Harris and Laplacian Region Detectors

Svetlana Lazebnik

January 30, 2008

This document presents the mathematical details of the scale- and affine-invariant Harris [4] and Laplacian [1, 2] region detectors. First, scale-invariant regions are obtained by performing automatic spatial and scale selection (Section 1). Next, the affine shape of these regions is estimated through an *affine adaptation* process (Section 2).

1 Spatial and Scale Selection

This section discusses the procedure for extracting *scale-adapted* local regions, i.e., interest points equipped with characteristic scales. The Laplacian detector extracts image regions whose locations and characteristic scales are given by scale-space maxima of the Laplace operator. The Harris detector uses the same operator for scale selection, but finds the locations of interest points as the local maxima of a "cornerness" measure based on the second moment matrix.

Let us begin by summarizing Lindeberg's procedure for spatial and scale selection using the Laplace operator [2]. Let $I(x)$ denote the image intensity as a function of position. To simplify the presentation, we will treat $I(x)$ as a continuous differentiable function. We can form a basic linear scale space by considering all possible images that result from convolving $I(x)$ with an isotropic Gaussian of standard deviation σ:

$$\begin{aligned} I(x;\sigma) &= G(x;\sigma) * I(x) = \int G(x - x';\sigma)\ I(x')\ \mathrm{d}x', \\ G(x;\sigma) &= \frac{1}{2\pi\sigma^2} \exp\left(-\frac{|x|^2}{2\sigma^2}\right). \end{aligned}$$

For any fixed scale σ, we can compute any combination of spatial derivatives of $I(x;\sigma)$ (effectively, these derivatives are computed by convolving the image $I(x)$ with the appropriate combination of derivatives of the Gaussian $G(x;\sigma)$). For instance, we may want to custom-design a differential operator tuned to certain kinds of image structures, such as edges, ridges, or blobs [2]. A desirable property for a scale-space differential operator is that it should always produce the same response to an idealized scale-invariant structure like a step edge. However, if we simply take a "blurred derivative" of a step edge, we will obtain weaker responses at larger scales. This motivates the definition of *scale-normalized* differential operators, whose output remains constant if the image is scaled (resized) by an arbitrary factor. One particularly useful normalized differential quantity is the scale-normalized Laplacian $\hat{\triangle} I(x;\sigma)$, which is one of the simplest scale selection operators [2]:

$$\hat{\triangle} I(x;\sigma) = \sigma^2 \left(\frac{\partial^2 I(x;\sigma)}{\partial x^2} + \frac{\partial^2 I(x;\sigma)}{\partial y^2} \right) \tag{1}$$

$$= \sigma^2 \left(\frac{\partial^2 G(x;\sigma)}{\partial x^2} + \frac{\partial^2 G(x;\sigma)}{\partial y^2} \right) * I(x) = \hat{\triangle} G(x;\sigma) * I(x)\,.$$

As an illustration of why the scale-normalized Laplacian is appropriate for selecting the characteristic scale of local image structures, consider the one-dimensional toy example of Figure 1. The left-most picture shows a one-dimensional binary image (signal):

$$I(x) = \begin{cases} 1, & -8 \leq x \leq 8 \\ 0, & \text{otherwise.} \end{cases}$$

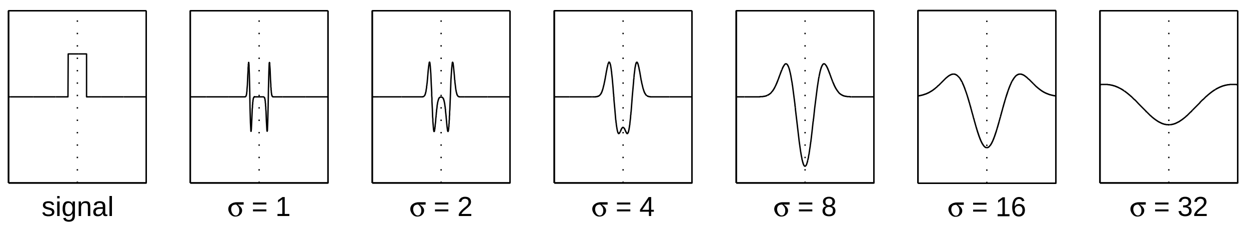

Figure 1: Scale selection: one-dimensional example (see text). The dotted vertical line is $x = 0$.

The next six pictures show the signal $I(x)$ convolved with one-dimensional scale-normalized Laplacian kernels $\hat{\triangle} G(x;\sigma)$ for different values of σ. The family of convolved signals $\hat{\triangle} G(x;\sigma) * I(x)$ has a clear global minimum at $\sigma = 8$ and $x = 0$. Thus, the scale-normalized Laplacian shows the strongest response at the characteristic scale of the signal.

Scale-selection operators like the Laplacian can also be used to find interest points in an image. We define a Laplacian *scale-space interest point* as a point where $\hat{\triangle} I(x;\sigma)$ *simultaneously* achieves a local maximum with respect to the scale parameter and the spatial variables. In the example of Figure 1, $(x, \sigma) = (0, 8)$ is a scale-space interest point. By computing such interest points, we treat scale and spatial selection as parts of the same problem.

The Harris detector is obtained by a simple modification of the Laplacian detector, where for spatial selection instead of the Laplacian we use the "cornerness" measure based on the *second moment matrix*, also known as the *local shape matrix* or the *auto-correlation matrix* [1, 4]:

$$M_I(x;\gamma,\sigma) = G(x;\sigma) * \left(\nabla I(x;\gamma)\ \nabla I(x;\gamma)^T \right)\,. \tag{2}$$

Note that $M_I(x;\gamma,\sigma)$ simultaneously lives in two scale spaces with parameters γ and σ. The *inner scale* γ is the scale at which smoothed image derivatives are computed, while the *outer scale* σ is the scale of integration — it defines the size of the window over which second moment information is accumulated. Mikolajczyk and Schmid [4] point out that the inner scale γ is less critical than the outer characteristic scale σ (though they do implement a procedure for selecting γ automatically). Intuitively, the inner scale describes the intrinsic resolution of the image measurement process — any information that is lost after blurring the image with γ is lost forever. In our implementation, we set γ to a constant value of 2.

Informally, the eigenvectors of the second moment matrix represent the dominant (orthogonal) edge orientations of the window, and the eigenvalues represent the amount of edge energy along these directions. The Harris detector looks for points like corners and junctions that have significant amounts of energy in two orthogonal directions. The Harris "cornerness" measure is defined as follows:

$$\det\left(M_I(x;\gamma,\sigma)\right) - \alpha\,\mathrm{tr}^2\left(M_I(x;\gamma,\sigma)\right)\,,$$

where α is a constant (0.05 in the implementation). Local maxima of this measure determine the locations of Harris interest points, and then the Laplacian scale selection procedure is applied at these locations to find their characteristic scales.

2 Affine Adaptation

Following spatial and scale selection, each interest point is described by a spatial location x and a characteristic scale σ. However, one could argue that a single characteristic scale parameter does not adequately describe the local geometry of a texture patch. After all, many textures are highly anisotropic, having different orientations and varying scales along different directions. To obtain a more accurate estimate of local shape for some point with location x and scale σ, we compute the second moment matrix $M_I(x;\gamma,\sigma)$, as defined by eq. (2). Consider the equation

$$(x-x_0)^T M(x-x_0) = 1,$$

where $M = M_I(x_0;\gamma,\sigma)$. It describes an ellipse centered at x_0, with principal axes corresponding to eigenvectors of M_I and axis lengths equal to the square roots of the inverse of the eigenvalues. This ellipse captures the local geometry of the corresponding texture patch, provided σ matches its characteristic scale. In the implementation, we scale the axes of each ellipse so that its area is proportional to the area of the circle with radius σ.

Suppose that we have computed an estimate M of the local shape matrix at some image location x_0. Without loss of generality, let us assume that x_0 is the origin. Given M, we can *normalize* the image patch centered at x_0 by computing a coordinate transformation U that would send the ellipse corresponding to M to a unit circle:

$$\hat{x} = Ux = M^{1/2}x$$

where $M^{1/2}$ is any matrix such that $M = (M^{1/2})^T(M^{1/2})$. Since M is a symmetric positive-definite matrix, it can be diagonalized as $M = Q^T DQ$ where Q is orthogonal and D is diagonal (with positive entries), so we can set $M^{1/2} = D^{1/2}Q$.

Second-moment matrices can also be defined in *affine Gaussian scale space* where local and integration scales are no longer described by single parameters γ and σ, but by 2×2 covariance matrices Γ and Σ [4]:

$$M_I(x;\Gamma,\Sigma) = G(x;\Sigma) * \left(\nabla I(x;\Gamma)\ \nabla I(x;\Gamma)^T\right)\,, \qquad \text{where}$$

$$\begin{aligned} G(x;\Sigma) &= \frac{1}{2\pi\sqrt{\det\Sigma}}\exp\left(-\frac{x^T\Sigma^{-1}x}{2}\right), \\ \nabla I(x;\Gamma) &= \nabla G(x;\Gamma) * I(x)\,. \end{aligned}$$

Consider two texture patches I and I' centered at locations x_0 and x'_0 (we can take x_0 and x'_0 as the origins of the respective local coordinate systems). Suppose that these patches are related by an affine transformation $x' = Ax$:

$$I(x) = I'(x') \ .$$

Then we can show [1, 4] that their second moment matrices are related as

$$\begin{aligned} M_I(x_0; \Gamma, \Sigma) &= A^T M_{I'}(x'_0; A\Gamma A^T, A\Sigma A^T) A \quad \text{or} \\ M &= A^T M' A \ . \end{aligned} \tag{3}$$

We can visualize the above transformation by thinking of the two ellipses $x^T M x = 1$ and $x'^T M' x' = 1$ related by the coordinate change $x' = Ax$. If we have computed M and M', we can normalize the local neighborhoods of I and I' as in (2):

$$\hat{x} = M^{1/2} x, \qquad \hat{x}' = M'^{1/2} x' \ .$$

In a realistic situation, A, the coordinate transformation between the image regions I and I', will be unknown. In fact, the two regions may not be related by an affine transformation at all. For this reason, (3) is really an equation where $M^{1/2}$ and $M'^{1/2}$ are known, and A must be determined. As it turns out, the solution is not unique [1]:

$$A = M'^{-1/2} R M^{1/2},$$

where R is an arbitrary orthogonal matrix. We can rewrite the above equation as

$$\begin{aligned} M'^{1/2} A x &= R M^{1/2} x \\ \hat{x}' &= R \hat{x} \ . \end{aligned}$$

We have arrived at the following conclusion: if the local coordinate systems of I and I' are related by an affine transformation, then the respective normalized systems are related by an arbitrary orthogonal transformation (a combination of rotation and reflection). Thus, we cannot obtain a complete registration between two image patches by using the local shape matrices alone, even if these can be computed exactly. As explained elsewhere in this dissertation, we can resolve the remaining ambiguity either by computing rotation-invariant descriptors of the patches, or by aligning their dominant gradient orientations.

Finally, note that the geometry estimate provided by the local shape matrix can be reasonably accurate only when the neighborhood over which the local shape matrix is computed matches the true shape of the texture patch. This is a bootstrapping problem that can be solved by an iterative approach [3, 4]. In our own implementation, the inclusion of an iterative scheme has not resulted in any measurable change in performance, and therefore we use just a single step of affine adaptation for all the experiments presented in this dissertation.

References

[1] J. Gårding and T. Lindeberg. Direct computation of shape cues using scale-adapted spatial derivative operators. *International Journal of Computer Vision*, 17(2):163–191, 1996.

[2] T. Lindeberg. Feature detection with automatic scale selection. *International Journal of Computer Vision*, 30(2):77–116, 1998.

[3] T. Lindeberg and J. Gårding. Shape-adapted smoothing in estimation of 3-d depth cues from affine distortions of local 2-d brightness structure. *Image and Vision Computing*, 15:415–434, 1997.

[4] K. Mikolajczyk and C. Schmid. Scale and affine invariant interest point detectors. *International Journal of Computer Vision*, 60(1):63–86, 2004.