Proceedings 1997 Image Understanding Workshop, May 11-14, 1997, New Orleans, LA

Integrated Vision and Sensing for Human Sensory Augmentation

Takeo Kanade and Vladimir Brajovic
School of Computer Science, Carnegie Mellon University,
5000 Forbes Avenue, Pittsburgh PA 15213
brajovic@cs.cmu.edu, tk@cs.crnu.edu
http://www.cs.cmu.edu/afs/cs/usr/brajovic/www/lab/vlsi.html

Abstract[1]

The Carnegie Mellon University MURI project sponsored by ONR performs multi-disciplinary research in integrating vision algorithms with sensing technology for low-power, low-latency, compact adaptive vision systems. These are crucial features necessary for augmenting the human sensory system and enabling sensory driven information delivery. The project spans four subareas ranging from low to high level of vision: (1) smart filters, based on the Acousto-Optic Tunable Filter (AOTF) technology; (2) computational sensor methodology, which integrates raw sensing and computation by means of VLSI technology; (3) neural-network based saliency identification techniques for identifying the most useful information for extraction and display; and (4) visual learning methods for automatic signal-to-symbol mapping.

1. Introduction

Automated vision and sensing research has made great strides in the last 30 years. Yet vision systems still lack attributes shared by most successful mass-market technologies — small size, low cost, low power and highly reliable performance. If computer vision processing had these characteristics, the potential applications would be nearly endless. Examples include: wearable smart vision systems for enhancing solder's situation awareness in the battlefield; head-up display vision enhancement systems for driving in bad weather and low visibility conditions; head-up display field telemedicine systems, and others. All these applications share common features — the applications are mobile and interact with the human sensory system. While today these scenarios are mostly futuristic speculations, some of the technologies they require have been partially demonstrated. Our research further develops these emerging technologies, and brings these visions closer to reality.

The CMU MURI project performs multi-disciplinary research spanning all levels of vision and sensing: dynamically tunable acousto-optic multispectral imaging [Brajovic and Kanade, 19971; VLSI-based computational sensors [Brajovic and Kanade, 1997]; neural network saliency detection [Pomerleau, 1997]; automatic visual acquisition of object models [Hebert et al., 1997]; domain-independent evolution-based learning for signal-to-symbol mapping [Glickman and Sycara, 19971; and learning coordination among multiple signal-to-symbol mapping agents [Teller and Veloso, 1997b]. We believe that the tight integration of vision algorithms and sensing technology will result in low-power, low-latency, compact, adaptive vision systems crucial for effective human sensory augmentation.

1.1. CMU Approach

The separation of sensing and processing, as a natural consequence of a conventional vision system comprising a camera and computer, results in several deficiencies. The two most critical features missing in this sens-and-process paradigm are *low latency processing* and *sensory adaptation.*

1. This research has been sponsored in full or in part by Office of Naval Research (ONR) under Contract N00014-95-1-0591. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies. either expressed or implied, of ONR or the U.S. Government.

Latency, or reaction time, is the time that a system takes to react to an event. The primary sources of latency in vision systems are the: ***data transfer bottleneck*** caused by the need to transfer an image from the camera to the processor, and the ***computational load bottleneck*** caused by the processor's inability to quickly handle a large amount of visual data. The detrimental effects of both bottlenecks scale–up with the image size. Often, the system "receives" the image data too late to cope with fast events or to provide sensory feedback to a human user. For example, during the frame time of a conventional camera, a person's gaze direction can shift by 18 degrees. To ensure that the viewer feels comfortable and natural in head–mounted display applications, for example, delays must be less than 10 to 20 msec.

Another aspect presently missing in machine vision is top-down sensory adaptation. Complex ad–hoc algorithms that try to extract relevant information from inadequate sensor data are inevitably unreliable. In fact, time and time again it has been observed that using the most appropriate sensing modality or setup allows recognition algorithms to be far simpler and more reliable. For example, the concept of active vision proposes to control the geometric parameters of the camera (e.g., pan, tilt, etc.) to improve the reliability of the perception [Aloimonos, 19921. It has been shown that initially ill–posed problems can be solved after the top-down adaptation of the camera's pose has acquired new, more appropriate image data. However, adjusting geometric parameters is only one level at which adaptation can take place. Another example of adaptation is multi–spectral imaging, which can eliminate confusion by providing sensor images appropriate **for** the task. Acquisition of appropriate sensor bands adaptively, however, is often difficult since most multi–spectral imaging devices have fixed spectral sensitivity, while the appropriate wavelengths to process vary as conditions and the task change. Therefore, a system that can adjust its operation at all levels, even down to the point of sensing, would be far more adaptive than one that tries to cope with the variations at the "algorithmic" or "motoric" level alone.

The two major shortcomings of the sense–and–process approach which are outlined above, along with the fact that this approach naturally leads to bulkier and less cost-effective systems, suggest that an alternative is needed. We are establishing a new paradigm in which sensing and vision processing are tightly coupled for fast, time-critical, adaptive operation.

The following sections describe basic techniques and technologies that the CMU team has worked on; we believe these are necessary for the success of a low–latency adaptive vision system for human sensory augmentation.

2. Multi-Spectral Imaging Filters

Contributors: L.J.Denes, M. Gottlieb, B.Kaminsky, P. Metes, Z.K. Kun, M. Capizzi, J. Hibner, D. Purta, A.M. Guzman

This program task incorporates the spectral (color) dimension into the visual reasoning process. A programmable optical filter is utilized at the system's front end to reduce the computational load and its resulting bottlenecks in future automated vision systems. Filtering the incoming scene according to its spectral composition can remove a large amount of undesirable background clutter prior to higher level processing. Figure 1 is a schematic representation of the process. Enhanced performance is anticipated in a variety of applications, including human sensory augmentation systems for driver assistance. Because of its ability to extract and track objects, this vision system will more closely mimic the human observer.

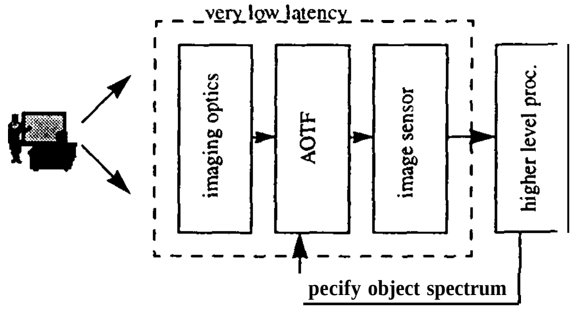

Figure 1: Object recognition using color discrimination.

We have assembled a multi-spectral imaging system operating in the visible to near IR range utilizing an existing acousto-optic tunable filter (AOTF). This configuration has been characterized, yielding design optimization information. Critical data include spatial and spectral resolution, out-of-band rejection, efficiency, field of view, and bandwidth. The design goal is efficient operation over nearly two octaves of wavelength, and superior image quality. Two major issues were successfully addressed. The first relates to the method of applying the multiple electrical RF control signals to the **AOTF** transducer to fully exploit the multispectral capability. Several approaches were analyzed, including multiple oscillators, spread spectrum techniques, and the use of an arbitrary waveform generator. Recent work has confirmed that the arbitrary waveform generator provides all of the flexibility required with no serious disadvantages. In addition, it is readily adaptable to computer control. The second issue addressed is how to best achieve object identification using color signature information. A fundamental issue arises because any background object with a broadband color distribution, e.g., white, will include the desired signature within its spectrum. Thus, these background objects may not be discriminated against the target object. To address this problem, we developed a processing technique using two video frames, in which the first frame grab contains a multispectral image whose spectral content lies outside the target color signature. This frame is inverted and then used as a spatial mask over the entire scene. The second frame grab includes only the target color signature and provides us with a gray scale. By using an appropriate threshold, the target image alone is displayed against a black background. Tests of laboratory scenes give encouragingly good results.

3. Computational Sensors for Low–Latency Adaptive Vision

Contributors: Vladimir Brajovic and Takeo Kanade

The computational sensor paradigm [Kanade and Bajcsy, 1993] has the potential to greatly reduce latency and provide top-down sensory adaptation to the vision system. By integrating sensing and processing on a VLSI chip, both transfer and computational bottlenecks can be alleviated: on–chip routing provides high throughput transfer; an on-chip processor could implement massively–parallel fine–grain computation providing high processing capacity which readily scales up with the image size. In addition, the tight coupling between processor and sensor allows for efficient top-down feedback that can control and adjust the sensor for further acquisition based on the preliminary results of the processing.

Our recent work has been concerned with efficient implementation of global operations over large groups of image data using a computational sensor paradigm [Brajovic and Kanade, 19941. Global operations are important because: (1) in perception, each decision is a kind of global, or overall, conclusion necessary for the coherent interaction with the environment, and (2) unlike local operations (e.g., filtering) which produce large amounts of preprocessed image data, global operations produce ***a few*** quantities for the description of the environment which can be ***quickly*** transferred and/ or processed to produce an appropriate action for a machine. The main difficulty with implementing global operations comes from the necessity to bring together all or most of the data in the input data set. We have formulated two mechanisms for implementing global operations in computational sensors: (1) ***sensory attention [Brajovic and Kanade, 1997]***, and (2) ***intensity–to–time processing paradigm [Brajovic and Kanade, 1996]***.

The ***sensory attention*** is based on the premise that salient features within the retinal image represent important global features of the entire image. This premise is attractive for two reasons. First, the main argument that has been used to explain the need for selective visual attention in brains is that, as there exist some kind of processing and communication limitations in the visual system, the same exists in machines. Attention “funnels” only relevant information and protects the limited communication and processing resources from the information overload. Indeed, the importance of selecting the relevant information from an image is now widely acknowledged in machine vision; some forms of attention mechanisms (e.g. selecting a correctly sized window within the image) are often employed in practical applications. Second, it has been shown that the visual attention improves performance, and is needed for maintaining coherent behavior while interacting with the environment (i.e. attention–for–action) [Allport, 19891. Location of such attention must be maintained in the environmental coordinates, thus maintaining coherence under ocular and head motion [Milanese, 19931. Unlike eye movement (i.e., ***overt*** shifts), the attention shifts (i.e. ***covert*** shifts) do not require any motor action, but occur internally on a fixed retinal image. For this reason, attention shifts are faster and play an important role in low–latency vision systems.

We have implemented sensory attention by fabricating and testing a ***tracking computational sensor.*** This track sensor optically receives a saliency map and continuously selects and tracks the peaks in the map. The location and intensity of the selected saliency peaks is reported on few output pins with low latency. These quantities are also used internally in a top-down fashion to aid tracking of the attended location. The chip is a 28 x 28 array of 60μ x 60μ cells, and is fabricated on a 2.2mm x 2.2mm die. When tracing bright, well-defined features, the sensor tracks targets moving across the retina at about 6900 cells/second.

The ***intensity–to–time processing paradigm*** is based on the notion that stronger signals elicit responses before weaker ones, thus allowing a global processor to make decisions based on only a few inputs at a time. The key is that some preliminary decisions about the retinal image can be made as soon as the first responses are received. The intensity–to-time processing paradigm is used for the VLSI implementation of a ***sorting computational sensor*** — a sensor that sorts input stimuli by their intensity as they are being sensed. The chip detects an image focused thereon and computes an image of indices. During the computation, the chip computes a cumulative histogram — one global quantity of the detected image — and reports it with low-latency on one of the pins before the image is ever read out. The cumulative histogram is used internally in a top-down fashion to generate indices within each pixel. The image of indices has a uniform histogram which has several important properties: (1) the contrast is maximally enhanced, (2) the available dynamic range of readout circuitry is equally utilized, i.e., the values read out from the chip use available bits most efficiently, and (3) the image of indices never saturates, and preserves the same range (e.g., from 1 to N) under varying conditions in the environment.

The adaptation of the dynamic range of the sorting sensor is illustrated in Figure 2 showing sequence of 93 images provided by the sorting sensor. By observing the wall in the background, we can see the effects of adaptive dynamic range: even though the physical wall does not change the brightness, it appears dimmer in those frames in which bright levels are taken by pixels which are physically brighter (e.g., subject's face and arm). When the subject turns and fills the filed-of–view with dark objects (e.g., hair) the wall appears brighter since it is now taking higher indices. Also, note that the maximum contrast is maintained in all the images since all images of indices have uniform histogram.

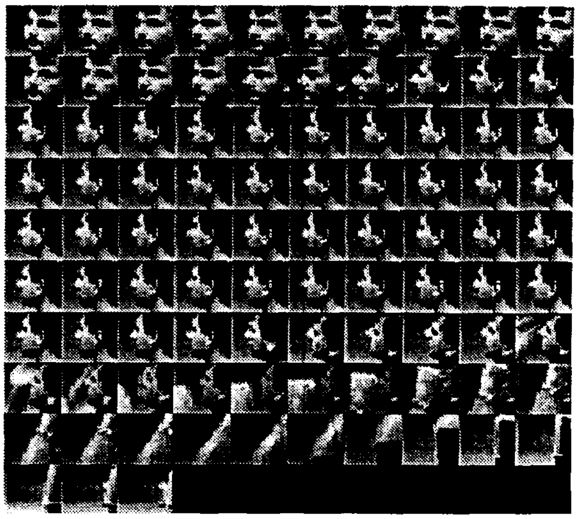

Figure ***2:*** Sequence of images of indices computed by the sorting sensors.

We continue to work on an improved sorting computational sensor with smaller pixels and a larger array. We also continue to work on developing new sensors based on the intensity-to-time processing paradigm. We have designed, and recently received a prototype of, a self-contained eye tracking sensor. We plan to test the sensor and apply it in several scenarios. In the near term, we will begin interfacing some of our computational sensors with smart AOTF filters.

4. Visibility Estimation from a Moving Vehicle

Contributor: Dean Pomerleau

Reduced visibility caused by fog, rain, snow, darkness and glare is a frequent contributing factor to traffic accidents [Najm et al., 1995]. In fact, some of the most serious of all highway incidents, sometimes involving dozens or even hundreds of vehicles, occur when reduced visibility conditions result in a chain reaction of crashes. Technologies typically employed to estimate visibility include: transmissometers, which measure the transmittance of the atmosphere over a baseline distance; and nephelometers, which measure the scattering coefficient caused by suspended particles in an air sample [National Weather Service 19961. Unfortunately, these systems suffer from several drawbacks as they are not always estimating visibility from the driver's point of view. The only way to automatically estimate the cumulative influence of these factors on the driver's ability to see potential obstacles ahead is to employ a sensing system which reasonably matches the driver's perceptual characteristics. We developed a system that accomplishes this match by using a CCD video camera pointing out the windshield of the vehicle, and processing the same features processed by the human driver to estimate visibility.

Manual visibility estimates are typically made by attempting to detect high contrast targets at various known distances. The farthest distance at which a target can be reliably detected is considered the visibility distance. Ideally, an automated visibility estimation system should work the same way. Unfortunately, it is very difficult to consistently find high contrast targets at various known ranges from a moving vehicle. Even the features that are supposed to be consistent on a roadway, the lane markings, vary greatly in their appearance, and are in fact frequently missing or obscured. The Rapidly Adapting Lateral Position Handler (RALPH) system [Pomerleau and Jochem, 1996] overcomes this difficulty when detecting the position and curvature of the road ahead in camera images by utilizing whatever features are visible on the roadway. These features may include lane markings, road/shoulder boundaries, tracks left by other vehicles, and even subtle pavement discolorations like the oil stripe down the lane center when necessary. Our visibility estimation system exploits RALPH's ability to find and track arbitrary road features. In short, the system estimates visibility by measuring the attenuation of contrast between consistent road features at various distances ahead of the vehicle.

The visibility estimation algorithm performs well under a wide variety of conditions. The rank ordering of six conditions tested corresponds reasonably well to one's intuitive notion of how difficult it is to see in these situations. Live vehicle tests in fog still need to be conducted (fog is rare in Pennsylvania, particularly during the winter when these experiments were conducted). However, the results from the simulated fog experiments and the live daytime tests in rainy conditions suggest that the algorithm should perform well, and report significantly reduced visibility under foggy conditions.

While all the work reported here has been done with a standard black and white CCD camera, we are investigating the potential for using alternative sensors for improved performance. For example, a high-dynamic range camera, such as a VLSI sorting computational sensor, would respond more like the human eye in extreme lighting conditions, and could therefore provide better visibility estimates. Another possibility would be to combine this visibility estimation technique with smart AOTFs for multispectral imaging. By testing the visibility at different wavelengths, it may be possible to select the best wavelength(s) for operation under the current conditions.

5. Multi–Agent Learning for Signal Classification in Vision

Contributors: Astro Teller and Manuela Veloso

A wide variety of machine learning mechanisms create multiple models that must be reconciled, chosen among, or in some cases, ***orchestrated*** In its most general form, this orchestration problem can be seen as part of the multi-agent learning problem.

There are many cases in which a task to be approached with machine learning techniques can be, or must be, solved in more that one "piece." Learning a team of robotic soccer players is a good example of a task that could conceivably be done as a single agent, but lends itself very naturally toward learning sub-solutions and then (or in addition) learning to ensure the mutual suitability of these sub-solutions. This insurance of mutual suitability is the ***orchestration problem.***

Evolutionary computation is a natural machine learning environment in which to find many, behaviorally distinct models. We focus on PADO, a evolutionary computation framework designed specifically for signal classification (e.g., [Teller and Veloso, 1997b]). As a process of divide and conquer, PADO evolves multiple pools of sub-solutions and then orchestrates one or more learned models from each pool.

The question we investigate is: "What opportunities are there for learning in the orchestration process and how much improvement can this learning provide?' While answering this question, our research demonstrated several things [Teller and Veloso, 1997b]. First, specific experiments on distinct signals demonstrated the feasibility of PADO's divide and conquer strategy; the failure of the evolved orchestration procedure suggested PADO's preferability to unconstrained learning. Second, the experiments provided a specific justification for maintaining a population; orchestration puts the options a population provides to good use. And finally, this work introduced specific techniques for orchestration learning and, through their successful application, demonstrated that orchestration is an important issue and that learned orchestration can provide dramatic generalization improvements.

6. Adaptive Acquisition of Search Control Knowledge in the Evolution of Face Recognition Neural Networks

Contributors: Matthew Glickman and Katia Sycara

Search algorithms for signal–to–symbol matching patterned after biological evolution are attractive for use in domains such as vision that have complex search spaces for a number of reasons. These include: (1) Their application does not explicitly require deep insight into the domain; (2) They are relatively straightforward to paralyze; and (3) their natural analog has resulted in entities of extraordinary complexity and robustness. However, the search performance in any particular domain is highly dependent on the interaction between the chosen representation of the space and the specific search operators employed. For evolutionary algorithms in particular, this interaction is a poorly understood process, leaving practitioners with few guidelines as to how to make the right choices to yield good performance.

One popular approach to improving the performance of search in a particular domain is to seek to incorporate pre-existing knowledge of the domain into the operators and representation. However, this approach is problematic for evolutionary search because of the aforementioned opacity of the interaction between the operators and the representation. This difficulty, popularly known as "the representation problem," is only compounded in more complex domains, presenting a formidable obstacle to the application of artificial evolution in precisely those domains in which they may be of the greatest utility.

Therefore, rather than seeking to find how pre-existing domain knowledge can be best exploited by evolution, our research is directed toward the automatic acquisition of such knowledge in operational form. The experiments reported herein demonstrate that information about a particular domain generated over the course of evolutionary search can be extracted, analyzed, and then employed to improve search in future runs.

The space explored is the weight space of fixed-topology, feed-forward artificial neural networks (ANNs) for face recognition. Over the course of adaptation, weight vectors, along with their self-adapted, variable mutation rate, were collected. These data were then used to train another ANN to predict the appropriate mutation rate for a given weight vector for the face-recognition domain in general. Finally the mutation rate-prediction networks were used to drive evolution on another face recognition task, resulting in networks with improved generalization performance.

Our preliminary results indicate that this approach is reliably feasible. Due to the fact that (1) the specific weight-vector/mutation-rate pairs chosen for training were selected via a simple, Darwinian selection process, and (2) that the target mutation rates contained in these data had also been adapted via this same selection process, the results reported here indicate that simple Darwinian selection is sufficient to generate a training signal from which domain/search-control knowledge may be extracted. This result indicates a promising direction for the successful application of artificial evolution in complex domains such as image understanding.

7. Visual Learning for Landmark Recognition

Contributors: Martial Hebert, Katsushi Ikeuchi, Yukata Takeuchi, Patrick Gros

Recognizing landmarks is a critical task for interaction of a machine with the environment. Landmarks are used for building maps of unknown environments. In this context, the traditional recognition techniques based on strong geometric models cannot be used. Rather, models of landmarks must be built from observations obtained using image-based techniques. This section describes building image-based landmark descriptions from sequences of images, and then recognizing the landmarks. This approach also addresses the more general problem of identifying groups of images with common attributes in sequences of images. We show that, with the appropriate domain constraints and image descriptions, this can be done using efficient algorithms.

Recognizing landmarks in sequences of images is a challenging problem for a number of reasons. The appearance of any given landmark varies substantially from one observation to the next. In addition, to variation due to different aspects, illumination change, external clutter, and changing geometry of the imaging devices are other factors affecting the variability of the observed landmarks. Finally, it is typically difficult to use accurate **3D** information in landmark recognition applications. For those reasons, it is not possible to use many of the object recognition techniques based on strong geometric models.

The alternative is to use image-based techniques in which landmarks are represented by collecting images which are supposed to capture the "typical" appearance of the object. The information most relevant to recognition is extracted from the collection of raw images and used as the model for recognition. This process is often referred to as "visual learning."

Progress has been made recently in developing such approaches. For example, in object modeling [Gross et al.], **2D** or **3D** model of objects are built for recognition applications. An object model is built by extracting features from a collection of observations. The most significant features are extracted for the entire set and are used in the model representation. Extensions to generic object recognition were presented recently [Carlsson, 19961.

Other recent approaches use the images directly to extract a small set of characteristic object images which are compared with observed views at recognition time. For example, the eigen-images techniques are based on this idea.

Those approaches are typically used for building models of a single object observed in isolation. In the case of landmark recognition for navigation, there is no practical way to isolate the object in order to build models. Worse, it is often not known in advance which of the objects observed in the environment would constitute good landmarks. Visual learning must therefore be able to identify groups of images corresponding to "interesting" landmarks and to construct models amenable to recognition out of raw sequences of images.

A similar problem, although in a completely different context, is encountered in image indexing, where the main problem is to store and organize images to facilitate their retrieval [Lamiroy and Gros, 19961 [Schmid and Mohr, 1996]. The emphasis in this case is on the kind of features used and the types of requests that can be made by the user. For image retrieval, actual systems (QBIC, JACOB, Virage...) are closer to smart browsing than to image recognition. Using criteria such as color, shape, regions, etc., the systems search for images most similar to a given image. The user can then interact with the system to define which of these images seems the most interesting, and a new set of closer images is displayed.

Our system tries to combine those two categories of systems. In a training stage, the system is given a set of images in sequence. The aim of the training is to organize these images into groups based on similarity of feature distributions between images. The size of the groups obtained may be defined by the user, or by the system itself. In the latter case, the system tries to find the most relevant groups, taking the global distribution of the images into account. In a second step, the system is given new images, which it tries to classify as either one of the learned groups, or belonging to the category of unrecognized images. Figure 4 shows indentifying landmarks from a moving vehicle.

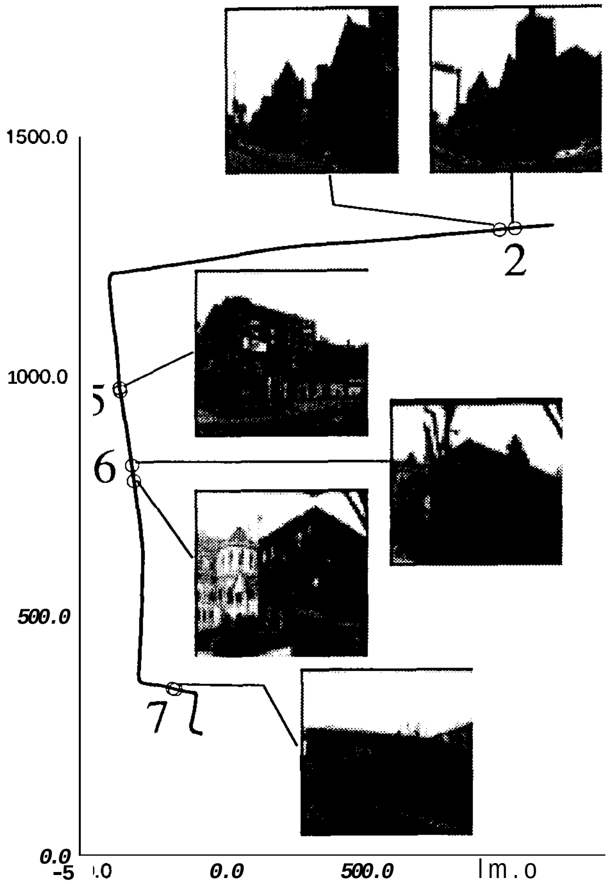

Figure 3: Overhead view of the path followed while collecting the images.(distances are indicated in meters.) Four landmarks are correctly identified, corresponding to groups 2, 5, 6, and 7 of the training sequence. Example images from the test sequence are shown for

The basic representation is based on distributions of different feature characteristics. All these different kinds of histograms are computed for the whole image and for a set of sub-images. Tests similar to Chi-square tests are used to compare these histograms and define a distance between images. This distance is then used to cluster the images in what are called groups. An agglomerative grouping algorithm is used at this stage. At each step of the algorithm, the clusters made are evaluated by an entropy-like function, whose maximum gives the optimal solution in a sense specified later. Each group is then characterized by a set of feature histograms. When new images are given to the system, it evaluates a distance between these images and the groups. The system determines to which group this image is the closest, and a set of thresholds is used to decide if the image belongs to this group.

The main goal of the work presented here was to explore the use of tools and methods in the field of image retrieval when applied to the problem of landmark recognition. It is clear that the global architecture of the system is close to that of object recognition systems [Gross et al.]: a training stage in which 3D shape, 2D aspects, or groups, are characterized is followed by a recognition stage in which this information is used to recognize the models, objects or groups in new images. The difference comes from the wide diversity of the images and from the groups which are not reduced to a single aspect of an object. The two challenging tasks which we concentrate on describing in the remainder of the paper are to define these groups more precisely as sets of images, and to automatically learn a characterization for each group: what remains invariant, what varies, and in which proportions.

8. Conclusion

CMU MURI performs cross -disciplinary research which will result in high performance vision systems adequate for “natural” human sensory augmentation and sensor driven information delivery. We are demonstrating progress in all levels of vision: from image formation and computational sensing to high level adaptive context–independent learning strategies. We believe that the tight integration of these techniques will provide opportunity for more efficient bottom–up and top-down

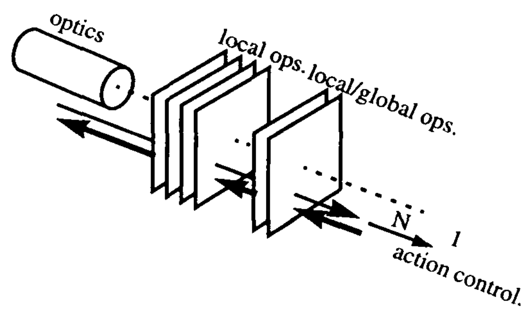

Figure 4: A vision system with tight integration of image formation, sensing and processing for adaptive low–latency applications.

control in vision processes which will result in low–power, low–latency, compact, reliable and adaptive vision systems (see Figure 4) crucial for effective human sensory augmentation.

References

[Allport, 1989] Allport, A. "Visual Attention", *Foundation of Cognitive Science,* M. Posner (ed.), MIT Press, 1989, pp. 631–682.

[Aloimonos, 1992] Aloimonos, J. (ed.), *Special Issue on Purposive, Qualitative, Active Vision,* CVGIP: Image Understanding, Vol. 56, No. 1, 1992.

[Brajovic and Kanade, 19941 Brajovic, V. and T. Kanade, "Computational Sensors for Global Operations", ***IUS*** *Proceedings,* pp. 621-630, 1994.

[Brajovic and Kanade, 19961 V. Brajovic and T. Kanade: "A Sorting Image Sensor: An Example of Massively Parallel Intensity-to-Time Processing for Low-Latency Computational Sensors," *Proceedings of the* ***1996*** *IEEE International Conference on Robotics and Automation,* Minneapolis, MN, April 1996, pp. 1638–1643.

[Brajovic and Kanade, 19971 Brajovic, V and T. Kanade: "Computational Sensors for Low–Latency Adaptive Vision," in IUW Proc., 1997.

[Carlsson, 1996] Carlsson, **S.** "Combinatorial Geometry for Shape Representation and Indexing," *Proc. International Workshop on Object Representation for Computer Vision.* Cambridge, England, April 1996.

[Brajovic and Kanade, 19971 Denes, l.J., M. Gottlieb, B.Kaminsky, P. Metes, Z.K. Kun, M. Capizzi, J. Hibner, D. Purta, and A.M. Guzman: "Multi-Spectral Imaging Filters," in IUW Proc., 1997.

[Glickman and Sycara, 19971 Glickman, M. and K. Sycara: "Adaptive Acquisition of Search Control Knowledge in the Evolution of Face Recognition Neural Networks," in IUW Proc., 1997.

[Gross et **al.]** P. Gros, **O.** Bournez and E. Boyer. *Using Local Planar Geometric Invariants to Match and Model Images of Line Segments.* To appear in Int. J. of Computer Vision and Image Understanding.

[Hebert et al., 19971 Hebert, M., K. Ikeuchi, Y. Takeuchi, P. Gros, "Visual Learning for Landmark Recognition," in IUW Proc., 1997.

[Kanade and Bajcsy, 19931 Kanade, T. and R. Bajcsy, "Computational Sensors: A Report from DARPA workshop", ***IUS*** *Proceedings,* 1993.

[Lamiroy and Gros, 19961 B.Lamiroy and P.Gros. Rapid Object Indexing and Recognition Using Enhanced Geometric Hashing. *Proc. of the 4th European* ***Conf.*** *on Computer Vision,* Cambridge, England, pages 59--70, Vol. 1, April 1996.

[Milanese, 1993] R. Milanese, *"Detecting Salient regions in an Image: from Biological evidence to computer implementation"*, Ph.D. Thesis, Dept. of Computer Science, U. of Genova, Switzerland, December 1993.

[Najm et al., 1995] Najm, W., Mironer, M. and Fraser, L. (1995) "Analysis of Target Crashes and ITS Countermeasure Actions". *Proc. of 1995 ITS America Annual Meeting,* pp. 931-940.

[National Weather Service 1996] *"Surface Weather Observations and Reports"* (1996) Federal Meteorological Handbook, 5th Edition, National Weather Service Publication FMH-1.

[Pomerleau and Jochem, 19961 Pomerleau, D. and Jochem, T. (1996) Rapidly Adapting Machine Vision for Automated Vehicle Steering. *IEEE Expert, Vol. 11,* ***No.*** *2.* pp. 19-27.

[Pomerleau, 19971 Pomerleau, D. "Visibility Estimation from a Moving Vehicle Using the RALPH Vision System", in IUW Proc., 1997.

[Schmid and Mohr, 19961 **C.** Schmid and R. Mohr. Combining Greyvalue Invariants with Local Constraints for Object Recognition. *Proceedings of the Conference on Computer Vision and Pattern Recognition,* San Francisco, California, USA. pages 872--877, June 1996

[Teller and Veloso, 1997b] Teller, A. and M. Veloso, "Learning Better Classification Teams," in IUW Proc., 1997.