CHAPTER

Corner, interest point, and invariant feature detection

6

Corner detection is valuable for locating complex objects and for tracking them in 2-D or 3-D. This chapter discusses this detection problem and considers which methods are best suited for the task.

Look out for:

- the ways in which corner features are useful
- the variety of methods available for corner detection—template matching, the second-order derivative method, the median-based method, the Harris interest point detector
- where the corner signal is a maximum: how detector bias arises
- how corner orientation may be estimated
- why invariant feature detectors are needed: the hierarchy of relevant types of invariance
- how feature detectors may be made invariant to similarity and affine transformations—SIFT, SURF, MSER, etc.
- the need for invariant detectors to embody multiparameter descriptors to help with subsequent matching tasks
- what criteria can be developed for measuring the performance of conventional and invariant types of feature and feature detector
- the histograms of oriented gradients (HOG) approach to feature detection.

Note the variety of methods available for performing interrelated detection tasks. However, different methods have different speeds, accuracies, sensitivities, and degrees of robustness: this chapter aims to bring out all these aspects of the problem.

6.1 INTRODUCTION

This chapter is concerned with the efficient detection of corners. It has been noted in previous chapters that objects are generally located most efficiently from their features. Prominent features include straight lines, circles, arcs, holes, and corners. Corners are particularly important since they may be used to locate and orientate objects and to provide measures of their dimensions; for example, knowledge about

Computer Vision. DOI: http://dx.doi.org/10.1016/B978-0-12-809284-2.00006-X

orientation will be vital if a robot is to find the best way of picking up an object, while dimensional measurement will be necessary in most inspection applications. Hence efficient, accurate corner detectors are of great relevance in machine vision.

We start this chapter by considering what is perhaps the most obvious detection scheme—that of template matching. Then we move on to other types of detectors, based on the second-order derivatives of the local intensity function; subsequently, we find that median filters can lead to useful corner detectors, with properties similar to those of the second-order derivative−based detectors. Next, we consider detectors based on the second moments of the *first* derivatives of the local intensity function. While this will complete the traditional approach to corner detection, it opens the door for consideration of the highly important invariant local feature detectors that have in the past decade or so been developed for matching widely separated views of 3-D scenes, including those containing rapidly moving objects. By the end of the chapter the vital task of considering performance criteria for the various types of corner and feature detector will also be undertaken.

6.2 TEMPLATE MATCHING

Following our experience with template matching methods for edge detection (Chapter 5: Edge Detection), it would appear to be straightforward to devise suitable templates for corner detection. These would have the general appearance of corners, and in a 3 × 3 neighborhood would take forms such as the following:

$$\begin{bmatrix} -4 & 5 & 5 \\ -4 & 5 & 5 \\ -4 & -4 & -4 \end{bmatrix} \quad \begin{bmatrix} 5 & 5 & 5 \\ -4 & 5 & -4 \\ -4 & -4 & -4 \end{bmatrix}.$$

The complete set of eight templates being generated by successive 90° rotations of the first two shown. An alternative set of templates was suggested by Bretschi (1981). As for edge detection templates, the mask coefficients are made to sum to zero so that corner detection is insensitive to absolute changes in light intensity. Ideally, this set of templates should be able to locate all corners and to estimate their orientation to within 22.5°.

Unfortunately, corners vary very much in a number of their characteristics, including in particular their degree of pointedness, internal angle, and the intensity gradient at the boundary. [The term “pointedness” is used as the opposite to “bluntness”, the term “sharpness” being reserved for the total angle η through which the boundary turns in the corner region, i.e., π minus the internal angle.] Hence it is quite difficult to design optimal corner detectors. In addition, corners are generally insufficiently pointed for good results to be obtained with the 3 × 3 template masks shown above. Another problem is that in larger neighborhoods, not only do the masks become larger but also more of them are needed to obtain optimal corner responses, and it rapidly becomes clear that the template matching approach is likely to involve excessive computation for practical corner detection. The alternative is to approach the problem analytically, somehow deducing the ideal response for a corner at any arbitrary orientation, and thereby bypassing the problem of calculating very many individual responses to find which one gives the maximum signal. The methods described in the remainder of this chapter embody this alternative philosophy.

6.3 SECOND-ORDER DERIVATIVE SCHEMES

Second-order differential operator approaches have been used widely for corner detection and mimic the first-order operators used for edge detection. Indeed, the relationship lies deeper than this. By definition, corners in grayscale images occur in regions of rapidly changing intensity levels. By this token they are detected by the same operators that detect edges in images. However, corner pixels are much rarer than edge pixels—by one definition, they arise where two relatively straight-edged fragments intersect. (We might imagine a 256 × 256 image of 64K pixels, of which 1000 (~2%) lie on edges and a mere 30 (~0.06%) are situated at corner points.) Thus it is useful to have operators that detect corners *directly*, i.e., *without unnecessarily locating edges*. To achieve this sort of discriminability it is clearly necessary to consider local variations in image intensity up to at least second order. Hence the local intensity variation is expanded as follows:

$$I(x,y) = I(0,0) + I_x x + I_y y + I_{xx}x^2/2 + I_{xy}xy + I_{yy}y^2/2 + \dots \tag{6.1}$$

where the suffices indicate partial differentiation with respect to x and y and the expansion is performed about the origin $\mathrm{X}_0(0,0)$. The symmetrical matrix of second derivatives is:

$$\mathcal{I}_{(2)} = \begin{bmatrix} I_{xx} & I_{xy} \\ I_{yx} & I_{yy} \end{bmatrix} \quad \text{where } I_{xy} = I_{yx} \tag{6.2}$$

This gives information on the local curvature at X_0. In fact, a suitable rotation of the coordinate system transforms $\mathcal{I}_{(2)}$ into diagonal form:

$$\tilde{\mathcal{I}}_{(2)} = \begin{bmatrix} I_{\tilde{x}\tilde{x}} & 0 \\ 0 & I_{\tilde{y}\tilde{y}} \end{bmatrix} = \begin{bmatrix} \kappa_1 & 0 \\ 0 & \kappa_2 \end{bmatrix} \tag{6.3}$$

where appropriate derivatives have been reinterpreted as principal curvatures at X_0.

We are particularly interested in rotationally invariant operators and it is significant that the trace and determinant of a matrix such as $\mathcal{I}_{(2)}$ are invariant under rotation. Thus we obtain the Beaudet (1978) operators:

$$\text{Laplacian} = I_{xx} + I_{yy} = \kappa_1 + \kappa_2 \tag{6.4}$$

and

$$\text{Hessian} = \det(\mathcal{I}_{(2)}) = I_{xx}I_{yy} - I_{xy}^2 = \kappa_1\kappa_2 \tag{6.5}$$

It is well known that the Laplacian operator gives significant responses along lines and edges and hence is not particularly suitable as a corner detector. On the other hand, Beaudet's "DET" operator does not respond to lines and edges but gives significant signals in the vicinity of corners: it should therefore form a useful corner detector. However, DET responds with one sign on one side of a corner and with the opposite sign on the other side of the corner: at the point of real interest—on the point of the corner—it gives a null response. Hence rather more complicated analysis is required to deduce the presence and exact position of each corner (Dreschler and Nagel, 1981; Nagel, 1983). The problem is clarified in Fig. 6.1. Here the dotted line shows the path of maximum horizontal curvature for various intensity values up the slope. The DET operator gives maximum response at positions P and Q on this line, and the parts of the line between P and Q must be explored to find the "ideal" corner point C where DET is zero.

Perhaps to avoid rather complicated procedures of this sort, Kitchen and Rosenfeld (1982) examined a variety of strategies for locating corners, starting from the consideration of local variation in the directions of edges. They found a highly effective operator which estimates the projection of the local rate of change of gradient direction vector along the horizontal edge tangent direction, and showed that it is mathematically identical to calculating the horizontal

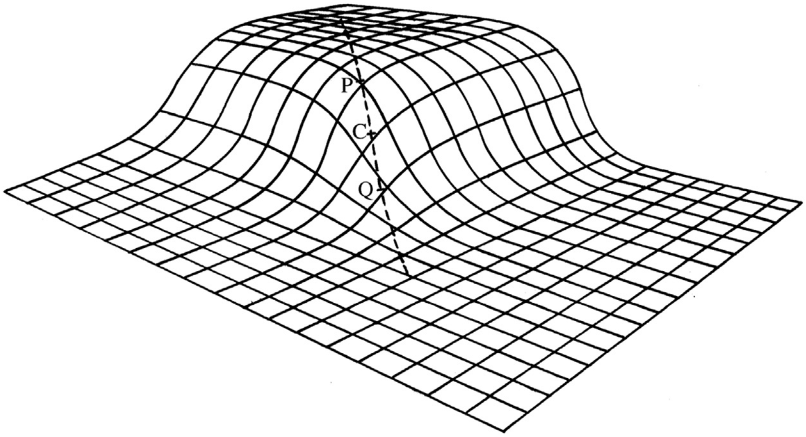


FIGURE 6.1

Sketch of an idealized corner, taken to give a smoothly varying intensity function. The dotted line shows the path of maximum horizontal curvature for various intensity values up the slope. The DET operator gives maximum responses at P and Q, and it is required to find the ideal corner position C where DET gives a null response.

curvature κ of the intensity function I. To obtain a realistic indication of the strength of a corner they multiplied κ by the magnitude of the local intensity gradient g:

$$\begin{aligned} C &= \kappa g = \kappa (I_x^2 + I_y^2)^{1/2} \\ &= \frac{I_{xx} I_y^2 - 2 I_{xy} I_x I_y + I_{yy} I_x^2}{I_x^2 + I_y^2} \end{aligned} \tag{6.6}$$

Finally, they used the heuristic of nonmaximum suppression along the edge normal direction to localize the corner positions further.

In 1983 Nagel was able to show that the Kitchen and Rosenfeld (KR) corner detector using nonmaximum suppression is mathematically virtually identical to the Dreschler and Nagel (DN) corner detector. A year later, Shah and Jain (1984) studied the Zuniga and Haralick (ZH) corner detector (1983) based on a bicubic polynomial model of the intensity function: they showed that this is essentially equivalent to the KR corner detector. However, the ZH corner detector operates rather differently in that it thresholds the intensity gradient and then works with the subset of edge points in the image, only at that stage applying the curvature function as a corner strength criterion. By making edge detection explicit in the operator, the ZH detector eliminates a number of false corners that would otherwise be induced by noise.

The inherent near-equivalence of these three corner detectors need not be overly surprising, since in the end the different methods would be expected to reflect the same underlying physical phenomena (Davies, 1988d). However, it is gratifying that the ultimate result of these rather mathematical formulations is interpretable by something as easy to visualize as horizontal curvature multiplied by intensity gradient.

6.4 A MEDIAN FILTER—BASED CORNER DETECTOR

An entirely different strategy for detecting corners was developed by Paler et al. (1984). It adopts an initially surprising and rather nonmathematical approach based on the properties of the median filter. The technique involves applying a median filter to the input image, and then forming another image that is the difference between the input and the filtered images. This difference image contains a set of signals that are interpreted as local measures of corner strength.

Clearly, it seems risky to apply such a technique since its origins suggest that, far from giving a correct indication of corners, it may instead unearth all the noise in the original image and present this as a set of "corner" signals. Fortunately, analysis shows that these worries may not be too serious. First, in the absence of noise, strong signals are not expected in areas of background, nor are they expected near straight edges, since median filters do not shift or modify such edges significantly (see Chapter 3: Image Filtering and Morphology). However, if a window is moved gradually from a background region until its central pixel is just over a convex object corner, there is no change in the output of the median filter: hence there is a strong difference signal indicating a corner.

Paler et al. (1984) analyzed the operator in some depth and concluded that the signal strength obtained from it is proportional to (1) the local contrast and (2) the "sharpness" of the corner. The definition of sharpness they used was that of Wang et al. (1983), meaning the angle η through which the boundary turns. Since it is assumed here that the boundary turns through a significant angle (perhaps the whole angle η) within the filter neighborhood, the difference from the second-order intensity variation approach is a major one. Indeed, it is an implicit assumption in the latter approach that first- and second-order coefficients describe the local intensity characteristics reasonably rigorously, the intensity function being inherently continuous and differentiable. Thus the second-order methods may give unpredictable results with pointed corners where directions change within the range of a few pixels. Although there is some truth in this, it is worth looking at the similarities between the two approaches to corner detection before considering the differences. We proceed with this in the next subsection.

6.4.1 ANALYZING THE OPERATION OF THE MEDIAN DETECTOR

This subsection considers the performance of the median corner detector under conditions where the grayscale intensity varies by only a small amount within the median filter neighborhood region. This permits the performance of the corner detector to be related to low-order derivatives of the intensity variation, so that comparisons can be made with the second-order corner detectors mentioned earlier.

To proceed we assume a continuous analog image and a median filter operating in an idealized circular neighborhood. For simplicity, since we are attempting to relate signal strengths and differential coefficients, noise is ignored. Next, recall (Chapter 3: Image Filtering and Morphology) that for an intensity function that increases monotonically with distance in some arbitrary direction $\tilde{x}$ but which does not vary in the perpendicular direction $\tilde{y}$, the median within the circular window is equal to the value at the center of the neighborhood. This means that the median corner detector gives zero signal if the horizontal curvature is locally zero.

If there is a small horizontal curvature κ, the situation can be modeled by envisaging a set of constant-intensity contours of roughly circular shape and approximately equal curvature, within the circular window, which will be taken to have radius a (Fig. 6.2). Consider the contour having the median intensity value. The center of this contour does not pass through the center of the window but is displaced to one side along the negative $\tilde{x}$ axis. Furthermore, the signal obtained from the corner detector depends on this displacement. If the displacement is D, it is easy to see that the corner signal is $Dg_{\tilde{x}}$ since $g_{\tilde{x}}$ allows the intensity change over the distance D to be estimated (Fig. 6.2). The remaining problem

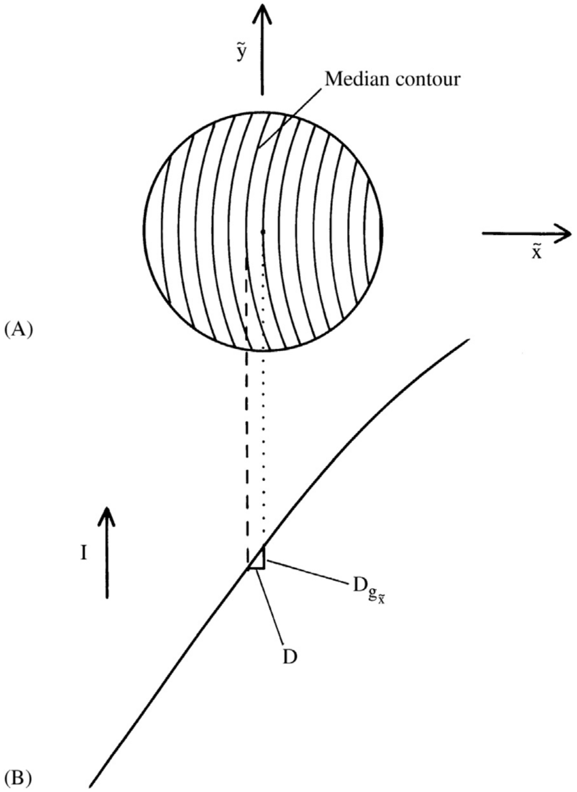


FIGURE 6.2

(A) Contours of constant intensity within a small neighborhood: ideally, these are parallel, circular, and of approximately equal curvature (the contour of median intensity does not pass through the center of the neighborhood); (B) Cross-section of intensity variation, indicating how the displacement *D* of the median contour leads to an estimate of corner strength.

© Springer 1988.

is to relate D to the horizontal curvature κ. A formula giving this relation has already been obtained in Chapter 3, Image Filtering and Morphology. The required result is:

$$D = \frac{1}{6}\kappa a^2 \tag{6.7}$$

So the corner signal is

$$C = Dg_{\tilde{x}} = \frac{1}{6}\kappa g_{\tilde{x}} a^2 \tag{6.8}$$

Note that C has the dimensions of intensity (contrast), and that the equation may be reexpressed in the form:

$$C = \frac{1}{12}(g_{\tilde{x}}a)\cdot(2a\kappa) \tag{6.9}$$

so that, as in the formulation of Paler et al. (1984), corner strength is closely related to corner contrast and corner sharpness.

To summarize, the signal from the median-based corner detector is proportional to horizontal curvature and to intensity gradient. Thus this corner detector gives an identical response to the three second-order intensity variation detectors discussed in Section 6.3, the closest practically being the KR detector. However, this comparison is valid only when second-order variations in intensity give a complete description of the situation. Clearly, the situation might be significantly different where corners are so pointed that they turn through a large proportion of their total angle within the median neighborhood. In addition, the effects of noise might be expected to be rather different in the two cases, as the median filter is particularly good at suppressing impulse noise. Meanwhile, for small horizontal curvatures, there ought to be no difference in the positions at which median and second-order derivative methods locate corners, and accuracy of localization should be identical in the two cases.

6.4.2 PRACTICAL RESULTS

Experimental tests with the median approach to corner detection have shown that it is a highly effective procedure (Paler et al., 1984; Davies, 1988d). Corners are detected reliably and signal strength is indeed roughly proportional both to local image contrast and to corner sharpness (see Fig. 6.3). Noise is more apparent for 3 × 3 implementations and this makes it better to use 5 × 5 or larger neighborhoods to give good corner discrimination. However, the fact that median

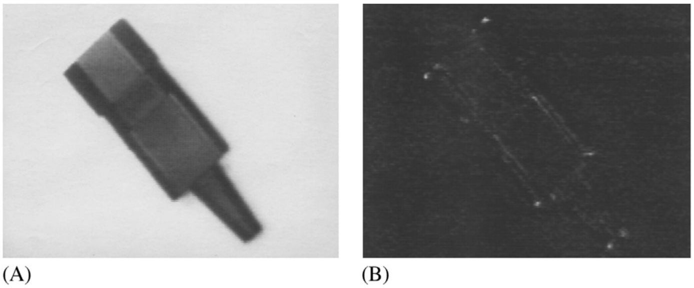


FIGURE 6.3

(A) Original off-camera 128 × 128 6-bit grayscale image; (B) result of applying the median-based corner detector in a 5 × 5 neighborhood. Note that corner signal strength is roughly proportional both to corner contrast and to corner sharpness.

operations are slow in large neighborhoods, and that background noise is still evident even in 5 × 5 neighborhoods, means that the basic median-based approach gives poor performance by comparison with the second-order methods. However, both of these disadvantages are virtually eliminated by using a "skimming" procedure, in which edge points are first located by thresholding the edge gradient, and the edge points are then examined with the median detector to locate the corner points (Davies, 1988d). With this improved method, performance is found to be generally superior to that for the KR method in that corner signals are better localized and accuracy is enhanced. Indeed, the second-order methods appear to give rather fuzzy and blurred signals that contrast with the sharp signals obtained with the improved median approach (Fig. 6.4).

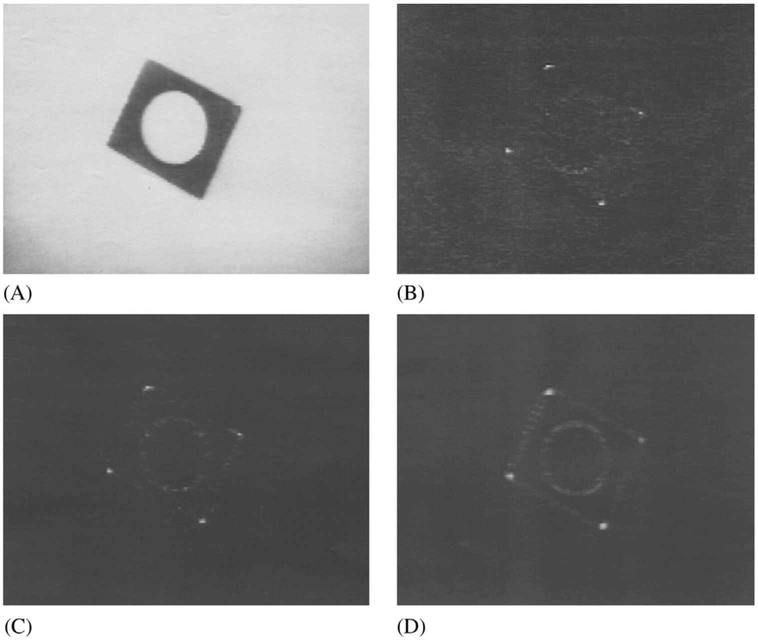


FIGURE 6.4

Comparison of the median and KR corner detectors: (A) original 128 × 128 grayscale image; (B) result of applying a median detector; (C) result of including a suitable gradient threshold; (D) result of applying a KR detector. The considerable amount of background noise is saturated out in (A) but is evident from (B). To give a fair comparison between the median and KR detectors, 5 × 5 neighborhoods are employed in each case, and nonmaximum suppression operations are not applied: the same gradient threshold is used in (C) and (D). *KR*, Kitchen and Rosenfeld.

© Springer 1988.

At this stage the reason for the more blurred corner signals obtained using the second-order operators is not clear. Basically, there is no valid rationale for applying second-order operators to pointed corners, since higher derivatives of the intensity function will become important and will at least in principle interfere with their operation. However, it is evident that the second-order methods will probably give strong corner signals when the tip of a pointed corner appears anywhere in their neighborhood, so there is likely to be a minimum blur region of radius *a* for any corner signal. This appears to explain the observed results adequately. However, note that the sharpness of signals obtained by the KR method may be improved by nonmaximum suppression (Kitchen and Rosenfeld, 1982; Nagel, 1983). Furthermore, this technique can also be applied to the output of median-based corner detectors: hence, the fact remains that the median-based method gives inherently better localized signals than the second-order methods.

Overall, the inherent deficiencies of the median-based corner detector can be overcome by incorporating a skimming procedure, and then the method becomes superior to the second-order approaches in giving better localization of corner signals. The underlying reason for the difference in localization properties appears to be that the median-based signal is ultimately sensitive only to the particular few pixels whose intensities fall near the median contour within the window, whereas the second-order operators use typical convolution masks which are in general sensitive to the intensity values of all the pixels within the window. Thus the KR operator tends to give a strong signal when the tip of a pointed corner is present anywhere in the window.

6.5 THE HARRIS INTEREST POINT OPERATOR

Earlier in this chapter, we considered the second-order derivative type of corner detector, which was designed on the basis that corners are ideal, smoothly varying differentiable intensity profiles. We also examined median filter–based detectors: these had a totally different modus operandi and were found to be suitable for processing curved step edges whose profiles were quite likely *not* to be smoothly varying and differentiable. At this point we consider what other strategies are available for corner detection. An important one that has become extremely widely used is the Harris operator. Far from being a second-order derivative type of detector, the Harris operator only takes account of first-order derivatives of the intensity function. Thus there is a question of how it can acquire enough information to detect corners. In this section we construct a model of its operation in order to throw light on this crucial question.

The Harris operator is defined very simply, in terms of the local components of intensity gradient I_x, I_y in an image. The definition requires a window region to be defined and averages $\langle\cdot\rangle$ are taken over this whole window. We start by computing the following matrix:

$$\Delta = \begin{bmatrix} \langle I_x^2\rangle & \langle I_xI_y\rangle \\ \langle I_xI_y\rangle & \langle I_y^2\rangle \end{bmatrix} \tag{6.10}$$

where the suffixes indicate partial differentiation of the intensity I; we then use the determinant and trace to estimate the corner signal:

$$C = \det\Delta/\text{trace}\,\Delta \tag{6.11}$$

While this definition involves averages, we shall find it more convenient to work with sums of quadratic products of intensity gradients:

$$\Delta = \begin{bmatrix} \Sigma I_x^2 & \Sigma I_xI_y \\ \Sigma I_xI_y & \Sigma I_y^2 \end{bmatrix} \tag{6.12}$$

To understand the operation of the detector, first consider its response for a single edge (Fig. 6.5A). In fact:

$$\det\Delta = 0 \tag{6.13}$$

because I_x is zero over the whole window region. Note that there is no loss in generality from selecting a horizontal edge, as det Δ and trace Δ are invariant under rotation of axes.

Next consider the situation in a corner region (Fig. 6.5B). Here:

$$\Delta = \begin{bmatrix} l_2g^2\sin^2\theta & l_2g^2\sin\theta\cos\theta \\ l_2g^2\sin\theta\cos\theta & l_2g^2\cos^2\theta + l_1g^2 \end{bmatrix} \tag{6.14}$$

where l_1 and l_2 are the lengths of the two edges bounding the corner, and g is the edge contrast, assumed constant over the whole window. We now find:

$$\det\Delta = l_1l_2g^4\sin^2\theta \tag{6.15}$$

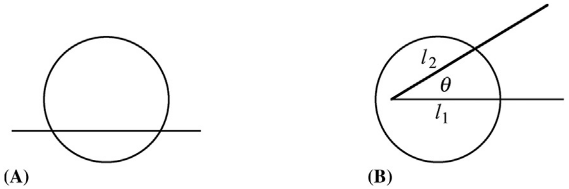


FIGURE 6.5

Case of a straight edge and a general corner. (A) A single straight edge appearing in a circular window. (B) A general corner appearing in a circular window. Circular windows are taken as ideal in that they will not favor any direction over any other.

© IET 2005.

and

$$\text{trace}\ \Delta = (l_1 + l_2)g^2 \tag{6.16}$$

so

$$C = \frac{l_1 l_2}{l_1 + l_2} g^2 \sin^2\theta \tag{6.17}$$

which may be interpreted as the product of (1) a strength factor λ, which depends on the edge lengths within the window, (2) a contrast factor g^2, and (3) a shape factor $\sin^2\theta$, which depends on the edge "sharpness" θ. Clearly, C is zero for $\theta = 0$ and $\theta = \pi$, and is a maximum for $\theta = \pi/2$, all these results being intuitively correct and appropriate.

There is a useful theorem about the sets of lengths l_1 and l_2 for which the strength factor λ, and thus C, is a maximum. Suppose we set $L = l_1 + l_2 =$ constant. Then $l_1 = L - l_2$, and substituting for l_1 we find:

$$\lambda = \frac{l_1 l_2}{l_1 + l_2} = \left[L l_2 - l_2^2\right]/L \tag{6.18}$$

$$\therefore \quad \mathrm{d}\lambda/\mathrm{d}l_2 = 1 - 2l_2/L \tag{6.19}$$

which is zero for $l_2 = L/2$, at which point $l_1 = l_2$. This means that the best way of obtaining maximum corner signal is to place the corner symmetrically within the window, following which the signal can be increased further by moving the corner so that L is maximized (Fig. 6.6).

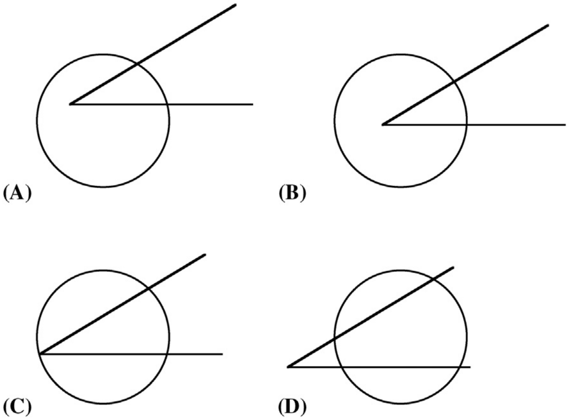


FIGURE 6.6

Possible geometries for a sharp corner being sampled by a circular window. (A) General case. (B) Symmetrical placement with $l_1 = l_2$ (see notation in Fig. 6.5B). (C) Case of maximum signal. (D) Case where the signal is reduced in size as the tip of the corner goes outside the window.

© IET 2005.

We also notice another fact—that if l_i is small (for either value of i), the corner signal at first increases linearly with l_i, and as noted earlier, the corner detector will ignore a single straight edge on its own.

Finally, the fact that we are exploring the properties of a symmetric matrix that can be represented using any convenient set of orthogonal axes means that we can find the eigenvalues and eigenvectors. However, it is illuminating to notice that these arise automatically when a symmetrically aligned set of axes is selected along the corner bisectors, as then the off-diagonal elements of the modified Δ matrix acquire two components $(L/2)g^2 \sin(\theta/2)\cos(\theta/2)$ of opposite sign and therefore cancel out. The on-diagonal elements are thus the eigenvalues themselves, and are $(L/2)g^2 \times 2\cos^2(\theta/2)$ and $(L/2)g^2 \times 2\sin^2(\theta/2)$. Again, if $\theta = 0$ or π, one or other of the eigenvalues is zero, so the determinant is zero and the corner signal vanishes; also, the maximum signal occurs for $\theta = \pi/2$.

6.5.1 CORNER SIGNALS AND SHIFTS FOR VARIOUS GEOMETRIC CONFIGURATIONS

In this section we seek the conditions for maximum corner signal for corners of different degrees of sharpness. We shall follow the observation of the previous section that maximum signal requires that $l_1 = l_2 = L/2$.

First, we take the case when $\theta = 0$: we have already seen that this leads to $C = 0$.

Next, when θ is small, i.e., less than $\pi/2$, we can go on increasing L by moving the corner symmetrically. The optimum is reached exactly as the tip of the corner reaches the far side of the window (Fig. 6.6). We could envisage the corner moving even further, but then the portions of the sides that lie within the window will be moved laterally, so they will become shorter, and the signal will fall (Fig. 6.6D).

Now take the case $\theta = \pi/2$. Then we can again proceed as above, and the optimum will still occur when the tip of the corner lies on the far side of the window (Fig. 6.7A). However, further increase of θ will result in a different optimum condition (Fig. 6.7B–D). In that case the optimum occurs for a reduced shift of the tip of the corner, the optimum occurring when the visible ends of the edges are exactly at opposite ends of a window diameter (Fig. 6.7D). Formally, we can see this in the symmetrical case ($l_1 = l_2$) from the following equation:

$$\lambda_{\text{sym}} = (L^2/4)/L = L/4 \tag{6.20}$$

so reduction of L will reduce λ_{sym} and C will fall. This situation continues until $\theta = \pi$, at which point C again falls to zero.

We can now calculate the corner shift produced by the Harris detector. Specifically, the detector places the maximum output signal at the center of the window in the cases where the signal is stated to be "optimum" above. The shift produced has a size equal to the radius a of the window for small corner angles,

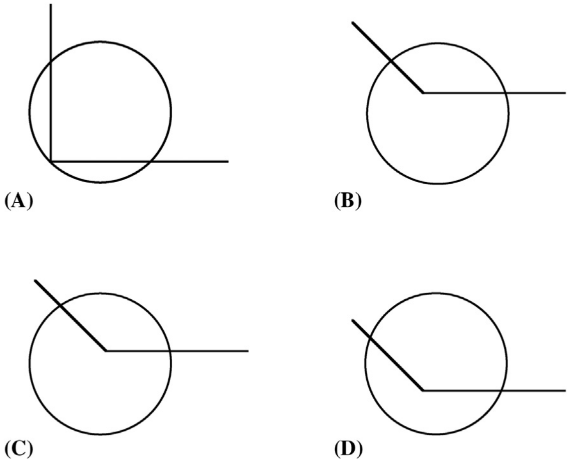


FIGURE 6.7

Possible geometries for right angle and obtuse corners. (A) Optimum case for a right-angle corner. This is the right-angle case corresponding to that shown in Fig. 6.6C. (B) General case for an obtuse corner. (C) Symmetrical placement with $l_1 = l_2$. (D) Case of maximum signal. In (D) the edges bounding the corner cross the boundary of the circular window at opposite ends of a diameter.

© IET 2005.

as then the tip of the corner is symmetrically placed on the boundary of the window. When θ rises above $\pi/2$, simple geometry (Fig. 6.8A) shows that the shift is given by:

$$\delta = a \cot(\theta/2) \tag{6.21}$$

Hence δ starts with value a at $\theta = \pi/2$, and falls to zero as $\theta \rightarrow \pi$ (Fig. 6.8B).

6.5.2 PERFORMANCE WITH CROSSING POINTS AND T-JUNCTIONS

In this section we consider the performance of the Harris operator on other types of feature, which are not normally classed as simple corners. Examples are shown in Figs. 6.9 and 6.10. It turns out that the Harris operator picks these out with much the same efficiency as for corners. We start by considering crossing points.

One of the most important points to notice is that many of the same equations apply as for corners, and in particular Eq. (6.17) still applies. However, l_1 and l_2 must now be taken as the sum of the edge lengths in each of the two main directions. Here there is an important point to note—that along the two edge directions the signs of the contrast values both reverse at the crossing point. Nevertheless, this does not alter the response, because in Eq. (6.17) the contrast g is squared.

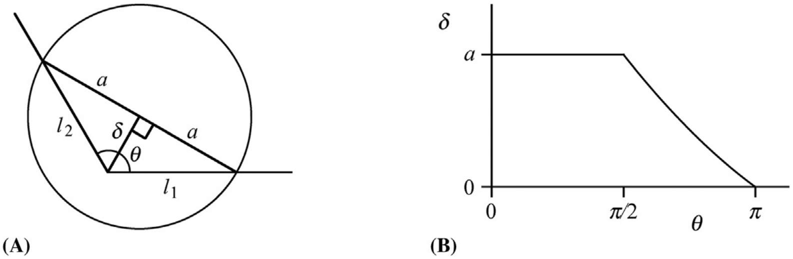


FIGURE 6.8

Geometry for calculating obtuse corner shifts and actual results. (A) Detailed geometry for calculating corner shift for the case shown in Fig. 6.7D. (B) Graph showing corner shift δ as a function of corner sharpness θ. The left of the graph corresponds to the constant shift of *a* obtained for sharp corners, while the right shows the varying results for obtuse angles.

© IET 2005.

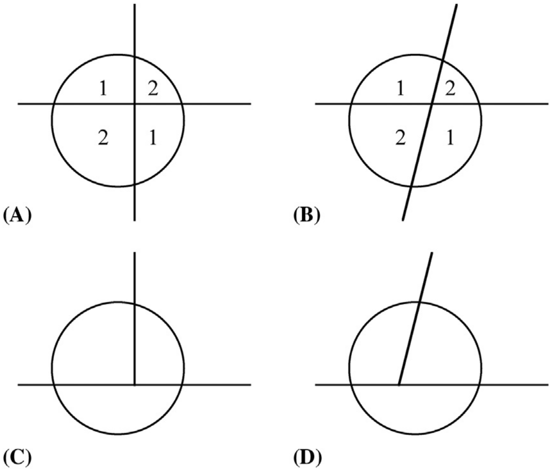


FIGURE 6.9

Other types of interest point. The types of interest point shown in this figure are those that cannot be classed as simple corner points. (A) Crossing point. (B) Oblique crossing point. (C) T-junction. (D) Oblique T-junction. In (A) and (B) the numbers indicate regions of equal, or different, intensity.

© IET 2005.

So when the window is centered at the crossing point, which is the tip of both constituent corners, the values of l_1 and l_2 are doubled.

Another relevant factor is that the corner configuration is now symmetric about the crossing point, so by symmetry this must also be the position of

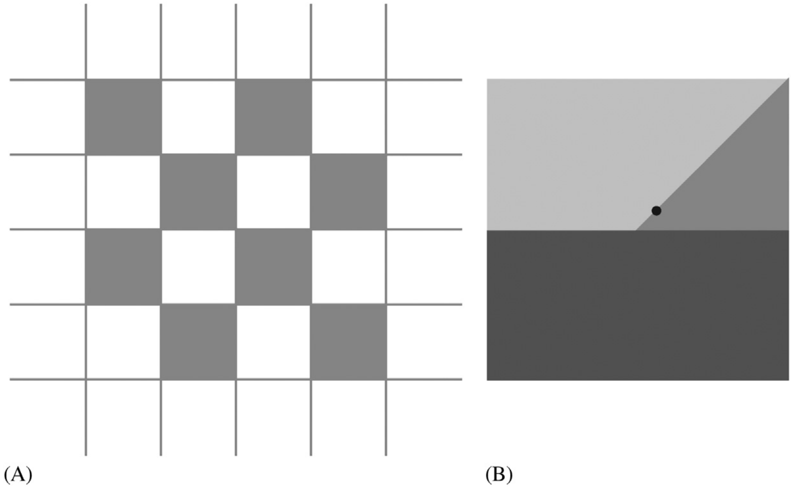


FIGURE 6.10

Effect of the Harris corner detector. (A) shows a checkerboard pattern that gives high responses at each of the edge crossover points. Because of symmetry, the location of the peaks is exactly at the crossover locations. (B) Example of a T-junction. The black dot shows a typical peak location: in this case there is no symmetry to dictate that the peak must occur exactly at the T-junction.

maximum signal. In fact, the global maximum signal must occur when there is a maximum length of both edges within the window, and they must therefore be closely aligned along window diameters. These remarks apply both for $\pi/2$ and for oblique crossovers (Figs. 6.9A,B and 6.10A).

We now consider another case that arises fairly often—the T-junction interest point. This can be either a $\pi/2$ or an oblique junction. Such cases are more general than corners and the crossing point junctions discussed above, in that they are mediated by three regions with three different intensities (Figs. 6.9C,D and 6.10B). A complete analysis of the situation for all these cases cannot be undertaken here. Instead, we consider the interesting case of a high contrast edge that is reached but not crossed by a low contrast edge. In this case, the additional intensity breaks the symmetry of the junction, so that not only does the corner peak not lie on the junction point but also there will be a small lateral movement of the peak. However, if the low contrast edge has much lower contrast than the other two, the lateral shift will be minimal. To calculate the corner signal, we first generalize Eq. (6.17) to take account of the fact that one line will have higher contrast than the other:

$$C' = \frac{l_1 l_2 g_1^2 g_2^2}{l_1 g_1^2 + l_2 g_2^2} \sin^2 \theta \tag{6.22}$$

where l_1 is taken as the straight edge with high contrast g_1 and l_2 is the straight edge with low contrast g_2. Proceeding as before we find that the optimal signal occurs where $l_1|g_1| = l_2|g_2|$. Interestingly, this can mean that the maximum signal occurs on the low contrast edge, in a highly asymmetric way (Fig. 6.10B). Part of the motivation of this study was the observation that the Harris operator peaks shown in the literature (e.g., Shen and Wang, 2002) often seem to be localized at such points, though this does not seem to have been remarked upon before 2005 when the author noted and explained the phenomenon (Davies, 2005). While apparently trivial it is actually important, as measurement bias can mislead and/or be the cause of error in subsequent algorithms. However, here the bias is known, systematic and calculable, and can be allowed for when the operator is used in practice.

Notice that the Harris operator is often called an "interest" operator, as it detects not only corners but also other interesting points such as crossovers and T-junctions, and we have seen that there is good reason why this happens. Indeed, it is difficult to imagine that a second-order derivative signal would give sizeable signals in these other cases, as the coherence of the second derivative would be largely absent, or even identically zero in the case of a crossover. Interestingly, the Harris operator often used to be called the Plessey operator, after the company at which it was originally developed.

6.5.3 DIFFERENT FORMS OF THE HARRIS OPERATOR

In this section, we consider the different forms the Harris operator can take. The form in Eq. (6.11) is due to Noble (1988) who actually gave the inverse of this expression and included a small positive constant in the denominator to prevent divide-by-zero situations. However, the original Harris operator had the rather different form:

$$C = \det \Delta - k(\text{trace } \Delta)^2 \tag{6.23}$$

where $k \approx 0.04$. Ignoring the constant, we find that the analysis presented above remains virtually unchanged, particularly concerning the optimal signal and the localization bias. The term involving k was added by Harris and Stevens (1988) in order to limit the number of false positives due to prominent edges. In principle, isolated edges should have no such effect, because as shown earlier, they lead to $\det \Delta = 0$. However, noise or clutter can affect this by introducing short extraneous edges, which interact with any existing strong edges to constitute pseudo-corners (in the absence of explanations in the literature, this seems to be the most reasonable interpretation of the situation): hence k has to be adjusted empirically to minimize the number of false positives. Searching the literature shows that in practice workers almost invariably give k a value close to 0.04 or 0.05: in fact, Rocket has investigated this, and has found that (1) making k equal to 0.04 rather than zero drastically cuts down the number of false positives due to edges; and (2) there appears to be an optimum value for k which is actually much closer to 0.05 than to 0.04, but definitely below 0.06, the k response function being a smoothly varying curve (Rocket, 2003). Nevertheless, we must expect the optimum value of k to vary with the image data.

Interestingly, in tests carried out using the Harris operator, the form given in Eq. (6.11) was used without any attempt to introduce a term in k (though divide-by-zero was taken care of), with the results shown in Fig. 6.11. Excessive numbers of false positives due to edges were not evident, though possibly this was so because of the lack of sensitivity to that effect with this particular type of data.

Finally, it should be pointed out that, when making direct theoretical comparisons between the Harris and other operators (such as the second-order derivative and median-based operators), the square roots of the expressions in Eqs. (6.11) and (6.17) will need to be taken to ensure that the result is directly proportional to edge contrast g.

6.6 CORNER ORIENTATION

This chapter has so far considered the problem of corner detection as relating merely to corner location. However, of the possible point features by which objects might be detected, corners differ from holes in that they are not isotropic, and hence are able to provide orientation information. Such information can be used by procedures that collate the information from various features in order to deduce the presence and positions of objects containing them. In Chapter 11, The Generalized Hough Transform, it will be seen that orientation information is valuable in immediately eliminating a large number of possible interpretations of an image, and hence of quickly narrowing down the search problem and saving computation.

Clearly, when corners are not particularly pointed (Fig. 6.12), or are detected within rather small neighborhoods, the accuracy of orientation will be somewhat restricted. However, orientation errors will seldom be worse than 45°, and will generally be less than 20°. (In fact, accuracy of corner location will also suffer. However, a way of overcoming this problem will be described in Chapter 11, The Generalized Hough Transform, by making use of the generalized Hough transform.) Although these accuracies are far worse than those (around 1°) for edge orientation (see Chapter 5: Edge Detection), they nevertheless provide valuable constraints on possible interpretations of an image.

Here we consider only simple means of estimating corner orientation. Basically, once a corner has been located accurately, it is a rather trivial matter to estimate its orientation from that of the intensity gradient at that location. This estimate can be made more accurate by finding the mean intensity gradient over a small region surrounding the estimated corner position, i.e., using the components $\langle I_x \rangle$ and $\langle I_y \rangle$.

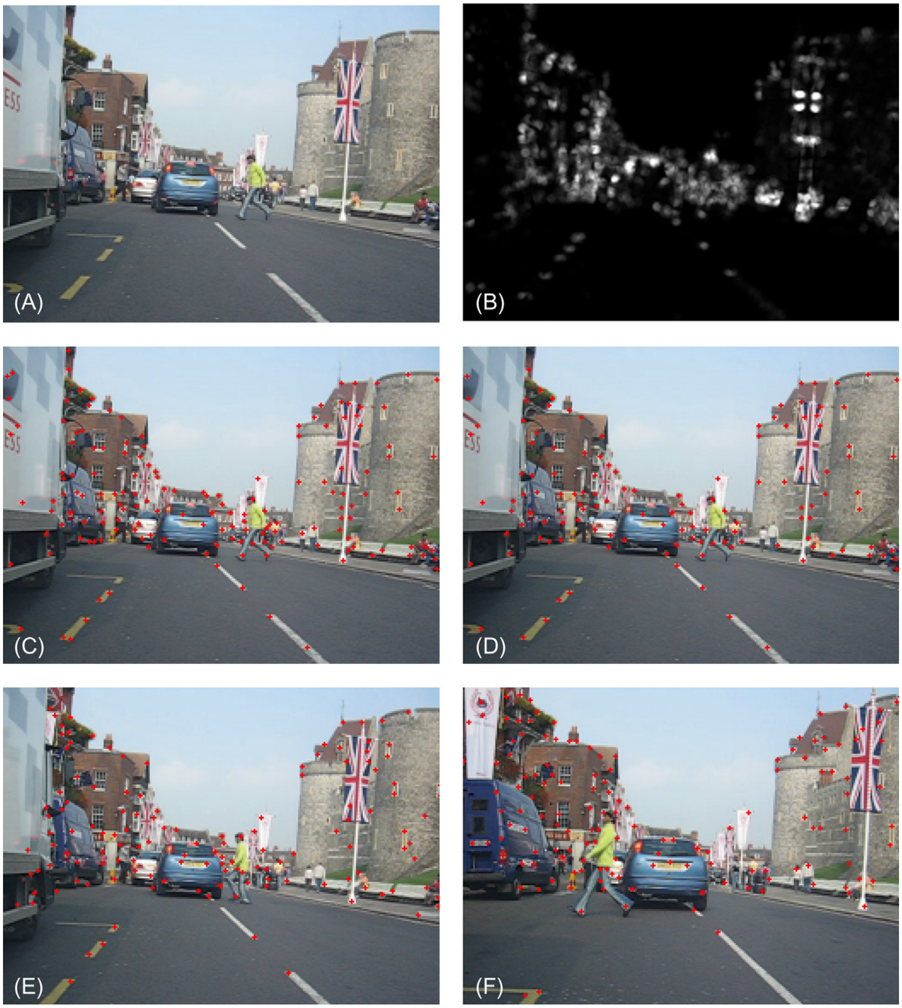


FIGURE 6.11

Application of the Harris interest point detector. (A) Original image. (B) Interest point feature strength. (C) and (D) Placement of interest points giving greatest response over a distance of 5 pixels (C) and 7 pixels (D). (E) and (F) Placements for later frames in the sequence (using maximum responses over a distance of 7 pixels). (D)–(F) show a high consistency of feature identification, which is important for tracking purposes. Notice that interest points really do indicate locations of interest—corners, people's feet, ends of white road markings, and castle window and battlement features. Also, the greater the significance as measured by the pixel suppression range, the greater relevance the feature tends to have.

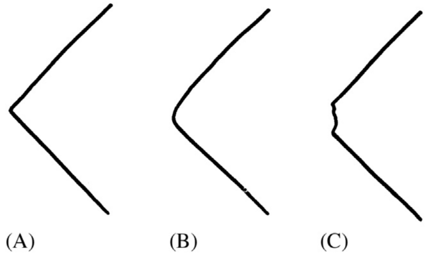


FIGURE 6.12

Types of corner: (A) pointed; (B) rounded; (C) chipped. Corners of type (A) are normal with metal components, those of type (B) are usual with biscuits and other food products, whereas those of type (C) are common with food products but rarer with metal parts.

© IEE 1988.

6.7 LOCAL INVARIANT FEATURE DETECTORS AND DESCRIPTORS

The discussion in the former sections covered corner and interest point detectors that were useful for general-purpose object location, i.e., finding objects from their features. The specifications of the detectors were that they should be sensitive, reliable, and accurate, so that there would be little chance of missing any objects containing them and that object location would be accurate. In the context of the object inference schemes described in Part 2—and particularly in Chapter 11, The Generalized Hough Transform—it does not matter if some features are missing or whether additional noise or clutter features arise, as the inferential schemes are sufficiently robust to be able to find the objects in spite of this. However, the whole context was the essentially 2-D situation where it was good enough to imagine that the objects were nearly flat, or had nearly flat faces, so that 3-D perspective types of distortion could be avoided. Even so, in 3-D, corners appear as corners from almost any viewpoint, so robust inference algorithms should still be able to perform object location. However, when viewing objects from quite different directions in 3-D, appearance can change dramatically, so it can become extremely difficult to recognize them, even if all the features are present in the images. Thus we arrive at the concept of viewing over wide baselines. In the case of binocular vision, which takes two views over quite a narrow baseline ($\sim$7 cm for human eyes), the difference between the views is necessary in order to convey depth information, but it is rarely so great that features recognized in one view cannot be re-identified in the second view (though when huge numbers of similar, e.g., textural, features occur, as when viewing a piece of material, this may not apply). On the other hand, when objects are viewed on a wide baseline, as happens after significant motion has occurred, the angular separation between the views may be as large as 50°. If even larger angular separations occur, there will be much less possibility of recognition. While this sort of situation can be tackled by memorizing sequences of views of objects, here we concentrate primarily on what can be discerned from local features that are seen in wide baseline views of up to ~50°.

At this point we have established the need to be able to recognize local features from wide baseline views as far apart as 50° so that objects can be recognized and tracked or found in databases without especial difficulty. Clearly, the corner and interest point detectors described thus far have no special provision for this. To achieve this aim, additional criteria have to be fulfilled. The first is that feature detection must be consistent and repeatable in spite of substantial change of viewpoint. The second is that features must embody descriptions of their localities so that there is high probability that the same physical feature will be positively identified in each of the views. Imagine that each image contains ~1000 corner features. Then there will be ~1 million potential feature matches between two views. While a robust inference scheme could perform the match in the case of flat objects, the situation becomes so much more demanding for general views of 3-D objects that matching might not be possible, either at all, or more likely, within a reasonable time—or without large numbers of ambiguities occurring. So it is hugely important to minimize the feature matching task. Indeed, ideally, if a rich enough descriptor is provided for each feature, feature matching might be reducible to one-to-one between views. At this stage we are extremely far from this possibility, as the corners that we have detected can so far only be characterized on the basis of their enclosed angle and intensity or color (and note that the first of these parameters will generally be substantially changed by the altered viewing angle). In what follows we consider in turn the two requirements of consistent, repeatable feature detection and feature description. But first, we have to define the various types of feature normalization that are involved.

6.7.1 GEOMETRIC TRANSFORMATIONS AND FEATURE NORMALIZATION

Broadly speaking, obtaining consistent, repeatable feature detection involves allowing for and normalizing the variations between views. The obvious candidates for normalization are scale, affine distortion and perspective distortion. Of these, the first is straightforward, the second is difficult, and the third is impractical to implement. This is because of the number of parameters that need to be estimated for each feature. Bearing in mind that local features are necessarily small, the accuracy with which the parameters can be estimated decreases rapidly with increase in their number. Fig. 6.13 shows how various transformations affect a 2-D shape. The following equations respectively define Euclidean, similarity (scale variation) and affine transformations:

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} r_{11} & r_{12} \\ r_{21} & r_{22} \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} t_1 \\ t_2 \end{bmatrix} \tag{6.24}$$

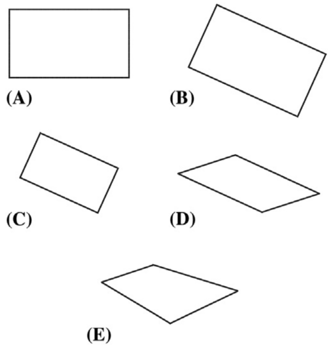


FIGURE 6.13

Effects of various transformations on a convex 2-D shape. (A) Original shape. (B) Effect of Euclidean transform (translation + rotation). (C) Effect of similarity transform (change of scale). (D) Effect of affine transform (stretch + shear). (E) Effect of perspective transform. Note that in (D) parallel lines still remain parallel: this is not in general the case after a projective transform, as indicated in (E). Overall, each of the transforms illustrated is a generalization of the previous one. The respective numbers of degrees of freedom are 3, 4, 6, and 8: in the last case each of the four points is independent and has 2 degrees of freedom, though there are constraints, such as convexity having to be maintained.

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} sr_{11} & sr_{12} \\ sr_{21} & sr_{22} \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} t_1 \\ t_2 \end{bmatrix} \tag{6.25}$$

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} t_1 \\ t_2 \end{bmatrix} \tag{6.26}$$

where rotation takes place through an angle θ, and the rotation matrix is:

$$\begin{bmatrix} r_{11} & r_{12} \\ r_{21} & r_{22} \end{bmatrix} = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \tag{6.27}$$

Euclidean transformations allow translation and rotation operations and have 3 degrees of freedom (DoF); similarity transformations include scaling operations and have 4 DoF; affine transformations include stretching and shearing operations, have 6 DoF, and are the most complex of the transformations that make parallel lines transform into parallel lines; projective transformations are much more complex, have 8 DoF, and include operations which (1) make parallel lines non-parallel, and (2) change *ratios* of lengths on straight lines. The steady increase in the number of parameters is what mitigates against estimation of perspective distortions in the feature points: in fact, it also tends to reduce accuracy for the scale parameter when estimating full affine distortion.

6.7.2 HARRIS SCALE AND AFFINE INVARIANT DETECTORS AND DESCRIPTORS

Before proceeding to consider the above ideas in more detail, note that feature detectors such as the Harris operator already estimate location and orientation, so normalization for translation and rotation is already allowed for. This leaves scale as the next candidate for normalization. Here, the basic concept is to apply a given feature detector at various scales, using larger and larger masks. In the case of the Harris operator, there are two relevant scales: one is the edge detection (differentiation) scale σ_D and the other is the overall feature (integration) scale σ_I. In practice these need to be linked together (this involves little loss of generality) so that $\sigma_I = \gamma\sigma_D$, where γ has a suitable value in the range 0–1 (typically ~0.5). σ_I then represents the scale of the overall operator. The approach is now to vary σ_I and to find the value that provides the best match of the operator to the local image data: the best match (extremum value) is the one representing the local image structure: it is intended to be independent of image resolution, which is arbitrary. In fact, the resulting "scale-adapted" Harris operator rarely attains true maxima over scales in such a ("scale-space") representation (Mikolajczyk and Schmid, 2004); this is because a corner appears as a corner over a wide range of scales (Tuytelaars and Mikolajczyk, 2008). To achieve an optimal scale for matching, a totally different approach is applied: that is to use the Harris operator to locate a suitable feature point, and then to examine its surroundings to find the ideal scale, using a Laplacian operator. The scale of the latter is then adjusted to determine, in a matched filter (i.e., optimum signal-to-noise ratio) way, when the profile of the Laplacian most accurately matches the local image structure (Fig. 6.14). The required operator is called a Laplacian of Gaussian (LoG). It corresponds to smoothing the image using a Gaussian and then applying the Laplacian $\nabla^2 = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}$ (see Chapter 5: Edge Detection). Also, as convolution

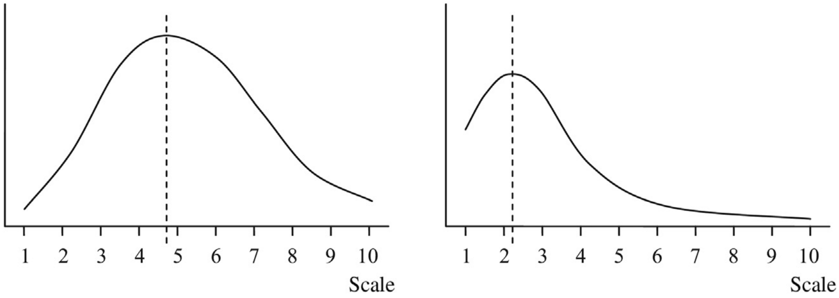


FIGURE 6.14

Scaling graphs for two objects which are being matched. The scaling graph on the left has an extremum at 4.7, and the one on the right has an extremum at 2.2. This shows that the best match occurs when the ratios of their scaling factors are approximately 2.14:1. The vertical scales of the graphs do not come into the optimization calculation.

($\otimes$) is associative, we have $\nabla^2 \otimes (G \otimes I) = (\nabla^2 \otimes G) \otimes I = LoG \otimes I$. Hence we arrive at the following combined isotropic convolution operator:

$$LoG = \frac{(r^2 - 2\sigma^2)}{\sigma^4(2\pi\sigma^2)} \exp(-r^2/2\sigma^2) = \frac{(r^2 - 2\sigma^2)}{\sigma^4} G(\sigma) \tag{6.28}$$

where

$$G(\sigma) = \frac{1}{2\pi\sigma^2} \exp(-r^2/2\sigma^2) \tag{6.29}$$

Having optimized this operator, we know the scale of the corner, and also its location and 2-D orientation. This means that when comparing two such corner features we can maintain translation, rotation, and scale invariance. To obtain affine invariance we estimate the affine shape of the corner neighborhood. Examining the Harris matrix Eq. (6.10), we rewrite it in the scale-adapted form:

$$\Delta = \sigma_{\mathrm{D}}^2 G(\sigma_{\mathrm{I}}) \otimes \begin{bmatrix} I_x^2(\sigma_{\mathrm{D}}) & I_x(\sigma_{\mathrm{D}})I_y(\sigma_{\mathrm{D}}) \\ I_x(\sigma_{\mathrm{D}})I_y(\sigma_{\mathrm{D}}) & I_y^2(\sigma_{\mathrm{D}}) \end{bmatrix} \tag{6.30}$$

where

$$I_x(\sigma_{\mathrm{D}}) = \frac{\partial}{\partial x} G(\sigma_{\mathrm{D}}) \otimes I \tag{6.31}$$

and similarly for $I_y(\sigma_{\mathrm{D}})$. These equations take full account of the differentiation and integration scales σ_{D} and σ_{I}. Then for each scale of the scale-adapted Harris operator, we repeat the process that was applied while determining the scale using the Laplacian, this time iteratively determining the best-fit ellipse (rather than circle) profile that fits the local intensity pattern. In fact, in spite of starting at separate scales, the resulting elliptic fits are, for well-defined corner structures, highly consistent and a robust average can be selected. The corresponding ellipse will (compared with the original circle) be stretched by different amounts in two perpendicular directions: the degrees of stretch and skew are the output affine parameters.

The final step is to normalize the feature by transforming it so that the elliptic profile becomes isotropic, circular, and therefore affine invariant (i.e., the affine deformation is nullified). This corresponds to equalizing the eigenvalues of the optimum scale-adapted second-order matrix, Eq. (6.30).

When comparing two corners we require invariant parameters. To obtain these descriptors, it is necessary to determine Gaussian derivatives of the local neighborhood of the interest points, computed on the transformed isotropic feature profile. Clearly, the Gaussian derivatives have to be adjusted for a standardized isotropic profile size, and they have to be normalized to intensity variations by dividing the higher order derivatives by the first-order derivative (i.e., the average intensity gradient in the neighborhood). In the work of Mikolajczyk and Schmid (2004), descriptors of dimension 12 were obtained by using derivatives up to fourth order. (There are two first-order, three second-order, four third-order, and five fourth-order derivatives: excluding the first-order derivatives, this leaves a total of 12 up to fourth order.) This set of descriptors proved highly effective for identifying corresponding pairs of features in widely different views of up to ~40° angular separation with better than 40% repeatability, and in the affine case up to ~70° with up to 40% repeatability. In addition, the localization accuracy for Harris–Laplace dropped off more or less linearly with angular separation, becoming excessive above 40°, whereas that for Harris–Affine remained at an acceptable level (~1.5 pixel error). The Harris–Laplace was described as having a breakdown point at a viewpoint change of 40°.

6.7.3 HESSIAN SCALE AND AFFINE INVARIANT DETECTORS AND DESCRIPTORS

Over the same period that scale and affine invariant detectors and descriptors based on the Harris operator were developed, investigations of similar operators based on the Hessian operator were being undertaken. Here it is useful to recall that the Harris operator is defined in terms of first derivatives of the intensity function I, while the Hessian operator (see Eq. (6.5)) is defined in terms of the second derivatives of I. Thus we can consider the Harris operator as being edge-based, and the Hessian operator as being blob-based. This matters for two reasons. One is that the two types of operator might, and do, bring in different information about objects and hence to some extent they are complementary. The other is that the Hessian is better matched than the Harris to the Laplacian scale estimator: indeed, the Hessian arises from the determinant and the Laplacian from the trace of the matrix of second-order derivatives (Eq. (6.2)). The better matching of the Hessian to the Laplacian results in improved scale selection accuracy for this operator (Mikolajczyk and Schmid, 2005). The other details of the Hessian–Laplacian and Hessian–Affine operators are similar to those for the corresponding Harris operators and will not be discussed in more detail here. However, it is worth remarking that in all four cases there are typically 200–3000 detected regions per image depending on the content (Mikolajczyk and Schmid, 2005).

6.7.4 THE SCALE INVARIANT FEATURE TRANSFORMS OPERATOR

Lowe's scale invariant feature transform (widely known as "SIFT") was first introduced in 1999, a much fuller account being given by Lowe (2004). While being restricted to a scale invariant version, it is important for two reasons: (1) for impressing on the vision community the existence, importance, and value of invariant types of detector; and (2) for demonstrating the richness that feature descriptors can bring to feature matching. For estimating scale, the SIFT operator uses the same basic principle as for the Harris and Hessian-based operators outlined above. However, it differs in using the Difference of Gaussians (DoG) instead of the LoG, in order to save computation. This possibility is seen by differentiating G with respect to σ in Eq. (6.29):

$$\frac{\partial G}{\partial \sigma} = \left(\frac{r^2}{\sigma^3} - \frac{2}{\sigma} \right) G(\sigma) = \sigma \, LoG \tag{6.32}$$

which means that we can approximate LoG as the DoG of two scales:

$$LoG \approx \frac{G(\sigma') - G(\sigma)}{\sigma(\sigma' - \sigma)} = \frac{G(k\sigma) - G(\sigma)}{(k-1)\sigma^2} \tag{6.33}$$

where use of the constant scale factor k permits scale normalization to be carried out easily between scales.

In fact, it is in the design of the descriptors that SIFT is particularly different from the Harris and Hessian-based detectors. Here the operator divides the support region, at each scale, into a 16×16 sample array and estimates the intensity gradient orientations for each of these. They are then grouped into sets of sixteen 4×4 sub-arrays and orientation histograms are generated for each of these, the directions being restricted to one of eight directions. The final output is a 4×4 array of histograms each containing entries for eight directions—amounting to a total output dimensionality of $4 \times 4 \times 8 = 128$.

The overall detector is found (Mikolajczyk, 2002) to be more repeatable than Harris–Affine and to retain a final matching accuracy above 50% out to a 50° angular separation. However, because of the limited stability of Harris–Affine, Lowe (2004) recommends the approach of Pritchard and Heidrich (2003) of including additional SIFT features with 60° viewpoint separation during training. We defer further discussion of the performance of this detector to Section 6.7.7 below.

6.7.5 THE SPEEDED-UP ROBUST FEATURES OPERATOR

The development of SIFT stimulated efforts to produce an effective invariant feature detector that was also highly efficient and required a smaller descriptor than the large one employed by SIFT. An important operator in this mold was the speeded-up robust features (SURF) method of Bay et al. (2006, 2008). This was based on the Hessian–Laplace operator. In order to increase speed, several measures were taken: (1) the integral image approach was used to perform rapid computation of the Hessian and was also used during scale space analysis; (2) the DoG was used in place of the LoG for assessing scale; (3) sums of Haar wavelets were used in place of gradient histograms, resulting in a descriptor dimensionality of 64—half that of SIFT; (4) the sign of the Laplacian was used at the matching stage; (5) various reduced forms of the operator were used to adapt it to different situations, notably an "upright" version capable of recognizing features within ± 15° of those pertaining to an upright stance, as occurs for outdoor buildings and other objects. By maintaining a rigorous, robust design, the operator was described as outperforming SIFT, and also proved capable of estimating

3-D object orientation within fractions of a degree and certainly more accurately than SIFT, Harris–Laplace, and Hessian–Laplace.

Of some importance for this implementation is the integral image approach (Simard et al., 1999), brought prominently to light by Viola and Jones (2001), but maybe not utilized as much as one might expect during the 2000s. This is extremely simple, yet radical in the levels of speedup it can bring. It involves computing an integral image I_Σ, which is an image that retains sums of all pixel intensities encountered in a single scan over the input image:

$$I_\Sigma(x,y) = \sum_{i=0}^{i \le x} \sum_{j=0}^{j \le y} I(i,j) \tag{6.34}$$

This not only permits any pixel intensity in the original image to be recovered:

$$I(i,\ j) = I_\Sigma(i,\ j) - I_\Sigma(i-1,\ j) - I_\Sigma(i,\ j-1) + I_\Sigma(i-1,\ j-1) \tag{6.35}$$

but also allows the sum of the pixel intensities in any upright rectangular block, such as those ranging from $x = i$ to $i + a$ and $y = j$ to $j + b$ within block D in Fig. 6.15, to be utilized:

$$\begin{aligned} \sum_{\mathrm{D}} I &= \sum_{\mathrm{A}} I - \sum_{\mathrm{A,B}} I - \sum_{\mathrm{A,C}} I + \sum_{\mathrm{A,B,C,D}} I \\ &= I_\Sigma(i,\ j) - I_\Sigma(i+a,\ j) - I_\Sigma(i,\ j+b) + I_\Sigma(i+a,\ j+b) \end{aligned} \tag{6.36}$$

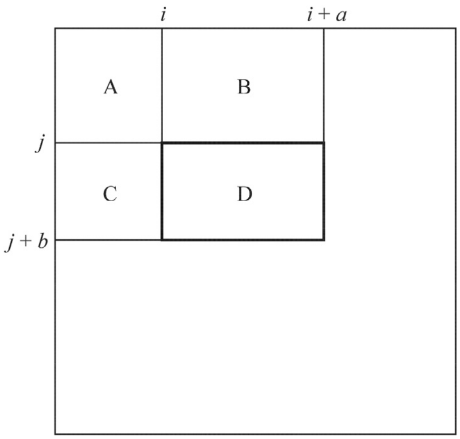


FIGURE 6.15

The integral image concept. Here block D can be considered as made up by taking block A + B + C + D, then subtracting block A + B and block C, in the latter case by subtracting A + C and adding A: see text for an exact mathematical treatment.

The method is exceptionally well adapted to computing Haar filters that typically consist of arrays containing blocks of identical values, for example:

$$\begin{bmatrix} -1 & -1 & 1 & 1 & 1 & 1 & -1 & -1 \\ -1 & -1 & 1 & 1 & 1 & 1 & -1 & -1 \\ -1 & -1 & 1 & 1 & 1 & 1 & -1 & -1 \\ -1 & -1 & 1 & 1 & 1 & 1 & -1 & -1 \end{bmatrix}$$

Note that once the integral image has been computed, it permits summations to be made over any block merely by performing four additions—taking a miniscule time that is independent of the size of the block. A simple generalization to the 3-D box filter (Simard et al., 1999) is also possible, and this is used in computing within scale-space in the SURF implementation.

6.7.6 MAXIMALLY STABLE EXTREMAL REGIONS

There is one class of invariant feature that does not fall into the pattern covered in the preceding sections—the invariant region type of feature. Among the most important examples of this type of feature is the maximally stable extremal region (MSER). The method analyses regions with increasing ranges of intensity and aims to determine those that are extremal in a particularly stable way. (Recall that finding extrema is a powerful general method for locating invariant features, as we have already seen in Section 6.7.2.)

The method (Matas et al., 2002) starts by taking pixels of zero intensity and progressively adding pixels with higher intensity levels, at each stage monitoring the regions that form. At each stage largest connected regions or "connected components" will represent extremal regions. (See Section 8.3 for a full explanation of connected components and their computation. Meanwhile, assume that a "connected component" means a region *containing everything that is connected to any part of that region*. Hence by definition, a connected component is an extremal region.) As more and more gray levels are added, the connected component regions will grow and some initially separate ones will merge. MSERs are those connected components that are close to stable (as conveniently measured by their area) over a range of intensities, i.e., each MSER is represented by the position of a local intensity minimum in the rate of change of the area function. Interestingly, relative area change is an affine-invariant property, so finding MSER regions guarantees both scale and affine invariance. In fact, because of the way that intensities are handled in this method, the results are also independent of monotonic transformation of image intensities.

While every MSER can be regarded as a connected component of a thresholded image, global thresholding is not carried out, and optimality is judged on the basis of the stability of the connected components that are located. Intrinsically, MSERs have arbitrary shapes, though for matching purposes they can be converted into ellipses of appropriate areas, orientations, and moments. Perhaps surprisingly for affine features, they can be computed highly efficiently in times that are nearly linear in the number of pixels; they also have good repeatability, though they are quite sensitive to image blur. This last problem is to be expected, given the dependence on individual gray levels and the precision with which connected components analysis is carried out; in fact, this difficulty has been addressed in a recent extension of the work (Perdoch et al., 2007).

6.7.7 COMPARISON OF THE VARIOUS INVARIANT FEATURE DETECTORS

While there are many more invariant feature detectors than we have been able to cover here (Salient regions, IBR, FAST, SFOP, . . .) and many variants of them (GLOH, PCA-SIFT, . . .), we next concentrate on comparisons between them. In fact, most of the papers describing new detectors make comparisons with older detectors, but often with limited data sets. Here we outline the conclusions of Ehsan et al. (2010), who compared SIFT, SURF, Harris–Laplace, Harris–Affine, Hessian–Laplace, and Hessian–Affine, using the following data sets: Bark, Bikes, Boat, Graffiti, Leuven, Trees, UBC, and Wall (viz. eight sequences of six images)—see Oxford Data Sets at http://www.robots.ox.ac.uk/~vgg/research/affine/ (accessed April 19, 2011). Table 6.1 presents the results in a modified form with SURF placed after Hessian–Laplace, as it is based on the latter.

Apart from Table 6.1, Ehsan at al. (2010) showed the results of using three different criteria for judging repeatability of feature detector performance. The first was the standard repeatability criterion:

$$C_0 = N_{\text{rep}}/\min(N_1, N_2) \tag{6.37}$$

where N_1 is the total number of points detected in the first image, N_2 is the total number of points detected in the second image, and N_{rep} is the number of repeated points.

Table 6.1 Comparison of Invariant Feature Detectors

Data sets	SIFT	Harris–Laplace	Hessian–Laplace	SURF	Harris–Affine	Hessian–Affine	Total
Bark	■■■	■	■	■■	■	■	9
Bikes	■	■■	■■■	■■■	■■	■■■	14
Boat	■■	■■	■■■	■■	■	■■	12
Graffiti	■	■	■	■	■■■	■■■	10
Leuven	■■■	■	■■	■■■	■	■■	12
Trees	■	■■	■■■	■■■	■■	■■	13
UBC	■■	■■■	■■■	■■■	■■■	■■■	17
Wall	■■■	■■	■■	■■■	■■	■■	14
total	16	14	18	20	15	18	101

The totals give some indication of the overall capabilities of the detectors, and of the complexity of the individual data sets. However, the detector totals must be interpreted in the light of the highest level of invariance achievable—viz. scale or affine.

They emphasized that it has been remarked (Tuytelaars and Mikolajczyk, 2008) that repeatability "does not guarantee high performance in a given application." They reasoned that this was due in part to comparing features within adjacent pairs of images rather than over whole image sequences: specifically, they recommended that each image should be compared taking the first frame of the sequence as a reference and using the following criterion:

$$C_1 = N_{\text{rep}}/N_{\text{ref}} \tag{6.38}$$

Nevertheless, they also proposed a more symmetric measure of repeatability:

$$C_2 = N_{\text{rep}}/(N_{\text{ref}} + N_{\text{c}}) \tag{6.39}$$

where N_{c} is the total number of points detected in the current frame. This proved to be a less harsh and more realistic criterion when compared with the trends of the observed ground truth for an image sequence. Further evidence for moving away from the standard repeatability criterion C_0 is that it rewards failure to detect features (because decrease in N_1 or N_2 will, if anything, *raise* the value of C_0). This suggests altering C_0 to use the maximum instead of the minimum. However, using either the maximum or the minimum tends to emphasize extreme results, leading to nonrobust measures. From this point of view the most appropriate measure has to be C_2. In fact, this criterion gave optimal results and low error probability measures when run against ground truth using Pearson's correlation coefficients (Ehsan et al., 2010). Using C_2, an important result was the dominance of the Hessian-based detectors, which is already very evident in Table 6.1 (derived from their table 2), where the three Hessian-based totals are 18, 18, 20, vis-à-vis 14, 15, and 16 for the others (interestingly, the situation is even more polarized in favor of Hessian-based detectors when the data sets giving the best and worst results—Bark and UBC—are ignored). Note that Tuytelaars and Mikolajczyk (2008) did not come out so strongly in favor of the Hessian-based detectors, but this was probably because their analysis of data sets was not so extensive; also, Ehsan et al. (2010) were the first to look at image sequences so rigorously using C_2.

The review by Tuytelaars and Mikolajczyk (2008) is of great value in evaluating performance using several disparate criteria, viz. repeatability, localization accuracy, robustness, and efficiency. Some of their results are shown in Table 6.2—notably those for all the feature detectors covered in Table 6.1 and those for the single-scale Harris and Hessian, and for the MSER detector (Matas et al., 2002) mentioned earlier. They also make the following valuable observations:

1. Scale invariant operators can normally be dealt with adequately by a robustness capability for viewpoint changes of less than 30°, as affine deformations only rise above those due to variations in object appearance beyond that level.
2. In different applications, different feature properties may be important, and thus success depends largely on appropriate selection of features.

Table 6.2 Performance Evaluation of Various Feature Detectors

Detector	Invariance	Repeat-ability	Accuracy	Robust-ness	Efficiency	Total
Harris	Rotation	■■■	■■■	■■■	■■	11
Hessian	Rotation	■■	■■	■■	■	7
SIFT	Scale	■■	■■	■■	■■	8
Harris–Laplace	Scale	■■■	■■■	■■	■	9
Hessian–Laplace	Scale	■■■	■■■	■■■	■	10
SURF	Scale	■■	■■	■■	■■■	9
Harris–Affine	Affine	■■■	■■■	■■	■■	10
Hessian–Affine	Affine	■■■	■■■	■■■	■■	11
MSER	Affine	■■■	■■■	■■	■■■	11

The totals give some indication of the overall capabilities of the detectors: however, they must be interpreted in the light of the highest level of invariance achievable (Column 2).

3. Repeatability may not always be the most important feature performance characteristic: not only is it hard to define and measure but robustness to small appearance variations matters more.
4. There is a need for work focusing on complementarity of features, leading either to complementary detectors or to detectors providing complementary features.

In the last respect, note that some ground work has recently been carried out by Ehsan et al. (2011) to measure the coverage of interest point detectors. They identify the recent SFOP scale invariant feature transform (Förstner et al., 2009) as the most outstanding detector in this respect, either used on its own or in conjunction with others. Ehsan et al.'s new criterion for coverage C is based on the harmonic mean so as not to overemphasize nearby features:

$$C = N(N-1)/\sum_{i=1}^{N-1}\sum_{j>1}^{N}(1/d_{ij}) \tag{6.40}$$

(In this formula d_{ij} is the Euclidean distance between feature points i and j.)

Note that a detector that has high coverage is not guaranteed to be sound: after all, a detector giving a random selection of feature points might fare well on this count. Hence a coverage criterion can only come into its own when it is used to select complementary types of feature and feature detector, or a detector that provides a good mix of types of feature, for which the output selection is known to be sound on other counts (repeatability, robustness, and so on).

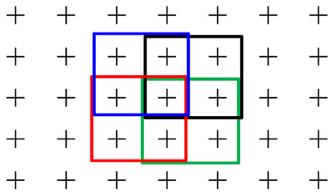
6.7.8 HISTOGRAMS OF ORIENTED GRADIENTS

This study of locally invariant feature detectors would be incomplete without including the histograms of oriented gradients (HOG) approach (Dalal and

Triggs, 2005), which appeared during much the same period as SIFT, SURF, and the other methods mentioned earlier. HOGs were designed for, and are well-matched to, the detection of human shapes. Basically, they focus on the straight limbs of the human body, which have many edge points aligned along the same direction—though the latter will naturally change with walking or other motions. The basis of the method is to divide the image into "cells" (sets of pixels) and to produce orientation histograms for all of them. Voting into the orientation histogram bins takes place with weighting proportional to gradient magnitude. The cells are combined into larger overlapping blocks, with the result that some of the blocks end up with larger signals indicating the presence of human limbs. Curiously, the HOG detectors cue mainly on silhouette contours and emphasize the head, shoulders, and feet.

Although the HOG approach and the SIFT operator have a certain resemblance in that they both use orientation histograms, they do so in different ways. In fact, the aim of SIFT is to provide orientation invariance by matching pairs of features by first canceling the difference in orientations between them (SIFT is intrinsically scale-invariant, though it is ultimately also translation and rotation invariant as its descriptor has full knowledge of the location and orientation of each feature). In contrast, the HOG approach does not provide orientation invariance but instead aims to present a cell map of orientation histograms over parts of an image, so that human or other shapes can be identified. Thus it should be described as a *regional image descriptor* rather than as a *local image descriptor*. However, it has the further advantage of providing a good measure of geometric and photometric invariance as it focuses only on local variations in orientation.

To provide illumination invariance, the histograms in the HOG operator are normalized with respect to image contrast. Dalal and Triggs achieved this by normalizing the contrast over neighboring (typically 2 × 2) blocks of cells. Interestingly, they found it better to do this several times for each cell by using different overlapping blocks of cells and treating the results as independent signals. In the case of 2 × 2 blocks of cells, this gave four results for each cell—as indicated by the four main squares in the following figure:



Overall, experiment showed that square cells of side 6–8 pixels together with block sizes of 2 × 2 or 3 × 3 cells worked best (in this context "best" meant giving the best match to the widths of human limbs in the particular image data set).

Another interesting detail was that nine histogram bins were used to cover the range 0°−180°, amounting to relatively fine quantization for orientation; in addition, no gain was found from using signed gradients when detecting humans (though the position was found to be reversed for car and motorbike recognition).

6.8 CONCLUDING REMARKS

Corner detection provides a useful start to the process of object location and to this end is often used in conjunction with the abstract pattern matching approaches discussed in Chapter 11, The Generalized Hough Transform. Apart from the obvious template matching procedure, which is of limited applicability, three main approaches have been described. The first was the second-order derivative approach, which includes the KR, DN, and ZH methods—all of which embody the same basic schema; the second was the median-based method, which turned out to be equivalent to the second-order derivative methods in situations where corners have smoothly varying intensity functions; and the third was the Harris detector which is based on the matrix of second moments of the *first* derivatives of the intensity function. Perhaps surprisingly, the latter is able to extract the same information much as the other two approaches, though there are differences, in that the Harris detector is better described as an interest point detector than as a corner detector. In fact, the Harris detector has probably been the most widely used corner and interest point detector of all, and for general purpose (non-3-D) operation this still seems to be the case—in spite of the advent of the SUSAN detector (Smith and Brady, 1997), which is known to be faster and more efficient, but somewhat less resistant to noise.

Interestingly, the situation presented above started changing radically from about 1998, when workers started looking for approaches to object location, which were not merely robust to noise, distortion, partial occlusion, and extraneous features, but were able to overcome problems of gross distortion due to viewing the same scene from widely different directions. This "wide baseline" problem, which is prominent with 3-D and motion applications, including tracking moving objects, became the driving force for radical new thinking and development. As we have seen, attempts were made to adapt the Harris operator to this scenario, making it invariant to similarity (scale) and affine transformations, though in the end somewhat more success was achieved by returning to the Hessian operator discussed earlier in this chapter. Alternative approaches included the MSER approach, which is by no means based on the location of any sort of corner or interest points. Indeed, it harks back to the thresholding methods of Chapter 4, The Role of Thresholding. But this is all to the good, as the underlying task is that of segmentation coupled with recognition and identification/matching: division of the subject into watertight topics such as thresholding, edge detection, and corner detection has limited validity or at least it is too restrictive to offer the best solutions to the real problems of the subject. In this context it is relevant that the newly evolved feature detectors embody multiparameter descriptors, making them far better suited not only to detection per se but also to the more exacting task of wide baseline 3-D matching.

Overall, in this chapter we have seen the corner detector approach transmogrify itself to overcome the problems of viewing objects from directions as far apart as 70°—and with a great deal of success. Remarkably, all this was achieved in little more than a decade—evidence that progress in this subject has been accelerating. Importantly, the old computer adage "garbage in—garbage out" is relevant, because feature detection forms a crucial link between the original pictures and their high level interpretation.

This chapter has studied how objects may be detected and located from their corners and interest points. It has developed both the classic approach to detector design and the more recent invariant approaches, which result in multiparameter feature descriptors to aid matching between widely separated views of objects.

6.9 BIBLIOGRAPHICAL AND HISTORICAL NOTES

The subject of corner detection has been developing for over three decades. The scene was set for the development of parallel corner detection algorithms by Beaudet's (1978) work on rotationally invariant image operators. This was soon followed by Dreschler and Nagel's (1981) more sophisticated second-order corner detector: the motivation for this research was to map the motion of cars in traffic scenes, corners providing the key to unambiguous interpretation of image sequences. One year later, Kitchen and Rosenfeld (1982) had completed their study of corner detectors based mainly on edge orientation and had developed the second-order KR method described in Section 6.3. The years 1983 and 1984 saw the development of the second-order ZH detector and the median-based detector (Zuniga and Haralick, 1983; Paler et al., 1984). Subsequently, the author's work on the detection of blunt corners (see Chapter 11: The Generalized Hough Transform) and on analyzing and improving the median-based detector appeared (Davies, 1988a,d, 1992a). Meanwhile, other methods had been developed, such as the Harris algorithm (Harris and Stephens, 1988; see also Noble, 1988). The Smith and Brady (1997) "SUSAN" algorithm marked a further turning point, needing no assumptions on the corner geometry, as it works by making simple comparisons of local gray levels: this is one of the most cited of all corner detection algorithms.

In the 2000s further corner detectors have been developed. Lüdtke et al. (2002) designed a detector based on a mixture model of edge orientation: in addition to being effective in comparison with the Harris and SUSAN operators, particularly at large opening angles, the method provides accurate angles and strengths for the corners. Olague and Hernández (2002) worked on a unit step edge function (USEF) concept, which is able to model complex corners well: this resulted in adaptable detectors that are able to detect corners with sub-pixel accuracy. Shen and Wang (2002) described a Hough transform−based detector: as this works in a 1-D parameter space it is fast enough for real-time operation; a useful feature of the paper is the comparison with, and between, the Wang and Brady detector, the Harris detector, and the SUSAN detector. The several example images show that it is difficult to be sure exactly what one is looking for in a corner detector (i.e., corner detection is an ill-posed problem) and that even the well-known detectors sometimes inexplicably fail to find corners in obvious places. Golightly and Jones (2003) present a practical problem in outdoor country scenery: they discuss not only the incidence of false positives and false negatives but also the probability of correct association in corner matching, e.g., during motion.

Rocket (2003) gives a performance assessment of three corner detection algorithms—the KR detector, the median-based detector, and the Harris detector: the results are complex, and the three detectors are found to have very different characteristics. The paper is valuable in showing how to optimize the three methods (not least showing that the Harris detector parameter k should be ~ 0.05), and also because it concentrates on careful research rather than "selling" a new detector. Tissainayagam and Suter (2004) give an assessment of the performance of corner detectors, with vitally important coverage of point feature (motion) tracking applications. Interestingly, it finds that, in image sequence analysis, the Harris detector is more robust to noise than the SUSAN detector, a possible explanation being that it "has a built-in smoothing function as part of its formulation." Finally, Davies (2005) analyzed the localization properties of the Harris operator: see Section 6.5 for the main results of this work.

While the above discussion covers many of the developments on corner detection, it is not the whole story. This is because in many applications it is not specific corner detectors that are needed but "interest point" detectors, which are capable of detecting any characteristic patterns of intensity that can be used as reliable feature points. In fact, the Harris detector is often called an interest point detector—with good reason, as indicated in Section 6.5.2. Moravec (1977) was among the first to refer to interest points and was followed by Schmid et al. (2000) and many others. However, Sebe and Lew (2003) and Sebe et al. (2003) call them salient points—a term more often reserved for points that attract the attention of the *human* visual system. Overall, it is probably safest to use the Haralick and Shapiro (1993) definition: a point being "interesting" if it is both distinctive and invariant—i.e., it stands out and is invariant to geometric distortions such as might result from moderate changes in scale or viewpoint (note that Haralick and Shapiro also list other desirable properties—stability, uniqueness, and interpretability). The invariance aspect is taken up by Kenney et al. (2003) who show how to remove ill-conditioned points from consideration, to make matching more reliable.

The subject of invariant feature detectors and descriptors has taken little over a decade to develop and over that period it has come a long way. It started with papers by Lindeberg (1998) and Lowe (1999) that indicated the way forward and provided basic techniques. It arose largely because of difficulties in wide baseline stereo work, and with tracking object features over many video frames—because features change their appearance over time and correspondences are easily lost. To proceed, it was necessary first to eliminate the relatively simple problem of features changing in size, thereby necessitating scale invariance (it being implicit that translation and rotation invariance have already been dealt with). Later, improvements became necessary to cope with affine invariance. Thus Lindeberg's pioneering theory (1998) was soon followed by Lowe's work (1999, 2004) on scale invariant feature transforms (SIFT). This was followed by affine invariant methods developed by Tuytelaars and Van Gool (2000), Mikolajczyk and Schmid (2002, 2004), Mikolajczyk et al. (2005), and others. In parallel with these developments, work was published on MSER (Matas et al., 2002) and other extremal methods (e.g., Kadir and Brady, 2001; Kadir et al., 2004).

Much of this work capitalized on the interest point work of Harris and Stephens (1988), and was underpinned by careful in-depth experimental investigations and comparisons (Schmid et al, 2000; Mikolajczyk and Schmid, 2005; Mikolajczyk et al., 2005). Next, the tide turned in other directions, in particular the design of feature detectors that aim at real-time operation—as in the case of the SURF approach (Bay et al., 2006, 2008).

A review article summarizing the main approaches was published by Tuytelaars and Mikolajczyk in 2008. However, that was by no means the end of the story. As outlined in Section 6.7.7, Ehsan et al. (2010) briefly reviewed the status quo on repeatability of the main features and detectors, and reported on experiments to assess it: their work included two new repeatability criteria that more realistically reflected the underlying requirements for invariant detectors. In addition, they presented new work (Ehsan et al., 2011) on the coverage of invariant feature detectors, reflecting poignant remarks made in Tuytelaars and Mikolajczyk's (2008) review. It is clear that, with the passing of the first decade of the 2000s, an even more exacting phase of development is under way, with more rigorous performance evaluation: it will no longer be sufficient to produce new invariant feature detectors; instead it will be necessary to integrate them much more fully with the target applications, following rigorous design to ensure that all relevant criteria are being met and that the tradeoffs between the criteria are much more transparent.

6.9.1 MORE RECENT DEVELOPMENTS

Since the Tuytelaars and Mikolajczyk's (2008) review, further relevant work on feature detectors and descriptors has emerged. Rosten et al. (2010) have presented the FAST family of corner detectors that is designed on a new heuristic to be especially fast while at the same time to be highly repeatable; they also review methods for comparing feature detectors and call for less concentration on how a feature detector should do its job than on what performance measure it is required to optimize. Cai et al. (2011) work on a "linear discriminant projections" procedure for reducing the dimensionality of local image descriptors and manage to bring the SIFT tally down from 128 to just 30. However, they warn that this seems to be achievable only by making the projections specific to the type of image data. With a similar motivation, Teixeira and Corte-Real (2009) quantize the SIFT descriptor to form visual words using a predefined vocabulary, though in this case the vocabulary is structured in the form of a tree; it is constructed using a generic data set related to the type of object tracking being performed. van de Sande et al. (2010) discuss the generation of color object descriptors. They find that the choice of a single-color descriptor for all categories of data is suboptimal: but for unknown data, "OpponentSIFT" (using three sets of SIFT features for the three opponent colors) showed the highest degree of invariance with respect to photometric variations. Zhou et al. (2011) proposed a method to perform descriptor combination and classifier fusion. They cast the problem of object classification into a learning setting which again means that the method is adaptive and not applicable to new data without retraining. Overall, we see that trying to reduce the original 128 SIFT features (or the equivalent) tends to make such methods specific to particular training data.

6.10 PROBLEMS

1. By examining suitable binary images of corners, show that the median corner detector gives a maximal response within the corner boundary rather than half-way down the edge outside the corner. Show how the situation is modified for grayscale images. How will this affect the value of the gradient noise-skimming threshold to be used in the improved median detector?
2. Prove Eq. (6.6), starting with the following formula for curvature:

$$\kappa = (\mathrm{d}^2y/\mathrm{d}x^2)/\left[1+(\mathrm{d}y/\mathrm{d}x)^2\right]^{3/2}$$

Hint: First express dy/dx in terms of the components of intensity gradient, remembering that the intensity gradient vector (I_x, I_y) is oriented along the edge normal; then replace the x, y variation by I_x, I_y variation in the formula for κ.