TIME-VARYING LINEAR AUTOREGRESSIVE MODELS FOR SEGMENTATION

Charles Florin[1], *Nikos Paragios*[2], *Gareth Funka-Lea*[1], *James Williams*[3]

[1]Imaging & Visualization Department, Siemens Corporate Research, Princeton, NJ,
[2]MAS - Ecole Centrale de Paris, Chatenay-Malabry, France,
[3]Siemens Medical System, AX Forchheim, Germany

ABSTRACT

Tracking highly deforming structures in space and time arises in numerous applications in computer vision. Static Models are often referred to as linear combinations of a mean model and modes of variation learned from training examples. In Dynamic Modeling, the shape is represented as a function of shapes at previous time steps. In this paper, we introduce a novel technique that uses the spatial and the temporal information on the object deformation. We reformulate tracking as a high order time series prediction mechanism that adapts itself on-line to the newest results. Samples (toward dimensionality reduction) are represented in an orthogonal basis, and are introduced in an auto-regressive model that is determined through an optimization process in appropriate metric spaces. Toward capturing evolving deformations as well as cases that have not been part of the learning stage, a process that updates on-line both the orthogonal basis decomposition and the parameters of the autoregressive model is proposed. Experimental results with a nonstationary dynamic system prove adaptive AR models give better results than both stationary models and models learned over the whole sequence.

Index Terms— Tracking, Segmentation, autoregressive

1. INTRODUCTION

Motion perception is a fundamental task of biological vision, with motion estimation and tracking being the most popular and well-addressed applications. To this end, given a sequence of images, one would like to recover the 2D temporal position of objects of interest. These applications often serve as input to high-level vision tasks like 3D reconstruction.

Tracking non-rigid objects is a task that has gained particular attention in computational vision, in particular with Kalman Snakes [1] and multiple hypotheses trackers [2, 3]. Constraints/models are imposed in the temporal evolution of the target and prediction mechanisms are used to perform tracking. Shape tracking with autoregressive dynamic models is a step forward in this direction, with different shape spaces being investigated. In [4], a first-order model is used to track cardiac cycles in echocardiographic sequences, while in [5] Fourier descriptors are used to describe shapes, and a linear dynamic model tracks their evolution on time. Tracking articulated structures is a well suited problem for autoregressive models and therefore in [6] a method based on a linear dynamic model is proposed. In [7], a static autoregressive model was developed to produce a prior for levelset-based segmentation. The main limitation of such models refers to their *time-invariant* nature. Consequently, either a complex heuristic is developed to mix models, or Markov fields are introduced for multimodality.

To address this issue, adaptive dynamic models such as adaptive Kalman Filter [8] integrate a Kalman Filter that estimates additive noise properties (mean and variance) to an AR model. Except for the noise properties, no adaption is made for the signal itself if its own properties change (e.g. new properties cannot be captured by the current feature space) or if the system dynamism (the regression) is modified. On the other hand, in the context of classification and learning, adaptive feature spaces such as adaptive Principal Component Analysis (PCA) [9][10] were developed to take into account the newest results to estimate the feature space. Nevertheless, even if the feature space is adapted to the most recent exemplars, in the context of segmentation and tracking, one also needs an adaptive predictive model to relate segmentation results across time, and to adapt the prediction scheme if changes in the dynamic system occur.

In this paper, we address such limitations and determine a predictive model that is incrementally adapted to changes both in the system dynamics (and not only the noise properties) and in the feature space. For that purpose, we propose an on-line technique for tracking based on higher order autoregressive models. Such a technique is based on dimensionality reduction of the parameters space using an orthogonal decomposition of the training set. Then, a linear autoregressive model is built in this space capable of predicting current states from the prior ones. Such a model and its feature space (the orthogonal decomposition of shapes) are updated on-line using new evidence. To this end, a proper geometric distance is used in a robust framework to optimize the parameters of the model.

The reminder of this paper is organized in the following fashion: in [SEC. (2)] we briefly present autoregressive models. Dimensionality reduction, and on-line learning are part of section [SEC. (3)], while tracking is presented in [SEC. (4)]. Results and discussion conclude the paper.

2. LINEAR AUTOREGRESSIVE MODELS

Time series models are very popular in a number of domains like signal processing. Let us assume a set of temporal observations $\mathbf{Y} = \{\mathbf{Y}_t; t \in [0,T]\}$, where each observation $\mathbf{Y}_t \in \Omega$ is a column vector of the N-dimension observation space Ω. Linear autoregressive models - of order k - consist of expressing the current observation as a combination of previous samples perturbed by some noise model:

$$\mathbf{Y}_t = \mathbf{H}\left[\mathbf{Y}_{t-1}^T\ \mathbf{Y}_{t-2}^T\ ...\ \mathbf{Y}_{t-k}^T\right]^T + \eta(\mu,\Sigma) \quad (1)$$

with N-by-kN matrix $\mathbf{H}$ called the *prediction matrix* and $\eta(\mu,\Sigma)$ being the *noise model* vector. For any matrix $\mathbf{M}$, $\mathbf{M}^T$ denotes the matrix transpose. In the most general case, one assumes that the input variable $\mathbf{Y}$ is defined in a high-dimensional space, and therefore some dimensionality reduction is to be performed. Without loss of generality, we assume either a linear or non-linear operator $\phi()$ (defined later in [SEC. (3.1)]) defines new observations of lower dimension $\mathbf{X} = \phi(\mathbf{Y})$. We further assume that such an operator is invertible, or otherwise stated: from a feature vector $\mathbf{X}$ one recovers the original observation $\mathbf{Y}$. In that case one restates the autoregressive model in a lower dimensional space, and [EQ. (1)] becomes:

$$\mathbf{X}_t = \mathbf{H}_\phi\left[\mathbf{X}_{t-1}^T\ \mathbf{X}_{t-2}^T\ ...\ \mathbf{X}_{t-k}^T\right]^T + \eta_\phi(\mu,\Sigma) \quad (2)$$

The estimation of such a model is done from a set of training examples and robust regression. Let us assume that $T >> k$ observations are available. Once such observations have gone through dimensionality reduction, we obtain an over-constrained linear system written:

$$\forall\, t \in [|k,T|],\ \mathbf{X}_t \leftarrow \hat{\mathbf{X}}_t = \mathbf{H}_\phi\left[\mathbf{X}_{t-1}^T\ \mathbf{X}_{t-2}^T\ ...\ \mathbf{X}_{t-k}^T\right]^T \quad (3)$$

The unknown parameters of such an over-constrained system is determined through a robust least square minimization:

$$\mathbf{H}_\phi = \text{argmin}_{\mathbf{H}}\left\{\sum_{t=k}^{T}\rho_\phi\ \ \mathbf{X}_t, \hat{\mathbf{X}}_t\ \right\} \quad (4)$$

where $\hat{\mathbf{X}}_t$ is the $\mathbf{H}_\phi$-predicted state vector, and ρ_ϕ is an error metric, in the observation space. Since the reduced space is potentially highly non-uniform (in the case of PCA, this means the variations along one mode are larger than the others), performing the minimization in the observation space greatly reduces the prediction error. The Euler-Lagrange equations of such a system lead to a linear problem that is solved in a straightforward fashion. The number of constraints used in such a procedure is determined off-line using the Schwartz's bayesian criterion.

The choice concerning ρ_ϕ is described in [SEC. (3.2)], and depends on the shape representation that is selected. Once such a metric has been defined and an important number of samples is present, one obtains the prediction matrix through mathematical inference.

Fig. 1. Registered training examples used for initial Principal Components Analysis.

3. SHAPE REPRESENTATION & METRIC

3.1. Contour Representation in a Low Dimension Space

When no topology constrains are given, implicit methods are popular shape representations. Let us consider a number of training examples to track $s = \{s_i, i \in [1,n]\}$. In [11] a distance transform representation ψ_i is considered for a given shape s_i, as explained in [EQ. (5)]:

$$\forall \mathbf{x} \in \Omega,\quad \psi_i(\mathbf{x}) = \begin{cases} 0, & \mathbf{x} \in s_i \\ +D(\mathbf{x}, s_i) > 0, & \mathbf{x} \in \Omega_i \\ -D(\mathbf{x}, s_i) < 0, & \mathbf{x} \in [\Omega - \Omega_i] \end{cases} \quad (5)$$

where Ω defines the image domain, and $D(\mathbf{x}, s_i)$ the Euclidean distance between point $\mathbf{x} \in \Omega$ and the exemplar's contour s_i. We call abusively "shape" the distance function ψ. Global registration between shapes is now performed by determining the affine transformation $\mathcal{A}$ that minimizes the integral of squared difference between the alignment shape's distance function and the reference distance function. The resulting aligned function is then represented by a column vector $\mathbf{Y}_i$ of dimension N after discretizing the image domain Ω with N control points.

Since N is usually too large for the computation of $\mathbf{H}_\phi$ in [EQ. (4)], Principal Component Analysis (PCA) is applied for an efficient dimensionality reduction. PCA refers to a linear transformation of variables that retains - for a given number m of operators - the largest amount of variation within the training data. Without loss of generality, a zero mean assumption is considered for the $\{\mathbf{Y}_i\}$ by estimating the mean vector $\overline{\mathbf{Y}}$ and subtracting it from the training samples $\{\mathbf{Y}_i\}$. The N-by-N covariance matrix $\overline{\Sigma} = \sum_{i=1}^{n}\mathbf{Y}_i\mathbf{Y}_i^T$ associated to the n training vectors $\mathbf{Y}_i$ is used for an Eigendecomposition. The N Eigenvectors $\mathbf{U}_q$ form an orthonormal basis onto which the vectors $\mathbf{Y}_i$ are projected. Only the m Eigenvectors associated to the highest Eignvalues are kept, so that the operator ϕ of [SEC. (2)] is defined by the affine transformation $\mathcal{A}$ and the projection from the N-dimension space to the m major Eigenvectors, and is invertible if one approximates the $N-m$ smallest Eigenvalues by 0. The projected vector $\phi(\mathbf{Y})$ is defined by the coefficients $\Lambda = \{\lambda_q\}_{q=1..m}$ so that:

$$\mathbf{Y}_i = \mathcal{A}(\overline{\mathbf{Y}}) + \sum_{q=1}^{m}\lambda_q\,\mathbf{U}_q. \quad (6)$$

Let us note $\mathbf{X}$ the feature vector $[\mathcal{A}, \Lambda]$ related to $\mathbf{Y}$.

3.2. Batch Learning & Euclidean Distance Metric

PCA decreases the problem's dimensionality leading to a highly non-uniform feature space (for the range of translation component is far superior to the one of the scale). In order to overcome such a limitation, we propose to use a metric defined in the original space (i.e. the observation space Ω) to recover the prediction mechanism in the reduced space (i.e. affine transformation from [SEC. (2)] and linear factors from [EQ. (6)]). The simplest metric between two level-sets is the L_2 norm between the two distance functions that correspond to the observation $\mathbf{Y}_t$ and the prediction $\hat{\mathbf{Y}}_t$.

$$\rho_\phi(\mathbf{X}_t, \hat{\mathbf{X}}_t) = \int_\Omega \left\| \mathbf{Y}_t(\mathbf{x}) - \hat{\mathbf{Y}}_t(\mathbf{x}) \right\|^2 d\mathbf{x} \quad (7)$$

refers to a well behaved distance between observations, and predictions and implicitly accounts for the range of parameters of the autoregressive model. This guarantees that the feature space and autoregressive models are optimum for the L_2 norm for the training set. However, in order to capture changes in shape or varying regression, an explicit on-the-fly update scheme is required.

Once new observations have been introduced to the process, the prediction matrix as well as the orthogonal basis are to be updated. Incremental principal component analysis is used for the basis, while an exponential forgetting method is more suitable for the prediction matrix.

4. ON-LINE ADAPTATION OF THE MODEL

4.1. Adaptation of the Orthogonal Basis using Incremental PCA

Incremental PCA[9, 10] consists of adding the latest observation to the PCA learning set. Thus, a new feature space is to be used to represent the state decomposition $\mathbf{X}$. Using these new variation modes, and the corrected state $\hat{\mathbf{X}}_t$, the transition model is then updated and ready to be used to predict the following state $\mathbf{X}_{t+1}$. The method presented in [9] can be summarized as follows: given a PCA at time t-1, mean $\overline{\mathbf{Y}}_{t-1}$, a set of eigenvectors $\mathbf{U}_{t-1} = [u_i]$, and their corresponding eigenvalues $\mathbf{D}_{t-1} = \text{diag}\,(d_1, d_2, ...)$, given a new state $\mathbf{Y}_t$, the PCA is updated at time t starting by the mean:

$$\overline{\mathbf{Y}}_t = \frac{(t-1)\overline{\mathbf{Y}}_{t-1} + \mathbf{Y}_t}{t}. \quad (8)$$

The eigenvector matrix is updated in a similar way (details can be found in [9]).

4.2. Adaptation of the Predictive Model

Once the prediction matrix has been estimated, new observations are introduced in the system toward decreasing the prediction error. To this end, one would like to find the lowest potential of

$$E_T(\mathbf{H}_\phi) = \min_{\mathbf{H}_\phi} \left\{ \sum_{t=k}^{T} \rho_\phi(\mathbf{X}_t, \mathbf{H}_\phi \left[\mathbf{X}_{t-1}\ \mathbf{X}_{t-2}\ ...\ \mathbf{X}_{t-k}\right]) \right\} \quad (9)$$

prior method	(1) Stationary AR		(2) Adaptive AR	
(a)	correctly seg.	95.20 %	correctly seg.	96.19 %
	under-seg.	2.76 %	under-seg.	2.32 %
	over-seg.	2.03 %	over-seg.	1.49 %
(b)	failed		correctly seg.	96.66 %
			under-seg.	2.08 %
			over-seg.	1.26 %
(c)	failed		correctly seg.	96.04 %
			under-seg.	1.47 %
			over-seg.	2.48 %

Table 1. Percentage of correctly segmented, oversegmented and undersegmented pixels for diverse prior, same energy. Method (1) uses the stationary AR, learned from the first frames. Method (2) uses the Adaptive AR described in this paper. Dataset (a) is the original dataset. Dataset(b) is the original dataset with a horizontal occlusion, and Dataset (c) presents a vertical occlusion.

In [12], the result is obtained by dividing the sum of squares into blocks, solving the problem for the first block and using this result as initialization once the following block is added to the previous block. Unlike the method presented here, [12] solved the Gauss-Newton iterations using Extended Kalman Filter for nonlinear measures $E(\mathbf{H}, \mu, \Sigma)$. Experiments have shown that few (a couple of dozen) Gauss-Newton iterations are required to achieve far better results than a simple time-invariant dynamic model. For non-linear time processes, the local approximation of $(H_{T+1}, \mu_{T+1}, \Sigma_{T+1})$ may not well correspond to the state transition in a very early time step. For that reason, *exponential forgetting* is introduced by multiplying the sum's terms in [EQ. (9)] with exponential weights $w_t = e^{-t/\tau}$, where τ is the *exponential forgetting* window size. The smaller τ the more reactive but also the more sensitive to noise is the non-stationary autoregressive model.

5. RESULTS & DISCUSSION

5.1. Comparison with stationary AR models

For comparison purposes we use the same dataset with and without digital occlusions, which shows a man silhouette walking and then running, and test different priors for level-sets [13] evolving according to the same energy. This energy corresponds to the sum of a data-driven term (a histogram-based Chan & Vese functional [14]), and a term associated to the shape prior provided by the dynamic model. The silhouette moves in front of a uniform light colored background so that segmentation errors do not interfere (one is not interested in the segmentation quality so much as in the dynamic system itself). Furthermore, to properly test the model adaptability, the system dynamic has to change; therefore the silhouette walks then runs. The first experiment consists in training the AR-PCA model on the whole sequence (58 frames); this technique loose track of the silhouette in all three cases (unoccluded, vertical occlusion and horizontal occlusion). The second experiment compares the results (see [TABLE (1)])

(a)

(b)

Fig. 2. (a) Vertical and (b) horizontal occlusions added to the original dataset

between the stationary AR model and the adaptive model described in the paper, when both are initialized with the first 18 frames. The stationary model learns the dynamic of the walking pace, but is unable to sustain the dynamic changes when the silhouette starts to run. On the contrary, the adaptive model learns these new dynamic on-line, and is able to make correct predictions. This proves the non-stationary AR model is more suitable to this problem than stationary models and even models trained on the whole sequence.

5.2. Robustness to occlusion

The main drawback one expects from the locally adaptive method is the potential accumulation of errors. To test that, we introduce digital occlusions, see [FIG (2)] (one horizontal occlusion that covers one third of the character during 20 frames, and then one vertical of the same width as the character) of the background mean color, and run the tracking scheme with stationary and adaptive priors. Once again, the results demonstrate that the adaptive model sustains these occlusions. Nevertheless, for larger occlusions, errors accumulate and the tracking is lost.

5.3. Discussion

While using available technology, the non-stationary approach totally changes the scope of AR models for shape priors. The present method benefits from the same advantages as stationary models, described in [7], but does not rely on any stationary assumptions. Datasets with time-varying dynamic, that stationary ARs are not able to track, are now successfully processed by the locally adaptive AR model. Furthermore, to some extent, the non-stationary models handle occlusions and missing data. To sustain larger occlusions, one might think of an occlusion detection scheme and a special heuristic to handle them to put the adaptation of the model on hold. Another possibility to increase robustness and reactivity, with the same Gaussian noise assumption, is to use the framework provided by Kalman Filter. Furthermore, in the case of occlusions, when the Gaussian assumption does not hold, one might be tempted to use a much heavier nonparametric representation for the distribution, such as Particle Filtering. A last interesting perspective might also be to incorporate the quality of the segmentation into the on-line learning (in [SEC. (4)]) to favor the time steps that gave the best results.

6. REFERENCES

[1] D. Terzopoulos and R. Szeliski, "Tracking with Kalman Snakes," in *Active Vision*, A. Blake and A. Yuille, Eds., pp. 3–20. MIT Press, 1992.

[2] M. Isard and A. Blake, "Contour Tracking by Stochastic Propagation of Conditional Density," in *ECCV*, 1996, vol. I, pp. 343–356.

[3] K. Toyama and A. Blake, "Probabilistic Tracking in a Metric Space," in *ICCV*, 2001, pp. 50–59.

[4] J. C. Nascimento, J. S. Marques, and J. M. Sanches, "Estimation of cardiac phases in echographic images using multiple models.," in *ICIP (2)*, 2003, pp. 149–152.

[5] C.-B. Liu and N. Ahuja, "A model for dynamic shape and its applications," in *CVPR (2)*, 2004, pp. 129–134.

[6] A. Agarwal and B. Triggs, "Tracking articulated motion using a mixture of autoregressive models," in *ECCV*, Prague, May 2004, pp. III 54–65.

[7] D. Cremers, "Dynamical statistical shape priors for level set-based tracking," *PAMI*, vol. 28, no. 8, pp. 1262–1273, August 2006.

[8] G. Doblinger, "An adaptive kalman filter for the enhancement of noisy ar signals," 1998.

[9] P. Hall and R. Martin, "Incremental eigenanalysis for classification," in *Proc. British Machine Vision Conference*, 1998, vol. 1, pp. 286–295.

[10] Y. Li, "On incremental and robust subspace learning," *Pattern Recognition*, vol. 37, no. 7, pp. 1509–1518, 2004.

[11] N. Paragios, M. Rousson, and V. Ramesh, "Matching Distance Functions: A Shape-to-Area Variational Approach for Global-to-Local Registration," in *ECCV*, 2002, pp. II:775–790.

[12] D. P. Bertsekas, "Incremental least squares methods and the extended kalman filter," *SIAM J. on Optimization*, vol. 6, no. 3, pp. 807–822, 1996.

[13] S. Osher and N. Paragios, *Geometric Level Set Methods in Imaging, Vision and Graphics*, Springer Verlag, 2003.

[14] T. Chan, B. Sandberg, and L. Vese, "Active Contours without Edges for Vector-Valued Images," *Journal of Visual Communication and Image Representations*, vol. 2, pp. 130–141, 2000.