Automation and Robotics in Construction XI
D.A. Chamberlain (Editor)

Identifying Key Features in a Building Using a Single Uncalibrated Camera

Alastair M. Paterson[a*], Geoff R. Dowling[b], Denis A. Chamberlain[a]

[a]Construction Robotics Unit [b]Dept. of Computer Science

City University, Northampton Square, London, EC1V OHB, UK

ABSTRACT

This paper describes the application of image processing to the automation of external building inspection. A robot is used to locate defects which need to have their positions accurately recorded. The first stage in this process is to locate the robot on the building and a method is proposed where a single image is all that is required to find the robot. A number of edge detectors have been examined for their edge quality when applied to images of buildings. It has been found that the choice of camera is important as different types handle sharp transitions differently. The commonly employed Hough Transform, is shown to be rather unsuitable for finding lines with this type of image and an alternative, which links pixels, is used to provide improved feature location. Results are given showing how significant features are highlighted. A comparison is made between a simple model, a building with few distinct features and a complex building consisting of many windows, balconies and decorative panels.

1. INTRODUCTION

Over the last few decades many tower blocks have been constructed for both offices and living accommodation. For a variety of reasons these buildings are showing signs of decay in alarmingly high numbers. Unless repair work is carried out, these buildings will ultimately have to be declared structurally unsafe. Consequently, there is a large demand for external building inspection. This work is currently done manually, often by people abseiling down sides of the building. Apart from being a hazardous procedure (people have been known to lean out of windows and try to cut ropes), errors are often made in identifying and recording defect position, leading to ambiguous results. The first task during a building inspection is usually a visual survey primarily for finding cracks and rust stains from reinforcement bars. A robot is being developed to find these defects, but in order to locate them the robot needs to know its own position. The work presented here shows how computer vision is being developed to aid the robot in its navigation in such a way that the operator has to perform the minimum amount of setting up. Commercial considerations also place an emphasis on low cost.

This work is supported by a SERC CASE award with Laing Technology Group Ltd. as the collaborating body.

* e-mail: A.M.Paterson@uk.ac.city

2. RELATED WORK

Unlike applications in the nuclear industry, where some automated inspection has been developed, the general structural inspection industry has yet to benefit from similar technology. Specialist areas such as the nuclear industry have large funds available for developing their own inspection equipment, but for standard building surveys very little exists. Whilst a range of wall climbing robots have been developed, there has been little useful research on their use in building inspection. To some extent there is a preoccupation with robot motion rather than practical provisions such as automatic position determination. As far as visual inspection is concerned, most development is associated with crack detection. Miuri et al.[1] is a good example of finding cracks in concrete, although they do not suggest how the crack's position might be accurately determined. Fukuda et al.[2] propose a robot that can navigate with stereo vision by recognising air conditioning diffuser units in ceilings, but this is very restricted. Much robot navigation is involved with automatic vehicles that can drive along roads. Here specific features are searched for such as the white lines, kerbs and other vehicles. The Global Positioning System of satellites can be used to fix a robot's position but this, at present, is a very costly alternative to vision. In general, machine vision in inspection is operated in a designed environment. Work done by Ala et al.[3] to examine masonry units, for example, minimises the amount of background information by providing suitable lighting and sufaces.

3. METHOD

Since one of the aims of this work is a minimum cost solution, it was decided to use only a single camera on the robot for visual inspection and defect positioning, relative to building features. A further camera is located on the ground to meet the needs of global positioning. The principle is based on the fact that given a single black-and-white photograph of the building and robot taken from most angles, it is possible to mark the robot's position on a diagram of the building by counting windows, floors et cetera. For the end user this minimises the amount of setting up required as no targets are placed on the building. The camera position and parameters are not required and no sensors need be placed on the robot's winching mechanisms.

To obtain the robot's approximate location, two stages of detection are required. The first is to identify the key features on the face of the building on which the robot is placed. The second is to detect a target placed on the back of the robot. Once these are known, the center of the target (or some other reference point) can be marked on a CAD diagram of the building. This position does not need to be precise, it merely needs to indicate at which feature the robot is located. Once known, the position is transmitted to the robot and it refines its location using its own camera.

Edge detection methods are used to extract the major features, principally windows and the robot target. A CAD model is used to generate information about the features such as how they relate to one another so that these features can be compared to those found in the image. Confidence is built up as more features are identified until satisfactory knowledge about the image exists to locate the robot.

3.1. Image Capture

Images are captured by a PC using a real-time frame grabber to minimise the effects of any vibration or movement. The images used have 256 grey levels and are cropped to 512 x 512 pixels.

3.1.1. Camera

For any vision system, the behavior of the optics and sensor is vital to the ability to perform high level recognition tasks. A low cost solution would be to use a camcorder but close examination of its image reveals distortions. The step response of the camera (found from imaging a white to black transition) was found to be crucial in edge detection. In one case, it determined whether or not an edge was detected. Camcorders have their output enhanced, resulting in an overshoot on either side of the edge, as shown in figure 1. This enhancement improves the visual appearance of the image but at the pixel level it can be treated as noise, either spreading an edge or eliminating it. A suitable camera should minimise the amount of overshoot without decreasing the gradient of the edge. By using a Super VHS (S-VHS) camera, signal quality is preserved between camera and frame grabber, since the luminance and the chrominance signals do not have to be modulated and demodulated, as is the case for composite video.

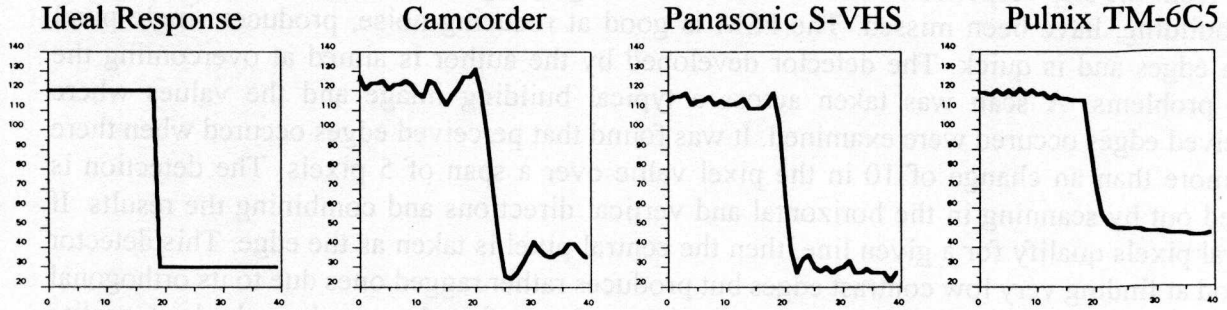


Figure 1. Camera Step Responses - Pixel Intensity Value vs. Position

3.1.2. Raw Image Enhancement

When viewing a video film we see a good image because the frames are effectively integrated over time. If a single frame is viewed we see that the image contains a fair amount of noise. For captured images then, it is necessary improve image quality before looking for edges. Edge detectors tend to be sensitive to noise and it is better to improve the raw image, than removing unwanted edges after high pass filtering. This is particularly important when using building images as there is a vast amount of desirable information, making it hard to distinguish between noise and features.

Two simple filtering operations have been studied: mean and median smoothing using a 3 x 3 pixel window. Median smoothing produces the best output as it preserves the sharpness of edges. Mean smoothing can blur features in close proximity, merging them into a single feature, but is quicker. Both of the operations could be implemented in hardware. Histogram equalisation was investigated and certainly can improve the image in some areas but it tends to highlight unwanted features such as the patterns on decorative panels. Contrast stretching can improve quality but again it can enhance noise, making further processing demands harder. Some combinations have yielded good results but lack robustness.

3.2 Edge Detection

The first step in building recognition is to determine the boundaries of building features such as window frames. This is usually done by looking for edges in the image defined by changes in grey level. There are a number of edge detectors available and an investigation of the most popular has been made to see how well they perform on the building images. For subsequent processing, it has been the edge quality that has been found to be the most important factor in evaluating the detectors, for example, whether the edges produced are continuous or broken. Speed must also be considered, but since the base camera will not be used frequently, speed is of a lower significance. The edge detectors evaluated were: Sobel[4], Roberts Cross[4], Canny[5], Finite State Machine (FSM)[6], and one developed by the author. Ideally the edge detector should produce single pixel width edges. The Sobel and Roberts Cross are unable to do this as they are gradient detectors and further processing is required to yield single pixel width edges. They are also affected by the range of constrasts across a typical building image. Many of the desired features can be missed because the edges are weak. This occurs, for example, when a window frame and concrete panel have similar grey levels. What is really required is a detector that shows edges that are *perceived* and this is where the FSM performs well. Instead of using the gradient directly, it looks at how the gradient is varying and only if it varies in a certain way is the edge said to exist. Although the FSM is fairly invariant to contrast, it can fail badly from the step response of certain cameras and highly significant edges, such as the edges of a building, have been missed. The FSM is good at reducing noise, produces single pixel width edges and is quick. The detector developed by the author is aimed at overcoming the FSM problems. A scan was taken across a typical building image and the values where perceived edges occured were examined. It was found that perceived edges occured when there was more than an change of 10 in the pixel value over a span of 5 pixels. The detection is carried out by scanning in the horizontal and vertical directions and combining the results. If several pixels qualify for a given line, then the central pixel is taken as the edge. This detector is good at finding very low contrast edges but produces rather ragged ones due to its orthogonal nature. It is however quick. The Canny edge detector was found to produce the best quality edges but is by far the slowest. However, the output does need to be thresholded to pick out the desired edges and a strategy must be found to automatically produce a suitable threshold.

3.3. Line Finding

As buildings are man-made structures, they typically possess many straight lines and it is the location of these lines that forms the basis of the recognition process. However, pixels resulting from edge detection do not possess any connectivity information and this needs to be found in order to locate lines. Two techniques have been investigated: the Hough Transform and pixel linking.

3.3.1. Hough Transform

A popular method of finding lines within a binary image is the Hough Transform (HT)[7]. Since buildings are essentially made up of straight lines it seemed a logical choice to use this. Unfortunately, for a number of reasons, it was found that the HT did not perform as well as expected for a number of reasons. The standard line parameterisation using ρ and θ results in a number of intersecting sine curves in Hough space. The large number of pixels in the source data caused interference of the peaks representing the lines. Consequently some lines were lost and other lines appeared where they should not be. This was improved by using an anti-aliasing filter[8]. In a typical building image, there may often be lines close together, and these lead to

corresponding peaks in the Hough space that can be indistinguishable. A peak filter[9] helped but did not entirely solve the problem. The HT has no sense of connectivity so whilst it wins on finding broken lines, it loses for the same reason if there are many separate pixels which by chance happen to lie on a straight line. Lines will often be placed where there are no perceived lines in the image. Similarly, a group of windows on a given floor, although seen as separate distinct objects, will have all their colinear horizontal features such as the sills, connected by the HT. Extra processing is needed to be done to break a given line into its component parts. The HT works much better with a few well seperated lines such as those found in masonry unit inspection[3]. The final and major area for failure is caused by the optics and the physics of perspective. To view an entire building usually requires a wide angle view and the image is consequently distorted. Lines passing through a vanishing point are in fact curved[10] and are only perceived as being straight due to the spherical nature of the eye. The result in either case is that long lines, typically the edges of a building, are slightly curved in an image. Curved lines will be missed by the HT due to peak spreading, so the HT can loose some of the most highly significant lines in the entire image or place them in the wrong place. Improvements to the HT to reduce these problems are suggested in section 5 below.

3.3.2. Pixel linking

An alternative way of finding lines is to find groups of connected pixels. The approach developed by West and Rosin[11] has been used to link pixels into open and closed groups. The open group contains lines, some of which may be colinear and thus joined to form long lines. This performs better than the HT as all pixels in an approximate line will be grouped even if the line is disorted. Also, lines two pixels apart can be resolved. The closed group contains sets of pixels that form closed loops, and are very likely to result from the features we are looking for, originating from circles, rectangles et cetera. At this point we are beginning to extract meaningful information from the image.

3.4. Shape detection

A view of a building will always have vanishing points and as a consequence any rectangles or circles on the building will not appear as such. It is therefore of little use to look for orthogonal or circular features. However, with a little extra processing of the open groups in pixel linking and combining the closed groups we can determine whether the group contains any vertices. By measuring the distance from the centre of the group to the group's pixels a signature[4] for the shape is generated. Vertices will show up as peaks. The vertices can be joined to create a description of the shape. If the edges of the shape comply with the principal vanishing points in the image, we can be confident that the shape is a real feature and is kept for the recognition phase. Similarly, any lines that are found to extend a significant distance across the image are also kept.

3.5. Building Recognition

How is it that we as humans can recognise an object quickly from many different angles even if it is considerably disorted? We can recognise a building from a very oblique view even if it is distorted or partially obscured. Many recognition processes require calibrated equipment and precise location so that the actual coordinates of key image points can be found. Photogrammetry is an example and would yield 3D information about a window which can be compared to the design dimensions. What is very important in human vision and recognition are the relationships between objects. The concepts of "next to", "above" and "contained in" are far

more relevant than actual measurements. After all, we do not have graduated scales in our eyes to give accurate position. Biederman et al.[12] suggest this is the way the brain works and have introduced "Geons" as the basis of recognition. For building recognition, a window can be made up of several sections. An approximate description of their relative positions and aspect ratios will remain virtually unchanged even if the image perspective is changed and the camera optics altered. The relationships from the image are compared with a list generated from the CAD model. The more features that match, the higher the confidence that the building has been correctly recognised. Once features are identified, their actual dimensions are used to provide a mapping from the image to CAD model so that the true coordinates of any image pixel can be found.

4. RESULTS

The software has been developed as a Windows 3.1 application to enable data to be transfered to other programs which are used for different robot operations. The system used is a 33Mhz 486DX machine with 8MBytes of RAM. To test ideas and to obtain a basic framework for the application, a model building has been constructed together with a robot to provide simple uncluttered images. Figure 2 shows a perspective view of the model and figure 3 shows an actual building similar to this. Figure 4 is of a large tower block with many complex features. All the *raw* images have been enhanced for printing but the Canny output is from the original. For the model image, most of the significant features have been highlighted by pixel linking. The real building images, however, show that there is considerable amount of unwanted information, particularly from windows where lights, reflections and curtains provide strong edges. Also the desired edges are frequently broken by shadows, because the face of the building, unlike the face of the model, is not flat.

5. CONCLUSIONS

The HT suffers from a number of problems when applied to a real image of a building. The area of investigation here is to reduce the number of lines and the amount of data in the parameter space. This can be approached by splitting the image up into smaller images and performing the HT on each of them[13]. This reduces the effects of image distortion. Reducing the θ range through which the values of ρ are calculated reduces the amount of data in the parameter space and speeds up execution. This can be justified since the majority of lines are in the horizontal and vertical directions, but allowances for perspective have to be made. Alternatively, the output of a gradient edge detector such as Sobel can be used to give a single value for θ.

A data list containing all the relevent features and their characeristics needs to be generated from a CAD program such as AutoCAD. The format will be the same for all applications and would normally be supplied by the surveyor. A line drawing of one face of the building is all that is required for closed features to be extracted. These are then stored as the template, to which the image features are compared.

Current work has shown that it is possible to extract some of the basic features necessary for recognising a building from a single image, indicating that this method has considerable potential for precise location and measurement. The most crucial part of this work is the extraction of valid edges, therefore, the camera and image capture system must be carefully chosen to produce steep edges with no overshoot. Further work is required to produce Canny

quality edges quickly and to reconcile process derived features with those of the CAD model. The CAD model also provides information to help filter out the unwanted information. A strategy needs to be found for automatically processing the raw images which produces consistent results during edge detection.

REFERENCES

1. M. Miura, T. Miyajima, S. Miura, K. Ogassawara, An Automated Measuring System of Cracks in Concrete which uses Image Processing Techniques and Artificial Intelligence Theory, ISARC 8th International Symposium on Automation and Robotics in Construction, Stuttgart, Germany, Vol. 2, 3-5 June 1991, pp 631-638
2. T. Fukuda et al., Development of Air Conditioning Equipment Inspection Robot with Vision Based Navigation System, 11th International Symposium on Automation and Robotics in Construction, Brighton, England, 24-26 May 1994
3. S.R. Ala, D.A. Chamberlain, T.J. Ellis, Real Time Inspection of Masonry Units, IEE 4th International Conference on Image Processing and its Applications, Maastricht, The Netherlands, 7-9 Apr. 1992 , pp 181-184
4. R.C. Gonzalez and P. Wintz, Digital Image Processing, 2nd edition, Addison-Wesley, ISBN 0-201-11026-1, 1987
5. J. Canny, A Computational Approach to Edge Detection, IEEE Transaction on PAMI, Vol 8, No 6, 1986, pp 679-698
6. A. Ginige, A Unified Approach to Image Feature Detection Using Finite State Machines, 4th International Conference on Image Processing and its Applications, Maastricht, The Netherlands, 7-9 Apr 1992, pp 486-489
7. R.O. Duda, P.E. Hart, Use of the Hough Transform to Detect Lines and Curves in Pictures, Communications of the ACM, Graphics and Image Processing Vol. 15, Jan 1972, No. 1, pp11-15
8. N. Kiryati, A.M. Bruckstein, Antialiasing the Hough Transform,The 6th Scandinavian Conference on Image Analysis, Oulu, Finland, June 1989, pp621-628
9. V.F. Leavers, Shape Detection in Computer Vision Using the Hough Transform, Springer-Verlag, ISBN 3-540-19723-0, 1992, Ch 4, pp70-75
10. B. Ernst, The Magic Mirror of M.C.Escher, Tarquin Publications, ISBN 0-906-212-45-6, 1985, pp49-51
11. P.L. Rosin, G.A.W. West, Segmentation of Edges into Lines and Arcs, Image & Vision Computing, Vol. 7, 1989, pp109-114
12. I. Biederman, E.E. Cooper, J.E. Hummel, and J.Fiser, Geon Theory as an Account of Shape Recognition in Mind, Brain, and Machine, Proceedings of the 4th British Machine Vision Conference, Guildford, England, Vol. 1, pp.175-186, September 1993
13. M. Mirmhedi, Transputer Configurations for Computer Vision, PhD Thesis, City University, London, England, 1991, Ch4, pp78-81

Figure 2. Model Building

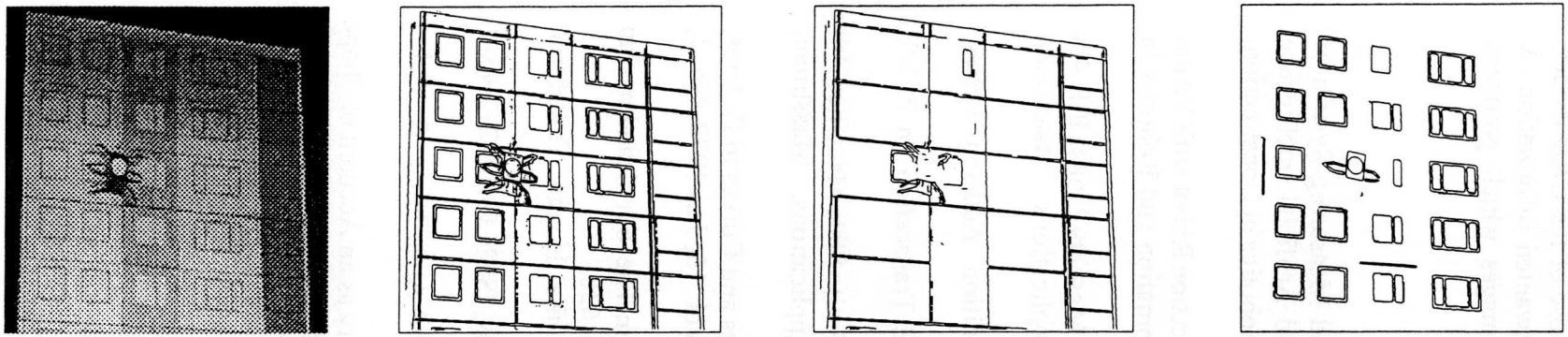


Figure 3. Tower Block in Bratislava

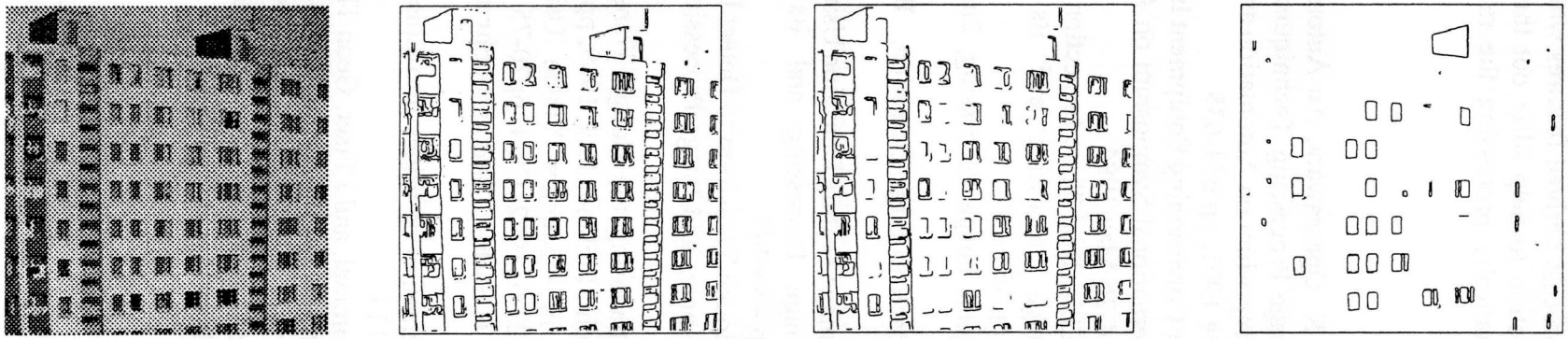
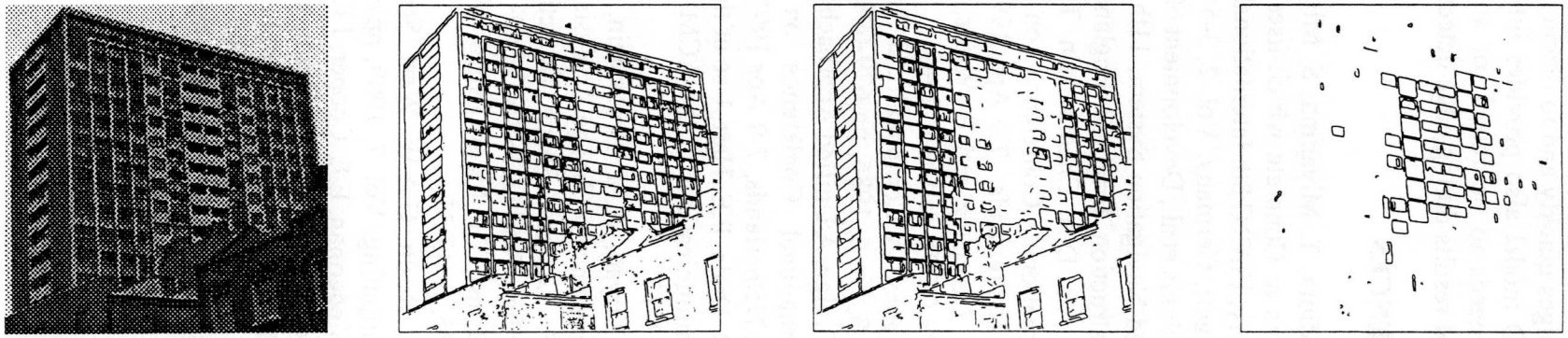


Figure 4. Tower Block in London



a. Image b. Canny Edge Detection c. Linked Open Edges d. Linked Closed Edges