)Q(y,z) = \pi(z) \quad (z \in E) \tag{1.2}$$

if such a probability exists (otherwise $\pi(y) = 0, \forall y \in E$). In this case, we say that the markov chain is ergodic. Then, if the chain is not periodic, we have:

$$\lim_{t\to\infty} \pi_t(y) = \pi(y) \tag{1.3}$$

1.2 Wide-sense stationary processes

Consider a process represented by an ergodic markov chain, with the stationnary probability as initial distribution, then its distribution does not change with a shift in time. We say that the process is stricly stationary if for all $p \in \mathbb{Z}$ and all subsets of k random variables, the distributions of $\{X(t_1), \dots, X(t_k)\}$ and $\{X(t_1 + p), \dots, X(t_k + p)\}$ are the same. This implies that the mean is independent of time:

$$\mu(t) = E\{X(t)\} = \mu \tag{1.4}$$

and the autocovariance (we always consider second-order stochastic processes) depends only on the difference in time:

$$Cov(X(t), X(s)) = E\{(X(t) - \mu(t))(X(s) - \mu(s))\} = R_X(t,s) = R_X(t-s) \tag{1.5}$$

In the following, we will observe processes which are not necessarily strictly stationary, and for which we usually just want to estimate the second order moments. Then we can simply suppose that the properties (1.4) and (1.5) are verified and moreover if the variance of the process is finite (second-order process), i.e.

$$E\{|X(t)|^2\} \leq \infty \tag{1.6}$$

we will say that the process is wide-sense stationary. Note that we will sometime omit the mean in the autocovariance expression; because it is independent of time, it is in fact equivalent to considering the zero-mean process $X(t) - \mu$.

1.3 Locally stationary processes

To assume that the process which represents a video signal is strictly or wide-sense stationary is too strong an assumption, even when we consider homogeneous sequences such as video conferences, without important changes between the successive frames. In fact we can just suppose that the stationarity of the processes is verified for short time intervals, which is an intuitive representation of locally stationary processes. In the following, we refer to [4] for the definition of locally stationary processes and for the estimation of covariance of such processes.

We first express the covariance $R(t,s)$ from the distance between t and s around a mid-point position:

$$R(t,s) = R_1(\frac{t+s}{2}, t-s) \tag{1.7}$$

When the process is wide-sense stationary, we know that:

$$R_1(\frac{t+s}{2}, t-s) = R_1(t-s)$$

For a locally stationary process, we have

$$\forall x \in \mathbb{R}, \exists l(x) : \forall \frac{t+s}{2} \in \left[x - \frac{l(x)}{2}, x + \frac{l(x)}{2}\right] : \qquad R_1(\frac{t+s}{2}, t-s) \approx R_1(x, t-s)$$

Moreover, we expect that if t and s are far apart and do not belong to the same interval of stationarity, $X(t)$ and $X(s)$ will be uncorrelated:

$\forall u \in \left[x - \frac{l(x)}{2}, x + \frac{l(x)}{2}\right]$, if $|v| = |t-s| > \frac{l(x)}{2}$:

$$E\{X(u+\frac{v}{2})X(u-\frac{v}{2})\} = R_1(u,v) \approx 0$$

1.4 Countable State Space Semi-Markov Processes [6]

To characterize precisely the model of video sequences which we will developp, in particular the periods where we want to consider the process to be stationary, we will need to use semi-Markov processes. We describe here a few characteristics of semi-Markov Processes on a countable state space, with continuous time (for a film we will then consider the discrete times of successive frames). For a complete presentation of markov and semi-markov processes, one can refer to [6] and [7].

Let $X = (X(t), t \in \mathbb{R}_+)$ be a stochastic process with countable state space E. In the following, we will only consider right continuous processes, i.e.:

$$\forall \omega \in \Omega, \forall t \in \mathbb{R}_+, \exists \delta = \delta(t, \omega) : X(t+h) = X(t) \ for \ 0 \leq h < \delta$$

Under this condition the process is said to be a jump process, and the sample paths are step functions.

A semi-Markov process is one that changes in accordance with a Markov chain but takes a random amount of time between changes. More specifically, consider a stochastic process with states $0, 1, \ldots$, which is such that, whenever it enters state $i, i \geq 0$:

1. The next state it will enter is state j with probability $P_{ij}, i, j \geq 0$.
2. Given that the next state to be entered is state j, the time until the transition from i to j occurs has distribution F_{ij}.

If we let $Z(t)$ denote the state at time t, then $\{Z(t), t \geq 0\}$ is called a *semi-Markov process.* Note that in our work, we will consider F_{ij} to be independant of the next state j, and simply write F_i. But we keep the formal notation for this presentation.

Let H_i denote the distribution of time that the semi-Markov Process spends in state i before making a transition.

$$H_i(t) = \sum_j P_{ij} Fij(t)$$

and let ν_i denote its mean. That is,

$$\nu_i = \int_0^\infty x \, dH_i(x)$$

If we let J_n denote the nth state visited, then $\{J_n, n \geq 0\}$ is a Markov chain with transition probabilities P_{ij}. It is called the *embedded Markov chain* of the semi-Markov process. We say that the semi-Markov process is irreducible if the embedded Markov chaine is irreducible as well. We also denote S_n the jump times, and T_n the times spent in the successive states, with the relation:

$$T_n = S_n - S_{n-1}, \ n \geq 1$$

We have $P(T_n \leq t | J_n = i) = H_i(t)$, and $Z(t) = J_n, \forall S_n \leq t < S_{n+1}$.

Let T_{ii} denote the time between successive transitions into state i and let $\nu_{ii} = E[T_{ii}]$. We also need the following definition: a random variable X is said to have a lattice distribution if there exists $d \geq 0$ such that $\sum_{n=0}^{\infty} P(X = nd) = 1$. The largest d having this property is said to be the period of X. We now give some results which will be important to compute probabilities on semi-Markov processes.

If the semi-Markov process is irreducible and if T_{ii} has a nonlattice distribution with finite mean, then:

$$P_i = \lim_{t \to \infty} P(Z(t) = i | Z(0) = j)$$

exists and is independent of the initial state. Furthermore:

$$P_i = \frac{\nu_i}{\nu_{ii}} = \lim_{t \to \infty} \frac{\text{amount of time in i during } [0, t]}{t} \tag{1.8}$$

That is, $\frac{\nu_i}{\nu_{ii}}$ equals the long-run proportion of time in state i.

Now suppose that the embedded Markov chain $J_n, n \geq 0$ is ergodic and let its stationary probability be $\pi_j, j \geq 0$.

$$\pi_j = \sum_i \pi_i P_{ij}, \quad \sum_j \pi_j = 1$$

Then with the conditions of the precedent equation, we have also:

$$P_i = \frac{\pi_i \nu_i}{\sum_j \pi_j \nu_j} \tag{1.9}$$

1.5 Estimating the mean and the covariance

In our work, we will have to extract information from the processes being considered, especially the mean and covariance. Considering a specific video sequence as a realization of a stochastic process has then a major drawback: we must estimate these parameters from only one realization of the process. In these conditions, estimating the mean and autocovariance remains simple when we consider the process as wide-sense stationary, but becomes difficult as soon as it is not stationary any more. In [4], Mallat, Papanicolaou and Zhang study the properties of locally stationary processes and give an algorithm to estimate the covariance of such processes.

1.5.1 Wide-sense stationary processes [10]

Mean

Let us consider a wide-sense stationnary process $(X(t), t \in \mathbb{N})$, of which we can only observe one realization. With N samples in this realization, we can have an estimator of the mean using:

$$\tilde{\mu} = \frac{1}{N}\sum_{n=0}^{N-1} X(n)$$

We say that the processus is ergodic for the mean when we have:

$$\lim_{N\to+\infty} E\{(\mu-\tilde{\mu})^2\} = \lim_{N\to+\infty} \frac{1}{N}\sum_{l=-N+1}^{N-1}(1-\frac{|l|}{N})R_X(l) = 0$$

$\tilde{\mu}$ is a good estimator of μ if the process is ergodic for the mean.

Autocovariance

For all k, the autocovariance $R_X(k)$ can also be estimated from the values of one realization from the following empirical mean:

$$\tilde{R}_X(k) = \frac{1}{N}\sum_{n=0}^{N-1-|k|}(X(n)-\tilde{\mu})(X(n+|k|)-\tilde{\mu})$$

1.5.2 General processes

If the process is not stationary, the mean will in general depend on time and the autocovariance cannot be written under the form (1.5). In this case the previous estimators cannot be used.

We will now consider processes that are *locally stationary*, so that on each period of stationarity the mean and covariance can be approximate with only one realization of the process. Intuitively speaking, a locally stationary process has subsequences which are stationary, and whose length may also be random. Estimating the covariance of locally stationary processes has been studied by Mallat, Papanicolaou and Zhang in [4] and they provide an estimation algorithm which can be used even for one realization, based on the search of a good basis approaching as close as possible to the Karhunen-Loève basis which diagonalizes the covariance matrix. Let us briefly summarize their approach.

Covariance Estimation of locally stationary processes with best basis search [4]

Covariance operator Define the *covariance operator* K for any $f \in L^2(\mathbb{R})$ by

$$Kf(n) = \sum_{m=-\infty}^{+\infty} R(m,n)f(m) \tag{1.10}$$

with R(m,n) the autocovariance of the process X. We consider the inner product

$$< f, X >= \sum_{n=-\infty}^{+\infty} f(n)X(n)$$

which is a linear combination of the process values at different times. K is a hermitian and positive operator. Also we have

$$E\{< f, X >< g, X >^*\} =< Kf, g >$$

We will denote by the term Karhunen-Loève basis an orthonormal basis which diagonalizes the covariance operator. In particular, the diagonal coefficients in a basis g_m of the matrix representing the covariance operator are given by

$$E\{| < X, g_m > |^2\} =< Kg_m, g_m >$$

Covariance estimation For a wide-sense stationary process R only depends on the difference of position n, and therefore we know that the covariance operator can be diagonalized by the complex exponentials

$$e_\omega[n] = e^{-in\omega}$$

In this case, the spectrum (the eigenvalues are called the *spectral power*) is given by the Fourier transform of $R(n)$. This spectrum characterizes completely the autocovariance via the inverse Fourier transform.

Considering for example a circulary stationary signal X of size N, we have the discrete Fourier basis wich diagonalizes its covariance operator:

$$\left\{ g_m[n] = \frac{1}{\sqrt{N}} exp(\frac{i2\pi mn}{N}) \right\}_{0 \le m < N}$$

The spectral power is given by:

$$\hat{R}(m) = \sum_{l=0}^{N-1} R(l) exp(\frac{-i2\pi ml}{N}) = E\{| < X, g_m > |^2\}$$

and is estimated with only one realization by computing the periodogram:

$$| < X, g_m > |^2 = \frac{1}{N} \left| \sum_{n=0}^{N-1} X(n) exp(\frac{-i2\pi mn}{N}) \right|^2$$

and we can deduce $R(n)$ from $\hat{R}(m)$ by inverse Fourier transform.
For locally stationary processes, we do not know a priori the Karhunen-Loève basis. But the idea is to search for a basis which approximate the Karhunen-Loève basis among a dictionnary of basis. Because a locally stationary process can be approximated by a stationary process on short periods of time, we search for a basis which almost diagonalizes this process among local cosine basis on these time periods. Because we do not know the intervals of stationarity, the dictionnary among which we search for the basis is composed of vectors of local cosine multiplied by windows of different sizes (in practice, these vectors are sorted into a tree, and the supports are intervals of size 2^{-j}).

Mallat, Papanicolaou and Zhang give then an algorithm which searches for the best basis by minimizing a negative quadratic sum. They prove that we thereby obtain an approximation of the spectral power, from which we can deduce the covariance matrix.

Chapter 2

Spatio-temporal model for video sequences

2.1 Assumptions and notations

Let Z be the whole pixel plane. We denote by $v(t, z)$ the value of pixel at time t. Let $U_k = \{u = (x, y) : (x, y) \in \{-k, \ldots, 0, \ldots, k\}^2\}$. In the following, we will usually assume that the number k has been chosen and simply note U. Then $N(z) = \{z + u, u \in U\}$ is the set of immediate neighbors of z, while $z+u(d)$ is an element of $N(z)$ in direction d. By convention, when $d = 0$ we will have $u(0) = (0, 0)$.We propose to consider $v_z = \{v(t, z), t \geq 0\}$ as a stochastic process called the *pixel process*, and determined by an underlying Markov chain $r_z = \{r(t, z), t \geq 0\}$ which takes its values in the set U and is defined by the following transition probabilities :

$$P[r(t + 1, z) = u'|r(t, z) = u] = p(u, u', z) \tag{2.1}$$

and for each z, u we have

$$\sum_{u' \in U} p(u, u', z) = 1 \tag{2.2}$$

We call r_z *the direction process* for each pixel position z, and we use it to determine the "direction of motion" in the video sequence. Note that this process depends on the pixel position z, so that the process may be different in different areas of an image.

We will assume that r_z is ergodic and that it has a stationary distribution that we denote by $p(u, z)$ which satisfies :

$$p(u, z) = \sum_{u' \in U} p(u', u, z)p(u', z) \tag{2.3}$$

Then we consider the direction process as strictly stationary.

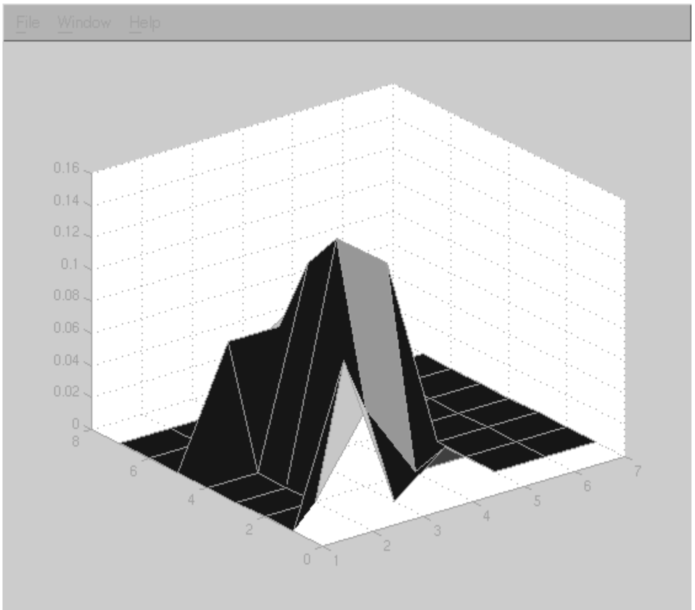


Fig 2.1.1: $p(u,z)$ for $u \in [-3,3]^2$, in a moving area

The *instantaneous average direction* is defined as

$$d(u,z) = \sum_{u' \in U} u'.p(u,u',z) \tag{2.4}$$

which tells us how the direction will become in one time step on the average at some point z, if the current direction is u. Similarly, the *stationary average direction* at point z is

$$d(z) = \sum_{u \in U} d(u,z).p(u,z) = \sum_{u \in U} u.p(u,z) \tag{2.5}$$

If we note $E(X)$ the mean of a random variable X, we have:

$$d(z) = E(r(t,z))$$

The mean does not depend on time, because we supposed the direction process stationary. We show the average direction (figure 2.1.3) for the sequence taxi (figure 2.1.4).

2.2 Validity of the model

It is quite obvious that we can't assume the direction process to be stationary for a whole video sequence. First of all, we will only consider homogeneous

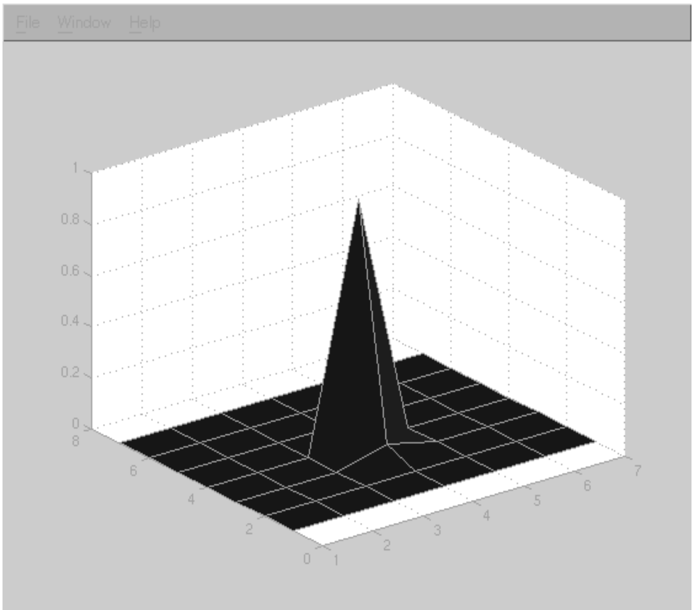


Fig 2.1.2: $p(u, z)$ for $u \in [-3, 3]^2$, in a still area

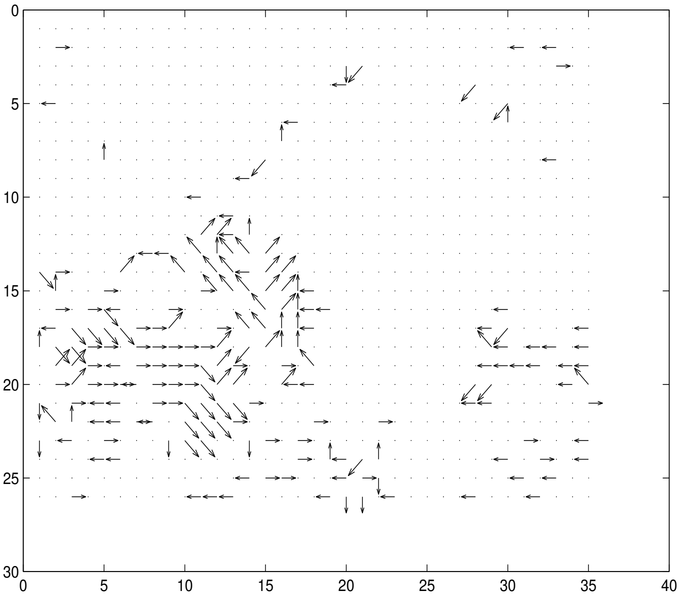


Fig 2.1.3: Average direction computed on 6 frames of the sequence taxi

Fig 2.1.4: One frame of the sequence taxi

sequences, that is to say without scene changes or too irregular movements. A good example of these sequences would typically be the video conference.

Even for those homogeneous sequences, we will be able to assume the stationarity only during short time periods, that is to say a few successive frames. Considering a range of stationarity (a group of pictures), and assuming that the direction process is ergodic and non-periodic, the equation (1.3) says that after a few frame, we should be able to approximate the distribution of $r_z(t)$ by the stationary distribution. This distribution will change during the next range of stationarity.

Thus estimating the parameters of the direction process and then the pixel process will naturally introduce us to the locally stationary processes.

2.3 Spatial and temporal dependency

We propose that the direction process be used to compute the pixel value using the relation :

$$v(t+1,z) = v(t,z-r(t,z)) + n(t+1) \tag{2.6}$$

where $n = \{n(t), t \geq 0\}$ is a sequence of independent and identically distributed random variables representing a noise process. We will denote by $q_n(j) = P[n(t) = j]$ the distribution of the noise random variable $n(t)$, for any j and t. It is reasonable to assume that it is of zero mean, and furthermore that it has a finite second moment.

Suppose that the direction process has a stationnary distribution (as explained before, with $p(u,z) = Prob(r(t,z) = u)$), we can see how the spatial characteristics of the image change with time to be compatible with the temporal model. We consider an image as a set of random variables $v(t,z), \forall z \in Z, \forall t$ each with a different probability distribution $P(v(t,z))$. A whole image at time t is described by $I(t) = \{v(t,z), \forall z \in Z\}$, and we also consider the joint probability

$$P(I(t) = I) = P(v(t,z_1) = i_1, ..., v(t,z_N) = i_N)$$

if a frame contains N pixels. If the noise can be neglected, we have:

$$\begin{aligned} P(v(t+1,z) = i) &= P(v(t,z - r(t,z)) = i) \\ &= \sum_{u \in U} P(v(t,z-u) = i / r(t,z) = u).p(u,z) \end{aligned}$$

We can consider that $\{v(z-u,t) = i\}$ and $\{r(t,z) = u\}$ are independant events (and verify it trough simulation by showing $P(v(t,z-u) = i, r(t,z) = u) = P(v(t,z-u) = v).P(r(t,z) = u))$ and thus:

$$P(v(t+1,z) = i) = \sum_{u \in U} P(v(t,z-u) = i).p(u,z) \qquad (2.7)$$

signifying that the probability distribution of the pixel values of a point z at time $t+1$ is determined by the distributions at time t at this point and his neighbors.

If we consider the effect of noise, with the noise process being independant of the pixel values and direction processes:

$$P(v(t+1,z) = i) = \sum_{\substack{u \in U \\ j \in L}} P(v(t,z-u) = i - j).p(u,z).q_n(j) \qquad (2.8)$$

with L the set of possible gray or color levels for each pixel.

2.3.1 Expression of the mean

$$\begin{aligned} E(v(t+1,z)) &= \sum_{i \in L} i.P(v(t+1,z) = i) \\ E(v(t+1,z)) &= \sum_{i \in L} i. \sum_{\substack{u \in U \\ j \in L}} P(v(t,z-u) + j = i).p(u,z).q_n(j) \\ E(v(t+1,z)) &= \sum_{\substack{u \in U \\ j \in L}} E\left[v(t,z-u) + j\right].p(u,z).q_n(j) \\ E(v(t+1,z)) &= \sum_{\substack{u \in U \\ j \in L}} E\left[v(t,z-u)\right].p(u,z).q_n(j) + E_n \end{aligned}$$

Finally:

$$E(v(t+1,z)) = \sum_{\substack{u\in U\\ j\in L}} E\left[v(t,z-u)\right].p(u,z).q_n(j) \tag{2.9}$$

because we assumed the mean of the noise $E_n = \sum_{j\in L} j.q_n(j)$ to be zero.

Moreover, we want now to consider $v(t,z)$ as a wide-sense stationary process in time, so that $E(v(t,z))$ becomes independant of t and we note:

$$E(v(t,z)) = \mu(z)$$

So (2.9) becomes

$$\mu(z) = \sum_{\substack{u\in U\\ j\in L}} \mu(z-u).p(u,z).q_n(j) \tag{2.10}$$

Note the importance of supposing the direction process stationnary (strict sense, i.e. $p(r(t,z))$ independant of time), in order to preserve the independance of time in (2.10). We will have to make further remarks about this result later (see 2.4). In particular, we see that we have here to make the assumption that the periods of stationarity are the same for different pixels. This can typically be a good approximation on a small blocks of pixels, but we can't for example reuse the equation (2.10) inside the sum for $\mu(z-u)$.

2.3.2 Autocorrelation function

As in chapter 1, let us note the autocovariance function:

$$E\left\{(X(t)-\mu(t))(X(s)-\mu(s))\right\} = R_X(t,s)$$

and the autocorrelation function:

$$E\left\{(X(t)X(s)\right\} = C_X(t,s)$$

which play the same role because we also have to compute the mean.

Considering the pixel process, if we already know the displacement vectors or the probability of such vectors, we are able to express $C(j,k)$ in function of $\{p(u,z), u \in U, z\}$ and a local spatial covariance. With the relation (2.6) of constant intensity (i.e $v(t+1,z) = v(t, z-r(t,z))$) we have :

$$C(t,t+k,z) = E(v(t,z)v(t+k,z)) = E(v(t,z)v(t+k-1, z-r(t+k,z)))$$

If we assume that that the spatial variations on one image at time t are independent of the directions of motion, we can write :

$$C(t,t+k,z) = \sum_{u\in U} E(v(t,z)v(t+k-1,z-u))P(r(t+k,z)=u)$$

Since we suppose that $r_z(t)$ is a stationary process, we have

$$P(r(t+k,z)=u)=p(u,z).$$

We can iterate this relation and we obtain :

$$\begin{aligned} C(t,t+k,z) &= \sum_{u\in U} E(v(t,z)v(t+k-1,z-u))p(u,z) && (2.11)\\ &= \sum_{u_1,\dots,u_k} E(v(t,z)v(t,z-u_1-\cdots u_k))p(u_1,z)\dots p(u_k,z-u_1-\cdots-u_k) \end{aligned}$$

If we consider a short period of time where $v(t,z)$ is stationary, $C(t,t+k,z)$ is independant of t, so:

$$C(k,z)=\sum_{u_1,\dots,u_k} E(v(z)v(z-u_1-\cdots u_k))p(u_1,z)\dots p(u_k,z-u_1-\cdots-u_k) \tag{2.12}$$

does not depend on t. Note that the quality of this approximation is decreasing with k, because we have to suppose the same intervals of stationarity on larger areas of the image in (2.12).

2.4 Complete Model

Before we apply the video model to the estimation of the quality of a compression algorithm, it can be useful to summarize the characteristics we have assumed so far. As we saw in the previous section, to easily compute parameters like the mean or the covariance, it is interesting to find in the processes some stationary periods.

So the following ideas will define our model:

- Our sequence is given by the *image process*, which is a set of *pixel processes*: $I(t)=\{v(t,z),z\in Z\}$. There are underlying *motion processes* $\{r(t,z),z\in Z,t\in\mathbb{N}\}$ that we compute from $I(t)$.
- We assume the stationarity of the direction process at point z on short periods of time. Thus we describe the first level of the model with the following semi-Markov process:

  the Markov chain $\{J_n,n\geq 0\}$ represents the successive states. each state value $i=0,1,\dots$ is characterized by a stationary distribution $p_i(u,z)$ for the motion process: there is a finite number of possible states. Moreover, we suppose that the chain (J_n) is stationary. We define P_{ij} the transition probabilities for the markov chain $(J_n)_{n\in\mathbb{N}}$, and π_i the stationary probability: $\pi_i=P(J_n=i)$.

  $(S_n)_{n\in\mathbb{N}}$ represent the dates of the jumps. We note $T_0=S_0=0,T_n=S_n-S_{n-1}$, the time spent in the successive states. We note the distribution functions H_i, with $H_i(t) = P(T_n \leq t | J_n = i)$, and take the other notations from 1.4.

- We will have to work independently on different blocks of the successive frames (typically squares of $8 * 8$ pixels). For each pixel of one block, at time t, the states J_n can be different. But the times S_n of jump are the same for all pixels of one block (in order to assure the compatibility with the following point, and practically we compute the motion by blocks).
- For each state J_n (state of stationarity for r_z which lasts T_n), we assume also that the pixel processes v_z are wide-sense stationary. The dependance between v_z and r_z is explained in 2.3. The relations (2.10) and (2.12) are valid as long as we stay in the same block.

Note now that each time we assume that a process is stationary (J_n during the whole sequence, r_z, v_z during the times T_n) we have in the same time a way to compute the probabilities ($p(u,z)$, π_i and P_i with (1.1) and (1.8) , considering a restriction on the appropriate periods of time) or the mean and covariance (for v_z and r_z by an empirical mean on the intervals T_n).

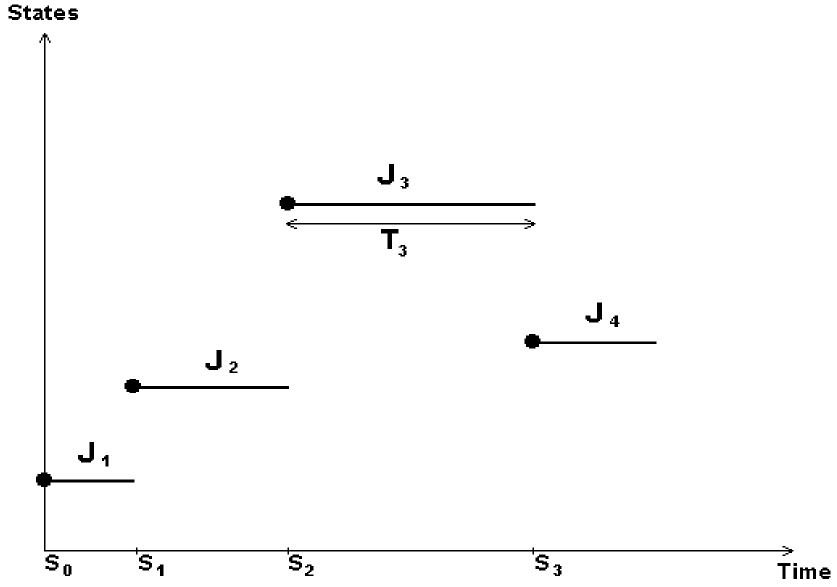


Fig 2.4.1: A sample path of the semi-Markov process

2.5 Practical determination of stationary periods

In order to determine the sojourn times in each state, we use the ideas and algorithm developped in [4] and [5], working with the norm of the motion vector as signal. Thus we obtain the approximate Karhunen-Loève basis of the signal, by minimizing an entropy function. This is of course an empirical method to determine the states where the motion process is supposed to be stationary, but it gives plausible results for our study. The algorithm we use is part of the WaveLab library implemented in Matlab (function *cptour*). It constructs the binary tree of the best basis and we use it to select the intervals of stationarity. Then we are able to compute all the values we need on the stationary periods.

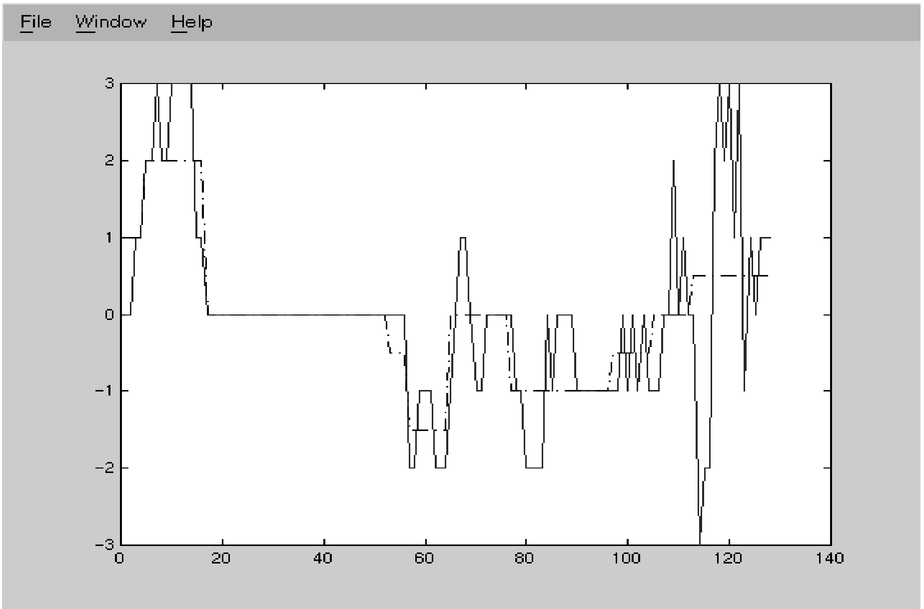


Fig 2.5.1: Speed profil of the salesman sequence and its mean computed on stationary periods

Figure 2.5.1 shows the evolution of the vertical speed for a block of the sequence Salesman. The algorithm gives us a binary tree representing the best basis, we compute the mean on these intervals and we discretize it.

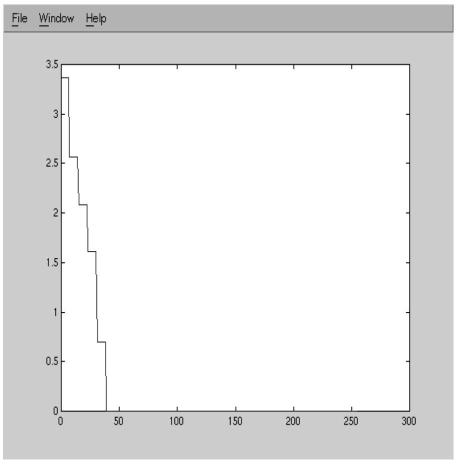
2.6 First study of sojourn times length

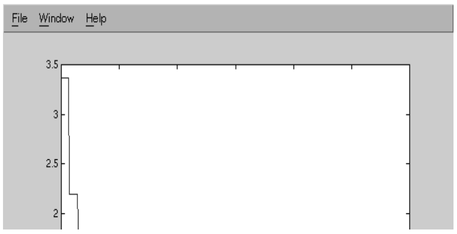
In this section we give an example of the study of the sojourn times length for a sequence of 4096 frames. We use the previously described algorithm to compute the different sojourn times in the successive states, which gives times of the form 2^j. In this approach, we characterize a state by the value of the mean of the speed, after quantization in order to obtain a finite set of states. So we approximate the considered semi-Markov process by the *mean speed process*. The uniform quantization of the average speed gives 10 different states, for which we study the law of the sojourn times.

If H_i is an exponential distribution (that is, the semi-Markov process becomes a Markov process) with parameter λ_i:

$$H_i(t) = P(T_n \leq t | J_n = i) = 1 - e^{-\lambda_i t}$$

we see that plotting the number of sojourn times T_n in state i which are greater than t, as a function of t, should give an exponential $e^{-\lambda_i t}$. The following figures (Fig. 2.6.1) give for some states the *log* of this number, and we see that in most cases, an approximation by a line, whose slope would give the coefficient $-\lambda_i$ is correct, what seems to justify an approximation by exponential distributions.


File
Window
Help
3.5
3
2.5
2
1.5
1
0.5
0
0
50
100
150
200
250
300


File
Window
Help
3.5
3
2.5
2