Image and Vision Computing 16 (1998) 593–595

Editorial

E.R. Davies[a], M. Atiquzzaman[b]

[a]Machine Vision Group, Department of Physics, Royal Holloway, University of London, Egham, Surrey, TW20 0EX, UK
[b]Department of Electrical & Computer Engineering, University of Dayton, Dayton, OH 45469-0226, USA © 1997 Elsevier Science B.V.

Introduction

While writing this editorial, we are very much aware that the papers published in this Special Issue are just a fraction of those that might, through the exigencies of time and opportunity, have been submitted and included. For example, many potential papers might too soon have been sent to other journals, or might have been minutes away from completion and thereby unable to live up to the deadlines for this issue. As a result, the papers that were finally included constitute a sample—a snapshot even—of the subject area. Nevertheless, with a total of ten papers, there is an argument that they must be representative in some sense, and should give a good cross-section of relevant techniques, application areas, and types of image data. Indeed, looking more closely at the situation, it is very likely that all the issues that come up in these pages have enormous generality and wide relevance to the subject. The important topics and issues include robustness against noise and occlusion, accuracy of interpretation and object location, capability for real-time operation, potential for solving both 2D and 3D recognition problems, capacity for integration with other relevant software modalities, and so on. Then there are the variants on the main projection-based transforms theme: how useful and different are structural indexing, geometric hashing, projection-based transforms, and indeed, the original Hough transform (HT)? And has the latter now been superseded in favour of its more up-to-date variants? We shall try to answer some of these questions, but first we shall look at the individual papers in more detail.

Perusing the papers

The paper by Kamat-Sadekar and Ganesan is a development of a fairly standard example of the HT for line segment detection. It starts from previous work which considers the ‘butterfly’ structure around the peak in (ρ,θ) parameter space, and makes a careful analysis of the information available when lines have multiple segments (a situation that applies to road markings, for example). The paper finds that there is sufficient information in the peak structure to define numerous collinear segments with quite high accuracy, and shows how the algorithm can be made adaptive to optimise the accuracy with which these segments can be defined. Although firmly based in the conventional formulation of an HT, the work provides a valuable extension which will make line detection possible in complex real images where occlusions abound. One of the valuable features of the technique is that it involves only one pass over the image data, so that the computational load is kept within strict bounds.

At the first sight the paper by Goulermas and Liatsis is a standard application of the conventional HT to circular object detection. However, the objects to be detected are typified by overlapping bubbles which appear in noisy real-world underwater images. Conventional HTs find considerable difficulty with such images, missing some overlapping components and also having restricted accuracy, though the basic search task is performed adequately. Thus a post-processor of some power is required, and for this purpose a carefully designed genetic algorithm (GA) implementation is found to solve the problem. Although GAs are normally considered to be slow, their use in conjunction with an HT is found to be well justified. As always with GAs, a key part of the design is the formulation of a suitable fitness function. An advantage of the approach is that in a different application such as straight line or ellipse location, the GA module would remain unchanged, although the HT component would have to be modified appropriately.

The paper by Olsen considers the computational load involved in setting up the generalised Hough transform (GHT) for effective 3D model-based vision. The objects to be detected are typified by staplers and other devices which have corners and straight edges. If the recognition task is divided into many small sub-tasks (as in a recent successful approach), randomisation can limit the number of sub-tasks to be solved, and the overall computation is $O(mn^3)$. In this approach the limitation is due to randomness, and the idea introduced in this paper is to eliminate

0262-8856/98/$19.00 © 1998 Elsevier Science B.V. All rights reserved.
PII S0262-8856(98)00103-6

much of the randomness by invoking perceptual grouping in a systematic way. In particular, corner features which are found to be linked by significant stretches of straight edge are taken as forming viable perceptual cues, and the GHT computation is reduced to $O(mn^2)$, giving two orders of magnitude improvement in a practical case. A specific advantage of the approach is that false positive rates are substantially reduced when perceptual grouping is used.

The paper by Johnson and Hebert uses a 3D surface matching and recognition technique which is developed from a combination of geometric hashing and structural indexing. The basic motivation is to get away from the use of a few possibly unreliable salient features on an object and instead to consider all the points on a surface, letting the matching algorithm decide at run-time which points are the best for recognition. The incisive new concept that is used is that of a spin-map. This is the 2D histogram that is formed by rotating the object about the surface normal L at a point on the object surface and accumulating object points in a plane passing though L. The spin-maps are used as signatures for the surface points and matched by a correlator which is able to cope well with occlusion and clutter. Problems associated with surface mesh descriptions have been solved, and the overall algorithm is able to perform in seconds, once the model hash-table has been generated (this can be done off-line), the complexity of the final matching being linear in all the relevant variables. The approach is sufficiently general to be able to cope with scene images produced by laser rangefinders, computerised tomography (CT) images, and landscapes, and thus is extremely powerful.

The purpose of the work described in the paper by Kim and Han is to perform stereo depth maps using a single line-scan camera moving along the direction of its sensor array in order to largely eliminate the correspondence problem, and to ease it further by utilising the continuity between images. By this means ambiguities can be eliminated, occlusions coped with, and depth accuracy boosted. To achieve this, a 2D ‘slit image’ is produced, which is a composite of all the 1D line-scan images. This has to be searched for straight lines corresponding to salient object points in the scene, and a standard HT is suitable for this purpose. This procedure permits the slit image to be segmented, the result providing an immediate analysis of the relative positions and depths of the objects in the scene, with the places where object occlusions appear and disappear being made explicit. An advantage of the approach is that region segmentation can be performed with sub-pixel resolution.

The paper by Kamath, Chaudhuri and Desai considers the problem of non-destructive testing. It starts from the Radon transform, which is the projection of an object in a range of directions to produce a set of intensity histograms. In CT, these are used for object reconstruction, but the raw data is usually noisy and reconstruction algorithms sometimes produce artifacts. In any case, in non-destructive testing reconstruction is unnecessary if a model-based approach can be used to detect defects bounded by lines, ellipses, etc. The paper considers how this may be achieved by direct analysis of the Radon transform in cases such as the non-destructive location of faults in metallic parts.

The paper by Fuh and Liu tackles the problem of recognising a large database of characters and symbols which can be subject to various deformations, including size modification, shearing, rotation, reflection and noise-induced distortion. An approach based on projection in multiple directions, similar to that in CT is used. Various methods are used to normalise the resulting histograms in order to eliminate the effects of pattern distortions. In addition, the projections are weighted to give pixels near the centres larger weights than those at the extremities of the patterns, because of the higher accuracies of their contributions to the histograms. Finally, by careful choice of angular quantisation and histogram resolution, recognition performance is optimised.

The paper by Fuh and Lin tackles the problem of obtaining unambiguous depth maps for objects on a worktable. Stereo matching problems are avoided by using light striping, but then ambiguities between different light stripes have to be eliminated by coding or other means. Here, the Fourier slice theorem used in CT is invoked. A solution is clearly obtainable by rotating the objects and taking a number of stripe images, since continuity can show the exact depth of each point. The problem then is how many images are required at various orientations. The paper explores the problem, and gives insight on how symmetrical objects can ease the problem, so that in the extreme case of cylindrically symmetrical objects only a single view is required. The method adopted is found to be able to cope successfully with substantial amounts of shadow and missing lines, and the spatial resolution is significantly better than for optical shuttering of the light stripes.

The paper by Gao, Niu, Zhao and Zhang is also concerned with computerised tomography, but in this case involves near infra-red (NIR) rather than X-radiation. In fact, biological tissue is a highly scattering medium at NIR wavelengths, so the simple exponential absorption law applicable for X-radiation is totally inappropriate. As a consequence, the algorithms needed for reconstruction of the sample are far more elaborate. The result of the work is a new strategy for the finite-element method solution of the optical tomography problem and involves using the Tikhonov-Miller regularisation procedure to give an acceptable approximation in an otherwise an ill-posed problem.

The work described in the paper by Startchik, Milanese and Pun recognises objects in a rather different way from HTs and other projection-based transforms. It describes a method of representing shapes which is invariant to projective transformations, and is thereby able to recognise objects that are severely distorted by translation, rotation and foreshortening: this is achieved by use of construction lines whose intersections with curves on an object are represented by invariant cross-ratio indices; for this procedure to work, the construction lines must themselves be selected in an invariant way, and for this purpose common tangents are deemed to be especially stable, leading to robust solutions. The above framework has the additional advantage of permitting the inclusion of illumination invariance, thus leading to realistic capabilities with real image data.

Some justification is perhaps required for including this last paper in a special issue on projection-based transforms. Such justification is not difficult to find. First, the motivation of the paper is how to cope with real 3D interpretation issues, especially those involved when images exhibit strong perspective. The paper is highly successful at this, solving in a totally different way problems similar to those involved in Olsen's work. A special issue should not be so limited in its scope as to ignore the limitations of one method and the possibilities of rather different approaches. Second, the paper by Startchik et al. implements a characterisation of the patterns that involves computing a number of cross-ratio values, and cross-ratios are parameters like any other; thus, in a certain sense, the set of precise numerical values of these parameters projects out a *single* location in parameter space, whereas other HT and index-related schemes *accumulate* votes at different locations in parameter space; furthermore, some such schemes provide only a minimum number of votes at any location in parameter space [1], so the work of Startchik et al. merely takes this concept to a limit.

Finally, it should be remarked that the concept of invariance is at the very heart of all considerations of pattern recognition. As a result, invariants are of especial interest in any approach to the subject, and in the future they could well assume even greater importance than at present. At the very least, invariants show what can be ignored in any image, and are thus of supreme value for the purpose of performing the all-important preliminary sift through the data [2].

The overall picture

Looking at the sequence of papers, we see that there is a spectrum from the plain conventional HT, though GA and perceptual grouping enhancements, though structural indexing and geometric hashing adaptations, to projective methods including the Radon transform and those using the Fourier slice theorem employed in CT work, and onto the cross-ratio and invariants approach already mentioned. We can also see examples of recognition involving familiar objects such as staplers, documentation symbolism, terrain identification, optical scattering by biological samples, X-ray scattering to locate cracks and hole defects in metal, and more.

Overall, it would seem that the 1980s craze for using HTs for everything has not died a death, but is alive and healthy; what is more, it has a niche and has integrated itself in the subject area (finding a niche is perhaps opportunistic: finding how to integrate spells survival). It is not one technique, but a range of techniques, including projection, clustering, transformation, and targeted at recognition in a range of ways. The approach Hough [3] invented 36 or so years ago—for it was as an invention rather than a scientific paper that it was introduced—is still here and still worth using, although over time it has matured and become more sophisticated, and its approach is far better understood theoretically than it was even a decade ago: a perusal of the many references in Olsen's paper will quickly confirm this. It seems that the decision to put together a special issue on this topic was a correct one to take. And the negative attitude of some that the HT is too slow to use in practice has perhaps evaporated, now that the reasons for slow speed have been enunciated more clearly (*and* observed in many other approaches to recognition!), and some means at least for overcoming them have been devised.

References

[1] D.H. Ballard, D. Sabbah, Viewer independent shape recognition, IEEE Trans. Pattern Analysis and Machine Intelligence, **5** (1983) 653–660.

[2] E.R. Davies, Machine Vision: Theory, Algorithms, Practicalities, 2nd ed., Academic Press, 1997.

[3] P.V.C. Hough, Method and means for recognizing complex patterns, US Patent 3069654, 1962.