MULTI-DOMAIN SUBSPACE REPRESENTATION OF IMAGE SEQUENCES

Klaus Diepold *Florian Obermeier* *Thomas Zeitler*
Technische Universität München
80290 Munich, Germany

ABSTRACT

We investigate three closely related approaches to employ subspace modeling techniques to digital video sequences. The goal for this modeling approach is to obtain a compact and faithful representation of the video sequence for computer vision applications and image understanding. We propose to use a multi-domain approach, where the subspace modeling is applied to data in the intensity domain (textures) and the motion domain (motion vector fields).

Experiments are described which show that especially a combination of motion estimation and subspace modeling of motion vector fields using singular value decomposition provides interesting results.

1. INTRODUCTION

We present some basic considerations and first investigation results for the use of subspace modeling techniques applied to image sequences. We will discuss how to represent digital video data by rank-reduced subspace models effectively. This approach holds some promise for solving various video analysis problems as is discussed in the article.

1.1. Multi-Domain Modeling of Video Sequences

Temporal intensity changes of pixels in an image sequence create the sensation of a moving picture. For image understanding purposes we aim to identify and track objects throughout a video sequence.

The temporal changes of pixel intensities may be attributed to the motion of objects in the scene. From a purely technical point of view motion of pixels amounts to address changes of the corresponding pixels. These address changes or, equivalently, the displacement of pixels define a domain transform of the intensity function $I(x,y,t)$. In the following we refer to this domain transform as the motion domain of the video sequence $I(x,y,t)$.

Alternatively, the temporal intensities changes of pixels may be attributed to changes of brightness or color of the pixels. Take a blinking street light as an example for such a temporal phenomenon. These changes of pixel intensities define a range transform of $I(x,y,t)$. In the following we refer to this range transform as the intensity domain of the video sequence $I(x,y,t)$.

The notion of multi-domain modeling presented here, combines the simultaneous modeling in the motion domain and the intensity domain. Note that there is no unique split between the two domains as the temporal variation of the function $I(x,y,t)$ can be explained as variations of intensity only and also in terms of address changes, i.e. by motion.

1.2. Reduced-Rank Subspace Modeling

Singular value decomposition (SVD) and subspace estimation techniques in combination with the associated low-rank approximation features have been used successfully in various fields in signal processing [2], [7] . The present modeling approach is based on the hypothesis that the function $I(x,y,t)$ can be represented efficiently by a combination of subspace models in the motion domain and in the intensity domain. A subspace model is considered to be efficient, if the subspace dimension necessary to faithfully explain the data has a low dimension. The multi-domain subspace model is considered to be efficient if the sum of the individual subspace dimensions is minimum. The minimality of the multi-domain subspace model provides a means to resolve the ambiguity of modeling a video sequence in the intensity and the motion domain.

Subspace models include the option to determine a low-rank approximation of the original data which serves as a mechanism to reduce the noise in video data.

The feasibility of such an approach has been investigated and the pertaining experimental results are presented for different methods of subspace modeling.

1.3. Singular Value Decomposition (SVD)

The singular value decomposition (SVD) of a m×n matrix A is given in terms of the factorization ([3], [6])

$$A = U \cdot \Sigma \cdot V^* = [u_1, ..., u_p] \cdot \begin{bmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_p \end{bmatrix} \cdot \begin{bmatrix} v_1^* \\ \vdots \\ v_p^* \end{bmatrix} \qquad \textbf{(1)}$$

with unitary matrices U and V ($U^* \cdot U=1$, $V^* \cdot V=1$) and the diagonal matrix Σ, which contains the singular values σ_i in decreasing order, i.e. $\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_p$, and rank(A) = p.

The columns of V, i.e. the vectors $v_1,\ldots,v_p$ constitute an orthonormal basis for the row space of **A**, while the columns of the matrix U, i.e. the vectors $u_1,\ldots,u_p$ form an orthonormal basis of A's column space. By selecting the first f singular values (f < p) and the associated singular vectors $u_1,\ldots,u_f$ and $v_1,\ldots,v_f$ results in the best rank-f approximation of the matrix **A.** This approximation is denoted by the matrix $\hat{A}$.

2. SUBSPACES ASSOCIATED WITH VIDEO DOMAINS

The set of all pixels in a sampled video sequence $I(x,y,t)$ forms a 'video cube' which is given by the vertical $x = \{1, 2,\ldots, n_v\}$, horizontal $y = \{1, 2,\ldots, n_h\}$, and temporal index $t=\{1, 2, \ldots, n_t\}$, where n_v and n_h denote the number of pixels in vertical and horizontal direction, respectively. The integer n_t denotes the number of frames in a given video sequence. In the following we will restrict our discussion to black and white video sequences by working on the luma component of color video only. The extension to color video is straightforward.

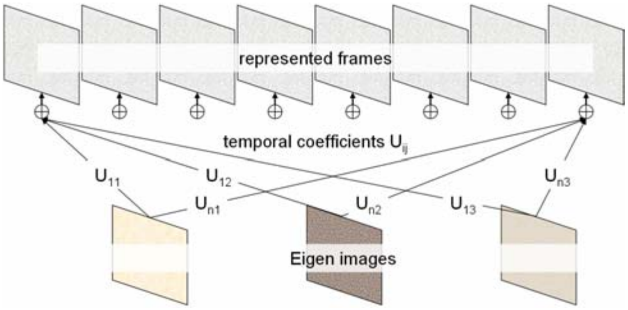

Figure 1: Subspace Modeling for Video data

2.1. Subspace in the Intensity Domain

Individual frames can be considered as vectors of dimension $n_v \cdot n_h$. All pixel values of a frame at time instant t are reformatted into a row vector

$$i_t = [I_{1,1,t}, I_{1,2,t}, \ldots, I_{2,1,t}, I_{2,2,t}, \ldots, I_{n_v,n_h,t}]. \quad \textbf{(2)}$$

The set of all video frames in a sequence form an n_t-dimensional subspace of the linear vector space of dimension $n_v \cdot n_h$, which will subsequently be referred to as the sequence space $\mathcal{X}$. We stack all reformatted frames of a specific sequence as the rows of a video sequence matrix X of dimension $n_t \times (n_v \cdot n_h)$

$$X = \left[- i_1 -, \quad \cdots, \quad - i_{n_t} -\right]^T. \quad \textbf{(3)}$$

We can easily identify that the sequence space $\mathcal{X}$ is spanned by the rows of the sequence matrix X. We can compute an orthonormal basis for the sequence space $\mathcal{X}$ using the singular value decomposition $X=U\cdot\Sigma\cdot V^*$ such that $\mathcal{X}$ is given in terms of the rows of the matrix V^*.

2.2. Subspace in the Motion Domain

In order to apply subspace modeling in the motion domain we need to account for the intensity changes that are caused by motion of objects or camera motion.

There exist various techniques to extract motion information from video sequences. In this investigation we compute dense motion vector fields using e.g. a Horn & Schunck type computation of optic flow [4].

We employ telescopic motion estimation, as illustrated in Figure 2. The frame in the middle of a group of pictures is chosen as the reference frame. Motion vector fields are computed for each frame in the group of pictures with respect to the reference frame. For this investigation the size of the group of pictures has been set to 10 frames.

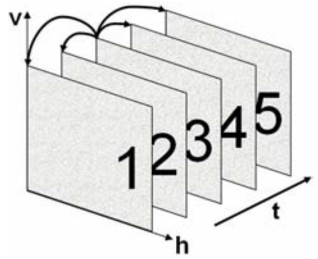

Figure 2: telescopic motion compensation

The resulting motion vectors fields are given in terms of the common coordinate system of the reference frame.

Motion vector fields consist of two components, a vertical and a horizontal component of the motion vectors. Both components of a motion vector can be combined to form a complex number, where the horizontal and vertical components of the motion vector for each pixel are represented by the real part and the imaginary part of a complex number. This is denoted by

$$D_{v,h,t} := D^v_{v,h,t} + j \cdot D^h_{v,h,t}, \quad \textbf{(4)}$$

where the letter D has been chosen with reference to the displacement of pixels.

Similar to the reformatting procedure for the video frames in the preceding section we reformat a single motion vector field to generate a complex-valued row vector

$$d_t = [D_{1,1,t}, D_{1,2,t}, \ldots, D_{2,1,t}, D_{2,2,t}, \ldots, D_{n_v,n_h,t}]. \quad \textbf{(5)}$$

The set of all motion vector fields d_t corresponding to a video sequence form an n_t-dimensional subspace of the linear vector space of dimension $n_v \cdot n_h$, which will subsequently be referred to as the motion space $\mathcal{M}$. We stack all reformatted motion vector fields of a specific sequence as the rows of a video sequence matrix M of dimension $(n_t\text{-}1) \times (n_v \cdot n_h)$

$$M = \left[- d_1 -, \quad \cdots, \quad - d_{n_t} -\right]^T. \quad \textbf{(6)}$$

The telescopic motion data for natural video sequences are expected to span a low dimensional vector space. This is based on the notion of temporal smoothness of motion vector fields due to the inertia of physical objects moving through space.

3. EXPERIMENTAL RESULTS

In this section we present some experimental results to evaluate the validity of the conjectures concerning the dimensionality of the video and motion subspaces. To this end we use the first ten frames of the sequences 'Akiyo', 'Foreman', 'Mobile', and 'Stefan' in CIF resolution.

3.1. Direct Intensity Modeling

The singular values for all four video test sequences, which are computed by applying the SVD to the corresponding video matrix X are displayed in Figure 3. From this plot it can be seen that a good energy compaction is achieved only for sequences with little or no motion. The row vectors v_i^* represent *Eigen* images of the sequence. The first three Eigen images of the sequence 'Akiyo' are shown in Figure 4. The first Eigen image represents the average of the sequence. The second and the third Eigen image are dominated by the motion of the eyes and the mouth, respectively. The temporal coefficients in Figure 5 show the contribution of the different Eigen images to the individual frames. The change in the coefficients tells a lot about the sequence. E.g. there is obviously less eye movement from frame 1 to 4 than there is between frame 4 and 6.

3.2. Modeling of Motion Compensated Intensity

In this experiment, the images in the sequences are motion compensated with respect to the reference frame. The motion compensated frames constitute the rows of the sequence matrix X. X is subsequently factorized by SVD. Figure 6 depicts the resulting singular values. The plot shows a substantially improved energy compaction in comparison to the result shown in Figure 3. That is equivalent to the observation that the subspace of motion compensated intensities has a lower dimensionality and hence can be faithfully represented with a smaller number of basis vectors. The accuracy of the motion estimation and compensation technique may have an impact on the achievable modeling quality [8].

3.3. Modeling in the Motion Domain

Subspace modeling in the motion domain accounts to computing the SVD of the motion matrix M. Plotting the singular values according to this modeling approach produces singular values, which are plotted in Figure 7. This plot demonstrates that a high degree of energy compaction is achieved, which implies a low dimensionality of the corresponding subspace of motion. Figure 8 shows the first Eigen vectors of motion belonging to the sequence ‘Akiyo’. The first Eigen vector mainly corresponds to movement of the eyes, the second Eigen vector mainly captures the movement of the mouth.

Moreover, important information about the sequences becomes visible, as the amount of motion and importance of different independent motion modes in the individual sequences can easily be derived from the singular values and the corresponding Eigen vectors.

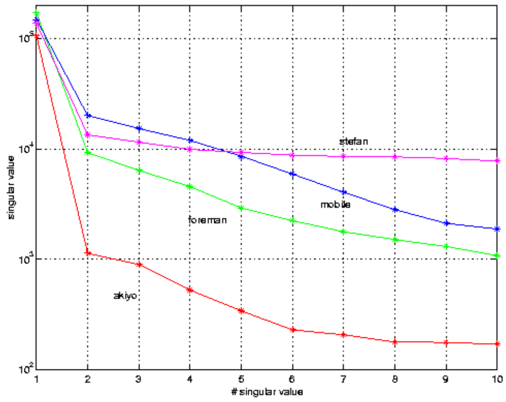

Figure 3: Singular values for direct intensity modeling

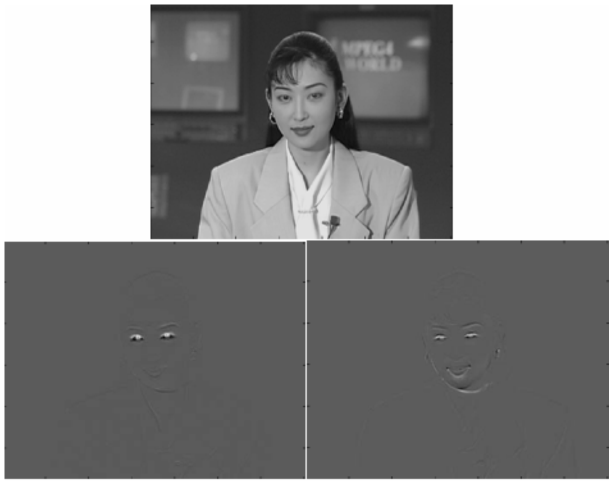

Figure 4: Eigen images of sequence corresponding to the first three singular values ‘Akiyo’

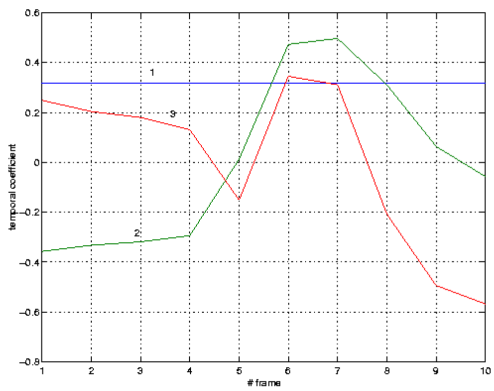

Figure 5: Temporal Coefficients for the first three Eigen vectors of the sequence ‘Akiyo’

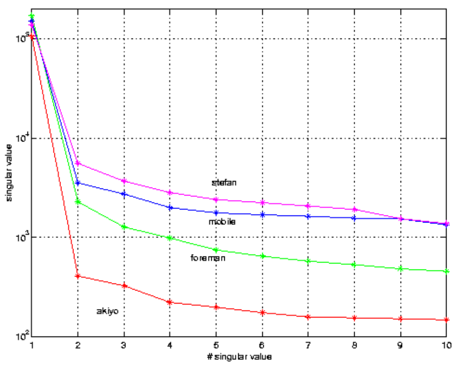

Figure 6: Singular values for motion compensated intensity modeling

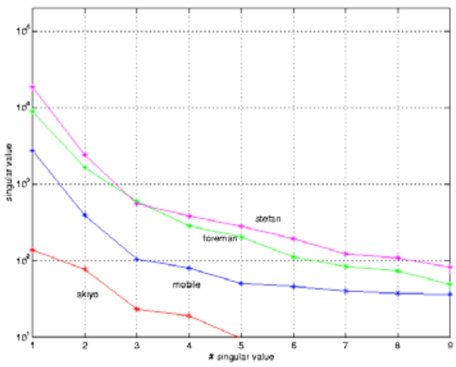

Figure 7: Singular values for modeling in the motion domain

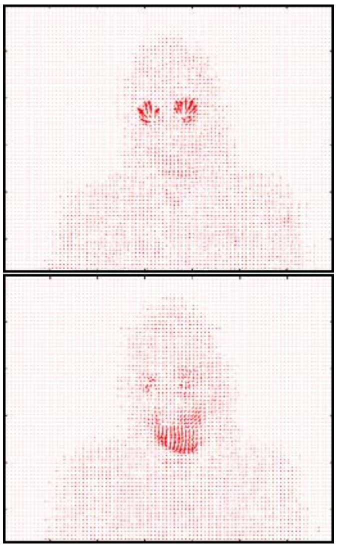

Figure 8: First and second Eigen vectors of motion modeling for sequence 'Akiyo'

4. TARGET APPLICATION SCENARIOS

Various technical challenges associated with the analysis of image sequences may benefit from the presented approach. The target applications include video analysis, computer vision tasks such as object recognition and segmentation, content-based search and retrieval of video material, computer animation, content based interactivity as well as multi sensor data fusion.

All application benefit from the ease of computation, the compaction of data, that comes along with possibly meaningful weightings (i.e. singular values).

Followed by a soft decision image analysis stage, the presented approach might by used to extract key information in a sensible way for multi sensor applications.

5. CONCLUSIONS

The information in a video sequence can be compacted by means of simple matrix computations in two domains, i.e. the intensity domain and the motion domain. Modeling in the intensity domain, it can be shown that a faithful representation with a low dimensional subspace works only for sequences with little motion. Motion compensation has been shown to improve the subspace modeling in the intensity domain.

However, the representation of video by describing the motion compensated texture seems incomplete. The motion information may also be represented in terms of a low dimensional subspace model. The experiments described in this paper also prove that dense motion vector fields can be described in a compact manner by subspace representation.

6. REFERENCES

[1] P. J. L. van Beek, A. M. Tekalp and a. Puri. *2-D Mesh and Motion Compression for Efficient Object Based Video Representation*, Proc. of IEEE Int. Conf. on Image Processing, Volume 3 pp. 440 - 443, Oct. 1997

[2] P. Dewilde and E. F. van Deprettre. *SVD and Signal Processing, Volume I - III*, Elsevier, 1988, 1991, 1995.

[3] G. H. Golub and C. F. van Loan. *Matrix Computations*, John Hopkins University Press, 1983.

[4] B.K.P Horn and B.G. Schunk. *Determining Optical Flow*, Artificial Intelligence, vol. 2, pp. 185-203, 1981.

[5] Kay A. Robbins, *Visualization of Scientific Video Data Using KL Decomposition*. IEEE Trans. on Visualization and Computer Graphics, Vol. 4, No 4, Oct.-Dec. 1998.

[6] Gilbert Strang. *Introduction to Linear Algebra*, WellesleyCambridge Press, 1993.

[7] R. Vaccaro, D. Tufts and G. Boudreux-Bartels. *Advances in Principal Component Signal Processing*, in SVD and Signal Processing I, North-Holland, 1988.

[8] G. Wolberg, *Digital Image Warping*, IEEE Computer Society Press, Los Alamitos, CA, third edition, 1994.