Proceedings of the 2005 IEEE International Conference on Robotics and Automation (ICRA 2005)
Barcelona, Spain, Apr. 18-22, 2005

Developing a Modular Active Spherical Vision System

Nicholas D. Jankovic and Michael D. Naish
Dept. of Mechanical and Materials Engineering
The University of Western Ontario
London, Ontario, Canada
ndjankov@uwo.ca, naish@eng.uwo.ca

***Abstract* - This paper introduces a modular, real-time, omnidirectional, active vision system, as well as a constructed prototype. By combining omnidirectional and active pan-tilt cameras, a robust vision system is created that builds on the strengths of each camera type. The system can be easily configured to provide nearly an entire spherical field of view and independently track several targets of interest within the environment. The novel design allows the camera modules to be stacked, creating a vertical sensor structure. This vertical arrangement also provides a simple solution to the epipolar geometry and triangulation for target localization. Applications for this modular system can range from simple mobile robot navigation to complex multi-target tracking and surveillance.**

Index Terms* - *Omnidirectional, modular, surveillance, active vision and multi-camera system.

I. Introduction

Throughout the evolution of surveillance and mobile robot systems, vision has played a central role in perceiving the environment. This is most often accomplished using standard perspective cameras, which provide a high level of detail at the cost of a relatively narrow field of view (FOV). To overcome the FOV limitation, cameras can be fitted to pan-tilt-zoom (PTZ) platforms to dynamically increase the FOV. Unfortunately, the instantaneous FOV is still very small (i.e., there is a very large "blind-spot"), rendering these systems less suitable for *detecting* moving targets.

Alternatives to the standard perspective camera are omnidirectional cameras, which have become increasingly popular in recent years. Depending on the design, these cameras typically provide a $360^o \times 180^o$ FOV or more. They have found extensive application in mobile robotics, as well as in visual surveillance tasks. Configurations for an omnidirectional camera may consist of a camera equipped with a fisheye lens [1] or a camera that views an image from a curved mirrored surface. The latter type is commonly referred to as a catadioptric camera [2].

By viewing a hemisphere of the environment these two camera types can help solve the "blind-spot" problem. Unfortunately, as a direct consequence of increasing the FOV, the level of image detail is greatly sacrificed. A trivial solution is to simply increase the camera resolution to compensate for the loss of detail. Another solution uses multiple wide-angle cameras, whose images are later stitched together to form an omnidirectional view. However, both approaches lead to increased computational costs. Much of the high level of detail is also wasted because large portions of the image may not contain any relevant information. For applications that require both a large FOV and the ability to resolve fine detail, an alternative approach that combines the benefits of both omnidirectional and perspective cameras may be advantageous. Multiple omnidirectional cameras can be combined to significantly reduce the total "blind spot" area, while one or more perspective cameras can extract fine detail whenever it is necessary.

There have been a number of omnidirectional vision systems developed in recent years that use perspective cameras. Many have been applied to mobile robots to deal with tasks such as map generation, navigation, obstacle detection and human interaction [3, 4, 5]. Similar systems have also been employed for surveillance tasks such as intruder detection and tracking. These consist of at least one ceiling or platform-mounted catadioptric camera and may use one or more active cameras [6, 7, 8, 9, 10].

The objective of this research is to create a system that uses both omnidirectional and perspective cameras in a modular and centralized fashion. A modular design simplifies system reconfiguration for specific applications. The camera modules are stacked vertically, providing several advantages over other systems: First, the configuration allows 360^o omnidirectional viewing in the horizontal plane. Second, the combined system can provide a spherical FOV. This is beneficial for surveillance applications, such as in large multi-level shopping malls or for robot navigation, where both ground and ceiling references can be used. Third, the vertical configuration minimizes unnecessary occlusions, such as seeing the active camera in the omnidirectional view. Fourth, it allows simple peripherally-guided active vision, allowing the system to resolve fine detail without resorting to excessively high-resolution omnidirectional cameras. Finally, when two cameras have overlapping fields of view, then target positions can be estimated using triangulation, for which there are three possible combinations.

1) Two omnidirectional cameras can roughly determine the location of a target. Triangulation is very quick, but accuracy is limited because the cameras only see the target with a relatively low level of detail.

2) Two active perspective cameras with highly detailed views can precisely extract a target's location using conventional stereo, but this requires two active cameras and takes more time for both to acquire the target.

3) One omnidirectional and one active perspective camera provide a good balance between triangulation accuracy, speed and utilization of camera resources.

II. System Requirements

The research objectives discussed in the previous section serve to define the following set of system requirements: The peripheral FOV needs to be maximized and occlusions minimized so that potential targets do not go undetected. Omnidirectional views should overlap to reduce "blind-spot" areas. The perspective camera must be able to pan, tilt and preferably zoom to resolve a high level of detail. It must rotate continuously and have unobstructed 360° access to the environment so that it may see *any* target viewed by the omnidirectional cameras.

Additional requirements include: low cost, low power consumption, compact size and lightweight modules. An all-digital system would help to reduce losses from analog to digital conversion and analog transmission. The modules must be capable of being connected in almost any combination. The system must operate in a real-time and synchronized manner so that the behavior is precise and predictable. It also needs to express some autonomy, such as being able track and prioritize detected targets, assigning the active camera(s) to the most significant ones so that more detailed information can be extracted.

III. Performance Specifications

Three different module types are available for system construction: a fisheye module, an active perspective module and a catadioptric module. By stacking these in different combinations, a number of unique systems can be created; some example configurations are shown in Fig. 1.

In order to create a spherical FOV with sufficient overlap, the individual FOV of the fisheye and catadioptric modules must be greater than 180°; thus, a FOV of about 200° is desirable. In order to capture adequate detail using an active camera module, its FOV should be 10 to 30 times less than that of the omnidirectional modules; however, the actual value depends largely on the intended application. Fig. 2 illustrates how the various fields of view would overlap for the configuration shown in Fig. 1b.

There are two "blind-spots" that can be seen. One occurs between omnidirectional cameras. This small area does not pose a significant problem because the views overlap at a known distance. The major "blind-spot" occurs behind the camera of the catadioptric module. This area cannot usually be avoided because the system must be rigidly secured at some point. Fortunately, it does not cause major problems for most applications because this area usually does not contain any targets of interest.

Some compromises have to be made when configuring the system. Increasing the range at which targets can be detected or detecting smaller targets requires higher camera resolution. High-resolution images can be used, but at the cost of increased computational power and bandwidth requirements. To lower the computational power and bandwidth demands, image resolution and colour depth must be reduced, sacrificing image detail.

The following specifications are selected in order to create a functional and cost-effective system using widely available components.

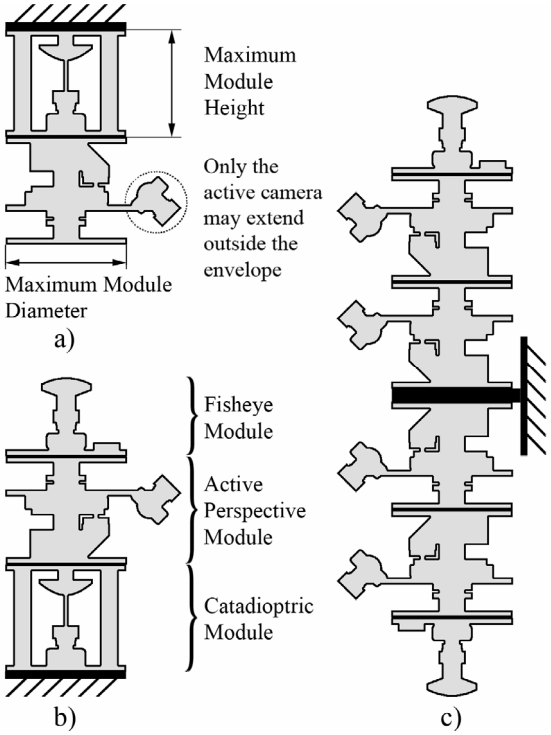

Fig. 1 Sample system configurations. a) a ceiling mounted system, b) a ground or robot mounted system and c) a wall mounted system; this system uses two stacks to effectively double the number of modules.

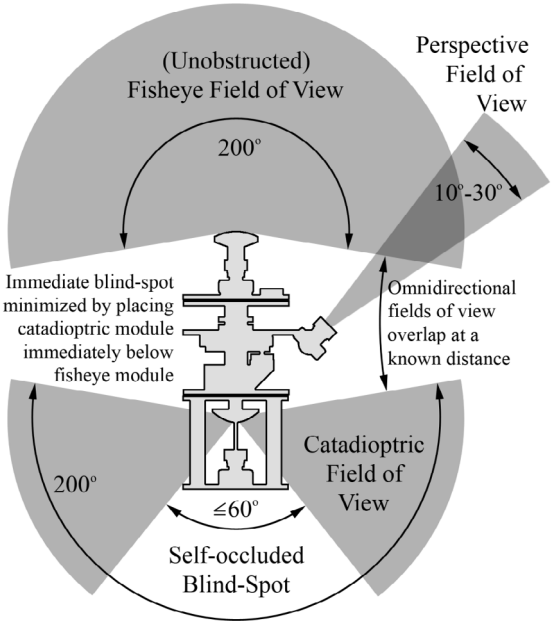

Fig. 2 Overlapping camera fields of view.

1) *Camera Resolution:* A minimum 640×480 pixel, colour CCD camera. This resolution is standard and it is commonly used for digital video. Image processing can be done very quickly. Higher resolutions are also an option.

2) *Frame Rate:* 15Hz for red-blue-green (RGB) or 30Hz for luminance and chrominance (YUV 411). These two image formats require the same bandwidth. These frame rates are also suitable for real-time applications and standard computers can handle the required bandwidth.

3) *Bus Type:* IEEE 1394a (Firewire). This serial bus architecture is established and well supported. It provides high bandwidth, 400Mbit/s, which is capable of supporting two cameras at the minimum resolution and frame rate. Firewire also allows cameras to be chained in series, which is necessary for a modular architecture.

4) *Module Diameter:* 0.16m or less (excluding the protruding active camera head). The diameter should be as narrow as possible, but must be wide enough to retain structural rigidity. The active camera head can extend outside the module providing unobstructed vertical access to the environment.

5) *System Height:* 1.0m or less. Since the modules are stackable, then it will grow in the vertical direction. One metre provides a reasonable upper limit, ensuring that the system remains compact and structurally rigid.

6) *Total Modules:* 6. Given practical limitations on size and bandwidth, a total of six modules can be stacked into one system. This allows sufficient module height, while allowing many different system configurations.

7) *Module Height:* Less than 0.16m. This number is based on the 1.0m system height limit, given six modules.

8) *Number of Buses:* Three are required to satisfy the minimum bandwidth for six cameras.

9) *Tracking Speed:* 1.5 rad/s (57 degrees/s) for each pan and tilt axis. This is equivalent to a typical person's walking speed of 1.5m/s (5.4km/h) at a range of 1m from the system. (Note that, most targets will be farther away.)

10) *Acquisition Time:* 1.0 second. This is the maximum time allowed for the active camera to saccade from one target to another. In the worst case, the new target is on the opposite side of the device, so half a revolution must be made. Assuming a triangular velocity profile, the maximum acceleration needs to be 12.6 rad/s^2, resulting in a maximum speed of 6.3 rad/s. One second provides a good balance between responsiveness and performance costs; moving any quicker would significantly increase the dynamic loads on the structure.

Theoretically, the system should be able to track any number of targets visible in the omnidirectional views; however, there are practical limitations. Numerous targets will cause an increased computational load, as well as increase the chance of occluding each other. Both effects cause system performance to degrade proportionately.

To satisfy the minimum requirement of having an active spherical vision system, three modules are needed. Using one catadioptric module (placed at the bottom) and one fisheye module (placed at the top) creates a good spherical FOV with minimum occlusions and sufficient overlap. The active camera is placed in between the other two, as shown in Fig. 1b.

IV. Basic Operation

In order for the surveillance system to function effectively, points in the omnidirectional images must relate to points in the perspective images. This requires both precise calibration and knowledge of the epipolar geometry between the two cameras. Fortunately, the vertically aligned structure helps simplify this task. When the principle axes of the omnidirectional cameras and the pan axes of the perspective cameras are all perfectly aligned, it allows the epipolar geometry to be solved trivially because epipolar lines appear as straight lines in the perspective views and as radial lines in the omnidirectional views.

A target point is seen on the image plane of an omnidirectional camera. This image point and the principal axis define a vertical plane that intersects the target in space. By means of prior calibration and reading the pan and tilt encoders, the pose of the active camera can be determined. Its image plane can then also be defined in space. The intersection of the vertical plane and the active image plane define a line in space. This forms the epipolar line when mapped back on to the active camera image. Sample images from all three cameras, with superimposed epipolar lines, are shown in Fig. 3.

The combination of an omnidirectional camera and an active camera creates a synergy such that a detected target in an omnidirectional image is able to guide the active camera to the appropriate viewing position. Since the cameras are all aligned vertically, the active camera position must simply line up with the correct radial epipolar line (i.e., $\phi_{pan} = \phi_{fish}$). If the active camera is located near the omnidirectional camera, then the inclination angle of the active camera is related to the radial distance of the target as seen in the omnidirectional view (i.e., $\gamma_{tilt} = f(r_{fish})$), thereby reducing the active camera search space. When both cameras are finally observing the same point, its location in space can be easily determined with simple geometry, as shown in Fig. 4.

V. System Design

A. Fisheye Module

The fisheye module is the simplest of the three modules. Its most significant feature is that it can provide a completely unobstructed omnidirectional field of view. In order to preserve this feature, the module can only be placed as the last element in the stack. One disadvantage is that a fisheye lens does not provide a true single effective viewpoint, but rather a locus of viewpoints along the principle axis. The locus can be determined by calibration, but suitable results are still obtained when a single viewpoint is assumed because the locus is small.

There are some design considerations necessary for this module, the most important being the field of view. The specification calls for a compact, 200^o FOV lens, but these are not common. Most fisheye lenses only view up to 180^o, although some fisheye lens converters can provide views up to 183^o. Alternatively, there is an SLR camera lens that can view 220^o, but it is 0.20m in diameter and weighs 5 kg. The other alternative is to design and fabricate a custom lens, but this may not be cost effective.

B. Catadioptric Module

The catadioptric module has a key advantage over the fisheye module: catadioptric cameras are easier to customize. Readily available mirrors usually provide views greater than 180^o and also provide higher resolution

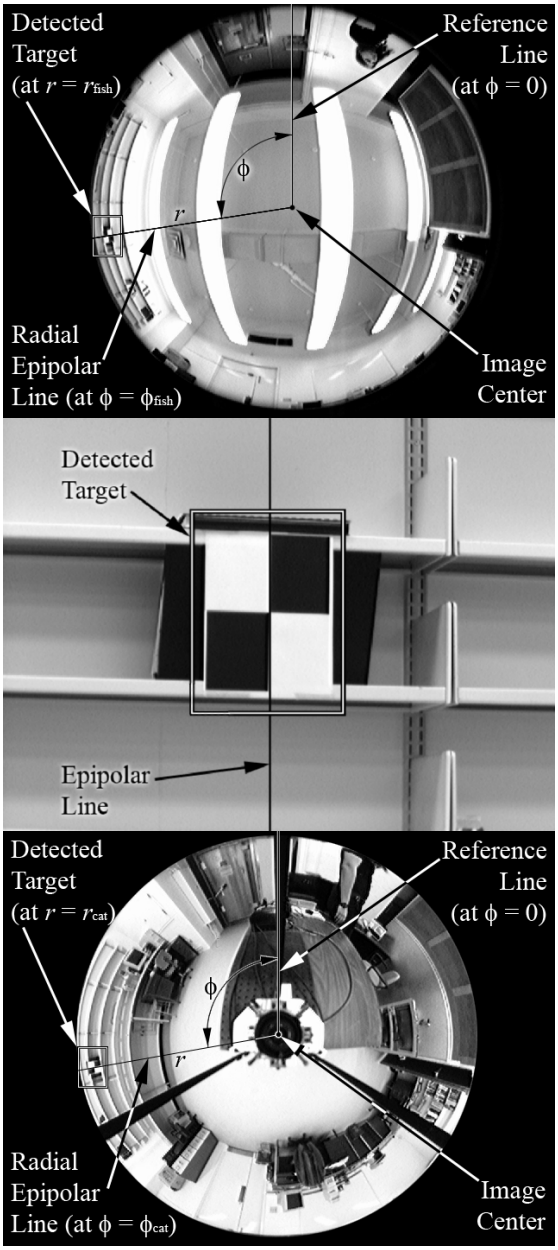

Fig. 3 Sample system images with epipolar lines: fisheye camera (top), active camera (center) and catadioptric camera (bottom). A checkerboard target is seen in both omnidirectional views, indicating sufficient overlap between their FOV's. The target position in the image can be defined using polar coordinates.

at the periphery. If standard mirrors are not suitable for the application, a custom one can be manufactured more easily and cost effectively than a fisheye lens. Unfortunately, the catadioptric configuration has one major disadvantage: it is inherently self-occluding at the image center (i.e., the camera sees its own reflection). Nevertheless, this permits other modules to be mounted on either side without significantly obstructing the FOV.

Some design considerations include having sufficient torsional stiffness, minimum occlusion and good image clarity. There are two common support configurations used: a vertical support beam or a clear acrylic tube. The vertical beam structure is simple to construct and provides

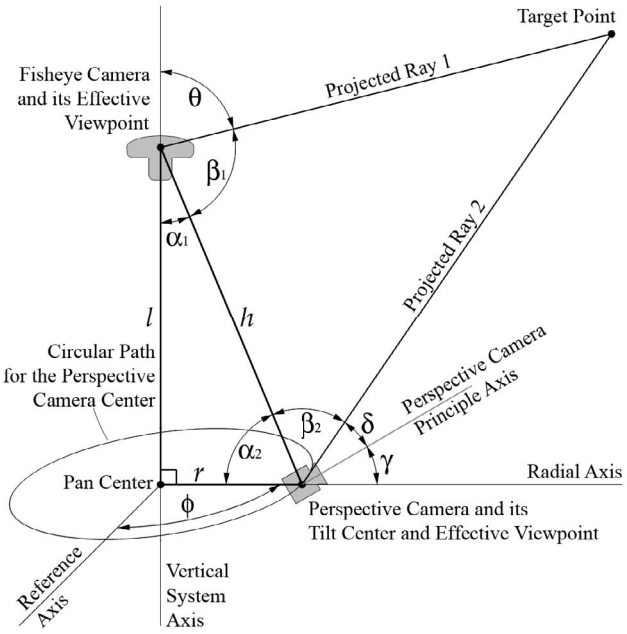

Fig. 4 Triangulation geometry for fisheye and perspective cameras (similar geometry exists for the catadioptric camera). Here, l and r are known from calibration and h, α_1 and α_2 can be easily calculated. θ is determined from the fisheye image and δ is from the perspective image. ϕ and γ are read from the pan and tilt encoders, respectively. Upon calculating β_1 and β_2, the target position can be determined using the cosine law.

better image clarity. However, it also creates additional occlusion and does not provide the best torsional stiffness. Rigidity is very important because it has to support the weight of up to five other modules, and potentially the dynamic torsional loads of up to five active modules as well. Conversely, a cylinder is very stiff, but introduces internal reflections. Even though the latter design allows for a potentially unobstructed omnidirectional view, it is not possible here because cables must connect to the modules both above and below. To limit the occlusion caused by a large cable bundle, it is better to distribute the cables so that there are only three narrow occlusions, thereby reducing the chances of missing a target.

Another design consideration pertains to the type of camera-mirror combination to be used. A single effective viewpoint provides beneficial properties and there are two common combinations that can provide this: a standard perspective camera viewing a hyperbolic mirror and an orthographic camera viewing a parabolic mirror. The latter camera requires the use of a bulky and expensive telecentric lens; therefore, the former configuration provides the most compact and cost-effective solution.

C. Active Perspective Module

The active perspective module (Fig. 5) is the most complex one in the system and many factors influence its design. The support structure introduces a number of design constraints. It must allow the camera to rotate continuously (via a slip ring) and also allow cables to connect to the other two modules. This requires mounting the slip ring to a stationary tube with a hollow core. The structure must support the static and dynamic loads of

itself and up to five other active modules, so the core must also be very strong. The type of slip ring and the pan motor system both affect the module height. Ideally, the components can be arranged in a flat and layered manner.

The drive system also has some unique features. Since the pan axis has a stationary hollow core, then it cannot be driven from an inline motor assembly. An offset spur gear stage is used to provide power, maintain position accuracy and reduce backlash.

It is often advantageous to have the camera rotate about its perspective center in an active camera system; this greatly simplifies later triangulation calculations. The vertical system design inherently does not allow the camera to pan through its center; however, the tilt axis does not have this restriction. By rotating the camera center through the tilt axis, the center is restricted to a circular path, as illustrated in Fig. 4. Reading the pan encoder position provides a means of locating the perspective center in space. This triangulation approach works best with monofocal lenses because the camera center of a zoom lens varies with the focal length. One solution is to implement a zoom lens with a stationary camera center. The other is to calibrate and then calculate the location of the center with every new camera position.

D. Module Connectivity

Having a system with modular cameras necessitates a modular hardware architecture to connect them. Bandwidth dictates that one Firewire bus can support a total of two cameras and two motion controller devices. Each module has three sets of communication and power connectors; however, only one connector set is used, the other two simply pass directly to the other side (in the case of the active and catadioptric modules). Each module can be connected in any one of three orientations (in 120° increments) to the previous module. This provides enough flexibility during installation to ensure that only two modules are connected to any one bus at a time. Fig. 6 illustrates how these three Firewire buses are utilized.

The fisheye and catadioptric modules are simple and straight forward, whereas the active module must also accommodate motion control. In the active module, there are two devices: a Firewire converter (for motion control) and a Firewire frame grabber (to acquire images). The converter simply sends motion commands to the controllers and receives incremental encoder readings. Even though it is highly desirable to employ an all-digital system, a frame grabber and an analog camera are selected out of necessity. The bandwidth required for a 400Mbit/s Firewire signal is quite high, making communication across a standard brush-conductor slip ring very troublesome. Alternatively, optical slip rings can easily sustain this high data transfer rate, but they are also prohibitively expensive.

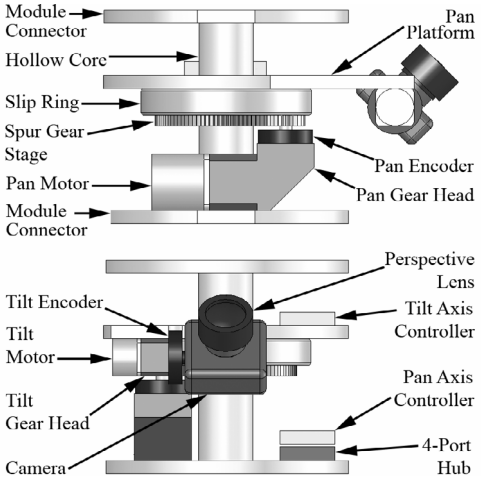

Fig. 5 Major active module components.

VI. Initial Prototype

A first generation prototype has been constructed in order to demonstrate proof of concept and provide a starting point for calibration, experimentation and software implementation. The configuration is identical to the one shown in Fig. 1b. This design was selected because it satisfies the minimum requirements for an active spherical vision system. The constructed system is shown in Fig. 7.

The catadioptric module uses a varifocal lens (set to a focal length of 3.6mm) paired with a Neovision H3G Hyperbolic mirror to give a 212° FOV. The fisheye module provides a 183° FOV using a 4mm monofocal lens with a Nikon FC-E8 fisheye lens converter, which provides the best balance between FOV, size and cost. The active module uses a 12mm monofocal lens and stepping motors to drive the pan and tilt axes. Stepping motors are chosen because they are cost-effective and provide precise incremental rotation. Planetary gear heads are used to increase the effective step resolution and increase torque at the output. The module uses a high torque stepper motor and a 10:1 planetary gear head with a custom-designed 4:1 anti-backlash spur gear stage to drive the pan platform, giving a total 40:1 reduction. The anti-backlash gear is used to eliminate any additional backlash. Similarly, the tilt stage uses a standard stepper motor and a 28:1 planetary gear head. 8000 pulses per revolution incremental optical encoders are used to provide closed loop feedback. The motors are controlled with an industrial PCI four-axis motion controller.

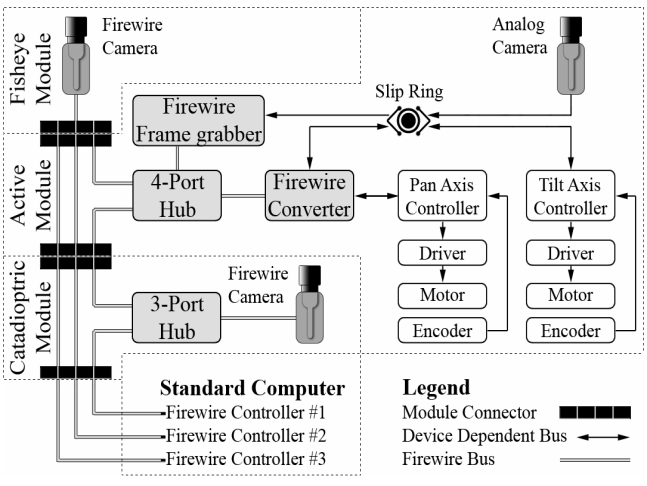

Fig. 6 Hardware architecture.

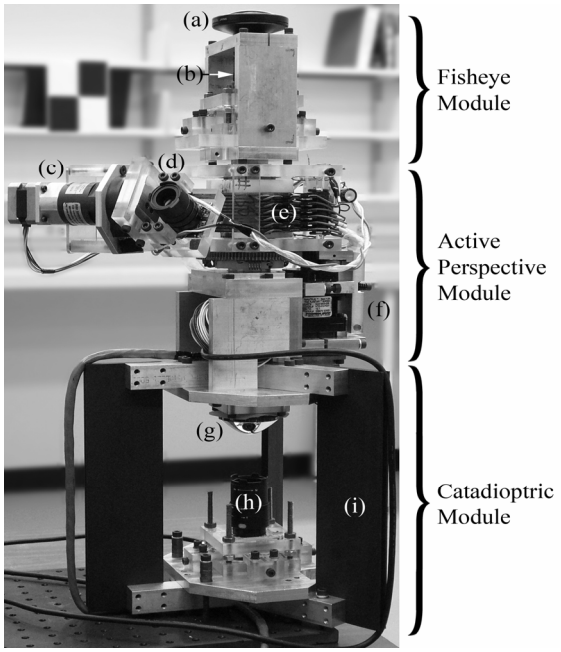

Fig. 7 Initial prototype. a) fisheye lens, b) 4mm camera, c) tilt motor, d) 12mm camera, e) slip ring, f) pan motor, g) hyperbolic mirror, h) 3.6mm camera, i) vertical support beam.

There are a number of differences between the constructed prototype and the ideal system. First, the active camera does not have any zoom capabilities; however, a high level of detail is still achievable with a 640×480-pixel camera. A target that appears as a 25-pixel cluster in the omnidirectional view can be seen as 16000 pixels in the active camera view, indicating a significantly higher level of detail. Second, the prototype is physically much larger than specified (about twice the size) because it was built using readily available and low-cost components, which produced a rather massive active camera platform. To limit the stress on the offset spur gear stage, the highest achievable acceleration is reduced to 1.0 rad/s^2, which increases the maximum target acquisition time to about 3.5 seconds. Third, the vertical beam design was chosen for the catadioptric support structure because it was simpler to manufacture and did not create any problems with internal reflections. Fourth, system modularity is rather limited. Firewire is used, but each camera connects directly to the computer. Each motor, encoder and limit switch also communicate directly with the motion controller, which increases the number of necessary cables.

Despite these limitations, the system meets the expectations of a first generation prototype. It serves as a platform to investigate the core technologies and gain insight that may not be apparent at the design stage.

VII. Conclusions

This paper describes the motivation behind the creation of a centralized and modular multi-camera spherical vision system. The spherical design minimizes the total "blind-spot" area so that virtually all targets are detected. The modular design allows maximum flexibility so that the system can be tailored to the desired application. A system can be constructed from three types of camera modules, stacked vertically. The fisheye module can view an unobstructed hemisphere, the catadioptric module provides a view greater than a hemisphere and the active perspective module resolves fine detail. The system geometry allows for simple peripherally-guided active vision and epipolar geometry. When two cameras view the same point, triangulation can be employed to estimate its location.

Of course, no system is without limitations. Even though the central design goal is to minimize occlusion, it cannot be eliminated completely due to functional limitations. For reasons of structural rigidity and bandwidth, the total number of modules per stack is limited to six. Allowing the stack to be any higher would make it unstable in outdoor environments or on a moving robot. The system is most applicable to surveillance. Larger robots may be able to use the system, but they would probably be limited to fewer than six modules to reduce weight and power consumption.

VII. Future Work

Future efforts will focus on system calibration and development of the software architecture for effective autonomous multi-target prioritization and tracking. Additional omnidirectional sensor modules such as omnidirectional laser range finders or sonar may be introduced to increase system functionality.

References

[1] B.Micusik and T.Pajdla, "Estimation of omnidirectional camera model from epipolar geometry," *Proc. IEEE Conf. on Comp. Vis. and Patt. Recog.,* vol. 1, Madison, WI, pp. 485-490, 2003.

[2] Y. Yagi, "Real-time omnidirectional image sensor for mobile robot navigation," *Proc. IEEE Conf. on Intelligent Control*, Vancouver, BC, pp. 702-708, 2002.

[3] Y. Yagi, K. Egami and M. Yachida, "Map generation for multiple image sensing sensor MISS under unknown robot egomotion," *Proc. IEEE/RSJ Int. Conf. on Intelligent Robot and Systems,* vol. 2, pp. 1024-1029, 1997.

[4] G. Adorni, L. Bolognini, S. Cagnoni and M. Mordonini, "A non-traditional omnidirectional vision system with stereo capabilities for autonomous robots," *Proc. of AI*IA 2001: Advances in Artificial Intelligence*, Springer-Verlag: Berlin, pp. 344-355, 2001.

[5] T. Wilhelm, H.J. Bohme and H.M. Gross, "A multi-modal system for tracking and analyzing faces on a mobile robot," *Robotics and Autonomous Systems*, vol. 48, pp. 31-40, 2004.

[6] M. Greiffenhagen, V. Ramesh, D. Comaniciu and H. Niemann, "Statistical modeling and performance characterization of a real-time dual camera surveillance system," *Proc. IEEE Conf. on Comp. Vis. and Patt. Recog.*, vol. 2, pp. 335-342, 2000.

[7] S. Guler, J.M. Griffith and I.A. Pushee, "Tracking and handoff between multiple perspective camera views," *Proc. of the 32nd App. Img. Patt. Recog. Workshop (AIPR'03)*, pp. 275-282, 2003.

[8] M. Doi and Y. Aoki, "Real-time video surveillance system using omni-directional image sensor and controllable camera," *Proc. of the SPIE - Real-Time Imaging VII*, vol. 5012, pp. 1-9, 2003.

[9] J.P. Barreto and H. Araujo, "A general framework for the selection of world coordinate systems in perspective and catadioptric image applications," *Int. J. of Comp. Vis.*, vol. 57, no. 1, pp. 23-47, 2004.

[10] G. Scotti, L. Marcenaro, C. Coelho, F. Selvaggi and C.S. Regazzoni, "A novel dual camera intelligent sensor for high definition 360 degrees surveillance," *Proc. of Intelligent Distributed Surveillance Systems (IDSS-04)*, pp. 26-30, 2004.