Towards a Versatile Computer Vision System

LUCIANO DA FONTOURA COSTA

IFQSC – Instituto de Física e Química de São Carlos
Universidade de São Paulo
Caixa Postal 369
13560 São Carlos, SP, Brazil
FAX: +55 (162) 71-3616
e-mail: luciano@uspfsc.ifqsc.usp.br

Abstract. This paper presents an ongoing project aimed at the development of a versatile and powerful model-based computer vision system. The basic processes and typical data representation to be used at each of its constituent levels as well as their possible implementations are also discussed in some detail.

1 Introduction

Although the current and future importance of computer vision can hardly be disputed, the development of computer vision systems – CVSs – with ability comparable to that of the primate vision system – PVS – has proven to be an elusive objective despite all the researches and developments in the field. The principal reasons commonly identified as barriers to general and powerful computer vision include: (a) lack of understanding about the PVS [Marr (1982)], which has been emulated from the very beginning of computer vision, (b) the memory and processing power allowed by the current integration technologies are believed to fall short of those which are actually needed and (c) the grey-level and spatial sampling resolutions are by far too coarse [Binford (1981)].

Another criticism to the developments in computer vision is that the performance of the involved techniques has not been properly and formally assessed as it has, for instance, in digital signal processing [Binford–Jain (1991)]. It is commonly believed that the achievement of more intelligent CVSs is conditioned to integrated advances in all of the previously identified fronts [Marr (1982)]; dedicated and simplified CVSs have proven to be a valuable test tube for experimenting with new principles, algorithms and architectures.

An important strategy to be considered for the development of CVSs is the integration of top-down and bottom-up approaches [Marr (1982)]; which implies that particular attention should be paid to the design and implementation of a control structure able to provide effective bidirectional data exchange between the various CVS processing levels. Not less important is the early identification of the bottlenecks in the adopted architecture, in order to allocate the hardware resources in such a way as to achieve a reasonable balance between the processing rates of each level [Costa–Slaets (1991)].

The current paper presents a discussion about a project aimed at developing and implementing a complete and versatile CVS, trying to follow the guidelines above while avoiding the identified shortcomings. Particular attention is focussed on the cross-fertilization principle, where insights are gathered from biological systems in order to design better computational techniques, which can eventually lead to advances also in biological vision. The CVS is intended to be designed and implemented gradually from a simple system for polyhedra recognition towards more powerful and versatile systems through the addition of new techniques and processing levels, more powerful hardware implementation, and consideration on other image features such as texture, colour, other curves, motion and even integration with other senses such as audition. It is important to observe that the present approach will not be limited to the emulation of natural mechanisms; for instance, it is easier to derive the distances of the objects in an image by using a telemetric system such as a those based on laser-scanning, e.g. [Gonzaga–Roda (1990)], rather than inferring it by stereoscopy.

The following sections provide an overview of the processes, typical data representation and architectures to be adopted in each of the intended CVS constituent levels; special attention is paid to the preliminary polyhedra recognition CVS presented in Figure 1.

2 Image Acquisition and Pre-Processing

In the PVS, the image is acquired and pre-processed by the eyes and elementary structures such as lateral

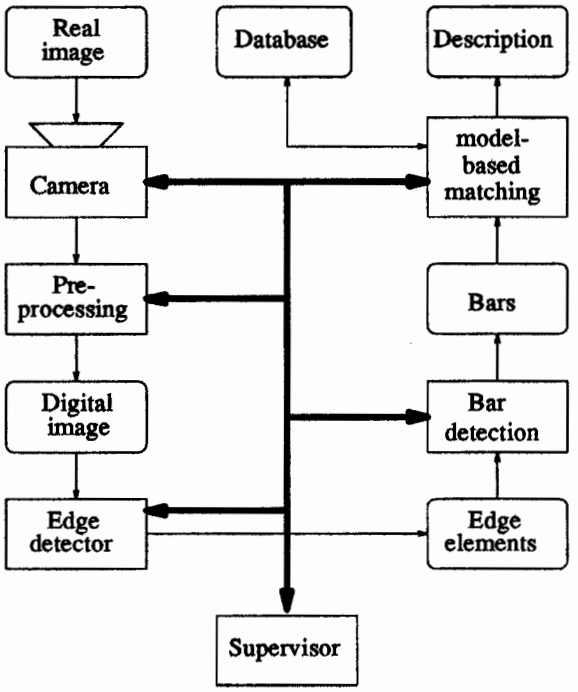

Figure 1: Block diagram of the preliminary CVS.

geniculate nuclei and superior colliculli. Particularly interesting features to be found at this level include: the eye tracking ability, the separate vision systems respectively defined by cones and rods and the very high spatial resolution for image acquisition: there are about 6.10^6 cones and 120.10^6 rods in each eye [Lindsay–Norman (1977)] providing a substantially better sampling than in conventional computer vision systems. An important property of the primate visual system is that it is rather less tolerant to translation than generally believed: the eye can only analyse the small part of the visual area projecting onto the fovea – the analysis of te whole image is achieved through eye scanning movements.

Typical computations in this level include image acquisition and pre-processing (e.g. noise reduction, histogram equalization), which implies a substantial amount of data to be stored and processed. However, such processings are usually local and relatively simple and thus suitable for parallelization. Special attention must be paid to designing ways of allowing the visual representations at this level to be effectively accessed by higher-levels processes.

In the preliminary CVS, a non-interlaced camera will be used to acquire images within the prespecified resolutions (preliminarly 256 grey-levels and 512x512 images) and the pixels will be sent directly to the pre-processing stage. Future developments could include texture/color encoding, cameras capable of much higher spatial and grey-level resolutions and telemetric scanning systems.

3 Low-Level

In the PVS, this level corresponds mainly to the processes that take place in the retina, whose function is believed to be specially oriented to edge detection. The importance of the information provided by the image edge elements can not be overstressed [Lindsay–Norman (1977), Marr (1982)], as illustrated by our ability to promptly recognize faces in cartoons and charicatures. Also interesting at this stage is the suggested multi-channel response of the retina to different spatial frequencies [Marr (1982), Goldstein (1989), Lindsay–Norman (1977)].

The low level on CVSs typically includes operators for edge element detection, 'thinning' and texture representation, this last being commonly based on Fourier transform. As in the previous level, the processes within this level must operate over a large amount of data, typically of $O(N^2)$, but, on the other hand, they are usually simpler than the processes at the higher levels and more suitable for implementation in parallel and/or dedicated structures.

The preliminary version will include a simple edge detector analogous to the Sobel operator, which will be implemented in dedicated systolic hardware, similar to that described in [Costa–Sandler (1990)], in order to allow an execution rate as high as 10^8 pixels per second. A 'thinning' processor may also be included in order to reduce the amount of redundant edge elements.

4 Intermediate-Level

In PVSs, intermediate processing occurs at the visual cortex and includes the detection of bars and corners which can be or not in motion [Goldstein (1989), Lindsay–Norman (1977)]. The amount of visual information, mainly the edge elements, is commonly of $O(N)$ at this stage. Nevertheless, the detection of bars and corners may still imply substantial processing requirements since the detection of curves with degree D typically implies processing complexity of $O(N^{D+1})$.

An effective and relatively simple technique for bar detection has been developed [Costa (1992), Costa–Sandler (1993)] which is based on a variant of the Hough transform for digital straight-line-segment detection [Illingworth–Kittler (1988)] and which includes two post-HT processings: (a) a connectedness analysis which can simultaneously confirm the straight bars indicated by the Hough transform and determine their endpoints and a (b) a merging stage which

links digital straight line segments which are broken or replicated during the Hough transform. This framework for digital straight line segment detection can be implemented without great difficulties in discrete hardware or VLSI in order to achieve high execution rate [Costa (1992), Costa–Sandler (1990)].

Future developments should include the addition of techniques for detection of circles and higher order curves as well as the application of the Hough transform to derive higher level information about the relative position of the bars, i.e. parallel/perpendicular bars and intersection between bars [Wahl–Biland (1986)].

5 High-Level

Unfortunately, unlike in the previous levels, there are few accurate biological evidences about the processes taking place beyond the visual cortex. There are indications about neural structures capable of hand and even face recognition, but almost nothing is kwown about how they are organized. It is nevertheless accepted that the cortical regions dedicated to language seem to play important role in vision; in fact more than half of the overall neocortex is connectect to vision processing [Blakemore (1990)].

CVS approaches typically adopted at this level are based on two principal underlying principles: syntactic processing, such as merging and parsing or the geometrical/combinatorial transformations and comparisons of the data structures supplied by the intermediate level with a database [Lin– Fraser (1991)].

The preliminary version of the intended CVS will incorporate a syntactic recognizer, possibly to be implemented in PROLOG or C++ running on a transputer network, which should be able to provide a description of objects in the original image (e.g. polyhedra, a book, a face, etc) based on the comparison of the digital bars supplied by its intermeditate level with entries in a database, which characterizes a model-based recognition approach. An ongoing alternative line of research consists in using neural networks for pattern recognition. Future developments will include the detection of solids bounded by arcs of circles or other curves as well as the information provided by additional features such as texture, color and motion.

6 Inference Level

According to the classification of vision levels adopted in this paper, this last level corresponds to the combination of the objects detected by the previous level in space and time, which, in powerful CVSs, can also lead to the determination of the environmental context and allow the confirmation of uncertainly classified objects, possibly by demanding new features to be extracted by the lower levels.

The respective PVS processes taking place at this level have so far been suggested by taking into account only psychological evidence [Goldstein (1989)] and [Lindsay–Norman (1977)]. Actual CVSs have not yet reached a stage of development which comprehensively incorporates the processes in this level, though inference and contextual analysis have been often addressed by artificial intelligence.

The preliminary CVS will not include any inference processing, since it is intended just for polyhedra detection. The inference mechanisms to be incorporated in later versions are still open to further research. It is likely that the assessment of the performance and needs of the preliminary system as well as the intended applications can contribute to the specification of this highest level.

7 The Overall Control Structure

A very important part of the PVS is *attention*; of which the closest corresponding structure in CVSs is the overall control system or supervisor. Attention seems to be an inherently selective and sequential process taking place under control of the highest levels in the PVS. It can be easily verified that we analyse images by parts, paying special attention to those parts which seem to convey more relevant information while trying to integrate such parts into a meaningfull whole. In such a processing strategy, the lower processing levels are often requested to reprocess parts of the image with refined accuracy.

This kind of selective attention can only be reasonably emulated in a CVS by providing effective interconnmunication between the CVS levels, which implies that the data structures in any level could be effectively accessed by any of the processes in the overall system, possibly under control of a supervisor. The many data access conflicts which are implied by the adoption of such strategy can be alleviated by partitioning the data structures in each level (the image, for instance, can be partitioned into uniform segments and the bars in separate lists). There is no definite strategy and architecture defined for the CVS supervision at the moment, these should evolve naturally during the analysis of the preliminary CVS overall operation. It is nevertheless expected that the CVS will primarily include mainly feed-forward interconnections between adjacent levels and thus perform in a bottom-up fashion, except for the post-HT connectedness analysis, which requires returns to the edge-detected image.

8 Concluding Remarks

This communication has discussed the principal aspects of an ongoing project aimed at designing and implementing a complete and versatile model-based computer vision system. The underlying phylosophy consists in trying to address as many of the above identified guidelines in order to overcome most of the shortcomings which have been identified in computer vision. The system is intended to be developed gradually from a simple system for polyhedra detection through the addition of new processes and consideration of other image features, which provides an ideal background for scientific initiation and post-graduation projects as well as possibilities for collaboration with other researchers and institutions with know-how in specific areas such as psychology, artificial intelligence and databases. Some contacts have already been established with research groups in Sweden (parallel image processors), Japan (neural networks, artificial intelligence and multiple-valued logic) and Brazil (neural networks).

Currently the system includes a complete intermediate vision system for digital bar detection, implemented in OCCAM-2 in a transputer network, which has already been successfullly applied to several tasks in visual inspection [Costa et al. (1991), Costa et al. (Sep. 1991)], communications [Costa–Sandler (Sep. 1991)] and image analysis [Costa (1992)]. Designs of high-performance architectures for edge element and digital bar detection are ready [Costa (1992), Costa–Sandler (1990)] and awaiting resources for their implementation, which will possibly be done by using discrete logic in order to allow easy modification and expansion of the system while keeping the cost low. Formal and comparative performance assessments of the involved techniques by taking into account their execution rate, accuracy and robustness, in a similar way to the binary Hough transform performance evaluation described in [Costa (1992)], are also intended to be performed in order to allow the identification of the CVS aspects deserving further attention.

9 References

T. O. Binford, Survey of model-based image analysis systems, *Artificial Intelligence*, **17**, 205–244, 1981.

C. Blakemore, Understanding images in the brain, *Image and Understanding*, 257–283, Cambridge University Press, 1990.

L. da F. Costa, *Effective detection of line segments with Hough transform*, PhD Thesis, King's College London, University of London, London, UK, May 1992.

L. da F. Costa and M. B. Sandler, A complete and efficient real time system for line segment detection based on the binary Hough transform, *Euromicro'90 Workshop on Real Time*, 205-213, Horsholm, Denmark, published by the IEEE Computer Society, June 1990.

L. da F. Costa and M. B. Sandler, Application of the binary Hough transform to image compression, *6th International Conference on Digital Processing of Signals in Communications*, 56–60, Loughborough, UK, Sep. 1991.

L. da F. Costa and D. R. Andrews and M. B. Sandler, Quality control of ultrasound transducers with the binary Hough transform, *19th International Symposium on Acoustical Imaging*, Bochum, Germany, 1991 (in press).

L. da F. Costa and X. Leng and M. B. Sandler and P. Smart, A system for semi-automated analysis of clay samples, *Review of Scientific Instruments*, **62**, 2163–2166, Sep. 1991.

L. da F. Costa and J. F. W. Slaets, On the efficiency of parallel pipelined architectures, *IEEE Transactions on Signal Processing*, **39**, 2086–2089, 1991.

L. da F. Costa, *Effective detection of digital bar segments with Hough transform*, Computer Vision, Graphics, and Image Processing: Graphical Models and Image Processing, **55**, 180–191, 1993.

E. B. Goldstein, *Sensation and Perception*, Wadsworth Publishing Co., 1989.

A. Gonzaga and V. O. Roda, Reconhecimento do formato de objetos tri-dimensionais, *Jornada EPUSP IEEE em Computação Visual*, 177–187, São Paulo, Brazil, Dec. 1990.

R. C. Jain and T. O. Binford, Dialogue: Ignorance, Myopia, and Naiveté in Computer Vision Systems, *Computer Vision, Graphics and Image Processing: Image Understanding*, **53**, 112–117, Jan. 1991.

W. Liu and D. A. Fraser, Algorithms and timing for identification of objects from 2-D images, *Concurrency: Practice and Experience*, **3**, 325–331, 1991.

P. H. Lindsay and D. A. Norman, *Human Information Processing – An Introduction to Psychology*, Harcourt Brace Jovanovich Publ., 1977.

D. Marr, *Vision*, W. H. Freeman, 1982.

R. J. Schalkoff, *Digital Image Processing and Computer Vision*, John Wiley and Sons, 1989.

F. M. Wahl and H. P. Biland, Decomposition of polyhedral scenes in Hough space, *8th Joint International Conference on Pattern Recognition*, 78–84, Paris, France, 1986.