AI Magazine Volume 12 Number 4 (1991) (© AAAI)

AAAI-90 Workshop on Qualitative Vision

William Lim, Philip Kahn, Daphna Weinshall, and Andrew Blake

■ *The AAAI-90 Workshop on Qualitative Vision was held on Sunday, 29 July 1990. Over 50 researchers from North America, Europe, and Japan attended the workshop. This article contains a report of the workshop presentations and discussions.*[1]

The AAAI-90 Workshop on Qualitative Vision was organized into seven sessions, with each session focusing on a specific topic. The first session was on the approaches and psychophysical bases of qualitative vision, addressing the question of what is qualitative vision. The second session presented work on motion and navigation. The topics of the next four sessions were qualitative shape extraction, qualitative feature extraction, qualitative vision systems and intelligent behavior, and high-level qualitative vision. To provide the participants with an opportunity to express their views on qualitative vision, the last session of the workshop was organized as an open discussion, again addressing the question of what is qualitative vision.

What Is QualitativeVision?

The first presentation in this session was given by J. Aloimonos on a purposive and qualitative vision system called MEDUSA. To avoid problems faced by a general-purpose vision system, Aloimonos proposed a more directed vision system based on task-specific processes. The system uses qualitative techniques (for example, comparing quantities or discrete classification) to implement specialized modules such as those for centering a moving object in the image, detecting the presence of a moving object, or detecting if an object is approaching the camera.

A robust qualitative cue based on motion parallax was discussed by A. Blake, R. Cipolla, and A. Zisserman. Motion parallax depends on the relative positions of two moving points. This cue has been found to be more robust than the absolute position of a single point (especially in the computation of relative depth, curvature on specular surfaces, and curvature on extremal boundaries). The rotation independence of motion parallax lends to its robustness.

A nonmetric representation of curved surfaces, that is, of psychophysical relevance, was presented by F. Reichel and J. Todd. The representation is based on order relations of depth and orientation among neighboring surface regions within an arbitrarily small neighborhood. This ordinal structure forms an intermediate representation level that balances the strengths and weaknesses of accurate but computationally expensive *metric structure* (as typified in the work on shape from shading) and the viewpoint-insensitive but less precise *nominal structure* (where a surface is defined in terms of a small number of shape primitives).

Motion parallax depends on the relative positions of two moving points.

Motion and Navigation

Three presentations were given in this session. Two approaches for detecting motion were discussed in the presentation by R. Nelson. The first approach uses *constraint ray filtering*, where an independently moving object is detected by looking for regions that exhibit motion inconsistent with the motion of the observer. Observer motion is estimated qualitatively by comparing against robust characteristics of a small set of prototype motion. The second method does not require the observer motion to be known. The motion characteristics of each point in the current image are predicted and compared with the actual motion recorded in the next frame. Inconsistencies between the predicted and recorded motion characteristics indicate potential regions of high acceleration. This method can only detect *animate motion*, where objects are undergoing rapid acceleration.

A different approach for estimating object motion from a sequence of images was presented by C. Lee and S. Samaddar. Their approach looks for a moving region that bounds the object. This region is obtained by thresholded differencing. The background is then recovered by replacing pixels, lying outside the moving region, in the first image with those in the current image. Once the background is recovered, the moving object is extracted by differencing with the recovered background. An object mask is then formed to be used for tracking the object through the image sequence. The change in the size of the mask is used for depth estimation.

The presentation by D. Coombs, T. Olson, and C. Brown was on how visually mediated gaze control can be used for segmenting an image. Their method for estimating disparity involves computing the echo delay obtained when the images (stereo pair) are spliced and treated as one image containing an echo. With this method, a real-time vergence algorithm can be developed by first looking for peak disparity and adjusting the angular velocities of the camera to compensate for it. Segmentation of an image is possible by varying the vergence of the camera through a range of vergence angles. The image is then run through a zero-disparity filter. Object regions are indicated by peaks in the plot of the number of vertical edges versus vergence angle.

Qualitative Shape Extraction

This session focused on the computation of similar qualitative shape features from motion. The presentation by D. Weinshall described how surface patches can be classified, directly from motion disparities, as convex, concave, hyperbolic, parabolic, and planar with a simple parallel computation. This computation also gives the direction of motion and requires a dense motion-disparity field. The computation is based on a simple result showing that three collinear points on a nonconvex (nonconcave) area undergoing rigid motion will bend toward (away from) the focus of expansion.

A similar result was shown by the work of N. da Vitoria Lobo and J. Tsotsos. Given three collinear image velocity measurements, it is possible to determine whether the points are in a collinear, convex, or concave relationship. Moreover, if the three image velocity measurements are not collinear, the authors showed how the relative pairwise depth for the three points can be computed. Both results, requiring known direction of three-dimensional motion, are carried over to the domain of shape from stationary stereo.

The presentation by A. Zisserman and R. Cipolla described the constraints placed on the surface-differential geometry by observing a surface curve from a sequence of positions. The first constraint is derived from the visibility of the curve. The second constraint is derived from a generalization of the three point results (discussed earlier) to image curves. In particular, the tracking of inflections on image curves (using snakes, for example) determines whether the surface is nonconvex or nonconcave.

Qualitative Feature Extraction

The main focus of this session was computational methods for the extraction of qualitative image features. A directed visual perception system that moves a sensor to find a target object was presented by L. Wixson and D. Ballard. In this system, a color histogram correlation is used for computing confidences that the modeled target object is contained within the current sensor field of view. These confidences establish an order over the gaze direction space that is used to command the next camera gaze point in the target search procedure.

R. Rao and R. Jain noted in their presentation that a central issue in computer vision is the transformation of image signals into symbolic representations that support reasoning. The authors discussed a method for the symbolic description of oriented textures based on differential methods. The techniques are used to symbolically describe texture in an image of fluid motion and turbulent shear flow fields.

The presentation by S. Haynes and R. Jain was on how relative order in depth can be extracted using senses of occlusion and motion, approximate direction of motion, and other rules. These relative depths provide a partial order over depth in space that can provide valuable visual guidance for navigation and perception. Local and occlusion assessments are first computed and then linked over time to establish relative depth relations.

Qualitative Vision Systems and Intelligent Behavior

Work on AI-flavored qualitative vision was presented in this session. The presentation by R. Howarth and A. Toal was on a project that attempts to build a record of vehicle movements over time. Their approach, based on the work of K. Forbus and M. Fleck, is to represent both space and time in terms of a cellular array and chart transitions between cells.

In their presentation, H. Narayanan and B. Chandrasekaran also used cellular representation. In this case, it is used for representing kinematics, for example, of gear trains. This approach is taken in the hope that situations that are too complex for symbolic analysis might yield to more direct spatial representation. The work presented in this area is still preliminary, with no one yet claiming to demonstrate mature reasoning systems driven by visual input.

E. Adelson and P. Anandan presented an interesting view of a classic psychophysical problem—the perception of transparency. Previous theories have relied on numeric tests derived from the multiplicative effect that successive layers of transparent material have on light intensity. However, the observation of order relations between intensities at an X-junction created by overlapping transparent materials proves to be a powerful predictor for transparency. The advantage is that any such test is robust to nonlinearity in the sensor-photoreceptor response to intensity.

High-Level Qualitative Vision

Four presentations were given in this session. The paper by W. Lim made a case for using qualitative vision to build a system that recognizes and automatically builds models of objects in the rocks world. Because objects in the rocks world are hard to model quantitatively, qualitative three-dimensional models are used instead. Such a model of a rock is built from successive views of the object captured using a mobile camera. With a few fixed positions to start, new camera positions are generated as more information is acquired about the rock from previous views.

An approach for shape reconstruction based on qualitative features was presented by S. Dickinson, A. Pentland, and A. Rosenfeld. A set of 10 volumetric primitives is used for modeling objects. Extracted two-dimensional image features (for example, contours, line groupings) are matched with possible faces bounded by contours. These faces are then matched with the possible face structures for the given set of volumetric primitives. Only qualitative descriptions of shapes of surfaces and contours are used in this approach (for example, planar or straight, convex, concave).

S. Yantis presented some empirical data studying the visual tracking of spatial configurations in noise. Humans can track several randomly moving points as long as these points can be grouped into a nonrigid, convex, virtual polygon. One way this grouping can be done is to avoid tracking the rapidly changing position of the individual points, instead tracking the slower changing properties, such as the location, size, and approximate shape of the virtual polygon.

A qualitative approach for studying, classifying, and interpreting temporal sequences of images was discussed by J-Y. Herve and J. Aloimonos. Rather than try to reconstruct the structure of the scene, the relation between changes (or catastrophes) in the nature of vector fields from one class to another and the occurrence of events in the scene are studied. This analysis is done by detecting the appearance and disappearance of cusps of projections of surfaces in velocity space to the x - y space. Because this approach uses a more global property (for example, stable singularities of the vector field), it can be more robust than approaches that rely only on more local vector values.

Open Discussion

The open discussion focused on the question of what is qualitative vision. This session followed a different format than the other sessions and was organized into two parts. The first part was composed of brief (5-minute) position statements by five invited speakers: J. Aloimonos, A. Blake, R. Nelson, T. Poggio, and J. Todd. This part was followed by an extended moderated discussion.

J. Aloimonos presented the view that qualitative vision techniques extract those features that support the use of vision to perform useful functions (that is, it is purposive). This approach need not be reconstructive; a much smaller subset can support useful vision-guided behaviors. How crucial such qualitative features are in achieving significant vision-guided behaviors was also discussed.

Rather than present a single view, A. Blake described a spectrum of views on qualitative vision. At one extreme, the photogrammetrists seek a precise and quantitative determination of physical properties. At the other end of the spectrum, AI attempts to ascertain symbolic information in the face of incomplete knowledge. Work on robust extraction techniques is closer to that of the photogrammetrists, but they allow more variability, error, and so on. The topological school is somewhere between, looking for salient relationships among less specifically defined entities.

A distinction between information and knowledge was noted in the view expressed by R. Nelson. In qualitative vision, only visual information that supports visual operations is relevant. Nelson provided the following definition: “Qualitative vision is the computation of iconic image properties having a stable relationship to functional primitives.” A visual task must be specified to determine what constitutes a functional primitive. Quantitative computations (for example, determining depth maps, fitting polygonal patches) do not have a strong functional role in the visual task domain. These functional visual icons are not sloppy. It is clear what they are, and they have a tightly bound relationship to the functional purpose; all other visual information can be ignored.

The problems with the use of quantitative methods for visual function were underscored by T. Poggio. In his view, qualitative vision is a method for determining associations between the visual input and the desired output without regard for the structure of the intervening process. J. Todd noted that qualitative vision provides abstract representations that are robust under varying conditions. Qualitative vision methods provide a reduction in data and often an increase in stability. Conversely, because qualitative vision methods discard some information, they lose some discriminability. The key to a working visual system is balancing the information requirements against the information loss.

In the extended discussion, there was general agreement that far less information is required to perform realistic vision than is required by quantitative and reconstructive approaches. Aspects of qualitative vision that were discussed include nonmetric, nonreconstructive and noninvertible, topology and semantics, functionally descriptive features, and reasonable assumptions of image and environmental structure.

The open discussion focused on the question of what is qualitative vision.

Concluding Remarks

For the greater part of the workshop, the term qualitative was used to mean nonmetric. In some cases, the emphasis was on removing some of the reliance on metric precision that is characteristic of photogrammetry. Such qualitative algorithms can be more robust because they do not rely on fine metric discriminations that can be undermined by issues of calibration, restrictive assumptions on geometry or image structure, and so on. Qualitative features also emphasize useful features that provide compact descriptions of objects by omitting fine detail. There was general agreement that qualitative approaches can be used for building vision systems that serve realistic purposes. Such approaches focus on the achievement of the better-defined goals of functional systems rather than on intractable problems introduced by traditional approaches that address general vision problems.

Acknowledgments

We would like to thank the reviewers for their time and guidance. We would also like to thank Advanced Decision Systems (ADS), Grumman Corporation, Massachusetts Institute of Technology, and Oxford University for supporting our efforts in organizing the workshop. In particular, we would like to thank ADS for printing the workshop proceedings and making them available to the research community.

Note

1. Copies of the proceedings can be obtained by writing AAAI-90 Workshop on Qualitative Vision, c/o Philip Kahn, Advanced Decision Systems, 1500 Plymouth Street, Mountain View, CA 94043-1230, pkahn@ads.com..

William Lim obtained his Ph.D. from the Massachusetts Institute of Technology. He is currently a senior research scientist at the CS/AI Laboratory in the Corporate Research Center of Grumman Corporation. His research interests are AI architecture for high-level control of mobile robots, self-awareness of intelligent agents, and qualitative vision.

Philip Kahn is a senior computer scientist at Advanced Decision Systems, Mountain View, California. He received his B.A. in computer science and economics and his M.S. in computer science from the University of California at Los Angeles, and he was a researcher in the Computer Vision Research Laboratory at the University of Massachusetts at Amherst. His interests include computer vision, behavioral robotics, active sensor control and processing, biological vision systems, and environmental representation and recognition.

Daphna Weinshall was a research associate at the Center for Biological Information Processing at the Massachusetts Institute of Technology and is now working at the IBM Thomas J. Watson Research Center. She obtained her Ph.D. in statistics from Tel Aviv University. Her research interests include vision and psychophysics.

Andrew Blake is a faculty member in the Robotics Research Group at the University of Oxford. His principal research interests are in computer vision and computational psychophysics. He has coauthored two books: *Visual Reconstruction* (MIT Press) and *AI and the Eye* (Wiley).