

Global Measures of Coherence for Edge Detector Evaluation

Simon Baker*

The Robotics Institute
Carnegie Mellon University
Pittsburgh, PA 15213

Shree K. Nayar

Department of Computer Science
Columbia University
New York, NY 10027

Abstract

We propose a class of benchmarks for edge detector evaluation that require no ground truth. Each benchmark consists of a large number of images of a carefully designed scene for which we enforce a constraint on the edges, for example, that they are co-linear. We sample the space of edge appearances as densely as possible by capturing the images under widely varying imaging conditions. Not only do we change the viewing geometry and the illumination direction, but we also vary the camera parameters and the physical properties of the objects in the scene. We show that the degrees to which the constraints hold in the output edge-maps can be used as highly discriminating measures of edge detector performance. The code, images, and results which form our benchmarks are all available from the website <http://www.cs.columbia.edu/CAVE/>. The code and images enable a user to compare any new detector against several previous ones with minimal effort.

1 Introduction

Assessing the performance of an edge detector is an important task that has recently received considerable attention. For example, Cho *et al.* [1997] proposed using a statistical technique known as *bootstrap*, Heath *et al.* [1997] used human ability to recognize the scene from the edge-map, Shin *et al.* [1998] used the performance of an edge-based structure from motion algorithm, Dougherty and Bowyer [1998] used human marked ground truth, and Steeger [1998] studied sub-pixel localization in an industrial inspection task. For a review of earlier work, the reader is referred to the excellent survey in [Heath *et al.*, 1997].

An important issue that was not addressed by any of these approaches is how to sample the space of edge appearances thoroughly. As we shall show (see Figure 3), the performance of most edge detectors is highly dependent upon the viewing direction, the illumination conditions, the camera settings, and the physical properties of the scene. In this paper, we propose a set of benchmarks for edge detector evaluation that allow us to do two things:

1. Average over a wide range of imaging conditions, and
2. Compare performance across different conditions.

*The research described in this paper was conducted while the first author was with the Department of Computer Science at Columbia University. It was supported by an ONR/DARPA MURI grant under ONR contract No. N00014-97-1-0553.

Each of our benchmarks is based upon a constraint on the edges in the scene, for example, that they are co-linear. After applying the edge detector, the degree to which the constraint holds in the output edge map is estimated and used as the measure of performance. For the co-linearity constraint mentioned above, the measure of performance is a measure of the extent to which the detected edges are co-linear. Since the benchmarks are functions of how well the detected edges satisfy some scene constraint, we refer to them as *global measures of coherence*. In this paper, we describe four different global measures of coherence.

To sample the space of edge appearances densely, we carefully construct a scene for which the appropriate constraint holds. We then use robots to capture a large number of images, independently varying the viewing direction, the illumination conditions, and the camera parameters. By constructing similar scenes from different materials, we can also vary the reflectance properties of the objects, as well as shape properties such as the roundedness of the surface normal edge discontinuities. (It is possible to sample the space of images just as widely using synthetic images, however these images would not take into account the sources of noise that corrupt real images.)

Because our scene constraints are all based upon projective geometric structures, such as points, lines, and conics, there is a close relationship between our benchmarks and the approach of Coehlo *et al.* [1992] who suggested evaluating edge detectors using their performance estimating projective invariants. Besides their not attempting to sample the space of edge appearances, the other major difference between our approach and theirs is that their technique requires the precise measurement of quantities in the scene. On the other hand, our approach requires *no ground truth*. Further advantages of our approach include:

Ease of use: Many of the evaluation methodologies that have been proposed for edge detectors may never be used solely because doing so is too time consuming. All a designer of an edge detector has to do to evaluate their detector using our benchmarks is download the code and images from the website <http://www.cs.columbia.edu/CAVE/>, modify their edge detector to output the edges in the correct format, and finally run a script. In all, at most 2-3 hours is required. All of the results for the four detectors that we tested are also available for comparison.

Representative of a class of applications: We feel that

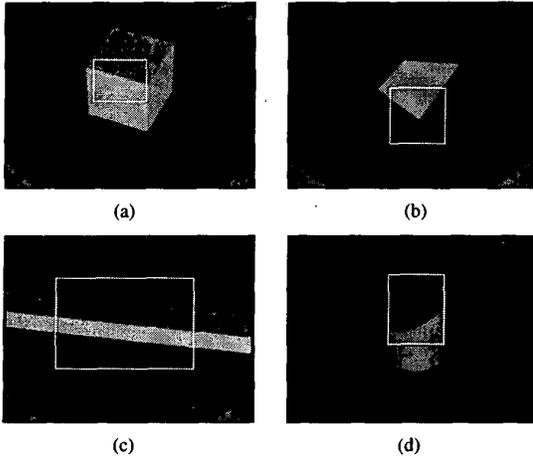


Figure 1: Cropped regions exhibiting the four constraints: (a) all edges are co-linear, (b) all edges intersect at a single point, (c) all edges are parallel in the scene, and (d) all edges lie on an ellipse. These constraints hold true irrespective of the imaging and illumination conditions.

it is probably impossible to devise an evaluation methodology that is appropriate for all applications. On the other hand, application specific techniques are of little use to the designer of a new edge detector. We feel that the correct balance is probably a collection of easy-to-use benchmarks, each designed for a subclass of applications. In this paper, we will argue that our benchmarks are appropriate for applications requiring precise sub-pixel localization and orientation estimation, good examples of which include the Hough transform, stereo matching, structure from motion, and the estimation of projective invariants.

2 Four Global Measures of Coherence

In this section, we introduce our four global measures of coherence. Each measure is derived from a constraint on the edges in the scene. The four constraints are illustrated in the cropped regions of Figure 1: in (a) all of the edges are co-linear, in (b) all of the edges intersect at a single point, in (c) all of the edges are parallel in the scene, and in (d) all of the edges lie on an ellipse. Note that there are a number of variants of our measures. We mention a couple that do not require knowledge of the orientations of the edges, but otherwise do not discuss the numerous alternatives.

As input to compute the measures, we assume the edge detector has yielded a set of edges:

$$E = \{e_i = (x_i, y_i, \theta_i) \mid i = 1, \dots, n\}, \quad (1)$$

where n is the number of edges. We assume that the i^{th} edge $e_i = (x_i, y_i, \theta_i)$ passes through the point (x_i, y_i) in the image, and the normal to this edge makes an angle θ_i with the positive y -axis.

2.1 All Edges are Co-Linear

The first constraint is that all of the edges are co-linear. Such a scene can be constructed by placing a convex polygonal object with uniform reflectance in front of a perfectly black background. If an image of such a scene is cropped so that only one depth or surface normal discontinuity is visible, all of the edges will be co-linear. See Figure 1(a) for an example image of such a scene.

The fact that co-linear edges in the scene project onto co-linear edges in the image relies upon the implicit assumption that the camera adheres to pure perspective projection. In fact, all of the measures rely upon this assumption. We applied Tsai's algorithm [Tsai, 1986] to estimate the radial distortion, but found it to be negligible ($\kappa_1 \approx -2.0 \times 10^{-6} \text{ pixels}^{-2}$) for the focal length used. (Note that measures similar to ours have been used in the past to perform camera calibration for radial distortion [Brown, 1971] [Becker and Bove, 1995].)

Given one of the edges $e_i = (x_i, y_i, \theta_i) \in E$, it is possible to estimate the line that all of the edges lie on. In the projective geometric notation of [Faugeras, 1993], the representation of this line is: $\mathbf{L}_i \equiv (l_i^1, l_i^2, l_i^3)$

$$\begin{aligned} &= (x_i, y_i, 1) \wedge (x_i + \cos \theta_i, y_i + \sin \theta_i, 1) \\ &= (-\sin \theta_i, \cos \theta_i, x_i \sin \theta_i - y_i \cos \theta_i). \end{aligned} \quad (2)$$

A minor difficulty that needs to be addressed at this point is that equality is only defined up to a constant multiplicative factor in projective spaces [Faugeras, 1993]. There are two aspects to this problem:

Sign of \mathbf{L}_i : Adding 180° to θ_i does not change the line, but reverses the sign of \mathbf{L}_i . Enforcing $\theta_i \in [0, 180^\circ)$ does not solve this problem on its own. Note that, a small perturbation, say from $\theta_i = 0.1^\circ$ to $\theta_i = 179.9^\circ$, results in a large change in l_i^2 . To solve the problem, we choose θ_i to be whichever of θ_i or $\theta_i + 180^\circ$ makes $l_i^3 \geq 0$. (This can also cause a problem when the line passes close to the origin, however we simply make sure that this does not occur in any of the images used in the benchmark.)

Scale of \mathbf{L}_i : One way to fix the scale of a line, that does not lie at infinity, is to require $(l_i^1)^2 + (l_i^2)^2 = 1.0$. This is already the case in Equation (2).

We would like to use the sum of the variances of the three line coordinates as the basis for the first measure. A natural question, however, is: how should the three components be weighted? Since the natural use of \mathbf{L}_i is to test whether a point $\mathbf{x} = (x, y, 1)$ lies on the line using $\mathbf{x} \cdot \mathbf{L}_i = 0$, a sensible choice for the ratio of the weights is $(Ex)^2 : (Ey)^2 : 1$, where Ex denotes the expected value of $|x|$ and Ey denotes the expected value of $|y|$. To avoid any dependence on the distance units, we define the first global measure of coherence to be: $\text{GMC}_1 =$

$$\frac{1}{Ex \cdot Ey} [(Ex)^2 \cdot \sigma^2(l^1) + (Ey)^2 \cdot \sigma^2(l^2) + \sigma^2(l^3)] \quad (3)$$

where for $k = 1, 2, 3$:

$$\sigma^2(l^k) = \frac{1}{n} \sum_{i=1}^n \left[l_i^k - \frac{1}{n} \sum_{j=1}^n l_j^k \right]^2 \quad (4)$$

is the variance of the k^{th} coordinate of \mathbf{L}_i . There are several ways that E_x and E_y could be estimated. We define:

$$E_x = \frac{1}{n} \sum_{i=1}^n |x_i| \quad \text{and} \quad E_y = \frac{1}{n} \sum_{i=1}^n |y_i|. \quad (5)$$

A variant of the first global measure of coherence can be estimated without using the angle θ_i . Suppose $e_i = (x_i, y_i, \theta_i) \in E$ and $e_j = (x_j, y_j, \theta_j) \in E$ are two edges. The line that passes through them is: $\mathbf{L}_{ij} \equiv (l_{ij}^1, l_{ij}^2, l_{ij}^3)$

$$\begin{aligned} &= (x_i, y_i, 1) \wedge (x_j, y_j, 1) \\ &= (y_i - y_j, x_j - y_j, x_i \cdot y_j - x_j \cdot y_i). \end{aligned} \quad (6)$$

Then, Equation (3) can be used to re-estimate the first global measure of coherence. However, the way the variances were computed in Equation (4) must be modified. If the two edges that are used to compute \mathbf{L}_{ij} are very close to each other, the estimate of \mathbf{L}_{ij} will be unnecessarily noisy. We only allow edges to contribute to the measure if they are more than 5 pixels apart. In particular, consider the set of edge index pairs: $P_1 =$

$$\{(i, j) \mid i < j \text{ and } (x_i - x_j)^2 + (y_i - y_j)^2 > 5^2\}. \quad (7)$$

Then, the variances for the second variant of GMC_1 are:

$$\sigma^2(l^k) = \frac{1}{|P_1|} \sum_{(i,j) \in P_1} \left[l_{ij}^k - \frac{1}{|P_1|} \sum_{(l,m) \in P_1} l_{lm}^k \right]^2. \quad (8)$$

2.2 All Edges Intersect at a Single Point

The second constraint is that all of the edges intersect at a single point. See Figure 1(b) for an example image of such a scene. Given two of the edges $e_i = (x_i, y_i, \theta_i)$, $e_j = (x_j, y_j, \theta_j) \in E$ that are not parallel, it is possible to estimate the point where the edges intersect. If $\mathbf{L}_i = (l_i^1, l_i^2, l_i^3)$ and $\mathbf{L}_j = (l_j^1, l_j^2, l_j^3)$ are the projective representations of the two edges, they intersect at the point:

$$\mathbf{P}_{ij} \equiv (p_{ij}^1, p_{ij}^2, p_{ij}^3) = (l_i^1, l_i^2, l_i^3) \wedge (l_j^1, l_j^2, l_j^3). \quad (9)$$

Since the lines \mathbf{L}_i and \mathbf{L}_j are not parallel, $p_{ij}^3 \neq 0$. Hence, \mathbf{P}_{ij} can be divided by its third coordinate to give the normalized point of intersection:

$$\bar{\mathbf{P}}_{ij} \equiv (\bar{p}_{ij}^1, \bar{p}_{ij}^2, 1) = \left(\frac{p_{ij}^1}{p_{ij}^3}, \frac{p_{ij}^2}{p_{ij}^3}, 1 \right). \quad (10)$$

Then, we define the second measure of coherence to be:

$$\text{GMC}_2 = \frac{1}{(E_x)^2} \sigma^2(\bar{p}^1) + \frac{1}{(E_y)^2} \sigma^2(\bar{p}^2) \quad (11)$$

where for $k = 1, 2$:

$$\sigma^2(\bar{p}^k) = \frac{1}{|P_2|} \sum_{(i,j) \in P_2} \left[\bar{p}_{ij}^k - \frac{1}{|P_2|} \sum_{(l,m) \in P_2} \bar{p}_{lm}^k \right]^2 \quad (12)$$

is the variance of the k^{th} coordinate of $\bar{\mathbf{P}}_{ij}$,

$$P_2 = \{(i, j) : |\theta^i - \theta^j| > 15^\circ\}, \quad (13)$$

is the set of edge index pairs that are not too close to parallel, and E_x and E_y are defined in Equation (5).

2.3 All Edges are Parallel

The third constraint is that all of the edges are parallel in the scene. See Figure 1(c) for an example image of such a scene. Unlike the previous two constraints, parallel edges in the scene do not always remain parallel in an image because of perspective foreshortening effects. However, if the camera adheres to perspective projection, parallel edges in the scene are projected onto edges that intersect at a single point in the image, namely the vanishing point.

We begin in the same way as Section 2.2 by estimating the intersection of the lines passing through two edges $e_i = (x_i, y_i, \theta_i)$, $e_j = (x_j, y_j, \theta_j) \in E$. If $\mathbf{L}_i = (l_i^1, l_i^2, l_i^3)$ and $\mathbf{L}_j = (l_j^1, l_j^2, l_j^3)$ are the two lines, they intersect at the vanishing point:

$$\mathbf{V}_{ij} \equiv (v_{ij}^1, v_{ij}^2, v_{ij}^3) = (l_i^1, l_i^2, l_i^3) \wedge (l_j^1, l_j^2, l_j^3). \quad (14)$$

The vanishing point \mathbf{V}_{ij} may or may not lie at infinity, so v_{ij}^3 may or may not equal 0. If \mathbf{V}_{ij} does lie at infinity, which is approximately the case in the benchmark images, the only useful information is the angle that it makes with the x -axis:

$$\phi_{ij} = \arctan \left(\frac{v_{ij}^2}{v_{ij}^1} \right). \quad (15)$$

The questions of weighting and distance units are not such an issue here, but the cyclic range of the arctan function is problematic; a vanishing point at one end of the range can easily be perturbed by noise to lie at the other end. The solution we adopted is to use the range $[-90^\circ, 90^\circ]$, making sure that the correct vanishing point never lies close to either 90° or -90° . Finally, we define the third global measure of coherence to be: $\text{GMC}_3 =$

$$\frac{1}{(180^\circ)^2} \cdot \frac{1}{|P_3|} \sum_{(i,j) \in P_3} \left[\phi_{ij} - \frac{1}{|P_3|} \sum_{(l,m) \in P_3} \phi_{lm} \right]^2 \quad (16)$$

where:

$$P_3 = \{(i, j) : |l_i^3 - l_j^3| > 20 \text{ pixels}\} \quad (17)$$

and $l_3^i = x^i \sin \theta^i - y^i \cos \theta^i$ is the third line coordinate. P_3 excludes any edge pairs that may lie on the same line.

2.4 All Edges Lie on an Ellipse

The fourth constraint is that all of the edges lie on an ellipse. An example of such a scene is presented in Figure 1(d). An ellipse is defined by $\mathbf{xAx}^T = 0$, where $\mathbf{x} = (x, y, 1)$ is a homogeneous vector of image coordinates, and:

$$\mathbf{A} = \frac{1}{2} \begin{pmatrix} 2 \cdot a_{11} & a_{12} & a_{13} \\ a_{12} & 2 \cdot a_{22} & a_{23} \\ a_{13} & a_{23} & 2 \cdot a_{33} \end{pmatrix} \quad (18)$$

is a 3×3 symmetric matrix, as usual defined up to a scale factor [Faugeras, 1993]. So, \mathbf{A} has just five independent parameters. The fact that the i^{th} edge $e_i = (x_i, y_i, \theta_i) \in E$ lies on the ellipse provides one constraint on \mathbf{A} :

$$(x_i, y_i, 1)\mathbf{A}(x_i, y_i, 1)^T = 0. \quad (19)$$

Since this constraint is linear in the six unknowns, the ellipse can be recovered using Gauss-Jordan elimination given the locations of five edges on the ellipse. We fix the scale factor and sign of \mathbf{A} by enforcing $a_{33} = 1.0$, while using benchmark images for which the ellipse does not pass too close to the origin to ensure that the sign of a_{33} cannot be perturbed accidentally.

Since $\mathbf{xAx}^T = 0$ is used to determine whether $\mathbf{x} = (x, y, 1)$ lies on the ellipse, we define the fourth global measure of coherence to be: $\mathbf{GMC}_4 =$

$$\frac{(Ex)^4 \sigma^2(a_{11}) + (Ex)^2 \sigma^2(a_{13}) + (Ey)^4 \sigma^2(a_{22}) + (Ey)^2 \sigma^2(a_{23}) + (Ex)^2 (Ey)^2 \sigma^2(a_{12})}{(Ex)^4 \sigma^2(a_{11}) + (Ex)^2 \sigma^2(a_{13}) + (Ey)^4 \sigma^2(a_{22}) + (Ey)^2 \sigma^2(a_{23}) + (Ex)^2 (Ey)^2 \sigma^2(a_{12})} \quad (20)$$

where $\sigma^2(a_{ij})$ is the variance of a_{ij} . Since it takes five edges to estimate the ellipse, the variances of the parameters a_{ij} are all computed over the set of quintuples of edges, no pair of which are closer than 30 pixels apart.

If the orientation of each edge is known, it is possible to estimate the ellipse parameters using just three edges. Each edge $e_i = (x_i, y_i, \theta_i) \in E$ provides two constraints on the parameters. The first one is $(x_i, y_i, 1)\mathbf{A}(x_i, y_i, 1) = 0$, as above. The second one is that the tangent to the ellipse has orientation θ_i . Differentiating $\mathbf{xAx}^T = 0$, setting $\frac{dy_i}{dx_i} = \tan \theta_i$, and reorganizing gives the tangency constraint:

$$2x_i a_{11} + a_{13} + y_i a_{12} + [2y_i a_{22} + a_{23} + x_i a_{12}] \tan(\theta_i) = 0. \quad (21)$$

Since this constraint is also linear in the ellipse parameters, three ellipse constraints and two tangency constraints can be used to estimate the a_{ij} . A second variant of the fourth measure of coherence can then be defined exactly as above, but with the variance now being computed across all triples of edges, no pair of which are closer than 50 pixels apart. Similarly, four ellipse constraints and one tangency constraint can be used to give a third variant.

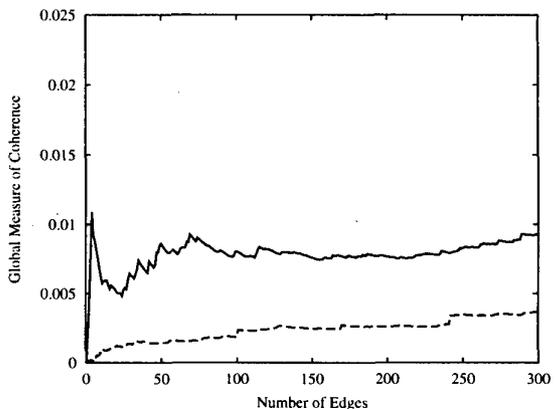


Figure 2: Plots of the first measure of coherence against the number of edges for two detectors. These plots were obtained by varying the thresholds inherent in the detectors. The dotted curve lies below the solid one and hence represents superior performance.

3 Computing Global Measures of Coherence

3.1 Comparing Detectors Across Parameters

Most edge detectors have various parameters or thresholds that must be set. For example, gradient based detectors usually threshold on the magnitude of the gradient. Because the value of any performance metric will depend upon these parameters, comparing detectors independently of their parameter settings requires some effort.

In the simplest case, there is only one parameter. As this parameter is varied, more or less edges will be detected, and the global measure of coherence will change. Usually, as more edges are detected, the measure gets worse. So, by varying the parameter, a curve can be plotted of the number of edges detected against the global measure of coherence. Two detectors can be compared by plotting these curves. The closer the curve lies to the abscissa the better the performance. See, for example, Figure 2.

For detectors with more than one parameter, the situation is more complicated. As the parameters vary, both the number of edges detected and the global measure of coherence change, forming a “cloud” of performance measurements [Dougherty and Bowyer, 1998]. Dougherty and Bowyer [1998] proposed searching the parameter space to convert this cloud of measurements into a curve defining the set of optimal parameter settings. For proof of concept, we simply used detectors with one threshold. Dougherty and Bowyer’s algorithm can be used for other detectors.

3.2 Sampling the Space of Edge Appearances

The appearance of an edge depends upon both the imaging conditions and the physical properties of the scene. There are several aspects to each of these variables:

- The pose of the objects.
- The illumination conditions.
- The camera settings: e.g. zoom, aperture, and focus.
- The reflectance properties of the objects.

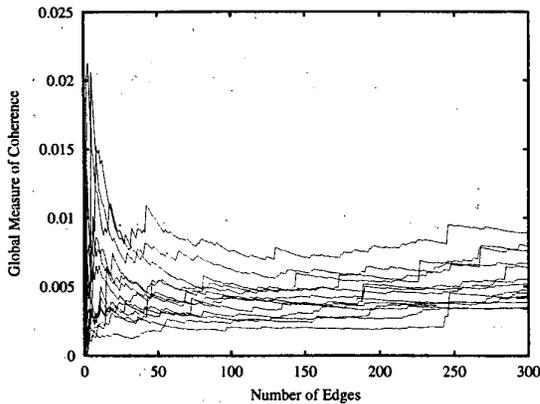


Figure 3: Plots of the first global measure of coherence against the number of edges for several different images, each captured with different pose and illumination direction. These plots demonstrate the huge range of performance that is possible for a single detector. As can be seen, the standard deviation of the measure is comparable to the average value of the measure.

- The discontinuity type: e.g. depth or surface normal.
- The roundedness of the edge.

In our experimental setup, any or all of these properties can be varied independently. The pose and illumination direction were varied using robots and turntables. The camera settings were changed by hand. The objects were manufactured with three different surface properties, one approximately Lambertian, one a diffuse metallic paint, and one a shiny plastic. The discontinuity type was varied by cropping different parts of the images. Lastly, for each type of reflectance property, multiple objects were created, each with edges having a different radius of curvature. The range of different performance levels across imaging conditions for a single detector is illustrated in Figure 3.

3.3 Averaging over a Collection of Images

As is illustrated in Figure 3, the variance of the global measures of coherence over the imaging conditions is typically quite large. On the other hand, the relative performance across the detectors is significantly less dependent upon the imaging conditions. So, instead of comparing the average measures across the detectors, we consider:

$$\text{RGMC}^{i,j} = \frac{\text{GMC}^j - \text{GMC}^i}{\text{GMC}^j + \text{GMC}^i} \quad (22)$$

where GMC^i is the global measure of coherence for detector i and GMC^j is the measure for detector j . The relative global measure of coherence $\text{RGMC}^{i,j}$ takes values in $[-1, 1]$. A positive value means that detector i is outperforming detector j , and vice versa.

4 Empirical Validation of the Measures

How do you empirically validate an evaluation methodology? Ideally, we would want to demonstrate a strong statis-

tical correlation between the performance of a large number of detectors on our measures and their performance in a wide range of applications. Doing so, however, would be hugely time-consuming. Therefore, as proof of concept, our goals are simply as follows:

- Demonstrate the range of performance: An evaluation methodology is useless unless the best detectors do substantially better than the worst. Our first goal, therefore, is to show that a wide range of different performance levels are possible. To do so, we consider both quite good detectors and relatively poor ones. On the other hand, if we only tested good detectors, we would never know the full range of performance.
- Demonstrate consistency between intuitive ranking and actual ranking: If the results of our measures disagree with the general perception of the quality of the detectors, there is something wrong. Our second goal, therefore, is to demonstrate that the results using our measures are in agreement with the general perception of the quality of the detectors tried. We make our task as easy as possible by selecting detectors for which the relative performance gap is as large as possible.
- Show that the measures are representative of certain tasks: As mentioned in the introduction, our measures are designed for tasks requiring precise sub-pixel localization and orientation estimation. To demonstrate that they are indeed representative of such tasks, we aim to show that performance drops off substantially for detectors that are known to give poor sub-pixel localization and orientation estimates.

4.1 Detectors Considered

We considered five detectors. For lack of space, we are only able to give very brief descriptions of the detectors:

1. The Roberts' cross operator, a very simple gradient operator [Pratt, 1991]. Because it uses a 2×2 window, the Roberts' cross operator should perform very badly.
2. The 3×3 Sobel operator [Pratt, 1991]. The Sobel operator should perform slightly better than the Roberts' cross operator, but still quite poorly.
3. A 5×5 Gaussian-smoothed gradient operator. This "Canny-like" operator should perform quite well.
4. The Nalwa-Binford detector, a model-matching detector based on a 5×5 window [Nalwa and Binford, 1986]. Like the other 5×5 operator, the Nalwa-Binford detector is expected to perform pretty well.
5. A model-matching detector that uses a 7×7 window [Baker *et al.*, 1998]. Since this detector uses the largest window, we expect it to perform the best.

4.2 Types of Comparison

There are two types of experiments that we performed:

1. Comparison of several detectors under the same set of imaging conditions: We vary some of the imaging conditions and compute the average performance of several detectors under those conditions. We then study the relative performance of the detectors.

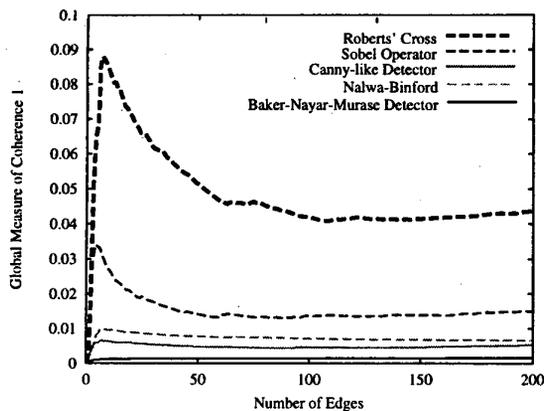


Figure 4: Plots of the first measure of coherence against the number of edges for five detectors. This measure clearly discriminates the better detectors from the poorer ones.

- Investigation of the change in performance of a detector as one aspect of the imaging conditions changes: We choose one aspect of the imaging conditions, for example, the focus setting of the camera. For several different settings of it, we vary some of the other aspects. We then compute the average performance of a single detector for each of the settings of the chosen aspect and study how the performance of the detector varies with that aspect of the imaging conditions.

4.3 Results

We only have room to present an illustrative sample of our results. The others are all similar and can be found at <http://www.cs.columbia.edu/CAVE/>. Figure 4 contains plots of the first global measure of coherence computed on average over 150 images, captured by varying both the pose and the illumination direction. As can be seen, the first measure captures a huge range of performance variation. Moreover, the ranking of the detectors is completely consistent with general belief. As described above, however, the variation across imaging conditions is large and the variance of the curves in Figure 4 is huge. In Figure 5 we present the relative measure, computed on average for the same set of 150 images. The performance is relative to Nalwa-Binford. Here, the standard deviation is considerably reduced, to almost exactly 0.4 for all four detectors and for all numbers of edges. Figure 5 demonstrates that by our first measure, the Baker-Nayar-Murase detector performs substantially better than the Nalwa-Binford detector, and the Roberts' cross detector far worse.

In Figure 6, we compare the performance of the Baker-Nayar-Murase detector across the two variants of the first measure of coherence. (The first variant fits a line to one edge location and orientation, whereas the second estimates the line from two edge locations.) We also compare the performance with and without sub-pixel localization turned on. As expected, these results demonstrate that sub-pixel

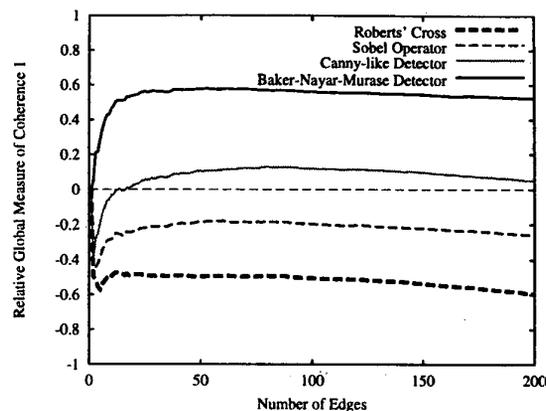


Figure 5: Plots of the relative measure of coherence for the same set of images used in Figure 4. The standard deviation of all the curves is approximately 0.4 for all numbers of edges.

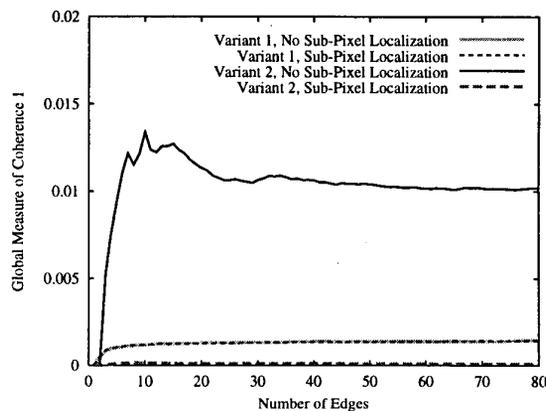


Figure 6: Plots of the two variants of the first measure of coherence for a detector, both with and without sub-pixel localization. The fact that only the performance of the second variant is hugely improved using sub-pixel localization indicates that the second variant of the measure is representative of tasks requiring precise sub-pixel localizations.

localization is very important for the second variant, but not for the first. Similar results show that orientation estimation is very important for the third global measure of coherence, whereas sub-pixel localization is not.

Figure 7 contains results obtained by varying the aperture of the camera for the third measure. As might be expected, the performance of both the Nalwa-Binford and Canny-like detectors gets significantly better as the aperture is widened from F4.0 to F1.8. Similar results show that performance also improves for more focused images.

5 Discussion

In this paper, we have presented global measures of coherence as a method of benchmarking an edge detector for applications that require precise sub-pixel localization and orientation estimation. The major advantages of our ap-

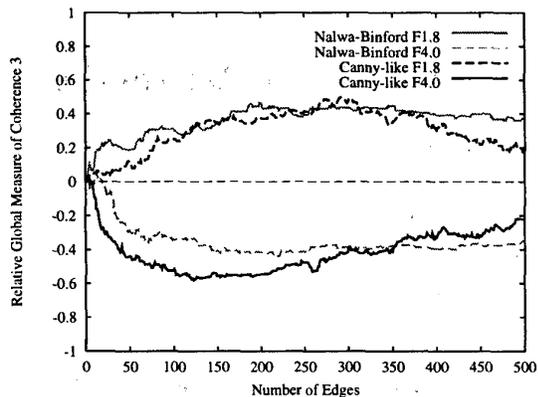


Figure 7: The third global measure of coherence plotted relative to the performance for an F2.8 aperture. As the aperture gets larger, the performance of both detectors gets better.

approach are: (1) it allows us to average over a large number of images that span a wide range of imaging conditions, and (2) it does not require ground truth. Our experimental results demonstrate that the measures discriminate many different levels of performance. Moreover, we have shown that our approach can be used to compare performance across different imaging conditions.

5.1 Limitations

Like all evaluation methodologies, global measures of coherence have limitations. Perhaps the most apparent is that the scenes used are very simple. Detecting the correct edges should be straightforward for any decent edge detector. Our benchmarks are explicitly designed for applications that use vision as a measurement tool, where determining the exact location and orientation of the edge is the most important goal. We would argue, however, that these applications are the ones for which applying an edge detector is the most appropriate.

A second limitation is that interpreting the absolute values of the measures is quite difficult because they are complex averages of the variances of projective quantities. Recent work on error propagation in projective spaces may help in understanding the precise meanings of our measures [Kanatani, 1993].

5.2 Future Work

We showed that our measures can be used to investigate how performance degrades as the camera parameters change. In particular, we varied the focus setting and the aperture. A large number of other experiments could have been performed. Conducting these experiments may lead to a far greater understanding of which physical effects actually make edge detection difficult. This may then lead to more realistic edge models, physically validated noise distributions, and eventually better edge detectors.

References

- [Baker *et al.*, 1998] S. Baker, S.K. Nayar, and H. Murase. Parametric feature detection. *IJCV*, 27(1):27–50, 1998.
- [Becker and Bove, 1995] S. Becker and V.M. Bove. Semi-automatic 3-D model extraction from uncalibrated 2-D camera views. In *Proc. of SPIE Visual Data Exploration and Analysis II*, volume 2410, pages 447–461, 1995.
- [Brown, 1971] D.C. Brown. Close-range camera calibration. In *Symp. on Close-Range Photogrammetry*, 1971.
- [Cho *et al.*, 1997] K. Cho, P. Meer, and J. Cabrera. Performance assessment through bootstrap. *IEEE Trans. on PAMI*, 19(11):1185–1198, 1997.
- [Coehlo *et al.*, 1992] C. Coehlo, A. Heller, J.L. Mundy, D.A. Forsyth, and A. Zisserman. An experimental evaluation of projective invariants. In J.L. Mundy and A. Zisserman, editors, *Geometric Invariants for Machine Vision*, chapter 4, pages 87–104. MIT Press, 1992.
- [Dougherty and Bowyer, 1998] S. Dougherty and K.W. Bowyer. Objective evaluation of edge detectors using a formally defined framework. In *Proc. of the 1998 Workshop on Empirical Evaluation Techniques in Computer Vision*, pages 211–234, 1998.
- [Faugeras, 1993] O. Faugeras. *Three-dimensional Computer Vision: A Geometric Viewpoint*. MIT Press, 1993.
- [Heath *et al.*, 1997] M.D. Heath, S. Sarkar, T. Sanocki, and K.W. Bowyer. A robust visual method for assessing the relative performance of edge-detection algorithms. *IEEE Trans. on PAMI*, 19(12):1338–1359, 1997.
- [Kanatani, 1993] K. Kanatani. *Geometric Computation for Machine Vision*. Oxford University Press, 1993.
- [Nalwa and Binford, 1986] V.S. Nalwa and T.O. Binford. On detecting edges. *IEEE Trans. on PAMI*, 8(6):699–814, 1986.
- [Pratt, 1991] W.K. Pratt. *Digital Image Processing*. Wiley-Interscience, 1991.
- [Shin *et al.*, 1998] M.C. Shin, D. Goldgof, and K.W. Bowyer. An objective comparison methodology of edge detection algorithms using a structure from motion task. In *Proc. of the 1998 Workshop on Empirical Evaluation Techniques in Computer Vision*, pages 235–254, 1998.
- [Steeger, 1998] C. Steeger. Analytical and empirical performance evaluation of subpixel line and edge detection. In *Proc. of the 1998 Workshop on Empirical Evaluation Techniques in Computer Vision*, pages 188–210, 1998.
- [Tsai, 1986] R.Y. Tsai. An efficient and accurate camera calibration technique for 3D machine vision. In *Proc. of CVPR 1986*, pages 364–374, 1986.