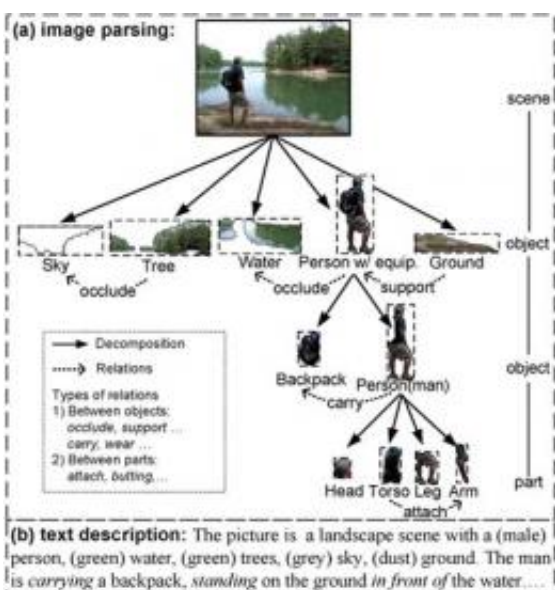


# New surveillance camera system provides text feed

June 3 2010, by Lin Edwards



Two major tasks of the I2T framework: (a) image parsing and (b) text description. Image credit: Benjamin Yao.

(PhysOrg.com) -- Scientists at the University of California in Los Angeles (UCLA) have developed a prototype surveillance camera and computer system to analyze the camera images and deliver a text feed describing what the camera is seeing. The new system aims to make searching vast amounts of video much more efficient.

The system was developed by Professor Song-Chun Zhu and colleagues Haifeng Gong and Benjamin Yao, in collaboration with the company

ObjectVideo, of Reston Virginia in the US. Dubbed I2T for Image to [Text](#), the system runs [video](#) frames through a series of vision algorithms to produce a textual summary of the contents of the frames. The text can then be indexed and stored in a database that can be searched using a simple text search. The system has been successfully demonstrated on surveillance footage.

The I2T system draws on a database of over two million images containing identified objects in over 500 classifications. The database was collected by Zhu starting in 2005 in Ezhou, China, with support from the Chinese government, but is still not large enough to allow the system to assess a dynamic situation correctly.

The first process in I2T is an image parser that analyzes an image and removes the background and identifies the shapes in the picture. The second part of the process determines the meanings of the shapes by referring to the image database. Zhu said that once the image is parsed transcribing the results into natural language “is not too hard.”

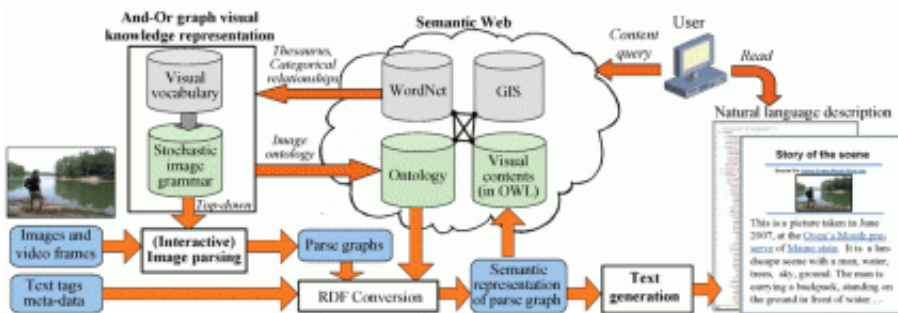


Diagram of the I2T framework. Image credit: Benjamin Yao.

The system also uses algorithms describing the movement of objects from one frame to another and can generate text describing motions,

such as “boat 3 approaches maritime marker at 40:01.” It can also sometimes match objects that have left and then re-entered a scene, and can describe events such as a car running a stop sign.

Professor Zhu said at the moment almost all searches for images within video is done using surrounding text, but the new system directly generates text from the images. He also added that the existence of YouTube and other video collections, and the expanding use of surveillance cameras everywhere show that being unable to efficiently search video is a major problem.

The I2T system is not yet advanced enough to recognize a large number of images instantly and is not ready yet for commercialization, but the researchers say it is close and needs only “minor tweaks.” The scientists also say they may be able to feed the text into a vocal synthesizer to increase its usefulness.

*You can now listen to all PhysOrg.com podcasts at*  
[www.physorg.com/podcasts-news/](http://www.physorg.com/podcasts-news/)

**More information:** -- Technical description of I2T: Image Parsing to Text Generation - [www.stat.ucla.edu/~zyyao/projects/I2T.htm](http://www.stat.ucla.edu/~zyyao/projects/I2T.htm)  
-- Research paper: Benjamin Yao, et al. I2T: Image Parsing to Text Description, Proceedings of IEEE [[pdf](#)].

© 2010 PhysOrg.com

Citation: New surveillance camera system provides text feed (2010, June 3) retrieved 20 March 2024 from <https://phys.org/news/2010-06-surveillance-camera-text.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is
---

provided for information purposes only.