On-the-fly Training

Javier Melenchón, Lourdes Meler, and Ignasi Iriondo

Enginyeria i Arquitectura La Salle, Universitat Ramon Llull
Communications and Signal Theory Department
Pg. Bonanova 8, 08022, Barcelona, Spain
{jmelen,lmeler,iriondo}@salleurl.edu
WWW home page: http://www.salleurl.edu/eng/elsDCTS/tsenyal/index.htm

Abstract. A new algorithm for the incremental learning and non-intrusive tracking of the appearance of a previously non-seen face is presented. The computation is done in a causal fashion: the information for a given frame to be processed is combined only with the one of previous frames. To achieve this aim, a novel way for simultaneous and incremental computation of the Singular Value Decomposition (SVD) and the mean of the data is explained in this work. Previous developed methods about computing the SVD iteratively are taken into account and a novel way to extract the mean from a factorised matrix using SVD is obtained. Moreover, the results are achieved with linear computational cost and sublinear memory requirements with respect to the size of the data. Some experimental results are also reported.

1 Introduction

The last years have witnessed extraordinary advances in computer and communications technology, leading to an increasing availability of information and processing capabilities of multimedia data [1], [2]. This fact is resulting in a higher and wider demand for easier access to information [3]. On one hand, this information is mainly stored in digital format, so its acces is limited to the user's ability to communicate with computers. On the other hand, it has been remarked the great expressive power of the natural language used in human-human communication, as well as its intrinsic multimodal features [4]. Consequently, the acces to digital information could be carried out using this natural language: reducing the necessity of knowing a specific way to interact with the computer and taking advantage of its expressive features. Moreover, multimodal interfaces with an audio visual system like a talking head could be used in order to speak to the user in natural language. As a result, talking heads used in multimodal interfaces seem to be a proper solution for making acces to information easier and more pleasing for human users.

As explained in [4], multimodal input analysis is necessary when working with multimodal interfaces and relies on interaction devices e.g. facial trackers. Some non-intrusive visual trackers can be used in this sheme because they retain information regarding to position, scale, orientation and appearance of the tracked element, e.g. [5], [6], [7] and [8]. Nevertheless, the whole sequence is needed by these algorithms to be processed off-line (they have a non-causal behaviour); as a result, a real time implementation of these methods is impossible, even without considering their computational cost. This temporal restriction is caused by the computation of a Singular Value Decomposition (SVD) over the whole observed data. Moreover, memory resources are greatly affected by this fact, limiting the duration of the observed sequence. Incremental SVD computation techniques as [9], [10], [11] and [12] may be useful in this case, but they do not take into consideration the mean of the data, which is crucial in the classification of the different gestures. Fortunately, this is taken into account in [13], although it does not propose a method to extract the mean information from a given SVD and it can only update the SVD from two other known SVD.

In this paper, a new method for updating SVD and mean information as well as extracting the mean of the data contained in a given SVD without increasing the cost order of either time or memory is presented in Sect. 2. The application of this new method is carried out in Sect. 3 by a causal algorithm for the tracking and learning of the facial appearance of a person. Experimental results are given in Sect. 4 and concluding remarks are explained in Sect. 5.

2 Incremental SVD with Mean Update

2.1 Fundamentals

The singular value decomposition of matrix $\mathbf{M}_{p\times q}=[\mathbf{m}_1\cdots\mathbf{m}_q]$ is given by:

$$\mathbf{M}_{p\times q} = \mathbf{U}_{p\times p}\mathbf{\Sigma}_{p\times q}\mathbf{V}^T_{q\times q} \ , \tag{1}$$

where $\mathbf{U} = [\mathbf{u}_1 \ \cdots \ \mathbf{u}_p]$ and $\mathbf{V} = [\mathbf{v}_1 \ \cdots \ \mathbf{v}_q]$ are orthonormal matrices; $\mathbf{u}_i$ are the eigenvectors of $\mathbf{M}\mathbf{M}^T$ and span the column space of $\mathbf{M}$; $\mathbf{v}_i$ are the eigenvectors of $\mathbf{M}^T\mathbf{M}$ and span the row space of $\mathbf{M}$; and $\mathbf{\Sigma}$ is a diagonal matrix with the singular values of either $\mathbf{M}\mathbf{M}^T$ and $\mathbf{M}^T\mathbf{M}$ in ascending order. Notice that if $\mathbf{M}$ is a rank r matrix, where $r \leq p$ and $r \leq q$, $\mathbf{\Sigma}$ has only r non-null singular values and (1) can be rewritten as the *thin SVD*: $\mathbf{M}_{p\times q} = \mathbf{U}_{p\times r}\mathbf{\Sigma}_{r\times r}\mathbf{V}^T_{q\times r}$. By the other hand, let $\mathbf{C}_{r\times q} = \mathbf{U}^T_{p\times r}\mathbf{M}_{p\times q}$ be the projections of the columns of $\mathbf{M}$ over the eigenspace spanned by $\mathbf{U}$. Using the *thin SVD* expression the projections matrix $\mathbf{C} = [\mathbf{c}_1 \ \cdots \ \mathbf{c}_q]$ can be written also as $\mathbf{C}_{r\times q} = \mathbf{\Sigma}_{r\times r}\mathbf{V}^T_{q\times r}$.

In other fields, like classification problems pointed by [13], a more suitable representation of $\mathbf{M}$ can be achieved including mean information $\overline{\mathbf{m}} = \frac{1}{q}\sum_{i=1}^{q}\mathbf{m}_i$ in (1), which has to be computed and substracted previously from $\mathbf{M}$ in order to be able to generate the SVD of $\mathbf{M} - \overline{\mathbf{m}} \cdot \mathbf{1}$:

$$\mathbf{M}_{p\times q} = \mathbf{U}_{p\times r}\mathbf{\Sigma}_{r\times r}\mathbf{V}^T_{q\times r} + \overline{\mathbf{m}}_{p\times 1}\mathbf{1}_{1\times q} \quad . \tag{2}$$

2.2 Updating SVD

Assuming an existing SVD (1), if new columns $\mathbf{I}_{p\times c} = [\mathbf{I}_1 \cdots \mathbf{I}_c]$ are added in order to obtain a new matrix $\mathbf{M}'_{p\times(q+c)} = \begin{bmatrix}\mathbf{M}_{p\times q} \ \mathbf{I}_{p\times c}\end{bmatrix}$, the SVD of $\mathbf{M}'$ can be updated from (1) using methods like [10] and [11], achieving:

$$\mathbf{M}'_{p\times(p+c)} = \mathbf{U}'_{p\times r'}\mathbf{\Sigma}'_{r'\times r'}\mathbf{V}'^T_{(q+c)\times r'} \quad . \tag{3}$$

Otherwise, if the representation of $\mathbf{M}'$ is chosen to be as (2) and $\overline{\mathbf{m}}'$ is set to $\frac{1}{q+c}\left(\sum_{k=1}^{q}\mathbf{m}_k + \sum_{l=1}^{c}\mathbf{I}_l\right)$ the SVD becomes:

$$\mathbf{M}'_{p\times(q+c)} = \mathbf{U}'_{p\times r'}\mathbf{\Sigma}'_{r'\times r'}\mathbf{V}'^T_{(q+c)\times r'} + \overline{\mathbf{m}}'_{p\times 1}\mathbf{1}_{1\times(q+c)} \quad . \tag{4}$$

Starting from (2) and matrix $\mathbf{I}$, (4) can be obtained using the method proposed by [13] if the SVD of $\mathbf{I}$ is previously computed and q and c are known beforehand. A new method for updating the SVD and the mean using only the new observations and previous factorization is presented in Sect. 2.3.

2.3 Updating SVD and Mean

Begining with an existing factorization of $\mathbf{M}_i$ as in (5), it is desired to obtain the SVD and mean of $\mathbf{M}_f$ shown in (6):

$$\mathbf{M}_i = \mathbf{U}_i\mathbf{\Sigma}_i\mathbf{V}_i^T + \overline{\mathbf{m}}_i\mathbf{1} \quad . \tag{5}$$

$$\mathbf{M}_f = \begin{bmatrix}\mathbf{M}_i \ \mathbf{I}\end{bmatrix} = \mathbf{U}_f\mathbf{\Sigma}_f\mathbf{V}_f^T + \overline{\mathbf{m}}_f\mathbf{1} \quad . \tag{6}$$

Defining $\hat{\mathbf{M}}_i$ (7) and centering new columns $\mathbf{I}$ around $\overline{\mathbf{m}}_i$ (8), it can be written:

$$\hat{\mathbf{M}}_i = \mathbf{M}_i - \overline{\mathbf{m}}_i \mathbf{1} = \mathbf{U}_i \boldsymbol{\Sigma}_i \mathbf{V}_i^T \ . \quad (7)$$

$$\begin{bmatrix} \mathbf{M}_i & I \end{bmatrix} - \overline{\mathbf{m}}_i \mathbf{1} = \mathbf{U}_f \boldsymbol{\Sigma}_f \mathbf{V}_f^T + \overline{\mathbf{m}}_f \mathbf{1} - \overline{\mathbf{m}}_i \mathbf{1} \ . \quad (8)$$

$$\begin{bmatrix} \mathbf{M}_i - \overline{\mathbf{m}}_i \mathbf{1} & \mathbf{I} - \overline{\mathbf{m}}_i \mathbf{1} \end{bmatrix} = \mathbf{U}_f \boldsymbol{\Sigma}_f \mathbf{V}_f^T + (\overline{\mathbf{m}}_f - \overline{\mathbf{m}}_i) \mathbf{1} \ . \quad (9)$$

$$\begin{bmatrix} \hat{\mathbf{M}}_i & \hat{\mathbf{I}} \end{bmatrix} = \mathbf{U}_t \boldsymbol{\Sigma}_t \mathbf{V}_t^T \ . \quad (10)$$

Note that (10) is the updated SVD from (7) when some new observations $\hat{\mathbf{I}}$ are added. This update can be done as [11] suggests:

$$\begin{bmatrix} \hat{\mathbf{M}}_i & \hat{\mathbf{I}} \end{bmatrix} = \begin{bmatrix} \mathbf{U}_i & \mathbf{Q}_i \end{bmatrix} \cdot \begin{bmatrix} \boldsymbol{\Sigma}_i & \mathbf{U}_i^T \hat{\mathbf{I}} \\ \mathbf{0} & \mathbf{Q}_i^T \hat{\mathbf{I}} \end{bmatrix} \cdot \begin{bmatrix} \mathbf{V}_i^T & \mathbf{0} \\ \mathbf{0} & \mathbf{1} \end{bmatrix} = \begin{bmatrix} \mathbf{U}_i & \mathbf{Q}_i \end{bmatrix} \mathbf{U}_d \boldsymbol{\Sigma}_d \mathbf{V}_d^T \cdot \begin{bmatrix} \mathbf{V}_i^T & \mathbf{0} \\ \mathbf{0} & \mathbf{1} \end{bmatrix} = \mathbf{U}_t \boldsymbol{\Sigma}_t \mathbf{V}_t^T \quad (11)$$

where QR-decomposition is done to $\hat{\mathbf{I}} - \mathbf{U}_i \mathbf{U}_i^T \hat{\mathbf{I}} = \mathbf{Q}_i \mathbf{R}_i$ to obtain an orthogonal basis $\mathbf{Q}_i$ for the reconstruction error. Next, the *mean update algorithm* can be executed starting from the knowledge of $\mathbf{V}_t^T = \hat{\mathbf{V}}_t^T + \overline{\mathbf{v}}_t \mathbf{1}$, where $\overline{\mathbf{v}}_t = \frac{1}{q+c} \sum_{k=1}^{q+c} \mathbf{v}_k$:

$$\begin{bmatrix} \hat{\mathbf{M}}_i & \hat{\mathbf{I}} \end{bmatrix} = \mathbf{U}_t \boldsymbol{\Sigma}_t \hat{\mathbf{V}}_t^T + \mathbf{U}_t \boldsymbol{\Sigma}_t \overline{\mathbf{v}}_t \mathbf{1} = \mathbf{U}_t \boldsymbol{\Sigma}_t \hat{\mathbf{V}}_t^T + \overline{\mathbf{m}}_t \mathbf{1} \ . \quad (12)$$

$$\begin{bmatrix} \hat{\mathbf{M}}_i & \hat{\mathbf{I}} \end{bmatrix} = \mathbf{U}_t \boldsymbol{\Sigma}_t \mathbf{R}_v^T \mathbf{Q}_v^T + \overline{\mathbf{m}}_t \mathbf{1} = \mathbf{U}_f \boldsymbol{\Sigma}_f \mathbf{V}_u^T \mathbf{Q}_v^T + \overline{\mathbf{m}}_t \mathbf{1} = \mathbf{U}_f \boldsymbol{\Sigma}_f \mathbf{V}_f^T + \overline{\mathbf{m}}_t \mathbf{1} \ . \quad (13)$$

$$\begin{bmatrix} \hat{\mathbf{M}}_i & \hat{\mathbf{I}} \end{bmatrix} + \overline{\mathbf{m}}_i \mathbf{1} = \mathbf{U}_f \boldsymbol{\Sigma}_f \mathbf{V}_f^T + \overline{\mathbf{m}}_t \mathbf{1} + \overline{\mathbf{m}}_i \mathbf{1} \ . \quad (14)$$

$$\begin{bmatrix} \mathbf{M}_i & \mathbf{I} \end{bmatrix} = \mathbf{U}_f \boldsymbol{\Sigma}_f \mathbf{V}_f^T + \overline{\mathbf{m}}_f \mathbf{1} \ . \quad (15)$$

It is assumed $\mathbf{Q}_v \mathbf{R}_v$ as the QR-decomposition of $\hat{\mathbf{V}}_t$, $\mathbf{U}_f \boldsymbol{\Sigma}_f \mathbf{V}_u^T$ as the SVD of $\mathbf{U}_t \boldsymbol{\Sigma}_t \mathbf{R}_v^T$ and $\overline{\mathbf{m}}_f = \overline{\mathbf{m}}_t + \overline{\mathbf{m}}_i$. Note that (15) and (6) are equivalent.

2.4 Mean Extraction from a Given SVD

The previous method can also be used to extract the mean information from an existing SVD, e.g. trying to express $\mathbf{S} = \mathbf{U}_t \mathbf{\Sigma}_t \mathbf{V}_t^T$ as $\mathbf{S} = \mathbf{U}_f \mathbf{\Sigma}_f \mathbf{V}_f^T + \overline{\mathbf{s}} \cdot \mathbf{1}$ setting $\left[\hat{\mathbf{M}}_i \; \hat{\mathbf{I}} \right] = \mathbf{S}$ and $\overline{\mathbf{m}}_t = \mathbf{0}$ in (12) to (15).

2.5 Time and Memory Complexity

The mean update presented in section 2.3 does not increase the order of resources required in methods of incremental SVD developed in [9], [11], [10], [12] and [13]. The computational cost becomes $O\left(qr^2 + pr^2\right)$ and the memory complexity is $O\left(pr + qr\right)$, as shown in Table 1.

3 On-the-fly Face Training

In this paper, *On-the-fly Face Training* is defined as the process of learning the photo-realistic facial appearance model of a person observed in a sequence in a rigorous causal fashion. This fact means that it is not necessary to take into account subsequent images when adding the information of the current one, which is considered only once. Note that the facial appearance is learnt in the

Table 1. Resource order requirements of the proposed SVD mean update algorithm

Operation	Comp. cost	Mem. requirements
$\mathbf{V}_{q\times r}^T - \left(\frac{1}{q}\sum_{k=1}^{q} \left(\mathbf{v}_k\right)_{r\times 1}\right) \mathbf{1}_{1\times q} \rightarrow \hat{\mathbf{V}}_{q\times r}^T$	$O\left(qr\right)$	$O\left(qr + r\right)$
$\hat{\mathbf{V}}_{q\times r} \rightarrow \left(\mathbf{Q}_v\right)_{q\times r} \left(\mathbf{R}_v\right)_{r\times r}$	$O\left(qr^2\right)$	$O\left(qr + r^2\right)$
$\left(\mathbf{U}_i\right)_{p\times r} \left(\mathbf{\Sigma}_i\right)_{r\times r} \left(\mathbf{R}_v^T\right)_{r\times r} \rightarrow \mathbf{T}_{p\times r}$	$O\left(pr^2 + r^3\right)$	$O\left(pr + r^2\right)$
$\mathbf{T}_{p\times r} \rightarrow \left(\mathbf{U}_f\right)_{p\times r} \left(\mathbf{\Sigma}_f\right)_{r\times r} \left(\mathbf{V}_u^T\right)_{r\times r}$	$O\left(pr^2\right)$	$O\left(pr + r^2\right)$
$\left(\mathbf{V}_f\right)_{q\times r} \rightarrow \left(\mathbf{Q}_v\right)_{q\times r} \left(\mathbf{V}_u\right)_{r\times r}$	$O\left(qr^2\right)$	$O\left(qr + r^2\right)$
Totals, assuming $p \gg r$ and $g \gg r$	$O\left(qr^2 + pr^2\right)$	$O\left(pr + qr\right)$

same order as the captured images, allowing a real-time learning capability in near future, as computational resources are constantly being increased.

3.1 Data Representation

An N image sequence $\mathbf{S} = [\mathbf{I}_1 \cdots \mathbf{I}_N]$ and a set of four masks $\mathbf{\Pi} = \left\{\pi^1, \ldots, \pi^4\right\}$, attached to four facial elements (like mouth, eyes or foerehead), are given. For each image $\mathbf{I}_t$, its specific mouth, eyes and forehead appearance are extracted with $\mathbf{\Pi}$, obtaining four observation vectors $\mathbf{o}_t^r$ (see Fig. 1). Therefore, four observation matrices $\mathbf{O}^r$ can be obtained from the application of the set of masks $\mathbf{\Pi}$ over the sequence $\mathbf{S}$. Dimensionality reduction of $\mathbf{O}^r$ can be achived using SVD [14]: $\mathbf{O}^r = [\mathbf{o}_1^r \cdots \mathbf{o}_N^r] = \mathbf{U}^r \mathbf{\Sigma}^r \left(\mathbf{V}^r\right)^T + \overline{\mathbf{o}}^r \mathbf{1}_{1\times N}$, where $\overline{\mathbf{o}}^r = \frac{1}{N}\sum_{k=1}^{N} \mathbf{o}_k^r$. Note that facial element appearances can be parameterized as $\mathbf{C}^r = \mathbf{\Sigma}^r \left(\mathbf{V}^r\right)^T$ (see Sect. 2.1). In the example proposed in this paper, faces composed of 41205 pixels could be codified with 35 coefficients, representing a reduction of more than 99.9% without any loss of perceptual quality (see Fig. 2).

3.2 Training Process

One major drawback of the parametrization presented in Sect. 3.1 consists in the image alignment of the sequence [5]. Unless all face images through the whole sequence have the same position, ghoslty results may appear and suboptimal dimensionality reduction will be achieved. The tracking scheme presented in this paper combines simultaneously both processes of learning and alingment.
First of all, the four masks π^r are manually extracted from the first image $\mathbf{I}_1$ of sequence $\mathbf{S}$ and the first observation vectors $\mathbf{o}_1^1, \ldots, \mathbf{o}_1^4$ are obtained. Next, the corresponding alignment coefficients $\mathbf{a}_1$ are set to $\mathbf{0}$; they represent the affine transformation used to fit the masks onto the face on each frame [5]. Using the tracking algorithm presented in [7] over the second image $\mathbf{I}_2$, observations

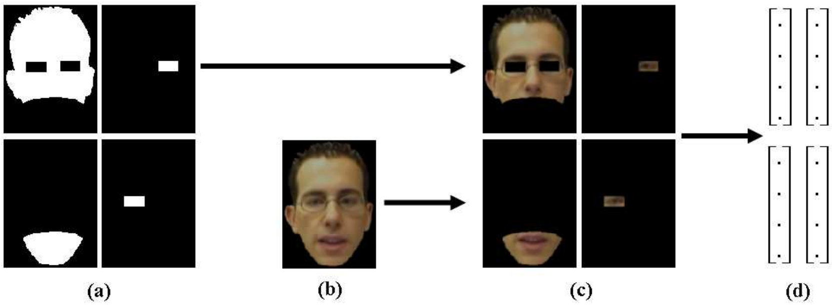

Fig. 1. (a) Masks π^r. (b) Image $\mathbf{I}_t$. (c) Regions $\mathbf{R}_t^r$, obtained from the application of each mask π^r over image $\mathbf{I}_t$. (d) Vectors $\mathbf{o}_t^r$ related to the defined regions.

Fig. 2. (a) Three observed frames of a subject's face. (b) The same synthesized frames after learning the appearance model of this person.

$\mathbf{o}_2^1, \ldots, \mathbf{o}_2^4$ and alignment coefficients $\mathbf{a}_2$ are stored. At this point, each facial element r can be factorized as $[\mathbf{o}_1^r\, \mathbf{o}_2^r] = \mathbf{O}_2^r = \mathbf{U}_2^r \mathbf{\Sigma}_2^r \left(\mathbf{V}_2^r\right)^T + \overline{\mathbf{o}}_2^r = \mathbf{U}_2^r \left(\mathbf{C}_2^r\right)^T + \overline{\mathbf{o}}_2^r$, where the current mean observation is generated by $\overline{\mathbf{o}}_2^r = \frac{\mathbf{o}_1^r + \mathbf{o}_2^r}{2}$, the eigenvectors of $\mathbf{O}_2^r \left(\mathbf{O}_2^r\right)^T$ are found in $\mathbf{U}_2^r$ and the texture parametrization of the r-th facial element in images $\mathbf{I}_1$ and $\mathbf{I}_2$ is obtained in $\mathbf{C}_2^r$. Once this initialization is done, the *On-the-fly Training Algorithm* (Figure 3) can be executed. Besides, only those columns of $\mathbf{U}_{t+1}^r$ and $\mathbf{V}_{t+1}^r$ whose values of $\mathbf{\Sigma}_{t+1}^r$ exceed a threshold τ are considered, keeping only those eigenvectors with enough information. The value of τ decreases from $0,5$ to $0,5 \cdot 10^{-3}$ in the first images (1 seconds at 25 im/s) in order to allow better face localization when almost no information is known

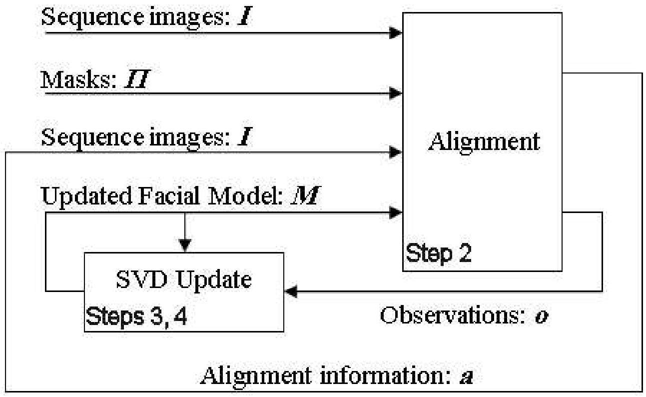

Fig. 3. Block diagram of the *On-the-fly Training Algorithm*

about its appearance. Notice that alignment parameters $\mathbf{a}$ can be used to extract gestural information in a multimodal input system [15].

On-the-fly Training Algorithm

In: $\mathbf{U}_2, \mathbf{\Sigma}_2, \mathbf{V}_2, \bar{\mathbf{o}}_2$, alignment coefficients $\mathbf{a}_2$ and set of four masks $\mathbf{\Pi}$

1. Set $k = 2$

2. Using $\mathbf{U}_k, \mathbf{\Sigma}_k, \mathbf{V}_k, \bar{\mathbf{o}}_k$, $\mathbf{a}_k$ and $\mathbf{\Pi}$, a new input image $\mathbf{I}_{k+1}$ is aligned, generating L observation vectors $\mathbf{o}_{k+1}^r$ and updating the alignment information to produce $\mathbf{a}_{k+1}$.

3. Obtain $\mathbf{U}_{k+1}, \mathbf{\Sigma}_{k+1}, \mathbf{V}_{k+1}$ and $\bar{\mathbf{o}}_{k+1}$ from $\mathbf{U}_k, \mathbf{\Sigma}_k, \mathbf{V}_k, \bar{\mathbf{o}}_k$, and $\mathbf{o}_{k+1}$ (4)-(8).

4. Trim $\mathbf{U}_{k+1}$ and $\mathbf{V}_{k+1}$ according to $\mathbf{\Sigma}_{k+1}$.

5. Set $k = k + 1$ and go to Step 2 until there is no more new images.

Out: $\mathbf{U}_f, \mathbf{\Sigma}_f, \mathbf{V}_f, \bar{\mathbf{o}}_f$ and alignment coefficients $\mathbf{a}_f$ for each image.

4 Experimental Results

The *On-the-fly Training Algorithm* has been tested over a short sequence and a long one, both recorded at a frame rate of 25 im/s. The short sequence consists of 316 images and it has been used to compare the results obtained from our *On-the-fly Training Algorithm* and its previous non-causal version [7]. Achieving the same quality in the results (see Fig. 4), the presented algorithm has reduced the execution time about 66% with respect to [7] and has required about 7 Mbytes in front of the 200 Mbytes consumed by [7] (see the comparison in Fig. 5). Later, if we focus on the long sequence (10000 frames), its processing requirements were impossible to met with the non-causal algorithm [7] because its huge memory cost of 6000 Mbytes, although massive storage systems (e.g. hard drives) were used; the *On-the-fly Training Algorithm* reduced the memory requirements to 17 Mbytes with a processing time of a little more than 10 hours (using a 2GHz processor) (see Fig. 5).

5 Concluding Remarks

In this paper, a new method for extracting the mean of an existing SVD is presented, without increasing either the cost order of memory or time. Thus, incremental computation of SVD preserving a zero data mean is allowed. Fields that can benefit from this method can be, e.g.: classification problems, where the mean information is used to center the data; incremental computation of covariation matices, which need to be centered around its mean; causal construction of eigenspaces, where the principal components of the data are included, as well as the mean information. With respect to the latter, the *On-the-fly Algorithm* is presented in this work. Given an image sequence and a set of masks, this algorithm is capable of generating a separate eigenspace for each facial element (learning all their appearance variations due to changes in expression and visual utterances) and effectively tracking and aligning them. Furthermore, longer sequences than previous methods [5], [7] can be processed with the same visual accuracy when no ilumination changes appear. Finally, we plan to add more

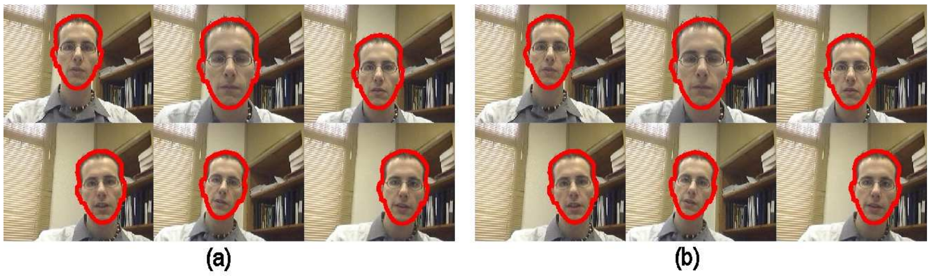

Fig. 4. Output tracking results of the learning process for: (a) the *On-the-fly Training Algorithm* and (b) the non-causal algorithm.

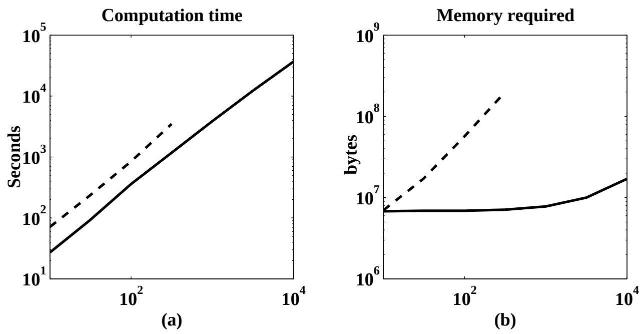

Fig. 5. Solid line represents the *On-the-fly Training Algorithm* performance while the dashed one belongs to the non-causal algorithm presented in the work of [7]. (a) Computation time in seconds. (b) Memory used in bytes.

robustness to this algorithm using methods like [5] and more work will be done in order to achieve real time perfomance, so specific appearance models can be obtained as a person is being recorded.

References

1. E. André: The generation of multimedia presentations. A Handbook of Natural Language Processing: techniques and applications for the processing of language as text, R. Dale, H. Misl and H. somers, Eds., Marcel Dekker Inc., 2000, 305-327

2. S. Robbe-Reiter, N. Carbonell and P. Dauchy: Expression constraints in multimodal human-computer interaction. Intelligent User Interfaces, 2000, 225-228
3. C. Stephanidis: Towards universal acces in the disappearing computer environment. UPGRADE, vol. IV, n. 1, 2003, pp. 53-59
4. E. André: Natural language in multimedia/multimodal systems. Handbook of Computational Linguistics, R. Miktov, Ed., Oxford Univ. Press, 2003, 650-669
5. F. de la Torre and M. Black: Robust parameterized component analysis: Theory and applications to 2d facial modeling. ECCV, 2002, 654-669
6. T. Ezzat, G. Geiger and T. Poggio: Trainable videorealistic speech animation. ACM SIGGRAPH, San Antonio, Texas, July 2002, 225-228
7. J. Melenchón, F. de la Torre, I. Iriondo, F. Alías, E. Martínez and Ll. Vicent: Text to visual synthesis with appearance models, ICIP, 2003, vol I, 237-240
8. D. Cosker, D. Marshall, P. Rosin and Y. Hicks: Video realistic talking heads using hierarchical non-linear speech-appearance models. Mirage, France, 2003
9. M. Gu and S.C. Eisenstat: A Stable and fast algorithm for updating the singular value decomposition. Tech. Rep. YALEU/DCS/RR-966, New Haven, 1993
10. S. Chandrasekaran, B. Manjunath, Y. Wang, J. Winkeler and H. Zhang: An eigenspace update algorithm for image analysis. GMIP, vol. 59, no. 5, 1997, 321-332
11. M. Brand: Incremental singular value decomposition of uncertain data with missing values. ECCV, 2002, I: 707 ff.
12. D. Scočaj and A. Leonardis: Incremental approach to robust learning of subspaces. ÖAGM, 2002, 71-78
13. P.M. Hall, D.R. Marshall and R. Martin: Adding and substracting eigenspaces with eigenvalue decomposition and singular value decomposition. ICV, vol. 20, no. 13-14, december 2002, 1009-1016
14. M. Kirby: Geometric Data Analysis: An Empirical Approach to Dimensionality Reduction and the Study of Patterns. John Wiley & Sons Inc., 2001
15. F. Keates and P. Robinson: Gestures and multimodal input. Behaviour and Information Technology, Taylor and Francis Ltd., 1999, 18(1), 35-42