2007 Applied Imagery Pattern Recognition Workshop

Designing to See vs. Learning to See: Comparing Analytical and Adaptive Methods

Cosmos Club, Washington DC, October 10-12, 2007

Sponsored by:
The Institute of Electrical and Electronics Engineers (IEEE)
&
Science Applications International Corporation (SAIC)
## Program at a Glance

### Wednesday 10 October

<table>
<thead>
<tr>
<th>Start</th>
<th>Event</th>
</tr>
</thead>
<tbody>
<tr>
<td>0800</td>
<td>Continental breakfast</td>
</tr>
<tr>
<td>0830</td>
<td>Overview: Designing to see vs. learning to see: why should we care?</td>
</tr>
<tr>
<td>0910</td>
<td>Session 1: Analytic methods for computer vision</td>
</tr>
<tr>
<td>1150</td>
<td>Poster Teasers</td>
</tr>
<tr>
<td>1200</td>
<td>Lunch</td>
</tr>
<tr>
<td>1300</td>
<td>Keynote: J. Albus, NIST</td>
</tr>
<tr>
<td>1400</td>
<td>Session 2: Learning methods for computer vision</td>
</tr>
<tr>
<td>1640</td>
<td>Panel: Putting it all together – Day 1</td>
</tr>
<tr>
<td>1730</td>
<td>Poster Session and Reception</td>
</tr>
</tbody>
</table>

### Thursday 11 October

<table>
<thead>
<tr>
<th>Start</th>
<th>Event</th>
</tr>
</thead>
<tbody>
<tr>
<td>0800</td>
<td>Continental breakfast</td>
</tr>
<tr>
<td>0830</td>
<td>TBD</td>
</tr>
<tr>
<td>0930</td>
<td>Session 3: Performance assessment and comparisons</td>
</tr>
<tr>
<td>1150</td>
<td>Poster Teasers</td>
</tr>
<tr>
<td>1200</td>
<td>Lunch</td>
</tr>
<tr>
<td>1300</td>
<td>Keynote: M. Riesenhuber, Georgetown University</td>
</tr>
<tr>
<td>1400</td>
<td>Session 4a: Hot topic: Super-resolution and recognition in video</td>
</tr>
<tr>
<td>1530</td>
<td>Session 4b: Hot topic: Object and component recognition</td>
</tr>
<tr>
<td>1640</td>
<td>Panel: Putting it all together – Day 2</td>
</tr>
<tr>
<td>1730</td>
<td>Poster Session and Buffet</td>
</tr>
</tbody>
</table>

### Friday 12 October

<table>
<thead>
<tr>
<th>Start</th>
<th>Event</th>
</tr>
</thead>
<tbody>
<tr>
<td>0800</td>
<td>Continental breakfast</td>
</tr>
<tr>
<td>0830</td>
<td>Keynote: A. Youssef, George Washington University</td>
</tr>
<tr>
<td>0930</td>
<td>Session 5a: Hot topics: Data fusion / Biometrics</td>
</tr>
<tr>
<td>1110</td>
<td>Broader Impacts Address</td>
</tr>
<tr>
<td>1210</td>
<td>Lunch</td>
</tr>
<tr>
<td>1310</td>
<td>Futures Panel: Where do we go from here?</td>
</tr>
<tr>
<td>1410</td>
<td>Discussion of Findings</td>
</tr>
<tr>
<td>1430</td>
<td>Closing Remarks</td>
</tr>
</tbody>
</table>
Detailed Program

Wednesday 10 October

0800  Continental breakfast

0830  Workshop Kickoff
       (R. Bonneau, General Chair)

0840  Overview: Designing to see vs. learning to see: why should we care?
       (M. Loew & M. Happel, Program Co-chairs)

Session 1: Analytic methods for computer vision
       (Chair: J. Irvine, SAIC)

0910  3D organization of 2D urban imagery
       (P. Cho, MIT Lincoln Laboratory)*

The quantity, quality and availability of urban imagery are rapidly increasing over time. Thousands of photos shot by inexpensive digital cameras in cities can now be accessed via the web. But usually no connection exists between retrieved urban thumbnails other than their having been collected within some metropolitan area. In this talk, we demonstrate the utility of 3D geometry derived from aerial LADAR data for relating 2D ground photos. This organizing principle enables intuitive navigation through vast imagery archives, reduction of overlapping photos into mosaics, and geometry-based image correlation independent of illumination conditions and temporal variations.

Working with New York and Baghdad data sets as representative examples, we first fuse aerial LADAR images of entire cities with satellite pictures and Geographic Information System (GIS) layers to produce comprehensive urban maps. Digital photographs may be mathematically inserted into these detailed world spaces via tiepoint identification. Reconstructed view-frusta then yield camera locations and pointing directions which may have been a priori unknown. They also allow knowledge to be transferred from the maps onto registered image planes. For example, absolute geolocations can be assigned to individual pixels, annotation labels can be moved from 3D to 2D, and GIS networks can be projected onto arbitrarily oriented ground photos. Moreover, such information naturally propagates among all images whose view-frusta intercept the same urban map location. We illustrate how many data mining and visualization challenges (e.g. identify all photos containing some stationary ground target, observe some structure from multiple perspectives, quantify match between two images, etc) become mathematically tractable once a 3D framework for analyzing 2D images is adopted.

We close with several future applications of this work to defense problems of national importance. They include urban mission planning for theater operations, video activity monitoring for homeland defense, and augmented reality systems for real-time situational awareness.

* This work was sponsored by the Department of the Air Force under Air Force Contract FA8721-05-C-0002. Opinions, interpretations, conclusions and recommendations are those of the authors and are not necessarily endorsed by the United States Government.
Methodology for automated image-to-image registration
(B. Kovalerchuk, BKF Systems, P. Doucette, ITT Advanced Engineering &
Sciences, R. Brigantic, Battelle Pacific Northwest Division, G. Seedahmed,
University of Florida, B. Graff, US Army Topographic Engineering Center)

Registration and alignment of feature (e.g., vector) and raster geospatial data is a difficult and
time-consuming process. Data to be integrated may have inaccurate and contradictory geo-
references or not have them at all. Different and unknown rotations, disproportional scales,
uncontrolled noise, different physical modalities, and other factors are fundamental challenges
for robust fusion and integration of such data. This is an area of extensive research and
development, but commercial software products that address current challenges do not yet exist.

This paper presents an approach for vector-to-raster registration. Candidate features are auto-
extracted and vectorized from imagery, which are the basis to compare against existing vector
layer(s) to be registered. Given that automated feature extraction (AFE) methods are imperfect,
the objective is to determine and gather a sufficient signal-to-noise ratio from AFE upon which
to base a registration process between vector data. The vector registration process is based on
an algebraic structural algorithm (ASA) in which structural components (e.g., angles, lengths
and areas) are used as similarity metrics for comparison between different vector data sets.

In this paper, we demonstrate and compare results with registering road vectors to commercial
imagery in semi-automated and fully automated operational modes. Our objectives include
exploitation of panchromatic and multi-spectral imagery.

Coffee Break

Adapting to change: The CFAR problem in advanced hyperspectral
detection
(A. Schaum, Naval Research Laboratory)

Newer, realistic models of targets and backgrounds used in hyperspectral detection do not
always lend themselves to a CFAR (constant false alarm rate) formulation. Several advanced
techniques are considered here. It is found that incorporating a particular empirically validated
method of target evolution permits an exact CFAR version of a large class of advanced
detectors based on elliptically contoured distributions. Other validated detectors are
considered, for which no closed form normalization exists to convert them to CFAR form. For
these a geometrical approach to achieving approximate CFAR performance is de-scribed and
analyzed.

Image differencing approaches to medical image classification
(D. Tahmoush, University of Maryland)

Image differencing is normally done by subtracting the low-level features like intensity in
images that are already aligned. This paper extracts high-level features in order to learn an
effective image differencing method for the diagnosis of breast cancer. Learning techniques are
challenging when examples are sets of features that lack any sort of meaningful ordering and
where spatial relationships are important. We demonstrate a technique that avoids arbitrary
spatial constraints and is robust in the presence of noise, outliers, and imaging artifacts, while
outperforming even commercial products in the diagnosis of breast cancer images. First, the
landmarks are found and ranked, and then the top candidates are sorted into a point set.
Second, the point sets of the two images are differentiated by learning parameters for a novel
distance metric that bypasses direct registration.
A technique that radiologists use to diagnose breast cancer involves finding potentially cancerous sites in the mammograms and then comparing the left and right breasts to reduce the effect of false positives and to produce a diagnosis. The symmetry of the human body is utilized to increase the accuracy of the diagnosis. We emulate this technique in an attempt to capture the diagnosis of the radiologist.

The image comparison process determines the presence of cancer 85% of the time, outperforming all other systems, thus making it a strong classifier which should significantly improve CAD systems. The results compare favorably with the state of the art in both academic and commercial approaches, achieving a 9% overall improvement over the best academic approach and a 120% improvement in non-cancerous cases over the best commercial approach while maintaining the same accuracy on cancerous cases.

1140 Session summary and open issues
(J. Irvine, SAIC)

1150 Poster teasers

1200 Lunch

1300 Keynote Address: “Toward a theory of perceptual learning”
(James Albus, National Institute of Standards & Technology)

Human level perception (indeed mammal and bird perception) requires a well designed system in which learning can take place. It seems reasonable to assume that evolution provides the design and learning adapts the system to the environment. Furthermore, learning is of at least two varieties: One that takes place over the lifetime of the organism, and another that remembers single events during everyday life. Finally, any adequate theory must explain the richness of immediate visual experience.

Session 2: Learning methods for computer vision
(Chair: G. Becker, Unisys Corporation)

1400 Accelerating Learning and Reducing Error using Mentoring for Collective Learning Systems
(A. Armstrong & P. Bock, The George Washington University)

Mentoring is a new selection policy for statistical learning systems that has been tested with a Collective Learning Automaton (CLA) which solves simple, but representative problems that increases in complexity and change over time, simulating early learning and psychological development.

Currently, selection policies for CLAs usually select a random response to an immature stimulus, i.e., a stimulus that does not yet have a high-confidence response associated with it. Although this policy is completely unbiased and encourages exploration of the decision space, these random selection policies ignore any other stimulus-response (SR) pairs, assuming they are irrelevant and unrelated to the current stimulus. However, it may be possible that other SR pairs are, in fact, not unrelated to the current stimulus and, if these SR pairs are already mature and confident, this fact can be statistically exploited to assist selections for immature stimuli. The mentoring policy hypothesizes that selecting responses to immature stimuli by using responses to different, but mature and confident stimuli is statistically a better strategy than simply selecting random responses over the long term. It is important to note that mentoring
does not require or use any explicit or implicit feature-vector comparisons in search of an appropriate response. Preliminary results show that mentoring significantly accelerates learning and reduces error. Additionally, mentoring accelerates the recovery of effective performance when the problem is changed.

1430 A mathematical architecture for molecular computing

1500 Coffee Break

1530 In Situ Adaptive Feature Extraction for Underwater Target Classification
(T. Cobb & J. Stack, Naval Surface Warfare Center Panama City)

Underwater targets imaged by side-look sonar typically have distinct highlight and shadow pixel regions that are used as input features for pattern classification algorithms. Usually, image data used to develop feature sets and train classification algorithms covers a wide variety of sea bottom environments. This approach to training classifiers biases feature extraction thresholds and learned classifier margins to an amalgamated sea bottom environment representing by the training data set and can reduce performance to newly sensed data from alternate environments. This research develops an adaptive technique for classification of underwater targets via in situ threshold optimization and compares its performance to non-adaptive classification using fixed thresholds learned from training data.

In this paper, sonar image pixels are modeled as K-distributed random variables. Feature values for regions of interest are determined by counting segmented shadow, highlight, and background pixels within the region using a set of thresholds determined via K-distribution parameter estimation. For in situ adaptive classification, threshold levels are linked to the K-distribution shape parameter and texture correlation of the local background region. Non-adaptive thresholds are set globally based on the entire set of training data.

Results show that adapting the feature extraction parameters per test image results in significantly improvements in correct classification rate over globally optimized parameters. For this research, the a priori mappings are learned from a statistically large high-fidelity synthetic data set (>100,000 images) and the performance is validated on experimentally collected data of over 3000 detected regions containing over 150 targets. Results are compared in terms of speed and correct classification rates between the adaptive classification technique and the classical approach of optimizing global thresholds over the same training dataset.

1550 The Framework of Markov Random Fields - Perspectives of Modelling, Run-Time Adaptation and Real-Time VLSI-Implementations
(S. Stilkerich, EADS Deutschland GmbH)

Algorithmic robustness in real-world scenarios, real-time processing capabilities and physical compactness of the system are the three essential and at the same time contradictory requirements modern image-processing systems have to fulfill to go significantly beyond state-of-the-art systems. Without suitable image processing and analysis systems at hand, which comply with the before mentioned contradictory requirements, devices for application scenarios of the next generation in aeronautic, space and automotive will not become reality.
A coherent approach, which simultaneously realizes (1) modeling capabilities of a priori knowledge, (2) capabilities of the modeled image processing system to adapt to a changing environment and (3) massively parallel VLSI implementations, is presented in this contribution. Markov Random Field (MRF) theory provides the theoretical sound basis for our proposed approach.

In detail we will discuss the following three topics in this paper:

1) Modeling capabilities of image processing systems in a statistical framework with the help of MRF’s. Models on regular 2-dimensional MRF grids with local statistical interdependencies. Extension on irregular MRF structures (graphical models).

2) Capabilities of these model classes to adapt ("learning") to a priori unknown image model details and environmental settings. Adaptation of the image processing model and its free parameters during run-time.

3) Realization of massively parallel VLSI architectures for the described class of MRF models. Systematic derivation (based on the constraints of MRF theory) of a generic architecture template.

For the purpose of illustration we will discuss our proposed approach by means of a data-driven segmentation model and show the theoretical model, simulation results and a massively parallel architecture. The contribution is finalized by a discussion of application scenarios in aeronautic and space (main business of EADS) to underpin the industrial and practical relevance of the proposed direction.

### 1610 New hybrid classification scheme for mining multisource geospatial data
(R. Vatsavai, B. Bhaduri, Oak Ridge National Laboratory, S. Shekhar, T. Burk, University of Minnesota)

Supervised learning methods such as maximum likelihood classification (MLC) and maximum a posteriori (MAP) are often used in land cover (thematic) classification of remote sensing imagery. ML/MAP classifiers rely exclusively on spectral characteristics of thematic classes whose statistical distributions are often overlapping. The spectral response distributions of thematic classes are dependent on many external factors including elevation, soil types, and atmospheric conditions present at the time of data acquisition. A second problem with statistical classifiers is the requirement of large number of accurate training samples, which are often costly and time consuming to acquire over large geographic regions. With the increasing availability of geospatial databases, the first limitation can be addressed by exploiting the knowledge derived from these ancillary datasets to improve classification accuracies even when the class distributions are highly overlapping. Likewise the small training sample problem can be addressed with newer semi-supervised techniques which utilize freely available unlabeled training samples to iteratively improve the parameter estimates of the underlying statistical model. Unfortunately these two techniques can not be easily fused as these are no convenient multivariate statistical models that can describe multisource geospatial databases (both continuous and discrete random variables). In this paper we present a new hybrid semi-supervised learning algorithm that effectively exploits freely available unlabeled training samples from multispectral remote sensing images and as well as the ancillary geospatial databases. We have conducted several experiments on real datasets, and our new hybrid approach shows over 12-18% improvement in overall classification accuracy over the conventional classification schemes.

### 1630 Session summary and open issues
(G. Becker, Unisys Corporation)
1640 Panel: Putting it all together – Day 1
(Session 1 & 2 Chairs and Speakers)

1730 Reception & Poster Session

Poster Session 1
(Chair: N. Gupta, ARL)

W-1 Shape from texture
(F. Galasso and J. Lasenby, University of Cambridge)

Shape from texture has received much attention in the past few decades. We propose a computationally efficient method to extract the 3D shape of developable surfaces from the spectral variations of a visual texture. Under the assumption of homogeneity, the texture is represented by the novel method of identifying ridges of its Fourier transform. Local spatial frequencies are then computed using a minimal set of selected Gabor Filters. In both orthographic and perspective projection cases, new geometric equations are presented to compute the shape of developable surfaces from frequencies. The results are validated with semi-synthetic and real pictures.

W-2 Automated registration evaluation system (ARES)
(D. Lewis, S. Bergeron, M. Kim, & P. Doucette, National Geospatial-Intelligence Agency)

This paper describes the Automated Registration Evaluation System (ARES) developed for the National Geospatial-Intelligence Agency. The goal of the ARES is to dramatically reduce the cost of evaluating different automated data (raster and/or vector) registration technologies. Under the first year of the ARES project, five different automated registration methods were selected for evaluation. These methods represented the state-of-the-art in automated image-to-image registration and image-to-3D feature registration.

A common problem when attempting to compare multiple, different automated registration applications is one of interpreting the final results in a consistent manner across all the different applications. The applications use different output formats to store the results from automated registration. It is impractical to convert the results from different vendors into a common format or framework for evaluation. The ARES approach decomposes the automated registration process into individual components.

Using this approach, ARES can focus the assessment of different automated registration applications to an individual step within the overall registration process. Other components of the overall process, such as output format, can be made common between the different individual applications. The use of this system engineering decomposition process to automated registration allows ARES to evaluate the results from several different automated registration applications in a consistent and impartial manner.

During the first year of the ARES project, over two terabytes of unclassified NITF imagery and GIS test data were acquired for testing. These data sets were representative of the data registration challenges facing the intelligence community (IC) today. ARES performed over 4000 individual registration test cases between the five different automated registration methods. ARES was able to ascertain the operating conditions and performance of each of these methods against a wide range of IC registration requirements.
Using a high-fidelity simulation framework for performance singularity identification and testing
(C. Scrapper, R. Madhavan, & S. Balakirsky, National Institute of Standards and Technology)

At the National Institute of Standards and Technology, we have been developing test methods to classify the performance characteristics of a system using quantitative metrics that facilitate the inter-comparison of experimental results. De facto standard testbeds provide a baseline that target specific aspects of the system, allowing researchers to assess the performance of various systems in different scenarios and environmental conditions.

A common way to evaluate the performance of a system is to compare the algorithmic outputs with ground truth to discover irregularities and artifacts that exist. In turn, these inconsistencies are used to identify divergences in the system’s performance and discover the errors it is prone to. Performance analysis at the algorithmic level provides developers insight into performance singularities (which we define as the point where an algorithm fails to be well-behaved). Performance singularity identification and testing provides real-time meta-data that allows developers to understand the impact of singularities on the overall performance of a system.

As an example of the concepts developed in this paper, we will present a navigation solution based on image registration algorithms and the methodology used for the identification and testing of performance singularities of this algorithm. This research will use the baseline control framework, Mobility Open Architecture Simulation and Tools (MOAST), and a high-fidelity simulation testbed, Urban Search and Rescue Simulation (USARSim).
Thursday 11 October

0800  Continental breakfast
0830  TBD

Session 3: Performance assessment and comparisons
(Chair: E. Williams, Penn State University)

0930  Combining Rule-based and Machine Learning Approaches for Shape Recognition
(G. Becker, Unisys Corporation)

My paper and presentation will present the pros and cons of rule-based template shape recognition and adaptive supervised shape recognition in a variety of applications. Rule-based template matching systems have long been used to identify shapes in images. These template approaches are easy to implement, don’t require training, and can be very effective in some applications. Templates also have drawbacks in that they lack flexibility and an easy way to deal with variability across images. Supervised shape recognition uses training examples to learn the degree of flexibility and variability necessary for a particular application. Supervised systems, however, require training examples which may be difficult or expensive to produce. Shape recognition examples will be shown from applications using x-ray images and photographs. Some interesting differences between these two image types will also be discussed.

0950  The impact of band-to-band registration on hyperspectral target detection performance
(J. Casey, Rochester Institute of Technology)

Many hyperspectral sensors make use of multiple focal planes or multiple spectrometers to span a broad spectral range. These configurations can lead to spatial misregistration between sets of spectral bands. Some sensor designs result in registration errors of fractions of a pixel, while others may cause a misregistration up to several pixels. Perfect registration is never attainable, but techniques can be employed to better register the data, hopefully to within one pixel. The misregistration inherent in the design of such sensors will have an impact on the signature of targets in a scene, thus affecting the performance of target detection algorithms which utilize the entire spectral range. Synthetic imagery can be produced using a scene model to simulate such registration errors in controlled amounts. Simulations can be a great tool to compare the performance of hyperspectral target detection algorithms under varying levels of misregistration because simulated imagery produced using a scene model provides a great level of control. This paper focuses on the comparative performance of several hyperspectral target detection algorithms for multiple targets in a simulated scene which has varying amounts of misregistration between the visible and near infrared channels. Results indicate that statistical target detection algorithms are more robust than geometric algorithms when attempting to detect targets in misregistered images. As the amount of misregistration increases, geometric algorithms suffer more false alarms at boundaries where different background endmembers meet one another. This is caused by spectral mixing of background endmembers in only one set of bands. Statistical target detection algorithms may perform better at background characterization because of this phenomena.
1040 Lambda-connectedness determination for image segmentation
(L. Chen, University of the District of Columbia)

Image segmentation is to separate an image into homogeneous regions as distinct and belonging to different objects. It is an essential step in image analysis and computer vision. Even though there is no unified theory for image segmentation, some practical methods have been studied for years such as thresholding, edge-based segmentation, region growing, clustering (e.g. $k$-mean or fuzzy $c$-mean), and split-and-merge segmentation to name a few. Lambda-connected segmentation is a type of region growing segmentation technique. It was proposed to find an object with the property of gradual variation. In this paper, we will compare the lambda-connected segmentation with maximum entropy-based threshold segmentation, Otsu's threshold segmentation, and variational principle-based segmentation. With combining the optimization technique, we proposed an automated method to determine the connectedness for lambda value. The experimental results on various real images have shown the promise of the new method.

1100 Analytical versus adaptive image formation using optical phased arrays
(R. Kendrick & J. Marron, Lockheed Martin Advanced Technology Center)

Current optical phased arrays produce images by adaptively phasing the output of several telescopes on a common focal plane. Image-based phasing techniques such as Phase Diversity, are used to maintain the phasing in real time. This requires both a computationally intensive algorithm for estimating the phasing errors as well as a means for rapidly adjusting the optical path length through each telescope. In this paper we will compare this adaptive technique of phasing multiple telescopes with the analytical technique of digital holography. Digital holography provides a means of digitally estimating and correcting the phasing errors between the multiple telescopes. The process can occur long after the data has been acquired which relaxes the requirements on the stability of the phased array as well as the mechanical complexity. Experimental results will be shown for adaptive and analytical image formation in remote sensing applications.

1120 Pub/sub information management for imagery pattern recognition
(G. Ramseyer, Air Force Research Laboratory)

Imagery pattern recognition (IPR) problems are becoming more complex as image resolutions continue to increase along with algorithm complexity. Systems for analysis are being developed on increasingly diverse and heterogeneous platforms, from hardware support using FPGAs, GPUs, and Cell processors to high performance computers, featuring multi-core processors and shared memory communications. Given the complexity of modern IPR systems, neither analytical nor adaptive methods are sufficient to understand the shortfalls of an approach to IPR for any particular architecture.

We use a unique test environment to implement an adaptive, experimental method that can be used to quickly estimate performance and perform validation. Initial results are then used in analysis, to determine the codes that demonstrate the best system performance while producing the best, or at least acceptable, results. Analytic theories can be easily tested, by adapting execution requests according to adaptive feedback from experiments. In many cases, results will expose weaknesses in the system that should be corrected. To enhance our iterative adaptive refinement process, we generate multiple test cases that give analysts insights into possible approaches to improve system performance. With the availability of high performance computing, for parallel codes, parameters are more quickly discovered.
Our novel approach to processing publishes execution requests which are received by subscribing services. Processing can be performed by a spectrum of implementations and techniques that provide varying fidelity of results. Adaptive methods refine software based on input regarding which techniques work best. We are considering extending the IM brokering software to perform content-filtering on the payload, in addition to brokering based on metadata that describes the requested computation. This refinement will improve our ability to match images with codes that are best suited. For example, we can content route very large images to codes that use FPGA-based approaches for acceleration. As another example, we can filter large images using IPR algorithms to route them to algorithms that can perform appropriate correction.

We will discuss theoretical limitations for the design of imagery pattern recognition systems using a pub/sub information management system and present our experimental performance results. The pub/sub paradigm for communication presents opportunities for future parallel imagery pattern recognition implementations.

1140 Session summary and open issues
   (E. Williams, Penn State University)

1150 Poster teasers

1200 Lunch

1300 Keynote Address: “Object recognition in cortex: Computational mechanisms and translations to biologically-inspired machine vision and hybrid brain-machine object detection systems”
   (Max Riesenhuber, Georgetown University)

Object recognition is a difficult computational problem. Nevertheless, the human visual system can rapidly and effortlessly recognize objects in cluttered scenes under widely varying viewing conditions, at a level of performance far beyond that of current machine vision systems. We have recently made significant progress in understanding the neurocomputational mechanisms underlying rapid object recognition in cortex, and are now at a point where we can leverage these insights for applications. I will illustrate the promise of this approach with two examples: a biologically-inspired object detection system that delivers state-of-the-art performance on benchmark machine vision object detection tasks such as the Caltech 101, and a hybrid brain-machine system that classifies images based on the analysis of single-trial EEG signals recorded from humans viewing streams of rapidly presented images.

Session 4a: Hot topic: Super-resolution and recognition in video
   (Chair: Harvey Rhody, RIT)

1400 Regional variance dependent sub-frame reduction for face detection in video streams
   (V. Asari, Old Dominion University)

Statistical learning based face detection systems search multiple sized sub-frames of an image or frame of a video stream with a classifier trained to detect face object. Regions of the image or frame with a regional variance below that of a face do not need to be searched as it is not possible for a face to exist in this space. A preprocessing system to eliminate these low variance regions of a frame is presented in this paper.
The system utilizes a top-down, iterative search of a quad-tree decomposition of the frame to accomplish this task. A frame is divided into quarters and the regional variance of each of these sections is computed. Any section with a regional variance below that of a statistically computed threshold will be eliminated and no further processing will be required. If a section is above the threshold, that section will be further broken down and analyzed using the same procedure. This procedure continues until frame sections reach a predetermined size. All regions not eliminated are then merged and this output is fed to the trained classifier to detect any possible face present. The presented technique shows promise in reducing the overall search area of a frame, thus increasing the speed of the detection system.

1420 Super-resolution of video sequences using local subpixel motion correction of pixels from surrounding frames
(J. Colombe & B. Necioglu, The MITRE Corporation)

We present a method for super-resolving frames of video sequences based on supra-Nyquist information collected from surrounding frames. A ‘target’ frame is super-resolved by sparse up-sampling, addition of pixels from several surrounding frames based on subpixel alignments between local image patches from source and target frames, and imputation of missing pixels in the resulting montage target frame. Results are demonstrated on decimated low-res video clips, controlled to have frequencies up to 1.5x Nyquist, and the super-resolution results are compared to the original hi-res ‘ground truth’ clips and bicubic-interpolated versions of the low-res clips. The use of local subpixel alignment of information from surrounding frames makes it possible to super-resolve scene contents that move in non-affine ways, for example faces with changing expression, or 3D volume rotation of surfaces within a scene, as well as differential super-resolution of foreground and background regions that move independently. This approach also offers useful 3D structure-from-motion information. Applications in biometric face ID and stereo vision relevant to security-oriented domains are discussed.

1440 Moving vehicle registration and super-resolution
(F. Wheeler and A. Hoogs, GE Global Research)

The goal of image super-resolution is to estimate a high-resolution image from several low-resolution images of the same scene. In typical super-resolution applications, images are registered with a whole-image transform such as a homography, affine transform, or shifts. Such whole-image transforms can accurately register images of static scenes. However, with simple whole-image transforms, moving vehicles are not registered, and will in fact be distorted by super-resolution processing.

In this paper we describe a method specifically for registering and super-resolving moving vehicles from aerial surveillance video and show sample results. The challenge of vehicle super-resolution lies in the fact that vehicles may be very small and thus frame-to-frame registration does not offer enough constraints to yield registration with sub-pixel accuracy. To overcome this, we first register the large-scale image backgrounds and then, relative to the background registration, register the small-scale moving vehicle over all frames simultaneously using a vehicle motion model. To solve for the vehicle motion parameters we optimize a cost function that incorporates both vehicle appearance and background appearance consistency.

Once this process accurately registers a moving vehicle, it is super-resolved. We apply both a frequency domain and a spatial domain approach. The frequency domain approach can be used when the final registered vehicle motion is well approximated by shifts in the image plane. The robust regularized spatial domain approach handles all cases of vehicle motion.

In many image exploitation applications, moving vehicles are of specific interest, either for manual viewing or for automatic classification or tracking. Image super-resolution in almost all cases produces restored images that appear more crisp and that are more pleasing to view.
However, the real value of super-resolution in operation is when some new information about an imaged object can be determined. In the results section we are able to clearly demonstrate this by revealing features of moving vehicles that are not discernible in single images.

1500 Coffee Break

Session 4b: Hot topic: Object and component recognition
(Chair: H. Rhody, RIT)

1530 Multi-method recognition of pecan weevils
(S. Ashaghatra & P. Weckler, Oklahoma State University)

Pecan Weevil is one of the most destructive pests of pecans and are considered as the most serious “late-season” pest because they attack the nut. Nut losses due to pecan weevils can cause significant financial loss and can cause total crop failure. The goal of this study is to identify pecan weevil among other insects that are naturally present in the pecan habitat by implementing several image processing techniques. This is the first step toward building a wireless imaging system that would be commercially available for farmers. Insect recognition based on template matching was used in this study. Over 205 pecan weevil insects were used as a training set. The testing set consisted of 75 insects which normally exist in pecan habitat as well as 30 pecan weevils. Correlation-based template matching and geometrical image descriptors were the two recognition procedures used. Four types of geometrical descriptors, namely, Fourier descriptors, Zernike moment invariants, string matching and regional descriptors were used. The algorithm consisted of applying the above five methods. Moment invariants gave the lowest type I error; however, the type II error for this method is very large. NCC and regional descriptors methods gave the lowest type I and type II error. Further, string matching and Fourier descriptors produced the highest type I error. Results indicated that a positive match from three of the five independent tests would yield reliable results. Therefore, the algorithm included all five methods as no single method would achieve the desired success rate for identifying pecan weevils.

1550 Contouring as a useful feature for component recognition in digital transmissive images
(C. Oertel & P. Bock, The George Washington University)

The ALISA statistical-learning technology has been extended to include a new Component Module (ACM) designed specifically to recognize components of objects in digital transmissive images (DTIs), such as those generated by x-rays, CT, MRI, etc. Because traditional reflective surface features are not revealed in DTIs, the ALISA Texture, Geometry, and Shape Modules will not work, and a new module must be designed using a different basis for feature extraction. To accomplish this, the ALISA Component Module is trained with contour maps of several carefully selected regions-of-interest (ROIs) that contain a component-of-interest (COIs) in a sample of training DTIs. Once trained, the ACM may be shown the contour map generated from an entire test image to identify and locate instances of the trained ROIs, even in the presence of dense obscuration (shielding) or occlusion. This typical ALISA train-and-test paradigm will only work if the contour feature consistently and uniquely represents the different ROIs and their internal structure under a wide range of image acquisition conditions. Preliminary published results [Oertel2006] have shown that a simple contouring algorithm yields good and robust performance for a limited application domain (Beretta 9mm machine pistols). However, it is important to optimize and generalize the method to demonstrate its utility in a variety of different application domains. First, there are many different ways to accomplish contouring, and several of these have been employed to improve performance. In addition, using several
different contouring methods as concurrent dimensions of the feature space has been investigated. The classification performance of the investigation of the utility of these alternative contouring methods and strategies are reported in this paper for two research goals: 1) to successfully classify several ROIs in the set of 40 real x-ray images of Beretta 9mm machine pistols (as a feasibility and operating-point pilots); and 2) to successfully classify all meaningful ROIs in an artificially-generated, but reasonably-sized cover of the significant component geometries that could be found in real DTIs under realistic conditions. The conclusions of the second goal identify a specific contouring feature-space that consistently yields good and robust classification performance for many different kinds of components of objects in digital transmissive images in many different application domains.

1610  Biologically inspired system for object categorization in cluttered scenes
(T. Peerasathein, M. Woo, & R. Gaborski, Rochester Institute of Technology)

Humans can recognize objects in a cluttered scene in a few hundred milliseconds or less. Furthermore, we can easily recognize instances of a category in a cluttered scene, such as, all human faces in a scene. Because of the large in-class variability of features it has proven to be particularly difficult to develop algorithms to recognize every instance in a category, such as, all cat faces or dog faces instead of a particular cat or dog face. A similar case can be made for other categories, such as, vehicles.

One approach to solve the categorization problem is to model the processing in the human visual system. The human visual system can be divided into two major pathways, commonly called the ‘what’ and ‘where’ pathways. The ‘what’ pathway recognizes an object in a scene, but not its specific location. In this paper we present a biologically inspired hierarchical ‘what pathway’ neural network. The network can successfully classify objects into categories independent of their location and size in the scene. For example, the system can determine if an automobile is in a cluttered urban scene. Results will be presented for detection of objects of varying sizes and locations for several different categories.

1630  Session summary and open issues
(H. Rhody, RIT)

1640  Panel: Putting it all together – Day 2
(Session 3 & 4 Chairs and Speakers)

1730 Buffet & Poster Session

Poster Session 2
(Chair: TBD)

T-1  Robust method for multiple face tracking using Kalman filter
(Z. Shaik & V. Asari, Old Dominion University)

A robust method for tracking faces of multiple people using Kalman filter with process noise covariance update is proposed in this paper. This method overcomes the problem of partial occlusion (in case of rigid and non-rigid objects) and total occlusion (in case of non-rigid objects). The method uses the non-parametric distribution of face and cloth for distinguishing each person. Face color distribution is obtained from the first frame by applying Viola-Jones face detection method. Cloth color distribution is obtained from respective person’s clothes, assuming that the body moves along with the face. The size, center location and velocity of
motion of the detected face being the parameters of the Kalman vector, the search region for the face in consecutive frames is reduced because of prediction area generated by Kalman vector. Any abrupt change in the motion of the person forces the face detection algorithm to be applied to the whole image to determine probable faces, and similarity comparison of face and skin model distinguish each person effectively. Gaussian distribution based skin detection algorithm in the normalized color space will determine the face region (in the predicted area in case of no abrupt change in motion) and update the color distribution model and also correct the predicted values. This helps the method to adapt to lighting invariance and pose invariance. The size and location of the face boxes, and decrease in similarity of Bhattacharyya coefficients, after occlusion prediction, help in occlusion detection. The tracker stops tracking (but kept in history) the occulted object with similarity value less than threshold. The location and reappearance of the occulted face will be determined by Kalman predicted values around the occulted face or object. Phase detection after occlusion and the similarity measures combined with cloth model helps to distinguish each person. A new face is registered if the similarities of cloth and skin model are below threshold value. The partially occlusion problem is solved by using cloth model combined with the skin model distribution. The proposed method integrates many features to overcome partial and total occlusion and is robust in terms of lightning and pose variance. Color being the major feature of this method it is computationally simple and suitable for real time environments. The experiments demonstrate the robustness and simplicity of this algorithm in the case of indoor and outdoor environments. The adapting feature to lighting variance (even the use of popular color spaces has illumination effect) and pose variance, adaptive noise covariance Kalman filter, multiple features for occlusion detection and recovery, computational simplicity describe the robustness of the proposed method compared to other methods. When occulted by a rigid object rather than face, the cloth distribution model is used to track the face. The ongoing work is improving the performance of the proposed method in the case of non-frontal face in the video including other features for accurate face tracking after occlusion.

T-2 Turbulence spectra in cardiovascular risk evaluation
(Y. G. Tirat-Gefen & J. C. Evans, George Mason University)

This paper discusses how the turbulence spectra, generated due to the flow of blood inside the arteries and other circulatory vessels in the body, can be used as a measurement of the level of obstruction due to possible plague formation. We show a correlation between the expected low and high frequency bands of the turbulence spectrum in a vessel and its average diameter. The combined effect of many vessels is evaluated. These power spectrum measurements can be used as a baseline for evaluating the progression of plague formation over a lifetime, potentially with a lower cost than current techniques.

T-3 Modified WFPD with variable window size
(L. Toledo & F. Cuevas, Centro de Investigaciones en Óptica)

We present an algorithm for the retrieval of the phase of a complicated single fringe pattern, based on the WFPD method. The pattern is divided into a set of partially overlapping sub-image, variable size windows. In these sub-images the estimated phase is modelled as a parametric analytic-function, and its parameters are optimized using a genetic algorithm (GA). We suppose that the estimated phase in the window is equal to the real phase in this region, and the estimated phase is then connected to the phase in an adjacent window by adding a DC bias to the analytic function. The overlapped phase similarity criterion is avoided in the present method, and it is substituted by a second order smoothness criterion. The demodulation errors from others windows did not affect the demodulation of a given window.
Active tracking of surface targets
(A. Waxman, D. Fay, P. Ilardi, P. Arambel, & J. Silver, BAE Systems Advanced Information Technologies)

We seek to establish a quality metric for sensor fused video based on the performance of a multisensor system that actively tracks multiple surface targets over an extended field of regard. Such a system must, in real-time, fuse multisensor/spectral imagery, detect targets reliably, track the targets for extended time periods as they move, stop, approach, cross, hide/emerge, over an extended field of regard.

The system must automatically control the pointing, zoom and modes of cameras so as to maintain and re-establish track as targets move outside the current field of view. All this must be done for an increasing number and density of targets in clutter. Individual performance metrics relating to these various stages of processing include percent detection vs. false alarm rate, track identity lifetime, track purity, track accuracy, field of regard, target revisit rate, and number of targets tracked. All of these metrics reflect the quality of the multisensor imagery and the algorithms (image fusion, target learning, detection, tracking, sensor resource management) for the multi-target tracking task. This same task applies to human observers tracking multiple surface targets in displays of multisensor/spectral video. One intuitively expects that the "better" the image quality, the more "effectively" one can track multiple targets. We are exploring this relationship between tracking and fused video as a quality metric. This paper introduces the computational fused video tracking system that forms the basis of this new effort.
Geometric structure of a scene can be reconstructed using many methods. In recent years, two prominent approaches have been digital photogrammetric analysis using passive stereo imagery and feature extraction from LIDAR point clouds. In the first method, the traditional technique relies on finding common points in two or more 2D images that were acquired at different spatial locations. More recently, similar approaches have been proposed where stereo mosaics are built from aerial video using parallel ray interpolation, and surfaces are subsequently extracted from these mosaics using stereo geometry. Although the LIDAR data inherently contain 2.5 or 3 dimensional information, they also require processing to extract surfaces. In general, structure from stereo approaches work well when the scene surfaces are flat and have strong edges in the video frames. LIDAR processing works well when the data is densely sampled. In this paper, we analyze and discuss the pros and cons of the two approaches and methods for combining them in planar surface model fitting. We also illustrate the benefits of this data fusion through three scenarios: when one or more edges are not clearly visible in the video frames, when the LIDAR data sampling density is low, and when the object surface is not planar. Examples are provided from the processing of real airborne data gathered using a combination of LIDAR and passive imagery taken from separate aircraft platforms at different times.

MODIS and AWIFS multi-sensor imagery data fusion for crop classification
(Z. Yang, U. S. Department of Agriculture)

The mission of the National Agricultural Statistics Service (NASS), Department of Agriculture is to timely and accurately provide agricultural statistics to the public. To reduce the time for delivery of the satellite-based crop coverage classification and acreage estimation results to USDA Agricultural Statistics Board from December to October, it is critical to start the process of producing these results as early as possible. NASS currently uses the India’s ResourceSat I, 56m AWIFS (Advanced Wide Field Sensor) sensors imagery for our operational crop classification and acreage estimation analysis. The major obstacle to achieving the goal of early delivery is the limitation of the availability of the AWIFS data in August, which is a critical period for early crop remote sensing and acreage estimation. The AWIFS sensors have 5 day revisit. However, if excluding the cloudy imagery, sometimes there is hardly enough data for a reliable multitemporal and multispectral crop classification and estimation. To solve this data availability issue, the additional data source is needed to compensate the AWIFS limitation...
so that we can have more quality data available in time for crop classification and hence to meet the early report deadline. In this paper, we propose to fuse NASA MODIS data, which has daily revisit and lower resolution, with the higher resolution and longer revisit time AWIFS data. The fusion of the extra MODIS data with AWIFS data is not for better visual interpretation, but for more useful information to be used for better classification. First, the fusion of spectral signature feature vectors from both MODIS and AWIFS images is performed using the decision tree based classification method in this paper. Secondly, the pixel based Principal Component Analysis method is used to fuse the MODIS and AWIFS images and then the fused images are fed into the decision tree classifier to perform a crop classification. In this paper, the third study conducted is that the AWIFS data and the fused images are fed into the decision tree classifier, which fuses the spectral signature vectors of multitemporal high resolution AWIFS images with that of the PCA pixel fused images. Finally, the classification results are compared in term of accuracy improvement. The initial results indicate that the MODIS data fused with AWIFS data improves the accuracy of the crop classification in the large crop area and it is concluded that further research needs to be performed.

1010 Coffee Break

1040 Fuzzy edge detection in biometric systems
(S. Vasikarla, American InterContinental University, Los Angeles)

We propose a fuzzy logic based edge detector for feature extraction in biometric systems such as fingerprint and palm print recognition. Edge detection is carried out by means of global (histogram of gray levels) and local (pixels within a window) information. The local information is fuzzified by employing a modified Gaussian membership function. Using the contrast intensification operator, the image is enhanced to the required level of visual quality by entropy optimization of fuzzification parameters. Furthermore, the local edge detection operator is applied on the enhanced image using parameters obtained from entropy optimization. Finally, simple edge thresholding is applied to produce the skeleton image. Results demonstrate that this edge detector is immensely suitable for applications such as palm print and fingerprint identification, as it does not distort the shape and is able to retain the important edges unlike other edge detectors.

1100 Session summary and open issues
(James Aanstoos, Mississippi State University)

1110 Broader Impacts Address: “Confronting Kant’s Conundrum: The Grand Challenge for Computer Vision”
(Peter Bock, George Washington University)

As the Western World finally lifted itself out of the abyss of ignorance of the Middle Ages, philosophers debated how knowledge could and should be acquired, raising issues that been forgotten or lost for a thousand years. In the early 17th century, analytical reasoning was already being challenged by statistical empiricism. By the middle of the 18th century, the intellectual war between the two paradigms was in full swing, fueled by the powerful essays of Immanuel Kant. The lines were clearly drawn, and eminent philosophers and mathematicians were taking sides, each side firmly believing that it would eventually prevail.

This war raged for more than 200 years, but by the middle of the 20th century, most experienced scientists and philosophers had concluded that there could be no victor in this war: neither paradigm can be applied without invoking the other, and both are equally endowed with inherent uncertainty.
Today, in the field of computer vision, our new and powerful tools for empirical neuroscience research have clearly validated this conclusion, using the primate brain as a model of highly sophisticated image and signal processing. We now know that the human brain acquires its visual data using genetically specified features, after which the surviving information adapts our personal knowledge bases, allowing us to learn to draw meaningful conclusions about what we see. We all experience both the wonder and the uncertainty of this hybrid mechanism every single day.

The grand challenge, then, is to apply both adaptive learning and analytical reasoning to build a computer vision system that is on a par with (or even better than!) our own. There are, of course, a few critical issues to keep in mind.

1210  Lunch

1310  Futures Panel: Where do we go from here?

Panelists: Raj Madhavan, National Institute of Standards and Technology
John Tangney, Air Force Office of Scientific Research
Jeffrey Kretsch, National Geospatial-Intelligence Agency
John Irvine, SAIC

1410  Workshop preliminary report: Discussion of findings
(M. Loew & M. Happel, Program Co-chairs)

1430  Closing remarks & preview of AIPR-2008
(R. Bonneau, General Chair)