Recent Development of Visual / Audio Communication

Charles Hsu, Nick Karangelen
Trident Systems Inc.,
Fairfax Virginia 22030, USA

Abstract

Recent advances in areas of both the Wavelet Transforms (WT) [1,2,3] representing human visual system and neural network [4,5] have resulted in improved visual and audio communication. These software techniques are capable of achieving quality performance and reducing computational complexity in order to process a real time live video via narrow band communication such as radio. The technical solution integrated with computers can potentially provide us many applications including remote sensors, security systems, commercial and home video teleconferencing. This paper describes a low cost board to support a video compression, fully software implementation for real-time video compression, audio compression. We also introduce a novel technology called independent component analyses (ICA). De-noise experimental simulation has been successfully demonstrated using ICA.

1. Introduce

In recent years, visualized communications have become an exploding market due to current advances in video / image compression, and the maturity of digital signal processors (DSP) [6]. Standards such as those of the Motion Picture Expert Group (MPEG), the Motion Joint Photographic Expert Group (Motion JPEG) and the International Telegraph and Telephone Consultative Committee (CCITT) H.261, and H.263 were ratified for the applications of digital videodisc, digital satellite TV, high definition TV, digital videophone, and video/teleconferencing, etc. However, these standards based on the Discrete Cosine Transform (DCT) cannot process high-resolution video, and audio signals in real-time with acceptable quality. In order to have better visual and audio communication, technology of Wavelet Transform is introduced and becomes an international standard, called JPEG 2000.

In this paper, we will summarize recent development in visual / audio communication using wavelet. WaveNet program is described in session 2. WaveNet is Army funded project to transmit video via narrowband. Session 3 will show the experimental results for image and video compression. Session 4 shows the videoconferencing application. Unsupervised learning ICA is introduced in Session 5 and Session 6 is the conclusion.

2. WaveNet Program

Trident has developed a group of real-time video & still imagery processing technologies to meet the US Army’s image compression through narrowband transmission. Developed with funding from the Army’s Night Vision Electronics Sensors Directorate (NVESD), WaveNet can improve the performance of standard compression techniques on any type of imagery data--video, still, FLIR, or SAR—without any decrease in quality. In addition, the Trident Wavelet Codec can process imagery and video data in real time. The same techniques employed in the WaveNet pre- and post-processors have been used to enhance the Trident Wavelet Codec.

During the development of the WaveNet system design several tradeoffs were considered to meet
various applications. WaveNet compression system consists of the following functions; Image capture and frame grabber, preprocessing for region of interest (ROI) identification, image compression, transmission of compressed image data, receive transmitted data, decode and restore image, filter the image and/or enhance the ROI, and finally display the image. The top-level functional block diagram (or one very similar) shown in Fig.1.

Even at this level of abstraction, we can see that several important design decisions need to be made. Trade-off studies of the candidate compression schemes will need to be conducted. Will this system be based on wavelet, DCT or some other type of compression? What type of pre filters will be used, ROI, edge enhancement, background normalization? Once an image has been restored what, if any pre display filters will be required? These and many other questions are typically flushed out in simulations of the candidate algorithms. For these simulations to be meaningful, they must be evaluated in terms of performance, complexity, & resultant image quality. Sufficient amounts of data must be produced to conduct perception experiments; video applications [7,8] must run fast enough for an observer to make a reasonable estimate as to judge the quality and latency of the images. In most applications the cast and schedule constraints require that many of these tradeoffs and evaluations be conducted prior to make a major investment in hardware. Pure software solutions are often prohibitively slow, and do not map well to final system implementation. True general-purpose boards are costly and in general require some specialized programming expertise.

### 3. Results for Image / Video Compression

WaveImage Still Image and Video application was illustrated. First, we compared the WaveImage lossless still image compression algorithm to JPEG and the LuraWave displayed in Table 1, where JPEG is the current still image compression standard mainly based DCT operations. LuraTech (http://www.lurawave.com) is a technology company in image and document compression software and development employed by industry & governmental agencies.
Table 1. Comparison of Lossless image compression technologies.

<table>
<thead>
<tr>
<th>Technology</th>
<th>Original size</th>
<th>Compressed size</th>
<th>Ratio</th>
<th>Resource</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wavelmage</td>
<td>307,200</td>
<td>152,024</td>
<td>2.02:1</td>
<td>OSD-W08 (Trident)</td>
</tr>
<tr>
<td>TIFF</td>
<td>307,200</td>
<td>164,608</td>
<td>1.86:1</td>
<td>Adobe Photoshop 5.0</td>
</tr>
<tr>
<td>LuraWave</td>
<td>307,200</td>
<td>153,572</td>
<td>2.00:1</td>
<td><a href="http://www.luratech.com">http://www.luratech.com</a></td>
</tr>
</tbody>
</table>

Next, the Wavelmage Lossy Compression algorithm is demonstrated as follows. Five different levels compression of lossy image were processed using different technologies. The results were compared in terms of PSNR measurement and the compressed file size. The detailed PSNR versus compression ratio is plotted in Fig. 2. The red indicates the JPEG result, the green line is for LuraWave compression, and the blue line shows the performance using Wavelmage. It can be observed from Fig. 2 that both Wavelmage and LuraWave have better performance than JPEG since the JPEG red curve is lower than the Wavelmage blue line and the LuraWave green line. In this comparison, Peak Signal Noise Ratio (PSNR, measured in dB) is the widely used image quality measurement. Higher PSNR value indicates better image quality.

![Fig.2. Still Image Compression Comparison among Wavelmage, LuraWave, & JPEG](image)

![Fig.3. Compressions (90:1) using JPEG, and DWT of WaveNet.](image)

It is very important to notice that application of WaveNet in video processing not only increases the frame rate but also improves the image quality. Because of the smoothing operation of the pre-processor of WaveNet, H.263 can compress the video more effectively and have more frames processed. Table 2 shows the difference between H.263 without WaveNet and H.263 with WaveNet in different transmission rates. The original input video is CIF (352 x 288) format and 30 frames per second (fps) for 15 seconds (totally 450 frames). We can observe from Table 1 that H.263 without WaveNet can process 24 frames and 4 frames under the transmission rate in 16 K bits per second (bps) and 4.8 K bps respectively.
Compared to H.263 with WaveNet, the performance is about three-time superior, which can process 73 frames and 11 frames under the same transmission rate in 16 K bps and 4.8 K bps respectively.

<table>
<thead>
<tr>
<th>Communication Channel Data Rate</th>
<th>Total Frames for 15 seconds H.263</th>
<th>WaveNet</th>
</tr>
</thead>
<tbody>
<tr>
<td>16 K bits/sec</td>
<td>24</td>
<td>73</td>
</tr>
<tr>
<td>4.8 K bits/sec</td>
<td>4</td>
<td>11</td>
</tr>
</tbody>
</table>

Table 2. Comparison between H.263 without WaveNet and H.263 with WaveNet

In addition, we have successfully developed audio compression using WT. The WT audio compression using can achieve 150 k bps for CD quality without any psycho acoustic analysis.

4. Application: Videoconferencing

In videoconferencing, quality is often judged by perception, rather than by measurement. Quality depends on a variety of factors, including picture resolution, frame refresh rate and artifacts. The last are the spurious, details that simply look wrong in the picture, such as the “blackness” in H.263 images. In addition, the audio must sound appropriate to the listener’s ears, and the audio and video must be synchronized. Finally, in an interactive system such as videoconferencing, there must be low latency between sending and receiving the signals. Due to the requirement of high fidelity, high compression, multi-resolution, multiple displays in teleconferencing, the technology of discrete biorthogonal subband wavelet transform [7-12] plays an important role. In order to achieve a symmetric N-N videoconferencing, every terminal needs to encode and decode the video images and the codec must be built in every terminal In addition, multi-resolution of DWT provides the need of sharing video information concurrently. For instance, in an N x N resolution system, four terminals can share the N x N resolution by decoding the information except the DWT coefficients at the first level. Follow the principle, \( \log_2(N) \) of the terminals can share the resolution.

Fig.4. Block diagram of the symmetric N-to-N teleconferencing system, the transmission data stream.
Recording to the construction of sharing video information, the mechanism of bandwidth in the local area visual communication network is remarkably significant in data transmission. A center clock is required to synchronize and distribute the bandwidth of the local area visual communication network. The WaveNet terminal is synchronized by a specific block via receiving the header from the center clock control unit. The header includes the status of WaveNet terminals, levels of DWT, the priority sequence of WaveNet terminals, and the identification code for each WaveNet terminal. The block diagram of the symmetric N-to-N videoconferencing system is shown in Fig. 4. In Fig.4, the security code is also included to secure conversation during videoconferencing. The transmission data stream and an example of 4 users in videoconferencing are shown in Fig.5.

![Fig. 5 An example of multiple users in conferencing](image)

5. Unsupervised Learning & Human Vision Principles

An unsupervised learning strategy of Artificial Neural Network (ANN) is to change the weight matrix $[W]$ of ANN to sieve or so-to-speak squeeze in parallel all the useful information from the time series of input vector $x(t)$ until the output vector $u(t) = [W] x(t)$ contains no more useful information at maximum entropy $H(u)$ shown in Fig.6. In other words, hopefully all the good stuff is already kept in the memory weight matrix $[W]$. This strategy is different to the supervised learning, because one can not assume any specific output goal for the input, except the most natural one “garbage-output for any useful information-input” by the strict definition of no supervision. Such an intelligent learning may be described by the motto of neurocomputers—“data-in & garbage-out” as opposed to the usual motto—“garbage-in & garbage-out” in traditional & non-intelligent computers. This new paradigm is useful for solving a statistically matrix inversion $[A]^T$ which mathematically underlies the Independent Component Analyses (ICA) [13-18] as follows: $u(t) = [W] x(t) = [W] [A] s(t)$, where $t$ stands for both time of signal and the scanning order of pixels. If the learning of weight matrix $[W]$ can achieves the maximum entropy $H(u)$ of the output $u$ or the linear slope portion of the maximum entropy sigmoidal neuron output $H(y)=H(\sigma(u))=H(u)$ which implies that all nth moments of the ANN output components $u=[u_1, u_2]$ of two sensor neurons are independent in terms of the normalized statistical histograms $\rho(u)$ defined as: $<u^r> = \int u^r \rho(u) du$. Specifically, the whitening of the second moment of the output shows: $<u(t)u^T(t)> = [W][A] <s(t) s^T(t)> [A]^T[W]^T = [I]$ This Oja’s sphering is equivalent to $[W] = [A]^{-1}$ provide that statistical de-correlation of sources $<s(t) s^T(t)> = [I]$ is true (if not, pre-whitening filter $[Wz] = <x(t)x(t)^T>^{-1/2}$ is often used by Bell-Sejnowski et. al. The fourth cumulant, the Kurtosis $K(u)$, is often used by Helsinki’s group to seek the statistical matrix inversion. $K(u) = <u^4> - 3 ( <u^2> )^2$ in terms of a single weight vector update: $dw/dt = dK/dw$. The other weight vectors are found by the projection pursuits.

We apply the ICA algorithm to remove the noise from very old music records. The music quality is amazingly improved.
WT and artificial neural networks are one of the popular technologies due to the demanded visual / audio communications. Image / video compression using WT is introduced and their results are compared to current industrial standard JPEG. Videoconferencing using WT is presented. Unsupervised leaning ICA was demonstrated by the noise elimination of the old records.

6. Conclusion

Reference