Facial Expression recognition based on dynamic textures

J. Trojanová and M. Železný
Department of Cybernetics, Faculty of Applied Sciences, University of West Bohemia, Czech Republic
trojana@kky.zcu.cz, zelezny@kky.zcu.cz

In this paper, we introduced a novel method for facial expression recognition in video scenes based on dynamic textures approach. The dynamic textures are presented as linear dynamic system that models spatial and temporal stochastic process. The dynamic textures are used for learning and recognition of facial expression. Originally dynamic textures were used for video scenes that exhibit certain stationarity properties in time (e.g. Sea-waves, smoke) [1].

Overview on dynamic textures

Brief overview is presented here, more details about the dynamic textures can be found in [1]. The sequence of images is referred as $\{I(t)\}_{t=1..\tau}$, where τ is time interval. The output of the system $y(t)\in R^m$, $t=1,..,\tau$ is used for representing the pixel intensity. The system is defined as follows :

$$x(t+1)=Ax(t)+v(t) \qquad x(0)=x_0;\ v(t)\sim N(0,Q)$$
$$y(t)=Cx(t)+w(t) \qquad w(t)\ \ N(0,R)$$

where $x(t)\in R^n$ is hidden state, $A\in R^{nxn}$ is the state transition matrix that represent the dynamics of the system, $C\in R^{mxn}$ is the output matrix, v(t) is the driving output to the system, which is assumed to be Gaussian white noise, w(t) is measurement noise.

Estimation of the dynamic textures parameters is done by identification of the system. For the sequence of τ frames the matrix $Y_1^t = [y(1),..,y(\tau)]\in R^{mx\tau}$ is defined. To obtain a unique solution of the system two assumptions about the model is made: canonical model of matrix C: $C^TC=I_n$ (I_n is the *nxn* identity matrix) and data dimensionality reduction m>>n, rank(C)=n. The estimation of the parameters of the models are:

$$\hat{A}=\Sigma V^T\begin{bmatrix}0 & 0\\ I_{t-1} & 0\end{bmatrix}V\left(V^T\begin{bmatrix}I_{t-1} & 0\\ 0 & 0\end{bmatrix}V\right)^{-1}\Sigma^{-1}; \qquad \hat{C}=U; \qquad X=\Sigma V^T$$

where matrix U and V are obtained by singular value decomposition of.

$$Y_1^t=U\Sigma V^T$$

Matrix $U\in R^{mxn}$, $U^TU=I$; $V\in R^{\tau xn}$, $V^TV=I$.

Preliminary testing

Each dynamic texture is characterized by parameters: matrix *A*, *C*. To compare textures difference, distance matrix using subspace angle is used. Comparison is as follows. Resolution of the video is 160x120, each frame is divided into 64 blocks. We concentrate on block series 27-30 that represent upper face.

The comparison between neutral state to facial action unit AU02 (rise eye borrows) was made, see Figure 1. Action unit is a term of facial muscle movement defined by Ekman in [2].

Future work

Since dynamic texture capture a dynamic process we assume that training on small number of speakers would reach efficient results even for speakers that were not included in training phase. The testing on larger dataset was not done yet. The database UWB-07-EFER (UWB stands for University of West Bohemia, 07 for year of recording, EFER for Emotions and Facial Expressions Recognition) was recently acquired. The video database consist of facial action units and expressions of 20 people, it is now being annotated.

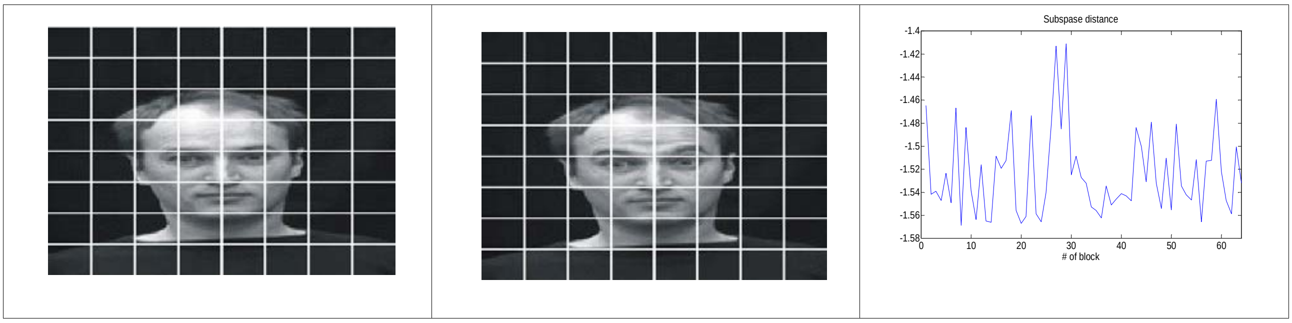

Figure 1. *Comarison between neutral face sequence and sequence that represent eye borrows rise. Image is split into 64 blocks. Plot shows the subspace distance between parameters A,C for all blocks. The biggest distance can be seen on blocks segments 27-30.*

Conclusion

The algorithm is efficient to learn the dynamics of the facial expression. The parameters of the model that represent data can be used as parameterization of the data. The method has shown promising results on preliminary testing.

References

1. Dorretto, G. A. Chiuso, Y.N. Wu and S. Soatto. (2003). Dynamic Textures, *International Journal of Computer Vision*, **51(2)**, 91-109.
2. Ekman P. and Friesen W. (1978) *Facial Action Coding System*, Consulting Psychologists Press Inc. Palo Alto, CA.