A Gradient-Based Evidence Measure for Image Matching

Daniel Scharstein*

Department of Computer Science
Cornell University
Ithaca, NY 14853, USA
schar@cs.cornell.edu

Abstract

We present a simple yet powerful method to perform point-to-point matching between two images. The method uses an *evidence measure*, whose value for a given displacement reflects both the similarity between two locations and the confidence in a correct match. The measure is based on the gradient fields of the images, and can be computed quickly and in parallel. Accumulating the evidence measure for different displacements allows (1) stable computation of correspondences without smoothing across motion boundaries, and (2) detection of dominant motions, which can serve as attention cues in active vision systems. The method works well both on highly textured images and on images containing regions of uniform intensities, and can be used for a variety of applications, including stereo vision, motion segmentation, and object tracking.

Keywords: Image Registration, Motion Computation, Stereo Matching, Active Vision

*This work was supported in part by National Science Foundation PYI grant IRI-9057928 and matching funds from Xerox Corp.

1 Introduction

A fundamental problem in computer vision is the so-called *correspondence problem*, that is, to establish point-to-point correspondences across a pair of images. Solving this problem is important for a large number of subsequent tasks, including computation of visual motion, recovering depth from stereo, and object tracking. Most algorithms for computing correspondence have the following *point-oriented* control strategy: For each location in one image, find the displacement that aligns this location with the best matching location in the other image.

The method presented in this paper uses a different approach: Given a certain displacement, find all the locations that match well. That is, we propose a *displacement-oriented* control strategy. Under the assumption that the visual motion between two images can be locally approximated by pure translation, near points corresponding to the same object have similar displacements. By accumulating evidence for matches under a given displacement, dominant motions can be detected, which can serve as attention cues in active vision systems.

Comparing locations in two images involves a *matching criterion*: a measure of goodness of a proposed match. A key observation in this paper is that most methods for computing correspondences have *two* underlying criteria:

- a *similarity* criterion that reflects how well two locations in the two images resemble each other;
- a *confidence* criterion that reflects the likelihood that a match is correct.

Existing methods often treat these two criteria separately. Our method uses a single measure, which — given a certain displacement — gives a (strong) positive response where points match with (high) confidence, a negative response where there is a clear mismatch, and zero response in regions where there is neither evidence for a match nor evidence against a match. The measure is based on comparing the *gradient fields* of the images.

This approach has the following advantages:

- The evidence measure, which is only based on the local gradients, can be computed quickly and in parallel.

- For a given displacement, the measure can be accumulated by averaging over a larger area. The average value represents evidence for or against a match, thus enabling the use of a displacement-oriented control strategy.

- Finding maxima in the accumulated measure is a stable way of computing correspondences without smoothing across motion boundaries.

We will discuss each of these properties in more detail later.

1.1 Related Work

A vast amount of work has been done on computing correspondences, and we will only list a few examples here. Generally, one can distinguish between *feature-based* and *area-based* approaches [Barnard and Fischler, 1982], which utilize the concepts of similarity and confidence in a different order.

Feature-based methods extract distinctive features (e.g., by doing edge detection) before the matching process and thereby try to decide off-line which locations in the image can be matched with high confidence. All other locations are ignored in the actual matching process, resulting in a sparse output. The measure of similarity used during matching is usually based on attributes of the features; for example, if the extracted features are the intensity edges, one could compare their orientation, length, and contrast. Grimson [1985] describes a feature-based stereo matcher. For examples of feature-based object tracking methods, see Huttenlocher *et al.* [1993] and Koller *et al.* [1993].

Area-based methods try to match all locations in the image, and the desired output is a dense field of displacement vectors. In order to cope with locations of little intensity change, small windows, or *areas* — which hopefully are not completely uniform — are matched as a whole. Usually, some kind of correlation measure (or the sum of squared differences) of the intensities is used to reflect the similarity of two windows. Anandan [1989] gives a good overview of area-based systems for motion computation. An example for an area-based method used for object tracking is described by Woodfill and Zabih [1991]. Cochran and Medioni [1992] describe a stereo matcher that uses a combination of an area-based and a feature-based method.

Comparing windows instead of single points allows matching with higher confidence, but is prone to problems if an object changes shape too much from one image to another, or if the window contains motion boundaries. Some area-based methods (e.g., Anandan [1989]) also compute the confidence associated with each match explicitly and use this information in a subsequent smoothing step of the computed displacements.

Assuming only Gaussian noise, using intensity differences as a cost to minimize is optimal [Anandan, 1989; Matthies *et al.*, 1989; Simoncelli *et al.*, 1991]. However, this assumption is easily violated: two cameras can differ in bias and gain, and intensities can depend on the position in the image due to *vignetting*. The gradient-based method presented in this paper is less sensitive to these problems. Non-parametric measures as used by Zabih and Woodfill [1994] are a different way of addressing these problems. Seitz [1989] uses local gradients for object recognition.

Aligning two images pixel by pixel is also referred to as *image registration*; a typical application in this area is the computation of elevation maps from a pair of satellite images, which is a central problem in photogrammetry [Moffitt and Mikhail, 1980]. Emphasis is placed on high accuracy, often on a sub-pixel level [Tian and Huhns, 1986], and matches are usually computed with an area-based method using correlation.

The idea of a control strategy that collects support for given displacements bears some similarity to Marr's model of the human stereo system involving a set of disparity pools [Marr, 1982]. Prazdny [1985] describes a stereo matching algorithm that collects support for different disparity hypotheses in a manner similar to the accumulation step in our method. His algorithm, however, requires an initial set of possible disparity hypotheses collected by explicit feature matching.

Coombs and Brown [1993] describe an active stereo vision system that finds points at the depth of fixation (the so-called *horopter*) by means of a feature-based *zero-disparity filter* (see also Coombs *et al.* [1992]). Olson and Lockwood [1992] describe a different way of disparity filtering using a multi-scale correlation method to extract points at zero disparity. Both approaches differ from the one described in this paper in that they do not return a measure that reflects the evidence for a match at a certain position.

A short version of this paper appears in ICPR '94 [Scharstein, 1994].

2 Measuring Evidence for Matches

In this section, we will describe the proposed method of comparing gradient fields in more detail. The particular measure we introduce has proven to work quite well, and is an example of a measure that can be used in a displacement-oriented control strategy. Possible other methods are discussed in section 6.

In the following, we will treat an image as a continuous intensity function $I(x, y)$; we will discuss dealing with discrete images in section 2.3.

2.1 Comparing Two Gradient Vectors

As mentioned above, our method combines the notions of similarity and confidence (or distinctiveness) into a single measure of *evidence* for or against a match at a certain location under a certain displacement, based on the two gradients at this location. In particular, if $\mathbf{g}_L$, $\mathbf{g}_R$ are the two gradient vectors to be compared, we use the average magnitude of the two gradients $\overline{m} = (|\mathbf{g}_L| + |\mathbf{g}_R|)/2$ at a certain point to represent confidence, and the (negated) magnitude of the difference of the two gradients $-d = -|\mathbf{g}_L - \mathbf{g}_R|$ to represent similarity. We define the *evidence* for a match to be the weighted sum of these two terms:

$$e = \overline{m} - \alpha d.$$

It turns out that $\alpha = 1$ is a reasonable choice for the weight parameter, since it yields a symmetric range $[-m, m]$ of values for e for the case of comparing two vectors of length m. (Evidence $e = m$ if the two vectors have the same direction, and $e = -m$ if the two vectors have opposite directions.) See Figure 1 for an illustration of different values of e for pairs of gradient vectors of length m and 0.

If both gradients are zero, one can't tell whether or not they match, and consequently $e = 0$. (This measure ignores the original intensities, although one can argue that they provide additional information. However, comparing absolute intensities has proven to be not very stable in practice.) Note that e can also be zero for two non-zero gradient vectors, for example, in the case of two vectors of equal length defining an angle of $60°$. Intuitively, this reflects the situation where the directions of gradients are too different to consider it a match, but not different enough to count it as a mismatch. Of course, the

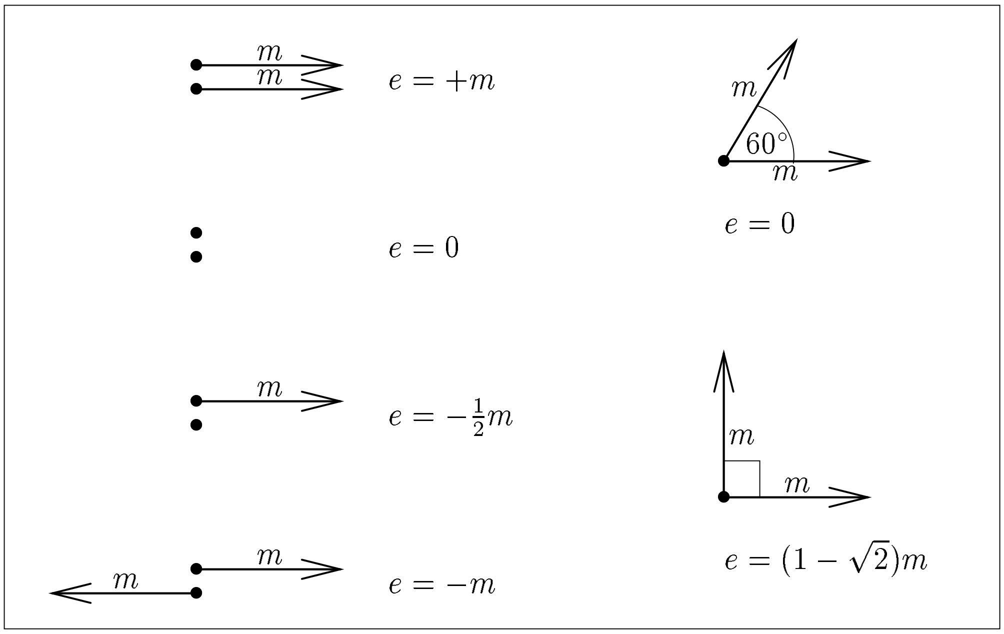

Figure 1: Value of evidence measure $e = \overline{m} - d$ for different pairs of gradient vectors

right value for this "angle of zero evidence" might depend on the application, in particular on how much rotation is possible in the motion between two images. By choosing a higher weight α for the gradient difference, one can reduce the angle for which $e = 0$, but our experiments indicate that changing the weight is not critical, and that $\alpha = 1$ is a reasonable general choice.

To display what values e takes on for different pairs of vectors, Figure 2 shows a contour plot of e for comparing any vector (x, y) to the unit vector $(1, 0)$. The contour lines are the locations of the endpoints of all vectors that yield the same value e.

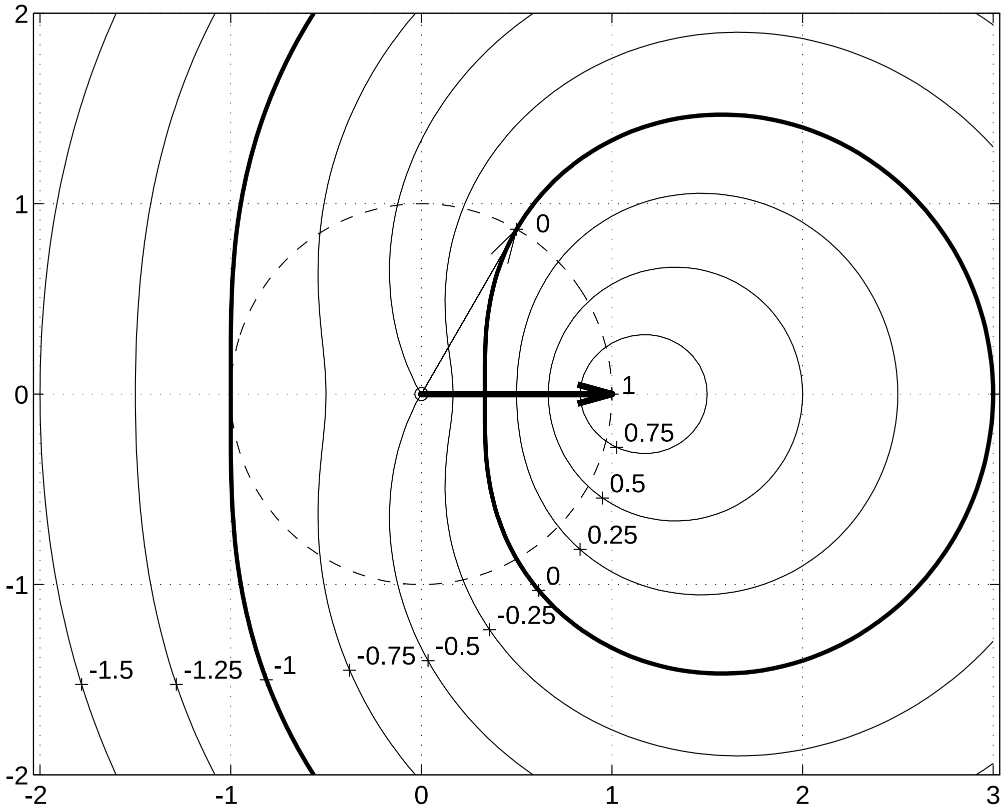

Figure 2: Contour lines of the evidence measure e for a match with the vector $(1, 0)$. The unit vector of angle $60°$ is shown as an example; note that its endpoint lies on the $e = 0$ curve.

2.2 Comparing Gradient Fields

We now extend the measure to entire images. Let $I_L(x,y)$, $I_R(x,y)$ be the two images, and let $\mathbf{G}_L$, $\mathbf{G}_R$ be their gradient vector fields[1]. That is,

$$\mathbf{G}_L = \begin{pmatrix} \frac{\partial I_L}{\partial x} \\ \frac{\partial I_L}{\partial y} \end{pmatrix}, \mathbf{G}_R = \begin{pmatrix} \frac{\partial I_R}{\partial x} \\ \frac{\partial I_R}{\partial y} \end{pmatrix}.$$

For a given displacement $\delta = (\delta_x, \delta_y)$ the *evidence* E_δ for a match at (x,y) under this displacement is

$$E_\delta(x,y) = \frac{|\mathbf{G}_L(x,y)| + |\mathbf{G}_R(x+\delta_x, y+\delta_y)|}{2} - |\mathbf{G}_L(x,y) - \mathbf{G}_R(x+\delta_x, y+\delta_y)|.$$

In this paper we only deal with displacements that are pure translations, constant at every point in the image. It is also possible to use more complex transformations, especially in situations where the possible motions between the two images are constrained. For example, in the case of a stereo setup with known calibration parameters, it would be useful to make δ a continuous transformation that keeps epipolar lines aligned.

2.3 Dealing with discrete images

In order to apply the method to real, discrete images, we approximate the gradients by differences. After an initial smoothing step with a Gaussian filter to compensate for quantization error and noise, the gradients in the x and y directions are computed by convolution with simple stencils $[-1\ 0\ 1]$ and $[-1\ 0\ 1]^T$. In the experiments reported here, we used a Gaussian filter with $\sigma = 0.5$ pixels. Also, we only consider displacements $\delta = (\delta_x, \delta_y)$ whose components are multiples of whole pixels, although it is possible to compute E_δ for non-integer displacements by interpolating the gradients.

It should be noted that, for a given displacement δ, E_δ can be computed very fast, since only a few floating point operations and a single square root is needed at each pixel. The square root is necessary to compute the magnitude of the gradient differences; the two magnitudes of gradients $|\mathbf{G}_L|$ and $|\mathbf{G}_R|$,

[1] L and R stand for left and right, suggesting a stereo vision application. Of course, the two images could also be taken by a single camera, either moving itself or observing a moving environment.

which do not depend on the displacement δ, only need to be computed once. The local nature of the computations makes the method ideally suited for a parallel implementation.

A sequential implementation on a Sparc station takes less than one second to compute E_δ for a 256×256 pixel image. In section 5 we will discuss ways to make the computation even faster.

3 Accumulating results

To find the best match for an isolated point, all we can do is to maximize E_δ at this point for all δ under consideration. Doing so independently for every point is not very stable and might produce a noisy and inconsistent displacement field.

To deal with this problem, motion computation methods usually make the assumption that nearby points have similar displacements, based on the observation that motion in real scenes varies smoothly almost everywhere. Furthermore, it is often assumed that motion can be described locally by pure translation, i.e., rotational components and effects of perspective foreshortening are small enough. Many point-oriented methods utilize the assumption of a smooth motion field *after* computing initial matches by smoothing the displacement field, often employing some confidence measure associated with each match to constrain the smoothing process [Horn and Schunck, 1981; Anandan, 1989]. The problem is that this tends to smooth over motion discontinuities, which contain important information about the scene geometry.

In contrast, our displacement-oriented method uses the assumption of a smooth motion field *while* finding the matches. The idea is that if a certain displacement δ aligns two matching objects, E_δ will have a strong positive response at the location of the match. By accumulating E_δ over a certain area (i.e., computing the average or smoothing with a Gaussian filter), dominant motions can be detected. That is, only the correct displacement E_δ will yield support for a match over a larger area, thereby creating a maximum among all δ under consideration.

Note that our method does not smooth over motion boundaries, since it is not assumed that *all* close pixels have the same disparity. A point on a motion boundary will give rise to a positive response for two different displacements, corresponding to the two different motions. If necessary, the local response at that point can help to break the tie.

In our implementation we found that, in order to accumulate E_δ, it worked best to use convolution with a Gaussian for a weighted average rather than just averaging over a rectangular window.

3.1 Finding interesting displacement ranges

It is also possible to accumulate E_δ over very large areas, such as a quarter of the image or even the entire image, to find an initial set of interesting displacements. Most displacements will only align a small subset of features, yielding a negative value for the accumulated E_δ. Only the displacements that align larger parts of the image will yield an above-average response, which can serve to select an initial set of displacements, for which the matching with smaller windows is undertaken. A scale-space approach could be used to speed up the initial selection of interesting displacements.

Peaks in the accumulated E_δ as a function of δ can also serve as attention cues for active vision systems.

4 Experiments

A striking experiment is to just observe E_δ for different displacements δ. As test data we use a stereo pair from the street image sequence[2], depicting a woman crossing a street. This image pair is an interesting example in that it contains large regions with little texture. Also, the absolute intensities are quite different between the two images. To illustrate the power of using maxima in the accumulated measure E_δ as attention cues, we have selected the displacements that yield the strongest response (maximal $\sum E_\delta$) in each quadrant of the image. Figure 3 shows the original image pair and plots of E_δ for the resulting four displacements δ. Gray corresponds to a value of 0, light to positive values, and dark to negative values. Note that these displacements align the dominant features in each quadrant. One can also see that the measure is not sensitive to the brightness difference between the original images.

[2]The street images were provided by Wilfried Enkelmann, Fraunhofer Institut für Informations- und Datenverarbeitung IITB, Karlsruhe, Germany.

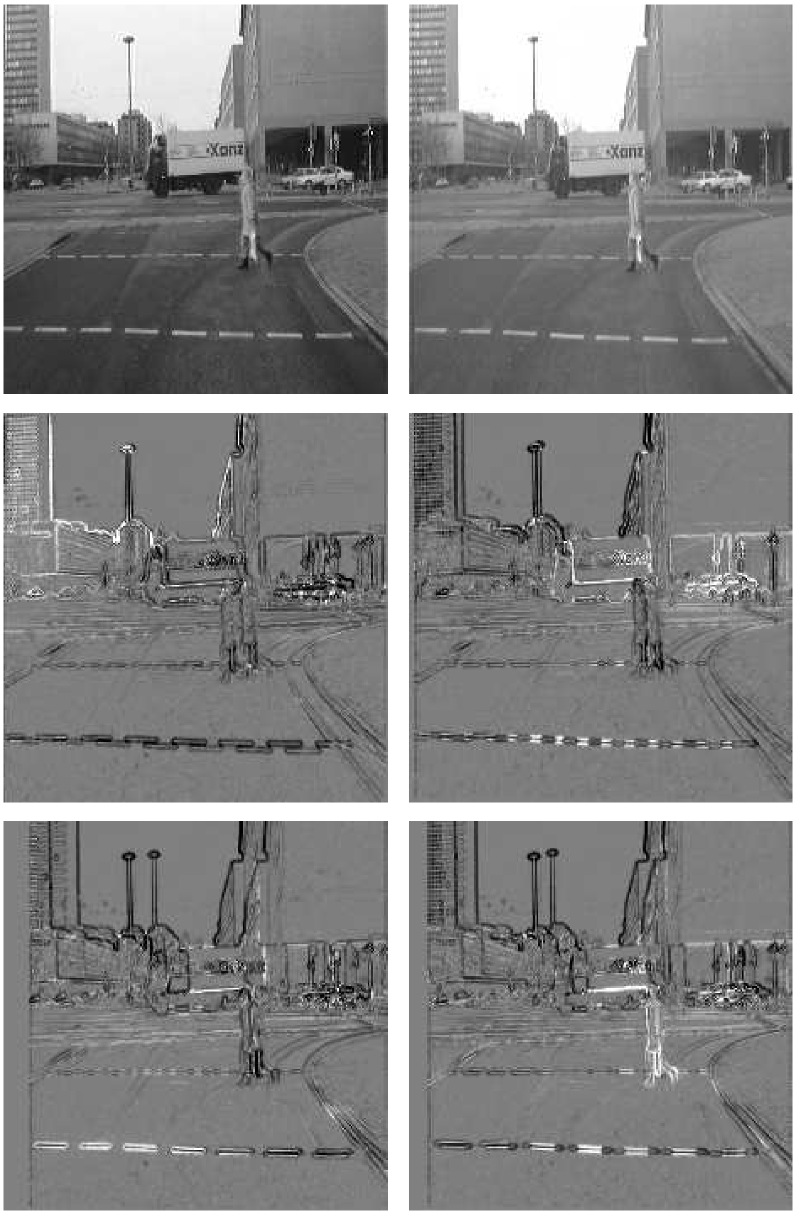

Figure 3: The street image pair and gray level plots of E_δ for the four displacements δ that maximize $\sum E_\delta$ in each of the four quadrants. Gray corresponds to a value of 0, light to positive values, and dark to negative values.

4.1 Stereo

We implemented a simple stereo matcher that uses the evidence measure to select matches. After smoothing the image pair with a Gaussian with $\sigma = .5$, we compute E_δ for a range of different δ. The measure is then accumulated by smoothing each E_δ with a Gaussian with $\sigma = 2$. The disparity at each point is taken to be the displacement that maximizes the accumulated measure at this point.

In the first experiment we ran the matcher on two highly textured images from the Stanford tree sequence[3], depicting an outdoor scene. The considered range of disparities is $\delta_x = 0 \ldots 12$. Simply picking maxima in the accumulated measure already gives surprisingly good results. Figure 4 shows the original image pair and a gray level plot of the computed disparities. Lighter shades of gray correspond to closer points, darker shades correspond to points farther away.

The second experiment shows how confidence can be incorporated in the matcher to be able to deal with images with less texture, where it is harder to find clear maxima in the evidence measure. An advantage of the measure we use is that the *value* of the achieved maximum is related to the gradient magnitude at that point, and thus represents the confidence for the match being correct. To demonstrate this, we will use the street image pair described above. Unreliable matches can be suppressed by setting a threshold for the actual achieved maximum at each point. Figure 5 shows two gray level plots of the computed disparities; in the image at the bottom all unreliable matches are displayed in black. The considered range of disparities is $\delta_x = -3 \ldots 21$. Note that while feature-based matchers try to decide beforehand which locations to match, our method allows the selection of reliable points after the matching process.

4.2 General motion

To test the method on general motion, we used two images from the cat sequence[4]. This sequence depicts a cat walking on a lawn in front of some bushes. We use frames 1 and 5 of this sequence. The camera follows the cat,

[3] The tree images were provided by SRI; we used images number 18 and 24 as right and left images respectively.

[4] The cat images were provided by John Woodfill.

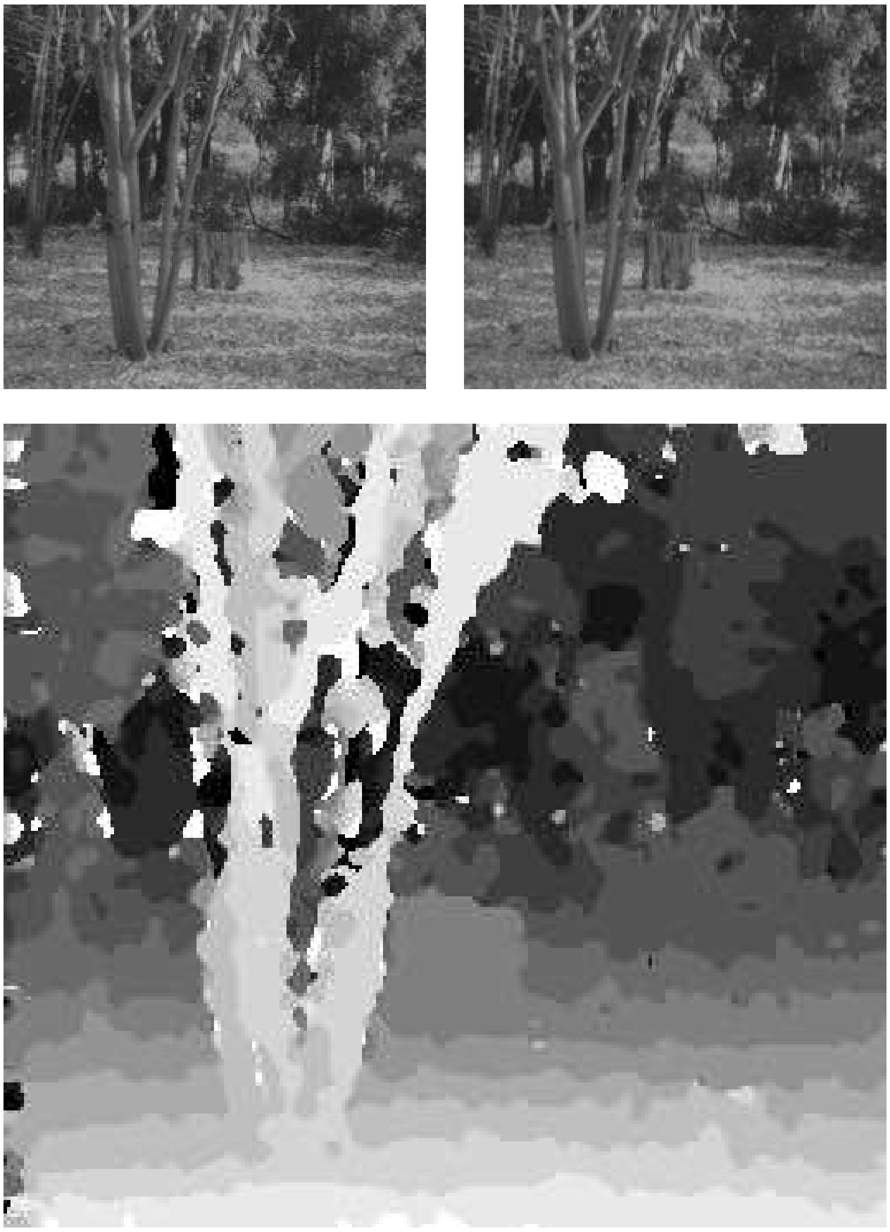

Figure 4: Disparities of the tree image pair. Gray levels correspond to disparities: lighter is closer, darker is farther away. The original image pair is shown at the top.

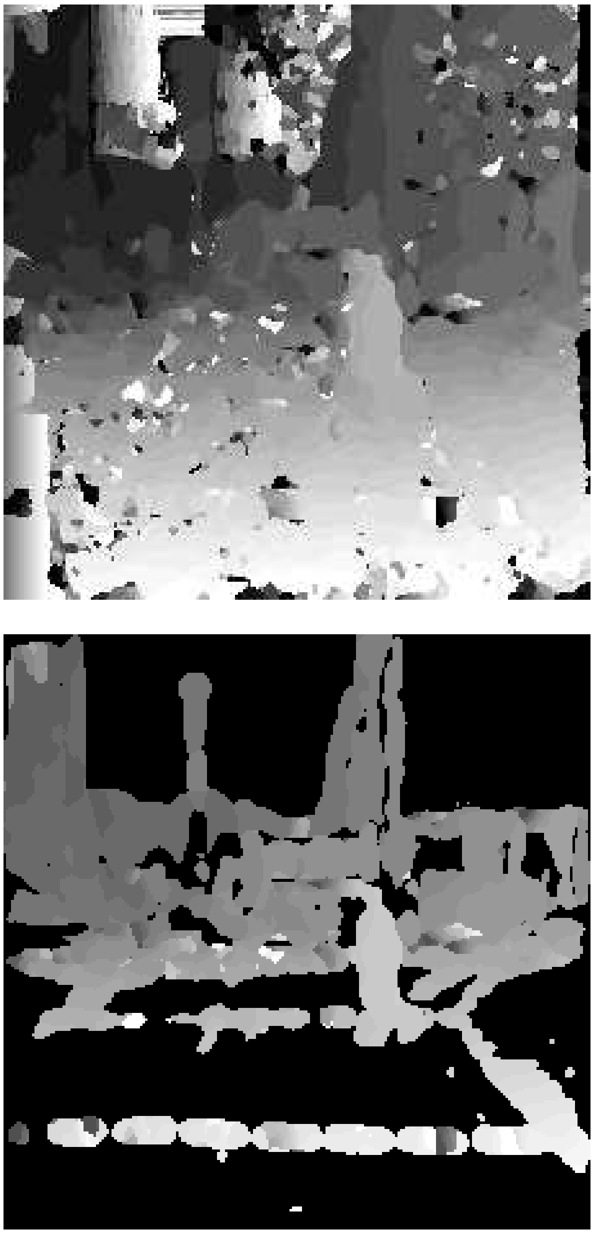

Figure 5: Disparities for the street image pair. Gray levels correspond to disparities: lighter is closer, darker is farther away. In the bottom image, uncertain matches are displayed in black.

so that the visual motion of the cat is almost only caused by its (non-rigid) change of shape, whereas the background moves by more than 10 pixels to the left. Like the tree images, the cat images are well textured, so we don't use the confidence information here. Figure 6 shows the original image pair and the x-components of the displacements that maximize the accumulated measure. The considered ranges are $\delta_x = -15 \ldots 4$, $\delta_y = -2 \ldots 1$; accumulation is done with a Gaussian filter with $\sigma = 2$.

5 Efficiency

As stated before, the computation of E_δ can be carried out very fast, and is easily parallelizable. In this section we outline two possible ways to improve efficiency.

First, we can gain speed by approximating the Euclidean norm L_2, which involves the computation of a square root, by simpler norms, such as the L_1 or the L_∞ norm.

These norms are not rotationally invariant; their relative error with respect to the Euclidean norm depends on the orientation of the vector. One can visualize this by comparing the unit circles of the different norms, which are a diamond, a circle, and a square for L_1, L_2, and L_∞ respectively. A much better approximation to the circle is given by an octagon, and the corresponding norm is only slightly more complicated. It can be expressed as a weighted sum of the L_1 and the L_∞ norm:

$$L_{\mathrm{oct}} = \gamma L_1 + (1-\gamma) L_\infty.$$

A weight $\gamma = \sqrt{2} - 1$ yields a regular octagon with a relative error of only 4% (compared to the Euclidean norm). Using L_{oct} instead of the Euclidean norm in the computation of E_δ can yield a speedup of about 20% (depending on the hardware).

Further speedup can be achieved in the accumulation of E_δ by approximating the Gaussian filter by a sequence of box filtering operations, as is described by Wells [1986].

Experiments have indicated that the matching process is insensitive to these changes, and that the method relies mostly on the qualitative "shape" of the evidence function e.

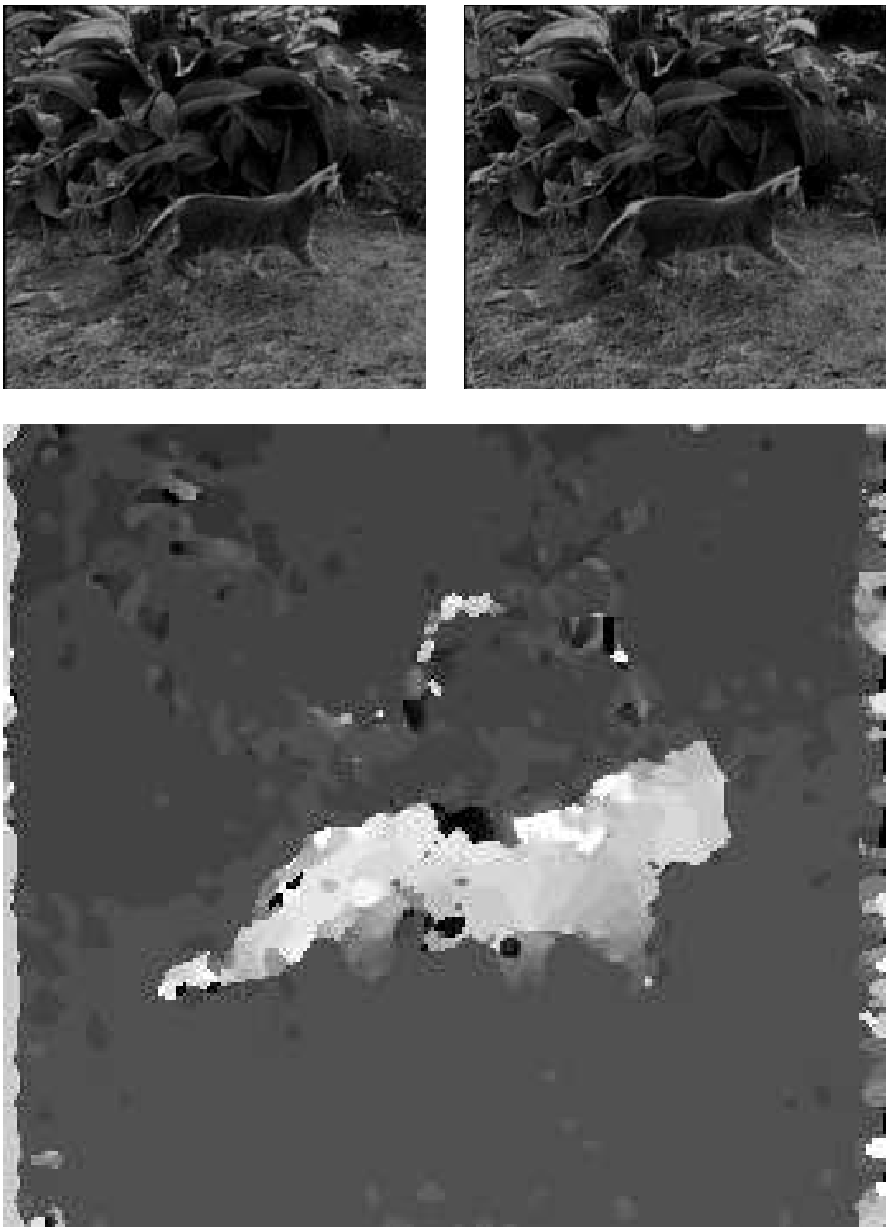

Figure 6: x-components of the best displacements for the cat images. The original image pair is shown at the top.

6 Discussion and Possible Extensions

A problem with the measure discussed here is that partially aligned intensity edges yield a positive response, which can make it hard to find the component of the displacement that is parallel to this edge. For example, in Figure 5 one can observe errors in the computed disparities of the street marks in the foreground of the scene. This is due to the so-called *aperture problem*, which states that, locally, only the component of displacement in the direction of the intensity gradient can be recovered.

A possible way to deal with this problem is to use a measure that is only sensitive to corners. We are currently investigating functions that combine gradients and traditional "cornerness" measures (such as discussed by Deriche and Giraudon [1993]). We are also investigating measures that combine original intensities, first order derivatives (gradients) and second order derivatives (Hessians).

An important observation is that derivatives of all orders can contribute to evidence *against* a match, while evidence *for* a match is harder to capture. That is, different absolute intensities as well as different gradients are indicators for a mismatch, while (purely local) identical intensities are not evidence for a match, and identical gradients only tell us about a match in the direction of the gradient. Other information needs to be taken into account to avoid these *false positives*. For example, in a calibrated stereo system in which the epipolar lines coincide with the scanlines, only the gradient in the x direction should be counted as evidence for a match (but the gradient in the y direction can still tell us about mismatches).

As already mentioned before, an obvious extension to the current method would be to incorporate it into a scale-space approach, and thereby make the evidence measure sensitive to a larger pool of displacements. Multiple scales can also be used in accumulating E_δ, to allow for varying levels of detail of image features.

7 Conclusion

We have presented a simple yet powerful method to perform point-to-point matching between two images. The method uses an *evidence measure* that is based on the gradient fields of the images and that combines the notions of *similarity* between two locations, and *confidence* for a correct match. The computation of the measure is simple and highly parallelizable.

For a given displacement, the measure can be accumulated over a larger area, to collect evidence for or against a match at this location. Using a *displacement-oriented* control strategy that accumulates evidence for a range of different displacements, dominant motions can be detected, which can serve as attention cues in an active vision system.

Finding maxima in the accumulated measure is a stable way of computing correspondences without smoothing across motion boundaries. The method works well both on highly textured images and on images containing regions of uniform intensities, and can be used for a variety of applications, including stereo vision, motion segmentation, object tracking, and active vision.

Acknowledgements

I am grateful to Dan Huttenlocher for his encouragement and many valuable discussions. Thanks also to Amy Briggs, Carlo Tomasi and Nick Trefethen who carefully reviewed earlier drafts of this paper and provided many helpful comments.

References

[Anandan, 1989] P. Anandan. A computational framework and an algorithm for the measurement of visual motion. *International Journal of Computer Vision*, 2(3):283–310, January 1989.

[Barnard and Fischler, 1982] S. T. Barnard and M. A. Fischler. Computational stereo. *ACM Computing Surveys*, 14(4):553–572, December 1982.

[Cochran and Medioni, 1992] S. D. Cochran and G. Medioni. 3-D surface description from binocular stereo. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 14(10):981–994, 1992.

[Coombs and Brown, 1993] D. Coombs and C. Brown. Real-time binocular smooth pursuit. *International Journal of Computer Vision*, 11(2):147–164, October 1993.

[Coombs *et al.*, 1992] D. Coombs, I. Horswill, and P. von Kaenel. Disparity filtering: Proximity detection and segmentation. In *Proc. 11th SPIE Conference on Intelligent Robots and Computer Vision: Algorithms, Techniques, and Active Vision*, pages 195–206, Boston, MA, November 1992.

[Deriche and Giraudon, 1993] R. Deriche and G. Giraudon. A computational approach for corner and vertex detection. *International Journal of Computer Vision*, 10(2):101–124, April 1993.

[Grimson, 1985] W. E. L. Grimson. Computational experiments with a feature based stereo algorithm. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 7(1):17–34, January 1985.

[Horn and Schunck, 1981] B. K. P. Horn and B. G. Schunck. Determining optical flow. *Artificial Intelligence*, 17:185–203, 1981.

[Huttenlocher *et al.*, 1993] D. P. Huttenlocher, J. J. Noh, and W. J. Rucklidge. Tracking non-rigid objects in complex scenes. In *Proc. Fourth International Conference on Computer Vision*, pages 93–101, Berlin, Germany, May 1993.

[Koller *et al.*, 1993] D. Koller, K. Daniilidis, and H.-H. Nagel. Model-based object tracking in monocular image sequences of road traffic scenes. *International Journal of Computer Vision*, 10(3):257–281, June 1993.

[Marr, 1982] D. Marr. *Vision.* W. H. Freeman and Company, New York, 1982.

[Matthies *et al.*, 1989] Larry Matthies, Takeo Kanade, and Richard Szeliski. Kalman filter-based algorithms for estimating depth from image sequences. *International Journal of Computer Vision*, 3(3):209–236, September 1989.

[Moffitt and Mikhail, 1980] F. H. Moffitt and E. M. Mikhail. *Photogrammetry.* Harper and Row, New York, 3rd edition, 1980.

[Olson and Lockwood, 1992] T. J. Olson and R. J. Lockwood. Fixation-based filtering. In *Proc. 11th SPIE Conference on Intelligent Robots and Computer Vision: Algorithms, Techniques, and Active Vision*, pages 685–695, Boston, MA, November 1992.

[Prazdny, 1985] K. Prazdny. Detection of binocular disparities. *Biological Cybernetics*, 52(2):93–99, 1985.

[Scharstein, 1994] D. Scharstein. Matching images by comparing their gradient fields. In *Proc. 12th International Conference on Pattern Recognition*, Jerusalem, Israel, October 1994.

[Seitz, 1989] P. Seitz. Using local orientation information as image primitive for robust object recognition. In *SPIE Visual Communications and Image Processing IV*, volume 1199, pages 1630–1639, 1989.

[Simoncelli *et al.*, 1991] E. P. Simoncelli, E.H. Adelson, and D.J. Heeger. Probability distributions of optical flow. In *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, pages 310–315, Maui, Hawai, June 1991.

[Tian and Huhns, 1986] Q. Tian and M. N. Huhns. Algorithms for subpixel registration. *Computer Vision, Graphics, and Image Processing*, 35:220–233, 1986.

[Wells, 1986] W. M. Wells, III. Efficient synthesis of gaussian filters by cascaded uniform filters. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 8(2):234–239, March 1986.

[Woodfill and Zabih, 1991] J. Woodfill and R. D. Zabih. An algorithm for real-time tracking of non-rigid objects. In *Proc. American Association for Artificial Intelligence Conference*, 1991.

[Zabih and Woodfill, 1994] R. Zabih and J. Woodfill. Non-parametric local transforms for computing visual correspondence. In *Proc. Third European Conference on Computer Vision*, volume II, pages 151–158, May 1994.