VIDEO SEQUENCE PROCESSING BASED ON SPATIOTEMPORAL INDEPENDENT COMPONENT ANALYSIS

Zhenhe Chen and Xiao-Ping Zhang

Department of Electrical and Computer Engineering, Ryerson University

Toronto, Ontario, CANADA M5B 2K3

Email: {zchen, xzhang}@ee.ryerson.ca

Abstract

Compression and content-based video retrieval (CBVR) are essential needs for efficient and intelligent utilizations of vast multimedia databases over the Internet. In video sequences, object based extraction techniques are gaining importance in achieving compression and performing content-based video retrieval. In this paper, we propose a novel technique to extract objects from video sequences based on spatiotemporal independent component analysis (stICA), wavelet analysis and edge detection. The frame data obtained after stICA analysis is further processed using wavelet, and edge detection techniques to improve the accuracy of the extracted object. Initial studies demonstrate high potential for stICA based object extraction technique in comparison with existing techniques. Nine frames captured from a video sequence are used in our experiments and high extraction accuracy was achieved.

Keywords: Content-based Video Retrieval (CBVR), Independent Component Analysis (ICA), Spatiotemporal Independent Component Analysis (stICA), Wavelets.

1. INTRODUCTION

Content-based video retrieval (CBVR) and compression are essential needs for efficient and intelligent utilizations of vast multimedia databases over the Internet. In video sequences, object based extraction techniques are gaining importance in achieving compression and performing content-based video retrieval.

There are many object extraction techniques, such as average-max filters [1] and key frame representation [2]. However, average-max filters in [1] need an obtained background before processing. Matching between key

frames is difficult and inaccurate when automatic machine search and retrieval are demanded.

The novelty of the proposed methods in this paper lies in the extraction of independent objects through a background separation technique and in the processing of every independent frame of video sequences. In this paper, Independent Component Analysis (ICA) and Spatiotemporal Independent Component Analysis (stICA) are applied to the video frames respectively. The data obtained after ICA and stICA are further processed by wavelet analysis and edge detection to improve the accuracy. Experiment results show satisfactory outputs from the proposed system.

2. METHODOLOGIES

2.1 Extracting objects by Independent Component Analysis (ICA)

The video object segmentation system introduced in this paper uses both Independent Component Analysis and Spatiotemporal Independent Component Analysis (stICA) to extract objects from selected video frames.

ICA is a statistical technique introduced in the 1980s [3] in the context of neural network modeling. Over this period, ICA based techniques gained importance in the fields of signal processing, medical application, neural networks, information theory and telecommunications. ICA technique is based on a mixing model given by

$$\mathbf{x} = \mathbf{A}\mathbf{s} \quad (1)$$

where $\mathbf{x} = [x_1, x_2, \dots, x_n]$ is the observed mixture output, $\mathbf{A}$ is the mixing matrix whose coefficients are a_{ij} , $i, j = 1, \dots, n$, and $\mathbf{s} = [s_1, s_2, \dots, s_n]$ is the independent source vector.

In this model, one needs to find $\mathbf{s}$ simply by knowing the mixture signals $\mathbf{x}$. There are two constraints in ICA: (1) Source signal $\mathbf{s}$ must be non-Gaussian and (2) the components of $\mathbf{s}$ are statistically independent. Therefore, if the mixture signals can be decomposed into non-

Gaussian and statistically independent signals, these independent signals form the estimation of source signals.

In [4], Hyvarinen gave a pedagogical procedure to demonstrate that maximizing the nongaussianity could give independent components. He also illustrated that ICA estimation by minimization of mutual information is equivalent to maximizing the nongaussianities of the estimates of the independent components. These are the most popular approximation methods used in ICA algorithms.

In our experiments, video frames are selected as observed mixture signals $\mathbf{x}$, and Informax algorithm [5] is applied on these signals. Extracted signals $\mathbf{s}$ are obtained and further processed by wavelet analysis which will be introduced in the following sections.

2.2 Extracting Objects by Spatiotemporal Independent Component Analysis (stICA)

Spatiotemporal Independent Component Analysis (stICA) [6] is a generalization of classic ICA to blindly separate the independent sources from their spatial and temporal mixtures. It was initially developed in functional Magnetic Resonance Imaging (fMRI). In ICA model (1), the source matrix $\mathbf{s}$ is assumed to be independent and there are no constraints on the other matrix. But in stICA, not only spatial signals (images) are a set of independent components, but the time courses are also a set of independent components. If we take a transpose of (1), we have

$$\mathbf{x}' = \mathbf{s}' \mathbf{A}' \quad (2)$$

Now, $\mathbf{s}$ also looks like a mixing matrix. Here $\mathbf{A}$ and $\mathbf{s}$ are generated by independent random variables.

Stone, *et al.* introduced this technique of stICA [6] on fMRI. In his work, he describes a $m \times n$ matrix containing a sequence of n images $\mathbf{X} = [\mathbf{x}_1, \dots, \mathbf{x}_n]$. Each image $\mathbf{x}_i$ is a $m \times 1$ vector. A linear decomposition into k modes is defined by a matrix factorization

$$\mathbf{X} = \mathbf{S} \mathbf{\Lambda} \mathbf{T}' \quad (3)$$

where $\mathbf{S} = [\mathbf{s}_1, \dots, \mathbf{s}_k]$, $\mathbf{T} = [\mathbf{t}_1, \dots, \mathbf{t}_k]$ and $\mathbf{\Lambda}$ is a diagonal matrix of scaling parameters. Independent image vectors $\mathbf{s}_i$ are the column of spatial images $\mathbf{S}$ and corresponding independent time courses $\mathbf{t}_i$ are the column of $\mathbf{T}$.

Singular value decomposition (SVD) can factorize $\mathbf{X} = \mathbf{UDV}'$, where $\mathbf{U}$ is a $m \times k$ matrix of $k \leq m$ eigenimages, $\mathbf{V}$ is a $n \times k$ matrix of $k \leq n$ eigensequences, and $\mathbf{D}$ is a diagonal matrix of singular values. In order to determine the independent $\mathbf{S}$ and $\mathbf{T}$, we assume two $k \times k$ unmixing matrices $\mathbf{W}_S$ and $\mathbf{W}_T$ exist such that $\mathbf{S} = \tilde{\mathbf{U}} \mathbf{W}_S$ and $\mathbf{T} = \tilde{\mathbf{V}} \mathbf{W}_T$, where $\tilde{\mathbf{U}} = \mathbf{UD}^{1/2}$ and $\tilde{\mathbf{V}} = \mathbf{VD}^{1/2}$. Now we have:

$$\mathbf{X} = \mathbf{SAT}' = \tilde{\mathbf{U}} \mathbf{W}_S (\tilde{\mathbf{V}} \mathbf{W}_T)' = \tilde{\mathbf{U}} \mathbf{W}_S \mathbf{W}_T' \tilde{\mathbf{V}}' \quad (4)$$

One can recover the image signals and time courses at the same time using minimization of mutual information similar to normal ICA approximation techniques. Please refer to [6] for details.

In the proposed techniques as shown in Figure 1, we select the video sequences as observed images $\mathbf{X}$ and apply stICA on them. Using SVD, video sequence frames are decomposed into two parts, eigenimages and corresponding eigen time courses.

$$\mathbf{X} = \mathbf{UDV}' = (\mathbf{UD}^{1/2})(\mathbf{VD}^{1/2})' = \tilde{\mathbf{U}}\tilde{\mathbf{V}}' \quad (5)$$

Informax criterion is used on both spatial and temporal matrices instantaneously: conjugate gradient minimization is implemented to approximate unmixing matrices $\mathbf{W}_S$ and $\mathbf{W}_T$, and find the independent components $\mathbf{S}$ and $\mathbf{T}$. Even though we extract the spatial signals $\mathbf{S}$, they are still coarser and will require fine tuning by postprocessing techniques as will be explained in the next sections.

2.3 Using Wavelet Analysis to locate Region of Interest (ROI)

As an excellent tool of image analysis, wavelet performs well in local edge detection. In general, large coefficients are introduced in the process of wavelet transformation when an edge transition is detected.

The extracted source images from both ICA and stICA are decomposed using Haar filter at the second level. Analyzed along the vertical direction, we measure the maximum of each column's absolute values. In our simulations, it can be noted that large variation in the coefficients are observed at the edges of the image.

But there are also some valleys between peaks, which can confuse decision making. To avoid missing information, moving overlap windows are used. Mean value is calculated for every four pixels. Overlap windows smooth the curve so that it covers all pixels. The experiments show that the valleys between two peaks are increased.

A high pass filter follows the moving windows and sets the pixels outside edges to be zeros, which form the background of the object.

Performing wavelet analysis along the horizontal direction by the same routine, rectangle regions of interest (ROI) as shown in Figure 3 and Figure 4 are obtained. In this way, ROIs which contain the objects of interest are then identified. The purpose of segmenting an ROI is to decrease the computational complexity for later postprocessing and to reduce the noise so that edges detection technique can achieve better results.

2.4 Outlining objects by Edge detection

For both ICA and stICA processed images, we apply edge detection such as the Sobel method in the rectangle ROIs. This gives us clear object edges as shown in Figure 5.

In comparing the outlined edge objects with the original video frame, final extracted objects are obtained, shown in Figure 6.

3. RESULTS AND DISCUSSIONS

In our experiments, a video sequence of 10 seconds duration is used and we randomly selected 9 frames as mixture signals for our system.

Each mixture is a 360×240 pixels image with 256 grayscale. In the first simulation, we applied the stICA algorithm. Nine images are obtained, but only the background picture is relatively clear and close to the original image. There are some blurs in the other eight pictures. Then this recovered background is used, and subtracted from all the original frames to get the initial processed images like Figure 3.

As mentioned earlier in section 2.3, our postprocessing begins with wavelet analysis along vertical and horizontal directions, which can successfully locate the region of interest to reduce the computation and complexity. For example, in Figures 3 to 5, we just need to consider an area of 179×51 , which is only 10.57% of the original area. And in Figure 4, it shows the ROI's location is so accurate that the interested object is completely tracked inside the rectangular area.

Following locating ROI, edge detection is applied. The stICA processed image helps us to overcome this problem by coarse detecting the edge and shows it as a black and white picture (Figure 5). Combining these outputs and original video frames, we can locate the edge of our interested objects (Figure 6).

All of the extracted objects are clear and identical. Now they are ready for security purpose usage or for content-based retrieval through the Internet.

In the second simulation, we apply the mixture data to classic ICA. Then the same postprocessing is done. The final result is given in Figure 7.

For comparison, the peak-signal-to-noise-ratio (PSNR) of the experiment outputs compared to the real object is calculated in Table 1. It can be seen that stICA based extraction is superior to the conventional ICA.

4. CONCLUSIONS

We present a new method to extract objects from video sequences. stICA, wavelet analysis and edge detection techniques are used. The traditional ICA approach is compared with the proposed technique.

The experiments prove that stICA method can restore a fixed camera background without supervision successfully and stICA's results are better than the classic ICA's. The reason is that stICA concerns both spatial and time sequences which are very important in video analysis, but ICA only concerns on spatial sequences.

Future work will involve improving the stICA algorithm to produce more accurate source images thereby reducing the post processing requirement. Other postprocessing techniques will also be experimented for better ROI detections.

References

- [1] A. Amer, "Object and Event Extraction for Video Processing and Representation in On-Line Video Application". Ph.D. thesis, INRS-Telecommunications, December 2001.
- [2] P. Gresele and T. Huang, "Gisting of video documents: A key frames selection algorithm using relative activity measure". *The 2nd Int. Conf. On Visual Information Systems*, San Diego, CA, USA, 1997.
- [3] J. HERAULT, C. JUTTEN, "Space or time Adaptive Signal Processing by Neural Networks Model". *Int. Conf. on Neural Networks for computing*, Snowbird Utah, USA, 1986.
- [4] Hyvarinen, Juha Karhunen, *Independent Component Analysis*. New York, NY: John Wiley & Sons, 2001.
- [5] T. Lee, M. Girolami and T. Sejnowski, "Independent component analysis using an extended infomax algorithm for mixed sub-Gaussian and super-Gaussian sources". *4th Joint Symposium on Neural Computation*, Institute for Neural Computation, vol. 7, pp. 132-139, June 1997.
- [6] JV Stone, *et al.*, "Spatial, temporal, and spatiotemporal independent component analysis of fMRI data". In R.G. Aykroyd K.V. Mardia and I.L. Drydent, editors, *Proceedings of the 18th Leeds Statistical Research Workshop on Spatial-Temporal Modeling and its applications*, pages 23 – 28. Leeds University Press, 1999.

Fig. 1. Block diagram of the proposed technique

Fig. 2. Original Frame

Fig. 3. ROI of stICA output image

Fig. 4. ROI of original image

Fig. 5. Edge detect using Sobel method

Fig. 6. Extracted object by stICA

Fig. 7. Extracted object by ICA

Table 1. Comparison between stICA and ICA

	PSNR
stICA	59.9256
ICA	57.7385