

Context Extraction Scene Understanding

ObjectVideo

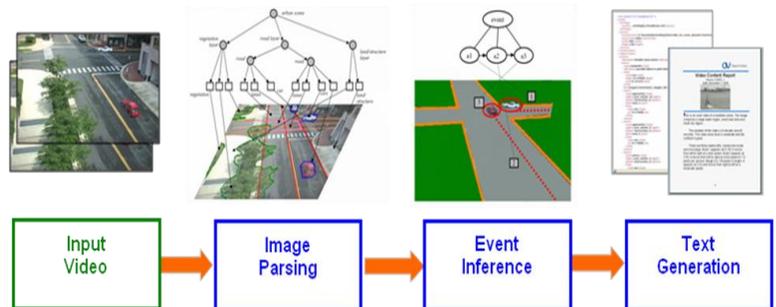
11600 Sunrise Valley Drive, Suite # 290
Reston, VA 20191

Contact: Mr. Paul Brewer

Phone: (703) 654-9300

Email: pbrewer@objectvideo.com

Website: www.objectvideo.com/



Command: ONR-SBIR

Topic: N06-147

PROBLEM STATEMENT

Our nation's most critical assets are protected by the video-based surveillance. However, Sandia National Labs has determined that humans are not well suited for the task of watching surveillance video and that the human attention span during such a task is less than 20 minutes.¹

Most critical video is never reviewed at all. And despite these facts, the deployment of video-based sensors is growing exponentially, as are the digital video management and storage platforms that support these cameras. The sheer volume of stored video data is simply unmanageable, and valuable information that is "hidden in the haystack" is simply inaccessible.

Intelligent video systems can automatically detect, classify and track persons, vehicles, and watercraft and infer threat potential through activity recognition. Such systems also offer the possibility of searching for patterns of intrusion or surveillance over an extended time course. Appropriate imagery is stored with tagging metadata. However, not everything in a scene can be tagged and in order to fully exploit surveillance imagery, there needs to be a capability for querying the database by either semantic queries for metadata search or by content-based image retrieval. Content-based retrieval tools and architecture that enable users of intelligent video surveillance systems to easily conduct video-based forensics and IMINT (IMagery INTelligence) from video imagery and display entire event histories for a given target or alerting event, regardless of whether the video data was tagged during acquisition.²

¹ Green, Mary W. (1999) The Appropriate and Effective Use of Security Technologies in U.S. Schools, *A Guide for Schools and Law Enforcement Agencies*, Sandia National Laboratories, http://www.ncjrs.gov/school/ch2a_5.html

² http://www.navysbir.com/n06_2/n062-147.htm

WHO CAN BENEFIT?

The capabilities that will be enabled by the technologies created by the Content Extraction Scene Understanding (CESU) project can benefit a wide range of applications that involve the use of video sensors, including:

- Critical infrastructure protection: NAVFAC, Naval Facilities Engineering Command, provides the security infrastructure for worldwide Naval facilities and is searching for technologies that provide better protection with minimum manning requirements. Other customers might include the USAF Integrated Base Defense Security System (IBDSS), the Pentagon Force Protection Agency and the Department of Homeland Security.
- ISR applications: Use of Unmanned Air Vehicles (UAVs) is on the rise. UAVs flew 258,502 mission hours in 2007, up from 27,201 in 2002. DoD is developing very large format sensors that can capture more video data than ever before. This mission application shares the same video overload issue.³
- Shipboard: The Fleet is also relying more heavily on camera technology and is developing new generations of periscope cameras, 360° mast-mounted cameras, etc. Potential users include the Littoral Combat Ship, PEO ships, and the Aegis Combat System.
- Maritime protection: Littoral waters, riverine environments and ports and harbors suffer from the need to monitor extremely large geographic areas with dozens or hundreds of cameras plus integrated radars, Automatic Identification System (AIS) receivers and other sensors. IVRS can abstract these data types and include them in user searches.
- Primes who might benefit from IVRS are Lockheed Martin, General Dynamics, SAIC, BAE, Boeing and Harris Corp.

BASELINE TECHNOLOGY

The current baseline for monitoring video is known as the “Mark I eyeball”, which is the colloquial phrase for relying on a human maintaining eyes on vigilance to, in this case, the “video wall” to identify interesting and threatening events. That is to say, video monitoring and retrieval is currently a manual process that begins and ends with people viewing video feeds either in real time or forensically by searching through mass video archives. Not only is human monitoring of security cameras ineffective, providing a watchstander on a 24x7 basis is cost prohibitive.

TECHNOLOGY DESCRIPTION

The CESU solution gives the human’s “Mark I eyeball” a means to extract relevant information and to convert that digital knowledge into searchable text reports that describe the video and enable efficient searches of mass video databases faster than real-time. A good analogy might be that it is similar to the difference between searching

³ <http://www.popsci.com/military-aviation-space/article/2008-02/warships-tomorrow>

"Approved for public release; distribution is unlimited."

through the card catalog and the library stacks versus conducting an iterative Google search.

CESU uses a stochastic attribute image grammar framework to extract semantic and contextual content. In this framework, a visual vocabulary is defined from pixels, primitives, parts, objects and scenes. The grammar provides a principled mechanism to list visual elements and objects present in the scene and describe how they are related. The relations can be spatial, temporal or compositional. Guided by bottom-up object and target detection, a bottom-up top-down strategy is used for inference to provide a description of the scene and its constituent elements. The visual content output is in a semantic representation format. A text generation system then converts the semantic information to text for automatic video annotation or as text reports read by human analysts or other users.

CESU will provide several search mechanisms including geo-registered rules-base searches, text-based queries, and context-based searches.

Table 1: Features, Advantages, and Benefits

Features	Advantages	Benefits
Creation of machine readable XML and human readable HTML video metadata	Turns video data into compact metadata that describes the video and is searchable through keyword- and rules-based queries.	Searchable video
Spatial database	100-900x faster method of searching video metadata	Responsive search engine enables users to iterate and “guide” additional searches.
Scene context	Automatically learn about the scene by watching over time. Can learn scene type (day/night, urban, maritime, UAV, etc), scene elements (roads, water regions, sidewalks, foliage) and normalcy (typical activities)	Automatic scene understanding and creation of scene metadata that represents the world in the language of the user.
Scene context inference	Enables users to search based on scene elements (“Search for a vehicle that drove onto <i>a sidewalk</i> ”)	More intuitive search elements.
Contextual scene labeling with cooperative and competitive edges	Labeling of objects (Vehicles, Buildings, etc.)	Enables more intuitive searches based on scene context.

CURRENT STATE OF DEVELOPMENT

ObjectVideo (OV) has developed a core video content indexing and retrieval system during the Phase II of the “Content Extraction Scene Understanding” SBIR project. The overall goal of the system is to make surveillance videos “Googleable”. The list of system capabilities that have already been developed include,

- Develop deformable templates learning for action recognition
- Incorporate shape detection for context scene labeling
- Develop tool to integrate GIS information
- Develop tool to integrate semantic web data (e.g. WordNet)
- Automatic generation of text to describe scenes and targets in video.
- Tools to allow the user to quickly search videos based on extracted data
 - Intuitive GUIs for query formulation
 - Report/Visualization Tools
- Integrate query parsing with SPARQL query engine
- Investigate use of special functions within SPARQL, (e.g. to compute geo-distance between two points)
- Investigate use of spatial database for geographic query and develop prototype software interfaces

"Approved for public release; distribution is unlimited."

Please see Table A for a list of milestones achieved, associated dates and TRL levels. OV's next objective is the evaluation of the system in a relevant maritime/urban environment. OV has selected the NAVSEA base in Port Panama City, FL for the prototype demonstration. Please see Table B for a complete schedule of upcoming tests and demonstrations.

ObjectVideo has a proven track record of taking product to market through Large Systems Integrators (LSI) and as a technology module to a manufacturer's product. In either case, ObjectVideo works with that partner to integrate the technology and to provide field engineering expertise, training and post-sales support.

Table A: Milestones achieved.

SBIR Timeframe	Milestone	TRL	Risk	Measure of Success	TRL Date
Phase II- Year 1	Core Video Content Indexing and Retrieval System Development	5	Low	User can query videos (static, UAV,USV) for targets and events of interest. An intuitive graphical interface is available for query formulation. Visualization tools developed to aid rapid understanding of retrieval results.	September 2008
Phase II- Year 1	Automatic Video Annotation by employing contextual cues and scene grammars	5	Low	A complete textual description of the scene and its contents is automatically generated. The user can read/search this description for video retrieval	September 2008
Phase II- Year I	Integration with GIS database and SPARQL query engine	5	Low	Video queries can be formed using street/city names and geodetic coordinates	July 2009

Table B: Objectives for year II of the SBIR project.

SBIR Timeframe	Required Tests, Demos & Next Steps	TRL	Target Date
Phase II- Year 2	Experimentation & Demonstration on NAVSEA Port Panama City Video data.	6	December 2009
Phase II- Year 2	Incorporation of user feedback from the NAVSEA demo. Experimentation/ testing on Tera-Byte Datasets.	7	April 2010
Phase II- Year 2	Prototype demonstration and evaluation	8	July 2010

"Approved for public release; distribution is unlimited."

REFERENCES

ONR-SBIR
Technical Point of Contact
703-696-4503

The following individuals are independent, 3rd party referrals, that can be contacted to discuss our technology:

Mary Ann F. Harrison, Ph.D. Senior Scientist Physical Sciences and Technology Scientific Research Group WVHTC Foundation Phone #: (304) 333-6432 Email: mharrison@wvhtf.org	Brian Banks, Ph.D. Chief Systems Engineer SAIC Pacific Technology Division P: 808 875 9286 Email: brian.banks@saic.com
--	--

ABOUT THE COMPANY

ObjectVideo is the leading provider of intelligent video software for security, public safety, business intelligence gathering, process improvement and other applications. ObjectVideo software is primarily available to the market through original equipment manufacturers (OEMs) that embed ObjectVideo® OnBoard™ - digital signal processor-based technology - into a variety of devices. ObjectVideo's world-wide partner network includes industry leading manufacturing and technology partners.

Intelligent video analytics software from ObjectVideo is based on years of advanced computer vision expertise, resulting in field proven technology that automatically extracts useful information from video. The practical uses for the data extracted from video are limitless - from real-time electronic notifications about perimeter breaches to determining department store shopping patterns, from detecting tail-gaiters at secure entryways to measuring traffic density in tunnels, or from triggering alarms on stolen equipment to saving energy in unused building space.

ObjectVideo provides original equipment manufacturers (OEMs) with targeted software solutions that fill the requirements for a variety of industries today:

- Homeland
- Transportation
- Banking
- Retail
- Gaming
- Buildings