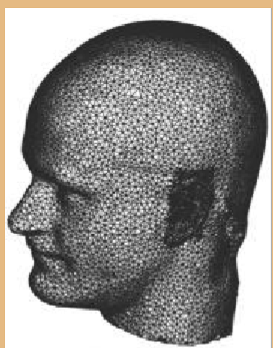
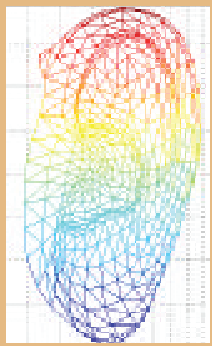
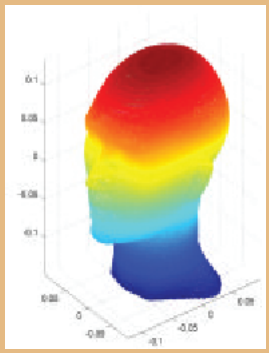
CENTER FOR AUTOMATION RESEARCH

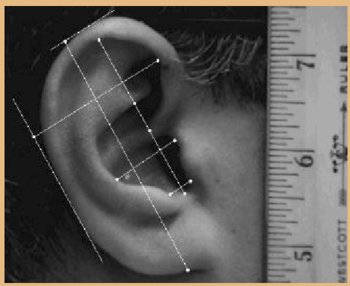
INSTITUTE FOR ADVANCED COMPUTER STUDIES

UNIVERSITY OF MARYLAND









Status Report

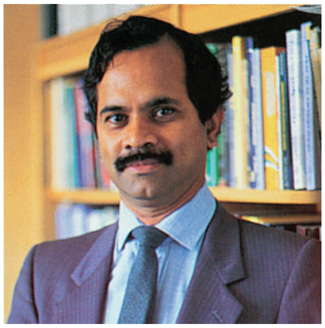
July 2001-Dec 2002

Contents

We are pleased to announce that Dr. David Jacobs, Associate Professor of Computer Science, has joined the Computer Vision Laboratory. We are pleased to announce the establishment of the Perceptual Interfaces and Reality Laboratory, directed by Dr. Ramani Duraiswami.

Cover photo: Meshes of the head and pinna used in the numerical computation of the Head Related Transfer Function. Results of a decompositional approach to computing the HRTF by separating the head and pinna. Anthropometry of the pinna used for correlating observed pinna features.

The Center for Automation Research



Rama Chellappa
Director

Center Staff:

Facilities Manager
Andrew Pilipchuk

Support Staff
Sailaja Akunuri
Janice Perrone
Millie Smith

Azriel Rosenfeld
Founding Director and Professor Emeritus

The Center for Automation Research is part of the Institute for Advanced Computer Studies in the College of Computer, Mathematical, and Physical Sciences at the University of Maryland in College Park. The Center's laboratories conduct research in the areas of computer vision, language and media processing, computer graphics and visual informatics, and perceptual interfaces and reality. Faculty associated with these laboratories have appointments in the Departments of Computer Science, Electrical Engineering, and Linguistics. The Center also maintains close relationships with other research groups at the University, including the Laboratory for Computer-Aided Control Systems Engineering in the Institute for Systems Research and the Laboratory for Basic Research in Sensory Systems in the Department of Psychology.

The Computer Vision Laboratory

Yiannis Aloimonos

Faculty
John S. Baras
Rama Chellappa
Larry S. Davis
David Jacobs
Azriel Rosenfeld
Hanan Samet

Visiting Faculty
Leila DeFloriani
Zoran Duric
Nathan Netanyahu
A.N. Rajagopalan

Research Faculty
Cornelia Fermüller
Volker Krüger
Reuven Meth
Gang Qian
Amit RoyChowdhury
Chandra Shekhar
Egemen Tanin
Isaac Weiss
Yaser Yacoob
Liang Zhao
Qinfen Zheng

Support Staff
Sara Larson

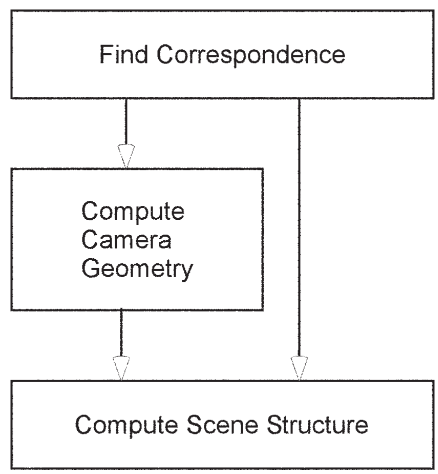
The Laboratory
The Computer Vision Laboratory is one of the oldest and largest research groups of its kind in the world. It was established in 1964 by Prof. Azriel Rosenfeld; its current head is Prof. Yiannis Aloimonos. The Laboratory has a staff of over 80, including faculty, research scientists, visiting faculty and researchers, graduate students, and support personnel. It conducts research on time-varying image analysis, three-dimensional scene analysis, purposive vision systems, biometrics, and geometrical data structures for vision and graphics. Recent research by the senior faculty associated with the Laboratory is summarized in the following subsections.

Facilities
The Laboratory's computing facilities consist of many SUN SPARCstations, Macintosh computers, and PCs with video input and output capabilities. The Laboratory also has facilities for video conferencing, video digitizing, recording, editing, and playback, as well as grayscale and color page scanning. Available software includes a wide variety of basic image processing and analysis programs. Facilities for image input include many solid-state and high-speed TV cameras and an infrared camera.

Yiannis Aloimonos
Our group continues to work on the basic principles governing visual space-time. We have concentrated on the problem of recovering 3D models of a (possibly changing) scene from multiple views of it, a central problem in computer vision. We highlight here two efforts that took place during the last year; one has to do with the correspondence problem and the other with eye design.

New Mathematics for Visual Correspondence
There exists a highly developed theory, under a variety of names such as multi-view geometry, structure from motion, etc., that given point and line correspondences in multiple views, solves for the view geometry (where are the cameras?) and subsequently for a 3D model of the scene. This state-of-the-art theory appears to be reaching its limitations. Results have many errors: too many distortions and holes in the 3D models, and the need for manual intervention due to the inability to deal with the correspondence problem.

The basic framework of structure from motion has, for the most part, been pursued as in the following diagram:



Finding correspondence means finding image features that were generated by the same world feature. These features can be points, lines, corners, or some other feature. Most researchers will say that correspondence is a "difficult step," but the truth is that no one has been able to solve this problem in three decades of computer vision research. Tracking has had some success in limited domains where there are obvious corners to track, but more natural scenes have proven impossible for current methods.

The last step of computing scene structure given camera geometry and correspondence lies more in the realm of graphics. That is, we want to know how to create the best possible model from the correspondence and geometry which we have calculated. Most of the computer vision community would say that the problem of scene reconstruction has been solved, and indeed it has been for points and lines, but not when the input is images. For example, the reconstruction problem for patches of various textures has not been addressed in detail at all. The depth of a patch with a homogeneous texture cannot be found, but if one is interested only in projecting to a nearby viewpoint, this doesn't matter. Texture is an integral part of reconstruction.

The middle step of camera geometry computation has been the subject of much research for the past two decades. Constraints have been discovered that relate point and line correspondences in multiple views to the camera geometry and the 3D scene, giving rise to a multi-view geometry theory.

The field has reached the stage where we believe we have solved the final two steps of the process, but we still do not have a good insight into how to generate the input to these stages. Despite the fact that we have a well-developed theory about points and lines, there still does not exist an automated system which calculates structure from pictures or video streams. To get past this we need to solve the correspondence problem. There is an additional problem: From point and line correspondences a 3D mesh is created and then the image texture is mapped onto it. This is too artificial. Somehow the texture does not participate in 3D model creation, except for finding points and lines. One of the major reasons for limited success has to do with the fact that points and lines cannot be accurately located in images. We have shown that, in general, there is a statistical bias in the location of points (line intersections), lines and image movement. Images, like any signal, are noisy, and the visual system has to perform its interpretation in the presence of this noise. Because of the noise, errors occur in the computation of features, their positions, orientations, sizes and movements. The computational processes are such that the errors in the estimation are systematic; in statistical terms, we say the estimation is biased. To avoid the bias would require accurate estimation of the noise parameters, but this, because of the large number of unknown parameters (the geometry and photometry of the changing scene), is in general not possible. This bias is responsible for a large number of illusions. We have investigated the bias and its relationship to the image texture. See, for example, Figure 1. By fixating at

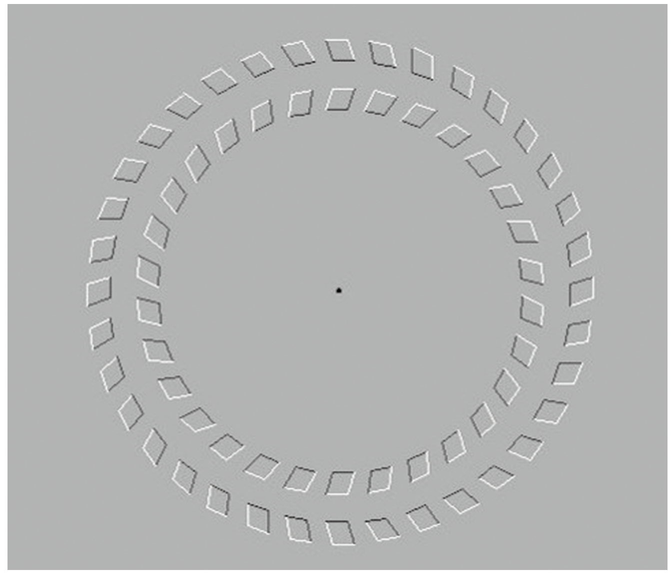


Figure 1. An illusion by Pinna.

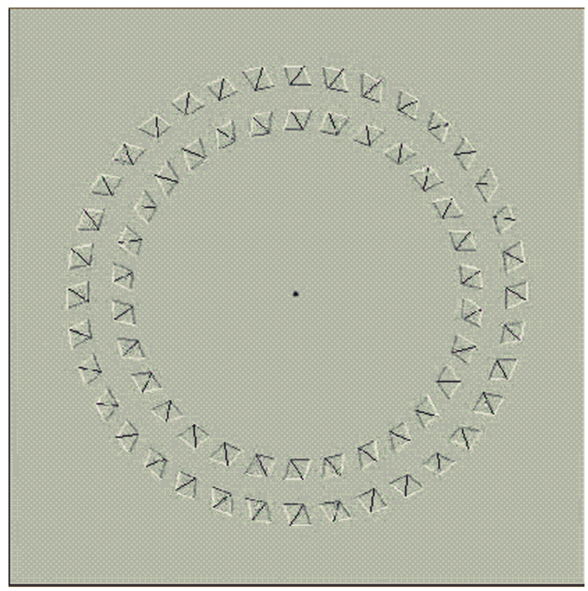


Figure 2. The illusion disappears.

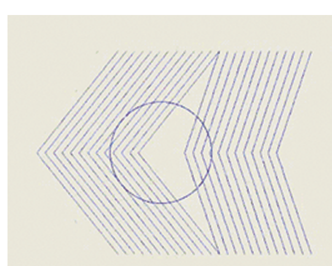


Figure 3. The circle distorts.

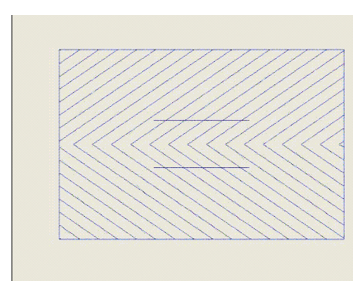


Figure 4. The lines are horizontal and parallel but do not appear so.

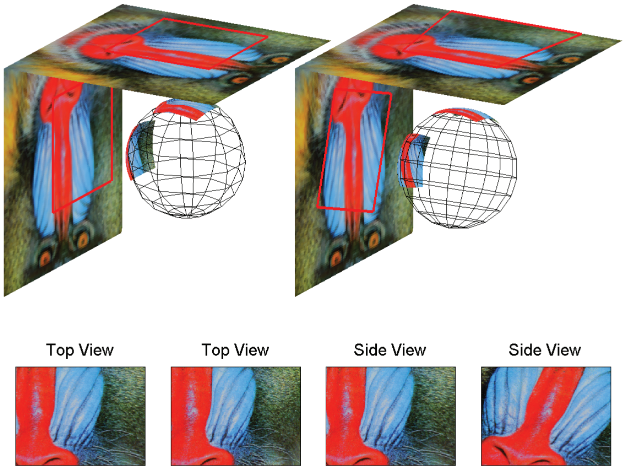
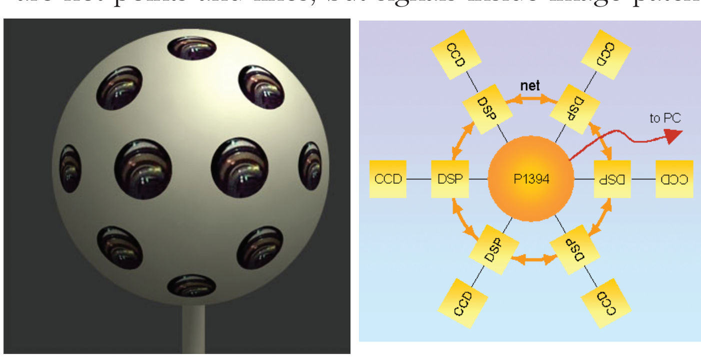


Figure 5. Views of a monkey.

the center and moving the figure back and forth along the optical axis, we experience an illusory rotational movement. The bias theory predicts this. By manipulating the image texture, one can manipulate the bias. For example, in Figure 2 the illusion disappears. Similarly, by changing the background in Figures 3 and 4 (as seen in the videos http://www.cfar.umd.edu/users/yiannis/dialogue100/video01.mpg and video02.mpg) we alter the perception of the lines or the circle.

This bias translates to three dimensions; this can be easily derived. Thus, both the camera geometry and the 3D model are distorted, resulting in inaccuracies. To achieve high accuracy, we need a new theory that is not based on points and lines, and can effectively deal both with texture and with the correspondence problem. We have developed the beginnings of such a theory. We have named it Harmonic Computational Geometry to emphasize the marriage of harmonic analysis and geometry. The theory introduces new atoms for the problem; these are not points and lines, but signals inside image patches, specifically their harmonic components. New constraints, the harmonic epipolar, the harmonic trifocal, and others, have been discovered that have a multilinear structure. The theory provides a tool for dealing with the correspondence problem (solves it or decides that it cannot be solved), and in conjunction with the existing points and lines theory, can lead to unprecedented accuracy in 3D reconstruction. The input to the process is the output of filters applied to image patches.



(a) (b)

Figure 6. (a) A golf ball size eye consisting of many cameras looking outward. (b) A system diagram of an Argus eye.

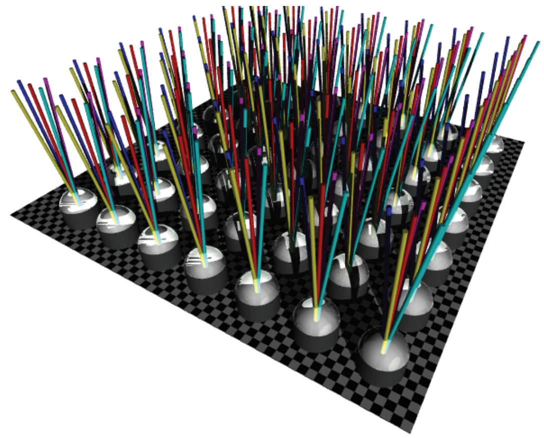


Figure 7. Many miniature cameras on a surface (here a plane) constitute a polydioptric eye.

New Eyes

For a long time the cameras we have been using, although they become more sophisticated technologically as time passes, are basically based on the same principle: the pinhole. Geometrically, one considers the light rays passing through a point and cuts (some of) these rays with a plane (the film). Thus, images are formed. Perhaps one of the reasons we stick to this model is that it is also a model for our eyes, human eyes, which we value and celebrate. But if we examine the animal world we will discover a large variety of eyes: the set of rays that is cut with a surface, and the nature of the surface and the distribution of photoreceptors on it, along with their optical properties, define different eyes. We have been studying general principles of eye design. What makes a good eye, or a good camera? It turns out that this is a question that can only be answered in a purposive manner. What is the eye going to be used for? That is, what kind of processing will be performed on the images collected by the eye? One can easily get lost in the complexities of this question. For this reason, we have concentrated on the problem of making 3D models of the world from multiple views, a significant part of the vision problem itself. We have discovered two principles. One has to do with the field of view and the other with the dioptric number.

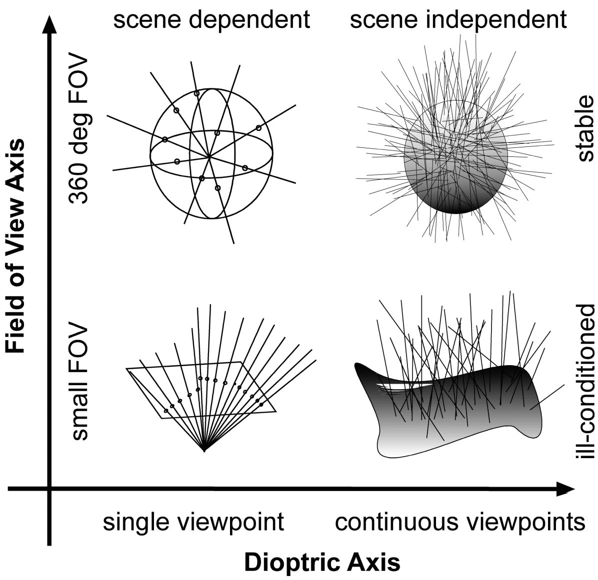


Figure 8. Eye diagrams.

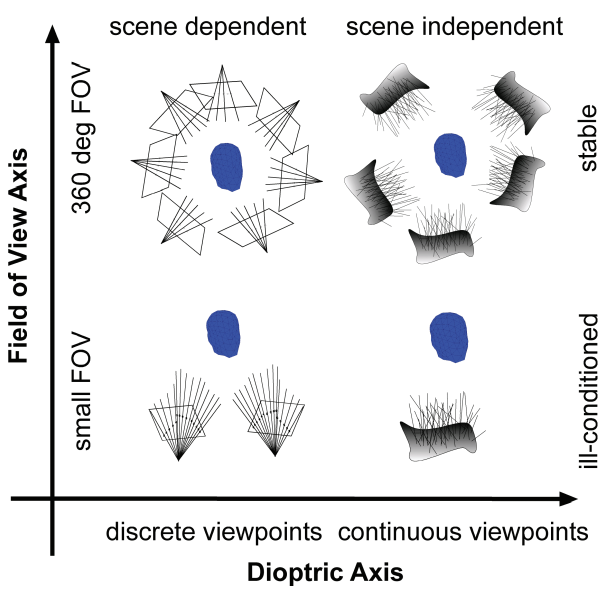


Figure 9. Eye diagrams.

For a conventional pinhole camera with a limited field of view it is hard to estimate 3D models of a scene from multiple views of it (e.g., video). The reason is that solving for camera geometry (where are the views?) is problematic because of the confusion between rotation and translation. This is shown in video03.avi, a frame of which is seen in Figure 5. Imagine that you are at the center of the sphere, looking at the ceiling. Both the ceiling and the wall are painted with a picture of a monkey. Whether you translate parallel to the ceiling or rotate around an axis parallel to it, the acquired videos are basically indistinguishable. But if at the same time you also look at the wall, the two motions can be easily distinguished. Thus, the problem of making 3D models is ill-posed for restricted fields of view but becomes well-posed for a full field of view (i.e., cutting the rays which converge to a point with a sphere). The problem, however, remains non-linear.

The dioptric number D amounts to the number of viewpoints (or pinholes) that the eye possesses. A single camera has D = 1, a stereo system has D = 2, a network of 100 cameras arranged on a surface has D = 100. As the dioptric number increases, the eye becomes capable of sampling the plenoptic function, in which case the problem of making 3D models becomes linear. These principles led us to start developing two families of new cameras, Argus eyes and polydioptric eyes (Figures 6 and 7). Figures 8 and 9 show "eye diagrams" which demonstrate the two principles discussed above, for both outward and inward looking systems of cameras.

Using this new framework enables us to extract more accurate models of a dynamic world. Figure 10 shows different novel views of a talking head reconstructed from multiple video sequences using dynamic multiresolution subdivision surfaces. For more details, see www.videogeometry.com.

Figure 10. Novel views of a talking head reconstructed from multiple video sequences.

Rama Chellappa

Prof. Chellappa's research deals with many asepcts of computer vision and image processing. Current interests include human identification at a distance using gait, body appearance, and facial information; human activity inference; three-dimensional model building from video, shape-encoded tracking; hyperspectral data analysis, and microsensor image understanding.

Distributed Tracking Using Multiple Cameras

To persistently track people and vehicles in a large area of interest, distributed imaging sensors must be used, since the field of view (FOV) of a single camera is limited. By using distributed cameras, when a target is leaving the FOV of one camera but still remains in the monitored area, other cameras should be able to pick up the target and continue the tracking task. There are two major challenges here. One is to automatically calibrate a camera network with multiple arbitrarily placed cameras, and the other is to find correct target associations between video streams from different cameras when multiple targets are present at the same time. Sequential Monte Carlo (SMC) algorithms have been used to solve the camera network self-calibration problem. The posterior distributions of the relative positions and orientations of the cameras are represented by related sample and weight sets. We have also developed an algorithm to solve the problem of target association or target cross-sensor hand-over. An example of distributed tracking via cross-sensor hand-over is shown below. In this example, two people were tracked using camera 1 and one person using camera 2. The tracking results are shown in Figures 11 and 12, respectively, with the tracked targets marked by bounding boxes.

Given a video clip, the intrinsic parameters of the camera recording the video may not be known, especially the focal length, which may even change during image capture. Moreover, there are special groups of camera motion sequences known as critical motion sequences (CMS) that lead to non-unique solutions for the camera intrinsic and extrinsic parameters. Successfully dealing with CMS is of great importance in practical applications since many camera motion trajectories are "close" to being critical in the L2 sense. Using SMC, we have developed a Bayesian self-calibration method for a moving camera with unknown focal length. No initial guess at the focal length is needed. The algorithm can also detect the presence of CMS by finding the posterior probability of the criticality of the camera motion sequence, and can properly process critical or quasi-critical motion sequences.

Multiple independently moving objects are often present in real-life videos. Assume that feature point trajectories tracked through input videos are used as measurements. The goal is to solve for the correct feature segmentation on each moving object, the relative motion between each moving object and the camera, and the 3D structure of each moving object. Feature segmentation is represented by a so-called validity vector, which has the same number of components as the number of feature points. Through SMC, samples of the validity vector related to a moving object evolve to a structure such that the entries corresponding to the features on the object will have large positive values while the remaining entries have small, negative values. Feature segmentation and object motion are then achieved together because the true 3D motion of an object can explain the 2D trajectories of the features belonging to this object. The structure is easy to get after the above two problems are solved.

Figure 11. Two tracked persons using camera 1.

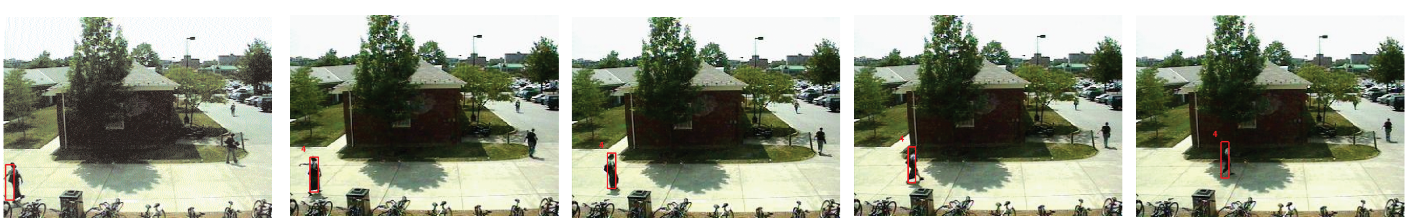


Figure 12. One tracked person using camera 2.

3D Modeling from Unconstrained Video

We have posed the 3D reconstruction problem in an estimation-theoretic framework using the optical flow paradigm for modeling the motion between the frames of the video sequence. We show how the statistics of the errors in the input motion estimates are propagated through the 3D reconstruction algorithm and affect the quality of the output. The derivation does not make standard statistical assumptions of Gaussianity and independence, We have shown that the 3D estimate is always statistically biased, and the magnitude of this bias is significant. We have analyzed the effect of different camera motion parameters on the bias, and concluded that it is most sensitive to the rotational motion. Existing results on the Cramer-Rao (CR) lower bound have been extended to derive a generalized CR bound, which incorporates the bias. In order to demonstrate our analysis in a practical application, we consider the problem of reconstructing a 3D model of a human face from a video. An algorithm has been developed that obtains a robust 3D model by fusing two-frame estimates using stochastic approximation theory and then combines it with a generic face model in a Markov chain Monte Carlo optimization procedure. Figure 13 shows two frames from the original video sequence and Figure 14 shows four novel views generated from the reconstructed 3D face model. We have addressed the question of how to automatically evaluate the quality of a 3D reconstruction from a video sequence, and developed a criterion using concepts from information theory. Finally, we have developed a probabilistic registration algorithm that extends the results of our work to create holistic 3D models from multiple video streams.

Figure 13. Two frames of the original video sequence which is the input to the SfM reconstruction algorithm.

Figure 14. Different views of the 3D model after texture mapping.

Human Identification at a Distance

Our early work on gait recognition involved computing the width of the outer contour of the binarized silhouette of a walking person as the image feature. In this method, a set of exemplars that occur during a walk cycle was chosen for each individual. Using these exemplars a lower-dimensional Frame to Exemplar Distance (FED) vector was generated. A continuous HMM was trained using several such FED vector sequences. This approach serves to compactly capture structural and dynamic features that are unique to an individual. The statistical nature of the HMM provides overall robustness to representation and recognition. Experiments were conducted on scenarios encompassing variations across time, clothing, and illumination change, and the performance of the method was found to be good.

One of the databases had the subjects walk with a ball in their hands. It was found that the method gave better results in this scenario as compared to normal walk cases. This suggests that for the purpose of recognition, certain parts of the body may be more favorable than others. In particular, the leg motion provides more discriminating evidence as compared to the evidence provided by hand and leg motion together. To test this hypothesis, we developed a fusion algorithm. This involves dividing the silhouette image into three regions, head, torso and legs. Information from leg swings alone was extracted for training the hidden Markov model. A score transformation is applied to the height of the person and the probability is fused with the evidence produced by the HMM to give the overall probability. It was found that this method. improved the recognition results.

In another study we examined the spatio-temporal pattern generated by the width feature for different walkers. A comparison of the temporal width plots across individuals reveals interesting structural and dynamic information. The differences in gait can be captured by matching the sequences of width vectors corresponding to different individuals. A pattern matching method based on a dynamic programming paradigm is most suitable in this case. This method, based on the Dynamic Time Warping (DTW) algorithm popular in speech recognition, is used to compensate for variations in walking speed, reflected in the number of frames for each gait cycle. Experiments with many data sets show that this method has the potential for serving as a filter.

The eigengait and HMM methods for gait-based human recognition assume that the humans walk in a fronto-parallel manner. In real applications, such restrictive walking patterns are not realistic. Recently, we have

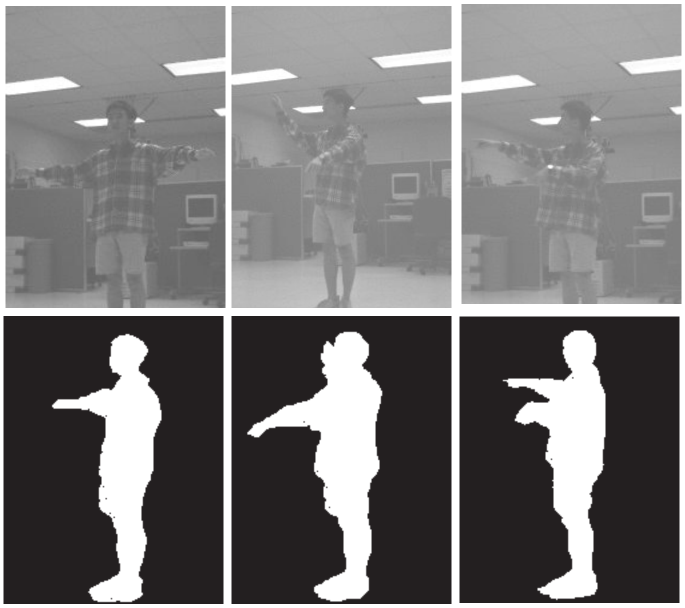


Figure 15. View synthesis using image-based visual hull.

investigated two approaches for handling wide variations in walking directions. One of the approaches uses the image-based visual hull approach for generating side views of walking humans from arbitrary directions. An example is given in Figure 15. Image-based visual hull approaches require at least four views of calibrated images to produce satisfactory results. Another approach uses a cardboard model of humans at a distance and generates side views of humans walking in arbitrary directions. Width vectors calculated from such side views can be used in HMM or eigengait methods. An example of the latter is shown in Figure 16.

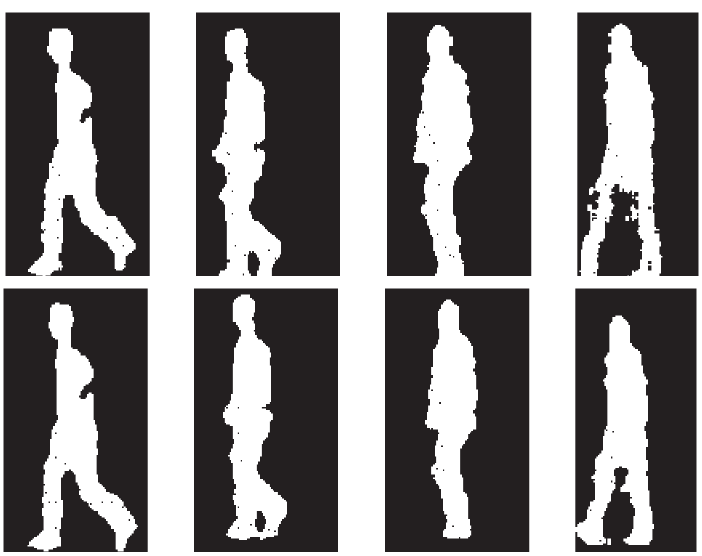


Figure 16. The top row represents different stances of a person walking at an angle of 30 degrees to the camera; the bottom row represents a fronto-parallel side view synthesized from the original video.

Figure 17. The top row shows gallery images. The bottom row shows sample frames from one probe video.

Robust Face Recognition from Video

Human recognition from video (see Figure 17) requires solving two problems, recognition and tracking, simultaneously. This leads to a parameterized time series state space model, representing both the motion and identity of the human. Sequential Monte Carlo algorithms have been developed to offer numerical solutions to this model. However, in outdoor environments, the solution is more likely to diverge, causing failures in both recognition and tracking. We have developed a robust algorithm for tackling this problem by incorporating the constraint of temporal continuity in the observations. Experimental results demonstrate improvements over the CONDENSATION algorithm.

Shape-Encoded Tracking

We exploit the fact that the structure of the human body is well known and can be approximated by an articulated blob model or other more accurate articulated models. Optical flow under scaled orthographic projection is used to relate the spatial-temporal intensity change of the image sequence to the human motion parameters. These motion parameters, including translation/rotation of the torso and rotation angles/velocities of the joints, are eventually solved through a group of linear equations to achieve global optimization in the sense of minimizing least square error. The physical limit of each joint can be easily added into the equations so that the problem becomes one of linear programming, which can be efficiently solved. In order to handle the error, we are also working on building a robust statistical model to do closed loop tracking, so that the error does not accumulate too much during tracking. Some examples of shape-encoded tracking of humans are shown in Figures 18 and 19.

Figure 18. Optical-flow based tracking of a human.

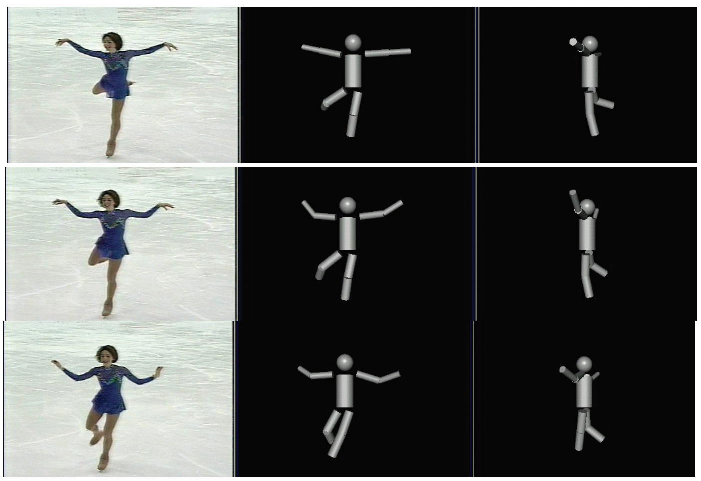


Figure 19. Particle filter based tracking of a human.

Human Activity Monitoring

Monitoring activities in a region from video data is an important surveillance problem. The goal is to learn the pattern of normal activities and detect unusual ones by identifying activities that deviate appreciably from the typical ones. We are using an approach based on statistical shape theory (derived from Kendall's shape model). In a low-resolution video each moving object is best represented as a moving point mass or particle. An activity can then be defined by the interactions of all or some of these moving particles over time. We model the configuration of particles by a polygonal shape formed from the locations of the points in a frame, and the activity by the deformation of the polygon over time. These parameters are learnt for each typical activity. Given a test video sequence, an activity is classified as abnormal if the probability of the sequence (represented by the mean shape and the dynamics of the deviations), given the model, is below a certain threshold. This approach gives very encouraging results in surveillance applications (Figure 20) using a single camera, does not depend on 3D information and is able to identify various kinds of abnormal behaviors.

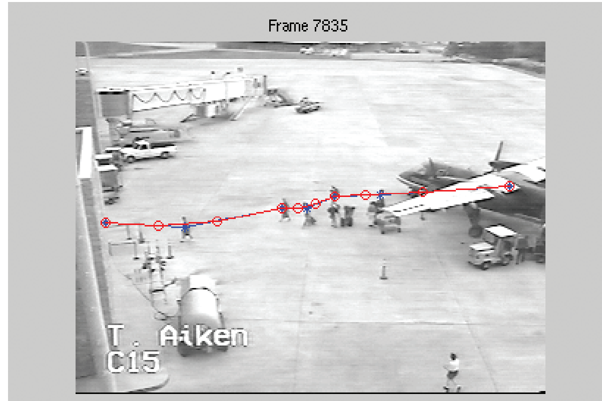


Figure 20. Activity-encoded particle filtering.

Automatic Target Detection Using Multi-Temporal Dual-Color IR imagery

We are developing automatic vehicle detection and recognition algorithms for infrared sequences collected by a dual-color sensor. It has been shown that the dual color IR sensor gives improvements in ATR performance. Traditional approaches use only still dual-color images. We expect that by exploiting the temporal variations in the dual-color IR sequences, intensity changes due to environmental factors (such as illumination, weather, etc.) can be accounted for, resulting in improved ATR performance. We have performed preliminary target segmentation/detection using eigenvectors corresponding to the top eigenvalues. Dual-color IR images are first adjusted for environment temperature variations (estimated from the mean of the input image). Then a Hotelling transformation is performed using the covariance matrix of the thermal variation normalized images. Figure 21 shows an example of truck segmentation. A silhouette verification based ATR algorithm can then be used to detect the target.

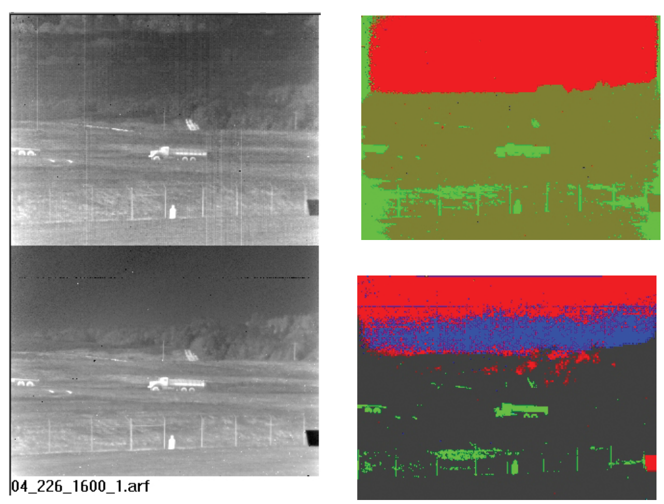


Figure 21. An example of target segmentation using top eigen-projections.

Larry S. Davis

During the past year we have focused on problems related to visual surveillance. We have investigated the use of kernel density estimation techniques in real-time visual surveillance and monitoring. We have shown how a variety of low- and intermediate-level vision problems, from background modeling and detection to tracking, can be robustly and efficiently addressed using efficient algorithms for the construction and evaluation of such probabilistic models. An example is shown in Figure 22. Here, color appearance models for people are constructed using kernel density estimation techniques. These models can then be used in conjunction with occlusion modeling to segment small groups of people into individuals.

This work has been extended to a multi-perspective framework. We have shown how one can combine wide baseline stereo analysis, general-purpose color segmentation, background modeling, and Kalman filter tracking into a unified Bayesian framework for building appearance models and tracking multiple people in cluttered environments. In Figure 23 we show the result of segmenting one frame from one view in a situation in which five people are in the fields of view of the cameras.

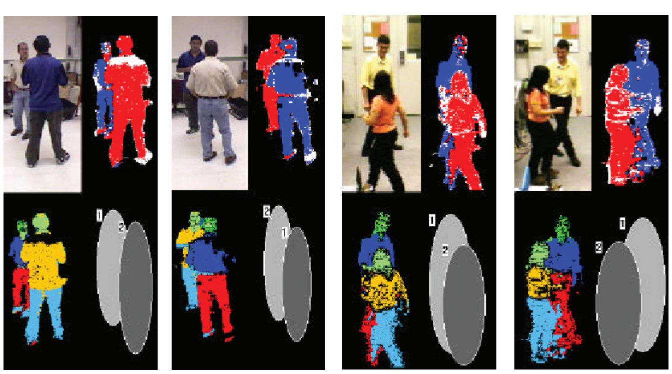


Figure 22. Construction of two color appearance models using kernel density estimation.

Figure 23. Segmentation of an image containing five people.

Figure 24. Reconstruction of a person from eight widely separated views.

We have developed an eigengait-based approach for the identification of people from a distance. We have shown how models for periodic motion analysis can be adapted to the problem of recognizing individuals based on their gaits. We have also studied methods of computing and utilizing cadence and stride for gait recognition.

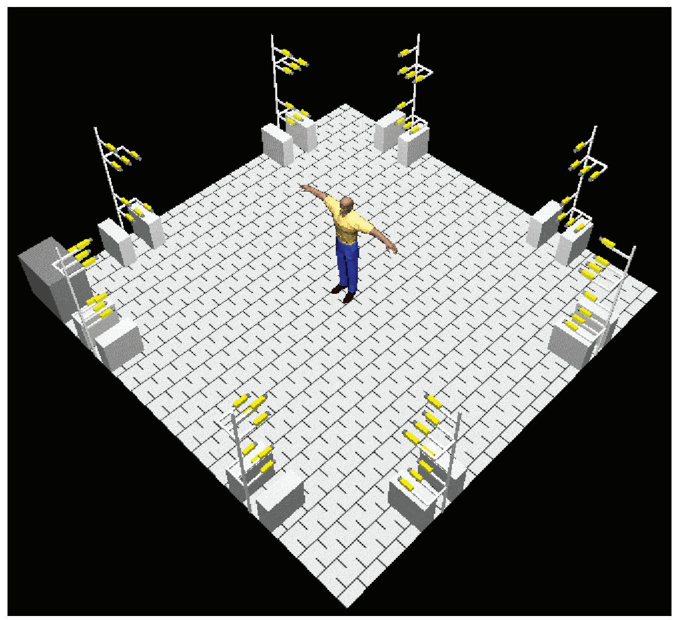
We are studying problems related to multi-view wide baseline stereo reconstruction, and have recently developed a stereo algorithm that works on a discrete voxelized representation of the scene. An iterative scheme is used to estimate the probability that a scene point lies on the true 3D surface. The novelty of this approach lies in the ability to model and recover surfaces which may be occluded in some views. This is done by explicitly estimating the probabilities that a 3D scene point is visible in a particular view from the set of given images. This relies on the fact that for a point on a Lambertian surface, if the pixel intensities of its projections into two views differ, the point must be occluded in one of the views. The image in Figure 24 is a VRML reconstruction of a person from eight widely separated views using this algorithm.

Keck Laboratory

The Keck Laboratory for the Analysis of Visual Movement ("Keck lab") is a multi-perspective imaging laboratory, containing 64 digital progressive-scan cameras organized as sixteen relatively short baseline stereo rigs (see the schematic diagram below, right). In each quadranocular rig, there are three monochromatic cameras and one color camera. Each group of four cameras is connected to a PC running Windows NT 1 which can collect imagery from the cameras at speeds of up to 85 frames per second (FPS). The PCs are all networked. The dimensions of the Keck laboratory are 7mx7mx3m. A panoramic view is shown below.

A primary goal in the design of the Keck lab was to maximize captured video quality, while using commonly available hardware for economy. To meet this goal, uncompressed video is captured using digital progressive-scan cameras directly to PCs. We chose Kodak ES-310 cameras with a resolution of 648x484x8. These cameras can operate at up to 85 FPS in full frame progressive-scan mode (speeds up to 140 FPS can be achieved using a smaller region-of-interest window). The ES-310 has a 10-bit digitizer for each pixel, in which the user can select which 8 bits are used for digital output, allowing for an effective gain in low-light conditions. The camera has a software-controllable iris and shutter speed, is capable of being externally triggered, and can be controlled via an external serial interface (RS 232 or RS 485). Each CCD has a two-tap digitizer to achieve fast A/D output. Forty-eight of the cameras are grayscale (ES-310), while the remaining sixteen are color (ES-310C). The color cameras are identical to the grayscale ones, except for the addition of a Bayer color filter mask on the CCD, and an IR filter. To process the images from the color cameras, color pixel (RGB) values need to be spatially interpolated.

Each acquisition PC runs Windows NT 4.0 SP4, and is equipped with four Matrox Meteor II digital frame grabbers. Each PC has five PCI slots. One slot was left free for any future cards (such as gigabit networking, which is currently being implemented). The video display, keyboard and mouse outputs for each of the PCs are aggregated using Apex monitor switches. The Keck lab was designed to capture uncompressed video sequences to both memory and disk.. The design of the lab allows capturing uncompressed video to memory at up to 100 MB/s, and capturing to disk at up to 50 MB/s. In order to achieve the required 50 MB/s disk throughput, three SCSI Ultra 2 Wide disks (Seagate Cheetah) are used in a RAID configuration.



David Jacobs

Prof. Jacobs' research focuses primarily on problems related to the process of recognizing objects in images. This process begins with perceptual organization, which groups together components of an image that belong to individual objects, as a precursor to recognizing these objects. To then determine the identity of an object requires an understanding of how the shape and material properties of objects determine their appearances in images. This leads to two key problems. First, how do we recognize objects under unknown, complex lighting conditions? Second, how do we recognize objects from various viewpoints? Finally, recognition also involves building and using a large database of different objects. To understand recognition we need to understand how to organize such a database for efficient and effective retrieval.

Recognition and Lighting

Variability in lighting has a large effect on the appearance of objects in images, as is illustrated in Figure 25. We can better recognize objects under diverse lighting conditions if we can understand this variability. We approach this problem in several ways. First we consider how to recognize Lambertian objects with known 3D structure. Second, we study non-Lambertian objects that have known structure but are made of materials with unknown reflectance properties. Third, we show how to recover the 3D structures of objects, using images taken under various, complex lighting conditions. Finally, we consider recognition when we have previously seen an object but don't know its 3D structure.

Figure 25. The same face under two different lighting conditions.

For convex, Lambertian objects, we have devised a simple way of characterizing the set of images produced under a range of lighting conditions. For light distant from an object, we can describe its intensity as a function of direction. We analyze lighting by representing these functions using spherical harmonics. This is analogous to Fourier analysis. To model the way surfaces turn light into an image we look at reflectance as a function of the surface normal. We show that reflectance functions are produced through the analog of a convolution of the lighting function using a kernel that represents diffuse reflection. This kernel acts as a low-pass filter with 99.2% of its energy in the first nine components, the zero, first, and second order harmonics. As a result, we can show that for non-shiny objects, only the diffuse, low-frequency components of the light affect the way the object looks. This explains previous empirical results that show that the set of images of an object can be represented by a low-dimensional subspace in the space of all images. Our results show how to build this representation analytically. Using these results we can construct algorithms for object recognition based on linear methods. We apply these algorithms to perform face recognition by finding the 3D model that best matches a 2D query image.

When objects can be shiny, the set of images they produce can be much more complex. For example, a convex mirror can produce any image. Shadows cast by part of a non-convex object on another part of the object add to this complexity. Recently, we have extended our results to these situations. We are currently working on building novel recognition algorithms that can handle Lambertian objects with unknown albedo, and non-Lambertian objects. We have also shown how to use our characterization of the images produced by Lambertian objects to perform photometric stereo, as illustrated in

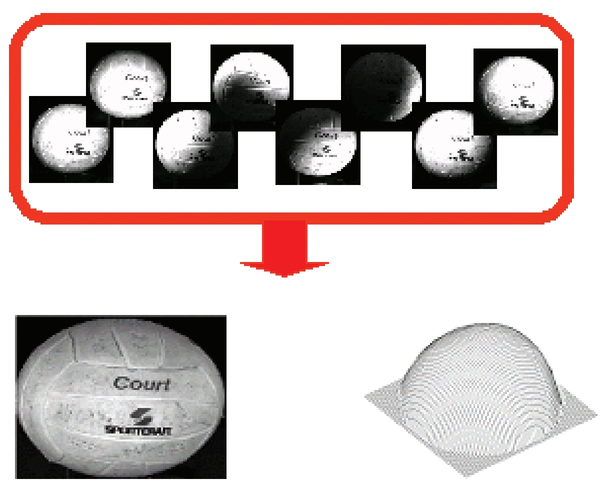


Figure 26. Top: a set of images. Bottom: the albedo (left) and shape (right) derived from photometric stereo.

Figure 26. That is, given a set of images of an object taken under different lighting conditions, we can recover its structure and albedo. Our approach differs from past work on photometric stereo in handling images in which the lighting is unknown. Also, we can handle images in which the lighting in each image contains multiple sources, both point and diffuse. Finally, we have investigated the problem of recognizing an object when we don't know its 3D structure, and must compare a new image directly to a previous one. We show how to devise an image comparison method that is relatively insensitive to lighting changes.

Geometry and Recognition

For some time we have been working to understand how to relate the 3D shape of an object to the shape of its 2D projection in an image. In recent work we attempt to understand how we can see a shape in one or more images and then recognize that shape in another image. For example, how can we decide that the silhouette of a monkey seems to come from a different object than the one that produced a previously seen silhouette of an elephant? And how can we decide that a silhouette of an elephant viewed from the front and from the side seem consistent with the same object?

We have been trying to understand the fundamental constraints on silhouette matching. That is, when is it possible that a set of silhouettes all come from the same object? We have solved a special case of this problem. In this case, objects do not move, and the camera is only allowed to rotate about a single axis parallel to the image, such as a mobile robot moving on a flat floor (Figure 27). Assuming orthographic projection, we can determine whether three images are consistent with a single object by solving a linear program.

Perceptual Organization and Memory

For some time we have worked on using statistical models of object shape and background clutter to perform

Figure 27. Our algorithm determines whether three silhouettes, such as these, could come from the same object.

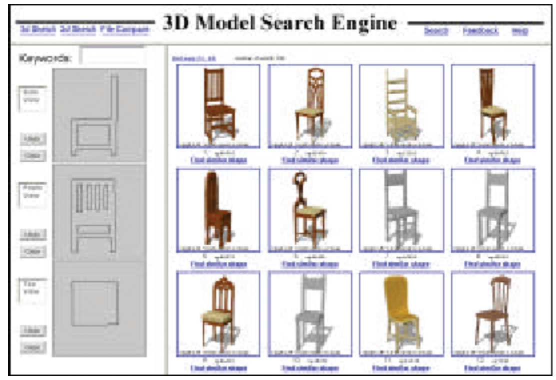


Figure 28. Left: a user has sketched a chair. The system retrieves graphics models with matching shapes.

figure/ground segmentation and perceptual grouping. For example, if an object boundary contains gaps, we can use a statistical model of object shape to find the most likely way of filling in these gaps.

In current work, we are integrating these statistical object completion processes into a recognition system that retrieves objects from a memory. The basic idea is to build a statistical model that summarizes the general properties of objects. Then, given partial information about an unknown object, this model is used to compute a probability distribution over the possible ways of filling in the missing information about the object. This information is then used to retrieve likely objects from memory. This process can be iterated until a matching object can be found.

We have implemented a system to solve simple word memory problems. The system finds a word that contains a set of letters in a specific order. For example, in the problem: '*x*m*l*', a matching word must contain 'x', 'm', and 'l' in that order; '*' can be replaced by zero or more missing letters. 'example' is a valid answer. The model displays patterns of performance similar to those displayed by people. We are currently extending this model to solve vision problems.

We have also been considering how to retrieve 3D computer graphics models from a large library of models. We have constructed a search engine that crawls the web, downloading computer graphics models, and then organizes them for text- or content-based retrieval. Figure 28 shows an example of content-based retrieval. The system can be accessed at: shape.cs.princeton.edu/search.html.

Azriel Rosenfeld

Prof. Rosenfeld's research has dealt with many aspects of image analysis and computer vision. This report summarizes recent work on fuzzy mathematics, digital topology, point-to-line mappings, handwriting, and stereo.

Fuzzy Mathematics

We have studied basic dissimilarity measures between fuzzy subsets of a finite set *S*, and have shown that both of the measures are metrics. The measures can also be generalized to fuzzy structures, such as fuzzy graphs, defined on *S*; the graph generalization is closely related to a recently defined measure of graph dissimilarity.

We have defined fuzzy end nodes in a fuzzy graph, and have shown that no node can be both a cut node and a fuzzy end node. In a fuzzy tree, every node is either a cut node or a fuzzy end node, but the converse is not true. We have also shown that any nontrivial fuzzy tree has at least two fuzzy end nodes, and have characterized fuzzy cycles that have no cut nodes or no fuzzy end nodes.

Digital Topology

We have generalized basic parts of digital geometry to sets P of ***tiles*** (compact, convex sets) in R^2 or R^3. Tiles are a generalization of pixels or voxels; they can have arbitrary shapes, are allowed to have overlapping interiors, and need not cover the space. The union of all the tiles of P is denoted by $U(P)$. The ***neighborhood*** $N_P(p)$ of a tile p is the union of the tiles of P that intersect p. p is called ***simple*** if deletion of p from P does not change the topology (in the homotopy sense) of $U(P)$. We have shown that if P satisfies a property called ***strong normality*** (*SN*), and deletion of p preserves the topology of $N_P(p)$, then p is simple. This may not be true if P is not *SN*; and even if P is *SN*, p may be simple even if deletion of P does not preserve the topology of $N_P(p)$.

Point to Line Mappings

In 1962, Hough used a linear point-to-line mapping (PTLM) to detect sets of collinear points in an image, by mapping the points into concurrent lines and detecting peaks where many lines intersect. In 1972, Duda and Hart pointed out that Hough's method is not practical, because the peaks need not lie in a bounded region. Since 1972, methods of detecting sets of collinear points have made use of nonlinear point-to-curve mappings that map collinear points into concurrent curves whose intersections do lie in a bounded range. We have shown that any PTLM that maps collinear points into concurrent lines must be linear, and that no such PTLM can map all the sets of collinear points in an image into peaks that lie in a bounded region; thus Duda and Hart's objection applies to any PTLM-based Hough transform.

Handwriting

We have developed methods of detecting forgeries of off-line signatures. Although a great deal of work has been done on off-line signature verification over the past two decades, the field is not as mature as on-line verification. Temporal information used in on-line verification is not available off-line and the subtle details necessary for off-line verification are embedded at the stroke level and are hard to recover robustly. We approach the off-line problem by establishing a local correspondence between a model and a questioned signature. The questioned signature is segmented into consecutive stroke segments that are matched to the stroke segments of the model. The cost of the match is determined by comparing a set of geometric properties of the corresponding sub-strokes and computing a weighted sum of the property value differences. The least invariant features of the least invariant sub-strokes are given the biggest weights, thus emphasizing features that are highly writer-dependent. Random forgeries are detected when a good correspondence cannot be found, i.e., the process of making the correspondence yields a high cost. Many simple forgeries can also be identified in this way. The threshold for making these decisions is determined by a Gaussian statistical model. Using the local correspondence between the model and a questioned signature, we perform skilled forgery detection by examining the writer-dependent information embedded at the substroke level and trying to capture unballistic motion and tremor information in each stroke segment, rather than as global statistics.

Stereo

We have developed a hierarchical stereo matching strategy using the Discrete Wavelet Transform (DWT). Both area- and feature-based methods are combined into a single process by taking advantage of the discrete wavelet decomposition. Image components extracted from the approximation, horizontal, and vertical channels of the decomposition are combined to do the matching at each level. The disparity at a coarse level is then propagated to the finer levels. Because the detail channel of the decomposition is discarded, noise is automatically reduced. Experiments using various kinds of image pairs show that the method is accurate, fast, and highly robust to noise.

Hanan Samet

Prof. Samet's research is concerned with many aspects of spatial databases and image databases with a heavy emphasis on representations. The applications are in geographic information systems (GIS), computer graphics, and image processing.

A significant part of our research efforts deals with the integration of spatial and nonspatial data into a database management system. Our approach to solving this problem is based on the observation that the problem is really one of sorting. The difference from other approaches is the realization that the geometric data must be sorted with respect to its extent rather than just as points, as is being done by most researchers. This research is supported by the QUILT GIS, which is a working geographic information system developed by us. In the past few years, we have been able to integrate parts of the QUILT GIS into a home-grown database system which we call SAND. One of the results of this work has been the development of a Browser, called the SAND Spatial Browser, which enables users to pose queries that combine spatial and nonspatial data. The ultimate goal of the SAND Spatial Browser project is the creation of a spatial spreadsheet that enables users to interact with spatial data with the same ease that they have been accustomed to with conventional spreadsheets. A prototype spatial spreadsheet has been developed in which each cell corresponds to an instance of the SAND Spatial Browser.

One of the most noteworthy features of the SAND Spatial Browser is the user interface which enables spatial queries to be specified graphically instead of requiring the use of SQL. Of course, an eventual goal is to develop a translator from the graphical specification to SQL, which is a non-trivial task.

In the past year we have continued our development of a JAVA-based version of the SAND Spatial Browser that is designed to enable its use over the internet. This raises challenging problems as to what is to be performed on the server and what is to be performed on the client.

For example, Figure 29 is an example screen shot from the JAVA-based SAND Spatial Browser. It shows the relation corresponding to mortality rates per 100,000 for bladder cancer for white males for the time period 1970-1994 obtained from the National Atlas of Cancer Mortality. We have also overlaid it with the result of a clustering-like operation that is available in SAND. In particular, we have shown a partition of the underlying space with respect to the 17 counties with the highest mortality rates so that each county in each partition is closer to the county with the high rate in that partition than to any other county with a high rate. The green dots indicate locations of high chlorine emissions obtained from EPA data. The goal of this query is to determine if there is some spatial correlation between counties with a high incidence of bladder cancer and large chlorine emissions. As can be seen, locations with a large amount of chlorine emission are not clustered around these counties. Thus these two events do not seem to be spatially correlated.

The scenario depicted in Figure 29 is analogous to a discrete Voronoi diagram, and is a form of clustering. This clustering operation is available in the SAND Spatial Browser and is achieved by executing an incremental ***distance semi-join*** operation, using an algorithm developed by us, where the input relation corresponding to the high chlorine emissions map is joined with the high incidence of bladder cancer map, and the join condition is based on proximity with the closest tuple pairs from the two sets being retained. Once the closest emissions-cancer pair (a,b) has been found, the next closest pair is found from the set of emissions tuples which excludes tuple a from participating. This process is continued until the closest high incidence of bladder cancer county has been found for each of the high chlorine emission locations.

We have also started to work with the data on the FedStats web site. This web site contains statistics gathered by federal agencies that is available to the general public over the world wide web. Our goal is to make the SAND Browser the tool of choice for examining such data. This also requires the ability to input the data into the SAND system. One of the approaches that we tried was to convert data in EXCEL format. However, this process was somewhat cumbersome and unreliable when accessing data from different sources in different formats. In order to access multiple data sources in real

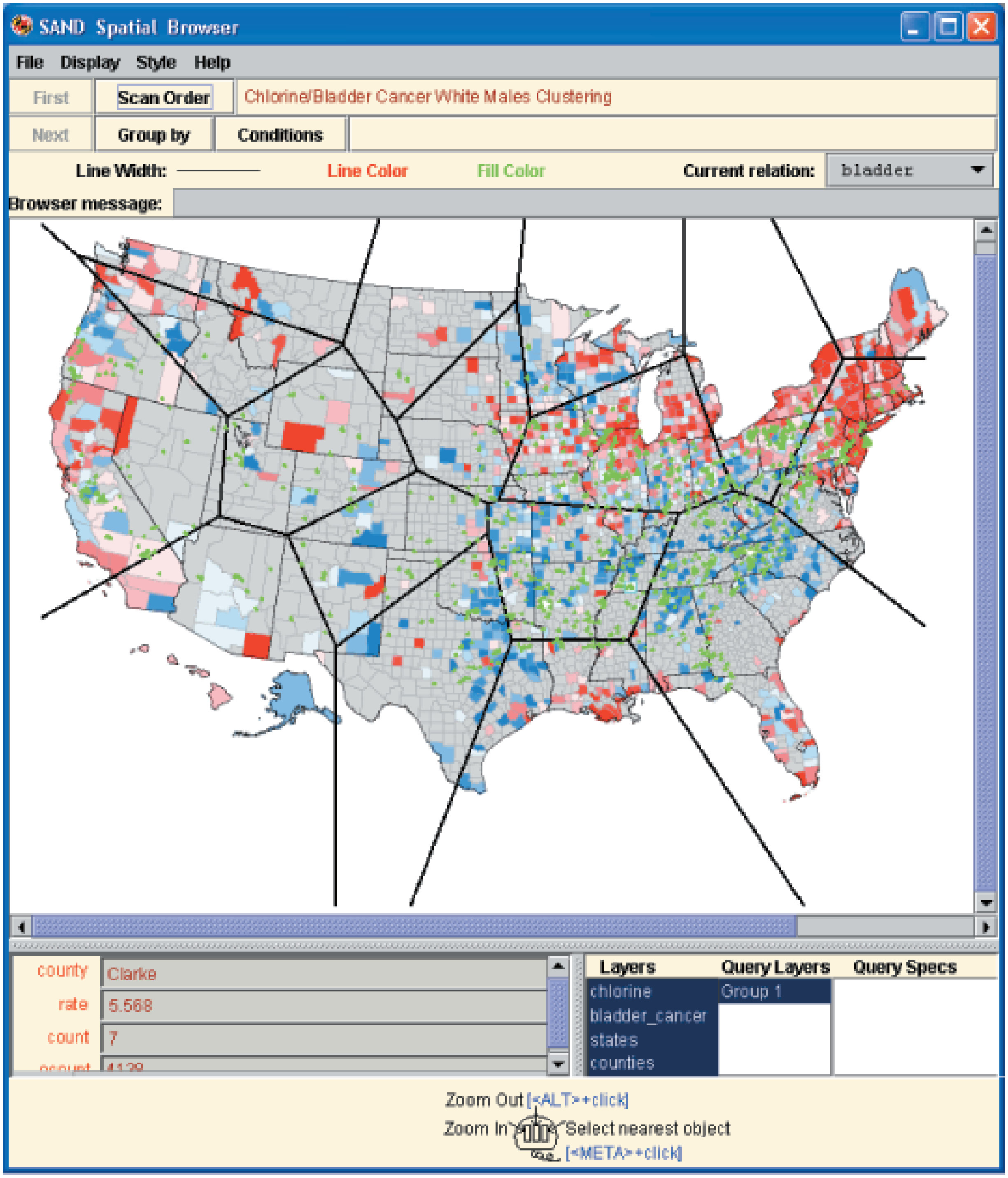


Figure 29. Sample screenshot of a possible user interaction with the SAND Internet Browser.

time reliably, we looked for another mechanism that would support data exchange by design. The XML protocol and its Document Type Definitions (DTDs) have emerged to become virtually a standard for describing and communicating arbitrary data, and GML is becoming increasingly popular for exchange of geographical data. Thus we have made SAND XML-compatible so that users can instantly retrieve spatial data provided by various agencies in the GML format via their web services and then explore, query, or process this data further within the SAND framework.

On a parallel track to the client-server environment, we are also investigating the use of the SAND Browser in a peer-to-peer environment. The motivation here is to be able to deal with FedStats applications where

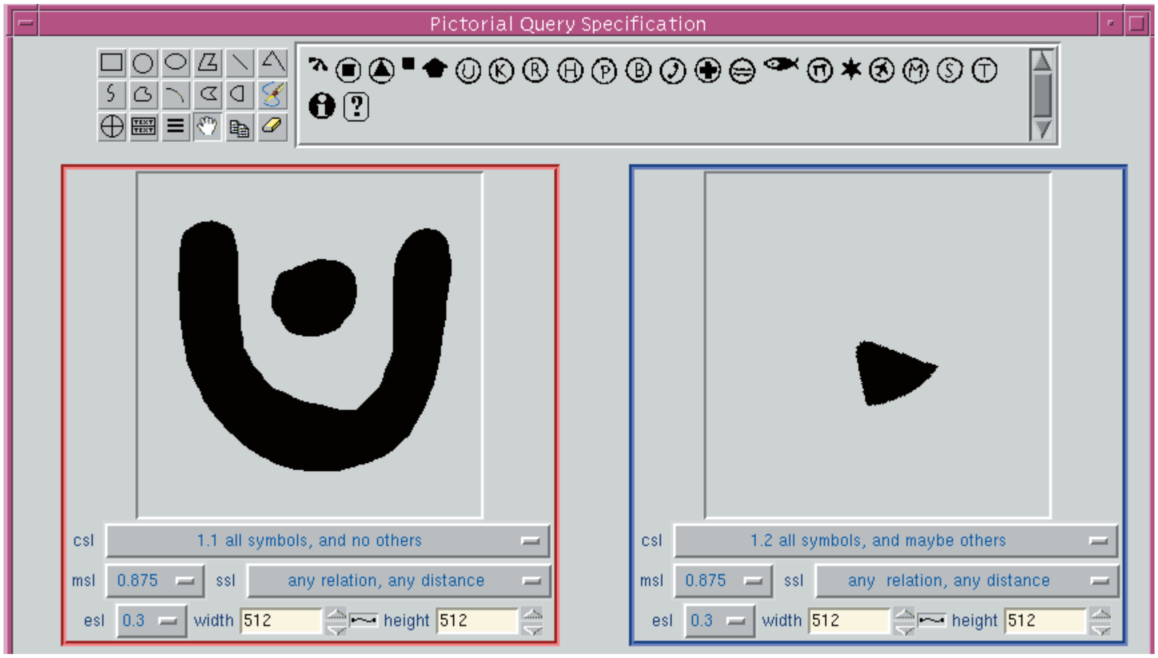


Figure 30. Sample pictorial query containing a u-bend and a circle on the left and a triangle shape on the right.

users have converted their data to the SAND internal format and now wish to store it back on the server. The server may be swamped and there may be no bandwidth left for the transaction We propose to store the data in other computers and upload it once the server is less busy. Another issue arises when some users want a data set that has already been converted. This data set can reside on several computers. In this case we wish to transfer it more quickly to the user by fetching different parts of the data set from different users. We are investigating methods for dealing with this scenario.

One of the most prominent features of the SAND Browser is the ability to perform ranking. This is an operation that enables us to retrieve data in the order of their distance (spatial) from another spatial object, and we can also restrict the domain from which the data are drawn. Our approach differs from conventional approaches in the sense that it is incremental. In essence, once we have obtained the k nearest items, if we want the next nearest (i.e., the $k+1^{st}$ nearest), then we can obtain it directly instead of having to recalculate all k+1 nearest neighbors. Our approach makes use of a priority queue on the objects. Neither the reference query object nor the objects stored in the database are restricted to points; they can be lines, areas, surfaces, etc. We are currently extending this work to deal with moving objects and hence handle kinetic representations.

Other research in the spatial database domain includes continued investigation of the use of the PK-tree to index spatial (non-point) data. The PK-tree is a dynamic multidimensional data structure that decouples the tree structure of the decomposition of the underlying space from that of the node hierarchy (i.e., the grouping of the nodes corresponding to the blocks resulting from the decomposition) that makes up the tree directory of the access structure. The PK-tree is a grouping technique that can be adapted to any spatial index that is based on a disjoint decomposition of space, usually regular, although this need not be the case. When used on secondary storage, it can yield a guaranteed minimum storage utilization of 50%, with the average being similar to that typically obtained with B-trees (i.e., 65-70%). The advantage of the PK-tree over many other spatial indexing structures is its simplicity of design and its large fanout. We have shown that the PK-tree is a viable competitor to widely used representations such as the R-tree. In fact, our experiments demonstrate that the PK-tree has comparable and often superior performance, to the R*-tree in terms of I/O cost for some conventional spatial queries (e.g., nearest neighbor and windowing), and significantly faster build times. We are also investigating other representations that are based on decoupling, such as the BV-tree.

In the image database area we have worked on two problems. The first was a comparative study of global and local methods for the shape analysis of logos in an image database. The methods are judged by using the shape signatures to define a similarity metric on the

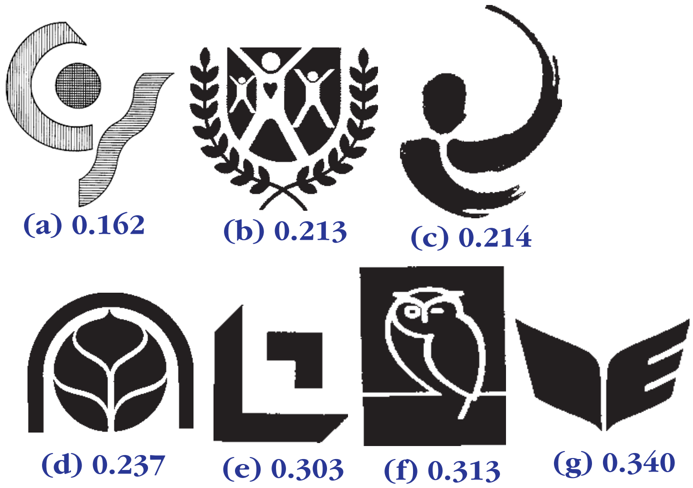


Figure 31. Results of the pictorial query in the left part of Figure 30 (seeking a U-Bend and a circle).

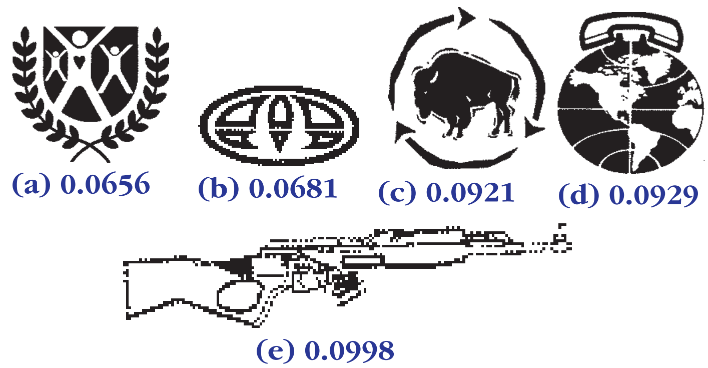


Figure 32. Results of the pictorial query in the right part of Figure 30 (seeking a triangle shape).

logos. As representatives of the two classes of methods, we used the negative shape method, which is based on local shape information, and a wavelet-based method, which makes use of global information. We applied both methods to images with different kinds of degradations and examined how a particular degradation affects the strengths and shortcomings of each method. The results were used to develop a new method which is based on the relative performances of the two methods and is robust with respect to all degradations examined.

The second problem was the development of a system that enables the pictorial specification of queries in an image database. The queries are comprised of rectangle, polygon, ellipse, and B-spline shapes. The queries specify which shapes should appear in the target image (termed ***csl***, or contextual similarity level), as well as spatial constraints on the distance between them, their relative position (termed ***ssl*** , spatial similarity level), and relative size (termed ***esl***, extent similarity level). The retrieval process makes use of an abstraction of the contour of the shape which is invariant to translation, scale, rotation, and starting point and is based on Fourier descriptors. These abstractions are used to locate logos in an image database. The system can also operate based on the classification of the shapes as one of the primitive shapes (i.e., rectangle, polygon, ellipse, and B-spline).

For example, Figure 30 shows a sample pictorial query containing two query images. The query image on the left consists of a u-bend and a circle, while the query image on the right consists of a triangle shape. In this case, we let csl=2 for each of the query images, which means that the database image must contain all components of the query image and possibly others, and that there are no spatial constraints. The sizes are supposed to be similar. The results are shown in Figures 31 and 32, ordered by their matching rank. All seven resulting logos do contain a u-bend and a circle. Note that the u-bend occurs at different rotations. Similarly, we have five results for the triangle shape.

The combined query seeks all result images that contain a u-bend and do not contain a triangle shape. The result of this query is shown in Figure 33. It is the same as Figure 31 except for the emblem-like image in Figure 31b which has a triangle component.

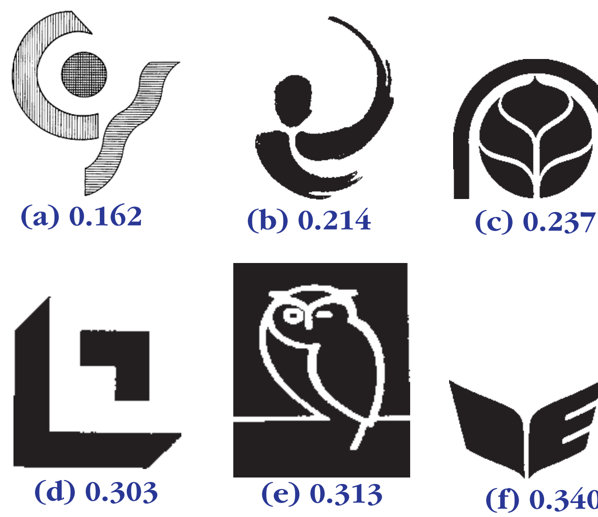


Figure 33. Results of the pictorial query on Figure 31 (seeking a U-Bend and no triangle shape).

Isaac Weiss

Isaac Weiss's research involves two main subjects: application of invariant theory to computer vision, and robust estimation.

Invariants are very powerful tools for object recognition, which is a central goal of computer vision. One example of how invariants are used for recognition is provided by the viewpoint problem. Different images of the same object often differ from each other because of the different viewpoints from which they were taken. Common recognition methods try to match an image of an object, seen from an unknown viewpoint, to an image stored in a data base. For this they need to find the correct viewpoint, a difficult problem that can involve search in a large parameter space of all possible viewpoints and/or finding point correspondences. Geometric invariants are shape descriptors, computed from the image, that remain unchanged under transformations such as changing the viewpoint. Thus they can be matched without search. Projective (viewpoint) invariants of curves and surfaces were an active mathematical field in the 19th century. However, in the vision community they have only in the last decade received major attention, after a paper by Dr. Weiss reviewed them and pointed out their importance.

Invariants are only a part of the tool-box needed for object recognition. We have begun work on recognizing objects in 3D range images. Our contribution to the project involves using a hierarchical graph-based invariant representation of objects, as well as using advanced statistical methods for deciding on the correct model to match to the observed object.

Objects can undergo other types of geometric changes, in addition to viewpoint changes. Among them is articulation, i.e. when an object has parts that move with respect to one another such as the turret of a tank. We have developed a method that recognizes such objects in range images, and also provides the correct articulation. This is done without extracting local features such as edges, which can be quite unreliable in these images.

Part of the difficulty of object recognition is the variability of objects. For instance, we can recognize apples even though different apples have slightly different shapes. We have applied invariants of such small deformations to characterize classes of objects such as apples and bananas, and to differentiate these classes from each other.

Invariants in fact have much wider uses than eliminating geometric unknowns. Images also depend on the physical process that creates them, which can involve visible light, infrared, radar, etc. The physical component has even more unknowns then the geometrical one.

For example, in an ordinary image, the amount of light falling on the film depends on physical quantities such as surface reflectance, illumination intensity and its spatial distribution, characteristics of the imaging system, etc. Recovering the original shape requires the elimination of all these unknowns. This is the "Shape from Shading" problem. Infrared images also depend on surface temperature. All these unknowns greatly complicate the recognition task.

Invariants of physical processes have been extensively studied in modern physics and there are various methods of finding them. Simple examples are the laws of conservation of energy and momentum, which are in fact invariants of the laws of motion. In current research some of these methods are being adapted for application to computer vision.

Robust estimation techniques are of very general significance, and of particular value in obtaining invariants. This is because invariants depend on accurate measurements of features such as point coordinates or curve parameters. Part of the progress that has been made is in the theoretical analysis of the problem, especially in finding high-order derivatives. Another part of our research is on using non-parametric methods, such as variants of bootstrap, which are harder to study analytically. Of special interest here is the problem of removing "outliers" from the given data. These outliers distort the results of simpler methods such as least-squares fitting, and removing them can greatly increase the reliability of the measurements.

Contracts and Grants

Sponsor	Dates	Amount	Title
DARPA/ONR	08/01/00 - 12/31/02	1,922,393	Model-Based Integrated Approaches for Remote Identification of Humans
NSF	09/15/99 - 08/31/02	1,096,011	High-Performance Systems for Shape and Action Modeling
MURI/ARO	08/01/02 - 07/31/07	1,000,000	Spectral Object Level Change Detection
DARPA/ARL/SRI	07/16/98 - 04/30/02	961,964	4D Model-Based Activity Monitoring
ARDA (VACE)	08/25/00 - 08/24/02	898,380	Activity Detection by Video Content Evaluation (ADVICE)
ARL/BAE Systems	05/31/01 - 09/30/02	713,100	Advanced Sensors CTA
NSF	06/15/01 - 05/31/03	616,172	Scalable Data Collection Infrastructures for Digital Government Applications
NSF	09/01/00 - 08/31/04	520,000	Real-Time Capture, Management and Reconstruction of Spatio-Temporal Events
NSF	09/01/99 - 08/31/03	500,000	NSF-CNpq Collaborative Research: Issues in the Development of Spatial Spreadsheets and Browsers
ARL/New Mexico State University	07/01/01 - 09/30/02	500,000	Advanced Decision Architectures CTA
DARPA/AFRL	09/20/99 - 09/19/02	494,301	Real time, Distributed Agents for Visual and Acoustic Sensor Data Processing
ARL/General Dynamics	06/11/01 - 09/30/02	468,000	Robotics CTA
DARPA/AFRL	02/24/97 - 02/25/02	366,589	3D Object Recognition from Multiple and Single Views
DARPA/ONR	07/01/02 - 06/30/03	364,799	Human Activity Analysis and Recognition using Distributed Cameras
ONR	01/01/01 - 12/31/03	300,000	Human Tracking and Verification in Video
DARPA/ONR	12/01/99 - 11/30/02	288,760	Gesture-Driven Control of Spaces and Objects in Collaborative Augmented Reality
NSF/UMAB	09/15/97 - 08/31/01	249,213	Three-Dimensional Reconstruction of Synapses
NSF	10/01/00 - 09/30/03	199,999	Alternative Camera Technology
ARL/General Dynamics	03/14/02 - 05/30/03	196,000	MDARS-E
DARPA/Object Video	06/01/02 - 12/30/03	166,667	Video Monitoring, Tracking, and Classification
NSF	03/15/99 - 02/28/03	166,278	Real-Time Urban Management System for Dynamic City Visualization and Decision Support
DARPA/General Dynamics	12/19/01 - 12/29/02	150,000	Perceptor - Phase II: Recognition in Multi-Hyperspectral Imagery
NSF	10/01/02 - 09/30/04	150,000	3D Description and Recognition of Human Activities
Philips Research Laboratories	06/01/98 - 05/31/02	147,442	Real Time Visual Surveillance
NASA	01/01/98 - 03/31/03	137,770	Enhanced Metadata Extraction for NASA's Regional Validation Center
Raytheon	08/01/02 - 08/31/03	125,000	E3D
NASA/Foster Miller	07/18/01 - 09/30/02	110,100	Crew Performance Analyzer - Phase II
NIST	09/28/01 - 09/27/02	85,112	Gesture-Based Control of Mobile Platforms
DARPA/General Dynamics	04/02/01 - 09/30/01	71,051	Perceptor--Phase I
Lockheed Martin	07/01/01 - 07/01/02	50,000	Context-Based Detection of Vehicles in Aerial Images
Honda	01/01/02 - 12/31/02	32,000	Human Detection in Night Scenes

Journal Papers and Book Chapters

1. M. Agrawal and L. Davis, Trinocular Stereo Using Shortest Paths and the Ordering Constraint, ***International Journal of Computer Vision***, 2002.

2. Y. Aloimonos, Computational Video, ***The Visual Computer***, in press.

3. A. Amir, A. Efrat, P. Indyk, and H. Samet, Efficient Algorithms and Regular Data Structures for Dilation, Location and Proximity Problems, ***Algorithmica*** 30, pp. 164-187, 2001.

4. P. Baker, A. Ogale, and C. Fermüller, The Argus Eye: A New Tool for Robotics, ***Robotics and Automation Magazine***, in press.

5. A. Banerjee and R. Chellappa, A Statistical Physical Model for Foliage Clutter in UWB SAR Images, ***J. Optical Society of America***, in press.

6. R. Basri and D. Jacobs, Projective Alignment with Regions, ***IEEE Trans. on Pattern Analysis and Machine Intelligence*** 23, pp. 519-527, 2001.

7. R. Basri and D. Jacobs, Lambertian Reflectance and Linear Subspaces, ***IEEE Trans. on Pattern Analysis and Machine Intelligence,*** in press.

8. P. Bhattacharya, A. Rosenfeld, and I. Weiss, Point-to-Line Mappings and Hough Transforms, in ***Digital and Image Geometry*** (G. Bertrand, A. Imiya, and R. Klette, eds.), pp. 185-206, Springer, Berlin, 2001.

9. P. Bhattacharya, A. Rosenfeld, and I. Weiss, Point-to-Line Mappings as Hough Transforms, ***Pattern Recognition Letters*** 23, pp. 1705-1710, 2002.

10. P. Bhattacharya, A. Rosenfeld, and I. Weiss, Geometric and Algebraic Properties of Point-to-Line Mappings, ***Pattern Recognition***, in press.

11. T. Brodsky and C. Fermüller, Self-Calibration from Image Derivatives, ***International Journal of Computer Vision*** 48, pp. 91-114, 2002.

12. R. Chellappa and B.S. Manjunath, Texture Classification and Segmentation, in ***Foundations of Image Understanding*** (L. Davis, ed.), pp. 219-240, Kluwer, Boston, 2001.

13. R. Chellappa and A. Rosenfeld, Image Processing and Analysis, in ***Encyclopedia of Physical Science and Technology***, pp. 595-630, Academic Press, San Diego, 2001.

14. P. Cucka, Z. Duric, E. Rivlin, and A. Rosenfeld, Qualitative Description of Camera Motion and Scene Depth from Histograms of Normal Flow, ***Proc. Indian National Science Academy*** A, pp. 145-156, 2001.

15. S. Der, Q. Zheng, B. Redman, R. Chellappa, and H. Mahmoud, View-Based Recognition of Military Vehicles in Ladar Imagery Using CAD Model Matching, in ***Image Recognition and Classification: Algorithms, Systems and Applications*** (B. Javidi, ed.), Marcel Dekker, New York, 2002.

16. Z. Duric, R. Goldenberg, E. Rivlin, and A. Rosenfeld, Estimating Relative Vehicle Motions in Traffic Scenes, ***Pattern Recognition*** 35, pp. 1339-1353, 2002.

17. Z. Duric, W. Gray, R. Heshman, F. Li, A. Rosenfeld, A. Schoelles, C. Schunn, and H. Wechsler, Integrating Perception and Cognitive Modeling for Adaptive and Intelligent Human-Computer Interaction, ***Proc. IEEE 90***, pp. 1272-1289, 2002.

18. C. Esperanca and H. Samet, Experience with SAND/Tcl: A Scripting Tool for Spatial Databases, ***Journal of Visual Languages and Computing*** 13, pp. 229-255, 2002.

19. C. Fermüller and Y. Aloimonos, Statistics Explains Geometrical Optical Illusions, in ***Foundations of Image Understanding*** (L. Davis, ed.), Kluwer, Boston, 2001.

20. C. Fermüller, P. Baker, and Y. Aloimonos, Visual Space Time Geometry: A Tool for Perception and the Imagination, ***Proc. IEEE 90***, pp. 1113-1135, 2002.

21. C. Fermüller, D. Shulman, and Y. Aloimonos, The Statistics of Optical Flow, ***Computer Vision and Image Understanding*** 82, pp. 1-32, 2001.

22. T. Funkhouser, P. Min, M. Kazhdan, J. Chen, A. Halderman, D. Dobkin, and D. Jacobs, A Search Engine for 3D Models, ***ACM Transactions on Graphics***, in press.

23. C. Huang, L. Davis, and J. Townshend, An Assessment of Support Vector Machines for Land Cover Classification, ***International Journal of Remote Sensing*** 23, pp. 725-749, 2002.

24. D. Jacobs, Linear Fitting with Missing Data for Structure-From-Motion, ***Computer Vision and Image Understanding*** 82, pp. 57-81, 2001.

25. D. Jacobs, Perceptual Organization as Generic Object Recognition, in ***From Fragments to Objects: Segmentation and Grouping in Vision*** (T. Shipley and P. Kellman, eds.), Elsevier, 2002.

26. D. Jacobs, What Makes Viewpoint Invariant Properties Perceptually Salient?: A Computational Perspective, in ***Perceptual Organization for Artificial Vision Systems*** (K. Boyer and S. Sarkar, eds.), Kluwer, 2002.

27. D. Jacobs, J. Malik, and R. Nevatia, Breakout Session Report: Principles and Methods, in ***Perceptual Organization for Artificial Vision Systems*** (K. Boyer and S. Sarkar, eds.), Kluwer, 2002.

28. D. Jacobs, D. Weinshall, and Y. Gdalyahu, Classification with Nonmetric Distances: Image Retrieval and Class Representation, ***IEEE Trans. on Pattern Analysis and Machine Intelligence*** 22, pp. 583-600, 2000.

29. B. Li and R. Chellappa, Face Verification Through Tracking Facial Features, ***J. Optical Society of America*** A 18, pp. 2969-2981, 2001.

30. B. Li and R. Chellappa, A Generic Approach to Simultaneous Tracking and Verification in Video, ***IEEE Trans. on Image Processing*** 11, pp. 530-544, 2002.

31. B. Li, R. Chellappa, Q. Zheng, and S. Der, Model-Based Temporal Object Verification Using Video, ***IEEE Trans. on Image Processing*** 10, pp. 897-908, 2001.

32. Z. Liu, D. Jacobs, and R. Basri, Convexity in Perceptual Completion, in *Perceptual Organization for Artificial Vision Systems* (K. Boyer and S. Sarkar, eds.), Kluwer, 2002.

33. A. Mittal and L.S. Davis, M2Tracker: A Multi-View Approach to Segmenting and Tracking People in a Cluttered Scene, *International Journal of Computer Vision,* in press.

34. H. Moon, R. Chellappa, and A. Rosenfeld, Performance Analysis of a Simple Vehicle Detection Algorithm, *Image Vision Computing* 20, pp. 1-13, 2002.

35. H. Moon, R. Chellappa, and A. Rosenfeld, Optimal Edge-Based Shape Detection, *IEEE Trans. on Image Processing*, in press.

36. J. Neumann and Y. Aloimonos, Spatiotemporal Stereo Using Multi-resolution Subdivision Surfaces, *International Journal of Computer Vision* 47, pp. 181-193, 2002.

37. J. Neumann and C. Fermüller, Plenoptic Video Geometry, *The Visual Computer*, in press.

38. J. Neumann, C. Fermüller, and Y. Aloimonos, Animated Heads: From 3D Motion Fields to Action Descriptions, in *Deformable Avatars* (N. and D. Thalmann, eds.), pp. 1-12, Kluwer, Boston, 2001.

39. J. Neumann, H. Samet, and A. Soffer, Integration of Local and Global Shape Analysis for Logo Classification, *Pattern Recognition Letters* 23, pp. 1449-1457, 2002.

40. G. Qian, R. Chellappa, and Q. Zheng, Robust Structure from Motion Estimation Using Inertial Data, *J. Optical Society of America* A 18, pp. 2982-2997, 2001.

41. A.N. Rajagopalan and R. Chellappa, Higher-Order Statistics-Based Detection of Vehicles in Still Images, *J. Optical Society of America* A 18, pp. 3037-3048, 2001.

42. A. Rosenfeld, From Image Analysis to Computer Vision: An Annotated Bibliography, 1955-1979, *Computer Vision Image Understanding* 84, pp. 298-324, 2001.

43. A. Rosenfeld, Summation, in *Foundations of Image Understanding* (L.S. Davis, ed.), pp. 1-32, Kluwer, Boston, 2001.

44. A. Rosenfeld and A. Nakamura, Two Simply Connected Sets that Have the Same Area are IP-Equivalent, *Pattern Recognition* 35, pp. 537-541, 2002.

45. A. Rosenfeld, P.K. Saha, and A. Nakamura, Interchangeable Pairs of Pixels in Two-Valued Digital Images, *Pattern Recognition* 34, pp. 1853-1865, 2001.

46. P.K. Saha and A. Rosenfeld, Local and Global Topology Preservation in Locally Finite Sets of Tiles, *Information Sciences* 137, pp. 303-311, 2001.

47. H. Samet, Object Representations, in *Foundations of Image Understanding* (L. S. Davis, ed.), pp. 181-217, Kluwer, Boston, 2001.

48. C. Shin, D. Doermann, and A. Rosenfeld, Classification of Document Pages Using Structure-Based Features, *Intl. J. Document Analysis and Recognition* 3, pp. 232-247, 2001.

49. S. Sirohey and A. Rosenfeld, Eye Detection in a Face Image Using Linear and Nonlinear Filters, *Pattern Recognition* 34, pp. 1367-1391, 2001.

50. S. Sirohey, A. Rosenfeld, and Z. Duric, A Method of Detecting and Tracking Irises and Eyelids in Video, *Pattern Recognition* 35, pp. 1389-1401, 2002.

51. B. Stuart, P. Baker, and Y. Aloimonos, Towards the Ultimate Motion Capture, in *Deformable Avatars* (N. and D. Thalmann, eds.), pp. 1-12, Kluwer, Boston, 2001.

52. I. Weiss and M. Ray, Model-Based Recognition of 3D Objects from Single Images, *IEEE Trans. on Pattern Analysis and Machine Intelligence* 23, pp. 116-128, 2001.

53. W. Zhao and R. Chellappa, Symmetric Shape from Shading Using Self-Ratio Image, *Intl. J. Computer Vision* 45, pp. 55-75, 2001.

54. W. Zhao and R. Chellappa, Image-based Face Recognition: Issues and Methods, in *Image Recognition and Classification: Algorithms, Systems and Applications* (B.Javidi, ed.), Marcel Dekker, New York, 2002.

Conference Papers

1. M. Agrawal and L. Davis, A Probabilistic Framework for Surface Reconstruction from Multiple Images, *IEEE Conference on Computer Vision and Pattern Recognition*, 2001.

2. P. Baker, C. Fermüller, Y. Aloimonos, and R. Pless, A Spherical Eye from Multiple Cameras Makes Better Models of the World, *IEEE Conference on Computer Vision and Pattern Recognition*, pp. I:576-583, 2001.

3. R. Basri and D. Jacobs, Photometric Stereo with General, Unknown Lighting, *IEEE Conference on Computer Vision and Pattern Recognition*, pp. II:374-381, 2001.

4. R. Basri and D. Jacobs, Lambertian Reflectance and Linear Subspaces, *International Conference on Computer Vision*, pp. II:383-390, 2001.

5. C. BenAbdelkader, R. Cutler, and L. Davis, View-Invariant Estimation of Height and Stride for Gait Recognition, *Conference on Audio- and Video-Based Biometric Person Authentication*, 2002.

6. C. BenAbdelkader, R. Cutler, and L. Davis, Stride and Cadence as a Biometric in Automatic Person Identification and Verification, *IEEE International Conference on Automatic Face and Gesture Recognition*, 2002.

7. C. BenAbdelkader, R. Cutler, and L. Davis, Motion-based Recognition of People in EigenGait Space, *IEEE International Conference on Automatic Face and Gesture Recognition*, 2002.

8. C. BenAbdelkader, R. Cutler, H. Nanda, and L. Davis, EigenGait: Motion-based Recognition of People Using Image Self-Similarity, *Conference on Audio- and Video-Based Biometric Person Authentication*, 2001.

9. C. BenAbdelkader and L. Davis, Detection of People Carrying Objects: A Motion-based Recognition Approach, *IEEE International Conference on Automatic Face and Gesture Recognition*, 2002.

10. F. Brabec and H. Samet, Remote Thin-Client Access to Spatial Database Systems, *Conference on Digital Government Research*, pp. 75-82, 2002.

11. R. Chellappa, S. Zhou, and B. Li, Bayesian Methods for Face Recognition from Video, ***IEEE International Conference on Acoustics, Speech and Signal Processing***, pp. 4068-4071, 2002.

12. H. Chen, P. Belhumeur, and D. Jacobs, In Search of Illumination Invariants, ***IEEE Conference on Computer Vision and Pattern Recognition***, pp. 254-261, 2000.

13. A.R. Chowdhury and R. Chellappa, Robust Estimation of Depth and Motion Using Stochastic Approximation, ***International Conference on Image Processing***, pp. I:642-645, 2001.

14. A.R. Chowdhury and R. Chellappa, Towards a Criterion for Evaluating the Quality of 3D Reconstructions, ***IEEE International Conference on Acoustics, Speech and Signal Processing***, pp. 3321-3324, 2002.

15. A.R. Chowdhury, R. Chellappa, S. Krishnamurthy, and T. Vu, 3D Face Reconstruction from Video Using a Generic Model, ***International Conference on Multimedia***, pp. I:449-452, 2002.

16. P. David, D. DeMenthon, R. Duraiswami, and H. Samet, SoftPOSIT: Simultaneous Pose and Correspondence Determination, ***European Conference on Computer Vision***, pp. 698-714, 2002.

17. A. Elgammal and L. Davis, Probabilistic Framework for Segmenting People Under Occlusion, ***International Conference on Computer Vision***, pp. 145-152, 2001.

18. A. Elgammal, R. Duraiswami, and L.S. Davis, Efficient Kernel Density Estimation Using the Fast Gauss Transform with Applications to Segmentation and Tracking, ***International Workshop on Statistical and Computational Theories of Vision***, 2001.

19. A. Elgammal, R. Duraiswami, and L.S. Davis, Efficient Non-Parametric Adaptive Color Modeling Using Fast Gauss Transform, ***IEEE Conference on Computer Vision and Pattern Recognition***, 2001.

20. C. Fermüller, P. Baker, and Y. Aloimonos, Geometry and Statistics of Visual Space Time, ***British Machine Vision Conference***, 2002.

21. C. Fermüller, H. Malm, and Y. Aloimonos, Bias in Visual Motion Processes: A Theory Predicting Illusions, ***Workshop on Statistical Methods in Video Processing***, 2002.

22. C. Fermüller, H. Malm, and Y. Aloimonos, The Uncertainty in Localizing Image Features, ***Workshop on Statistical Methods in Video Processing***, 2002.

23. A. Folkers and H. Samet, Content-based Image Retrieval Using Fourier Descriptors on a Logo Database, I***nternational Conference on Pattern Recognition***, pp. III:521-524, 2002.

24. D.W. Jacobs, P. Belhumeur, and I. Jermyn, Judging Whether Multiple Silhouettes Can Come from the Same Object, I***nternational Workshop on Visual Form***, pp.532-541, 2001.

25. D. Jacobs, B. Rokers, A. Rudra, and Z. Liu, Fragment Completion in Humans and Machines, ***Neural Information Processing Systems***, 2001.

26. A. Kale, N. Cuntoor, and R. Chellappa, A Framework for Activity-Specific Human Identification, ***IEEE International Conference on Acoustics, Speech and Signal Processing***, pp. 3660-3663, 2002.

27. B. Li, R Chellappa, and H. Moon, Monte Carlo Simulation Techniques for Probabilistic Tracking, ***Conference Record of the Thirty-Fifth Asilomar Conference on Signals, Systems, and Computers***, pp. 75-82, 2001.

28. H. Liu, R. Chellappa, and A. Rosenfeld, Accurate Dense Optical Flow Estimation Using Adaptive Structure Tensors and a Parametric Model, ***International Conference on Pattern Recognition***, pp. I:291-294, 2002.

29. H. Liu, A. Rosenfeld, and R. Chellappa, Two-Frame Multi-Scale Optical Flow Estimation Using Wavelet Decomposition, ***IEEE International Conference on Acoustics, Speech and Signal Processing***, pp. 3588-3591, 2002.

30. A. Mittal and L.S. Davis, M2Tracker: A Multi-View Approach to Segmenting and Tracking People in a Cluttered Scene Using Region-Based Stereo, ***European Conference on Computer Vision***, 2002.

31. H. Moon, R. Chellappa, and A. Rosenfeld, 3D Object Tracking Using Shape-Encoded Particle Propagation, ***International Conference on Computer Vision***, pp. II:307-314, 2001.

32. H. Moon, R. Chellappa, and A. Rosenfeld, Tracking Human Activities Using Shape-Encoded Particle Propagation, ***International Conference on Image Processing***, pp. I:357-360, 2001.

33. J. Neumann and Y. Aloimonos, Spatiotemporal Analysis of Human Faces***, DAGM Conference***, pp. 61-68, 2001.

34. J. Neumann and Y. Aloimonos, Animated Heads Through Subdivision Surfaces, ***DAGM Conference***, 2001.

35. J. Neumann, C. Fermüller, and Y. Aloimonos, A Hierarchy of Cameras for 3D Photography, ***International Symposium on 3D Data Processing, Visualization and Transmission***, 2002.

36. J. Neumann, C. Fermüller, and Y. Aloimonos, A Hierarchy of Cameras for Identification Tasks, ***Workshop on Automatic Identification Technologies***, pp. 62-67, 2002.

37. J. Neumann, C. Fermüller, and Y. Aloimonos, Eyes from Eyes: Analysis of Camera Design Through the Plenoptic Function, ***Workshop on Omnidirectional Vision***, 2002.

38. J. Neumann, C. Fermüller, and Y. Aloimonos, Polydioptric Cameras: New Eyes for Structure from Motion, ***DAGM Conference***, 2002.

39. J. Neumann, H. Samet, and A. Soffer, Integration of Local and Global Shape Analysis for Logo Classification, ***Visual Form 2001*** (C. Arcelli, L. P. Cordella, and G. Sanniti di Baja, eds.), pp. 769-778, Springer, Berlin, 2001.

40. G. Qian and R. Chellappa, Structure from Motion Using Sequential Monte Carlo Methods, ***International Conference on Computer Vision***, pp. II:614-621, 2001.

41. G. Qian and R. Chellappa, Bayesian Self-Calibration of a Moving Camera, ***European Conference on Computer Vision***, pp. II:277-293, 2002.

42. G. Qian, R. Chellappa, and Q. Zheng, A Bayesian Approach to Simultaneous Motion Estimation of Multiple Independently Moving Objects, I*nternational Conference on Pattern Recognition*, WA I.9, 2002.

43. G. Qian, R. Chellappa, and Q. Zheng, Bayesian Structure From Motion Using Inertial Information, *IEEE International Conference on Image Processing*, pp. III:425-428, 2002.

44. H. Samet and A. Kochut, Octree Approximation and Compression Methods, *Data Processing Visualization and Transmission Conference*, pp. 460-469, 2002.

45. E. Tanin, F. Brabec, and H. Samet, Remote Access to Large Spatial Databases, *ACM Workshop on Geographic Information Systems*, pp. 5-10, 2002.

46. E. Tanin and H. Samet, APPOINT: An Approach for Peer-to-Peer Offloading the INTernet, *Conference on Digital Government Research*, pp. 99-105, 2002.

47. C. Yang, R. Duraiswami, and L.S. Davis, Superresolution Using Preconditioned Conjugate Gradient Method, *International Conference on Image and Graphics*, 2002.

48. C. Yang, R. Duraiswami, and L.S. Davis, Near-Optimal Regularization Parameters for Applications in Computer Vision, *International Conference on Pattern Recognition*, 2002.

49. S. Zhou and R. Chellappa, Probabilistic Human Recognition from Video, *European Conference on Computer Vision*, pp. 681-697, 2002.

50. S. Zhou and R. Chellappa, A Robust Algorithm for Probabilistic Human Recognition from Video, I*nternational Conference on Pattern Recognition*, pp. I:226-229, 2002.

51. S. Zhou, V. Krüger, and R.Chellappa, Face Recognition from Video: A Condensation Approach, *IEEE International Conference on Automatic Face and Gesture Recognition*, pp. 221-226, 2002.

Technical Reports

1. W. Zhao, R. Chellappa, A. Rosenfeld, and P.J. Phillips. Face Recognition: A Literature Survey. Technical Report: CAR-TR-948, CS-TR-4167R, August 2002.

2. Gisli R. Hjaltason and Hanan Samet. Incremental Similarity Search in Multimedia Databases. Technical Report: CAR-TR-956, CS-TR-4199, November 2000.

3. Daniel DeMenthon, Phil David and Hanan Samet. SoftPOSIT: An Algorithm for Registration of 3D Models to Noisy Perspective Images Combining Softassign and POSIT. Technical Report: CAR-TR-969, CS-TR-4257, August 2001.

4. Amit Roy Chowdhury and Rama Chellappa. Stochastic Approximation and Rate-Distortion Analysis for Robust Structure and Motion Estimation. Technical Report: CAR-TR-971, CS-TR-4261, June 2001.

5. Azriel Rosenfeld and Reinhard Klette. Digital Straightness. Technical Report: CAR-TR-970, CS-TR-4279, August 2001.

6. Amit Roy Chowdhury and Rama Chellappa. Statistical Bias and the Accuracy of 3D Reconstruction from Video. Technical Report: CAR-TR-972, CS-TR-4303, November 2001.

7. Jan Neumann, Cornelia Fermüller and Yiannis Aloimonos. Eyes from Eyes: Analysis of Camera Design Using Plenoptic Video Geometry. Technical Report: CAR-TR-973, CS-TR-4318, December 2001.

8. Ross G. Cutler, Ramani Duraiswami, J. Hench Qian and Larry S. Davis. Design and Implementation of the University of Maryland Keck Laboratory for the Analysis of Visual Movement. Technical Report: CS-TR-4329, UMIACS-TR-2002-11, 2002.

9. Philip David, Daniel DeMenthon, Ramani Duraiswami, and Hanan Samet. Evaluation of the SoftPOSIT Model-to-Image Registration Algorithm. Technical Report: CAR-TR-974, CS-TR-4340, UMIACS-TR-2002-22, July 2002.

10. Shaohua Zhou, Volker Krüger, and Rama Chellappa. Probabilistic Recognition of Human Faces from Video. Technical Report: CAR-TR-976, CS-TR-4359, UMIACS-TR-2002-39, April 2002.

11. Amit Roy Chowdhury and Rama Chellappa. Statistical Analysis of 3D Modeling From Monocular Video Streams. Technical Report: CAR-TR-977, CS-TR-4383, UMIACS-TR-2002-63, July 2002.

12. Omid David Tabibi and Nathan S. Netanyahu. Verified Null-Move Pruning. Technical Report: CAR-TR-980, CS-TR-4406, UMIACS-TR-2002-86, October 2002.

13. Patrick Baker and Yiannis Aloimonos. Rotations, Lines, and Multilinear Constraints. Technical Report: CAR-TR-981, CS-TR-4420, UMIACS-TR-2002-100, November 2002.

The Language and Media Processing Laboratory

David Doermann

Faculty
Rama Chellappa
Larry Davis
Bonnie Dorr
Doug Oard
Philip Resnik
Azriel Rosenfeld
Amy Weinberg

Visiting Faculty
Jaako Sauvola

Research Faculty
Daniel DeMenthon
Rebecca Hwa
Huiping Li

Support Staff
Denise Best

Founded in 1996 under a contract from the Department of Defense, the Laboratory for Language And Media Processing (LAMP) seeks to facilitate research and education in analysis and processing of multimedia information sources including documents, images and video. A contract was recently awarded to maintain LAMP's core objectives and to foster collaboration between language and media analysis researchers. The Laboratory is led by David Doermann, who directs research in document and video processing, and Amy Weinberg, who directs research in natural language processing. The Laboratory supports a variety of professional and infra-structural activities including:

- Basic and applied research in language processing, document analysis and video analysis
- Support of colloquia and workshops
- On-site collaboration with agency or affiliated personnel

The core document image and video analysis work remains focused on the problems surrounding access to large heterogeneous collections. In many environments, large static and dynamic collections of documents, images and videos are being gathered or created, yet these sources remain inaccessible without techniques to automatically index and retrieve the information that they contain. As technology moves us toward simplified creation and use of multimedia documents, we will see an even greater increase in the need to transmit, browse or otherwise process these collections efficiently.

Our recent efforts in document and video analysis have allowed us to develop an environment to effectively design and test new algorithms. We have developed a number of prototypes. Document images scanned from low-quality hard copies provide a challenging task for conventional systems, including page segmentation, labeling and optical character recognition (OCR). Likewise, video analysis efforts are focusing on the ability to provide genre classification of a series of detections in the textual, auditory and visual domains. As a natural extension to previous research, we are enhancing this environment to treat higher-level problems. The primary theme of the research is to provide automatic access to information sources by addressing issues involved in initial processing, organization, manipulation and retrieval.

Document Projects

Segmentation and Identification of Noise and Handwritten Annotations (see Figure 34)
The ability to fully segment a document into functionally different parts has been an ongoing goal of document analysis research. In the case where the content is presented in different fonts, or is handwritten as opposed to machine printed, different analysis algorithms may be required for interpretation. We have found annotations of particular interest in the processing of correspondence and related business documents.

This research focuses on the problem of segmenting and identifying handwritten annotations on noisy document images. In many types of documents such as correspondence, it is not uncommon for handwritten annotations to be added as part of a note, correction, clarification, or instruction, or for initials or a signature to appear as an authentication mark. It is important to be able to segment and identify such handwriting so we can 1) locate, interpret and retrieve it efficiently in large document databases, and 2) use different algorithms for printed/handwritten text or signature recognition.

A common approach to identifying handwriting consists of two processes: 1) a segmentation process which divides the text into regions at an appropriate level (word, line or zone) and 2) a classification process which identifies segmented regions as handwritten. The challenge is that in extremely noisy environments, noise often mixes with text, resulting in unreliable identification. Rather than removing noise simply by filtering, we treat noise as a distinguished class and model noise based on selected features. The classified noisy regions are further refined by a Markov Random Field based post-process which uses the contextual information to enhance the classification result. Experiments show that our approach can identify handwriting in extremely noisy documents, and improve general page segmentation results.

Bilingual Dictionary Access (see Figure 35)
A collaborative group of researchers at the University of Maryland and Johns Hopkins University are working to develop and demonstrate an integrated set of techniques for building rapidly retargetable systems to support translingual access to character-coded electronic texts, images of printed documents, and recorded news broadcasts. Our techniques enable the creation of working systems for previously unforeseen language requirements within one to two weeks using readily available resources. We are demonstrating a useful capability to search document images in a new language by users who are not familiar with that language, based on a limited quantity of print or electronic resources and the use of a language specialist for a brief time.

Our efforts center around the ability to 1) rapidly acquire lexical information from multilingual dictionaries, and 2) use parallel text from the Bible to bootstrap and improve OCR in low-density languages. We are developing a configurable dictionary parser that provides the ability to physically segment a scanned image of a dictionary page, and then use user-guided knowledge of the structure of dictionary definitions to extract a term list. The term list can then be used with existing automatic and semi-automatic translingual detection techniques to facilitate cross-language IR and machine translation.

Multilingual OCR Evaluation
Characterizing the performance of OCR systems is crucial for monitoring technical progress, predicting OCR performance, providing scientific explanations for system behavior, and identifying open problems. We are addressing issues of evaluation as well as problems associated with generating appropriate "ground truth" to support evaluation.

In multilingual environments, OCR has emerged as an important information technology, thanks to the increasing need for cross-language information access. While many research groups and companies have developed OCR algorithms for various languages, it is difficult to compare the performances of these OCR algorithms across languages. In order to bridge the gap, we are beginning to look at alternative sources of parallel text, such as Bibles, WWW pages (through STRAND, developed by Phil Resnik at UMD), and multilingual dictionaries, as datasets for comparing OCR accuracies across languages. Recently, we have done evaluations in Arabic, Chinese, and French, and have developed a set of tools that allow us to rapidly migrate to other languages. An ongoing focus is on the evaluation of follow-on processes with respect to OCR, including entity extraction, information retrieval, and machine translation. We have developed capabilities for generating ground truth from raw text in any font supported by MS-Windows, using generated documents, in conjunction with degradation modeling, to extend evaluation capabilities.

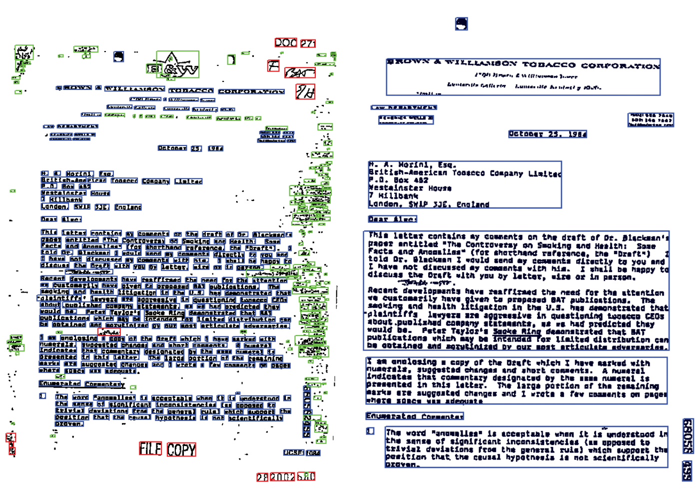


Figure 34. Original and segmented documents.

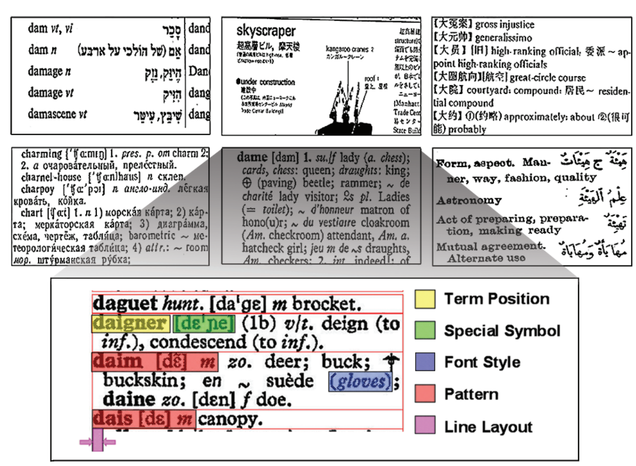
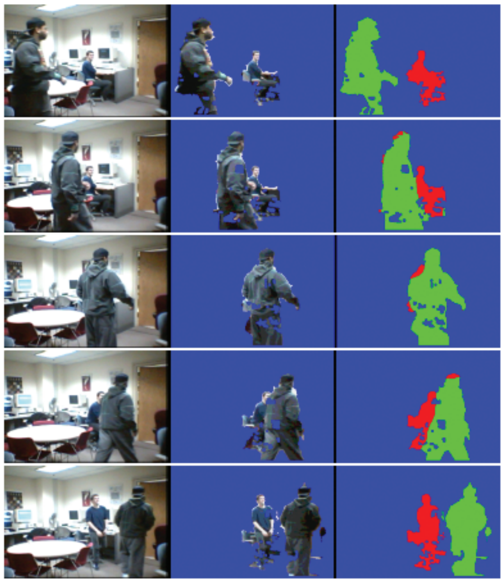


Figure 35. Segmentation and labeling of bilingual dictionaries.



People segmentation under occlusion

Figure 36. Segmentation Under Occlusion.

In additional to multilingual OCR evaluation, we have developed a toolset for page segmentation evaluation (PSET) and tools for ground truthing in multilingual character sets (TrueViz and ViPER).

Video Projects

Genre-Based Video Classification

The recent advances in computing power and network transmission have rekindled interest in how to index, browse and manipulate digital video efficiently. Manual annotation of video is expensive and tedious and will not be able to keep up with the rapidly increasing volume of video data. In our work we are addressing the problem of automatically classifying video sequences into different genres such as sports, news, commercials, or cartoons, as well as specific classes such as types of sports. If the video genre is correctly identified, content-based retrieval can be added to classical text content queries. For example, the user can input "news" to get all news video clips.

To achieve this goal, we extract features at both the frame and shot levels. Special attention is paid to content descriptors including faces, text blocks, indoor/outdoor classification, landscape, and speech. Low-level features, such as motion, color and edges, are also explored to assist the classification. Our focus is on the combination of these features to robustly identify each video type.

We have developed an extension of this system and are participating in the 2002 TREC-Video track. The goal is to integrate the results of various feature detectors with text-based queries from video OCR and automatic speech recognition, and visual queries from sample images and video clips.

Human Activity Analysis (see Figure 36)

The ability to detect, track and recognize people and vehicles and to observe how they interact with each other and their environment in video captured, compressed and transmitted under a wide variety of adverse conditions is a task fundamental to general content-based video surveillance systems or content filtering systems.

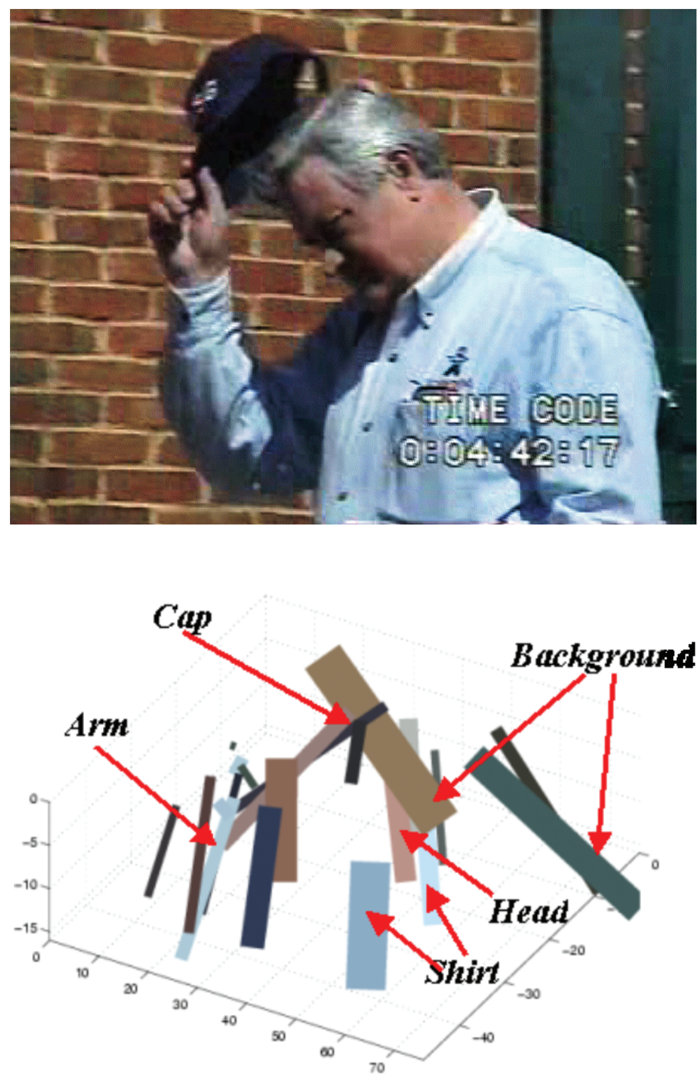


Figure 37. Space-time Descriptors.


Best-Case Results for Nearest-Neighbor Learning

Best-Case Results for Nearest-Neighbor Learning

Figure 38. Model and labeling result visualization.

We are developing an integrated system of algorithms for detecting, tracking, recognizing and filtering human activity in video streams, which we call ADVICE: Activity Detection by VIdeo Content Evaluation. Our integrated effort extends previous work on human activity and addresses the following core issues:

- Detection and tracking of people viewed by a passive or active camera.
- Integrated identification and tracking of people.
- Modeling, classification, and recognition of human activities.
- Content-based filtering.
- Evaluation.

This project is enabling technologies that facilitate content-based filtering of humans and human activities in two complementary information sources: surveillance video and telecommunication video. Our work attempts to overcome obstacles that inhibit the application of algorithms for human and human activity detection and recognition to video whose quality has been degraded by environmental, capture, and transmission effects. Furthermore, we are providing a language and mechanism by which researchers can define filtering scenarios interactively via the WWW, run them on either our data or submitted samples, and view results via streaming media.

Space-Time Video Representations

Simple descriptors often fail to adequately describe the visual content of videos, yet for the past decade, researchers have concentrated on extracting low level “features” from individual frames of videos, such as color, texture and shape. We are taking a different approach, based on finding features that characterize whole video sequences. Such features are extracted by analysis of whole stacks of video frames, which compose a space-time (ST) volume of pixels. Pixels with similar color and motion characteristics in this volume are grouped by an operation called space-time segmentation.

To visualize the segmentation, we can represent the segmented regions in a 3D X-Y-Time space. Slanted space-time regions are characteristic of motion. The labeling of regions consistent from frame to frame amounts to tracking the regions, and analyzing video frames as a block makes region tracking simpler than when it is based on analyzing individual frames. Each space-time region can be simplified as a straight-line segment with the average color of the region. We call such lines video strands. Figure 37 shows the video strands in the space-time volume that shows a man taking off his cap. The space-time segmentation and video strands are computed by a clustering technique called mean shift analysis, and matched to precomputed models. These and related techniques are being applied to the problem of identifying duplicate sequences in video, recognizing specific activities in surveillance footage, and indexing video by example.

Other recent and ongoing projects of interest include:

- Logical Document Analysis and Classification
- Pervasive Networking for Multimedia - providing basic media access and analysis capabilities through mobile devices
- Remote Video Access - integrating basic video surveillance and analysis capabilities for access through low-bandwidth and wireless channels
- MALACH - Multilingual Access to Large Spoken Archives
- ViPER - Video Performance Evaluation Resource
- Silicon Wafer Defect Inspection.

Additional details of selected projects can be found on our www pages: http://lamp.cfar.umd.edu. A logical document analysis example is shown in Figure 38.

Contracts and Grants

Sponsor	Dates	Amount	Title
DOD	09/30/96 - 12/31/01	3,493,054	Language and Media Processing Laboratory
Navy	01/04/00 - 04/03/05	391,000	Translingual Information Access
Mitre Corp.	11/20/01 - 01/30/03	110,997	Evaluation and Improvement of MT Using Parallel Corpora
Hitachi Ltd.	10/01/01 - 09/20/02	81,082	Video Frame Ranking Toolbox
DOD	11/04/99 - 09/30/01	75,000	Quantitative Performance Evaluation of OCR-Based Systems
Philips	06/01/00 - 09/30/01	55,000	Intelligent Set-Top Processing
Panasonic	04/01/01 - 03/31/02	35,000	Automated Modeling and Analysis of Structured Documents

Journal Papers

1. K. Guo, D. Doermann, and A. Rosenfeld. Forgery Detection by Local Correspondence, ***International Journal of Pattern Recognition and Artificial Intelligence*** 15, pp. 579-641, 2001.

2. S. Mao and T. Kanungo, Empirical Performance Evaluation of Page Segmentation Algorithms: Methdology and Application, ***IEEE Trans. on Pattern Analysis and Machine Intelligence*** 23, pp. 242-256, 2001.

Conference Papers

1. K. Darwish, D. Doermann, R. Jones, D. Oard, and M. Rautiainen, TREC-10 Experiments at Maryland: CLIR and Video, ***TREC***, 2001.

2. P. David, D. DeMenthon, and R. Duraiswami, SoftPOSIT: Simultaneous Pose and Correspondence Determination, ***European Conference on Computer Vision***, 2002.

3. D. Doermann, H. Ma, B. Karagol-Ayan, and D. Oard, Lexicon Acquisition from Bilingual Dictionaries, ***SPIE Photonics West Electronic Imaging Conference***, pp. 37-48, 2002.

4. P. Gupta, D. Doermann, and D. DeMenthon, Beam Search for Feature Selection in Automatic SVM Defect Classification, ***International Conference on Pattern Recognition***, 2002.

5. P. Gupta, D. Doermann, and D. DeMenthon, Support Vector Machines for Automatic Defect Classification, ***International Conference on Pattern Recognition***, 2002.

6. T. Kanungo, C. Lee, J. Czorapinski, and I. Bella, TRUEVIZ: A GroundTRUth/Metadata Editing and VIsualiZing Toolkit for OCR, ***SPIE Conference on Document Recognition and Retrieval,*** 2001.

7. H. Li and D. Doermann, Text Quality Estimation in Digital Video, ***SPIE Conference on Document Recognition and Information Retrieval***, pp. 232-243, 2002.

8. V.Y. Mariano, J. Min, J.-H. Park, R. Kasturi, D. Mihalcik, D. Doermann, and T. Drayer, Performance Evaluation of Object Detection Algorithms, ***International Conference on Pattern Recognition***, 2002.

9. Y. Qi, D. Doermann, and D. DeMenthon, Hybrid Independent Component Analysis and Support Vector Machine Learning Scheme for Face Detection, ***International Conference on Acoustics, Speech, and Signal Processing***, 2001.

10. M. Rautiainen and D. Doermann, Temporal Color Correlograms in Video Retrieval, ***International Conference on Pattern Recognition***, 2002.

11. C. Wolf and D. Doermann, Binarization of Low Quality Text Using a Markov Random Field Model, ***International Conference on Pattern Recognition***, 2002.

Technical Reports

1. Rebecca Green, Lisa Pearl, and Bonnie J. Dorr. Mapping Lexical Entries in a Verbs Database to WordNet Senses. Technical Report: LAMP-TR-068, CS-TR-4230/UMIACS-TR-2001-18, MDA-9049-6C-1250, 2001.

2. Rebecca Green, Lisa Pearl, Bonnie J. Dorr, and Philip Resnik. Lexical Resource Integration Across the Syntax-Semantics Interface. Technical Report: LAMP-TR-069, CS-TR-4231, UMIACS-TR-2001-19, MDA-9049-6C-1250, 2001.

3. Mari Olsen, David Traum, Carol Van Ess-Dykema, and Amy Weinberg. Implicit Cues for Explicit Generation: Using Telicity as a Cue for Tense Structure in a Chinese to English MT System. Technical Report: LAMP-TR-070, CS-TR-4248, UMIACS-TR-2001-33, MDA-9049-6C-1250, 2001.

4. Noah A. Smith. Detection of Translational Equivalence. Technical Report: LAMP-TR-071, CS-TR-4253, UMIACS-TR-2001-36, MDA-9049-6C-1250, 2001.

5. Rebecca Hwa. On Minimizing Training Corpus for Parser Acquisition. Technical Report: LAMP-TR-073, CS-TR-4258, UMIACS-TR-2001-40, IRI-9712068, N6600197-C-8540, 2001.

6. Nizar Habash, Bonnie Dorr, and David Traum. Efficient Language Independent Generation from Lexical Conceptual Structure. Technical Report: LAMP-TR-074, CS-TR-4262, UMIACS-TR-2001-43, MDA-904-96-C-1250, 2001.

7. Nizar Habash and Bonnie Dorr. Large Scale Language Independent Generation Using Thematic Hierarchies. Technical Report: LAMP-TR-075, CS-TR-4280, UMIACS-TR-2001-59, MDA-904-96-C-1250, 2001.

8. Nizar Habash. Nuun: A System for Developing Platform and Browser Independent Arabic Web Applications. Technical Report: LAMP-TR-076, CS-TR-4281, UMIACS-TR-2001-60, MDA-904-96-C-1250, 2001.

9. David Traum and Nizar Habash. Generation from Lexical Conceptual Structures. Technical Report: LAMP-TR-077, CS-TR-4282, UMIACS-TR-2001-61, MDA-904-96-C-1250, 2001.

10. Clare Cabezas, Bonnie Dorr, and Philip Resnik. Spanish Language Processing at University of Maryland: Building Infrastructure for Multilingual Applications. Technical Report: LAMP-TR-078, CS-TR-4285, UMIACS-TR-2001-64, MDA-904-96-C-1250, 2001.

11. Nizar Habash. A Reference Manual to the Linearization Engine oxyGen Version 1.6. Technical Report: LAMP-TR-079, CS-TR-4295, UMIACS-TR-2001-73, MDA-904-96-C-1250, 2001.

12. Necip Fazil Ayan and Bonnie J. Dorr. Generating a Parsing Lexicon from Lexical-Conceptual Structure. Technical Report: LAMP-TR-081, CS-TR-4332, UMIACS-TR-2002-14, MDA-904-96-C-1250, 2002.

13. Bonnie J. Dorr, Lisa Pearl, Rebecca Hwa, and Nizar Habash. Improved Word-Level Alignment: Injecting Knowledge about MT Divergences. Technical Report: LAMP-TR-082, CS-TR-4333, UMIACS-TR-2002-15, MDA-904-96-C-1250, 2002.

14. Nizar Habash and Bonnie Dorr. Handling Translation Divergences in Generation-Heavy Hybrid Machine Translation. Technical Report: LAMP-TR-083, CS-TR-4341, UMIACS-TR-2002-23, MDA-904-96-C-1250, 2002.

15. Necip Fazil Ayan and Bonnie Dorr. Creating Parsing Lexicons from Semantic Lexicons Automatically and its Applications. Technical Report: LAMP-TR-084, CS-TR-4352, UMIACS-TR-2002-32, MDA-904-96-C-1250, 2002.

16. Adam Lopez, Michael Nossal, Rebecca Hwa, and Philip Resnik. Word-level Alignment for Mutlilingual Resource Acquisition. Technical Report: LAMP-TR-085, CS-TR-4354, UMIACS-TR-2002-34, FCPO.810548265, RD-02-5700, 010418-7712, 2002.

17. Rebecca Hwa, Philip Resnik, and Amy Weinberg. Breaking the Resource Bottleneck for Multilingual Parsing. Technical Report: LAMP-TR-086, CS-TR-4355, UMIACS-TR-2002-35, FCPO.810548265, RD-02-5700, N60010028910, 010418-7712, 2002.

18. Puneet Gupta. Automatic Localization and Classification of Silicon Wafer Defects. Technical Report: LAMP-TR-087, CS-TR-4356, CAR-TR-975, 2002.

19. Nizar Habash and Bonnie Dorr. Handling Translation Divergences: Combining Statistical and Symbolic Techniques in Generation-Heavy Machine Translation. Technical Report: LAMP-TR-088, CS-TR-4369, UMIACS-TR-2002-49, FCPO.810548265, 010418-7712, 2002.

20. Philip Resnik and Noah A. Smith. The Web as a Parallel Corpus. Technical Report: LAMP-TR-089, CS-TR-4381, UMIACS-TR-2002-61, 2002.

21. Daniel DeMenthon and Remi Megret. Spatio-Temporal Segmentation of Video by Hierarchical Mean Shift Analysis. Technical Report: LAMP-TR-090, CS-TR-4388, UMIACS-TR-2002-68, CAR-TR-978, 2002.

22. Bonnie Dorr and Terry Gaasterland. Constraints on the Generation of Tense, Aspect, and Connecting Words from Temporal Expressions. Technical Report: LAMP-TR-091, CS-TR-4391, UMIACS-TR-2002-71, 2002.

23. Bonnie Dorr and Tiejun Zhao. Domain-Specific Term-List Expansion Using Existing Linguistic Resources. Technical Report: LAMP-TR-092, CS-TR-4399, UMIACS-TR-2002-81, 2002.

24. Bonnie Dorr and Dekang Lin. Construction of a Chinese-English Verb Lexicon for Embedded Machine Translation in Cross-Language Information Retrieval. Technical Report: LAMP-TR-093, CS-TR-4400, UMIACS-TR-2002-80, 2002.

25. Remi Megret and Daniel DeMenthon. A Survey of Spatio-Temporal Grouping Techniques. Technical Report: LAMP-TR-094, CS-TR-4403, UMIACS-TR-2202-83, CAR-TR-979, 2002.

The Graphics and Visual Informatics Laboratory

Amitabh Varshney

The University of Maryland's Graphics and Visual Informatics Laboratory (GVIL) aims to promote research and education in computer graphics, scientific visualization, and virtual environments. The goal of GVIL is to improve the efficiency and usability of visual computing applications in science, engineering, and medicine. The scope of GVIL's research covers design of algorithms and data structures for reconciling realism and interactivity for very large graphics datasets, rapid access to distributed graphics datasets across memory and network hierarchies, and study of the influence of heterogeneous display and rendering devices over the visual computing pipeline. GVIL activities involve development of visual computing tools and technologies and tools to support the following driving applications: protein folding and rational drug design, navigation and interaction with mechanical CAD datasets, and ubiquitous access to distributed three-dimensional graphics datasets. GVIL is directed by Professor A. Varshney

Facilities

GVIL has a wall-sized multi-projector responsive display, several Dell workstations with 3D graphics cards, and printers. GVIL is equipped with a video-editing suite which includes a Sony digital camcorder, ability to capture and edit digital and analog video, and real-time 3D graphics rendering. GVIL also has 3D graphics modelbanks and software for 3D modeling, rendering, and animation, and is on a gigabit ethernet.

Research Overview

Research in GVIL deals with a range of issues in visual computing, ranging from the applied to the theoretical. Our research is driven by the applications of computer-aided Proteomics and computer-aided engineering design. In the following sections we first discuss our driving applications and then give an overview of our ongoing research in software systems and algorithms.

Molecular Graphics

Protein folding and rational drug design are two representative Grand Challenge problems in the study of biological molecules that are hard to model, difficult to analyze, and computationally intensive. Virtual environments offer a powerful interaction environment for exploring such datasets in real time, enabling superior insights into the underlying biochemical processes. An important goal of such virtual environments is to provide a high-bandwidth human-computer interface to convey the rich multi-dimensional information space. The recent successes in human genome sequencing have taken us a step closer to the goal of designing novel therapeutic drugs. One of the goals of GVIL is to develop visual informatics tools and technologies that will give scientists deeper insights into understanding the relationships between form and function in biological proteins.

Molecular Electrostatics

Electrostatic interactions are of central importance in many biological processes. Experiments have shown that electrostatics influences nearly all biochemical reactions, including macromolecular folding and conformational stability. Electrostatics also determines the structural and functional properties of biological samples, such as their shapes, binding energies, and association rates. Electrostatic interactions are one of two kinds of long-range biological interactions; the other being van der Waals interactions, which fall off much faster in space. Due to the central role played by electrostatic interactions, their successful modeling has great practical importance in rational drug design and protein folding.

Quantum-mechanical methods and classical electrostatics are two possible ways to model the electrostatic properties of biological samples. Quantum-mechanical methods, although more accurate, demand extraordinary computation, and thus at present can only be applied to very small molecules. Classical electrostatics is more computationally tractable and is valid across large scales of biological interest, excluding, of course, quantum-mechanical scales. Classical electrostatics models the electrostatic interactions of bio-molecules as the interactions between partial atomic charges (also called net atomic charges). The electrostatics of a molecule depends not only on its 3D structure and charge distribution, but also on its environment and solvent. To limit the computational cost, the solvent is always treated as a continuum with average properties, while in the most important part, the solute, molecules are

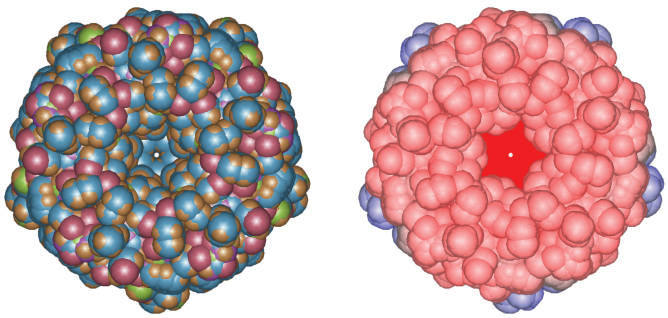


Figure 39. A mechanosensitive channel in the E. Coli membrane:(a) Van der Waals radii atoms; (b) Colored by the electrostatic potential.

described in atomic detail. Electrostatic interactions in solution can then be described by the non-linearized Poisson-Boltzmann equation (PBE).

We are developing a new method of efficiently computing and displaying electrostatic potentials by explicitly generating and incorporating the solvent interface This allows us to localize the discontinuities faster and more accurately. Figure 39 shows the surface of a single mechano-sensitive channel in the membrance of the E-coli bacterium and the computed electrostatic potential on the surface, where red represents the negative potential and blue represents the positive potential.

Shape Complementarity

Many drug development processes have begun with large-scale random screening of candidate inhibitors. These initial discoveries are improved to find new drugs. As molecular structure determination techniques and computational methods progress, protein-docking methods using structure-based molecular complementarity have become a feasible substitute for random screening in the drug design process.

Among many factors involved in protein-protein interactions, such as electrostatics, hydrophobicity, and hydrogen bonding, shape complementarity is of major concern. A complete search of all possible geometric fits of two flexible molecules takes too much time because of the extremely large number of degrees of freedom. The goal of our ongoing research in shape complementarity is to develop fast and reliable methods for finding docking sites and corresponding transformations to align the two molecules into complementary conformations (see Figure 40).

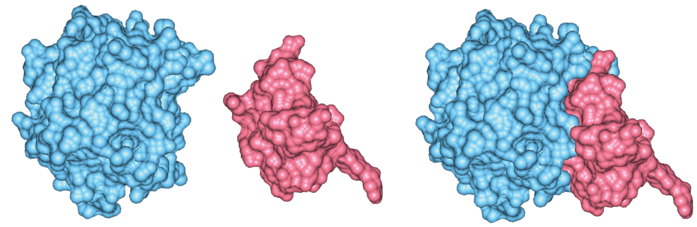


Figure 40. Chymotrypsinogen complex with human pancreatic secretory trypsin inhibitor.

Interactive Subsurface Scattering Using Local Illumination

Photo-realistic image synthesis remains one of the primary goals of graphics. For this it is necessary to model the interaction of light with objects in a physically correct manner. Several illumination models have been developed for image synthesis. Most of them model the bi-directional reflectance distribution function (BRDF). A good BRDF model, either derived or measured, can give highly realistic visual effects. However, the basic assumption of BRDF models, that light enters and exits the object surface at the same point, is sometimes not valid. For example, when subsurface scattering is involved the light enters the object at one point and exits at another. BRDF models are inadequate to simulate the appearance of materials with high subsurface scattering. Thus, BRDF models are only approximations to the more general bi-directional surface scattering reflectance distribution function (BSSRDF).

Subsurface scattering for single scattering events has been modeled in layered surfaces as analytical expressions in terms of one-dimensional linear transport theory. For multiple scattering events, a dipole point source diffusion approximation is used. The shortcoming of these approaches is that they require a global illumination solution, i.e., ray tracing, which is very slow and generally non-interactive. We have developed a simple lighting model to incorporate subsurface scattering effects within the local illumination framework. By observing that subsurface scattering is relatively local, we make approximations to the BSSRDF model and model the subsurface scattering effects by using only local illumination. Our model is able to capture the most important features of subsurface scattering: reflection and transmission due to single or multiple scattering.

Figure 41. Increasing the subsurface scattering (from left to right) in the Venus model.

We have modified the traditional local illumination model into a two-stage process. In the first stage we compute the reflection and transmission of light on the layered surfaces. The second stage involves bleeding the scattering effects to produce the final result. Using this approach, we can achieve interactive frame rates with an order of magnitude speedup compared with subsurface scattering simulated by the state-of-the-art global ray tracing method. Our results for the Venus model appear in Figure 41.

Point-based Rendering

As the complexity of graphics datasets continues to increase beyond display complexity, image-based and point-based rendering is beginning to emerge as a viable alternative to conventional triangle-based rendering. In the recent past we have explored real-time hardware-assisted image-based rendering. Given a collection of depth images representing an environment from fixed viewpoints and view directions, our approach first constructs an image-space simplification of the scene as a pre-process, and then reconstructs a view of the scene from arbitrary viewpoints and directions in real time. We achieve speed through the use of commonly available texture-mapping hardware and partially rectify the visibility gaps in a best-effort manner through morphing.

More recently we have explored point-based rendering as an alternative to image-based and triangle-based rendering. Points offer many benefits over image-based and triangle-based models: (1) they are more efficient at modeling and rendering complex environments, (2) they are amenable to being organized by a seamless hierarchical structure which can be used to efficiently control frame-rates and visual quality, and (3) their zero-connectivity characteristics can be used for efficient streaming for remote rendering. Current point primitives store only limited information about their immediate locality, such as normal, sphere of influence and disk of influence on the tangent plane. These primitives are then rasterized with flat shading and in some cases followed up with a screen-space filtering. Since the primitives are flat-shaded, such representations require very high sampling to obtain good rendering quality. In other words, the rendering algorithm dictates the sampling frequency at the modeling stage. This is clearly undesirable, as it may prescribe very high sampling even in areas of low spatial frequency, causing two significant drawbacks: (1) slower rendering due to increase in rendering computation and related CPU-memory bus activity, and (2) large disk and memory utilization.

In our recent research we have proposed using a novel rendering primitive that combines the modeling brevity of points with the rasterization efficiency of polygons. The surface is represented by a sampled collection of differential points, each with embedded curvature information that captures the local differential geometry in the vicinity of that point. This information gives a good approximation of the surface distribution in the vicinity of each sampled point, which is then used for rendering the point and its approximated vicinity. The total surface area that a point is allowed to approximate is bounded by the characteristics of the surface at that point. This scheme offers the potential of controlling the local sampling density based on surface curvature characteristics and offers more efficiency than rasterization of planar polygons. We use it to directly rasterize the local shape represented in each differential point. We further do screen space blending of these primitives to obtain superior consistency of the rendered surface. This offers other benefits such as texture mapping and environment mapping of the surface, as shown in Figure 42.

Progressive, View-Dependent Transmission of Point Data

Advances in 3D digitization have resulted in the proliferation of 3D models that are available for download on the Internet. As 3D digital libraries, virtual shopping malls, and online 3D games become more widespread on the Internet, the ability to download and view 3D graphical models over low-bandwidth networks becomes critical.

Compressed 3D models are more suitable for transmission over low-bandwidth channels than their uncompressed counterparts. Compression of triangulated 3D models has been an active area of research among computer graphics researchers since 1995 and

Figure 42. Interactive point-based rendering with (a) Refraction mapping for the Stanford bunny, and (b) Reflection mapping for a human head placed at the University of Maryland's 'M' traffic circle.

significant strides have been made. Work has also been done on progressive compression of 3D models. The key idea is to construct a sequence of decimation operations (such as edge collapse or vertex removal) to simplify an input triangle mesh progressively to a base mesh. The base mesh coupled with the decimation sequence in reverse give a multi-resolution hierarchy of meshes, each more accurate than its ancestors in the hierarchy. Efficient encoding of the base mesh and decimation operations gives a progressive scheme for transmitting large 3D models. Such a scheme allows the user to receive a coarse mesh followed by refinements to it, thereby avoiding long delays in obtaining a first view of the model.

In our approach, progressive transmission of such data proceeds by selectively transmitting the refinements from the server based on the viewing and lighting parameters at the client. The goal is to send the refinements in order of decreasing visually perceptible changes. For example, refinements to regions of the scene that are far from the viewer's location are less likely to result in major visually perceptible changes than those near in a perspectively viewed scene. Our work focuses on building a system that can adaptively determine which refinements to stream based on the viewing and lighting parameters. We have implemented a system for the progressive, view-dependent streaming of a 3D model of points from a server to a client over a network. Our system builds a hierarchy of the point data and uses the view parameters to choose, for each region of the scene, the appropriate level of the hierarchy to transmit.

Tiled Wall Displays

Recent growth of interest in virtual environments has been accompanied by a corresponding increase in the types of devices for sensory feedback, especially visual feedback. At present, the various display devices can be broadly classified into three main categories: (a) head-coupled displays, including head-mounted displays, (b) spatially-immersive displays, including CAVE and tiled displays, and (c) virtual model displays, including immersive and responsive workbenches. Our target application is conformational steering that multiple users can collaboratively use. The head-coupled displays are ill-suited for tasks that take several hours (due to eye and neck fatigue). Similarly, the virtual model displays are not suited for applications that are rich in visual information content and require a high resolution display or require a wide field of view and large area. Spatially-immersive displays, such as wall-sized tiled displays, allow long periods of work, offer a very high field of view and resolution, and afford a strong self-presence. Several groups have reported that wall-sized tiled displays are more supportive to collaboration and learning than regular monitors. Users stay longer in such displays, move and discuss the datasets more, and treat such displays as 'murals' that they repeatedly touch, inspect, walk around and see from different viewpoints.

We have built a tiled-display system that achieves geometric alignment for 3D graphics applications by pre-warping 3D objects on the fly. Our system consists of commodity off-the-shelf components. As shown in Figure 43, it includes one Sony DCR VX2000 digital video camera, one Da-lite portable wall-sized (6'x8') display screen, four Proxima Ultralight X350 DLP projectors arranged in a 2x2 array behind the display wall, and one dual-processor 933MHz Dell Precision 620 workstation installed with two Appian Jeronimo 2000 video cards. Each video card has two graphics adapters with 3D hardware acceleration.

Our prototype system achieves real-time 3D geometric alignment with hardware acceleration for each graphics pipeline. It accomplishes this by pre-warping 3D objects

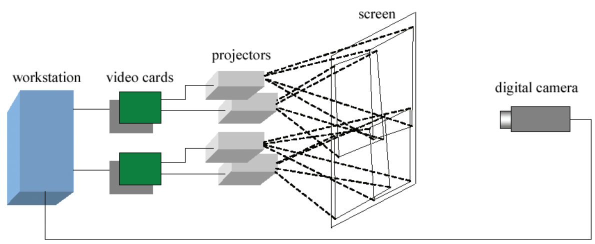


Figure 43. High-level system block diagram of the Maryland Display Wall.

at every frame followed by conventional 3D graphics rendering. The pre-warping transformations are shear transformations obtained by segmenting the projector footprints using a digital camcorder. After the geometric calibration, one can see the crisp seams between the projected areas. We use varying-opacity feathering polygons to remove the seams between overlapping regions. The cost of feathering is acceptable compared with hardware-blending solutions. We use ultrasonic technology (Mimio electronic whiteboard system) to track user gestures. A user holding a marker can walk up to the display wall, select a part of the protein sidechain and rotate it. The complete system is shown in Figures 44 and 45.

Figure 44. A system snapshot behind the display wall.

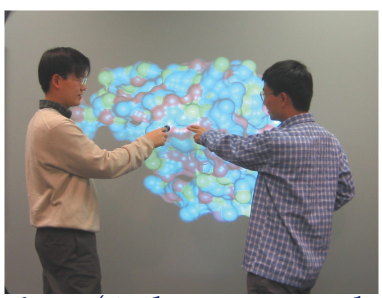


Figure 45. Chang Ha Lee and Zhiyun Li discussing the user interface for displaying the HIV protease on the Maryland Display Wall.

Contracts and Grants

Sponsor	Dates	Amount	Title
NSF	9/1/00 - 8/31/03	450,000	Visualization and Interaction with Large Graphics Datasets over Networks
NSF	1/1/99 - 12/31/02	262,609	Multimodal Interaction with Biological Molecules in Virtual Environments

Journal Papers and Book Chapters

1. C. Chang, Z. Li, A. Varshney and Q. J. Ge, Hierarchical Image-Based and Polygon-Based Rendering for Large-Scale Visualizations, in ***Hierarchical and Geometrical Methods in Scientific Visualization*** (H. Hagen, G. Farin, and B. Hamann, eds.), Springer-Verlag, in press.

2. A. Kalaiah and A. Varshney, Differential Point Rendering, in ***Rendering Techniques*** 01, Springer-Verlag, pp. 139-150, 2001.

3. A. Kalaiah and A. Varshney, Modeling and Rendering of Points with Local Geometry, ***IEEE Trans. on Visualization and Computer Graphics***, in press.

Conference Papers

1. C. Chang, A. Varshney, and Q. J. Ge, Haptic and Aural Rendering of a Virtual Milling Process, ***ASME Design for Manufacturing Conference***, Pittsburgh, PA, 2001.

2. C. Lee and A. Varshney, Representing Thermal Vibrations and Uncertainty in Molecular Surfaces, ***SPIE Conference on Visualization and Data Analysis***, San Jose, CA, 2002.

3. S. T. Li, Q. J. Ge, and A. Varshney, Web-based Interactive Design of Freeform Motion, ***ASME Design Automation Conference***, Pittsburgh, PA, 2001.

4. Z. Li and A. Varshney, A Real-time Seamless Tiled Display for 3D Graphics, ***Seventh Annual Symposium on Immersive Projection Technology***, Orlando, FL, 2002.

Books, Monographs, and Edited Collections

1. B. Chen, A. Kaufman, K. Mueller, and A. Varshney, State of the Art in Data Representation, IEEE Visualization Course Notes, October 2002.

2. J. Cohen, D. Luebke, M. Reddy, A. Varshney and B. Watson, Advanced Issues in Level of Detail, ACM SIGGRAPH Course Notes, August 2001, August 2002.

3. T. Ertl, K. Joy, and A. Varshney, eds., ***IEEE Visualization 2001 Proceedings,*** IEEE Computer Society, ISBN 0-7803-7200-X, October 2001.

4. D. Luebke, M. Reddy, J. Cohen, A. Varshney, B. Watson, and R. Huebner, ***Level of Detail for 3D Graphics***, Morgan-Kaufman, San Francisco, ISBN 1-55860-838-9, 2002.

5. A. Varshney, guest ed., Special Issue on Visualization, ***IEEE Trans. on Visualization and Computer Graphics***, Vol. 7, No. 3, July - September, 2001.

6. A. Varshney and Z. Pan, guest eds., Special Issue on Virtual Reality, ***International Journal of Image and Graphics,*** Vol. 2, No. 4, October - December, 2002.

The Perceptual Interfaces and Reality Laboratory

Ramani Duraiswami

Faculty
Rama Chellappa
Larry Davis
Shihab Shamma

Visiting Faculty
Bayya Yegnanarayana

Research Faculty
Daniel DeMenthon
Yaser Yacoob
David Harwood
Dmitry Zotkin

The Perceptual Interfaces and Reality Laboratory (PIRL) combines research activities in the areas of Perceptual Interfaces and Perceptual Reality. Perceptual Interfaces are concerned with extending human-computer interaction to use all modalities of human perception. These include vision, gesture, audition, and touch. The goal of Perceptual Reality is the creation of virtual and augmented versions of the world that produce percepts that seem identical to the real world.

Research Overview

Humans sense their environments using sight, sound, smell, touch and taste. These input streams are processed in various ways to extract cues about the environment. An amazingly detailed scene is perceived, and multiple objects can be segmented from the scene. These objects are imbued with multiple attributes, including visual (color, texture, shape, and location), auditory (pitch, timbre, intensity, and location), and touch (force, roughness). Conventional signal and image processing techniques are hard-pressed to extract the same attributes from the signals and ascribe to them the same labels which humans are effortlessly able to. Clearly, biological systems that are able to process such input have much to teach us.

While our research goals are relatively broad, our initial focus has been on the auditory modality. This is partially to provide a complementary strength to established research programs in vision at our Center. A fundamental aspect of our perception of auditory space is the location of an auditory source. Our audio research has focused on both the localization of a real source using machines (microphone arrays) and on rendering virtual auditory spaces where sources are perceived at prescribed locations.

Microphone arrays can be used to record an audio scene in a manner that allows the spatial relationships among sources in the scene to be understood, and that allows the spatial location of a particular source to be determined. They also allow sound from one source to be preferentially enhanced via beamforming. Applications include speech interfaces to computers where the talker is not tethered to a mike. A major focus of our research has been in developing novel algorithms for source localization and beamforming in reverberant environments, such as are customarily encountered. A microphone array developed at PIRL was used as the auditory modality in a SIGGRAPH 2000 demonstration of an interactive virtual toy. The toy's emotional responses were based on our research on emotion recognition (see Figure 46).

Microphone arrays and cameras can be used to multimodally sample the same space. Applications of this include creating videoconferencing systems that provide optimal sound and video output and maintain an awareness of a scene as a human cameraman would. A low-level task associated with this is the joint tracking of an acoustically active source via vision and sound, and tracking through occlusions in either chan-

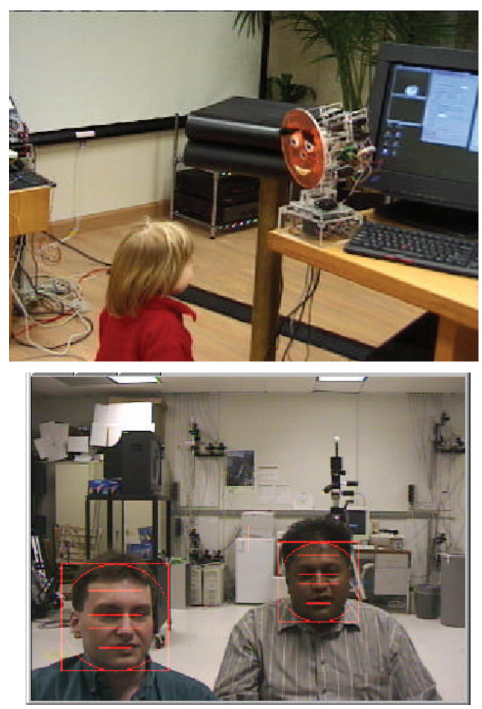


Figure 46. The top panel shows a child interacting with a "VToy". The bottom panel shows our "smart video-conferencing" system in which speakers are located using both face detection in the video scene, and acoustic source localization.

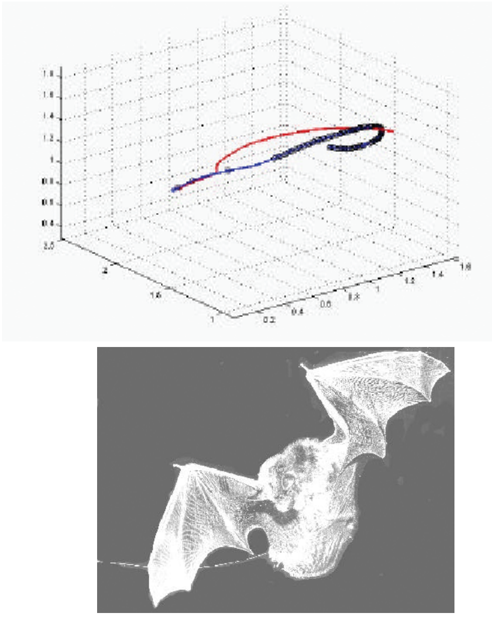


Figure 47. Tracking a bat. The computed track of the bat and the food. Circles show audio estimates while the line shows video estimates.

nel. Other associated probems included face detection to control camera zoom, tracking of people to allow a camera to pan and follow them as they move, and distributing the audio and video computation over multiple machines and processes to achieve a scalable architecture for "smart videoconferencing" (see Figure 46).

A serendipitous collaboration that resulted from this research is a project with the Bat Lab in the Department of Psychology. The goal of Bat Lab researchers was to track bats as they hunted prey using audio, and to show that bats are able to discriminate prey from distractors in a dark room using the audio modality alone. Infrared imaging and microphone arrays were used to track the bats, along with algorithms that we developed (see Figure 47).

Auditory Source Localization by Humans and the Head- Related Transfer Function

While determining the location of an audio source via a microphone array requires a minimum of four microphones, and even with much larger numbers of microphones still faces difficulties due to noise and reverberation, humans are effortlessly able to localize sources in such environments using just two ears. The mechanisms that allow them to do this must be understood, especially if we wish to create virtual worlds with prescribed auditory source locations. Scattering of sound by the human body (especially the external ears) provides powerful monaural and binaural cues for source location. The wavelength of audible sound, extending from 2 cm to 20 m, is comparable to the dimensions of the bodies of humans, and of our external ears. It turns out that our external ear is a very specially shaped "antenna" that presents a different shape to waves arriving from different directions, and accordingly has a different frequency-dependent scattering cross-section to these waves. This position- and frequency-dependent scattering behavior is typically encoded as a Head Related Transfer Function (HRTF), which represents the ratio of the complex sound pressure level received at the entrance to the ear canal, to the level that would have been received at the location of the head in the listener's absence. Knowing the HRTF, one can, in principle, reintroduce the cues created by a source at a particular location. However, the HRTF exhibits significant person-to-person variation. This is because the shapes of our bodies and ears are extremely individual (indeed, ear shapes were used for forensic work before fingerprints), and scattering off these shapes produces individual HRTFs. The HRTF is usually obtained via tedious measurements in which the sound is presented to the listener from different locations. However, it can also be obtained using scientific computing, by solving the wave equation with a discretized mesh of the surface of the person's head, torso and ears, obtained using computer vision. Our innovative approach to computing the HRTF is funded by an NSF ITR grant awarded in 2000. We have recently developed extremely fast algorithms based on the fast multipole methods for the numerical solution of the wave equation.

Creating Virtual Auditory Spaces

While the HRTF is important for "rendering" virtual audio, it turns out that it is only one component, and some other unsolved problems remain. The first set of difficulties arise due to the fact that humans are mobile, and the configuration of the scattering surfaces, and their positions relative to the expected location of the source, change as they move. This points to two key difficulties: first, the HRTF will change as users move; second, the users will expect the rendered source to remain still even when they move. Thus, the audio scene that is rendered must involve dynamic modification of the HRTF according to the user's movement, and must track the locations of the user's ears. Second, the sound that is received at the ears includes both direct sound from the source and sound that arrives after multi-path scattering from the walls and other surfaces in the environment. The scattered sound reaching the ears will thus arrive at different times than the direct sound, and will also have different spectral features due to scattering and attenuation, since the scattering surfaces have dimensions comparable to the wavelengths of sound at lower frequencies. The received sound thus encodes information about

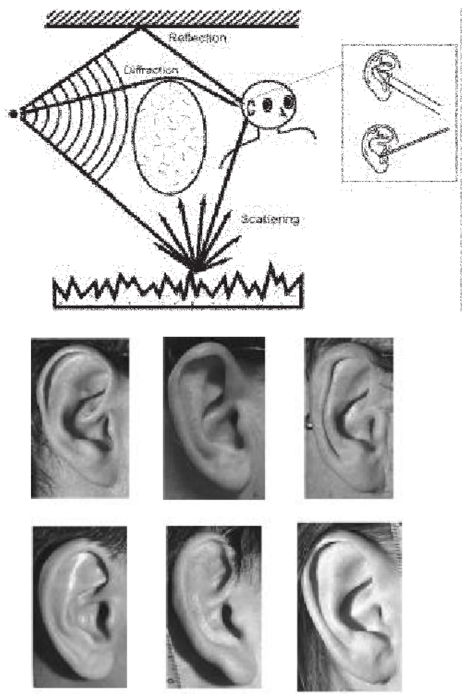


Figure 48. (a) Sound scattering off the environment and off our bodies provides cues to the location of the source and the size and composition of the environment. (b) Scattering off the body is extremely individual, since people's sizes and especially the shapes of their ears exhibit considerable variability leading to extremely individual HRTFs.

the environment, and also about the source location, especially its range (see Figure 48).

Room properties are encoded in a room impulse response. Calculating this response is difficult, as it involves solution of the wave equation for large complex geometries over a large frequency range. For high frequencies, approximate image models can be used. These models are similar to those used in computer graphics, but there is no guarantee of accuracy and no means to determine whether these models are doing an adequate job. We are investigating approaches to efficiently computing the room impulse response directly using physical models, and have developed methods based on the fast multipole method that speed up the computations by an order of magnitude.

Further, creating virtual auditory spaces requires development of smart signal processing algorithms to achieve real-time composition of time delay cues, HRTF cues, early room reflections and late room reflections. We have developed a system for creating a virtual auditory space that accounts for all these. An interactive "shoot-'em-up" game in which targets are localized using both head-mounted displays and audio cues has been developed.

Audio User Interfaces for the Visually Impaired and for General Users

Recent research at PIRL seeks to use our ability to create virtual auditory spaces to create auditory displays (funded via a seedling grant by DARPA's Augmented Cognition Program) and auditory user interfaces for the visually impaired (to be funded by a 2002 ITR award). Our auditory user interface will incorporate elements from all dimensions of the human auditory system that are neurologically significant, including:

- **Sound spatialization in three dimensions** using individualized/customizable HRTFs and environmental models that result in stable spatialization of multiple sources;
- **Tone and pitch modification** via use of a neuromorphic model that transforms sound to its neurological correlates;
- **Timbre modification** using the same neuromorphic model;
- **Sound intensity** manipulation;
- **Environmental characteristics** modification.

We are exploring how it may be possible to achieve perceptually distinguishable mappings of various types of data to this auditory space.

Textual Information Access for the Visually Impaired

Another portion of our research is concerned with creating prosthetic devices for the vision and hearing impaired, by mapping inputs from one modality into equivalent ones in another, so that computationally augmented input streams can be created with extra content from the missing modality. In a collaborative project with Johns Hopkins University we are creating seeing-eye computers that can detect text, perform OCR on it, and read it via text-to-speech (see Figure 49).

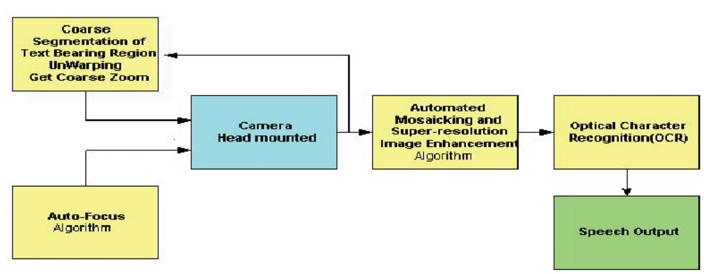


Figure 49. Schematic of the TIAVI system used to convert text recorded via a video camera into audible sound heard by a visually impaired listener.

Contracts and Grants

Sponsor	Dates	Amount	Title
NSF	09/00-09/05	$2,999,995	Personalized Spatial Audio via Scientific Computing and Computer Vision
NSF	09/02-09/07	$1,800,000	Customizable Auditory User Interfaces for the Visually Impaired and the Sighted
NSF	04/00-04/03	$700,000	Texture Information Access for the Visually Impaired
NSF	09/02-09/05	$450,000	Fast Multipole Translation Algorithm for Solution of the 3D Helmholtz Equation.
ONR	04/02-04/03	$175,000	Customizable Auditory Display

Journal Publications

1. V.R. Algazi, R.O. Duda, R. Duraiswami, N.A. Gumerov, and Z. Tang, Approximating the HRTF at Low Frequencies Using Simple Geometric Models of the Head and Torso, ***J. Acoust. Soc. Am.***, in press.

2. N. A. Gumerov and R. Duraiswami, Multiple Scattering from N Spheres Using Multipole Reexpansion, ***J. Acoust. Soc. Am.,*** in press.

3. D. Zotkin, R. Duraiswami, and L. Davis, Joint Audio-Visual Tracking Using Particle Filters, EURASIP ***Journal on Applied Signal Processing,*** 2002.

4. D. Zotkin, R. Duraiswami, and L. S. Davis, Creation of Virtual Auditory Spaces, ***IEEE Trans. Multimedia***, in press.

Conference Papers

1. R. Duraiswami, N. Gumerov, D. Zotkin, and L. Davis, Efficient Evaluation of Reverberant Sound Fields, ***IEEE Workshop on Applications of Signal Processing to Audio and Acoustics,*** pp. 203-206, 2001.

2. R. Duraiswami, D. Zotkin, and L. Davis, Active Speech Source Localization by a Dual Coarse-to-Fine Search, ***International Conference on Acoustics, Speech, and Signal Processing,*** 2001.

3. R. Duraiswami, D. Zotkin, and L. Davis, Multimodal 3-D Tracking and Event Detection Via the Particle Filter, ***Workshop on Event Detection in Video***, 2001.

4. N.A. Gumerov, Rectified Diffusion and Heat Transfer to Vapor-Gas Bubbles. ***International Conference on Multiphase Flow***, 2001.

5. N.A. Gumerov, Acoustic Cavitation Thresholds and Stable Oscillations of Vapor-Gas Bubbles, ***Meeting of the Acoustical Society of America***, 2001.

6. N.A. Gumerov and R. Duraiswami, Modeling the Effect of a Nearby Boundary on the HRTF, ***International Conference on Acoustics, Speech, and Signal Processing***, 2001.

7. N.A. Gumerov and R. Duraiswami, Multiple Scattering from N Spheres, ***IEEE AP-S International Symposium and URSI National Radio Science Meeting***, 2002.

8. N.A. Gumerov and R. Duraiswami, Fast Multipole Translations for the Helmholtz Equation, ***SIAM Annual Meeting,*** 2002.

9. N.A. Gumerov, R. Duraiswami, and Z. Tang, Numerical Study of the Influence of the Torso on the HRTF, ***International Conference on Acoustics, Speech, and Signal Processing***, 2002.

10. N.A. Gumerov, R. Duraiswami and D.Zotkin, Room Transfer Functions for Convex Polyhedral Rooms, ***Joint Meeting of the Acoustical Societies of America and Mexico***, 2002.

11. N.A. Gumerov, Z. Tang, R. Duraiswami, R.O. Duda, V.R. Algazi, and S.T. Raveendra, Modeling the Effect of the Torso on the Head-Related Transfer Function Via the Boundary Element Method, ***Meeting of the Acoustical Society of America***, 2001.

12. I.Haritaoglu, A.Cozzi, D. Koons, M.Flickner, D. Zotkin, R. Duraiswami, and Y. Yacoob, Attentive Toys, ***Proc. ICME***, pp. 1124-1127, 2001,

13. A. Zandifar, R. Duraiswami, A. Chahine, and L.S. Davis, A Video Based Interface to Textual Information for the Visually Impaired, ***International Conference on Multimodal Interfaces***, 2002.

14. D. Zotkin, R. Duraiswami, and L. Davis, Creation of Virtual Auditory Spaces, International Conference on Acoustics, Speech, and Signal Processing, pp. 2113-2116, 2002.

15. D. Zotkin, R. Duraiswami, and L. Davis, Customizable Auditory Displays, ***International Conference on Auditory Display,*** pp. 167-176, 2002.

16. D. Zotkin, R. Duraiswami, L. Davis, V. Raykar, and A. Mohan, Virtual Audio System Customization Using Visual Matching of Ear Parameters, ***International Conference on Pattern Recognition,*** 2002.

17. D. Zotkin, R. Duraiswami, H. Nanda, and L.Davis, Multimodal Tracking for Smart Videoconferencing, ***International Conference on Multimedia***, 2001.

Our Recent Graduates

Ph.D's

Name	Year	Thesis Topic
Ross Cutler	2000	On the Detection and Analysis of Oscillatory Motions in Video Sequences
Katherine Guo	2000	Forgery Detection by Local Correspondence
Geoffrey Hazel	2000	Texture, Shape and Context for Automatic Target Detection and Classification in Spectral Imagery
Gisli Hjaltason	2000	Incremental Spatial Algorithms
Vikrant Kobla	2000	Automated Analysis of MPEG-Compressed Video
Baoxin Li	2000	Human and Object Tracking and Verification in Video
Huiping Li	2000	Text Processing and Understanding in Digital Video Databases
Vasanth Philomin	2000	Real-Time Generic Object Detection and Tracking for "Smart" Vehicles
Robert Pless	2000	Video Linking
Manjit Ray	2000	Object Recognition Using Three-Dimensional Invariants
Christian Shin	2000	The Roles of Document Structure in Document Image Retrieval and Classification
Thanarat Chalidabhongse	2001	A Real-Time System for Detecting and Tracking People in Video
David Kuijt	2001	An Object-Oriented Approach to Parallel Spatial Indexing of Vector-Format Polygons
Hankyu Moon	2001	Shape-Encoded Particle Filtering for Object Detection and Tracking
Gang Qian	2001	Robust Methods for Structure from Motion Using Video Sequences
Yonatan Wexler	2001	Tensorial Methods for Image Synthesis
Motilal Agrawal	2002	Three Dimensional Reconstruction from Multiple Images
Chiraz Benabdelkader	2002	Gait as a Biometric for Person Identification in Video
Ahmed Elgammal	2002	Efficient Kernel Density Estimation for Real Time Computer Vision
Anurag Mittal	2002	Video Analysis Under Severe Occlusions
Amit RoyChowdhury	2002	Statistical Analysis of 3D Modeling from Monocular Video Streams
Dmitry Zotkin	2002	Algorithms for Acquisition and Rendering of Sounds in Perceptual User Interfaces

M.S.'s

Name	Year	Thesis Topic
Yuan Qi	2000	Learning Algorithms for Audio and Video Processing--Independent Component Analysis and Support Vector Machine Based Approaches
Ping Yu	2000	Mobility in an IP Environment
Doe-Wan Kim	2001	A Point Matching Algorithm for Automatic Groundtruth Generation for Microfilm Images
Qigong Zheng	2001	Document Degradation Models and their Use in Image Restoration
Puneet Gupta	2002	Automatic Redetection and Classification of Defects

Faculty


John S. Baras


Leila DeFloriani


Daniel DeMenthon


Bonnie Dorr


Cornelia Fermüller


David Harwood

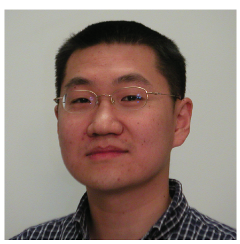

Rebecca Hwa

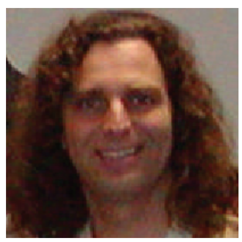

Volker Krüger


Huiping Li


Reuven Meth

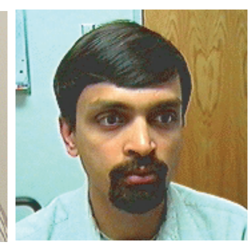

Doug Oard


Gang Qian

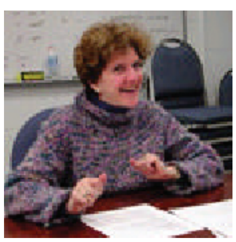

Philip Resnik


Amit RoyChowdhury


Shihab Shamma

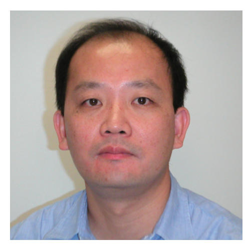

Chandra Shekhar


Egemen Tamin


Amy Weinberg


Yaser Yacoob


Liang Zhao


Qinfen Zheng

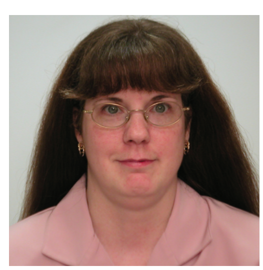

Dmitry Zotkin

Staff


Sailaja Akunuri


Denise Best


Sara Larson


Janice M. Perrone


Andy Pilipchuk

Center for Automation Research
University of Maryland
College Park, MD 20742-3275
(301) 405-4526
http://www.cfar.umd.edu

Computer Vision Laboratory
University of Maryland
College Park, MD 20742-3275
(301) 405-4526
http://www.cfar.umd.edu./cvl

Language and Media Processing Laboratory
University of Maryland
College Park, MD 20742-3275
(301) 405-6444
http://documents.cfar.umd.edu/LAMP

Graphics and Visual Informatics Laboratory
University of Maryland
College Park, MD 20742-3275
(301) 405-6761
http://www.cfar.umd.edu/gvil

Perceptual Interfaces and Reality Laboratory
University of Maryland
College Park, MD 20742-3275
(301) 405-6710
http://pirl.umd.edu

Other Laboratories Affiliated with the Center

Laboratory for Computer Aided Control Systems Engineering
(Institute for Systems Research)

Participating Faculty:
Eyad Abed, Raymond A. Adomaitis, John S. Baras, P.S. Krishnaprasad, William S. Levine, Steve Marcus, Gary W. Rubloff, Andre L. Tits, Evanghelos Zafiriou

Areas of Interest:
Optimization-based control system CAD, process simulation, real-time control, systems design and optimization; applications: simulation and control of semiconductor manufacturing processes, chemical process control, gas turbine engine control

Laboratory for Basic Research in Sensory Systems (Department of Psychology)

Participating Faculty:
Casper J. Erkelens (Utrecht Bio-Physics Research Institute), Eileen Kowler (Rutgers University), Zygmunt Pizlo (Purdue University), Robert M. Steinman

Areas of Interest:
Study of coordinated visuo-arm-motor action of human being manipulating objects within arm's reach, with high spatial and temporal resolution. For visualizations of this system see http://brissweb.umd.edu.

UNIVERSITY OF MARYLAND 18 56

Produced by University Printing Services for the Center for Automation Research 11/02. Design by Sailaja Akunuri.

Printed on Recycled Paper