It should also be noted that there is a rich literature of least squares algorithms designed to introduce numerical stability in awkward cases. The work described here could routinely be extended to incorporate methods such as UDU factorization.

REFERENCES

An Agile Stereo Camera System for Flexible Image Acquisition
ERIC KROTOK, FILIP FUMA, AND JOHN SUMMERS

Abstract—An agile stereo camera system designed for flexible image acquisition under a wide variety of viewing conditions and scenes is presented. A host processor sends commands to three microprocessors controlling, through 11 servomotors, the position and optical parameters of the cameras, and the scene illumination.

Manuscript received December 8, 1986; revised June 30, 1987. This work was supported in part by NSF CER, DEC Corporation, IBM Corporation, and LOR Star Corporation. This communication was presented in part at the International Conference on Pattern Recognition, Paris, France, October 1986.

E. Krotos is with the Robotics and Artificial Intelligence Group, L.A.A.S., du C.N.R.S., 7, Avenue du Colonel Roche, 31077 Toulouse Cedex, France. F. Fuma and J. Summers are with the GRASP Laboratory, Department of Computer and Information Science, University of Pennsylvania, Philadelphia, PA 19104-6389.

0882-4967/88/0200–0108$01.00 © 1988 IEEE
I. INTRODUCTION

Vision systems are essential for automating tasks such as object recognition, part manipulation, inspection, and measurement. All vision systems are limited by the quality of images they acquire, so choices of cameras and illumination sources generally are made with great care. To make these decisions easier, we present an agile stereo camera system developed as a flexible image acquisition device to automatically procure a steady stream of high-quality images of a wide variety of scenes.

Typical camera systems used in computer vision and robotics applications have fixed optical and position parameters which must be laboriously tuned for different viewing conditions and/or different scenes. Here we will briefly review some of these systems.

The "hand–eye" system developed at Stanford University [1] consists of a pan/tilt head, a lens turret for controlling focal length, color and neutral-density filters mounted on a "filter wheel," and a vidicon camera whose sensitivity is programmable. This system is limited in processing and I/O bandwidth (it used a PDP-6), and in resolution (four bits), and has only one camera. Moravec's [8] "slider stereo" system allows a single camera to translate along an axis on a mobile vehicle.

POPEYE is a grey-level vision system developed at Carnegie-Mellon University [1]. It is a loosely coupled multiprocessor system with a 68000 microprocessor, a frame grabber and buffer, an array processor, dedicated image preprocessing units, and a programmable transform processor. Image positioning is achieved with a pan/tilt head and motorized zoom/focus lens. Altogether, this is a very powerful and flexible system although it has only one camera and is not capable of translational movements.

The WABOT robot system [9], developed at Waseda University in Japan, has two cameras with computer-controlled focal length and scan line selection. The cameras rotate with the robot's trunk (pan) and within the trunk (vergence). It is difficult to determine from the paper what functions actually have been implemented and tested.

Kuno et al. [6] report an innovative stereo camera system whose interocular distance, yaw, and tilt are computer controlled. The cameras are mounted on a specially designed linkage; the distance between them may vary from 10 to 40 cm. With variable interocular distance a certain flexibility in processing is achieved: the larger the distance, the more precisely disparity information from stereo can be converted to absolute distance information; the smaller the distance, the easier the solution to the correspondence problem. It is not apparent from the paper how this flexibility is to be exploited, nor how the three degrees of freedom will be controlled.

Montgat and Pelsue [7] document a camera system with a variable focal length lens. Mirrors mounted on an X–Y deflector, two galvanometer scanners in an orthogonal configuration allow the camera to pan and tilt up to ±25°.

While these are important contributions, none provides the full flexibility of many primitive biological systems, and none approaches the abilities of the human oculomotor plant. In particular, there appear to have been no attempts to design and build camera systems with the kinematic and functional capabilities of the human head and neck.

In this communication we present a far more flexible camera system than those reported in the literature. First we describe the design and construction of the camera system hardware. Then we discuss the design, implementation, and performance of the hardware and software controlling each of the devices. Finally, we review some examples of how the camera system currently is employed and conclude with a critical discussion of the overall system.

II. CAMERA SYSTEM HARDWARE

The photograph in Fig. 1 illustrates the hardware configuration of the camera system, which is actuated by eleven servomotors: five adjust the position and orientation of the two cameras, and six adjust the optical parameters of the two lenses. In addition, special-purpose hardware illuminates ten independent lamps under computer control.

![Fig. 1. Photograph of camera system.](image)

![Fig. 2. Positioning mechanisms. Figure schematically illustrates two cameras mounted inside stationary gantry. They are able to translate horizontally and vertically and rotate by panning and tilting.](image)

In this section we discuss in detail the positioning mechanisms, the stereo cameras, the motorized lenses, and the lighting system.

A. Positioning Mechanisms

To acquire useful images of a large variety of scenes, the cameras must be agile enough to see different objects, which may be partially or completely occluded. To achieve this agility, we mount the cameras on a neck-like mechanism which has five degrees of freedom. Four of these are provided by the camera platform and one by the vergence platform.

Fig. 2 schematically illustrates the camera platform mechanisms.
A heavy steel gantry supports the platform, affording four degrees of positioning freedom: two translations and two rotations. Each of the four axes is actuated by an ac servomotor driving a gear mechanism. The horizontal and vertical axes' gears drive a split-band pulley actuator to translate the platform. To compensate gravity, the vertical axis has a constant force spring, which in practice suffers from problems of 'sticktion' and spring nonlinearities. The pan and tilt motors drive gears rigidly attached to the platform to rotate it. The following joint actuation is possible:

- translation $x(0 \leq x \leq 36.8 \text{ cm})$;
- translation $y(0 \leq y \leq 25.4 \text{ cm})$;
- pan (yaw) rotation by $\alpha(0 \leq \alpha \leq 55^\circ)$;
- tilt (pitch) rotation by $\beta(0 \leq \beta \leq 49^\circ)$;
- no roll ($\rho = 0$).

Four wire-wound potentiometers mounted on shafts driven by spur gears sense the position of the platform within the gantry.

Fig. 3 diagrams the mechanical design of the vergence platform, which allows the cameras to rotate toward or away from each other, delimiting the field of view common to the two cameras. Each camera is mounted on a separate aluminum plate, which rotates on aluminum pivots riding on Teflon blocks mounted at the front and back ends of the base plate. The distance between the pivots is fixed at 12.8 cm. To rotate the cameras, an appropriate motion of the motor is needed: clockwise movement of the screw causes convergence, and counterclockwise movement causes divergence. Movement of the screw drives a threaded block riding in a dovetail slot on the base plate. Bearings mounted on this block roll along guides attached to the camera plates, spreading the rear of the camera plates. A spring holds the rear ends of the camera plates together, resulting in a positive response to any motor movement and very little backlash.

The kinematics of the vergence platform allow rotation around each pivot by $\gamma(-1.0 \leq \gamma \leq 5.5^\circ)$. The rotations are coupled and their magnitudes are antisymmetric. This pair of rotations is called a vergence and may be either convergent or divergent. The cameras can converge approximately $5.5^\circ$ and diverge approximately $1.0^\circ$. The incremental camera rotation $\Delta \gamma$ ranges from $0.04^\circ$ to $0.08^\circ$ per motor turn, depending upon the position in the curved bearing guides. A conservative estimate of the total positioning error is on the order of $\pm 0.02^\circ$; in practice, there is less error, especially for small movements of the screw.

Kinematically the camera system is similar to the human head and neck. Referring to the coordinate system in Fig. 4, the neck muscles and skeleton allow rotations around $k$ (pan, yaw), around $j$ (tilt, pitch), and around $l$ (roll). The upper spine does not permit significant translations in any direction, which are generally accomplished by actuation of the shoulder and waist joints, and by locomotion. The camera platform does not allow roll, but does allow translations along the $j$ and $k$ axes. In a young person, the eye muscles can rotate the eyes from the primary position about $45^\circ$ to $50^\circ$ to left, right, and downwards, and about $40^\circ$ to $45^\circ$ upwards [2, p. 191]; the camera platform is slightly more restricted in these
TABLE I
LENS SPECIFICATIONS

<table>
<thead>
<tr>
<th>Item</th>
<th>Lens Attribute</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>mount</td>
<td>1 in (C type)</td>
</tr>
<tr>
<td>2</td>
<td>focal length f</td>
<td>17.5–105 mm</td>
</tr>
<tr>
<td>3</td>
<td>minimum focusing distance</td>
<td>1.3 m</td>
</tr>
<tr>
<td>4</td>
<td>aperture ratio (f/#)</td>
<td>1.8</td>
</tr>
<tr>
<td>5</td>
<td>minimum aperture</td>
<td>1 mm</td>
</tr>
<tr>
<td>6</td>
<td>maximum aperture</td>
<td>58 mm</td>
</tr>
<tr>
<td>7</td>
<td>C-mount to image plane</td>
<td>20.01 mm</td>
</tr>
<tr>
<td>8</td>
<td>mass (approx.)</td>
<td>0.5 kg</td>
</tr>
<tr>
<td>9</td>
<td>diagonal field angle</td>
<td>7.8–44.7°</td>
</tr>
<tr>
<td>10</td>
<td>horizontal field angle</td>
<td>4.8–28.2°</td>
</tr>
<tr>
<td>11</td>
<td>vertical field angle</td>
<td>6.2–36.1°</td>
</tr>
<tr>
<td>12</td>
<td>exit pupil position</td>
<td>≤ 420 mm</td>
</tr>
<tr>
<td>13</td>
<td>magnification</td>
<td>1–6×</td>
</tr>
</tbody>
</table>

Items 1–4, 7, and 9–13 and FUJINON specifications. Items 5 and 6 were computed from α = f/f.#.

In items 10–12 the field angle of the lens is not identical to the field angle of the photodetectors, which cover a smaller active optical area. The field angle α is computed as α = 2 × arctan (D/2f) for D, a linear dimension of the CCD chip.

rotations. In addition, the vergence platform allows a coupled rotation around k, converging and diverging, similar to forward-directed human vergence movements. In all, the camera system has two more translational and one less rotational degree of freedom than the human neck.

B. Stereo Cameras
To determine visually the distances of unknown objects, information from at least two different viewpoints (characterized by the sensor position, orientation, and/or optical parameters) is necessary. We choose to mount two cameras (Fairchild CCD222) on the platform. Video amplifiers and clock circuitry (Fairchild CCD3000) read the video signals from the CCD chips. A real-time digitizer and frame buffer (Ikons RDS3000) acquires 488 × 380 8-bit image data from each of the two channels of video. All video signals are 1-V peak-to-peak RS-170 compatible with external sync.

C. Motorized Lenses
To operate under a wide variety of illuminations, object reflectances, and object distances, the camera lenses must be easily adjustable. For different lighting conditions, the aperture must be able to change its diameter. To accommodate various object distances and image resolution requirements, the lens must be able to change its magnification (focal length, zoom). Of course, the images must be sharply focused, so the lenses must have adjustable focusing distances.

Table I summarizes the lens specifications of the two motorized lenses (FUJINON C6 × 17.5×). Attached to each zoom, focus, and aperture ring are variable-speed dc servomotors and optical shaft encoders; a total of six servomotors and six encoders are mounted in the cameras.

D. Lights
Different objects have very different reflectance properties, so flexible control of the scene illumination is crucial. Our approach is to use a large number of illumination sources with independent intensity control. To accomplish this, an 8085 microprocessor drives 10 ac outlets connected to lamps. Input is provided through an RS-232 line. The power output is varied by changing the duty cycle through optocouplers and power triacs.

III. DEVICE CONTROLLERS
Special-purpose devices control the camera platform, the lenses, and the lights. These device controllers are microprocessor systems,

conceptually located between a host processor and the actuation and sensing devices. The host processor (a VAX 11/750 running ULTRIX) is used to develop software and to download commands. The microprocessor is responsible for handling all interactions with the host, and controlling the platform motion, lens parameters, and lights’ intensities.

A. Camera Platform Controller
Fig. 5 provides an overview of the camera platform controller [10], which is based on an 8085 microprocessor performing three tasks: communicating with the host processor, running control algorithms, and actuating the four ac servomotors.

Communication with the host takes place over a serial line. A motor based on a shift-reduce parser receives commands from the host (or directly from a terminal) and translates them into read and write commands to memory-mapped analog motor driver boards.

The monitor is interrupt driven for high performance and table driven for generality and flexibility. The low-level control scheme is from the pd regime. The difference between measured and desired positions (position error) is computed. From a history of past positions, a velocity is calculated (not measured, since there are no tachometers). An initial drive signal is computed by using the position error as an index into a drive signal table. If the position error is large, then this value becomes the final drive signal. If the position error is small, then we avoid overshoot by ensuring that the velocity is not too large by retrieving a velocity damping value from a table using the velocity as the index. The final drive signal is the sum of the initial drive signal and the looked-up damping value. For generality, each motor has its own table of driving signals.

The phase difference between two ac currents, the reference and drive signals, determines the ac servomotor speed. The reference signal always is applied to every motor; the drive signal is only sent when it is desired to move a motor. The microprocessor actuates a servomotor by writing to a motor-driver board, which converts the
written value into a drive signal, stepped up by a transformer to the
proper voltage, and then applied to the servomotor. Wire-wound
potentiometers measure the joint position. The potentiometer output
is sent to an 8-bit A/D converter, whose output the microprocessor
then reads.

Using this control scheme, the horizontal and vertical axes have
position errors of approximately 2.5 mm, while pan and tilt have
angular position errors of less than 1° [10]. This accuracy is
acceptable for many applications.

B. Lens Controller

The lens controller adjusts the focal lengths, focusing distances,
aperture diameters, and vergence angle between the cameras. As
shown in Fig. 6, the controller consists of an 88/25 single board
computer (whose CPU is an 8088), seven motor-driver boards, an
interface board, and a front panel of switches. Like the platform
controller, the lens controller has three main duties: communication,
control, and actuation.

The lens controller has two RS-232 serial lines for communication.
One line communicates with either a host processor or human
operator via a terminal. The data received from the host port are input
to a command interpreter, which is based on a deterministic finite
automaton. The other serial line is a diagnostic port, to which host
input, calibration, and error messages are sent.

The controller has two modes, on line and local, selected by a
switch on the front panel. When the lens controller is on line, all host
input is processed as described earlier. Once in local mode, all host
input is ignored, and the controller listens to the front panel switches
controlling the focus, zoom, aperture, and vergence motors. For
precise manual control there is also a speed switch with settings for
slow, medium, and fast. An interactive debugger allows one to drive
any of the motors, to check a motor’s position, to clear a motor’s
position counter, and to (re)compute all of the motors’ control
parameters.

The low-level control scheme uses a table lookup implementation
of proportional control. Each motor has its own error class and drive
tables which are used to compute a drive signal. The error class table
maps position errors onto 16 error classes which are used as an index
into the drive table to determine the final drive signal. The error class
and drive tables are loaded based on motor profiling information
obtained from the calibration and debugger routines. For flexibility,
these tables can be changed on the fly from the host.

Each of the seven motor-driver boards actuates a dc servomotor. A
motor is moved by writing a byte to a register on the appropriate
motor board. This number is interpreted as the seven low bits
representing the number’s magnitude, while the eighth bit determines
the sign. The board then sends a pulsewidth-modulated drive signal
to the motor. The width of the driving pulse is proportional to the
magnitude of the number sent, and the sign bit determines the polarity
of the signal. The motor board also keeps a 16-bit count of the motor
position. Every time the motor turns, a signal is generated that either
increments or decrements the position register, depending upon the
direction of motor revolution.

C. Light Controller

The light controller has two main duties: communication with the
host and actuation of the lamps through ac power circuitry. No
intelligent” control is necessary.

The input data are a stream of byte pairs. The first byte of each pair
signifies the channel to be adjusted. The second byte represents one
of 50 possible intensity values. The controller accepts data at 9600
Bd, but it can only process 100 commands per second. However, the
controller can buffer 50 commands internally so that the 100
commands per second limit only has to be followed within a deviation
of 50 commands. This imposes no practical limits on the experimenter
since it is difficult to imagine a scenario requiring strobe lights.

IV. DISCUSSION

The camera system provides a very flexible image acquisition
device. Although primitive in many respects, it affords sensing
capabilities which many animals lack, as the following comparison
suggests. Attached immediately behind the eyeballs of most animals
are muscles. Though all vertebrates have six muscles per eye, they
use them to perform rather different eye movement. Amphi
forms generally do not move their eyes. Fish and reptiles do move them, but
move each eye independently in most circumstances. Birds may move their eyes a little, but generally lack room in their skull for such movements. They do scan their surroundings, shifting their foveas about, but they must do this by turning their heads rather than their eyes. Mammals alone among vertebrates move their two eyes together to look in the same direction. While other mammals show less eye movement than the higher primates with their distinct foveas, they generally attempt to use their binocular vision, bringing both eyes to bear on the same object. In any case, they are unable to move the two eyes independently. Therefore, by these comparisons the camera system can be considered to be fairly advanced, and in this section we illustrate the benefits of its flexibility by describing some of its uses (see also [3]).

An experimenter can operate the camera system with varying levels of sophistication. The most direct method of setting the seven lens controller parameters is the front panel of manual switches. A less direct method involves interactively entering commands with numeric parameters to the host processor, which parses the input and commands the appropriate device controller to actuate the motors or lamps. Programmed function keys on terminals allow commonly performed operations to be executed with a single keystroke. The most sophisticated mode of operation is fully automatic, where the camera system—autonomously, dynamically, and adaptively—executes tasks with no intervention by the experimenter.

As an example of the last mode of operation, we have implemented a completely autonomous "wake-up" procedure to dynamically adjust the aperture diameters and lamp intensities so that the image contrast is as sharp as possible. Thus because of the variable aperture size and programmable lights, it is always possible to capture high-contrast images. Other examples of autonomous operation are described by Krotkov [4].

Since the platform is mobile within the gantry, it is possible to acquire images from a wide range of viewpoints. This is useful for tracking objects moving out of the current field of view and for viewing objects which are occluded from certain viewpoints. The ability to "take another look" is also useful for verifying information derived from other viewpoints. As an example, consider applying the camera system to the task of visually guiding a robot gripper moving through a crowded environment. The ability of the cameras makes possible dynamic repositioning to overcome occlusion of the gripper and redundant position estimation over time.

The adjustable focusing distance makes it possible to acquire sharply focused images for many scenes under a wide variety of viewing conditions. The quality of focus in an image or image region significantly affects the results of many vision algorithms, and sharp detail is quite important for analyzing textures, detecting edges, and stereo matching. Focus can also be a depth cue, and with an adjustable focusing distance it is possible to compute range from focusing [5].

The zoom lenses accommodate both a large range of object distances and a spectrum of features and textures. We can take advantage of the variable focal lengths by using a small magnification to identify regions of interest, and a greater magnification for detailed processing of interesting regions. We now employ this approach for automatic focusing: identifying interesting regions at low resolution and then zooming in for fine focusing.

The vergence mechanism allows control over the disparity objects at any distance and defines the common field of view of the two cameras. Both of these properties add constraints which can be used advantageously in solutions to the correspondence problem faced by stereo algorithms.

These uses of the camera system clearly demonstrate its virtues, but it has equally clear limitations. It was built at a low cost, largely from surplus equipment and spare parts. As a consequence, some of its components could be more precise and reliable. As an example of the former, we find that the 8-bit resolution of the four platform potentiometers is insufficient for highly accurate platform positioning. As an example of the latter, we find that the ac servomotors slightly change their operating characteristics as they heat up due to the constant presence of the 110-V reference signal. In the future we plan to replace the entire ac system with a dc system modeled on the lens controller.

In summary, we have presented a camera system designed and built as a flexible image acquisition device for computer vision and robotics research and applications. We described the physical camera system, the device controllers, and the control algorithms. It has proven to be a useful tool, and in the future we expect to employ its versatility as a valuable resource in our computer vision and robotics research.

ACKNOWLEDGMENT

Our thanks to Prof. Ruzena Bajcsy, Dr. Samuel Goldwasser, Dr. Ralf Kories, Milan Pantelic, David Zygmunt, Matt Donham, and Nathan Ulrich. This work was performed at the University of Pennsylvania, Philadelphia.

REFERENCES