

**MASSACHUSETTS INSTITUTE OF TECHNOLOGY
LINCOLN LABORATORY**

ADVANCED CONCEPTS COMMITTEE

PROPOSAL

Title: Automatic text recognition within imagery in the wild

Date: November 8, 2012

Authors: Peter Cho (G106)

Summary of technical concept:

Billions of digital images currently accessible on the internet represent an unprecedented and generally untapped source of intelligence. But imagery analysts cannot keep pace with the accelerating rates at which photos and videos are being shot. So exploitation algorithms are needed by the Intelligence Community to automatically interpret electronic pictures and flag those of special interest for subsequent human analysis.

The appearance of text in uncooperatively-gathered pictures often provides valuable clues about their topic domains, scene contents and even camera geolocations. Optical character recognition in benign contexts such as book-scanning is regarded as a solved problem in some academic circles. But machine reading of text in pictures shot at random perspectives, unknown zoom-levels and uncontrolled illumination conditions remains an outstanding technical challenge. We propose to extend existing machine-learning techniques and to develop new computer vision algorithms for recognizing text within arbitrary digital images.

We specifically plan to merge extremal region shape ideas developed by Neumann et al (CVPR 2012) with unsupervised feature learning approaches investigated by Coates et al (ICDAR 2011). The former can rapidly reject regions within images where no text exists. The latter can robustly classify nominated bounding boxes as belonging to text/non-text categories and subsequently establish character identity. We will also impose color, scale and simple language model consistency constraints upon extracted words. Algorithm performance will be quantified on a quasi-random set of 1K+ manually-truthed photos and video frames downloaded from the web. Our goal is to reliably recognize text in megapixel-sized images on a laptop computer at a rate exceeding one picture per minute. If successful, our results could be scaled to exploit large quantities of open-source imagery in the future.

Summary of programmatic concept:

As the quantity and quality of digital imagery exponentially increase, the impact of automatic text search for images could rival that of Google text search for documents. During recent preliminary discussions, National Geospatial Agency and Air Force representatives have expressed strong interest in our ideas. We also plan to interact with NSA, CIA and other potential intelligence agency sponsors in the coming year.

History and status:

Our program will build upon major advances made in machine intelligence (e.g. unsupervised feature learning, support vector machines) and computer vision (e.g. extremal region extraction, codebook generation) over the past decade. It will also provide an automated exploitation module for the Image Search System developed from 2010-2012 through the Lincoln Imaging Sciences Center initiative.

Requested support:

We request \$85K IOE + \$15K OP = \$100K to support this research program from Jan 1, 2013 through Dec 31, 2013.