Proceedings of the IASTED International Conference
Signal and Image Processing
October 27-31, 1998, Las Vegas, Nevada - USA

REAL-TIME TRACKING SYSTEM USING C80 DSPs AND A BINOCULAR ROBOTIC HEAD

M. CASTRILLÓN-SANTANA[1], C. GUERRA-ARTAL, J. HERNÁNDEZ-SOSA, J. ISERN-GONZÁLEZ,
A. DOMÍNGUEZ-BRITO, F. HERNÁNDEZ-TEJERA, J. CABRERA-GÁMEZ
Departamento de Informática y Sistemas
Universidad de Las Palmas de Gran Canaria
Las Palmas
SPAIN

Descriptive keywords: Motion Detection and Estimation, Active Vision, Applications of image processing to Robotics.

Abstract

An active vision system to perform tracking of moving objects in real time is described. The main goal is to obtain a system integrating off-the-self components. These components includes a stereoscopic robotic-head, as active perception hardware; a DSP based board SDB C80, as massive data processor and image acquisition board; and finally, a Pentium PC running Windows NT that interconnects and manages the whole system. Real-time is achieved taking advantage of the special architecture of DSP. An evaluation of the performance is included.

Introduction

Active Vision Systems can be considered as dynamical systems that integrates control camera parameters, sensor motion and visual processing to simplify, accelerate and perform robust visual perception. Under this paradigm are faced up problems that were unmanageable with previous works in Image Processing. This is feasible thanks to the possibility of adaptation offered through mechanical devices such as robotic heads and motorized lens and also, to the benefits of specific image processing hardware.

Research and Development in Active Vision Systems [1], [2] has been a mainstream area in Computer Vision in the last years, mainly by its potential application in different scenarios where real-time performance is required such as robot navigation, surveillance, and visual inspection, among many others. Several systems have been developed during last years using robotic-heads for this purpose, or in general, sensors with egomotion.

Image processing procedures in continuous real-time computer vision applications have a bottleneck in the computation of large amount of input data, since they only have a reduced and bounded amount of time to issue results. Real-time conception depends on the latency of the events observed. Here, Real-time is assumed to be frame-rate, that is 25 frames per second.

There are three obvious main ways to overcome this temporal constriction: first, using high performance hardware to deal with incoming data; second, reducing by any filtering method the amount of input data to process, and third, optimizing image processing procedures

During the last years, many solutions have been developed based on high performance hardware. At the beginning, this hardware was special purpose and expensive, but in recent years this hardware is becoming less specific as well as less expensive. Most of currents Active Vision Systems have based their design on hardware such as Transputers or Digital Signal Processors (DSPs) networks, commonly, using VME bus to interconnect the system.

In the last three decades, different image processing algorithms has been developed. However, if real time constriction is imposed, it implies a strong limitation in the process that can be performed by a specific hardware architecture solution. Also, special purpose hardware forces the development and adaptation of procedures to get a computationally efficient solution that optimize all the architecture features.

The use of filtering techniques over the input visual data is a solution that can be used in analogy with biological vision. One of the filtering techniques used is based on a variable resolution retina. In this case, the resolution of the retina decreases with the distance to the center of the sensor. So, it is possible to establish a separation of the retinal area in two zones: foveal or high-resolution area and peripherical or low-resolution area. Using this configuration, control of perception is necessary, mainly

[1] Correspondence to Edif. Dptal. Informática y Matemáticas,Campus de Tafira,35017, Las Palmas de Gran Canaria, Gran Canaria, SPAIN. Email: modesto@mozart.dis.ulpgc.es; www: http://mozart.dis.ulpgc.es.

by means of an attention mechanism, in order to keep the foveal zone coincident with the interest area, and obtain a higher resolution over points of interest. This higher resolution area size is the result of a tradeoff between amount of data to process and amount of time available. This windowed area of the image is labeled as fovea [5][6].

In this work, a global solution is faced up, as many cases in Active Vision, using a combination of the three previously mentioned improvements.

The active vision system presented in this paper has been developed using off-the-shelf components: last generation C80 DSP, a PCI bus Pentium processor and a stereoscopic robotic-head. These elements compose a heterogeneous hardware platform that has allowed to design and build a cost effective system whose first prototype is able to perform movement detection and tracking of mobile objects in real-time.

Tracking

Biological Visual Systems are inherently active and are able to perform tasks of different level of complexity. Basic tasks such as movement detection, light detection or target tracking. More complex tasks such as object recognition, face identification or visual guidance. In many cases complex tasks make use and are supported by mechanisms provided by simple tasks.

For artificial systems and focusing on simple tasks, it is noticed these systems must act in response to events denoted as variations in the visual aspects of the environment. This kind of behavior is completely reactive and does not need high level processing. An example of such situation could be a movement of an element in a previously static environment.

Tracking is a basic process in visual systems whose main goal is to keep the gaze on an object of interest previously located and fixed, pursuing it when the object is moving in the field of view [8]. It is also desirable that tracking keeps the object of interest centered in the image, moving adequately the sensor. In order to increase the robustness of the tracking system and to obtain a smooth tracking trajectory. It is also helpful to incorporate movement estimation [9] to the process.

It should be noticed that this basic process does not manage situations such as occlusions or very fast movements. Higher level processes should manage such situations. Existing methods to perform tracking can be classified as[6]:

- Filter-and-follow methods based on the filtering of the image previous filter, which give as result only highlighted areas in the image where the object of interest is. To do this, it is necessary the object to be distinctive enough to obtain good results. Unfortunately, this is not common in the majority of the objects in the real world. [10][11][12]
- Local area correlation methods perform a correlation over the image; this method offers a maximum value where the image is more similar to the searched pattern.[13][14]
- Feature correlation methods look for object's features and then try to identify those features in following images [15].

The system presented makes use of the second approach, i. e., correlates a pattern over the image. This approach repeats the correlation over the whole fovea whose size depends dramatically on the hardware available. With all this, tracking has been designed based on these three following steps:

1. A movement is *detected*. In this stage a proper tracking process is not being performed yet. In the detection process of the developed system, the movement zone of the image is detected and extracted, obtaining as model an image window containing this moving object. So, the model is iconic. The system does not search for a known model. It just waits for something. This will be the model for next step. More elaborated model based approaches are common in the literature.

2. The attention window is focused, by means of a saccadic movement, on the area where the movement was detected. This area is *fixed;* that is, the cameras have been pointed to the same point in space [7].

3. Pursuing is performed mean successive detection of the object and fixation processes. In this work, detection is performed mean a simple correlation algorithm. Once this algorithm has detected the object a movement of fixation is carried out. Approaches for performing this detection process are basically model based. The differences between a saccadic movement of the second step and a normal fixation movement are that in the first one the movement of the camera is so fast that it would provoke a blurred image. So, processing image is not performed during a saccadic movement. Meanwhile in the second kind of movement, these are much slower and processing image can be done.

Architecture

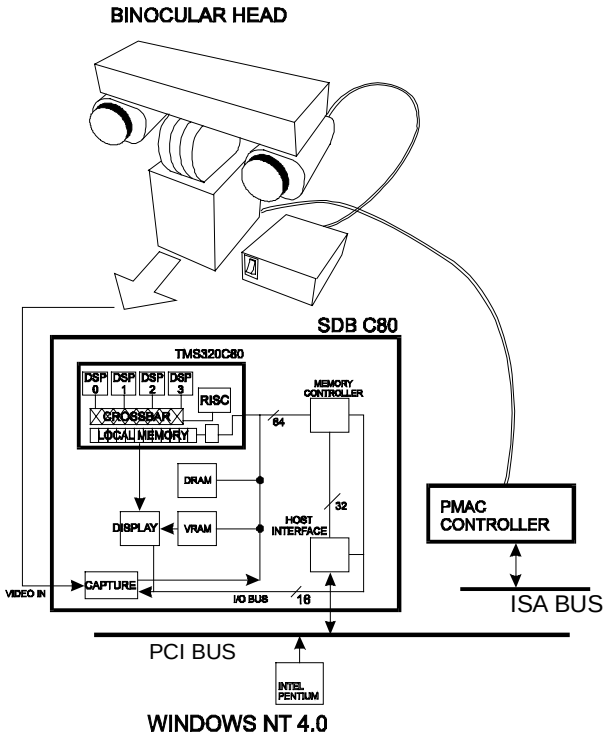
The active vision system presented has been developed using heterogeneous hardware and would perform tracking in real-time. As it was mentioned above hardware affects directly the implementation of the prototype. The main feature of this system is the design of an architecture based on the interaction of several hardware components to compose a full-fledged perception-action system. As described in the figure the system can be decomposed in the following elements:

- C80 DSP for heavy image processing.
- Pentium processor for system management.
- A stereoscopic robotic-head as active hardware.

Also, this architecture can be also decomposed based on the task each hardware component performs. That is, perception subsistem, action subsystem and communication. On one hand, the perception subsystem consists of a couple of TMS320C80 (C80) development boards [16] to perform image acquisition and tracking of the target object. It means this subsystem is responsible of the image processing of the system. Of course there is a wide range of hardware solutions for image processing: from expensive custom VLSI design or new attraction reconfigurable FPGAs, through traditionally based on DSPs or transputers, up to economic general-purpose processors. DSPs and transputers has been chosen for most experimental active head-eye systems due to the ratio processing power/cost offered. The launch of this DSP with its architecture and capabilities for image processing offers a new point of view for this kind of systems. C80 is a single-chip parallel processor that incorporates following characteristics:

- Four identical fixed point DSPs or Parallel Processors (PPs) that offers a splitting ALU.
- A floating point RISC Master Processor (MP) which is in charge of controlling the other four DSPs and communicating with the host PC.
- 50 Kbytes SRAM cache memory with 32 K shared among the processors.
- A high speed Crossbar Switching Network that connects PPs cache and processors.
- An integrated Transfer Controller that allows many image-oriented transfers.

On the other hand, the action subsystem is a commercial motorized robotic head. It offers four mechanical degrees of freedom: pan, tilt and two vergences, plus other six optical degrees: iris, zoom and focus for both lenses. All of these degrees are controlled via a commercial controller board that provides tools for reading robot position and commanding a new pose or lens configuration. Actual position of the robotic head is a critic feature of the system in order to related robot pose and image processing results. The action task carries out basically the translation of results coming from the perception subsystem into movements of the head.



Both subsystems need a layer for communication each other. A PC with a Pentium processor that hosts the perception and action subsystems carries out the control of the operation cycle using a PCI bus. Also the Pentium processor would be responsible of solving coordination problems the perception and the action subsystems. Synchronization is a clear example of this circumstance. It is obvious that processing is performed after the acquisition of the image has happened. It is also obvious that is necessary to put in correspondence the Kinematics State of the head and the moment an image was acquired. In any other case the system could make use of inconsistent data. It should be noticed the system should be able to adapt any of the head degrees of freedom according to processing results. Whenever exists an inconsistency in the data the system can produce unexpected results.

Concluding the perception subsystem acquires and performs the visual processing on the image. The heavy active vision algorithms (detection and correlation) are carried out on each image by a C80 chip, while estimation, robotic-head commanding and synchronization are responsibility of PC.

System Cycle of Operation

Based on the hardware described in the previous section, a prototype has been implemented to perform real-time tracking. System tracking life cycle is as follows:

1. Movement Detection:

   Detection of any movement in a new image, fixing a target when a main movement is observed. This module focuses the attention of the system in those areas where motion is detected. In this prototype it is supposed that movement detection is performed when robotic head is stopped, this restriction simplifies the problem due to movement detection is obtained by subtracting two consecutive images and undersamplig the output image calculated. Current work faces the problem with robotic head egomotion

2. Object Fixation:

   Both cameras of the robotic head moves in a saccadic mode, accommodating lens and pointing out the object to pursuit.

3. Pursuing:

   Once the object has been detected and fixed a pursuing state takes place. Considering the implementation of tracking algorithms more in detail, two major tasks are distinguished: correlation and target's trajectory estimation. Target movement is estimated and the mechanical elements of the system are adapted in consequence.

   3.1. A new image is acquired.
   3.2. The correlation is performed on each PP for a quarter of the image searching the target. The correlation algorithm searches a previously chosen pattern only on the fovea of the image just for improving the system performance.
   3.3. MP integrates partial results and stores best match.
   3.4. Best match indicates if the object is lost. In that case, the system starts its operation at the first step.
   3.5. When the object is still controlled, PC, who is responsible of estimating target's future position, reads position results from each image. After estimating new position PC commands mechanical devices. Trajectory estimation is necessary for pursuing an object; this technique allows the system moving in advance, i.e., anticipating object movement. For this purpose an alpha-beta filter which is an adaptation of a second order Kalman filter [7] for certain conditions (Kinematics Systems and temporal invariant behavior) is executed using the previous position, velocity and acceleration data to predict the trajectory of the object in short term.
   3.6. According to the chosen policy the target is updated and 3.1 is next step.

As it was mentioned previously correlation is the main image processing task of the system. This is a fast and simple technique for tracking with some known problems. Simple improvements increase efficiently the performance of the system just using a pattern updating politics. In order to optimize the use of the given architecture for image processing algorithms, it has been designed a parallel correlation algorithm [17] that exploits C80's specific architecture for this purpose. Dividing in a special way the rows of the fovea by the number of PPs, each PP works completely balanced. Using this philosophy all PPs work with exactly the same amount of data located in their cache memory and without idle times. The MP RISC processor integrates the partial results issued by the PPs.

Another point of interest in this schema is the fovea. Commonly fovea is located in the center of the image, but in this work a relocatable fovea has been considered in order to make use of processing power. Using this approach it is possible select visual attention [18], so the system can follow an object close to the border of the input image providing the mechanical devices more time for its response.

Experiments

The correlation algorithm allows a variable size for the searching area and a variable pattern size. It is possible testing the performance of different size configurations. Some experiments were performed on real world images, mainly tracking people. Performance of correlation with a pattern of 24x24 pixels in an area (fovea) of 80x80 pixels, results in 31 milliseconds per frame, this gives a range over 25 images per second, which is real-time, as we have defined here. The extra time, 9 milliseconds, can be used for other tasks.

It should be known that mechanical devices do not offer this response time, and the relocatable fovea cushions slower movements of the robotic head. The robotic head can be commanded every 25 milliseconds, however the time it takes to reach the commanded point is longer.

A previous detection module captures the object to follow mean a calculus of optical flow. This calculus takes 90 milliseconds over an image of 176x120 pixels running on a Pentium Pro 200mhz.

On the figure some frames of a tracking sequence are shown. The person is walking in an indoor environment. On this example, it can be seen that the correlation algorithm is quite robust even when the person is rotating over his vertical axis.

Conclusions

C80 DSP is actually an engine that makes feasible performing tracking in real-time. The system presented based on off-the-shelf components shows a good behavior for real-time tracking in 3D using a C80 DSP board for each camera to perform correlation of the pattern to follow.

A continuos operation system has been designed integrating very different components, and making a PC running Windows NT 4.0 master of this real-time tracking system.

This proposed architecture has been designed to add other non-reactive processes to perform more complex tasks. Future works will face up the development of tasks that use the results coming from this tracking system such as people identification and recognition.

Acknowledgements

This work has been partially supported by Spanish CICYT Project TAP9-0288.

References

[1] Aloimonos, J. Weiss, I. and Bandyopadhyay, A., Active Vision, *International Journal of Computer Vision*,1(4):333-356, 1987.

[2] Bajcsy R, Active Perception, *Proceedings of the IEEE Workshop on Computer Vision,* 76(8):996-1005, 1988.

[3] Kourosh Pahlavan, Active Robot Vision and Primary Ocular Proceses, *CVAP*, 1993.

[4] Ulf M. Cahn von Seelen, Brian C. Madden, A Modular Architecture For An Active Vision System Using Off-The-Shelf Components, *GRASP Laboratory, Department of Computer and Information Science, University of Pennsylvania*, 1996.

[5] F.Panerai, C. Capurro, G. Sandini, Space variant vision for an active camera mount, *TR 1/95, LIRA-lab-DIST University of Genova*, Italy, 1995.

[6] M. Wessler, L.A.Stein, Robust Active Vision from Simple Symbiotic Subsystems, *Technical Report, MIT*, 1997.

[7] Kjell Brunnström, Jan-Olof Eklundh and Tomas Uhlin, Active Fixation for Scene Exploration, *International Journal of Computer Vision*, vol. 17, pp. 137--162, 1996.

[8] Crowley J. L., Christensen H. I. (eds.), Vision as Process. Springer-Verlag, Berlin, 1995.

[9] Yaakov Bar-Shalom, Xiao-Rong Li, *Estimation and Tracking: Principles, techniques and software*, Artech House, Boston, 1993.

[10] Anne Wright. A high speed low latency portable vision sensing system. *In D.P. Casasent, ed., Intelligent Robots and Computer Vision XII: Algorithms and Techniques,* vol. 2055, 263-270, SPIE August 1993.

[11] H. Kimura and J. J. E. Slotine. Adaptive visual tracking and gaussian network algorithms for robotic catching. *In Advances in Robust and Nonlinear Control Systems*, 67-74, 1992.

[12] S.A. Brock-Gunn, G.R. Dowling and T.J.Ellis, Traking using colour information, *TR TCU/CS/1994/7, City University London*, 1994.

[13] I.D.Reid and D.W. Murray, Tracking foveated corner clusters using affine structure, *Proceedings of the Fourth International Conference on Computer Vision*, pages 76-83, Berlin, IEEE Computer Society, 1993.

[14] E. C. Hildreth, Computations underlying the measurement of visual motion, *Artificial Intelligence*, 309-354, 1984.

[15] H.Inoue, T. Tachikawa and M. Inaba, Robot Vision system with a correlation chip for real time tracking, optical flow and depth map generation, *Proceedings of the 1992 IEEE International Conference on Robotics and Automation,* 1621-1626, France, IEEE Computer Society, 1992.

[16] *TMS320C8x Software Development Board Technical Reference*, Texas Instruments, 1997.

[17] C. Guerra-Artal, M. Castrillón-Santana, J.D. Hernández-Sosa, J. Isern-González, J. Cabrera-Gámez, F.M. Hernández-Tejera. A C80 DSP-based Active Vision System for Real-Time Tracking. *Proceedings ICSPAT*, 1998.

[18] Posner, M.I., Orientation of attention, *Quart. J. Experimental Psychology*, 32:3-25, 1980.