**SANDIA REPORT**
SAND2005-6057
Unlimited Release
Printed October 2005

# Multi-scale Saliency Search in Image Analysis

Alexander Slepoy, Alexander Backer, and Anthony Campisi

Prepared by
Sandia National Laboratories
Albuquerque, New Mexico 87185 and Livermore, California 94550

Sandia is a multiprogram laboratory operated by Sandia Corporation,
a Lockheed Martin Company, for the United States Department of Energy's
National Nuclear Security Administration under Contract DE-AC04-94AL85000.

Approved for public release; further dissemination unlimited.

Sandia National Laboratories

Issued by Sandia National Laboratories, operated for the United States Department of Energy by Sandia Corporation.

**NOTICE:** This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government, nor any agency thereof, nor any of their employees, nor any of their contractors, subcontractors, or their employees, make any warranty, express or implied, or assume any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represent that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government, any agency thereof, or any of their contractors or subcontractors. The views and opinions expressed herein do not necessarily state or reflect those of the United States Government, any agency thereof, or any of their contractors.

Printed in the United States of America. This report has been reproduced directly from the best available copy.

Available to DOE and DOE contractors from
U.S. Department of Energy
Office of Scientific and Technical Information
P.O. Box 62
Oak Ridge, TN 37831

Telephone: (865) 576-8401
Facsimile: (865) 576-5728
E-Mail: reports@adonis.osti.gov
Online ordering: http://www.osti.gov/bridge

Available to the public from
U.S. Department of Commerce
National Technical Information Service
5285 Port Royal Rd.
Springfield, VA 22161

Telephone: (800) 553-6847
Facsimile: (703) 605-6900
E-Mail: orders@ntis.fedworld.gov
Online order: http://www.ntis.gov/help/ordermethods.asp?loc=7-4-0#online

SAND2005-6057
Unlimited Release
Printed October 2005

# Multi-scale Saliency Search in Image Analysis

Alexander Slepoy and Alexander Backer
Multiscale Computational Materials Methods Department

Anthony Campisi
Visualization & Data Department

Sandia National Laboratories
PO Box 5800, MS1110
Albuquerque, NM 87185

**Abstract**

Saliency detection in images is an important outstanding problem both in machine vision design and the understanding of human vision mechanisms. Recently, seminal work by Itti and Koch resulted in an effective saliency-detection algorithm. We reproduce the original algorithm in a software application Vision and explore its limitations. We propose extensions to the algorithm that promise to improve performance in the case of difficult-to-detect objects.

## 1. Introduction

A fundamental problem facing computer vision is information selection, which historically motivated the development of multi-resolution image representations such as pyramids and wavelets. However, straight applications of low-pass or band-pass filters are insufficient to extract interesting information, and it was not until the 1980's when specific efforts toward more powerful models began, inspired by neuro-physiological research (Milanese, 1993). Saliency detection is therefore concerned with finding where the "focus of attention" belongs within an image, and how it changes (or shifts) to another location.

Control of the focus of attention is believed to consist of a two-component framework (James 1890/1981; Nakayama and Mackeben, 1989). The first component is a fast, bottom-up primitive mechanism that selects certain areas of an image based on their saliency, while the second is a slower, top-down, approach which allows volitional, cognitive control of the "attentional spotlight" (Itti, Koch, 2000).

Koch and Ullman (Koch and Ullman, 1985) were the first to suggest the use of a "saliency map" for finding active locations in an image. A saliency map is an explicit 2-D image that encodes those areas which are found to be salient with the most salient location being represented by the brightest point, i. e. strongest response, in the image. There is some debate as to whether such a topographic map exists within the brain (Koch & Itti, 2000), however, for a computer algorithm it is a convenient structure.

Recently, Itti and Koch have developed an original algorithm for saliency search within input images. This exciting algorithm appears to reproduce human performance on a variety of images. We implement their algorithm in a Sandia code Vision. In this report, we describe our implementation, and demonstrate its performance on various images. As we shall show, the algorithm is surprisingly robust, and capable of handling realistic, cluttered scenes in an efficient manner with human-like, or superior, performance.

## 2. Model

The algorithm consists of several stages of filters that reduce the original information to channels and scales, then extract salient portions from each fraction, and, finally, recombine it to arrive at a consensus. The basic channels consist of color, intensity, and orientation. Scale decomposition is described below in Figure 1.

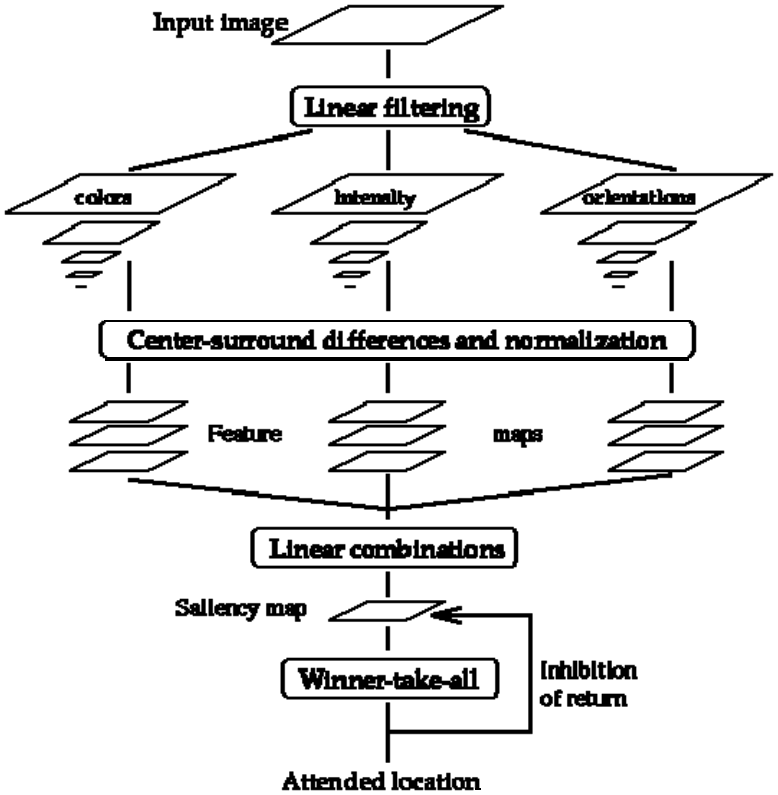


Fig. 1. Schematic diagram of the model used by Itti & Koch. Visual features are computed using linear filtering at 8 spatial scales, followed by center-surround differences, which compute local spatial contrast in each feature dimension for a total of 42 maps. A difference-of-Gaussians filter is iteratively applied to each map to instantiate competition for salience within each feature map. After competition, the feature maps are combined to create a single conspicuity for each feature (intensity, color, orientation). The three conspicuity maps are then combined to create the final saliency map (Itti & Koch, 2000). Our model differs from that of Itti & Koch, by simply selecting the most conspicuous point according to its relative intensity in the saliency map, instead of implementing a winner-take-all neural network.

1. Channels.

The original image is decomposed into a set of channels. The channels are of three types: intensity, color, and orientation.

   a. Intensity is a channel that considers the brightness of a pixel in a total amplitude sense.
   b. Four basic color channels (Red, Yellow, Blue, and Green) form a superposition of double opponency: R-G, B-Y. These choices are biologically motivated.
   c. Four orientation channels correspond to filters with kernels oriented at 0, 45, 90, and 135 degrees. This discrete representation is able to capture the critical distinctions in orientation.

2. Scales.

   a. Gaussian pyramids.
      A Gaussian pyramid consists of progressively lowpass filtering followed by subsampling (Burt & Adelson, 1983). The Gaussian pyramid is created to a depth of nine levels, with level 0 having a scale of 1:1 (the original input image) and level 8 being 1:256.

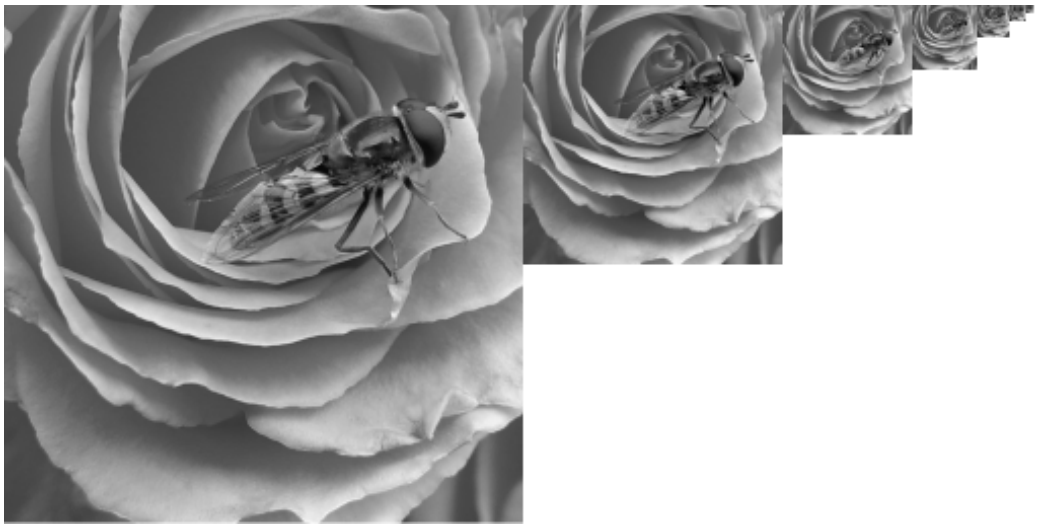

Fig. 2. Gaussian pyramid.

b. Laplacian pyramids.
Traditionally a Laplacian pyramid is created by taking the differences between adjacent levels in the Gaussian pyramid (Burt & Adelson 1983), however, for our purposes it suffices to create the Laplacian pyramid by taking the original image (in this case filtered by intensity), applying a lowpass filter, and taking the difference between the original and the low pass filtered images. The process is repeated to a depth of nine, by subsampling the low pass filtered image. Once the Laplacian pyramid has been created, the four orientation pyramids (0, 45, 90, and 135 degrees) are now generated by applying an oriented filtered to each image in the pyramid.

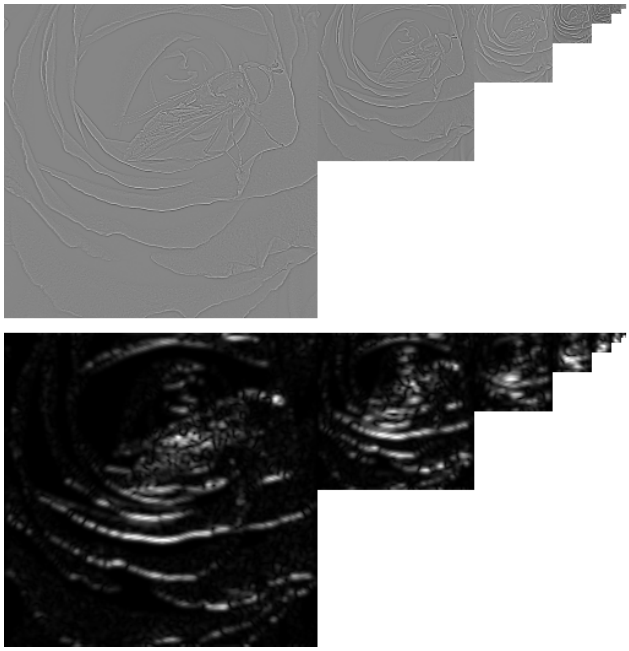

Fig. 3. Laplacian pyramid (top), and corresponding 0 degree orientation pyramid

c. Center-surround pyramids.

Taking the center at a fine scale, and subtracting the surround at a coarse scale accomplish this. The center corresponds to level c = {2, 3, 4} in the pyramid, and the surround to the level s = c + d, with d = {3, 4}. Thus six scales are created (2-5, 2-6, 3-6, 3-7, 4-7, 4-8).

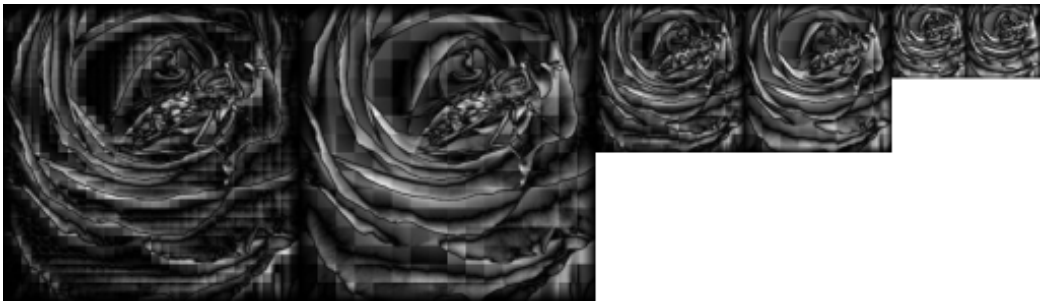

Fig. 4. Center surround pyramid, for Intensity channel.

3. Winner-take-all

a. Strategy of competitive selection.

b. Kernel of selection.

A large 2-D difference-of-Gaussians (DoG) filter is applied to reduce the noise from cluttered areas, and allow a peak to arise in each map. The DoG filter yields strong local excitation, which is counteracted by broad inhibition from neighboring locations.

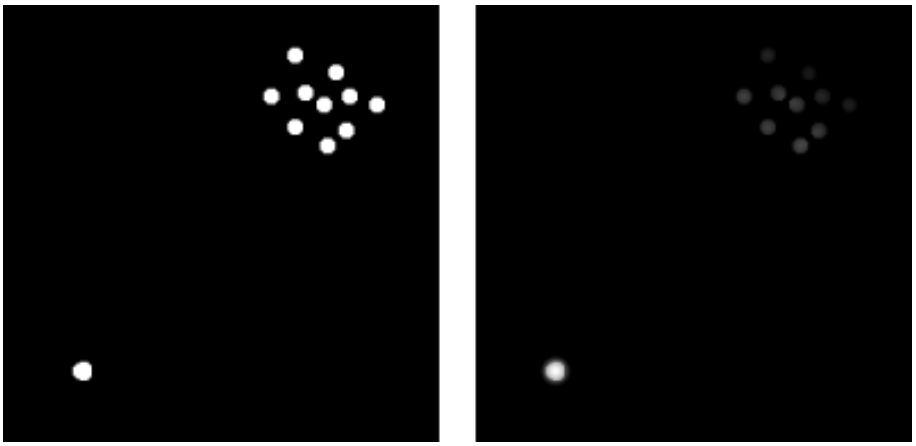

Fig. 5. Example of DoG filter in operation. On the left we have the original image, on the right we have the image after applying a 2-D difference-of-Gaussians kernel iteratively (15 times). Notice that the collection of “dots” in the upper right-hand corner has been suppressed, allowing the lone peak in the lower left-hand corner to emerge as the “winner.”

4. Synthesis of channel

a. Conspicuity Maps.

After each map has been subjected to the DoG filter for 5 iterations, they are summed together within their respective channels creating three conspicuity maps (intensity, color, and orientation).

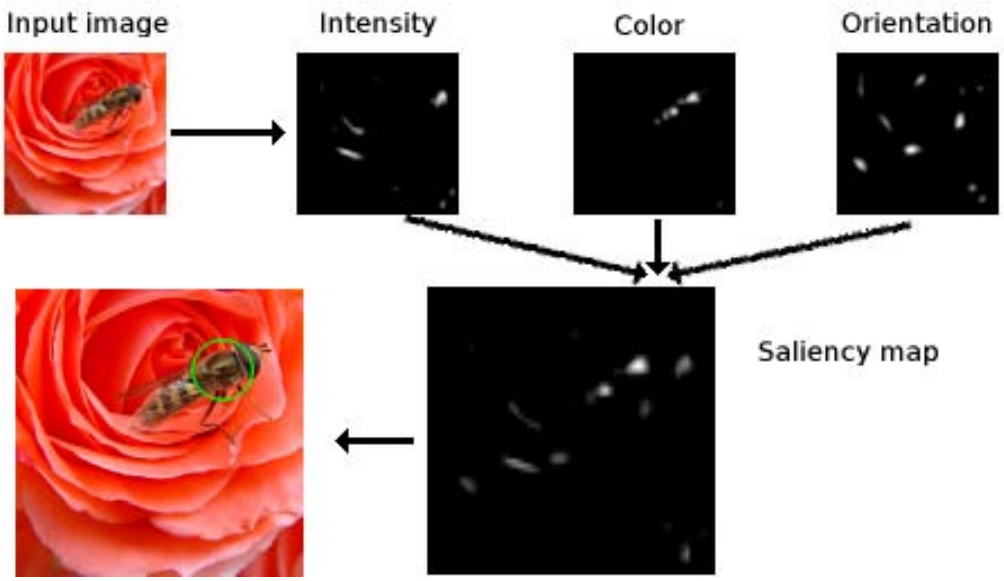


Fig. 6. Example of the model running on a 1024 x 1024 pixels color image. The three conspicuity maps (intensity, color, orientation) are the sum on their respective feature maps derived from the input image at different spatial scales. The three conspicuity maps are added together to create the saliency map. The most salient point in the image has been circled.

b. Saliency Map.
The three conspicuity maps then have the DoG filter applied to them and are weighted, before being summed to create the final saliency map. The DoG filter is then applied to the saliency map to achieve our final map.

c. Finding Multiple Areas of Interest.
Multiple areas of interest can be found in such a way that we can create a heirarchy of interest points within the image. An operation that can be performed directly to the saliency map without having to run the entire algorithm over again would be ideal. Thus once a salient point has been identified, we apply a variant of the DoG kernel, but only at that point. In this case we have strong local inhibition coupled with long range excitation, which suppresses the most salient point allowing another point to become salient (see figure 7). The process is applied recursively to find a user defined number of salient points.

## 3. Results

A variety of real and computer-generated images were tested on the algorithm, and the results indicate that the algorithm detects locations of interest in the image in a manner that makes functional sense. However, it should be noted that it is difficult to objectively define criteria for its performance on such images. Although there does exist exstensive information on visual scanpaths (the paths a subject's eyes follow when confronted with an image), it is unclear to what extent the precise trajectory followed is related to covert (bottom-up) shifts of attention or overt (top-down) shifts (Itti & Koch, 2000).

Itti and Koch, however, devised a test involving 44 images of camouflaged military vehicles in a natural setting. The algorithm was compared to 62 human observers, and the number of attentional shifts that the algorithm took before finding the vehicle were measured. On average the algorithm found the target in less time than the human observers.

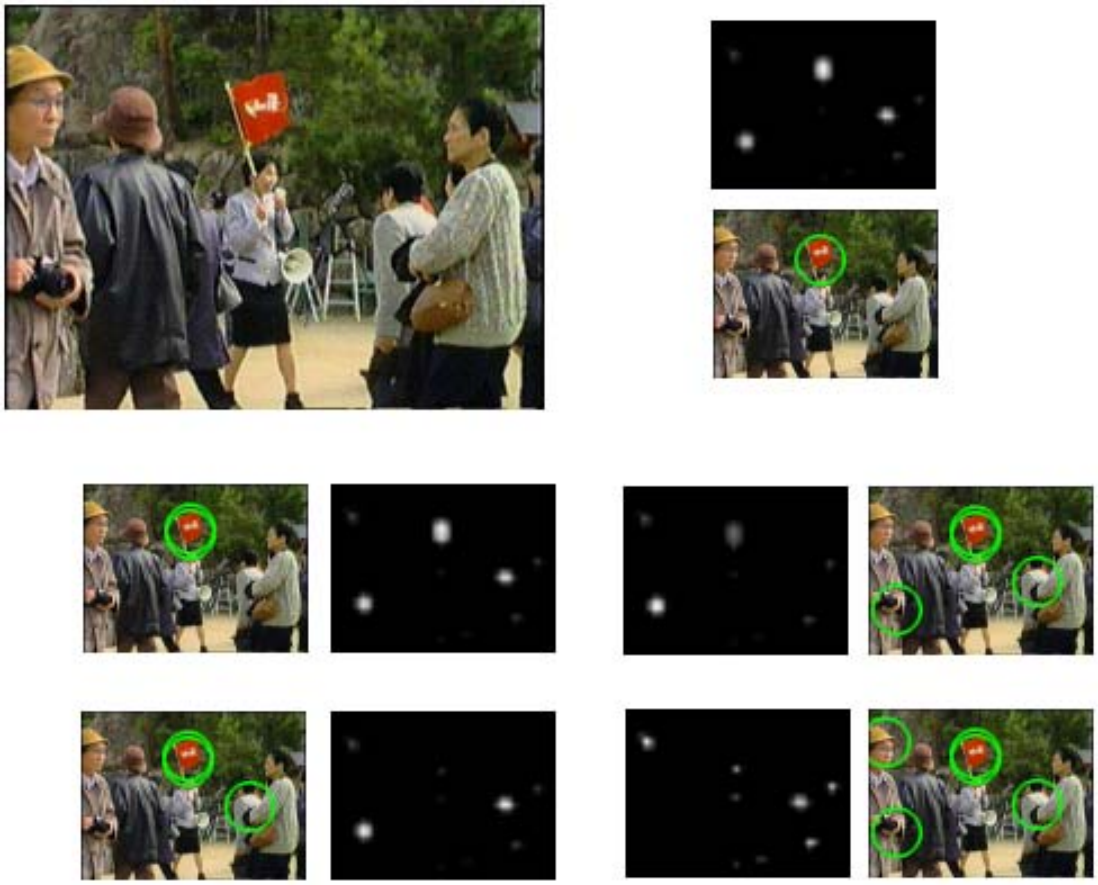

Fig. 7. Example of our program runnin on a 640 x 480 pixels image. Multiple areas of interest are detected quickly, by a simple recursive suppression method.

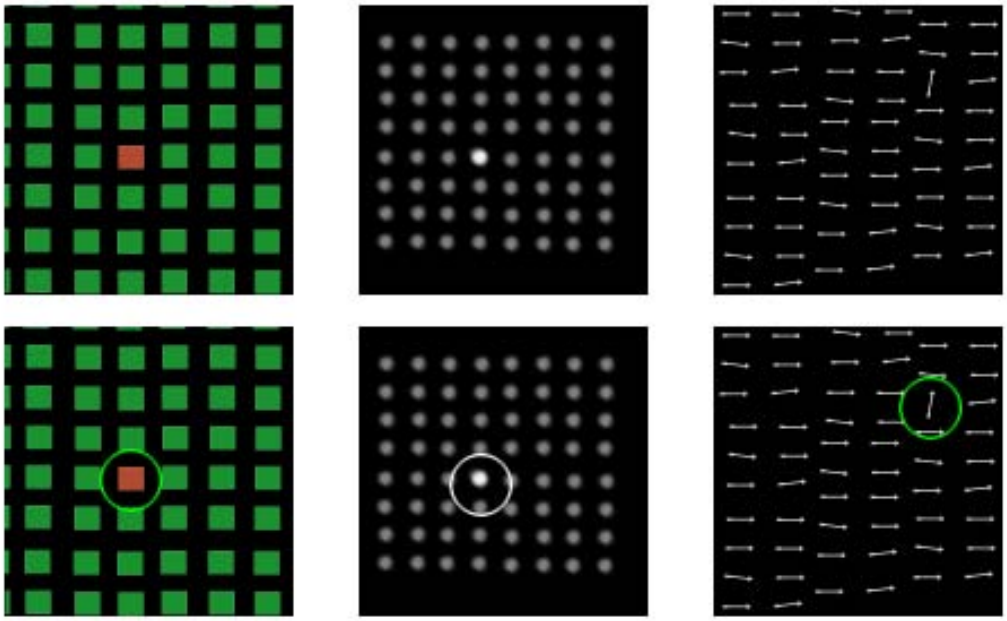

Fig. 8. Computer Generated images, representing each feature type, from left to right: color, intensity, orientation. The results of the algorithm are at bottom, showing that it correctly identified the areas that are most salient.

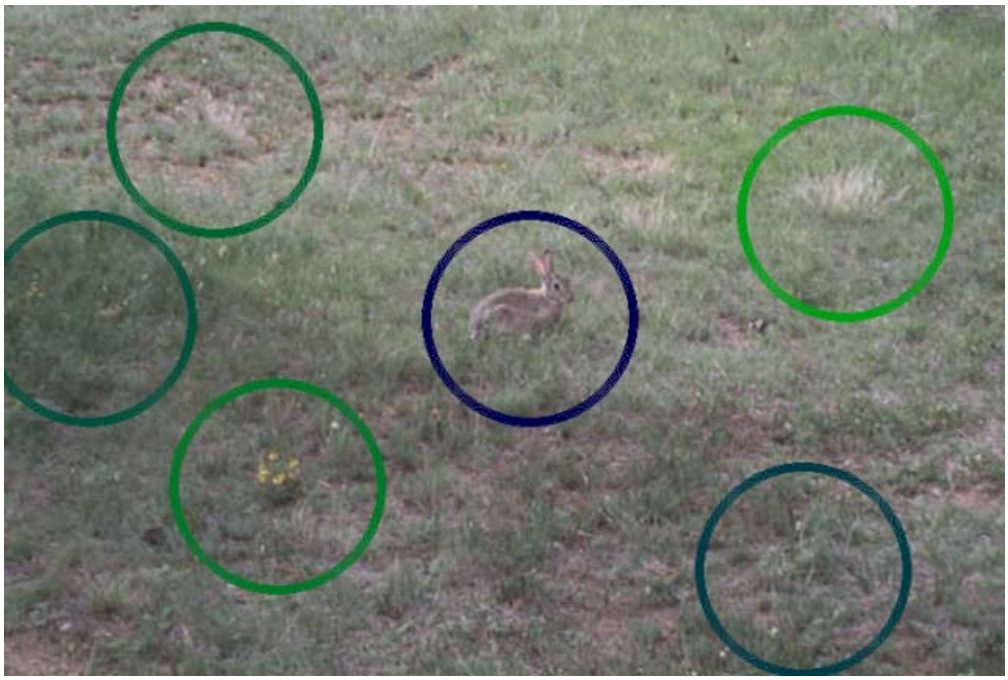

Fig. 9. The algorithm found the camouflaged rabbit, but it took more iterations, as might be expected.

Fig. 10. The original image, at top, was taken at a Civil War reenactment at Socorro, NM in 2004. The results (bottom), show that the program found the most conspicuous points to be the people and horses in plain view, with one exception.

## 4. Discussion

General structure of the algorithm is instructive. It is constructed of several sequential components:

1. Channel splitting and analysis.
2. Scale splitting and analysis.
4. Competitive selection.
5. Synthesis of scales.
6. Synthesis of channels.
7. Competitive selection.

We can identify aspects of automatic model construction and subsequent application in a boot-strap fashion. Model construction consists of partitioning/segmentation of information and extraction of the dominant features. Model is then constructed by recombination of the dominant features and exercised via competitive selection mechanism. Competitive selection is reminiscent of a global optimization procedure.

It is possible to add more channels to increase the capabilities of the algorithm. For example a Stereo channel would be instrumental in breaking camouflage (see Figure 11). A motion channel, for movies instead of static images, could also be included to provide for object tracking. It is well established that the human eye is incredibly sensitive to motion. Other static channels can be added as well, such as infrared.

We suggest that much of the technology applied here has strong parallels in the field of computational modeling of materials and systems, particularly, in the field of multi-scale modeling. We hope to use these parallels to develop original approaches and insights into these important areas.

Fig. 11. This image is of a pile of rocks, in a field of rocks. With two eyes (i. e. stereo vision) the pile is immediately apparent, however, with only one eye (or a camera lens) the pile disappears into the background.

References:

James, W. (1890/1981). *The principles of psychology.* Cambridge, MA: Havard University Press.

Milanese, Ruggero (1993). Detecting salient regions in an image: from biological evidence to computer implementation; Chapter 3: visual attention in machines. *University of Geneva*, 43

Burt, P., & Adelson, E. (1983). The laplacian pyramid as a compact image code. *IEEE Transactions on Communications, 31,* 532-540.

Koch, C. & Ullman, S. (1985). Shifts in selective visual attention: towards the underlying neural circuitry. *Human Neurobiology, 4,* 219-227.

Itti, L. & Koch, C. (2000). A saliency-based search mechanism for overt and covert shifts of visual attention. *Vision Research, 40,* 1489-1506.

Nakayama, K., & Mackeben, M. (1989). Sustained and transient components of focal visual attention. *Vision Resarch, 29,* 1631-1647

Distribution:

| | | |
|---|---|---|
| 5 | MS 1110 | A. Campisi, 1435 |
| 5 | MS1110 | A. Slepoy, 1435 |
| 1 | MS1110 | J. B. Aidun, 1435 |
| 1 | MS0310 | M. D. Rintoul, 1412 |
| 2 | MS0123 | LDRD Office, 1011 |
| 2 | MS9018 | Central Technical Files, 8945-1 |
| 2 | MS0899 | Technical Library, 4536 |