Joint 48th IEEE Conference on Decision and Control and
28th Chinese Control Conference
Shanghai, P.R. China, December 16-18, 2009

Methodologies for the Adaptive Compression of Video Sequences

Angelo Cenedese and Riccardo Marcon

Abstract— **In this work, a procedure for the video compression and transmission is presented, based on a Singular Value Decomposition approach, whose controls are obtained as the output of a constrained optimization problem that refers to the compression ratio as the optimization functional and an image quality index as the performance constraint. The tools of estimation theory allow to obtain a polynomial approximation of these indexes in a static fashion via least square technique, and adaptively, with the concurrent estimation of both the order of the polynomial functions and the salient function parameters through the use of Kalman Filters. The implementation of the whole system and some simulations of real video sequences are presented to validate and assess the proposed procedure.**

I. INTRODUCTION

The problem of video coding and compression is well known and widely studied in the image processing and communication community, and in the technical literature many standards are proposed to approach and solve the issue, that mainly regard computational aspects and in some cases software/hardware integrated architectures.

Conversely, this work has a methodological purpose in the tradition of system theory disciplines, and the main objective is the presentation of a theoretical framework where the typical issues raised by the quest for the "optimal" compression algorithm are approached.

In particular, the proposed method to perform video data compression for transmission is based on the Singular Value Decomposition (SVD) [1][2] of video sequences, and the solution of an optimization problem, where the control variables are the parameters of the SVD, and the optimization is carried out with respect to figures of merit such as the Compression Ratio and the image quality, expressed with different metric functions. Some specific aspects such as the approximation of these performance indexes and the online estimation of the relevant parameters are studied, and the implementation of the whole system is developed.

More in detail, the paper is organized as follows: In Sec. II the notation is introduced and the formalization of the problem with respect to the SVD coding/decoding procedure is sketched. In the main Sec. III, the compression algorithm is described, and a least square static and a Kalman Filter based dynamic approaches are presented. Sec. IV summarizes the structure of the procedure and Sec. V presents some simulation results on real sequences, taking into account the performance indexes Compression Ratio and reconstruction quality, but also giving some insights into the computational cost of the algorithm. Some conclusions and future work perspective are finally given in Sec. VI.

This work was supported by the Italian Ministry of Research and University through the PRIN programme 2006 "Smart Environments: Event Understanding, Sensor Adaptive Configuration and Multimodal Interfaces"

A. Cenedese is with the Department of Engineering and Management, University of Padova, 36100 Vicenza, Italy angelo.cenedese@unipd.it

R. Marcon is with the Department of Information Engineering, University of Padova, 35131 Padova, Italy marcon.riccardo@gmail.com

II. SVD COMPRESSION

As a first point, a preliminary summary of the adopted notation is in order. Be the frame sequence of interest represented as a temporal series of N frames, $\{I_t(x, y, n), n = 1, \dots, N\}$ with $t \in \mathbb{Z}$; each frame $I_t(x, y, n)$ has size $L_x \times L_y$ pixels and is subdivided into a grid of M blocks $b_{i,t}(x, y, n), i = 1, \dots, M$ of size $L \times L$ pixels, that form M-block families $\{b_{i,t}(x, y, n), n = 1, \dots, N\}, i = 1, \dots, M$.

The rationale behind the proposed algorithm to compress and transmit video sequences resides in a two step procedure, respectively performing the Singular Value Decomposition for each block (frame) and transmitting the principal components together with the related coefficients, instead of transmitting straightforwardly all the frame pixels (see also [3]). The steps for coding and decoding, that are schematically shown in Fig. 1, are described in next subsections.

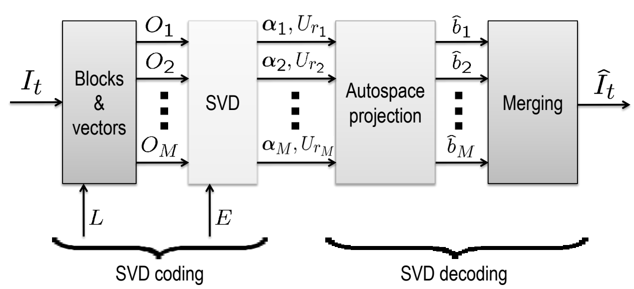

Fig. 1. Coding/Decoding procedure for SVD compression.

A. Coding

Starting from the sequence $\{b_{i,t}(x, y, n), n = 1, \dots, N\}$, each block $b_{i,t}(x, y, n)$ is reshaped as a vector $\mathbf{b}_i(n) \in \mathbb{R}^{L^2}$ of pixel intensities so as to construct the matrix

$$O_i = \begin{bmatrix} \mathbf{b}_i(1) & \mathbf{b}_i(2) & \cdots & \mathbf{b}_i(N) \end{bmatrix} \in \mathbb{R}^{L^2 \times N},$$

that is then decomposed according to the SVD procedure, $O_i = U_i \Sigma_i V_i^\top$ [4]: In the orthogonal matrix

$$U_i = \begin{bmatrix} \mathbf{u}_1 & \cdots & \mathbf{u}_{r_i} & \mathbf{u}_{r_i+1} & \cdots & \mathbf{u}_{L^2} \end{bmatrix} \in \mathbb{R}^{L^2 \times L^2}$$

the first r_i columns, corresponding to the non-null singular values $\{\sigma_1, \dots, \sigma_{r_i}\}$ in Σ_i, represent the principal components of O_i (rank $O_i = r_i$). An effective data compression is obtained by truncating the U_i matrix to the first $r_i^* \leq r_i$ columns: The resulting vector space

$$\mathcal{U}_{r_i^*} = \text{span}\{\mathbf{u}_1, \dots, \mathbf{u}_{r_i^*}\}$$

retains an amount of energy data of

$$E(r_i^*) = \frac{\sum_{k=1}^{r_i^*} \sigma_k^2}{\sum_{k=1}^{r_i} \sigma_k^2}. \tag{1}$$

The projection of the vectors $\mathbf{b}_i(n)$, $n = 1, \dots, N$ over the vector space $\mathcal{U}_{r_i^*}$ gives raise to the definition of the coefficients

$$\boldsymbol{\alpha}_i(n) = \left(U_{r_i^*}^\top U_{r_i^*}\right)^{-1} U_{r_i^*}\mathbf{b}_i(n) \in \mathbb{R}^{r_i^*}, \tag{2}$$

where

$$U_{r_i^*} = \begin{bmatrix}\mathbf{u}_1 & \cdots & \mathbf{u}_{r_i^*}\end{bmatrix} \tag{3}$$

are the principal components of interest.

B. Decoding

Principal components (3) and coefficients (2) are transmitted, and during the decoding phase they are recombined to reconstruct each frame block; the reconstruction of the i-th block of the n-th frame is given by a linear combination of eigen-images

$$\hat{\mathbf{b}}_i(n) = U_{r_i^*}\boldsymbol{\alpha}_i(n).$$

Finally, by grouping together all block estimates, the frames and the sequence $\hat{I}_t(x, y, n)$ is obtained.

In literature, several metrics are studied to evaluate and assess the accuracy and the goodness of the reconstruction, and those based on pixel-to-pixel comparison are among the most popular, such as the Mean Squared Error (MSE):

$$MSE_t \triangleq \frac{1}{N}\sum_{n=1}^{N}\left[\frac{1}{L_x L_y}\sum_{x=1}^{L_x}\sum_{y=1}^{L_y}\Delta I_t(x, y, n)^2\right], \tag{4}$$

being $\Delta I_t(u, v, n) = I_t(u, v, n) - \hat{I}_t(u, v, n)$ the difference between the original and the reconstructed images respectively. In the following, the peak to noise ratio will be used, whose relation with the MSE is as follows:

$$PSNR_t \triangleq 20\log\left(\frac{255}{\sqrt{MSE_t}}\right). \tag{5}$$

This aspect will then be considered in Sec. III-E.

III. Adaptive Compression Algorithm

Firstly, the Compression Ratio (CR) is introduced, being the ratio between the compressed data volume (byte) and the original data volume (byte):

$$CR = \frac{\sum_{i=1}^{M}\left(4L^2 r_i^* + 4N r_i^*\right)}{N L_x L_y}. \tag{6}$$

Here and in the following, no quantization is adopted; images are composed of 8-bit pixels, and all other data are 32-bit floating point numbers.

The proposed algorithm acts so as to minimize at each time step the cost of the data transmission CR, while keeping the image quality over a desired threshold; this latter parameter is estimated by now through the $PSNR$ (5).

To this aim, the optimal solution of the following constrained problem is sought:

$$\boldsymbol{\xi}_{opt} = \arg\min_{\boldsymbol{\xi}=(\xi_1,\xi_2)\in\Xi} CR(\boldsymbol{\xi}), \tag{7}$$

where the control quantities are given by $\xi_1 = E$, transmitted energy percentage (1) (thus related to the parameter r_i^*), and $\xi_2 = L$, block size; the constraint is given by the following expression

$$\Xi \triangleq \left\{(\xi_1, \xi_2) \in \mathbb{R}_+ \times \mathbb{N} \;\middle|\; \left\{\begin{array}{l} 0 \leq \xi_1 \leq 1 \\ 1 \leq \xi_2 \leq \min\{L_x, L_y\} \\ PSNR \geq \tau_{PSNR} \end{array}\right.\right\},$$

where τ_{PSNR} is the minimum quality threshold that can be accepted. In this context, the constrained optimization problem is solved numerically through the MATLAB™ fmincon routine.

To complete the definition of the problem (7) both the cost function and the constraint should be expressed in terms of $\boldsymbol{\xi} = (\xi_1, \xi_2)$: In this respect, two polynomial approximation models for the CR and $PSNR$ functions are derived using the Minimum Squares approach, and the following quadratic expressions are assumed in the first instance on an empirical base

$$\hat{CR}(\boldsymbol{\xi}) = a_0 + a_1\xi_1 + a_2\xi_2 + a_3\xi_1\xi_2 + a_4\xi_1^2 + a_5\xi_2^2 \tag{8}$$

$$PS\hat{N}R(\boldsymbol{\xi}) = b_0 + b_1\xi_1 + b_2\xi_2 + b_3\xi_1\xi_2 + b_4\xi_1^2 + b_5\xi_2^2. \tag{9}$$

The model parameter identification is now performed either statically or dynamically.

A. Static Approach

In the static approach, a first step involves the simulation of the compression for a certain number of values $\boldsymbol{\xi}^{(k)} = \left(\xi_1^{(k)}, \xi_2^{(k)}\right)$, $k = 1, \dots, K$ and the computation of the CR and $PSNR$ correspondent values. These measurements are then exploited for the estimation of the (a_j, b_j) parameters.

A comment on notation is now in order. Here and in the remainder of the paper the subscripts (CR or $PSNR$) referred to all quantities are omitted when possible to favor the clarity of the procedure and avoid weighing down the reading: Measurements $\mathbf{y}_{CR}$ and $\mathbf{y}_{PSNR}$ are indicated simply with $\mathbf{y}$, parameters θ_{CR} and θ_{PSNR} with θ, and so on.

The measurements $\mathbf{y}$ (both for CR and $PSNR$) are supposed to be produced by the following model:

$$\mathbf{y}(t) = S_K\boldsymbol{\theta}(t) + \mathbf{e}(t), \tag{10}$$

where the output measurement vector is either

$$\mathbf{y} = \mathbf{y}_{CR} = \begin{bmatrix}CR(\boldsymbol{\xi}^{(1)}) & \cdots & CR(\boldsymbol{\xi}^{(K)})\end{bmatrix}^\top$$

or

$$\mathbf{y} = \mathbf{y}_{PSNR} = \begin{bmatrix}PSNR(\boldsymbol{\xi}^{(1)}) & \cdots & PSNR(\boldsymbol{\xi}^{(K)})\end{bmatrix}^\top,$$

for the two quantities, respectively; the parameter vector $\boldsymbol{\theta}$ particularizes into

$$\boldsymbol{\theta}_{CR} = \begin{bmatrix} a_0 & a_1 & a_2 & a_3 & a_4 & a_5 \end{bmatrix}^\top \quad (11)$$

$$\boldsymbol{\theta}_{PSNR} = \begin{bmatrix} b_0 & b_1 & b_2 & b_3 & b_4 & b_5 \end{bmatrix}^\top, \quad (12)$$

and $\mathbf{e}$ ($\mathbf{e}_{CR}$ and $\mathbf{e}_{PSNR}$) is a zero-mean random error vector; finally the state matrix $S_K \in \mathbb{R}^{K\times 6}$ is given by

$$S_K = \begin{bmatrix} 1 & \xi_1^{(1)} & \xi_2^{(1)} & \xi_1^{(1)}\xi_2^{(1)} & \left(\xi_1^{(1)}\right)^2 & \left(\xi_2^{(1)}\right)^2 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & \xi_1^{(K)} & \xi_2^{(K)} & \xi_1^{(K)}\xi_2^{(K)} & \left(\xi_1^{(K)}\right)^2 & \left(\xi_2^{(K)}\right)^2 \end{bmatrix}. \quad (13)$$

The minimum square estimate $\hat{\boldsymbol{\theta}}$ ($\hat{\boldsymbol{\theta}}_{CR}$ and $\hat{\boldsymbol{\theta}}_{PSNR}$) is obtained resorting to

$$\hat{\boldsymbol{\theta}}(t) = \left(S_K^\top S_K\right)^{-1} S_K^\top \mathbf{y}(t).$$

In this approach, although $K \geq 6$ is sufficient to produce a solution, practically a very large number ($K \gg 6$) of simulations is needed because the function profiles has to adapt on a vast region in the Ξ domain and avoid overfitting.

B. Dynamic Adaptive Approach

In this second approach, the parameters of the profile functions are not computed from scratch at each time-step t, but are recursively obtained through a Kalman Filter procedure [5][6], as shown in the schematic drawing of Fig. 2. Since the dynamics of parameter change is not known

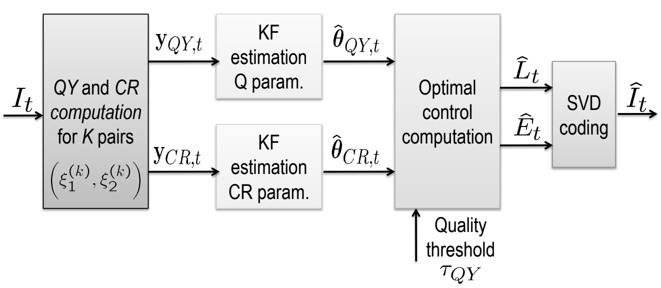

Fig. 2. Diagram of the proposed algorithm. QY indicates a generalized quality index ($PSNR$ or those presented in III-E).

a-priori, a random walk model is adopted for $\boldsymbol{\theta}$ ($\boldsymbol{\theta}_{CR}$ and $\boldsymbol{\theta}_{PSNR}$), the state equation being

$$\boldsymbol{\theta}(t+1) = A\boldsymbol{\theta}(t) + \mathbf{v}(t) \quad (14)$$

where $A = I_6$, $\mathbf{v}(t)$ ($\mathbf{v}_{CR}(t)$ and $\mathbf{v}_{PSNR}(t)$) is a zero-mean white noise with covariance matrix Q (Q_{CR} and Q_{PSNR}); these latter values are subject to a tuning procedure (Sec. III-D), because they are strictly related to the rate of change of $\boldsymbol{\theta}$ (the faster the change of $\boldsymbol{\theta}$, the bigger the value of Q). The output equation for $\mathbf{y}(t)$ ($\mathbf{y}_{FC}(t)$, $\mathbf{y}_{PSNR}(t)$) is

$$\mathbf{y}(t) = C_t\boldsymbol{\theta}(t) + \boldsymbol{w}(t), \quad (15)$$

where C_t is taken as S_K in (13) with $K = 6$, $C_t \in \mathbb{R}^{6\times 6}$, and $\mathbf{w}$ is a zero-mean noise of variance R. The output matrix C_t is time-variant, obtained by taking at each timestep a random set of $(\xi_1^{(k)}, \xi_2^{(k)})$ in order to keep the estimation procedure valid for the whole Ξ domain.

The Minimum Variance Estimator of the parameter $\boldsymbol{\theta}$ for the so defined linear time-variant model, given the measurements $\{\mathbf{y}(s); 0 \leq s \leq t\}$, is obtained as [7][8]:

1) Initial Conditions

$$\hat{\boldsymbol{\theta}}(0|0) = \mathbb{E}\left[\boldsymbol{\theta}(0)\right]$$
$$P(0|0) = \mathrm{Var}\left[\boldsymbol{\theta}(0)\right]$$

2) Prediction

$$\hat{\boldsymbol{\theta}}(t|t-1) = \hat{\boldsymbol{\theta}}(t-1|t-1)$$
$$P_{t|t-1} = P_{t-1|t-1} + Q$$

3) Correction

$$\hat{\boldsymbol{\theta}}(t|t) = \hat{\boldsymbol{\theta}}(t|t-1) + L_t\left[\mathbf{y}(t) - C_t\hat{\boldsymbol{\theta}}(t|t-1)\right]$$
$$P_{t|t} = P_{t|t-1} - P_{t|t-1}C_t^\top\Lambda_t^{-1}C_tP_{t|t-1}$$

where $P_{t|t-1}$ and $P_{t|t}$ are the a-priori and a-posteriori error variances,

$$P_{t|t-1} = \mathrm{Var}\left[\hat{\boldsymbol{\theta}}(t|t-1) - \boldsymbol{\theta}(t)\right]$$
$$P_{t|t} = \mathrm{Var}\left[\hat{\boldsymbol{\theta}}(t|t) - \boldsymbol{\theta}(t)\right],$$

Λ_t refers to the variance of the innovation process $\mathbf{e}(t) = \mathbf{y}(t) - C_t\hat{\boldsymbol{\theta}}(t|t-1)$,

$$\Lambda_t = C_tP_{t|t-1}C_t^\top + R$$

and the filter gain L_t is defined as

$$L_t = P_{t|t-1}C_t^\top\Lambda_t^{-1}.$$

From the asymptotic analysis of the Kalman Filter for time-invariant models ($C_t = C$) [6], condition for the convergence of $P_{t|t-1}$, $t \to +\infty$, to the definite-positive stabilizing solution $\bar{P}$ of the Discrete time Algebraic Riccati Equation (DARE)

$$P = A\left[P - PC^\top\left(CPC^\top + R\right)^{-1}CP\right]A^\top + Q \quad (16)$$

is the detectability of (A, C) and the reachability of (A, Q). This analysis is extended to the case of time-variant system and the mentioned condition is ensured by carefully selecting the tuning matrices: In particular, by choosing a definite-positive Q, the reachability matrix

$$\mathcal{R} = \begin{bmatrix} Q & AQ & \cdots & A^5Q \end{bmatrix}$$
$$= \begin{bmatrix} Q & Q & \cdots & Q \end{bmatrix}$$

is full rank; similarly, having C (C_t) invertible yields a full rank observability matrix:

$$\mathcal{O} = \begin{bmatrix} C \\ CA \\ \vdots \\ CA^5 \end{bmatrix} = \begin{bmatrix} C \\ C \\ \vdots \\ C \end{bmatrix}.$$

C. Polynomial Approximation of CR and $PSNR$

To improve the performance of the system, the assumption made on the order of the polynomial approximation (11)-(12) is now discussed, and the cardinality of the parameter sets will be indicated with p.

To find an optimal value for p (hence the order of the polynomial approximation) an identification procedure is carried out on a given measurement set $\mathbf{y}$

$$\mathbf{y} = \begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \end{bmatrix} \in \mathbb{R}^K,$$

where the first $K/2$ data $\mathbf{y}_1$ are used for the model identification ($\boldsymbol{\theta}_p \in \mathbb{R}^p$), and the remaining $K/2$ measurements $\mathbf{y}_2$ are employed to assess the accuracy of the computed model; also, it is assumed that $\mathbf{y}_1$ and $\mathbf{y}_2$ are equal mean, and the total covariance of $\mathbf{y}$ is $\sigma^2 I_K$. Therefore, the prediction error on future data given by

$$\boldsymbol{\varepsilon} = \mathbf{y}_2 - S_K \hat{\boldsymbol{\theta}}_p(\mathbf{y}_1)$$

shows [7][8]

$$\mathrm{var}[\boldsymbol{\varepsilon}] = K\sigma^2 \left(1 + \frac{p}{K}\right),$$

linearly dependent on p; σ^2 is not a-priori known, but can be computed using the Minimum Variance Unbiased Estimator [7][8]

$$\frac{K}{K-p}\hat{\sigma}_p^2 = \frac{1}{K-p} ||\mathbf{y}_1 - S_K \hat{\boldsymbol{\theta}}_p(\mathbf{y}_1)||^2.$$

The optimal order can be then computed by minimization of the Final Prediction Error (FPE)

$$\mathrm{FPE}(p) \triangleq \hat{\sigma}_p^2 \frac{1 + p/K}{1 - p/K} \tag{17}$$

constraining p in $[p_{\min}, p_{\max}]$. It follows:

$$p_{opt} = \arg\min_p \mathrm{FPE}(p).$$

Operatively, a set of models with increasing p, $p \in [p_{\min}, p_{\max}]$, is identified by computing $\hat{\sigma}_p^2$ at each time t and constructing the FPE function (17), leading to $p_{CR}(t)$ e $p_{PSNR}(t)$. The choice of a single order p for both CR and $PSNR$ is performed by choosing the most frequent value in this family.

D. Kalman Filter Tuning

In the model (14) the noise variance Q can be computed, in the hypothesis of i.i.d. noise, through the ergodic theorem

$$\lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} [\boldsymbol{\theta}(t+1) - \boldsymbol{\theta}(t)] [\boldsymbol{\theta}(t+1) - \boldsymbol{\theta}(t)]^\top = Q.$$

In practice, both in the static and in the dynamic parameter estimation, $\boldsymbol{\theta}(t)$ is not known, and only the estimate $\hat{\boldsymbol{\theta}}(t)$ can be inferred; therefore the noise variance is at timestep T approximated by

$$\hat{Q}_T = \frac{1}{T} \sum_{t=1}^{T} \left[\hat{\boldsymbol{\theta}}(t) - \hat{\boldsymbol{\theta}}(t-1)\right] \left[\hat{\boldsymbol{\theta}}(t) - \hat{\boldsymbol{\theta}}(t-1)\right]^\top$$

and, recursively, by

$$\hat{Q}_{T+1} = \frac{T\hat{Q}_T + \left[\hat{\boldsymbol{\theta}}(T+1) - \hat{\boldsymbol{\theta}}(T)\right]\left[\hat{\boldsymbol{\theta}}(T+1) - \hat{\boldsymbol{\theta}}(T)\right]^\top}{T+1}.$$

In this respect, some learning sequences are exploited through the static estimation procedure to derive a fixed variance value for the Kalman Filter. The output equation in this case is

$$\mathbf{y}(t) = S_K \boldsymbol{\theta}(t) + \mathbf{e}(t),$$

where $\mathbf{e}(t)$ is a gaussian random vector, $\mathbf{e}(t) \in \mathcal{N}\left(0, R(t) = \sigma^2(t) I_K\right)$. Being p the dimension of $\boldsymbol{\theta}(t)$,

$$\hat{\sigma}^2(t) = \frac{1}{K-p} ||\mathbf{y}(t) - S_K \hat{\boldsymbol{\theta}}(t)||^2. \tag{18}$$

By using these learning variances in the output equations given by (15), it follows:

$$R = \left[\frac{1}{T} \sum_{t=1}^{T} \hat{\sigma}^2(t)\right] I_K,$$

for both CR and $PSNR$.

E. Choice of the Quality Index

The choice of the quality index is of paramount importance for the derivation of the optimal values for the control variables, in order to provide a good visual quality at a possibly minimum cost. So far the quality index that has been considered (widely exploited in the literature) is simply the signal to noise ratio $PSNR$ (5). It is reasonable, though, to explore different possibilities; in particular, in this work, the idea of taking into account the human vision system suggests metrics where only the energy in the bandwidth of the Human Vision System (HVS) is considered [9][10].

In the hypothesis standing for the vision system of being linear and isotropic for low contrasted monochromatic images in absence of fast dynamics, a simple function characterizes the system. The human sensitivity threshold function has been studied [11][12], and by taking the inverse of such a function the spatial frequency response $H(r)$ is approximated. A symmetric HVS model is given by

$$H(r) = (0.2 + 0.45r)e^{-0.18r} \tag{19}$$

where the radial frequency is $r = \sqrt{u^2 + v^2}$, being u and v spacial frequencies. The principal spectral contribution for the images to be analyzed appears at low frequencies and the $H(r)$ filter reduces the static component ($r = 0$) and highlights its neighbor contributions: This is related to how the human system is sensitive to low contrast differences [13]. The $PSNR$ function has been modified in order to accommodate the HVS sensitivity by computing the energy of the error between reconstructed and original images, thus obtaining the "Peak Signal to HVS Noise Ratio":

$$PSHVSNR_t \triangleq 20 \log \left(\frac{255}{\sqrt{HVSMSE_t}}\right), \tag{20}$$

where the MSE weighted by the HVS function is

$$HVSMSE_t \triangleq \frac{1}{N}\sum_{n=1}^{N}\left[\sum_{u=1}^{L_x}\sum_{v=1}^{L_y} H(u,v)^2 \Delta F_t(u,v,n)^2\right], \tag{21}$$

$\Delta F_t(u,v,n) = \left|F_t(u,v,n) - \hat{F}_t(u,v,n)\right|$ being the difference between the Fourier transforms of $I_t(u,v,n)$ e $\hat{I}_t(u,v,n)$ respectively.

Further modified versions of (20) and (21) yield for example the "Normalized HSV MSE" [10]

$$HVSNMSE_t \triangleq \frac{1}{N}\sum_{n=1}^{N}\frac{\sum_{u=1}^{L_x}\sum_{v=1}^{L_y} H(u,v)^2 \Delta F_t(u,v,n)^2}{\sum_{u=1}^{L_x}\sum_{v=1}^{L_y} H(u,v)^2 \left|F_t(u,v,n)\right|^2}, \tag{22}$$

or the $logHVSNMSE$

$$logHVSNMSE_t \triangleq -10\log HVSNMSE_t \tag{23}$$

where small index variations are enhanced.

IV. Implementation of the Full Procedure

In this section the procedure for the adaptive compression of the video sequence is summarized, with reference to the schematic drawings of Figs. 1-2. The system receives sequences I_t, $t \in \mathbb{Z}_+$, of N frames, that are compressed adaptively to obtain $\hat{I}_t$. The following phases are highlighted:

1) the SVD compression is simulated for K pairs $\boldsymbol{\xi}^{(k)} = \left(\xi_1^{(k)}, \xi_2^{(k)}\right)$, $k = 1\ldots, K$ and the order p is obtained through the optimal procedure described before;
2) performance vectors $\mathbf{y}_{QY} \in \mathbb{R}^K$ and $\mathbf{y}_{CR} \in \mathbb{R}^K$, are composed, whose components are the quality index and compression ratio values for each pair $\boldsymbol{\xi}^{(k)}$; the cardinality of these two sets in the static case is much bigger than in the dynamic case ($K \gg p$);
3) the parameters of the polynomial approximation, $\boldsymbol{\theta}_{QY} \in \mathbb{R}^p$ e $\boldsymbol{\theta}_{CR} \in \mathbb{R}^p$, are estimated exploiting measurements $\mathbf{y}_{QY}$ e $\mathbf{y}_{CR}$ through either the minimum square approach or the Kalman Filter; in this latter case, the filters adapt their estimates according to the rate of variability of the parameters;
4) the constrained optimization problem (7) is solved numerically given a minimum threshold τ_{QY} for the reconstruction quality;
5) the optimal quantities $\hat{E} = \xi_{1,opt}$ and $\hat{L} = \xi_{2,opt}$ are used as control variables for the SVD compression of input sequence I_t to obtain $\hat{I}_t$.

V. Simulations and Experimental Results

The algorithms have been tested over several sequences. Here are reported results from 26 sequences of $N = 50$ frames ($L_x \times L_y = 576 \times 720$ pixel). The simulations have been performed on a notebook equipped with CPU Intel Core Duo T2300 at 1.6 GHz, 2 GB RAM at 666 MHz, running Windows XP Professional Ed.; the development environment is MATLAB™ 7.5.0 (version R2007b).

The order of the polynomial approximation model has been identified through the procedure of Section III-C, choosing among a set of complete polynomial expressions $f(\xi_1,\xi_2)$, ranging from $p = 3$, $f(\xi_1,\xi_2) = a_0 + a_1\xi_1 + a_2\xi_2$, to $p = 11$, $f(\xi_1,\xi_2) = a_0 + a_1\xi_1 + a_2\xi_2 + a_3\xi_1\xi_2 + a_4\xi_1^2 + a_5\xi_2^2 + a_6\xi_1^2\xi_2 + a_7\xi_1\xi_2^2 + a_8\xi_1^2\xi_2^2 + a_9\xi_1^3 + a_{10}\xi_2^3$. An example of FPE curves (17) for $PSNR$ and CR is given in Fig. 3, where the two orders related to $p = 6$ and $p = 10$ are highlighted. The former corresponds to the empirical quadratic approximation (11)-(12), while the latter represent the actual optimal value p_{opt} that emerges consistently in the analysis of all sequences ($t = 1,\ldots,26$) for both $PSNR$ and CR and gives rise to the following polynomial expression:

$$\begin{aligned} f(\xi_1,\xi_2) &= a_0 + a_1\xi_1 + a_2\xi_2 + a_3\xi_1\xi_2 + a_4\xi_1^2 + a_5\xi_2^2 \\ &+ a_6\xi_1^2\xi_2 + a_7\xi_1\xi_2^2 + a_8\xi_1^2\xi_2^2 + a_9\xi_1^3. \end{aligned}$$

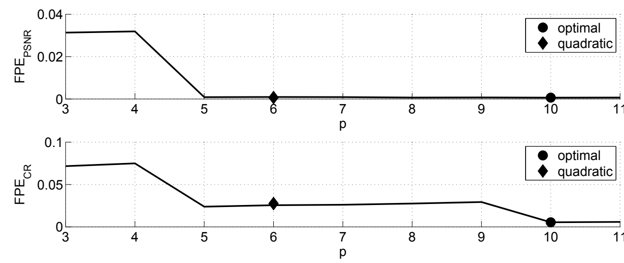

Fig. 3. FPE function example for $PSNR$ and CR profiles, as a function of the order p. The quadratic order ($p = 6$) and the optimal order ($p = 10$) are shown with diamond and circle respectively.

The implementation of the compression procedure is performed, and at each time step t the optimal control values $\hat{E}$ and $\hat{L}$ are computed. The expressions (5)-(6) for the interesting quantities are given in Figs. 4-5, referring respectively to the low-order quadratic polynomial approximation and the high-order optimal one.

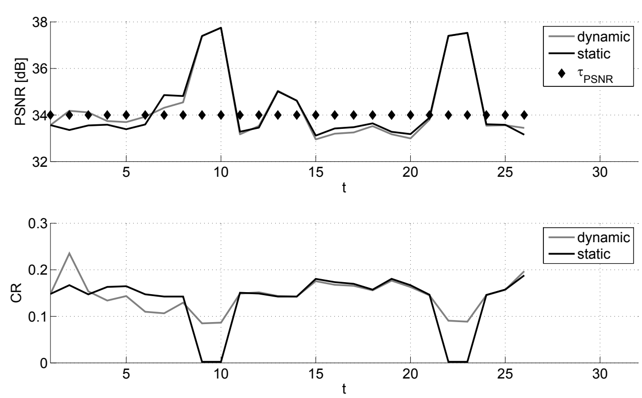

Fig. 4. $PSNR$ and CR behavior. Quadratic polynomial approximation ($p = 6$ parameters): Static estimation (black) vs dynamic estimation (gray) of the parameters; the quality threshold τ_{PSNR} is shown as a dotted line.

The static and the dynamic procedures performs similarly, though for high-order approximations the static procedure is more demanding in terms of computational effort, since it requires the inversion of the $K\times p$ S_K matrix ($K \gg p$), while

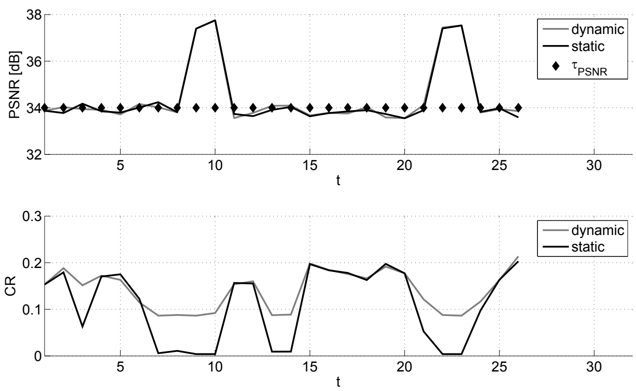

Fig. 5. $PSNR$ and CR behavior. Quadratic polynomial approximation ($p = 10$ parameters): Static estimation (black) vs dynamic estimation (gray) of the parameters; the quality threshold τ_{PSNR} is shown as a dotted line.

in the Kalman filter approach the state matrix is square, $p \times p$. Moreover, exploiting the high-order optimal approximation allows reaching better results in meeting the requirements on $PSNR$ also with a slightly better attained compression.

On the whole, the preliminary experiments show that both the static and the dynamic procedure are effective in producing a good quality ($PSNR \approx \tau_{PSNR}$) compressed video sequence ($CR < 1$), although more extensive validation is needed to gain some insight into the performance sensitivity with respect to the video sequence (dynamics time-scale, periodicity, ...) and to understand whether it is possible to state a trade-off between the static and the dynamic solutions.

To complete the study, it is worth giving some details of the performance with respect to the different quality indexes (5), (20), and (23). In general, it can be observed that the computation of $\hat{I}_t$ using the $PSNR$ metrics is on average faster but the requirement in bytes is higher (higher CR), for a chosen threshold τ_{QY}; on the other side, the CR performance is basically equivalent if measured with $PSHVSNR$ or $logHVSNMSE$. As far as the complexity is concerned, the algorithm execution time is mainly devoted to the SVD compression procedure, and only a small fraction is employed for the calculation of the quality index. Moreover, the SVD compression time is insensitive to the energy E values, while decreases quadratically as the block dimension L grows. Tab. I shows that the quality index computation time is constant with respect to any choice of control pair (E, L) (time is measured for a chosen sequence I_t), and the computation of the Fourier transform makes the calculation of $PSHVSNR$ (one FFT computation) and $logHVSNMSE$ (two FFT computations) more demanding if compared to $PSNR$, and basically takes all the algorithm execution time.

VI. Conclusions

In this work, an approach to video compression and transmission has been explored, based on the SVD decomposition according to the computation of the optimal choice for the SVD control quantities. To this aim a constrained optimization problem is defined and numerically solved.

index	mean [s]	var [s^2]
PSNR	0.4189	$4.2187 \cdot 10^{-4}$
PSHVSNR	4.0035	0.0011
logHVSNMSE	7.8812	$4.8718 \cdot 10^{-4}$

TABLE I

Execution time w.r.t different QY indexes.

The performance indexes that are considered in this framework are the compression ratio CR and the image quality index QY. In particular, the estimation of both the order and the parameters of the polynomial approximation of these figures of merit have been studied and implemented in the algorithm: The former is obtained through the minimization of the final prediction error, while the latter are estimated either statically via Least Squares or dynamically by using a pair of Kalman Filters running in parallel.

There are still many aspects that it is worth studying with future research: Firstly, some insight into the optimization problem itself is needed, beyond the numerical solution. Moreover, the following dual problem should be addressed:

$$\boldsymbol{\xi}_{opt} = \arg \max_{\mathrm{CR} \leq \tau_{CR}} \mathrm{QY}(\boldsymbol{\xi}).$$

Also different approximation techniques for the performance indexes and other techniques for the recursive estimation of the relevant parameters can be employed. Finally, a more complete comparative examination of the video quality index would be of interest.

References

[1] U. Söderström, "Very low bitrate facial video coding based on principal component analysis," Ph.L Thesis, online:www.diva-portal.org/umu/abstract.xsql?dbid=895[online], 2006.

[2] U. Söderström and H. Li, "Full-frame video coding for facial video sequences based on principal component analysis," in *Proceedings of the Irish Machine Vision and Image Processing Conference (IMVIP)*, 2005, pp. 25–32.

[3] A. Cenedese, R. Frezza, E. Campana, G. Gennari, and G. Raccanelli, "Building a normality space of events - a pca approach to event detection," in *Proc. of the 3rd International Conference on Computer Vision Theory and Applications (VISAPP2008), Funchal, Madeira, Portugal*, 2008, pp. 551–554.

[4] G. Golub and C. VanLoan, *Matrix Computations*. Baltimore: Johns Hopkins University Press, 1996.

[5] B. Anderson and J. Moore, *Optimal Filtering*. Englewood Cliffs, NJ,: Prentice-Hall, 1979.

[6] A. Jazwinski, *Stochastic Processes and Filtering Theory*. Academic Press - London, 1970.

[7] L. Ljung, *System Identification - Theory For the User*. Prentice Hall, 1999.

[8] T. Söderström and P. Stoica, *System Identification*. Prentice Hall, 1989.

[9] N. Nill, "A visual model weighted cosine transform for image compression and quality assessment," *IEEE Transactions on Communication*, vol. 33, no. 6, pp. 6551–6557, 1985.

[10] A. Eskicioglu and P. Fisher, "Image quality measures and their performance," *IEEE Transactions on Communication*, vol. 43, no. 12, pp. 2959–2965, 1995.

[11] L. Levi, "Vision in communication," in *Progress in Optics*, E. E.Wolf, Ed., vol. 8, 1970, pp. 345–374.

[12] D. Kelly, "Visual contrast sensitivity," *Opt. Acta*, vol. 24, pp. 107–130, 1977.

[13] J. Brailean, B. Sullivan, C. Chen, and M. Giger, "Evaluating the em algorithm for image processing using a human visual fidelity criterion," in *Proc. IEEE Internat. Conf. Acoust. Speech Signal Process*, 1991, pp. 2957–2960.