Multi-Core DSP Enables Advanced Target Discrimination and Tracking

David Abrams
CTO
TrueSentry, Inc.
111 N. Chestnut St. Suite 200
Winston-Salem, NC 27101
www.truesentry.com
dave.abrams@truesentry.com

Bruce Schulman
VP of Applications
Cradle Technologies, Inc.
1197 Borregas Ave.
Sunnyvale, CA 94089
www.cradle.com
bschulman@cradle.com

ABSTRACT

Video surveillance must automatically detect targets and suspicious behaviors in order to effectively monitor large networks of cameras. Single processor designs have limited previous attempts at video content analysis and computer vision. We present a cost-effective multi-core approach that utilizes 16 DSP and 8 RISC processors on one chip, working in parallel to analyze video feeds for use in smart network cameras.

Keywords

Video content analytics, computer vision, tracking, digital signal processor, smart camera, embedded systems

INTRODUCTION

Homeland Security systems need to meet the challenges of complex environments like sea ports, subways, airports, trains, buses, and other transportation systems. In this paper, we present a software and system framework that is designed to take advantage of a parallel processing design for embedded systems. Parallel processing of surveillance camera video feeds enables cost-effective video content analysis using a new multi-core digital signal processor (MDSP) chip design.

This breakthrough technology makes it possible to employ multi-modal, semi-parametric kernel density estimates for pixels and objects in the scene, providing much better target discrimination than current commercial tracking systems.

Objects are analyzed in parallel with full probabilistic geometric and color model for each. Objects are discriminated regardless of occlusions, merging and splitting of groups.

In addition, we present a framework for building a probability model of typical paths through a scene over many weeks of time. This enables the system to learn typical tracks and identify non-standard behavior. Also, an audio analytics stage is included for filtering out background noise and segmenting audio events.

Finally, we present an analytics demo, which generates an alert notification that makes automated text-to-speech phone calls to emergency response teams, as well as to a 911 computer aided dispatch (CAD) system.

ANALYTICS IN PUBLIC SAFETY

Computer vision systems have been developed for use in transportation systems like cars and buses [1] and sea port monitoring [2]. Surveillance site monitoring [3] with object tracking has become more common in commercial systems.

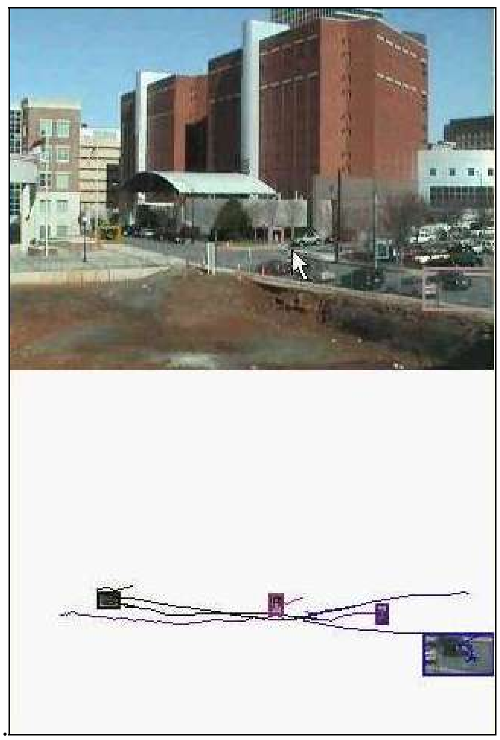

Figure 1: Government building surveillance: multiple-object tracking through occlusions in a complex outdoor scene

Our system uses a multi-modal background model and multiple-hypothesis tracker for accurate object tracking in outdoor scenes. The example above demonstrates four objects being tracked through occlusions and intersections. The following section describes our framework for at-the-edge DSP analytics using course-grain parallel processors with shared memory.

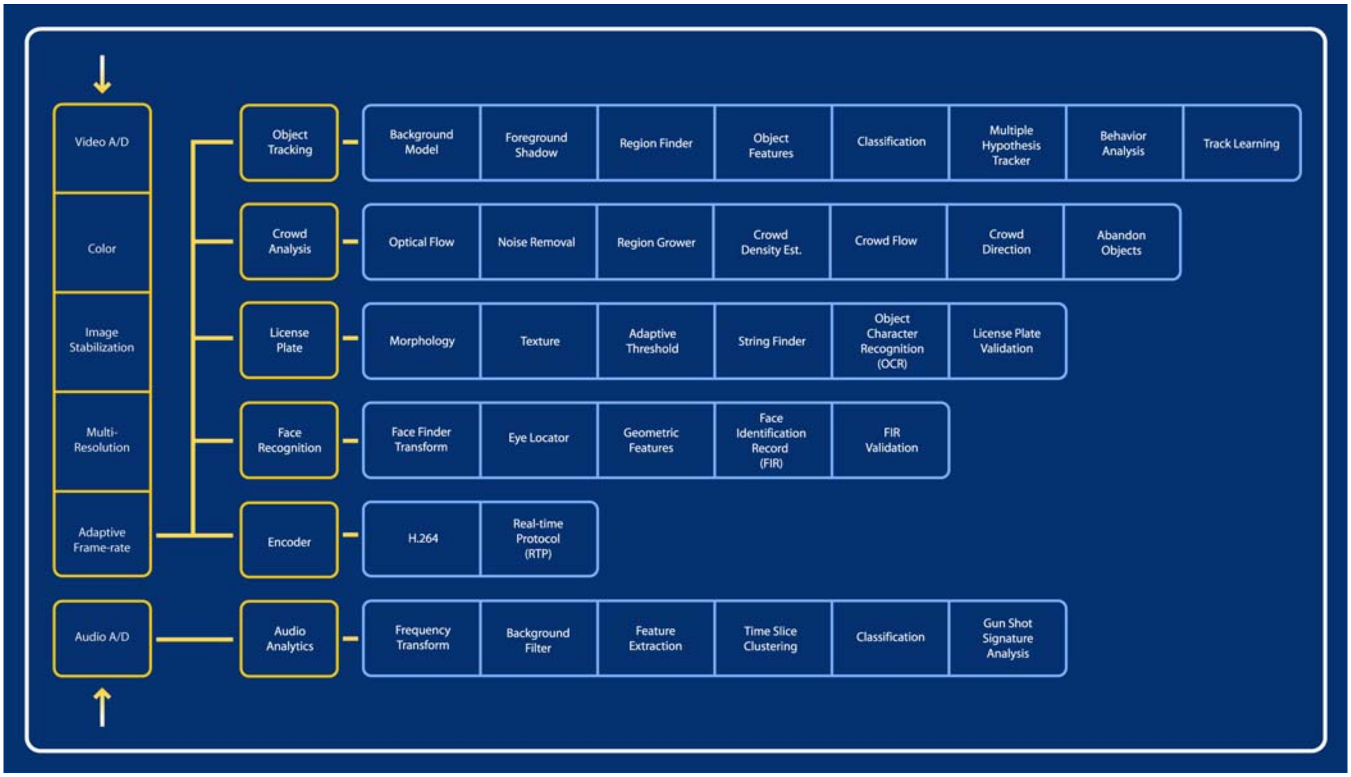

Figure 2: Embedded Analytics Pipeline for Homeland Security

PARALLEL EMBEDDED ANALYTICS

In Figure 2 we outline the algorithmic framework for analytics to process a single camera's video and audio content, including:

- Object tracking and learning – enter/exit zones
- Crowd analysis – crowd density and flow
- License plate recognition – string finder
- Face recognition – generate FIR of each face
- H.264 encoding – compress live video
- Audio signature analysis – gun shot detector

The framework is designed to compute multiple algorithms in parallel on the same camera feed. Each stage of the pipeline uses an asynchronous message to notify a controller with the results. In addition, multiple DSP's can be used to compute one stage (e.g., object tracking) on a single frame and use an efficient merge technique to compose the result. Careful memory management is required for this parallel technique to be effective.

Rate-Adaptive Video

After each frame is captured and pre-processed, we employ a novel adaptive frame-rate stage to share the next frame amongst each algorithm. This zero-copy image buffer locking mechanism employs hardware semaphores to give each analytics stage varying frame-rates. For example, we typically encode H.264 video at 30 fps, object tracking at 15 fps, and face recognition [4] could run at 1 fps. Controlling the temporal resolution of the analytics pipeline is essential because it enables run-time configuration changes for a public safety deployment.

Packed Probability Model

We use a packed data structure to represent multi-modal, semi-parametric kernel density estimates [5]. This is a much more memory efficient strategy for multi-dimensional probability models than the typical 3D color histograms used in many computer vision algorithms.

Scheduler and Memory Management

A job scheduler is used to queue analytics tasks for an available DSP. A tiered memory structure uses DMA transfers to move an image patch from SDRAM into local shared memory on the chip. Once in local memory, each DSP uses read/write FIFO's to provide pipelined, zero-latency access to the image data, background model, track database, object structures, and other data structures. We developed an efficient DSP register caching system in order to implement the traditional computer vision graph algorithms in DSP assembler.

Shared Memory

Although the dynamic job scheduler provides run-time flexibility for deploying the algorithms, we have found it necessary to carefully manage the precious shared memory by overlapping algorithms. For example, crowd analysis could get a hardware semaphore lock on the object tracker's output from background subtraction as an input to the optic flow stage. We can avoid extra DMA transfers by sharing the same local buffer.

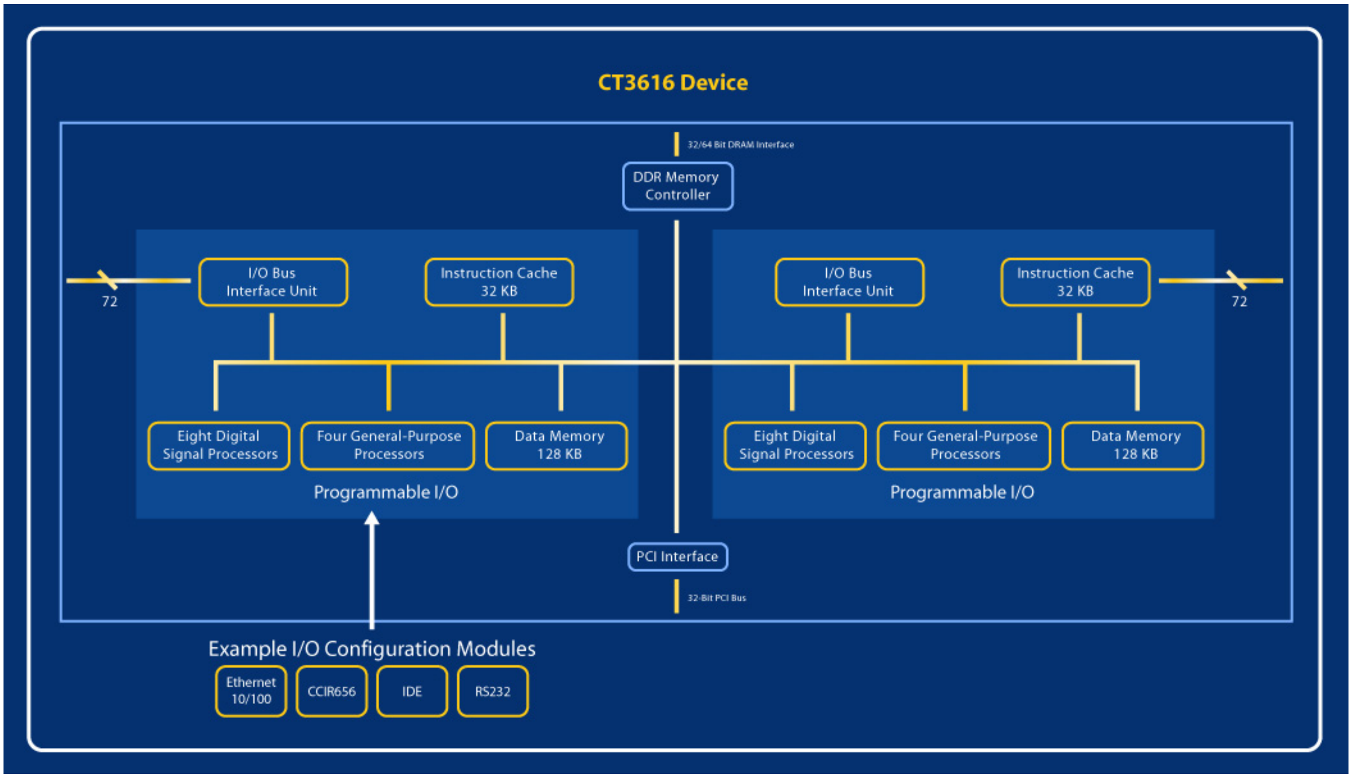

Figure 3: Compute Engine: CT3616 block diagram

MDSP COMPUTE ENGINE

The above compute engine shows two quads each with 8 DSP and 4 RISC processors connected to 128 KB shared memory. Hardware semaphores enable synchronization control. Each DSP can utilize packed instructions (e.g., sum-absolute-difference (SAD) for macro-block motion-estimation) for packed pixels computations. The RISC processor is typically used for scheduling DSP's, DMA transfers, and managing the algorithms.

We analyze audio/video with multiple, independent control threads, allowing multi-channel, multi-algorithm processing to scale in performance with more parallel elements. This can be used to increase the channel density for video analysis and compression on a single chip or increase the number and sophistication of analytics algorithms.

Analytics at the source

The advent of cost-effective DSP analytics enables us to move computation to the edge of the network. This avoids problems with traditional analytics running on a network server, because the performance is adversely affected by compression artifacts and image enhancement methods in today's security and network cameras. Our architecture gives the algorithms access to the original video.

Mega-Pixel Sensors

At-the-source computation can also utilize high resolution from mega-pixel image sensors. Current analog CCTV connected to a DVR is limited to 30 fps 720x480 NTSC video. Increasing the resolution allows each camera to cover a larger field-of-view keeping the spatial resolution constant. Analytics algorithms can run on a cropped window or multi-resolution pyramid from the mega-pixel image sensor.

Bandwidth Conservation

The encoder can provide H.264 video on high resolution images. The framework enables significant bandwidth conservation because we can dynamically increase the video encoder frame-rate, resolution, and lossy-compression parameters when analytics identifies a behavior and reduce it when no interesting activity is sensed.

Large Camera Deployments

Embedded analytics has significant economic advantages over deploying network servers in the field for Homeland Security deployments like borders, police cars, buses, traffic intersections, etc. For example, the original software in this application needed a dual Xeon CPU 2.8Ghz to process 4 channels. After porting to a single PCI board based on the compute engine the power savings are dramatic:

- 350 Watts – dual Xeon 2.8 GHz
- 10 Watts – single board computer with MDSP

These power savings greatly reduce unit cost, which enables more surveillance coverage in metropolitan and border deployments.

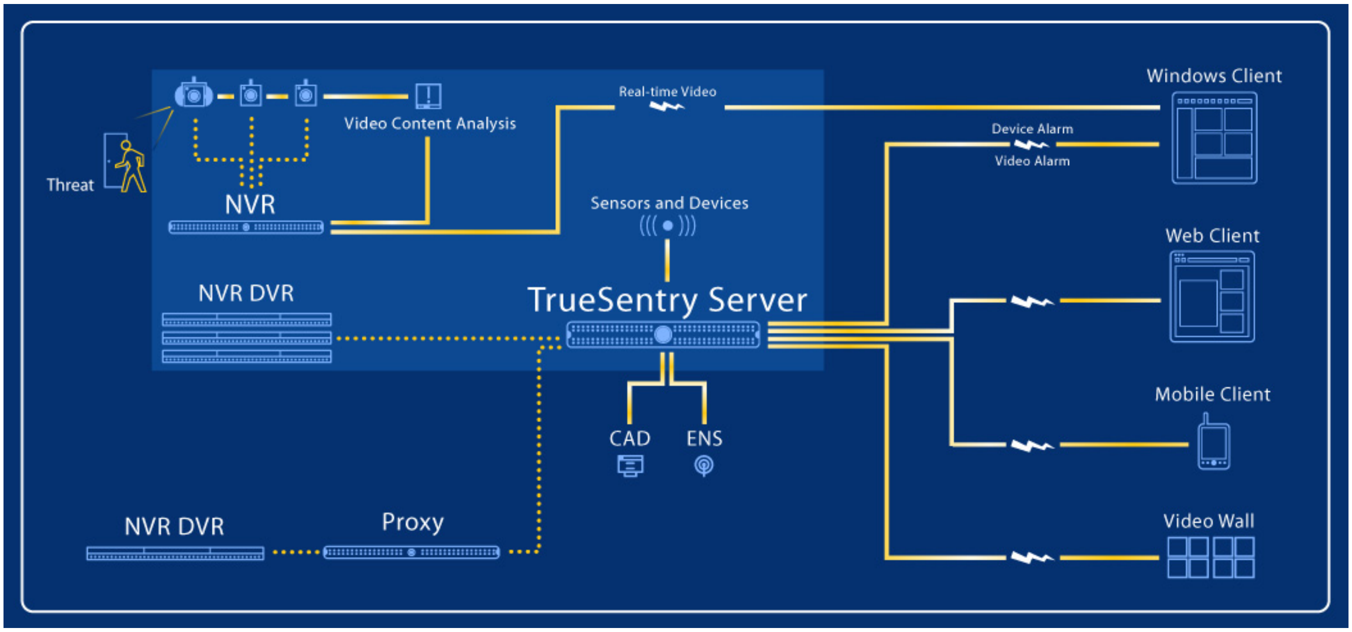

Figure 4: Command and Control Platform: TrueSentry connects analytics to 911 computer aided dispatch.

THREAT DETECTION AND RESPONSE

Video and audio content analytics are integrated into a public safety command and control platform in Figure 4. It shows how events detected via analytics are evaluated by a centralized server with a rules-engine.

Rules Engine

When an event is validated, a user-programmable action script is executed that can control devices (e.g., access control system, pan/tilt/zoom a camera), alarm video call-up on video wall, auto-upload video incidents to a central server for long-term storage. Devices and sensors, like fire panels; elevator panic buttons; and other building automation systems, can send alarms as well.

911 Computer Aided Dispatch

The system is integrated with a 911 computer aided dispatch (CAD) system for police and fire dispatch with GIS map overlay. The nearest fire, police, and emergency units appear on the GIS map for easy dispatch. The extended notification system (ENS) sends text-to-speech alerts to thousands of first responders in an emergency and tracks acknowledgements.

CONCLUSION

We presented a novel approach to audio and video content analysis for Homeland Security applications. The system utilizes an analytics pipeline designed for a new multi-core DSP chip. This provides the architecture for cost-effective, low-power smart cameras connected via the network to a command and control platform.

We employ rate-adaptive video for each stage in the pipeline: object tracking, crowd analysis, license plate recognition, face recognition and audio signature analysis for gunshot detection. A packed probability model is used along with efficient memory management of shared local chip memory to execute multiple analytics algorithms in parallel.

False alarm rates are greatly reduced with better analytics. This creates a force-multiplier for public safety personnel. Threats are intelligently identified, and a rules-engine notifies public safety personnel via 911 computer aided dispatch and first responder alert notification.

ACKNOWLEDGMENTS

We would like to thank Peter Quintas and Eugene Chua for their excellent contributions.

REFERENCES

1. Zehang Sun, George Bebis, Ronald Miller, "On-Road Vehicle Detection: A Review," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 5, pp. 694-711, May, 2006.
2. Chris Stauffer, et. al., "SeeCoast port surveillance," *Proc of SPIE Vol. 6204: Photonics for Port and Harbor Security II*, 2006.
3. Chris Stauffer, Eric Grimson, "Learning Patterns of Activity Using Real-Time Tracking," *IEEE Transactions on Pattern Recognition and Machine Intelligence (TPAMI)*, 22(8):747-757, 2000.
4. James Wayman, Anil Jain, Davide Maltoni, Dario Maio, *Biometric Systems: Technology, Design and Performance Evaluation*. Springer 2004.
5. John Shawe-Taylor, Nello Cristianini, *Kernel Methods for Pattern Analysis*. Cambridge University Press 2004.