GPGPU Based Real-Time Object Detection & Recognition System

Dan Parks[1], Archit Jain[2], John McInerney[2], and Laurent Itti[1,2]

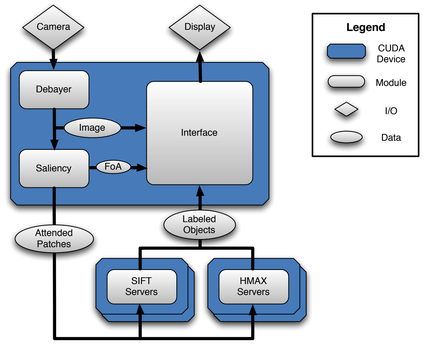
1. Neuroscience Graduate Program, University of Southern California, USA
2. Department of Computer Science, University of Southern California, USA

Introduction

CUDA Graphics Cards

- Vision algorithms computationally intensive
- Difficult to compare or tweak different algorithms.
- GPGPU allows speedups of 10-50x
- Implemented a GPGPU-based detection and recognition system
- Detection of regions is done using saliency (Itti, Koch, 1998)
- Recognition backend uses either HMAX (Riesenhuber, Poggio 1999) or SIFT (Lowe, 2004)
- Framework allows rapid testing and improvement of novel detection and recognition algorithms

1. Full system using CUDA



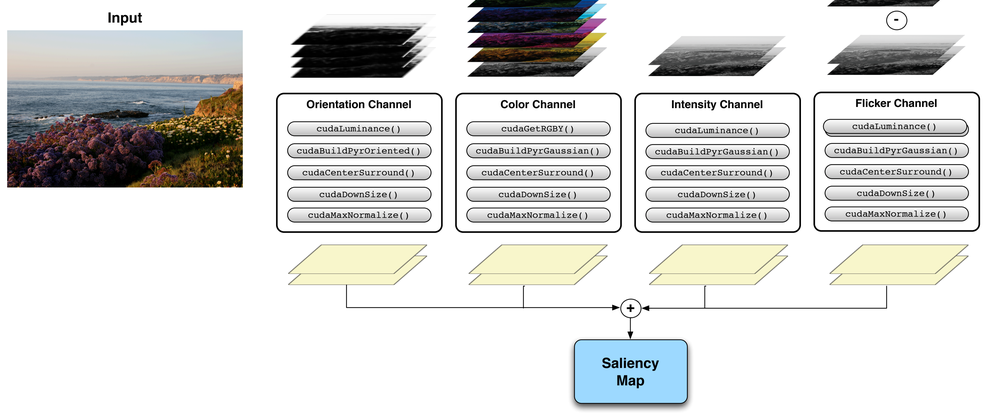
- System divided into two types of components:
- Display component displays the input stream and runs saliency (**Figure 3)**
- Recognition component gets image from the network and processes it using HMAX (**Figure 4)**, or SIFT and responds with a label.
- Display component displays the labels overlaid with the video.
- Any number of recognition servers can be running in the background in parallel.

2. Real time interface

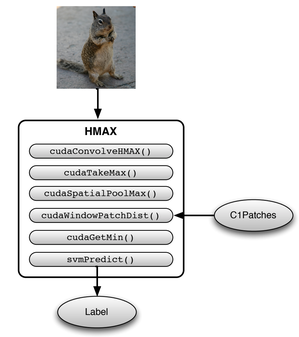
To download this poster see:
ilab.usc.edu/publications/doc/Parks_etal10vss.pdf
or take a picture of this QR Code:

To access our code see:
ilab.usc.edu/toolkit

3. Detection using saliency

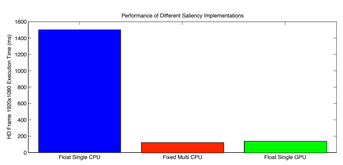


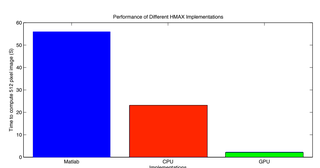
4. Recognition using HMAX



5. Speed Improvements

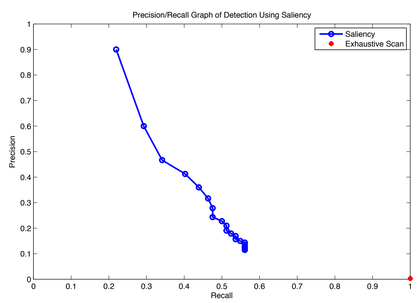
- Saliency sped up by 10.8x compared to single CPU floating point implementation
- HMAX sped up by 10x (Chikkerur, 2008)





6. Performance Of Saliency In Detection

- Saliency reliably detects objects in video within the first few saccades
- Precision drops off quickly thereafter
- Nonsalient objects are left, and performance stalls



Discussion

- Provides a realtime platform upon which to build and test detection, tracking, and objection recognition algorithms.
- Top down "cognitive" systems can utilize this underlying framework
- Prior knowledge of relations between objects in both space and time can be learned and utilized to enhance recognition. This information can then be further refined into recognition of events, and not just objects (e.g. ingress/egress by a person, a traffic collision). Further research is planned in this area.

References: Chikkerur (2008) CUDA implementation of a biologically inspired object recognition system
Itti, Koch (1998) A saliency-based search mechanism for over and covert shifts of visual attention. *Vision Research* 40:1489-1506.
Lowe (2004) Distinctive Image Features from Scale Invariant Keypoints. *International Journal of Computer Vision* 60:91-110.
Reisenhuber, Poggio (1999) Hierarchical Models of Object Recognition in Cortex. *Nature Neuroscience* 2:1019-1025.