Advanced Visual Sensor Systems *

Terrance E. Boult
Rick Blum
Ele. Eng. & Comp. Sci. Dept.
Lehigh Univ., Bethlehem, PA 18015
[tboult|rblum]@eecs.lehigh.edu

Shree K. Nayar
P.K. Allen & J.R. Kender
Department of Computer Science
Columbia Univ., NYC, NY 10027
[nayar|allen|jrk]@cs.columbia.edu

Abstract

This report is on the Lehigh/Columbia MURI contract. While the original focus was on sensors for manufacturing, the natural evolution of our basic research has led us to more general problems in more generic settings. As a multi-faculty multi-disciplinary project much of the work is naturally done in smaller subgroup. The major results over the past year were on 3D modeling/sensor planning, omni-directional imaging, and reflectance/image/noise modeling. There were more focused results in deformable models, feature detection, appearance matching, and video segmentation. Note that some projects (e.g., the outdoor 3D model building) bring together many of the above topics. This report provides short summaries of our significant contributions with citations to related papers. Length of presentation herein does not reflect level of effort nor our view of its significance – many of the most important areas have papers elsewhere in these proceedings.

1 Automated Site Modeling

3-D models of outdoor environments, known as site models, are used in many different applications that include city planning, urban design, fire and police planning, surveillance and virtual reality modeling. Creating site models for urban scenes which contain large structures (i.e., buildings) that encompass a wide range of geometric shapes and contain high occlusion areas is quite challenging.

These models are typically created by hand in a painstaking and error prone process. We are building a system to automate this procedure that extends our previous work in 3-D model acquisition using range data [Reed-1998, Reed and Allen-1998, Reed *et al.*-1997, Allen and Yang-1998]. This is an incremental volumetric method that can acquire and model multiple objects in a scene and correctly merge models from different views of the scene. Models are built from successive sensing operations that are merged with the current model being built, called the composite model. The merging is performed using a regularized set intersection operation. These partial, composite models can serve as input to our sensor planning system that can reduce the number of views needed to fully acquire a scene. Planning each sensing operation reduces the data set sizes and acquisition time for a complex model. Most existing systems do not use planning, but rely on human interaction or the need for large overlaps in images to assure adequate coverage of the scene. Our planner can incorporate different constraints including visibility, field-of-view and sensor placement to find the next viewing position that will reduce the model's uncertainty. The system has been tested on

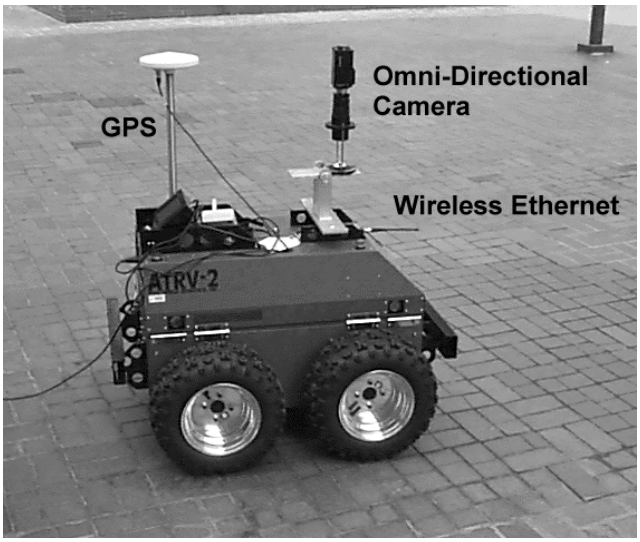


Figure 1: Mobile robot for automated site modeling.

*This work supported by ONR/ARPA MURI program ONR N00014-95-1-0601. Several other agencies and companies have also supported parts of this research.

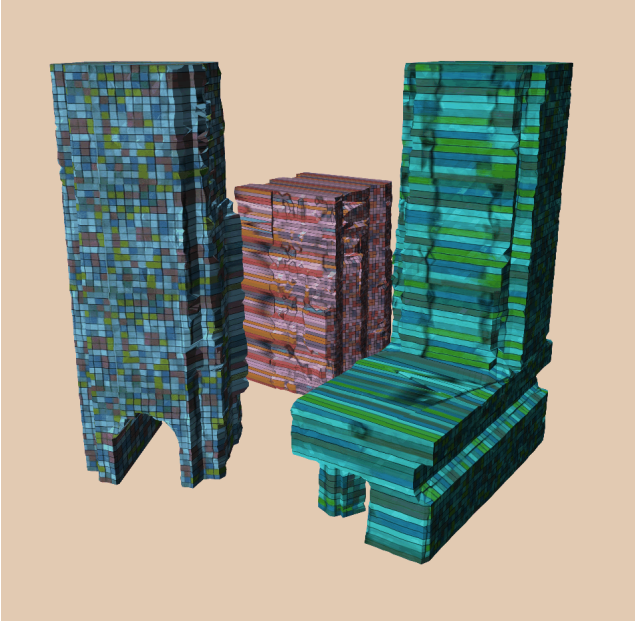


Figure 2: Recovered 3-D models of a blocks world city. All 3 objects were recovered at once. Sensor planning algorithms were used to reduce the number of range scans and find unoccluded viewing positions.

indoor models and we are now extending it to outdoor scenes with multiple objects.

At each step of the process, a partial solid model of the scene is created. The faces of this model consist of correctly imaged faces and faces that are occlusion artifacts. We can label these faces as "imaged" or "unimaged" and propagate/update these labels as new images are integrated into the composite model. The faces labeled "unimaged" are then the focus of the sensor planning system which will try to position the sensor to allow these "unimaged" faces to be scanned. The set intersection operation must be able to correctly propagate the surface-type tags from surfaces in the models through to the composite model. Retaining these tags after merging operations allows viewpoint planning for unimaged surfaces to proceed.

Figure 2 is a recovered 3-D solid model of a simulated city scene made up of 3 toy buildings placed on our laser scanner turntable. Four initial views were taken at 90° intervals, and this merged partial model was used by the view planner to choose the next view to reduce the model's uncertainty. This process of a partial model driving the planner for the next view was used until 8 more images were taken, reducing the model's uncertainty to a small threshold. The model is accurate and the method can recover structure that is occluded.

For automating this task outdoors, we are equipping a mobile vehicle with sensors and algorithms to accomplish this task. A picture of the vehicle is shown in figure 1. The equipment consists of an RWI ATRV mobile robot base, a spot range scanner (80 meter range spot scanner with 2-DOF scanning mirrors for acquiring a whole range image, not shown in image), centimeter accuracy on-board GPS, color cameras for obtaining photometry of the scene, and mobile wireless communications for transmission of data and high level control functions.

The planning algorithms can be used to navigate the mobile robot scanner system to a position for a new scan that will reduce the uncertainty in the scene. Given a partial model of the scene, we can tag the surfaces acquired by the scanner in such a way as to know what regions are imaged and which are occlusion surfaces. We can then use these occlusion surfaces to find unobstructed viewpoints. Once we compute these viewpoints, we can then command the mobile system to navigate inside a visibility volume and take a new scan, thus completing a partial model and "filling in the blanks" to build the complete model. Details can be found in this proceedings [Allen *et al.*-1998].

2 Interactive Sensor Planning

The automated site modeling system will be merging the range data with 2-D imagery to enhance the models. This also requires a view planning component. In cluttered and complex environments such as urban scenes, it can be very difficult to determine where a camera should be placed to view multiple objects and regions of interest. We have built an interactive sensor planning system [Stamos and Allen-1998] that can be used to select viewpoints subject to camera visibility, field of view and task constraints. This work builds upon our earlier work in sensor planning [Abrams-1997]. Application areas for this method include surveillance planning, safety monitoring, architectural site design planning, and automated site modeling. Given a description of the sensor's characteristics, the objects in the 3-D scene, and the targets to be viewed, our algorithms compute the set of admissible view points that satisfy the constraints. The system first builds topologically correct solid models of the scene from a variety of data sources. Viewing targets are then selected, and visibility volumes and field of view cones are computed and intersected to create viewing volumes where cameras can be placed. The user can interactively manipulate the scene and select multiple target features to be viewed by a camera. The user can also select can-

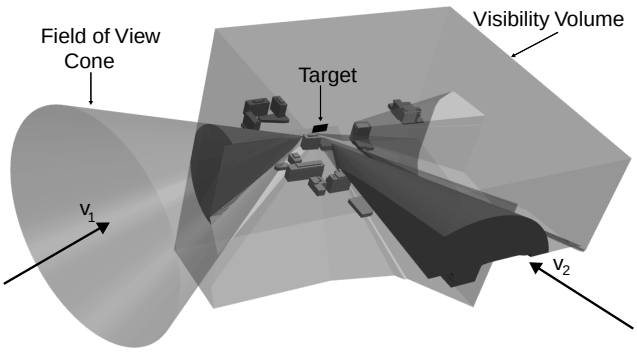


Figure 3: Planning camera positions in an urban scene. The target can be any face in the scene (multiple targets can also be used). The Visibility Volumes (transparent polyhedral volumes), the Field of View Cones for the direction $\mathbf{v_1}$ (transparent cones) and the Candidate Volumes (intersection of the visibility volumes with the field of view cones) for the viewing direction $\mathbf{v_1}$ (left partial cones) and for the directions $\mathbf{v_2}$ (right partial cone) are displayed.

didate viewpoints within this volume to synthesize views and verify the correctness of the planning system. The interactive system allows us to generate, load and manipulate different types of scenes and interactively select the target features that must be visible by a camera. The results of the sensor planning experiments are displayed as 3-D volumes of viewpoints that encode the constraints. Virtual cameras placed in those volumes provide a means of synthesizing views in real-time and evaluating viewpoints. The system can also be used to provide animated "fly throughs" of the scene.

Here is a brief example of how the system works for planning camera positions for surveillance of multiple targets in an urban scene. As input, the planner needs models of buildings that are watertight B-Reps (i.e. no dangling faces). In most cases, existing urban/city models are graphics models that have no need to be topologically consistent and geometrically correct, since 2-D viewing is the main application. Those models are not guaranteed to be complete since they lack topological information. We have built a tool that can take an incomplete model and recover the adjacency information between the faces to build a correct solid model. Once these solid models are built, targets can then be selected in the scene and a set of viewing volumes computed that determine visibility and field-of-view for viewing multiple targets in an urban scene. Intersecting these volumes for each viewing constraint results in the locus of viewpoints which are guaranteed to be occlusion-free and able to see targets within the field of view. Figure 3 shows a city model that was generated from an incomplete graphics model of Rosslyn VA. and was translated by the system to a valid solid model which the planner can use. Overlaid on the city model are the viewing volumes generated for different viewpoints on a selected target face in the scene. The object models and targets can be interactively manipulated while camera positions and parameters are selected to generate synthesized images of the targets that encode the viewing constraints. We are currently extending this system to include resolution constraints. Details can be found in these proceedings [Allen *et al*.-1998].

3 A Theory of Catadioptric Image Formation

Many applications of computational vision require a large field of view. Examples include surveillance, teleconferencing, and model acquisition for virtual reality. A number of other applications, such as egomotion estimation and tracking, do not strictly require a large field of view, but their robustness generally increases as the field of view gets wider. Although conventional imaging systems typically have very limited fields of view, using them in conjunction with a mirror can yield a wide field of view. This approach of using mirrors in combination with conventional imaging systems is usually referred to as *catadioptric image formation.* A large number of designs for wide field of view catadioptric imaging systems have been proposed in the literature. See, for example, [Rees-1970], [Charles *et al*.-1987], [Nayar-1988], [Yagi and Kawato-1990], [Hong-1991], [Goshtasby and Gruver-1993], [Yamazawa *et al*.-1993], [Bogner-1995], [Nalwa-1996], and [Nayar-1997].

As noted in [Rees-1970], [Yamazawa *et al*.-1995], [Nalwa-1996], and [Nayar-1997], it is highly desirable that an imaging system, catadioptric or otherwise, have a single effective viewpoint (center of projection). The reason a single effective viewpoint is so desirable is that it is both a necessary and sufficient condition for the generation of geometrically correct perspective images from the captured image(s). Subsequently, these unwarped images can be processed using the vast array of techniques that assume perspective projection. Moreover, if the images are presented to a human, as in [Peri and Nayar-1997], they need to be perspective images in order to appear undistorted.

Various different mirror shapes have been proposed for catadioptric sensors, including cones, spheres, hyperboloids, and paraboloids. Some of these mirror

shapes lead to a single effective viewpoint, whereas others do not. In [Baker and Nayar-1998b], we derived the complete class of catadioptric sensors with a single effective viewpoint that can be constructed using a single conventional camera. We showed that there is a 2-parameter family of mirror shapes that can be used, that being the class of swept conic sections. Within this set of solutions, several prove to be degenerate and hence are impractical, whereas others do lead to practical sensors. We evaluated all of the catadioptric sensors mentioned above in light of this result and showed which designs have a single viewpoint and which do not.

An important property of any sensor that images a large field of view is its resolution. The resolution of a catadioptric sensor is not, in general, the same as that of the sensor(s) used to construct it. In [Baker and Nayar-1998b], we also derived the relationship between the resolution of a conventional imaging system and that of a derived catadioptric sensor.

Another optical property that is affected in a catadioptric sensor is focusing. It is well known that a curved mirror increases image blur [Hecht and Zajac-1974]. In [Baker and Nayar-1998b], we analyzed this effect for catadioptric sensors. We showed how a focus look-up table can be computed for a catadioptric sensor. The results show that the focus setting of a catadioptric sensor using a curved mirror may be substantially different from that needed in a conventional sensor. Moreover, even for a scene of constant depth, widely different focus settings may be needed for different parts in the scene.

4 Remote Reality

The full viewing nature of omni-directional image lead quite naturally to develop what we call remote-reality. Virtual reality puts the user in an immersive "virtual" world. In remote reality there is nothing virtual about it; remote reality is an immersive video from a remote location. We have developed software that allows a user, who wears a (COTS) head-mounted-display with head-tracker, to look around within the omni-directional video stream. For more details see [Boult-1998c, Boult-1998b, Boult-1998a, Boult *et al.*-1998] or our website.

To be immersive the motion and video must be natural, which means 30fps video, fast reaction (15fps or better) to head movements and a smoothly moving viewpoint. The last part of this is inherent in the ParaCamera design, there is only one viewpoint. The other two are achieved, at 320x240 resolution, on a 233Mhz processor with inexpensive COTS parts.

Figure 4: Omni-directional camera on Remote Control Car and Remote Reality Driver.

When used with taped video, remote reality has direct application in training, mission rehearsal, route planning and complex plant management. With live video transmission it can be used in vehicle operation and in urban maneuvers (e.g., we have demonstrated mounting a paracamera on a remotely controlled vehicle and driving it using the remote-reality system). In an urban action, the vehicle could be driven from a remote location while a small unit follows behind the vehicle. Using a wearable remote reality system the soldier could independently and simultaneously look around the vehicle's location while remaining radio-silent. It also applies to remote underwater or airborne vehicles where a pilot could use remote reality to fly the vehicle while others look around for points of interest. The 3D nature of these vehicles makes it even more important that the the pilots have instant response as they turn their head. At the same time their nature limits transmission bandwidth; with remote reality a single video stream would support both the pilot and multiple spotters that could in any direction, including down ravines or side streets, for potential threats/targets

While the para-camera is now commercially available, we are still working on new camera designs, especially for the remote-reality applications described above. At Columbia, they are addressing both size reduction and environmental issues, see [Nayar and Boult-1998]. At Lehigh, most of the new designs are addressing physical and environmental constraints, and need to mix both optical and mechanical issues to maintain quality. These cameras developed include two different versions of under-water omni-directional cameras, two cameras designed for less obvious vehicles mounting, and two environmentally stable systems (used in the fall VSAM demo). Figure 5 shows two of these systems. We have also developed a dual-mode systems that uses a single camera which can be used in either omni-directional

mode or in a traditional pan-tilt-zoom mode.

Figure 5: Some custom omni-directional cameras for vehicle and underwater use.

5 Small Vision Systems

While the COTS laptop market has made advancements with PCMCIA video cards and some internal capture boards, the response to market pressure has resulted in systems that cannot capture/process full resolution video. They provide zoomed-port video, where full resolution video is overlaid on the screen, single frame capture (at full resolution) and 320x240 processing across the bus.

With the resolution demands of the omni-directional image systems, it was natural for us to convert our past efforts on LVS, the laptop-vision-system, to smaller, more robust and/or compact systems capable of full resolution processing. This has been mostly a systems engineering issue working with commercial-off-the-shelf parts, integrating them and porting our software. We have been exploring different points in the price/power/speed spectrum for potential applications ranging from security to wearable remote reality. The current systems include:

- a commercially Power-PC403-based system in a 10cm x 10cm x 25cm package providing 640x480 gray-scale imaging, 16MB memory and 10MB networking.
- a low-power (10watts) Pentium based PC104+ system based for wearable operation with dimensions 14cm x 14cm x 7cm. (Total system draw $\approx$ 20watts).
- a low cost commercial strong-ARM system, 22cm x 15cm x 4cm with $<$ 10 watt total diskless operation.
- two compact-PCI Pentium based systems, one in a NEMA-4 ruggedized housing.

This is an ever changing arena with a myriad of tradeoffs to address. People interested can find more detail, including system photos, on our website.

6 Panoramic Pyramids

Because wide-field of view images pack a large field of view into a single image, the resolution of that image is important for tracking and security applications. Thus many researchers are looking at moving to higher resolution cameras. As imaging technology advances higher resolution cameras are becoming common place, (e.g., HDTV Cameras produce approximately 2Kx1K 30bit pixels at 30fps). At lower frame rates, 2Kx2K and 4K x4K cameras are now commonly available. At these frame sizes, the data rates become phenomenal. An HDTV camera produces over 1.8Gbits (225MBytes) per second. At these rates just transferring the data into the machine is a serious issue.

Multi-resolution techniques, (a.k.a. Pyramid algorithms) have been used in a wide range of vision applications to reduce computational demands, and address, inherent scale issues. Unfortunately, building a proper pyramid is, in and of itself, a potentially costly operation requiring a pre-filtering convolution before down sampling. Presuming a 7x7 separable convolution before down sampling, we need 14 multiplies and 13 additions for each of 3x640x480 pixels at 30fps for approximately 750 MOPS to make the first layer of the pyramid, with 1/4 of that for each additional layer, the total is about 1000 MIPS just to form the pyramid. While this can be done with today's processors, it's quite taxing and has lead other DARPA IU projects to build special purpose hardware. The computational demand becomes more a more significant issue when we consider HDTV or larger images, for which the large data rates demand "intelligent" processing. Building a good pyramid would require 12,000 MIPS for HDTV and 70,000 MIPS for 4k x 4k 30fps imagery. None of todays specialized hardware, even the most advanced DSPs, can currently build such pyramids in real-time.

While coarse-to-fine processing is important and desirable, just the building of the pyramid to enable coarse-to-fine processing would require more computing power than available in todays workstations let alone that available in lower power units we might run from batteries. It might seem that high-resolution high-speed imaging is still well beyond us. However, with some catadioptric help, it is almost within reach today with panoramic pyramids.

Recall that in our omni-directional camera, a parabolic mirror is imaged by an orthographic lens to produce the image. The combination of orthographic projection and the parabolic mirror provides a single viewpoint, at the focus of the parabolic sur-

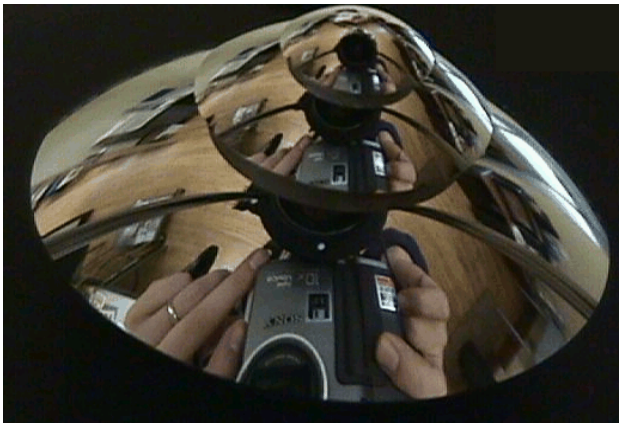


Figure 6: A three layer physical panoramic pyramid imaged from side to show mirror stack.

face. For any parabolic mirror size, the resulting paraimage contains the full hemispheric view. The size of the omni-directional image depends on the image magnification and the size of the mirror. The *panoramic pyramid*, proposed by T. Boult at Lehigh, uses a set of parabolic mirrors stacked one on top the other. Each mirror produces a paraimage and can be unwarpped to provide a perspective view in any direction within a panoramic slice. In the pyramid in figure 6, the mirrors were chosen so that the generated omni-directional images were 1/2 the resolution of the next finer resolution. Mirrors can, however, provide any resolution reduction desired. e.g. 4-to-1, 10-to-1 or even 6.4-to- 1 (say to reduce a 4kx3k image to normal 640x480 video).

The resulting panoramic pyramid system allows course-to-fine processing in the truest sense: after processing the pixels in the lower resolution regions, only the higher resolution pixels in the regions of interest need be transferred from the imaging array. We do not even need to transfer, let alone process, the pixels in low interest areas.

While we have built a panoramic pyramid prototype, there are numerous research issues still to be addressed especially aspects of noise, handling the increased demands on depth-of-field, system calibration and general real-time software development. Initial prototyping will use existing high-resolution cameras and PC. We will also be pursuing direct integration of high resolution imagers and DSP's to allow full programmability and "on camera processing".

7 Image Fusion

In recent years, image fusion approaches based on multiscale decomposition (MSD) began to emerge and receive increased attention. In our research,

Figure 7: A panoramic pyramid image of size 480x480. The "edges" of the mirrors only distort 1 pixel. (Because of alignment errors in the mounting the smallest mirror is displaced slightly downward.)

a generic image fusion framework (see Figure 1) based on multiscale decomposition (MSD) is studied [Zhang and R. S. Blum-, Zhang and R. S. Blum-97, Zhang and R. S. Blum-1997]. This framework provides freedom to choose different MSD methods and different fusion rules. It can not only describe the existing image fusion schemes developed in previous research, but also introduces some new approaches. These new approaches have been shown to outperform previous approaches. Our study was focused on how to use the MSD data of the source images to produce a fused MSD representation which should be more informative to the observer (human or computer). Studies of this type have been lacking. Experiments show that region-based fusion and multiscale MSD coefficients grouping, two new ideas we propose, can almost always improve the image fusion performance for any of the MSD methods we considered. Some natural extensions to the MSD-based image fusion research are also suggested based on this framework.

8 Blind Image Quality Estimation

Some new techniques are proposed for estimating the quality of a noisy image of a natural scene [Zhang and R. S. Blum-1998]. In our techniques, a mixture model is used in conjugation with the expectation-maximization (EM) algorithm to model images obtained by edge detection of the images under consideration. This approach yields an accurate represen-

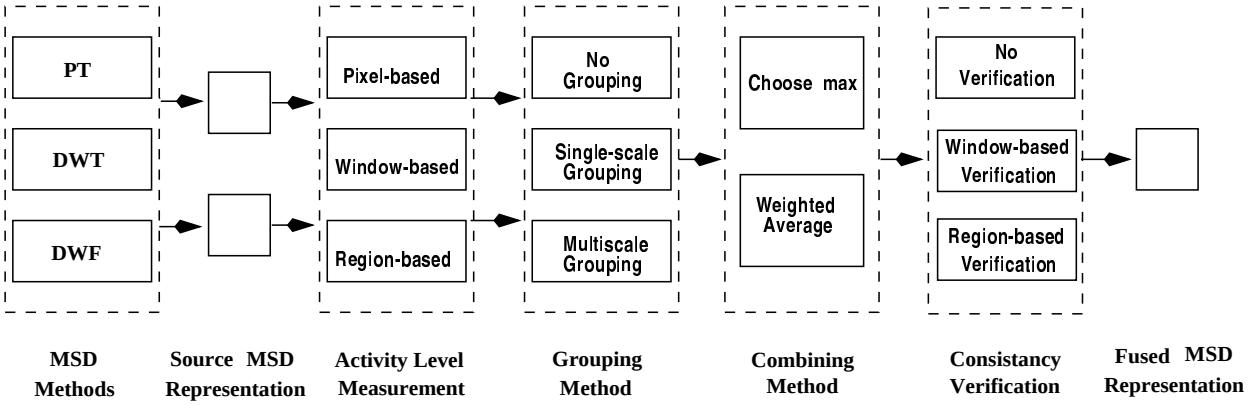


Figure 8: The generic framework of image fusion schemes

tation which is useful in the process of fusing several images to obtain a higher quality image. Quality measures of this type are needed for fusion, but they have not received much attention to date. Analytical justifications are given which explain why these techniques work. Experimental results indicate that the techniques work well in practice. These techniques need only the images to be evaluated and do not use detailed information about the formation of the image. The focus is on the case where the image is only corrupted by additive Gaussian noise, which is independent from pixel to pixel, but some cases with signal-dependent noise or blurring are also considered. Some practical examples are investigated in our research which illustrate the applicability of these techniques.

9 Quantitative Evaluation of Super-Resolution

No matter what the resolution of imaging sensors we build, there is always a desire to get a little more resolution from those images. In the first two years of this MURI project, we addressed this continuing need by developing and demonstrating efficient techniques producing super-resolution images. The techniques differed from previous work by concentrating more on earlier stages in particular the quality of image warping inherent in such techniques. We developed a new technique that integrated a model intra-pixel blurring into the process. Over the past year, we undertook what we believe is the first quantitative "task-based" evaluation of a super-resolution technique, [Chiang-1998]. The quantitative analysis had two components. The first used image sequences of text and the evaluation used a commercial OCR system with character recognition rates as the metric. The results showed that our new technique was significantly better than interpolation techniques and better than the existing super-resolution techniques against which we tested it. The initial qualitative evaluation compared our technique to the best known state-of-the art ([Irani and Peleg-1991, Irani and Peleg-1993]) and compared multiple instances of our approach (using both bilinear and our new warping techniques, with and without deblurring). Only the most successful were considered for quantitative evaluation. The first part of the quantitative evaluation used an OCR system as a back-end and "recognition rates as a metric". For "video-conferencing" quality cameras the experiments showed results ranging from 63% recognition (for uncorrected bilinear interpolation) to 97% raw recognition for new super-resolution techniques. With better cameras and more pixels per character, improvement is less significant, although measurable (97% versus 99%).

Text recognition, however, is a somewhat artificial problem with almost binary data. Our second quantitative evaluation was based on object recognition/pose estimation. The experiment used the SLAM system [Nene *et al.*-1994] which was developed in earlier years of this MURI. SLAM is an appearance based recognition system that uses a parametric eigenspace. For the experiments, SLAM was trained using a 50mm lens on 36 views of 12 objects. For testing, sequences of 8 to 32 images were obtained using a 12 mm lens and different techniques were used to scale the images before using them as input to SLAM. Different poses of the same object provide minimally different "images" where the detail added by super-resolution matters[Chiang-1998]. The results show that even in this more "vision" oriented application, super-resolution helps in most cases. In the cases where super-resolution did not outperform standard cubic-convolution interpolation, it was also the case that both methods did better if we blurred the image. (We believe this is because the images/algorithms were dominated by

aliasing artifacts which are reduced with blurring).

10 Reflectance and Texture of Real World Surfaces

Characterizing the appearance of real-world surfaces is important for many computer vision algorithms. The appearance of any surface is a function of the scale at which it is observed. When the characteristic variations of the surface are subpixel, all local image pixels have the same intensity determined by the surface *reflectance.* The variation of reflectance with viewing and illumination direction is captured by the BRDF (*bidirectional reflectance distribution function*). If the characteristic surface undulations are instead projected onto several image pixels, there is a local variation of pixel intensity, referred to as image *texture*. The dependency of texture on viewing and illumination directions is described by the BTF (*bidirectional texture function*).

We have measured the BRDF of over 60 samples of rough, real-world surfaces. Although BRDF models have been widely discussed and used in vision (see [Nayar *et al.*-1991],[Tagare and DeFigueiredo-1993],[Wolff-1994],[Koenderink *et al.*-1996],[Oren and Nayar-1995]) the BRDFs of a large and diverse collection of real-world surfaces have never before been obtained. Our measurements comprise a comprehensive BRDF database (the first of its kind) that is now publicly available at *www.cs.columbia.edu/CAVE/curet*.

Exactly how well the BRDFs of real-world surfaces fit existing models has remained unknown as each model is typically verified using a small number (2 to 6) of surfaces. Our large database allowed us to evaluate the performance of known models. Specifically, the measurements are fit to two existing analytical representations: the Oren-Nayar model [Oren and Nayar-1995] for surfaces with isotropic roughness and the Koenderink et al. decomposition [Koenderink *et al.*-1996] for both anisotropic and isotropic surfaces. Our fitting results form a concise BRDF parameter database that is also publicly available at *www.cs.columbia.edu/CAVE/curet*. These BRDF parameters can be directly used for both image analysis and image synthesis. In addition, the BRDF measurements can be used to evaluate other existing models [Nayar *et al.*-1991],[Tagare and DeFigueiredo-1993],[Wolff-1994] as well as future models.

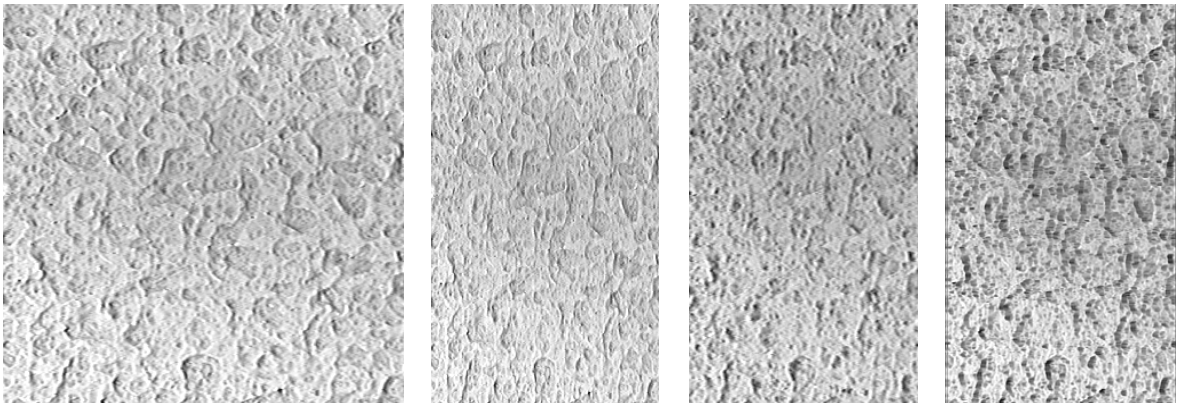
While obtaining BRDF measurements, images of each real-world sample were recorded. These images prove valuable since they comprise a texture database, or a BTF database, with over 12,000 images (61 samples with 205 images per sample). Current literature deals almost exclusively with textures due to albedo and color variations on planar surfaces (see [Wang and Healey-1996],[Chatterjee-1993],[Kashyap-1984]). In contrast, the texture due to surface roughness has complex dependencies on viewing and illumination directions. These dependencies cannot be studied using existing texture databases that include few images (often a single image) of each sample (for instance, the widely used the Brodatz database). Our texture database covers a diverse collection of rough surfaces and captures the variation of image texture with changing illumination and viewing directions. This database is also available at *www.cs.columbia.edu/CAVE/curet*.

11 Histogram Model for 3D Textures

The term *image texture*, or simply *texture,* usually refers to the digital image of a textured surface. In order to understand image texture, the nature of the *surface texture* must be specified. Image texture can arise not only from surface albedo variations (2D texture) but also from surface height variations (3D texture). The distinction between 3D texture and 2D texture is explored in recent work [Dana *et al.*-1996],[Dana *et al.*-1997],[van Ginneken *et al.*-],[Koenderink and van Doorn-1996],[Stavridi and Koenderink-1997],[Leung and Malik-1997]. While there is a large body of work dealing with algorithms for the analysis and synthesis of 2D texture, comparable work for 3D texture is quite sparse. Since the appearance of 3D texture depends on the illumination and viewing direction in a complicated manner, it is useful to refer to image texture as a *bidirectional texture function.*

Modeling and synthesizing this bidirectional texture is key to achieving robust texture recognition and segmentation as well as photorealistic texture rendering. A fundamental representation of texture is the histogram of pixel intensities. For 3D texture, just as image texture is bidirectional, the histogram is a *bidirectional histogram*. Changes in the histogram of 3D texture with illumination and viewing directions are indicative of the surface structure. The work of [van Ginneken *et al.*-] also addresses histograms of 3D texture by investigating the physical mechanisms underlying bidirectional histograms from a large variety of surfaces and by using statistical simulations to generate histograms of Gaussian rough surfaces.

We have developed an analytical model of the bidirectional histogram of image texture [Dana and Nayar-1998]. For arbitrary surfaces, developing



Near Frontal View | Simulated Oblique View | Actual Oblique View | Simulated Oblique View

Texture-Mapping | ***Texture-Morphing***

Figure 9: Illustration of 3D texture synthesis using texture-morphing and the histogram model. From left to right the images are 1) plaster from a near frontal view, 2) a simulated oblique view obtained by applying standard texture-mapping to the frontal view image, 3) the actual oblique view of the plaster sample, 4) a simulated oblique view by applying texture-morphing to the frontal view image. The simulated view obtained by texture-morphing is significantly better than texture-mapping.

such a model is extremely difficult. So, for tractability, we assumed the imaged surface has an isotropic random-slope profile and constant-albedo Lambertian reflectance. This model has been shown to be a good approximation for a variety of natural and man-made surfaces found in ordinary scenes. Our model is based on a geometric/photometric analysis of the interaction of light with the surface. We show the accuracy of the model by fitting to the histograms of real 3D textures from the Columbia-Utrecht texture database [Dana *et al.*-1997].

The model can be used in applications for both computer vision and computer graphics. The parameters obtained from the model fits are roughness measures which can be used in texture recognition schemes. In addition, the model has potential application in estimating illumination direction in scenes where surfaces of known tilt and roughness are visible. We demonstrate the usefulness of our model by employing it in a novel 3D texture synthesis procedure called texture-morphing. We show that results obtained using texture-morphing are superior to those obtained with standard techniques of texture-mapping (see Figure 9).

11.1 Training and Evaluating Deformable Models

A deformable model ("snake," in 2D) is a parameterized shape that is adjusted to minimize an objective function of it and an image, in order to fit the boundary of a depicted object. This work concerns the learning, from ground truth, of the objective function that guides the model. It is being applied to two very different image modalities, cardiac ultrasound and abdominal CT images; the work is expected to readily transfer to other domains, such as image or range data of industrial parts. In contrast to the assumptions of existing methods, the initial domains are realistic in that the desired objects are not necessarily the ones with the sharpest boundary edges.

Since one must choose the features to be learned, as well as the parameterized function that is fit to their distribution, the research also concentrates on evaluating the performance of a resulting objective function [Fenster and Kender.-1998]. The method developed is equally applicable to more typical deformable models, which have not been trained; it can and has been used to evaluate such models.

For the learning part of the work, we have derived objective functions from observed distributions of image intensity and gradient strength at different scales (blurs) for points on the shape boundary. Several related function families were tested for fitting the distributions, including joint Gaussians in intensity and gradient, identically distributed at every boundary point; separate distributions for different "sectors" of the boundary; and a 2D Gaussian with a covariance parameter. Each was tested at three scales.

For the evaluation part of the work, we have tested the objective functions' statistical behavior for shapes with varying degrees of closeness to the correct one, where "close" is measured by chamfer (av-

erage) distance, and "varying degrees" are given by a perturbation model (Figure 10). We have devised novel measures of performance, which include the incidence of "false positives", that is, the number of similar but incorrect shapes scoring better than ground truth, and the correlation coefficient of shape dissimilarity with objective function value.

Experimentation has shown that the simplest distribution model, the joint Gaussians, yielded a vast improvement over the traditional objective function, which maximizes gradient strength. Sectoring the contour into separately learned segments provided a further improvement for abdominal CTs. Twice as coarse a scale was found to be optimal in the ultrasound domain as in the CT domain. These and other results are presented in an accompanying paper in this volume.

12 Optimal Weighting Functions for Feature Detection

More often than not, existing feature detectors give equal importance to each of the pixels. However, some pixels may well provide more reliable information than others. For instance, when estimating the parameters of a corner, it seems intuitively reasonable that the center-most pixels provide almost no reliable information, whereas the pixels on the periphery are vital for high performance. In [Baker and Nayar-1998a], we studied how to weight the contributions of the image data to maximize the performance of a feature detector.

Model matching feature detectors, such as [Hueckel-1971], [Nalwa and Binford-1986], [Rohr-1992], and [Baker *et al.*-1998], are one of the predominant types of feature detectors. Such detectors assume an ideal parametric model of the feature in question. A feature is detected by a model matching detector if there exist parameter values such that the ideal feature instance and the image data are "sufficiently similar." To measure the degree of similarity, the detector requires a matching function. In [Baker and Nayar-1998a], the problem of deciding how to weight the contributions of the pixel intensity data was formulated as one of selecting the matching function to maximize detector performance.

The selection of the matching function for a feature detector had never before been studied in a systematic manner. In fact, most model matching detectors simply use the Euclidean L^2 norm without even mentioning the decision. Examples include [O'Gorman-1978], [Hummel-1979], [Morgenthaler-1981], [Zucker and Hummel-1981], [Nalwa and Binford-1986], [Rohr-1992], and [Baker *et al.*-1998]. Other model matching feature detectors have used weighted L^2 norms, but in all such cases the weighting function was chosen in a completely ad-hoc manner. See, for example, [Hueckel-1971], [Hueckel-1973], and [Hartley-1985].

As with any other matching problem, there is a trade-off between the complexity of the matching function and its performance. In feature detection, efficiency is of vital importance. In fact, one of the primary reasons that the Euclidean L^2 norm has been used so often in the past is because it can be evaluated efficiently. Therefore, we began by showing how the feature detection algorithm of [Baker *et al.*-1998], which uses the Euclidean L^2 norm, can be extended to use an arbitrarily weighted L^2 norm with essentially no additional computational cost. This result allowed us to consider the entire class of weighted L^2 norms as possible matching functions without sacrificing efficiency.

Next, we proposed optimality criteria for two key aspects of feature detection performance: *feature detection robustness* and *parameter estimation accuracy.* We also showed how these two criteria can be combined to form other optimality criteria that are more appropriate for specific applications. We analyzed the optimality criterion for parameter estimation under the simplifying assumption that the feature manifold is linear. We showed that, for a fairly general noise model, the optimal weighting function assigns a weight to each pixel that is inversely proportional to the variance of the noise at that pixel.

Extending this analysis to the general non-linear case for any of the optimality criteria proved intractable. Instead we proposed a numerical algorithm that can be used to find the optimal weighted L^2 norm for arbitrary parametric features and almost any conceivable optimality criterion. We applied this algorithm to three important features: the *step edge*, the *corner*, and the *symmetric line.* The results do indeed show that the center-most pixels are of little importance when estimating the parameters of the corner. They also show, as was widely believed, that the center-most pixels are the most important for the sub-pixel localization of a step edge. These experimental results demonstrate the major contribution of our work which is a general purpose method of automatically finding optimal weighting functions for parametric feature detectors.

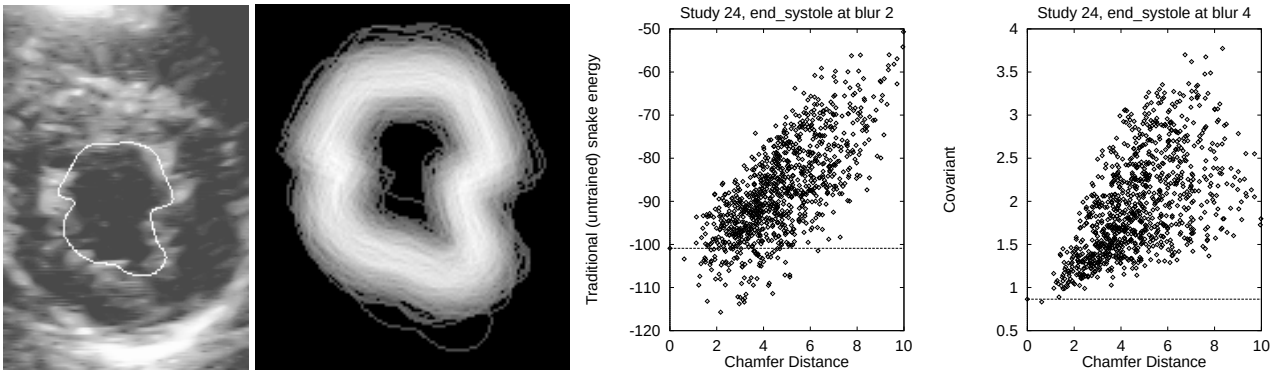


Figure 10: (a) *An echocardiogram with ground-truth blood pool outline;* (b) *1,000 perturbed versions of it (enlarged);* (c) *incorrectness* vs. *summed edge strengths;* (d) *incorrectness* vs. *learned Gaussian density of intensity and edge strength (5 parameters). Raw edge strength was even less correlated to correctness at blur 4 (not shown). The training (on 24 images, averaging 246 contour pixels each) did unusually well on this image, although with the CT data such improvement was typical. In this image, the trained model scored the ground truth better than all but one very close perturbation.*

13 Appearance Matching with Partial Data

Appearance matching based on linear subspace methods have found many important applications in computational vision including face recognition [Turk and Pentland-1991], real-time 3D object recognition [Nayar *et al.*-1996], and planar pose measurement [Krumm-1996]. Appearance matching methods generally use image brightness values directly, without relying on the extraction of low-level cues such as edges, local shading, and texture. This approach relies on the fact that brightness values can capture both geometric and photometric properties of the objects of interest. Formally, appearance matching is most often based on the generic linear model:

$$A\underline{x} = \underline{b} \qquad (1)$$

where, A is the design matrix that represents the subspace, $\underline{b}$ is the intensity array, and $\underline{x}$ are the subspace coordinates that characterize the image. Array $\underline{b}$ can contain the entire set of pixels in the image, or only a subset of the pixels. Respectively, A will consist of the entire set of rows or only the subset that corresponds to the pixel subset.

In many cases we would like to be able to recognize an image based on only part of its data. Such an example would be to recognize an occluded image. Clearly, occlusion degrades some of the brightness values. We would not like to use corrupted pixels for recognition. Occlusion is usually spatially localized, hence the subset used should be localized as well, within a square. We would like the window to be large and representative of the image. However, the probability of overlap between the window and the occluded region, which can be at a random location in the image, increases with the size of the window. Another application of subset selection would be to make recognition faster. This is because recognition time is linearly proportional to the size of the pixel set.

A number of attempts have been made to address recognition with occluded or partial data [Murase and Nayar-1995] [Moghaddam and Pentland-1995] [Krumm-1996] [Brunelli and Messelodi-1993] [Leonardis and Bischof-1996]. Though these approaches are interesting, none succeeds to address the underlying problems with adequate depth. The first three techniques select windows in an image with ad–hoc arguments. Further, they can only be applied to specific types of images. The last two initially select a small subset of pixels randomly and then prune it with iterative algorithms. However, an iterative algorithm is not guaranteed to converge to the desired solution. In addition, recognition based on very small subsets is not reliable.

Similar problems to the ones described above have been investigated more thoroughly in the general context of statistics [Hotelling-1944] [Ehrenfeld-1955] [Huber-1981] [Cook and Weisberg-1982]. Although useful, their results have limitations, for example, some assume that measurements can be repeated, others deal only with closed form mathematical expressions, or suggest algorithms practical for only small data sets.

Recently, we have derived optimal criteria [Hadjimetrious and Nayar-1998] for selection of a subset of pixels, through sensitivity analysis of the corresponding subset of rows of the design matrix A. For this, it turns out that, the correlation among the rows of A should be minimal. Furthermore, the rows

should have similar magnitudes. Our results have led to a number of selection criteria, which are used to implement two practical algorithms [Hadjimetrious and Nayar-1998]. The algorithms are widely applicable, and have been shown to select subspaces with near optimal properties. The first algorithm judiciously selects square windows. The second algorithm judiciously selects subsets of pixels from the entire image. The algorithms were tested with noisy images. They demonstrate superior recognition performance when compared to algorithms that select pixel subsets randomly. In addition, they substantially reduce recognition time with a relatively small decrease in recognition performance [Hadjimetrious and Nayar-1998].

13.1 Structural Analysis of Extended Video Sequences

In extended video sequences, individual frames are grouped into shots, which are defined by a sequence taken by a single camera. Similarly, related shots group into scenes, which are defined by a single dramatic event taken by a small number of related cameras. This hierarchical structure is deliberately constructed, as it is dictated by the limitations and preferences of the human visual and memory systems.

We have devised and demonstrated three novel high-level segmentation results derived from these considerations, some of which are analogous to those involved in the perception of the structure of music [Kender and Yeo-1998]. First and primarily, we have shown a method for measuring probable scene boundaries, by calculating a short-term human memory-based model of shot-to-shot "coherence". The detection of local minima in this continuous measure permits robust and flexible segmentation of the video into scenes, without the necessity of first aggregating shots into similarity clusters. Secondly and independently of the first, but also based on these memory models, we have shown a one-pass on-the-fly shot clustering algorithm. Third, we have shown partially successful results on the application of these two new methods to the next higher, "theme" or "act", level of video structure.

The work has been applied to several genres, including half hour sitcoms, and movie-length dramatic action films. These human perception-based methods are robust: although not designed for it, they also appears to apply equally well to the next lower level of detail, namely, the detection of the shot boundaries themselves.

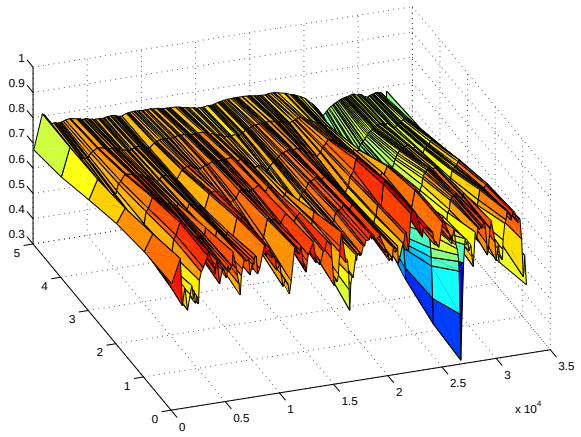


Figure 11: The coherence of a half-hour sitcom video. The X dimension indicates frame number, the Y dimension the log of the capacity of the memory buffer, and the Z dimension the coherence. Scene boundaries are given by local minima.

References

[Abrams, 1997] S. Abrams. *Sensor Planning in an active robot work-cell.* PhD thesis, Dept. of CS, Columbia Univ., Jan. 1997.

[Allen and Yang, 1998] P.K. Allen and Ruigang Yang. Registering, integrating, and building CAD models from range data. In *IEEE Intl. Conf. on Robotics and Automation*, 18-20 1998.

[Allen *et al.*, 1998] P.K. Allen, M. Reed, and I. Stamos. View planning for site modeling. In *Proc. DARPA IUW*, 1998.

[Baker and Nayar, 1998a] S. Baker and S. Nayar. Optimal weighting functions for feature detection. In *Proc. of the 1998 DARPA IUW*. 1998.

[Baker and Nayar, 1998b] S. Baker and S. Nayar. A theory of catadioptric image formation. In *Proc. of IEEE ICCV*, pages 35–42, 1998.

[Baker *et al.*, 1998] S. Baker, S. Nayar, and H. Murase. Parametric feature detection. *Intl. J. of Computer Vision*, 27(1):27–50, 1998.

[Bogner, 1995] S. Bogner. Introduction to panoramic imaging. In *Proc. of the IEEE SMC Conf.*, pages 3100–3106, Oct. 1995.

[Boult *et al.*, 1998] T. Boult, Chen Qian, Weihong Yin, Ali Erkin, Peter Lewis, Chris Power, and Ross Micheals. Applications of omnidirectional imaging: Multi-body tracking and remote reality. In *Proc. IEEE Workshop on Computer Vision Applications*, 1998.

[Boult, 1998a] T. Boult. Remote reality. In *Proc. of ACM SIGGRAPH 1998*, 1998. Tech. Sketch.

[Boult, 1998b] T. Boult. Remote reality demonstration. In *Proc. of IEEE CVPR*, 1998.

[Boult, 1998c] T. Boult. Remote reality via omnidirectional imaging. In *Proc. of the DARPA IUW*, 1998.

[Brunelli and Messelodi, 1993] R. Brunelli and S. Messelodi. Robust estimation of correlation: An application to computer vision. IRST Tech. Report, 1993.

[Charles *et al.*, 1987] J.R. Charles, R. Reeves, and C. Schur. How to build and use an all-sky camera. *Astronomy Magazine*, April 1987.

[Chatterjee, 1993] S. Chatterjee. *Markov Random Fields: Theory and Applications*, chapter Classification of natural textures using Gaussian Markov random fields. Academic Press, 1993.

[Chiang, 1998] M.C. Chiang. *Imaging-Consistent Warping and Super-Resolution*. PhD thesis, Dept. of CS, Columbia Univ.. Advisor: T.Boult, May 1998.

[Cook and Weisberg, 1982] R. Cook and S. Weisberg. *Residuals and Influence in Regression*. Chapman and Hall, 1982.

[Dana and Nayar, 1998] K.J. Dana and S.K. Nayar. Histogram model for 3d textures. In *CVPR*, 1998.

[Dana *et al.*, 1996] K.J. Dana, B. van Ginneken, S.K. Nayar, and J.J. Koenderink. Reflectance and texture of real-world surfaces. Tech. report, Columbia Univ. CS, 1996.

[Dana *et al.*, 1997] K.J. Dana, B. van Ginneken, S.K. Nayar, and J.J. Koenderink. Reflectance and texture of real world surfaces. In *CVPR*, 1997.

[Ehrenfeld, 1955] S. Ehrenfeld. On the efficiency of experimantal designs. *Annals of Mathematical Statistics*, 26:247–255, 1955.

[Fenster and Kender., 1998] S. Fenster and J.R. Kender. Sectored snakes: Evaluating learned-energy segmentations. In *Intl. Conf. on Computer Vision*, pages 420–426, 1998.

[Goshtasby and Gruver, 1993] A. Goshtasby and W.A. Gruver. Design of a single-lens stereo camera system. *Pattern Recognition*, 26(6):923–937, 1993.

[Hadjimetrious and Nayar, 1998] E. Hadjimetrious and S. Nayar. Appearance matching with partial data. In *IUW*, Monetery, CA, Nov. 1998. DARPA.

[Hartley, 1985] R. Hartley. A Gaussian-weighted multiresolution edge detector. *Computer Vision, Graphics, and Image Processing*, 30:70–83, 1985.

[Hecht and Zajac, 1974] E. Hecht and A. Zajac. *Optics*. Addison-Wesley, 1974.

[Hong, 1991] J. Hong. Image based homing. In *Proc. of the IEEE Intl. Conf. on Robotics and Automation*, 1991.

[Hotelling, 1944] H. Hotelling. Some improvements in weighing and other experimental techniques. *Annals of Mathematical Statistics*, 15:297–306, 1944.

[Huber, 1981] P.J. Huber. *Robust statistics*. Wiley, 1981.

[Hueckel, 1971] M.H. Hueckel. An operator which locates edges in digitized pictures. *J. of the Association for Computing Machinery*, 18(1):113–125, 1971.

[Hueckel, 1973] M.H. Hueckel. A local visual operator which recognizes edges and lines. *J. of the Association for Computing Machinery*, 20(4):634–647, Oct. 1973.

[Hummel, 1979] R.A. Hummel. Feature detection using basis functions. *Computer Graphics and Image Processing*, 9:40–55, 1979.

[Irani and Peleg, 1991] Michal Irani and Shmuel Peleg. Improving resolution by image registration. *CVGIP: Graphical Models and Image Processing*, 53(3):231–239, 1991.

[Irani and Peleg, 1993] Michal Irani and Shmuel Peleg. Motion analysis for image enhancement: Resolution, occlusion, and transparency. *JVCIP*, 4(4):324–335, 1993.

[Kashyap, 1984] R.L. Kashyap. Characterization and estimation of two-dimensional arma models. *IEEE Tran. on Information Theory*, 1984.

[Kender and Yeo, 1998] J.R. Kender and B.L. Yeo. Video scene segmentation via continuous video coherence. In *IEEE CVPR*, 1998.

[Koenderink and van Doorn, 1996] J.J. Koenderink and A.J. van Doorn. Illuminance texture due to surface mesostructure. *J. Optical Soc. Am. A*, 1996.

[Koenderink *et al.*, 1996] J.J. Koenderink, A.J. van Doorn, and M. Stavridi. Bidirectional reflection distribution function expressed in terms of surface scattering modes. In *ECCV*, 1996.

[Krumm, 1996] J. Krumm. Eigenfeatures for planar pose measurement of partially occluded objects. In *IEEE CVPR*, pages 55–60, 1996.

[Leonardis and Bischof, 1996] A. Leonardis and H. Bischof. Dealing with occlusions in the eigenspace approach. In *IEEE CVPR*, pages 453–458, 1996.

[Leung and Malik, 1997] T. Leung and J. Malik. On perpendicular texture or: Why do we see more flowers in the distance. In *CVPR*, 1997.

[Moghaddam and Pentland, 1995] B. Moghaddam and A. Pentland. Probabilistic visual learning for object detection. In *IEEE ICCV*, pages 786–793, 1995.

[Morgenthaler, 1981] D.G. Morgenthaler. A new hybrid edge detector. *Computer Graphics and Image Processing*, 16:166–176, 1981.

[Murase and Nayar, 1995] H. Murase and S. Nayar. Image spotting of 3d objects using parametric eigenspace representation. In *Proc. of the 9th Scandinavian Conf. on Image Analysis*, pages 323–332, 1995.

[Nalwa and Binford, 1986] V.S. Nalwa and T.O. Binford. On detecting edges. *IEEE Transactions on PAMI*, 8(6):699–814, 1986.

[Nalwa, 1996] V.S. Nalwa. A true omnidirectional viewer. Tech. report, Bell Laboratories, Holmdel NJ. USA, February 1996.

[Nayar and Boult, 1998] S.K. Nayar and T.E. Boult. Omnidirectional vision technology. In *Proc. of the DARPA IUW*, 1998.

[Nayar *et al.*, 1991] S.K. Nayar, K. Ikeuchi, and T. Kanade. Surface reflection: physical and geometrical perspectives. *IEEE Tran. on PAMI*, 13(7):611–634, 1991.

[Nayar *et al.*, 1996] S. Nayar, S.A. Nene, and H. Murase. Real–time 100 object recognition system. In *IEEE Intl. Conf. on Robotics and Automation*, pages 2321–2325, 1996.

[Nayar, 1988] S. Nayar. Sphereo: Recovering depth using a single camera and two specular spheres. In *Proc. of SPIE: Optics, Illumination, and Image Sensing for Machine Vision II*, 1988.

[Nayar, 1997] S. Nayar. Catadioptric omnidirectional camera. In *Proc. of CVPR 1997* , pages 482–488, 1997.

[Nene *et al.*, 1994] S.A. Nene, S. Nayar, and H. Murase. Software library for appearance matching (slam). In *DARPA IUW*, 1994.

[O'Gorman, 1978] F. O'Gorman. Edge detection using Walsh functions. *Artificial Intelligence*, 10:215–223, 1978.

[Oren and Nayar, 1995] M. Oren and S. K. Nayar. Generalization of the lambertian model and implications for machine vision. *Intl. J. of Comp. Vision*, 14:227–251, 1995.

[Peri and Nayar, 1997] V. Peri and S. Nayar. Generation of perspective and panoramic video from omnidirectional video. In *Proc. of the 1997 DARPA IUW*, 1997.

[Reed and Allen, 1998] M. Reed and P.K. Allen. 3-D modeling from range imagery: An incremental method with a planning component. *Image and Vision Computing*, (to appear) 1998.

[Reed *et al.*, 1997] M. Reed, P.K. Allen, and I. Stamos. Automated model acquisition from range images with view planning. In *IEEE CVPR*, 1997.

[Reed, 1998] M. Reed. *Solid Model Acquisition from Range Imagery*. PhD thesis, Dept. of CS, Columbia Univ., May 1998.

[Rees, 1970] D.W. Rees. Panoramic television viewing system. US Patent No. 3,505,465, April 1970.

[Rohr, 1992] K. Rohr. Recognizing corners by fitting parametric models. *Intl. J. of Computer Vision*, 9(3):213–230, 1992.

[Stamos and Allen, 1998] I. Stamos and P.K. Allen. Interactive sensor planning. In *CVPR*, pages 489–494, 1998.

[Stavridi and Koenderink, 1997] M. Stavridi and J.J. Koenderink. Surface bidirectional reflection distribution function and the texture of bricks and tiles. *Applied Optics*, 1997.

[Tagare and DeFigueiredo, 1993] H.D. Tagare and R.J.P. DeFigueiredo. A framework for the construction of reflectance maps for machine vision. *CVGIP: Graphical Models and Image Processing*, 57(3):265–282, 1993.

[Turk and Pentland, 1991] M. Turk and A. Pentland. Eigenfaces for recognition. *J. of Cognitive Neuroscience*, 3(1), 1991.

[van Ginneken *et al.*,] B. van Ginneken, J.J. Koenderink, and K.J. Dana. Texture histograms as a function of irradiaion and viewing direction. IJCV-submitted.

[Wang and Healey, 1996] L. Wang and G. Healey. Illumination and geometry invariant recognition of texture in color images. In *CVPR*, 1996.

[Wolff, 1994] L.B. Wolff. A diffuse reflectance model for smooth dielectrics. *J. Optical Soc. Am. A*, 1994.

[Yagi and Kawato, 1990]
Y. Yagi and S. Kawato. Panoramic scene analysis with conic projection. In *Proc. of the Intl. Conf. on Robots and Systems*, 1990.

[Yamazawa *et al.*, 1993] K. Yamazawa, Y. Yagi, and M. Yachida. Omnidirectional imaging with hyperboloidal projection. In *Proc. of the Intl. Conf. on Robots and Systems*, 1993.

[Yamazawa *et al.*, 1995] K. Yamazawa, Y. Yagi, and M. Yachida. Obstacle avoidance with omnidirectional image sensor HyperOmni Vision. In *Proc. of the IEEE Intl. Conf. on Robotics and Automation*, pages 1062–1067, 1995.

[Zhang and R. S. Blum,] Z. Zhang and R. S. Blum. A categorization and study of multiscale-decomposition-based image fusion schemes. Submitted to Proc. of IEEE.

[Zhang and R. S. Blum, 1997] Z. Zhang and R. S. Blum. Multisensor image fusion using a region-based wavelet transform approach. In *Proc. of the DARPA IUW*, pages 1447–1451, 1997.

[Zhang and R. S. Blum, 1998] Z. Zhang and R. S. Blum. On estimating the quality of noisy images. In *Proc. IEEE ICASSP98*, vol. 5, pages 2897–2900, 1998.

[Zhang and R. S. Blum, 97] Z. Zhang and R. S. Blum. A region-based image fusion scheme for concealed weapon detection. In *Proc. 31th Annual Conf. on Information Sciences and Systems*, pages 168–173, 97.

[Zucker and Hummel, 1981] S.W. Zucker and R.A. Hummel. A three-dimensional edge operator. *IEEE Transactions on PAMI*, 3(3):324–331, 1981.

This paper appeared in the 1998 DARPA Image Understanding Proceeding and is copyrighted.