A Bayesian Approach to Video Expansions on Parametric Over-Complete 2-D Dictionaries

Oscar Divorra Escoda and Pierre Vandergheynst
Signal Processing Institute (ITS)
Swiss Federal Institute of Technology in Lausanne (EPFL)
Ecublens, 1015 Lausanne, Switzerland
Email: {oscar.divorra, pierre.vandergheynst}@epfl.ch

***Abstract*— In this work, we explore a framework for the sparse representation of video sequences by means of spatio-temporal functions able to exploit the 2D nature of images and the temporal smoothness associated to object trajectories. Decomposition over redundant dictionaries formed by 2D functions capable to exploit image geometry, has shown to be well adapted for efficient sparse image approximations. Video representation by means of temporally evolving sets of such 2D functions seems thus a natural extension toward video approximation techniques. In the present paper we study the deformation of a geometry oriented image expansion based on Matching Pursuits (MP) [1], to obtain a parametric representation of frames transformation through time. We consider a modified MP approach based on a Bayesian decision criteria to deform geometrical primitives in a predictive fashion from frame to frame. Indeed, since motion stability is not guaranteed using a pure MP, a Bayesian framework is introduced to regularize motion among expansion terms of frames representations.**

I. Introduction

Video representations are often based on a signal model where objects and regions follow smooth geometrical transformations through time [2]. Considering images as sets of regions separated by contours (the piecewise image model [3]) makes geometry a key component of the information contained in natural images. Studies based in such assumptions underline the importance of geometrically adapted image representations [4], [3], [5]. By extension, excepted from covering and uncovering effects, sequences can be seen as sets of geometrical 2D primitives that evolve through time. Thus, meaningful representations capable of describing geometry in sequences, should benefit from 2D adaptivity while tracking at the same time temporal transformation. In this work, we study the deformation of a geometry oriented image expansion based on Matching Pursuits (MP) [1], [6] to obtain a parametric representation of frames transformation through time. We consider a modified MP approach based on Bayesian decision criteria to deform geometrical primitives through time. A Markov Random Field (MRF) framework is introduced to regularize motion among terms of frame decompositions. Results present the behavior of a greedy algorithm together with the use of redundant dictionaries to track the temporal evolution of 2D atoms. This is done for two cases, i) without considering *a priori* information in the greedy selection criteria and ii) considering it. Finally a Rate-Distortion measurement is performed to evaluate the decrease of entropy of the sequence parametric description versus the increase of approximation drift introduced by the modified greedy criteria.

This work was supported by the Swiss Federal Office for Education and Technology grant number 6044.1 KTS

The paper is structured as follows: In Section II the investigated framework for image and video representation is exposed. Section III introduces the Bayesian framework used in combination with a greedy algorithm. Results are presented in Section IV. Finally, conclusions are drawn in Section V.

II. Video Expansions Over Redundant Parametric Dictionaries

A. Image Modeling

Efficient modeling of the wide variety of geometrical features on images suggest the use of highly redundant sets of functions. In this paper we will model images as short (i.e. sparse) linear superpositions of atoms taken out from a huge, usually very redundant, library ($\mathcal{D}$) of functions g_γ usually referred to as a dictionary. Hence,

$$\hat{I} = \sum_{\gamma_n \in \Omega} c_{\gamma_n} \cdot g_{\gamma_n}, \tag{1}$$

where n is the summation index, c_γ corresponds to the projection coefficient for every atom g_γ and Ω is the subset of selected atom indexes from dictionary $\mathcal{D}$. In order to adapt our representation to catch the geometry of natural images, $\mathcal{D}$ is defined as $\mathcal{D} = \{g_\gamma : \gamma \in \Gamma\}$. Each atom $g_\gamma = U_\gamma g$ where U_γ is a geometrical transformation applied to the mother function g. This transformation consists in translation on the plane, anisotropic scaling and rotation. Anisotropic Refinement (AR) atoms [1] have been chosen as generating function g due to their geometry oriented structure.

B. Proposed Video Representation

We consider an approach where 2D spatial primitives g_γ^t (where t indicates the temporal dimension) obtained in the expansion of a reference frame of the form of (1) are tracked through time from frame to frame. Indeed, we would like to jointly represent image geometrical structures *and* their temporal evolution.

First, a reference frame is decomposed on the geometric redundant dictionary described above. Given the non uniqueness of the possible expansions and the computational complexity needed to find the optimal one, we choose to iteratively approximate the reference frame by means of the sub-optimal Matching Pursuit algorithm [6]. This retrieves at each step the atom g_γ that best approximates the signal in a L_2 norm sense. Once a new expansion term has been found, MP subtracts this from the signal in order to proceed in the following iteration to retrieve a new term from the residual.

Given a set of images belonging to a sequence, the changes suffered from frame I_t to I_{t+1} are modeled as the application of an operator F to the image I_t such that $I_{t+1} = F_t(I_t)$ and $I_{t+1} = \sum_{\gamma_n \in \Gamma} F_t^{\gamma_n}\left(c_{\gamma_n}^t \cdot g_{\gamma_n}^t\right)$

This poses a complex optimization problem to solve:

$$\min_{F_t} \left\| I_{t+1} - \sum_{\gamma_n \in \Omega} F_t^{\gamma_n}\left[c_{\gamma_n}^t \cdot g_{\gamma_n}^t\right] \right\|_2 \quad \text{subject to} \quad \text{Cost}(F_t) \leq \xi, \tag{2}$$

where F_t represents the set of transformations F_t^γ of all atoms that approximate each frame, and Cost represents a given constraint subject to the sequence model. In order to make it feasible, an approximate solution is obtained by the use of a greedy algorithm. Every frame is expanded by means of a modified matching pursuit. A similar approach to the used for the reference frame is used to retrieve the new set of g_γ^{t+1} (and the associated parametric transformation F_t). However, at every greedy decomposition iteration some new criteria needs to be considered in order to establish the relation with the expansion of the reference frame. First, only a subset of functions of the general dictionary is considered as candidate functions to represent the deformed atom. This subset is defined according to the past geometrical features of the atom in the previous frame, such that only a limited set of transformations (translation, scale and rotation) are possible. As shown in Sec. IV, the simple constraint of limiting possible atom transformations, and the simplicity of dictionary functions [1] turns into a lack of regularity (stability) of the atom motion. The use of some *a priori* information in the selection criteria of the greedy algorithm is considered in the following section. Indeed, we will assume that atoms that contribute to the same structure in an image cannot have very diverse motion.

III. A Bayesian Approach for Tracking Image Primitives

A. Weak MP Optimization Functional

In order to include in the MP algorithm a regularity measure, a more flexible version of the selection criteria is considered (Weak Greedy Algorithm -WGA- [7]). Instead of selecting the function giving the biggest scalar product at every iteration, we select the most probable function with respect to a certain motion. The selection of the atom that gives the maximum scalar product is equivalent to select the most probable atom given that all transformations have equal *a priori* probability. However, in the case of smooth motion, there will be a lot of transformations that are unlike and even impossible. Hence, at every greedy prediction iteration the atom selected will correspond to:

$$g_{\gamma_n'}^{t+1} = \arg\max_{g_{\gamma_n}^{t+1},\ \gamma \in \Gamma} \left\{ p\left(\mathcal{R}_n^{t+1} f, g_{\gamma_n}^t \mid \Delta\gamma_n, \Delta c_n\right) \cdot p\left(\Delta\gamma_n, \Delta c_n\right) \right\}, \tag{3}$$

such that $\Delta\gamma_n$ is the temporal parameter variation of the nth term of the modified MP expansion, and $\gamma_n^{t+1}, \gamma_n^t \in \Gamma$. The first probability term expresses matching probability of a transformed atom, from frame at time t into the frame at time $t+1$, constrained to a given motion (change in translation, rotation scale and projection coefficient) of an atom. The second probability term introduces in the functional the *a priori* knowledge based on a MRF of the possible transformations, i.e. some transformations will be more likely than others. This establishes a relation between nearby atoms transferring the more reliable motion obtained from higher energy atoms (first in the MP expansion) to weaker ones.

The matching probability $p\left(\mathcal{R}_n^{t+1} f, g_{\gamma_n}^t \mid \Delta\gamma_n, \Delta c_n\right)$ can be defined as a function of an estimated residual error energy $\left\|\hat{\mathcal{R}}_{n+1}^{t+1} f\right\|^2$ for the retrieval of function g_{γ_n} at iteration n. Atoms are assumed to deform under consistent motion transformation. Thus, no change in the coefficient will be considered (except for scale changes) in the estimation of the most probable motion:

$$\hat{\mathcal{R}}_{n+1}^{t+1} f = \mathcal{R}_n f^{t+1} - \overline{\left\langle \mathcal{R}_n^t f, g_{\gamma_n}^t \right\rangle} g_{\gamma_n}^{t+1}, \tag{4}$$

where $\overline{\left\langle \mathcal{R}_n^t f, g_{\gamma_n}^t \right\rangle}$ is normalized according to a possible rescaling of $g_{\gamma_n}^{t+1}$ with respect to $g_{\gamma_n}^t$.

Assuming Gaussianity (by the central limit theorem [8]) and independence of error samples $\mathcal{R}_{n+1}^t f(x,y)$ [9], the following conditioned optimization criteria can be proved:

$$p\left(\mathcal{R}_n^{t+1} f, g_{\gamma_n}^t \mid \Delta\gamma_n, \Delta c_n\right) \approx \frac{C_1}{\sqrt{\left\|\hat{\mathcal{R}}_{n+1}^{t+1}\right\|^2}}, \tag{5}$$

where C_1 is a constant.

The probability $p(\Delta\gamma_n, \Delta c_n)$ imposes the model that constraints the transformation $F_t^{\gamma_n}$ of $g_{\gamma_n}^t$ and the associated coefficient. Earlier atoms are trusted to generate the MRF for the future appearing atoms. Anyway, when no *a priori* indicator of the motion of a primitive is available, an initial tentative needs to be performed. The functions in use for the generation of our dictionary have a relatively simple shape. Similarly to the well known "aperture" problem, AR atoms may not be able to retrieve the appropriate translation in the direction parallel to contour gradients (usualy represented by the smooth part (Gaussian) of AR atoms). In addition, possible additional influences due to greedy sub-optimality [10] can be a problem as well. Thus, the whole pixmap of the original image inscribed in the support of that primitive is used for

a first estimate. This is, the cross-correlation (matching) for every possible geometric transformation of the zero mean and normalized versions of the pixmap and the frame that we want to approximate is used.

$\Delta\gamma_n$ components ($\Delta\vec{d}$, $\Delta\vec{s}$, $\Delta\theta$), considered to be independent random variables, and the temporal variation of the coefficient (Δc_n) are dependent on the temporal transformation $\Delta\gamma_n$. Their probability function is assumed to be of the form of a MRF. That is, they may be modeled by a Gibbs distribution [11], which allows to transform (3) into:

$$\Delta\gamma_n = \arg\min_{\Delta\gamma_n}\left\{\frac{1}{2}\log\left(\left\|\hat{\mathcal{R}}_{n+1}^{t+1}\right\|^2\right) + \lambda_{\Delta c_n}E_{\Delta c_n}(\Delta c_n) + \lambda_{\Delta\vec{d}_n}E_{\Delta\vec{d}_n}\left(\Delta\vec{d}_n\right) + \lambda_{\Delta\vec{s}_n}E_{\Delta\vec{s}_n}(\Delta\vec{s}_n) + \lambda_{\Delta\theta_n}E_{\Delta\theta_n}(\Delta\theta_n)\right\} \tag{6}$$

where $E_x(x)$ is a potential function that characterizes the MRF and how neighboring variables are related and each λ_x configures the contribution to the functional of each one of the terms. All λ_x are related to the statistics of the assumed Bayesian model and will be sequence dependent.

Temporal variations of coefficients Δc_n should be small in ideal tracking of a primitive. In any case, coefficients may not change sign. Changes to coefficients should be driven mainly due to the change of scale of the approximating function. To induce its temporal regularity, a normalized quadratic distance between the coefficients at time t and $t+1$ is considered for $E_{\Delta c_n}(\Delta c_n)$. Displacement, change of scale and rotation potentials, are measured as the euclidean distance between the value under test and the most likely (ML) transformation estimated from previous MP iterations at every image location. Hence, they can be represented as:

$$\begin{aligned} E_{\Delta\vec{d}_n} &= (d_x^n - \hat{d_x}^n)^2 + (d_y^n - \hat{d_y}^n)^2 \\ E_{\Delta\vec{s}_n} &= (s_x^n - \hat{s_x}^n)^2 + (s_y^n - \hat{s_y}^n)^2 \\ E_{\Delta\theta_n} &= (\theta^n - \hat{\theta}^n)^2, \end{aligned} \tag{7}$$

where $\hat{d}$, $\hat{s}$ and $\hat{\theta}$ correspond to the ML estimates. These estimates are nothing else than a weighted average of the deformation of previous atoms that overlap in the spatial location where this potentials are defined.

Finally, notice that an atom may become unuseful due to changes in the image sequence. A threshold is defined on the projection coefficients in order to detect it. These atoms are then reintroduced in the frame description by means of a full search without taking into account any kind of regularization.

IV. Experimental Results

In this section an example based in a synthetic sequence and one based in a natural one are presented. The examples show mainly the MRF translational fields obtained in the process of parametric representations. These evidence the effect of regularization on the parametric descriptions. An additional result on the natural sequece Foreman is presented to show the effect of regularization in the prediction drift and the average R-D.

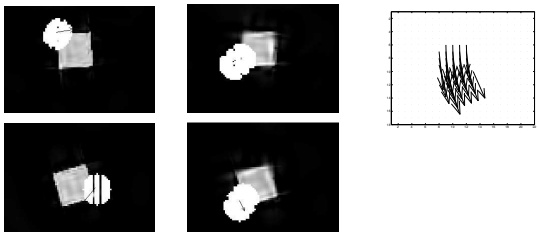

Fig. 1. Affine motion of a synthetic model (square). The white bat corresponds to the foot-print of a selected atom in two temporal instants. Left is the non-regularized prediction. Middle is the regularized prediction. Right most reliable motion of the regularized solution. Rotation and displacement can be appreciated.

The example corresponding to the motion associated to a particular atom in a synthetic sequence can be found in Fig. 1. The sequence corresponds to a translating and rotating square. We consider a particular atom, represented in the picture by a white mark that has the shape of its support. In both columns, we see the representation of the square by means of an expansion of 50 coefficients with the footprint of the function support superimposed. In the left column we display the corresponding past and present positions of the atom for the non regularized case, i.e. the selected atom is fully driven by the search of the highest projection coefficient absolute value. On the right, the atom is steered considering the *a priori* of rigid motion. At the bottom of Fig. 1 we can see the motion associated to atoms of the right column. However the synthetic model considered above is very simple and constrained. Fig. 3 shows a comparison between a non regularized result (left), and a regularized one (right) of the highway image. A clear influence of the regularization and motion initialization is reflected in the flow related to the atoms motion. In the upper right figure and the one below, a clear relation can be established between atoms that participate in the cars approximation and their motion. In the example where the truck appears, the influence between neighboring atoms located in the wood area in the background can not be avoided, i.e. the moving atoms of the truck *push* in some measure the atoms representing the background. Interdependence among neighboring primitives is responsible for their strong interaction. In order to have an objective measure of the regularization effects, we consider the R-D curve obtained for a simple coding scheme applied to the parametric representation of the Foreman sequence. For this purpose, we apply the simple coding scheme described in [12] to a group of 3 GOPs of 16 frames (a reference frame will be inserted at the beginning of each GOP). Given the regularized atom prediction criteria of our algorithm, spatial and temporal regularity are imposed among atoms. Hence, correlation of atoms at time t with their evolved version at time $t+1$ will be exploited by only encoding the temporal differences of parameters and coefficients (i.e. the set of F_t^γ of Eq.2). When an atom is refreshed , this is obtained by doing a full search in the whole image. Atoms that have been refreshed

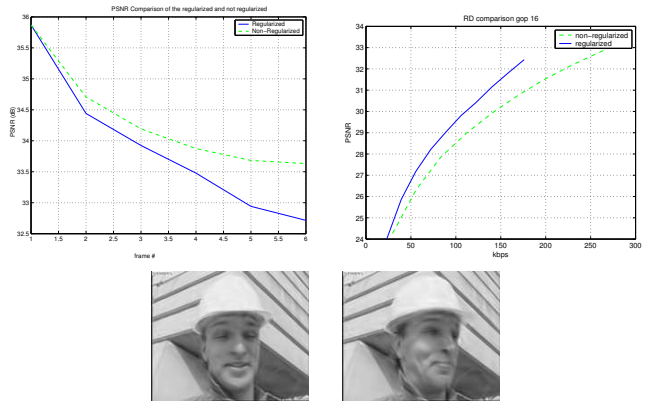

Fig. 2. Up (left): Curves representing the loss of frame approximation accuracy due to the regularization. Up (right): R-D comparison of the regularized and non-regularized Foreman sequences (16 frames/GOP). Rate variation is obtained by changing the number of terms (atoms/frame) considered in the sequence reconstruction. Down: Images corresponding to the reconstructed frames 1st and 6th from the predicted foreman sequence using 500 atoms/frame.

will also be coded by just sending the difference with respect to the atom they replace in the previous frame. Finally, an arithmetic coding of the differential data is performed. The curves on Fig. 2 show the gain obtained in terms of R-D of the regularized Bayesian matching with respect to the non-regularized one. As expected, regularization turns into a reduction of the entropy of the parametric representation of frame to frame variations. Furthermore, reduction in entropy is high enough to compensate in average the drift (i.e., WGA trades between regularity and signal approximation) presented by the regularized sequence with respect to the non-regularized.

V. Conclusions

The results show light on the possibility of tracking geometrical primitives through sequences using over-complete geometrically oriented dictionaries. The use of simple matching pursuits to track the transformation of primitives from frame to frame is revealed to generate very instable parametric sequence descriptions despite that signal approximations may be good. Experimental results and theory [10] justify the need for *a priories* to help (or fully drive) the decision criteria of the greedy algorithm when general over-complete dictionaries are in use. Additional effort is required to define more robust approaches involving structured dictionaries where sets of neighboring atoms could be considered to move together in a rigid way. Furthermore, suboptimalities should be avoided by the retrieval of a global optimum for the optimization problem of Eq. 2.

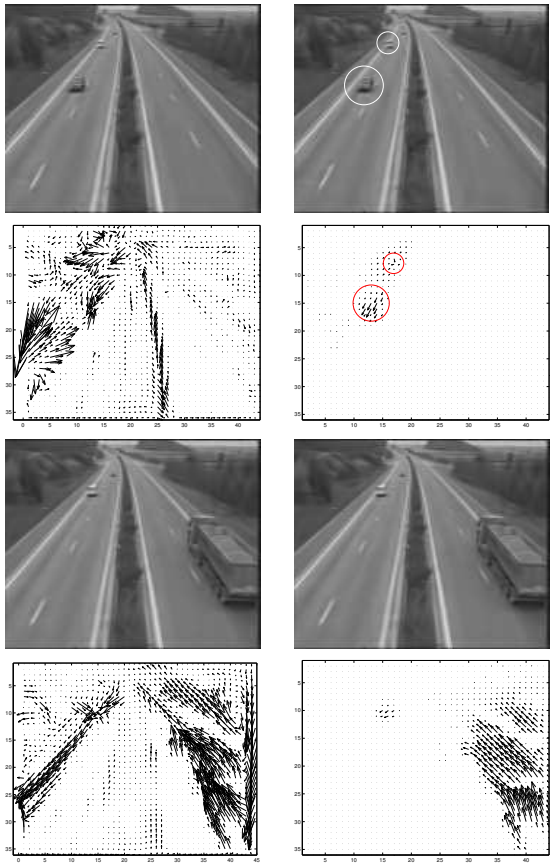

Fig. 3. Natural sequence motorway. Left column: non-regularized solution. Right column: regularized tracking. First and third rows: Respective reconstructions with 500 atoms. Second and forth rows: Most reliable primitives motion

References

[1] P. Vandergheynst and P. Frossard, "Efficient image representation by anisotropic refinement in matching pursuit," in *ICASSP*, vol. 3, Salt Lake City, May 2001.

[2] Y. Wang, J. Ostermann, and Y. Zhang, *Digital Video Processing and Communications*. Prentice Hall, 2001.

[3] M. N. Do, P. L. Dragotti, R. Shukla, and M. Vetterli, "On the compression of two-dimensional piecewise smooth functions," in *ICIP*, Thesalonica, October 2001.

[4] E. J. Candès and D. L. Donoho, "Curvelets - a surprisingly effective nonadaptive representation for objects with edges." *Curves and Surfaces, L. L. S. et al., ed., Nashville, TN, (Vanderbilt University Press)*, pp. 123–143, 1999.

[5] R. Figueras i Ventura, L. Granai, and P. Vandergheynst, "R-D analysis of adaptive edge representations," in *MMSP*, Virgin Islands, December 2002.

[6] S. G. Mallat and Z. Zhang, "Matching pursuits with time-frequency dictionaries," *IEEE Trans. on Signal Proc.*, vol. 41, no. 12, pp. 3397–3415, December 1993.

[7] V. N. Temlyakov, "Weak greedy algorithms," Department of Mathematics, University of South Carolina, Columbia, Tech. Rep., 1999.

[8] A. Papoulis, *Probability, Random Variables, and Stochastic Processes*, 3rd ed. McGrawHill, 1991.

[9] T. Aach, A. Kaup, and R. Mester, "Combined displacement estimation and segmentation of stereo image pairs based on Gibbs random fields," in *ICASSP*, 1990.

[10] O. Divorra Escoda, P. Vandergheynst, and M. Bierlaire, "Video representation using greedy approximations over redundant parametric dictionaries." LTS-2/ITS EPFL, Tech. Rep. ITS-2004.019, 2004, online: http://lts2www.epfl.ch/ divorra/publications.php.

[11] S. Geman and D. Geman, "Stochastic relaxation, gibbs distributions, and the bayesian restoration of images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 6, pp. 721–741, 1984.

[12] O. Divorra Escoda and P. Vandergheynst, "Video coding using a deformation compensation algorithm based on adaptive matching pursuit image decompositions," in *ICIP*, Barcelona, September 2003.