Real-Time Implementation of Full-Search Vector Quantization on a Low Memory SIMD Architecture

Antonio Gentile1,2, Huy Cat1, Faouzi Kossentini2, Filippo Sorbello2 and D. Scott Wills1

1 School of Electrical and Computer Engineering
Packaging Research Center
Georgia Institute of Technology
Atlanta, Georgia 30332-1250

2 Department of Electrical Engineering
University of Palermo
V.le delle Scienze, 90128 Palermo, Italy

3 Department of Electrical Engineering
University of British Columbia
Vancouver BC V6T 1Z4

Abstract

Vector quantization (VQ) has become a popular technique for image compression. While conventional unstructured VQs have the potential of achieving the best theoretical performance, they are also very demanding in terms of both storage and computational requirements. A significant amount of current research on VQ implementations addresses increasing the speed of image encoding, which is one of the more computationally expensive operations. This is typically accomplished by imposing structures, exploiting properties of the distance measure, or developing efficient and fast implementations. This research proposes a parallel implementation of a full-search VQ encoding algorithm using a low memory, fine grain SIMD pixel processor (SIMPil) being developed at Georgia Tech. This implementation fully exposes the available parallelism of the encoding process and exploits the processing and I/O capabilities of the processor, resulting in a system that can perform real-time image and video compression. The proposed implementation encodes a large region of the original image at once, replacing each constituent input block with its corresponding VQ codeword index. Preliminary simulation results indicate that the proposed implementation is capable of sustain real-time frame rates. In the simulation, 93,058 clock cycles are required to encode a single 64x64 region. The image of Lena (256x256) requires 16 passes to be completely encoded, for a total of about 1,472,089 clock cycles. With a 50 MHz processor, a 256x256 image frame will be encoded in 29.4 mS, supporting a frame rate of >30 frames/sec. The table below compares three VQ implementations on different hardware platforms. The performance is computer for four by four blocks. The MasPar performance is compensated for its larger codebook. A prototype SIMPil implementation is currently being fabricated by MOSIS in 0.8 μm CMOS.

<table>
<thead>
<tr>
<th>Distortion Measure</th>
<th>Platform</th>
<th>Clock Rate</th>
<th>Frame Size</th>
<th>Codebook Size</th>
<th>Encoding Time</th>
<th>Performance (blocks/sec)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Euclidean</td>
<td>2048 node MasPar 1</td>
<td>10 MHz</td>
<td>128 x 128</td>
<td>512</td>
<td>1.79 Secs</td>
<td>1,144</td>
</tr>
<tr>
<td>Euclidean</td>
<td>32 T800 nodes , 20 MHz</td>
<td>512 x 512</td>
<td>256</td>
<td>4.8 Secs</td>
<td>5,413</td>
<td></td>
</tr>
<tr>
<td>Absolute</td>
<td>256 node SIMPil</td>
<td>50 MHz</td>
<td>256 x 256</td>
<td>256</td>
<td>29.4 mS</td>
<td>159,000</td>
</tr>
</tbody>
</table>

* Research partially supported by the National Science Foundation contract #EEC-9402723 and by the Consiglio Nazionale delle Ricerche (Italy)