Independent Component Analysis (ICA) in Real and Complex Fourier Space: An Application to Videos and Natural Scenes

By

Nimit Kumar* and Shantanu Sharma**

{nimitk@iitk.ac.in, shsharma@iitk.ac.in}

A Project Report Submitted for the Course CS698: The Computational Brain

***Department of Electrical Engineering**

****Department of Computer Science & Engineering**

Indian Institute of Technology, Kanpur

ABSTRACT

Complex ICA is a recent advance in Blind Source Signal Separation techniques tailored for use in separating convoluted temporal signals. Thus, the use of ICA in the time domain is highly limited to such applications, where the mixtures are assumed to instantaneous mixture of the sources. In [1], an approach for blind source separation of convolved mixtures is proposed in the frequency domain. The problem in Frequency Domain then naturally demands a complex space analysis of the problem because of the complex components of the Fourier Series. In this project, we develop a Complex ICA framework for analyzing Videos and their Independent Components in the Fourier Domain.

1. Introduction

Complex Independent Component Analysis is effectively used in analyzing brain signals and in particular electroencephalographic (EEG) modeling. Another application of Complex ICA is in images. However, application of Complex ICA on videos is still an area of active research. Spatio-temporal dynamics of videos is apposite for the Complex ICA framework since signal superpositions are modeled as convolutions. The input signal superposition may also be frequency-dependent, allowing for distinct signal sources at different frequencies.

2. Motivation

Videos are image streams that are often well studied in their Fourier Domain. We analyze the possible independent components of the video clips. Thus the overall project objective is to:

1. Develop a general-purpose framework for Fourier Domain representation of videos and image streams.
2. Develop a Complex Independent Component Analysis application.
3. Use Complex ICA to study Videos in their Fourier Domain.

3. Theoretical Background

3.1 Real and Complex Bell & Sejnowski Rule

In Bell & Sejnowski rule, the weights are updated by using an "Infomax" algorithm, which works by maximizing the information content or entropy. The update equation of Bell & Sejnowski Rule is as follows:

$$\Delta W^{\mathrm{T}} \propto \{\boldsymbol{I} - 2g(\boldsymbol{y})\boldsymbol{y}^{\mathrm{T}}\} W^{\mathrm{T}}$$

In this case, the source signals have a positive kurtosis. The nonlinearity g(y) is defined by:

$$g(y) = \tanh(y)$$

In the Complex Bell and Sejnowski Rule, the weights are updated using a complex Infomax algorithm, which works by maximizing the information content or entropy in the Complex Space. The update equation for the rule is given below:

$$\Delta W = \eta \left[I + \underline{\varphi}(\underline{u})\underline{u}^H \right] W$$

The nonlinearity in Complex Bell & Sejnowski Rule is given by the following set of equations:

$$y = g(u_{\mathrm{Re}} + ju_{\mathrm{Im}}) \triangleq \tanh(u_{\mathrm{Re}}) + j\tanh(u_{\mathrm{Im}})$$

And

$$\varphi_i(u) = -(y_{i\mathrm{Re}} + y_{i\mathrm{Im}}) - (y_{i\mathrm{Re}} \;\; y_{i\mathrm{Im}})e^{-j2\theta_y}$$

3.2 Real and Complex Amari-Cichocki-Yang Rule

The Amari-Cichocki-Yang Rule assumes a Riemannian structure of the problem. Starting from a random initial matrix $\boldsymbol{W}_0$, the weights are updated using the following learning rule:

$$\Delta \boldsymbol{W}^{\mathrm{T}} \propto \{\boldsymbol{I} - g(\boldsymbol{y})\boldsymbol{y}^{\mathrm{T}}\}\boldsymbol{W}^{\mathrm{T}}$$

In this case, $\boldsymbol{g}(\boldsymbol{y})$ is the nonlinearity defined by:

$$g(y) = 29/4y^3 - 47/4y^5 - 14/3y^7 + 25/4y^9 + 3/4y^{11}$$

In the complex space, the Amari Rule changes to the following update equation:

$$\tilde{\nabla} W = (I - (\mathrm{v} \cdot \mathrm{u}^{\mathrm{H}}))W$$

Where

$$v_i = sign(u_i)\frac{g'(|u_i|)}{g(|u_i|)}$$

And

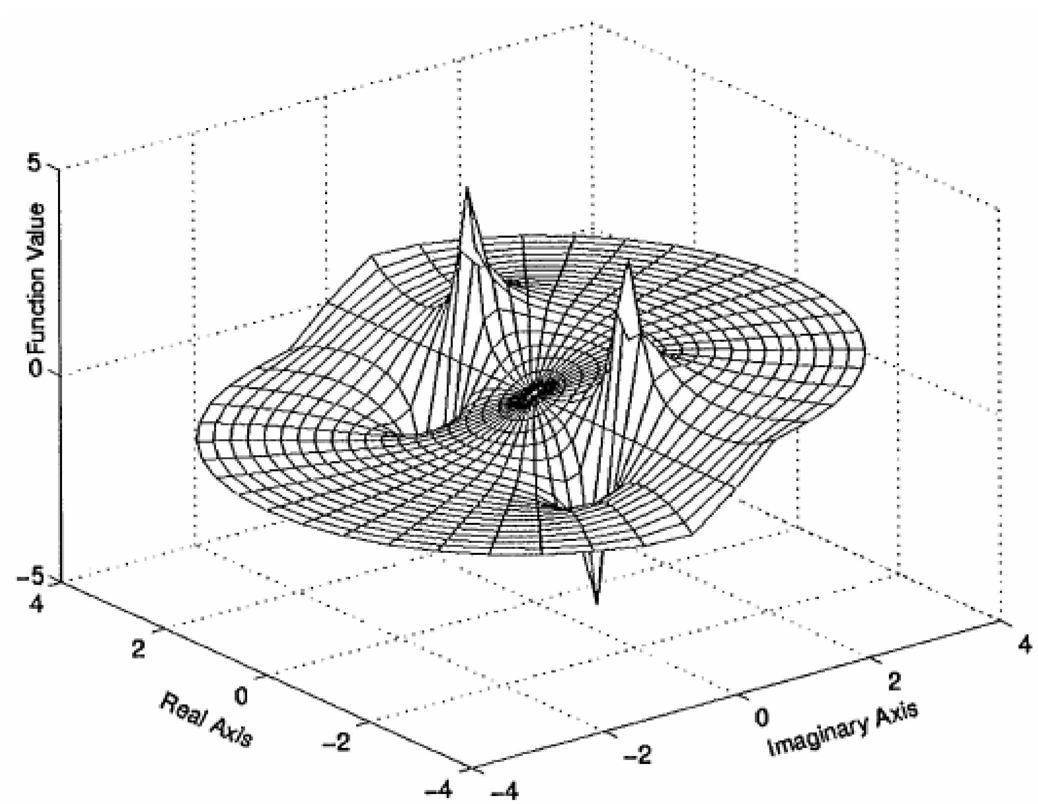
$$\frac{g'(x)}{g(x)} = \frac{1-e^{x}}{1+e^{x}}$$

3.3 Characteristics of a Complex Activation Function

The nonlinear activation function used in ICA should qualify the following properties:

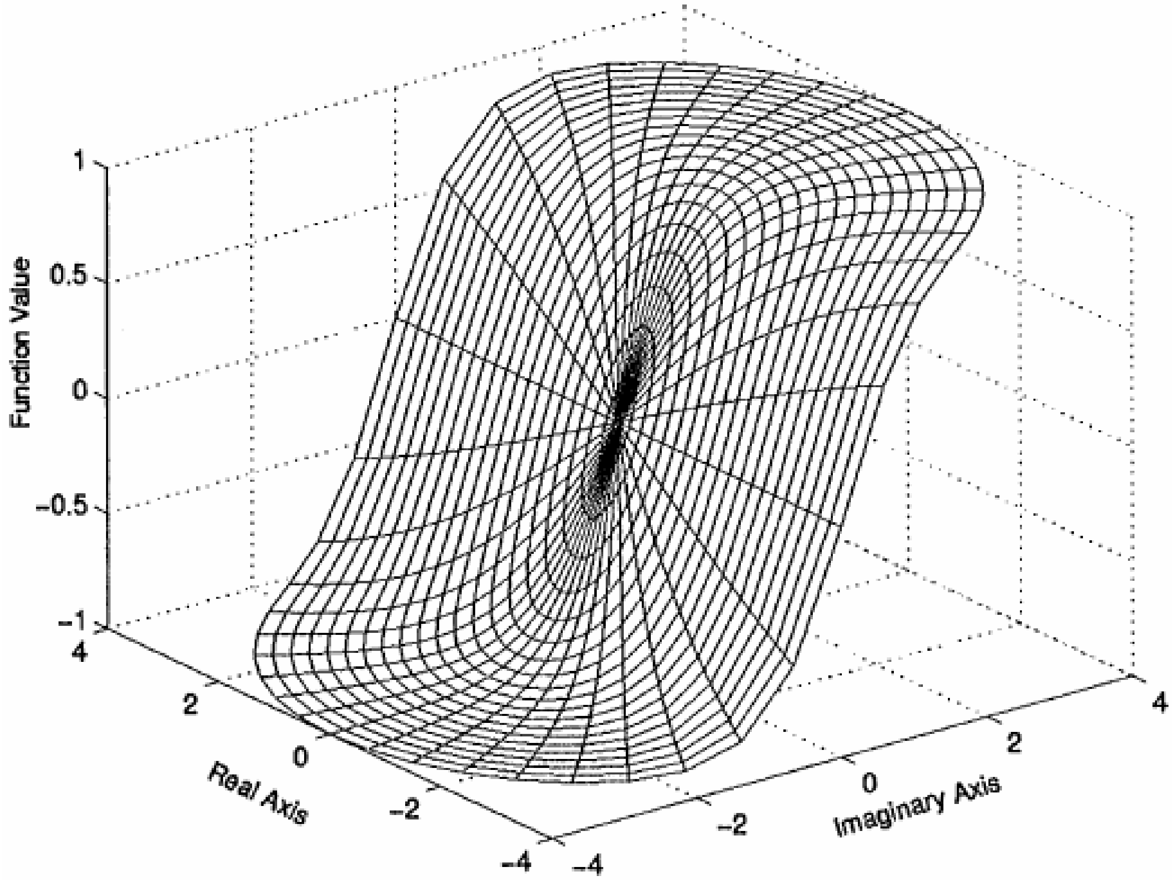
1. The function must be bounded.
2. The function should have a closer fit to the Complex Distribution Function (similar to the PDF in real case).

For instance the ***tanh***(z) activation function is not bounded at z = (k+½)πi as illustrated in the following diagram:



Thus a new activation function was proposed by Smaragdis (1998) defined as:

$$y = g(u_{\mathrm{Re}} + j u_{\mathrm{Im}}) \triangleq \tanh(u_{\mathrm{Re}}) + j \tanh(u_{\mathrm{Im}})$$

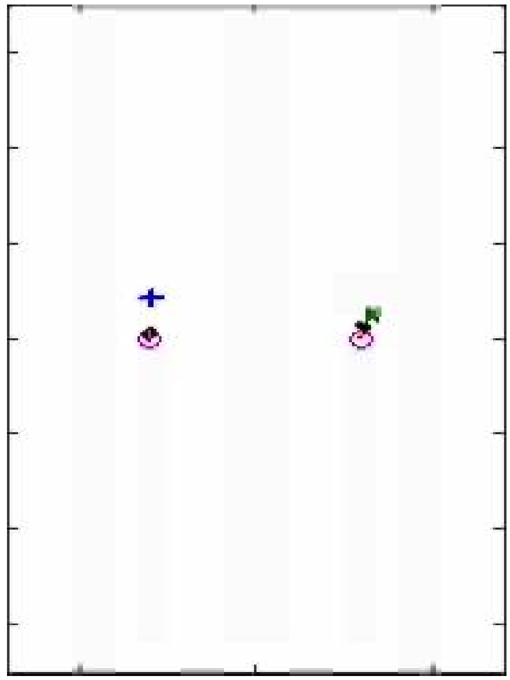


This function is continuous and also fits the complex distribution function closely. Therefore, it is an acceptable complex activation function.

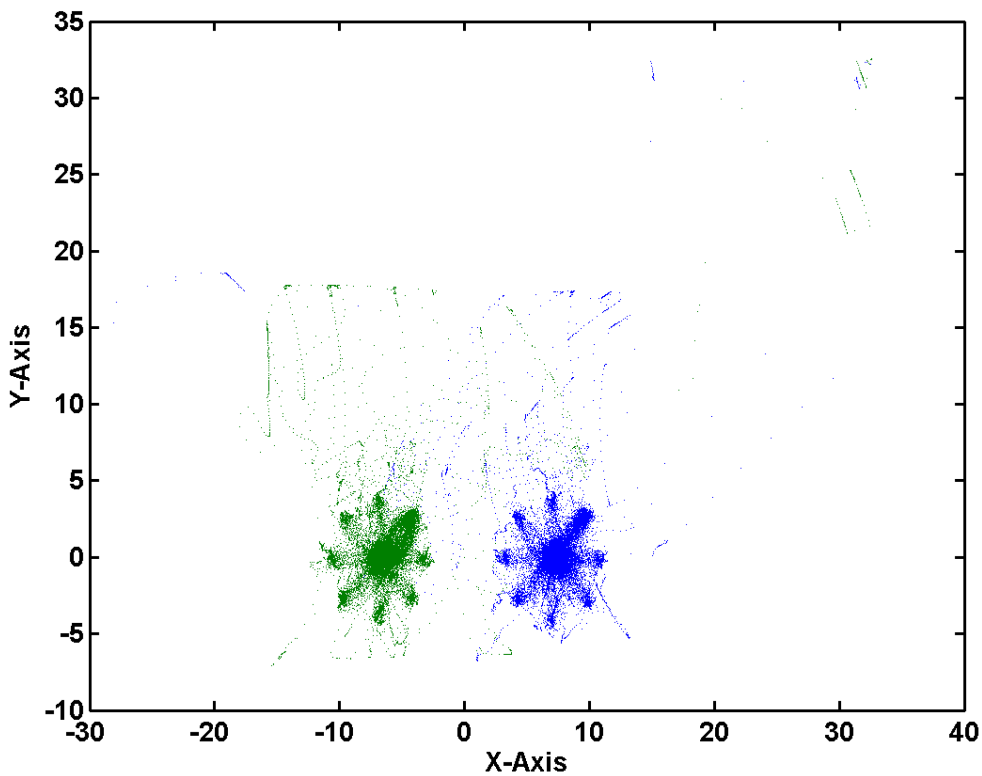
4. Methodology

4.1 A Short Note on Human Vision System

Experiments on gaze tracking suggest that the human vision is saccadic in nature. The following two figures illustrate this by results of an experiment analyzing the gaze of monkey:



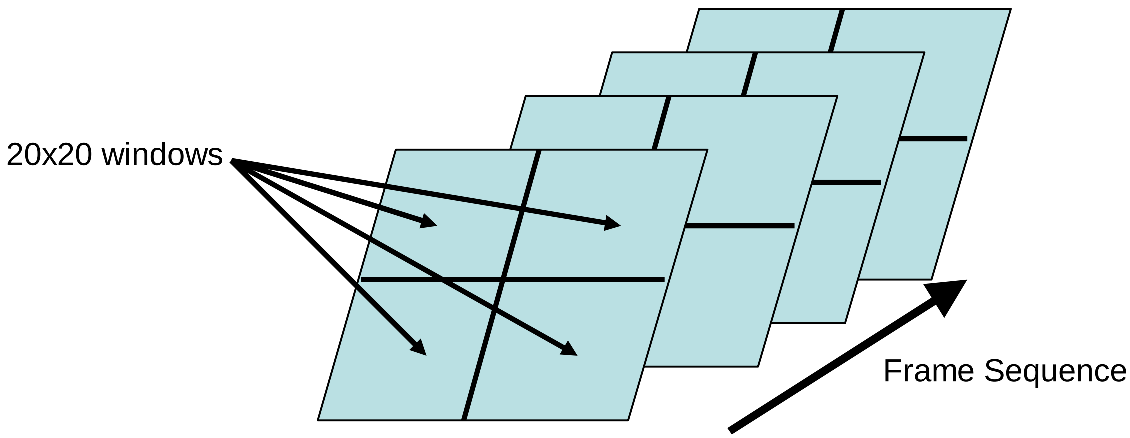
The two-eyes of a monkey trying to fixate and the consequent Saccadic Motion. (Data from: http://www.cs.huji.ac.il/~shpigi/)



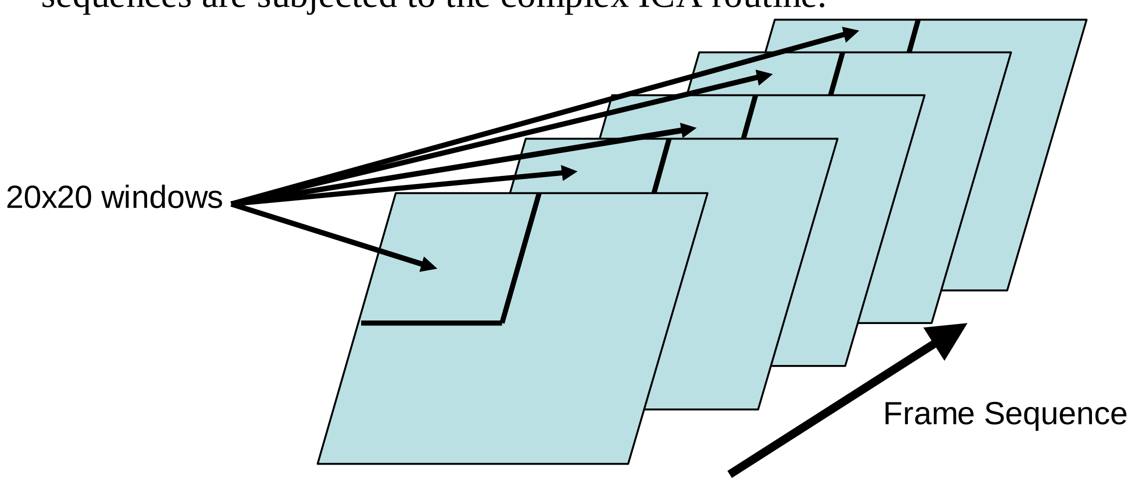
The two-eyes of a monkey trying to fixate and the consequent Saccadic Motion. (Data from: http://www.cs.huji.ac.il/~shpigi/)

We therefore infer that humans actually process more than one scene around (or frames). However, the region of attention is very limited (the frame window being small). This leads to two possible schools of thought on human perception of natural videos:

- Each scene (one full video frame) is analyzed one after the other. Each frame is broken down into 20x20 windows as shown and the process goes by analyzing all the windows in a frame and then moving to the next frame. The extracted windows are subjected to Sejnowski's ICA algorithm.



- A part of the video (a small window) is attended to at one time, thus discretizing the space into attentive regions. The video is broken down into 20x20 windowed-frames. The windowed-sequences are subjected to the complex ICA routine.



The first can be solved using real-ICA and extending Sejnowski's work for Natural Images. However, the second approach demands a complex-ICA formulation. In this example, we observe that not all problems are linear instantaneous mixtures of the sources. In some cases, the observations are convolutions of the sources, for instance:

- Electroencephalogram activities
- Multichannel Blind Equalization
- Video Transmission

Thus, a method to separate convolved signals was felt. Convolution in time domain is equivalent to multiplication in frequency domain. Fourier Transformations were performed in Complex domain to separate the convolved singnals.

4.2 ICA in Real Domain Applied to Natural Videos

Following the Sejnowski approach to natural scenes, we use the frame-based windowing scheme with the Real ICA method to extract the IC filters.

Algorithm –

- Prepare the training set by applying a 20x20 window on some frames.
- Make a 400-dimensional vector with each window.
- Apply Real-ICA to the obtained training set.
- Obtain the 400x400 dimensional weight matrix
- Given any subsequent frame, apply a 20x20 window and get the IC filtered image.

The method is heavily time consuming with increasing number of frames. However, for better accuracy a large training set is required. The following observations are made regarding this frame-based approach:

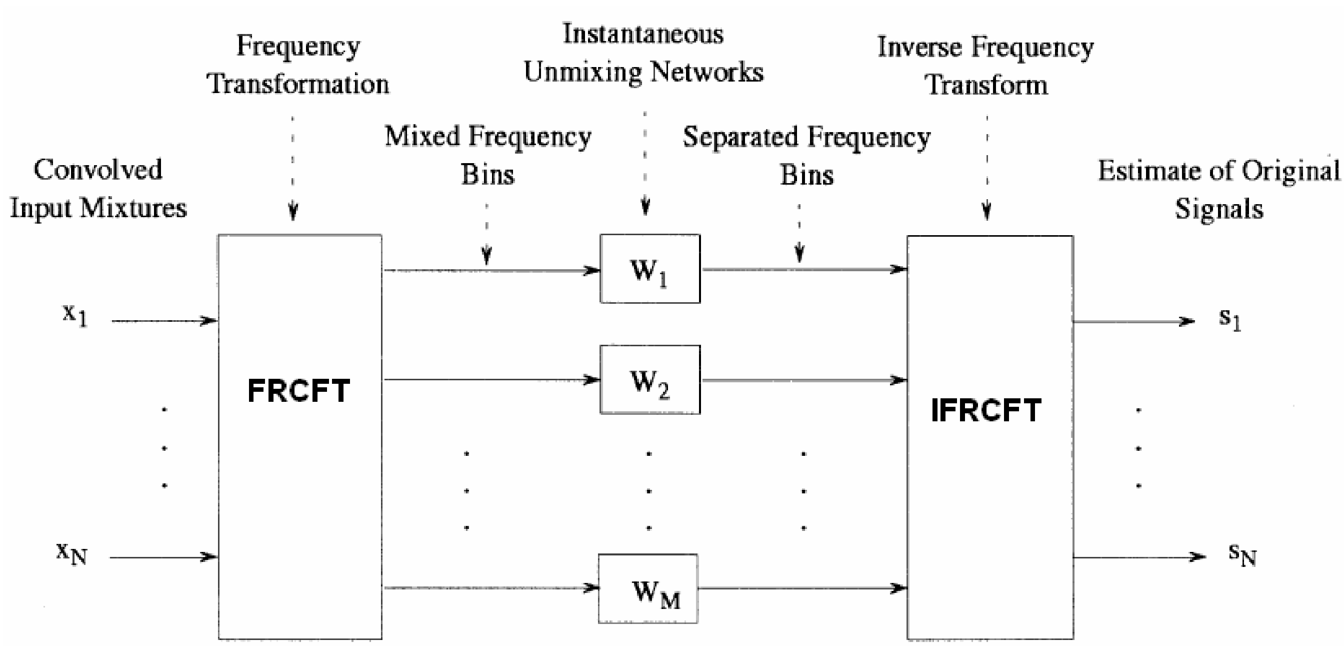
1. As in Sejnowski's work, we obtained filters that resemble gradient/texture features.

2. On using a larger window (as compared to 12x12 in Sejnowski's work), the variations are more subtle.

3. On using 19200 training samples, a computational time of more than 15hrs was required to extract the ICA.

4.3 ICA in Complex Domain Applied to Natural Videos

For applying the complex ICA, videos are modeled as 3-Dimensional images. Frame-Row-Column Fourier Transform (FRCFT) is 3D Fast Fourier Transform method. We use this to transform the video into the complex Fourier Domain, and then the complex ICA is applied. It is useful to work in Fourier space as we get orthogonal basis functions. This enhances the ICA capabilities as no extra effort is required to orthogonalize the coordinates. The flow-chart of this three-step process is show in the following diagram:



The following observations are made regarding this video-based approach:

1. The method yielded similar filters, that though difficult to comprehend look like texture filters.

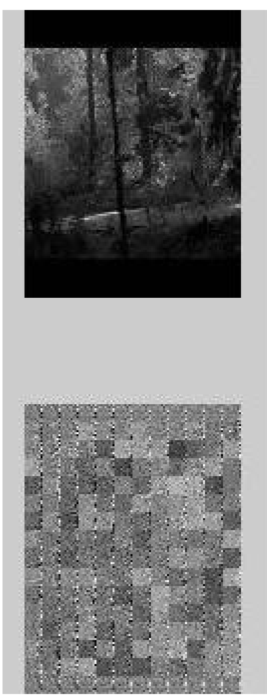
2. These are equally computational in both memory and time.

3. The FRCFT and IFRCFT both require considerable amount of memory.

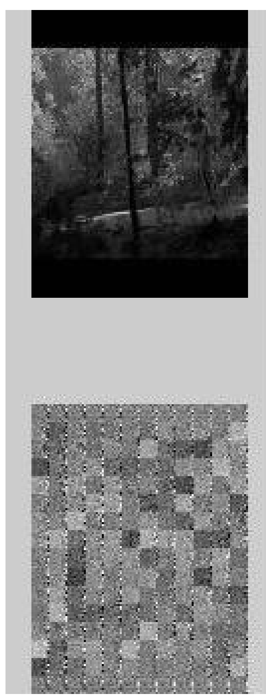
5. Experimental Results

Number of frames in the movie: 11
Size of the movie (in pixels): 320x480
Size of frame windows (in pixels): 20x20
The 11 Independent Components obtained from the movie-clip are shown below:

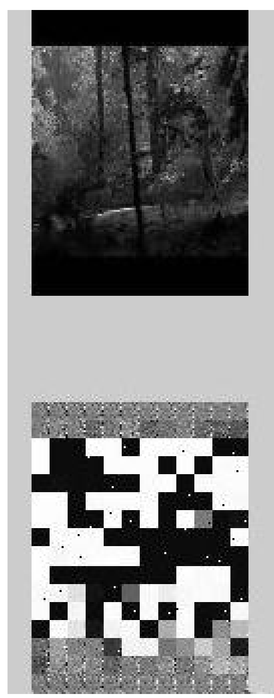
The independent component are thus found to represent textures in video frames. This observation is anticipated because the video can be reconstructed using a linear combination of the textures stored in the independent components.



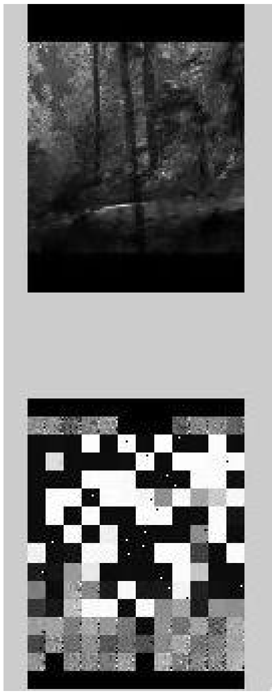
1



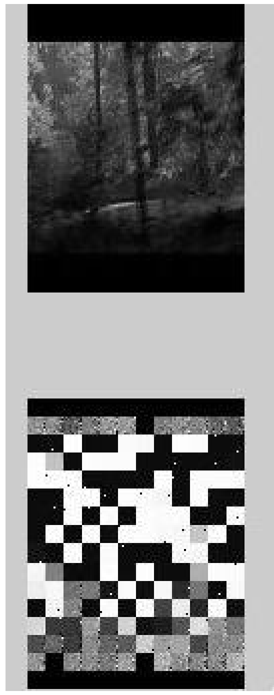
2



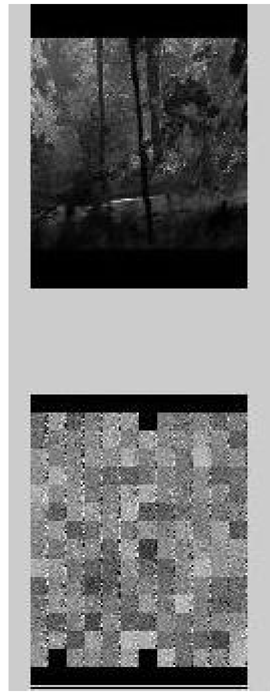
3



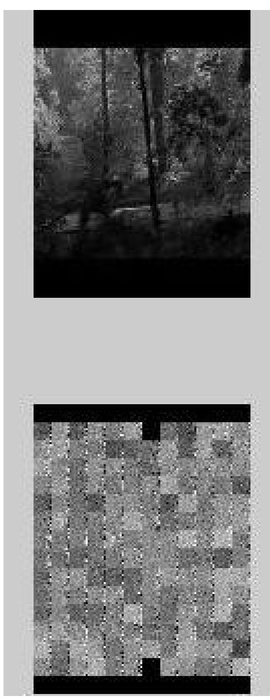
4



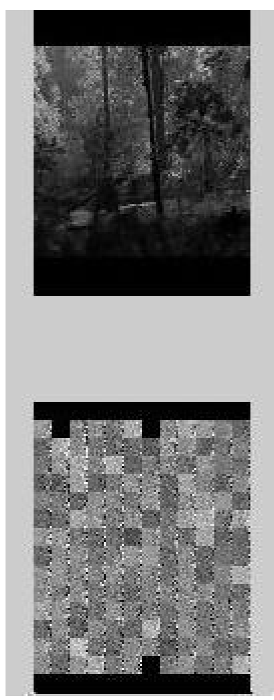
5



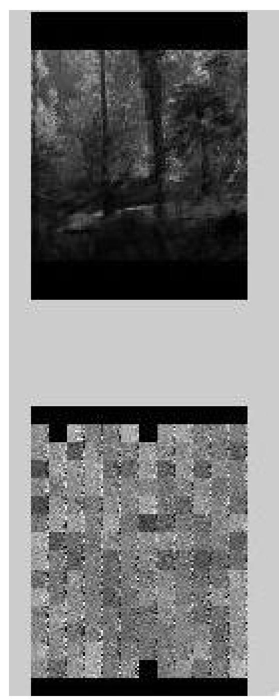
6



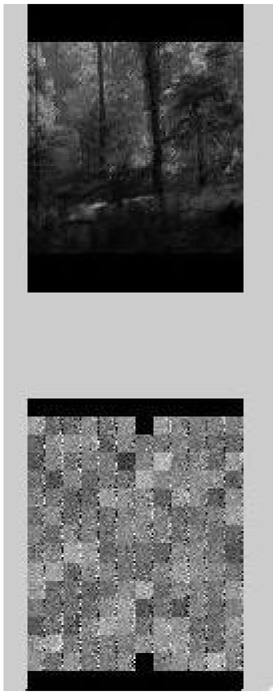
7

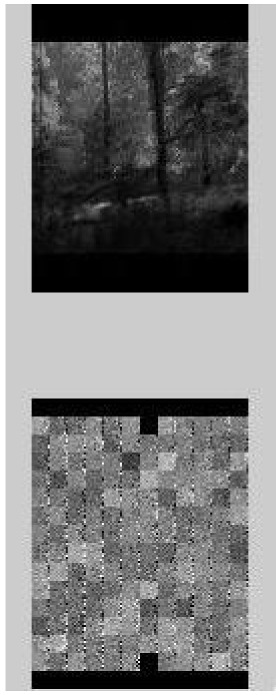


8



9


10


11

CONCLUSION

We have developed a new method for analysis of human understanding of videos. Two approaches are implemented: firstly, a frame-based approach using real ICA and secondly, a video-based approach, using complex ICA. Both methods lead to texture filters that relate to the Hubel & Wiesel visual receptive fields map. The code can be used as a library for ICA developments and video-processing on MATLAB. The work needs to be validated with bigger test sets and larger training samples. The image quality is a major reason for relevant results. There are other uses such as in Video multiplexing, which have not been considered, and can be looked upon in future.

The approach is compute-intensive in nature and also requires large amount of memory. The work needs to be validated with bigger test sets and larger training samples. The image quality is a major reason for relevant results. There are other uses such as in Video multiplexing, which have not been considered, and can be looked upon in future.

REFERENCES

1. P. Smaragdis, "Blind Separation of convolved mixtures in the frequency domain," *Neurocomputing*, Vol. 22, pp.21-34, 1998.
2. J. Anemuller, T.J. Sejnowski, and S. Makeig, "Complex independent component analysis of frequency-domain electroencephalographic data," *Neural Networks*, Vol. 16, pp. 1311-1323, 2003.
3. D.W. Dong, "Spatiotemporal Inseparability of Natural Images and Visual Sensitivities," In *Computational, Neural & Ecological Constraints of Visual Motion Processing*, J.M. Zanker & J. Zeil (Eds.) pp. 371-380, 2001.
4. J. Hurri, and A. Hyvarinen, "A Novel Temporal Generative Model of Natural Video as an Internal Model in Early Vision," In *Proceedings of the First International Workshop on Generative-Model-Based Vision (GMBV)*, 2002.
5. A.J. Bell, and T.J. Sejnowski, "The Independent Components of Natural Scenes are Edge Filters," *Vision Research*, Vol. 37, pp. 3327-38, 1997.
6. J.V. Stone, "Blind Source Separation Using Temporal Predictability," *Neural Computation*, Vol. 13, pp. 1559-1574, 2001.
7. H.L. Borgne, and A.G. Dugue, "Sparse-Dispersed Coding and Images Discrimination with Independent Component Analysis," In *Third International Conference on Independent Component Analysis and Signal Separation* (ICA'2001), San Diego, California, December 9-13, 2001.
8. H.L. Borgne, A.G. Dugué, A. Antoniadis, "Representation of images for classification with independent features," *Pattern Recognition Letters*, Vol. 25, pp 141-154, 2004.
9. H.L. Borgne, N. Guyader, A.G. Dugue, J. Hérault, "Classification of images: ICA filters Vs Human Perception," In *Proceedings of the Seventh International Symposium on Signal Processing and its Applications* (ISSPA) Vol. 2, pp. 251-254, 2003.
10. Y.M. Cheung, "Temporal Principal Component Analysis - Advances in Dual Auto-regressive Modeling for Blind Gaussian Process Identification," In *Proceedings of the IEEE International Conference on Systems, Man and Cybernetics* (IEEE-SMC), 2002.
11. F.R. Bach, M.I. Jordan, "Beyond independent components: trees and clusters," *Journal of Machine Learning Research*, Vol. 4, pp. 1205-1233, 2003.