

Distinctiveness Maps for Image Matching

Roberto Manduchi
Jet Propulsion Laboratory
California Institute of Technology
Pasadena, CA 91109
manduchi@jpl.nasa.gov

Carlo Tomasi
Computer Science Department
Stanford University
Stanford, CA 94305
tomasi@cs.stanford.edu

Abstract

Stereo correspondence is hard because different image features can look alike. We propose a measure for the ambiguity of image points that allows matching distinctive points first and breaks down the matching task into smaller and separate subproblems. Experiments with an algorithm based on this measure demonstrate the ensuing efficiency and low likelihood of incorrect matches.

1. Introduction

The crux of stereo matching is image ambiguity. If two features in the same image look alike, it may be impossible to find their corresponding features in the other image based only on local appearance, and global reasoning must intervene. For instance, the columns of the colonnade in Canaletto's palazzo ducale in figure 2 are very similar to one another, and knowing which column in one image matches which column in the other may involve counting columns from some distinctive reference point.

Most existing stereo algorithms discover ambiguity *post facto*. They first generate match candidates based on local similarity measures, and label features with multiple matches as ambiguous. A second stage then attempts to resolve ambiguous cases by imposing global consistency constraints. This generate-and-test approach works hard to produce a large number of match candidates, and then works harder to eliminate the bad ones.

In contrast, we propose a measure of image distinctiveness that allows sorting image points in order of increasing ambiguity before matching begins. If distinctive, that is, low-ambiguity points are matched first, the correspondence problem is broken down into a number of smaller ones. If the left-to-right ordering of features is preserved across images, a safe assumption in most cases, then each subproblem is restricted to pairs of corresponding epipolar line seg-

ments that lie between two of the given matches. Thus, the "safe" matches constrain the less safe ones, resulting in both fewer incorrect matches and a greater efficiency.

It is important to distinguish distinctiveness (or its opposite, ambiguity) from what is called "interest" in the computer vision literature [2, 6]. Interest operators are local, and detect image points that have sufficient texture for matching. Very interesting points can be highly ambiguous, like the edges in a periodic pattern. On the converse, distinctive points are not necessarily rich in texture. Interesting points ensure good match *accuracy*; distinctive points ensure low probability of *mismatch*. If the correspondence problem is formalized in terms of the minimization of a cost function, inaccuracy is equivalent to poor localization of the global minimum; mismatch is equivalent to choosing a wrong local minimum. Perhaps due to the difficulty of finding an adequate model, the analysis of mismatches has received much less attention in the literature than the analysis of match accuracy.

Loosely speaking, the ambiguity of a point is characterized by the difference in appearance from the most similar other point on the same epipolar line. Thus, an image location is ambiguous if there is some other location that looks similar to it. While ambiguity is measured in a single image, it is used for matching stereo pairs. It therefore stands to reason that the similarity metric used for measuring ambiguity should be the same as the one used for stereo matching. We give a precise definition of distinctiveness, the opposite of ambiguity, in section 2.

Distinctiveness maps may be used to speed up stereo algorithms by means of a hierarchical scheme. If the most distinctive points in an epipolar line are matched first, then the segments of epipolar line lying between two consecutive distinctive points may be matched independently by virtue of the ordering principle [3]. A fast divide-and-conquer strategy based on such observation is presented in section 3. Section 4 has the conclusions.

2. Distinctiveness maps

The yellow and blue plumage of a toucan produces a visually distinctive blotch amidst the green of a jungle. In a stereo image pair of this jungle scene, the toucan is trivial to match. If we consider an epipolar line cutting through the bird’s plumage, determining correspondences for the remaining pixels is probably hard, since everything is green and most leaves look the same. This example shows why some image locations are easy to match, while others are not. Distinctive features are unique, and look like nothing else in the picture, or at least along the epipolar line. Ambiguous points, on the other hand, are similar to many others. Their local appearance is inadequate for determining stereo correspondence.

At the same time, the distinctive features help matching the ambiguous ones as well. In fact, foliage that is on the left of the toucan in the left image matches foliage that is on the left of the bird also in the right image. This is the *ordering constraint* [3], which is violated only in rare cases like with a thin pole in the foreground, well away from the background. Barring these extreme cases, the ordering constraint can be used to leverage distinctive features in order to facilitate the establishment of correspondence for the more ambiguous image locations. In fact, before matching the toucan, every pixel in the left epipolar line can in principle be a match candidate for every pixel in the right one. After matching the toucan, on the other hand, the correspondence problem is broken into two smaller ones: one is for the two segments of epipolar line to the left of the toucan, the other is for the two segments to its right. Candidate matches that take pixels from both sides of the toucan are disallowed. In short, if distinctive features can be matched first, divide-and-conquer can be applied to stereo matching.

The distinctiveness of a point is not an absolute measure, but is subordinate to the chosen matching strategy. Section 2.1 defines the basic parameters of interest for stereo algorithms. Section 2.2 introduces our definition of distinctiveness, and section 2.3 presents some examples of distinctiveness maps for the case of correlation-based matching.

2.1. Basic parameters

Stereo algorithms can be roughly characterized in terms of the following parameters:

1. *The local descriptors*, which are vectors that encode the local profile of the image. More precisely, the local descriptor of the image at point x is a vectorial transformation of the brightness values within an analysis window W_a centered in x . Ideally, descriptors are invariant with respect to the geometric transformations of interest.

2. *The perceptual metric*, which measures the similarity of image points by the distance of the corresponding descriptors.
3. *The search window W_s* , which determines the largest disparity that can be measured by the algorithm. A large baseline requires a large search window, which implies high computational cost and high probability of mismatches.

For example, SSD-based correlation algorithms measure the euclidean distance between local descriptors formed directly by the values of the pixels within the analysis window. Filter-based algorithms [5][4][8] generalize the correlation idea, and represent local brightness profiles by means of vectors formed by the output of a bank of filters. Kass [5] and Jones and Malik [4] use banks of multiscale/multioriented filters, and use L_2 or L_1 perceptual metric. Tomasi and Manduchi [8] measure the local Taylor expansion of the brightness and use an ad-hoc perceptual metric for the fast and robust computation of nearest neighbors in the descriptors’ space.

2.2. Distinctiveness: a formal definition

Two points in two different images are similar when their perceptual distance is small. The same concept applies to two points of the same image, suggesting the following definition of distinctiveness:

Definition 1. (*distinctiveness in the discrete case*)

The distinctiveness of an image point x is equal to its perceptual distance to the most similar other point in the search window.

We may also define the *ambiguity* of a point as the inverse of its distinctiveness. If within the search window there is another point which looks exactly like x , then x is infinitely ambiguous: the risk of mismatch for such a point is very high.

This simple definition of distinctiveness must be modified for the continuous case, where the notion of “most similar other point” may not make sense. Let $d_x(s)$ be the perceptual distance between points x and $x + s$. Consider the set of maximally connected regions of W_s formed by the points where the gradient of $d_x(s)$ vanishes. We pick any one point from each such region, excluding the one containing the origin, to form the set of “characteristic local extrema” M .

Definition 2. (*distinctiveness in the continuous case*) The distinctiveness of the image point x is defined as

defined as

$$D(x) = \begin{cases} \min_{s \in M} d_x(s) & \text{if } M \text{ is not empty} \\ 0 & \text{if } M \text{ is empty} \end{cases} \quad (1)$$

Note that even points in segments of constant brightness (such as stripes or blobs) may be distinctive, as long as they are structurally different from the background. Such points are not considered interesting by standard local feature operators; in fact, they can be precious “anchor points” for *reliable* (albeit not necessarily *accurate*) matches.

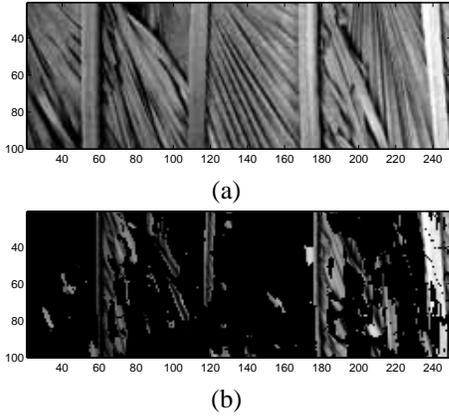


Figure 1. (a) Original image. (b) SSD-based distinctiveness map ($W_a=5 \times 5$ pixels; $W_s=21 \times 21$ pixels).

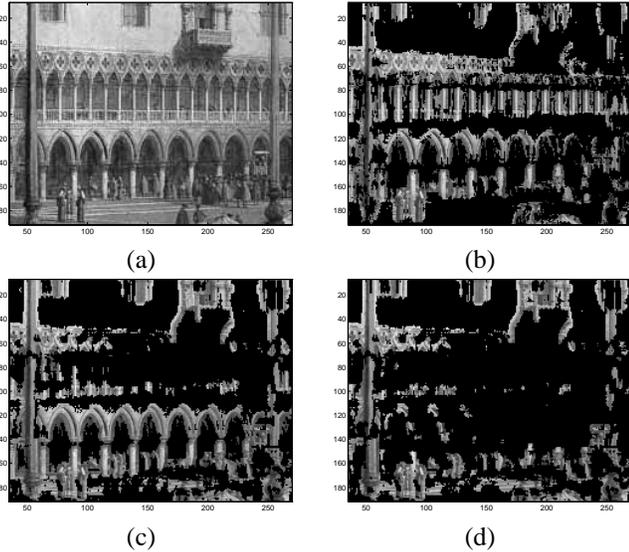


Figure 2. (a) Original image. (b),(c),(d) SSD-based distinctiveness map ($W_a=7 \times 7$ pixels; (b) $W_s=1 \times 21$ pixels; (c) $W_s=1 \times 41$ pixels; (d) $W_s=1 \times 61$ pixels).

2.3. An example: SSD based matching

SSD-based matching techniques are very popular for stereo matching. The surface $d_x(s)$, which measures the perceptual difference between x and the points within the search window, corresponds to the auto-SSD function

$$SSD_x(s) = \sum_{\bar{x} \in W_a} (l(x + \bar{x}) - l(x + \bar{x} + s))^2 \quad (2)$$

The auto-SSD profile around a point x contains precious information about the expected goodness of match. For example, its flatness in correspondence of the origin measures the expected match accuracy. On the other side, the risk of mismatch can be estimated from the distinctiveness of x , which is equal to the height of the smallest minimum of $SSD_x(s)$ beside the one in the origin.

We have applied the SSD-based distinctiveness operator to the image of figure 1(a) with an analysis window W_a of 5×5 pixels and a search window W_s of 21×21 pixels. In this image, the vertical stripes stand out distinctly, while the oblique edges form a periodic pattern, more prone to mismatch. In figure 1(b) we show the image points with distinctiveness above the average. The measured distinctiveness map agrees with our expectations: only the vertical stripes and the most outstanding oblique patterns are ranked distinctive.

As pointed out earlier, our definition of distinctiveness is subordinate to the choice of a particular matching system. Most stereo algorithms match epipolar lines, which is equivalent to constraining the search window height to just one pixel. We have computed the distinctiveness map of the Canaletto image in figure 2(a) using one-dimensional search windows. Figures 2 (b), (c) and (d) show the points with distinctiveness above the average for search windows of 1×21 , 1×41 and 1×61 pixels respectively (an analysis window of 7×7 pixel has been used in all three experiments.)

It is interesting to analyze the results in correspondence of the periodic patterns formed by the colonnades. Using the 1×21 search window, the columns in the upper row are ranked distinctive. However, since their repetition period is smaller than 41 pixels, they become ambiguous when the larger windows are used. A stereo algorithm with a search window of 41 pixels or more would be prone to mismatch these points. When the largest search window is used, also the wider columns in the lower row are ranked ambiguous. However, the flagpole on the left, as well as the window high above, are ranked distinctive in all three cases.

3. An application: hierarchical stereo

The selection of image features is at the basis of a number of classical stereo algorithms [1],[7]. After feature ex-

traction, matches are computed in a sequence of two steps.

First, each feature point in one image of the stereo pair is assigned a number of “candidate matches” in the other image. The selection is done according to criteria of spatial proximity and perceptual closeness. In other words, the candidate matches are the most similar features that lie inside the search window in the other image.

Second, the chain of correspondences which maximizes a global quality measure while satisfying the criteria of uniqueness, ordering and smoothness is selected from the pool of candidates. The global quality measure is a function of the perceptual distances of the candidate matches in the chain. This “disambiguation” task can be computationally very expensive, even if dynamic programming implementations are used [1][8].

In order to reduce the computational load of the process, we propose a hierarchical scheme that matches a set of distinctive points first. Once these highly distinctive points have been matched, the process is divided into a number of smaller subprocesses, by virtue to the ordering constraint. In other words, given any two consecutive distinctive points x_i, x_{i+1} , and their corresponding matches y_i, y_{i+1} found in the other image, the points in the segment $[x_i, x_{i+1}]$ are matched only with points in the segment $[y_i, y_{i+1}]$. It can be easily proved that this divide-and-conquer strategy effectively reduces the overall computational load.

Our hierarchical scheme can be implemented on top of almost any existing stereo algorithm. For our experiments we have adopted the algorithm of Tomasi and Manduchi [8], which uses the intrinsic curve representation of scanlines to determine candidate matches. An intrinsic curve is the path formed by the descriptor as one moves along the scanline, and therefore is invariant to image shift. Finding a candidate match then becomes a nearest neighborhood problem in the descriptors’ space, which can be solved efficiently by means of a suitable representation of the curves. For the same reason, finding the distinctiveness of a point (which corresponds to finding the “nearest neighbors” in the same curve) is a very fast operation.

We have tested the hierarchical stereo algorithm on two stereo pairs, the “Clorox” pair from Stanford University (figure 3) and the “Castle” pair from CMU (figure 4) (the images have been preciously subsampled by two along the horizontal and vertical axes.) The stereo pair “Clorox” is characterized by a very articulated depth field, with occlusions at the borders of the objects. The pair “Castle” shows patches with periodically repeated patterns.

The computed disparity maps are represented with pseudocolors.¹ The upper part of each image in the figures (above the epipolar scanline drawn in black) represents the left image in the stereo pair, the lower part is the right im-

age. No postprocessing has been performed on the computed disparities. The algorithm uses an adaptive resampling strategy that concentrates matches where the signal “business” is high [8]. This is the main reason for the sparseness of the computed disparity values, another reason being that a match is accepted only when its quality is above a certain threshold.

Beside being computationally efficient, the least-ambiguous-first technique reduces the risk of mismatches. This is shown here by way of examples in correspondence of the image patches highlighted in the figures. The original full-rate sampling period has been retained in these figures. In the case of figure 3, the periodic pattern corresponding to the keys of the calculator is a potential occasion for mismatch. This appears clearly by plotting the intensity in the two scanlines, as in the first plot of the figure (the solid line corresponds to the left image, the dashed line to the right image). The second plot shows the normalized distinctiveness of the points in the left scanline. It is rather difficult to correctly match by hand the peaks of the two signals, unless one uses the dark calculator’s edge as a reference point. As expected, this is where the distinctiveness has its maximum.

For the piece of scanline shown in the figure, we first selected and matched five highly distinctive features, obtaining the correct disparity estimates depicted with dashed lines in the third plot. Then, the scanline segments between the correspondences found in this first stage have been matched independently, producing the disparity values depicted with solid lines in the plot. No mismatch occurred with this procedure. As a counterexample, we repeated the experiment and selected five highly ambiguous features in the first stage; the results are shown in the fourth plot. Because of ambiguity, such features have been mismatched, constraining the subsequent stage to detect wrong matches.

A similar study case is shown in figure 4 for the “Castle” couple. Here, the textured pattern on the houses’ facades is interrupted by a flat white area, corresponding to the houses’ roofs. The distinctiveness map reveals such a region, and the match is best performed starting from these more distinctive points.

4 Conclusions

We have shown in this paper that distinctiveness, and not interest, is the appropriate criterion for feature selection in stereo matching. Distinctiveness is global, and subsumes the local notion of interest. Distinctive points are conceptually similar to outliers in a statistical model; they are the features that stand out most clearly in the image, and represent reliable “anchor points” for matching. Based on this intuition, we have proposed an hierarchical stereo algorithms that matches distinctive points first. This early commitment strategy can reduce the computational load effectively, at

¹Color versions of the figures contained in this paper can be found at <http://vision.stanford.edu/public/publication/tomasi/tomasiciap99.pdf>.

the same time minimizing the probability of mismatches.

5. Acknowledgments

The work described in this paper was performed while Roberto Manduchi was at Stanford University, sponsored by a fellowship from the University of Padova, Italy. Publication support was provided by the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise, does not constitute or imply its endorsement by the United States Government or the Jet Propulsion Laboratory, California Institute of Technology.

Carlo Tomasi was supported by NSF grant IRI-9506064 and DoD grant DAAH04-96-1-0007.

References

- [1] H. Baker and T. Binford. Depth from edge and intensity based stereo. In *Proc. of the Seventh Int. Joint Conf. on Artif. Intell.*, pages 631–636, 1981.
- [2] S. Barnard and B. Thompson. Disparity analysis of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2(4):333–340, July 1980.
- [3] O. Faugeras. *Three-Dimensional Computer Vision — A Geometric Viewpoint*. MIT Press, Cambridge, Massachusetts, 1993.
- [4] D. Jones and J. Malik. A computational framework for determining stereo correspondence from a set of linear spatial filters. In *Proc. of EECV'92*, pages 395–410, Santa Margherita Ligure, Italy, 1992.
- [5] M. Kass. Computing stereo correspondence. Master's thesis, M.I.T., 1984.
- [6] H. Moravec. Towards automatic visual obstacle avoidance. In *Proc. 5th Int. Joint Conf. on Artificial Intell.*, pages 661–669, Cambridge, Massachusetts, 1977.
- [7] S. Pollard, J. Mayhew, and G. Frisby. PMF: A stereo correspondence algorithm using a disparity gradient limit. *Perception*, 14:449–470, 1985.
- [8] C. Tomasi and R. Manduchi. Stereo matching as a nearest neighbor problem. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(3):333–340, Apr. 1998.

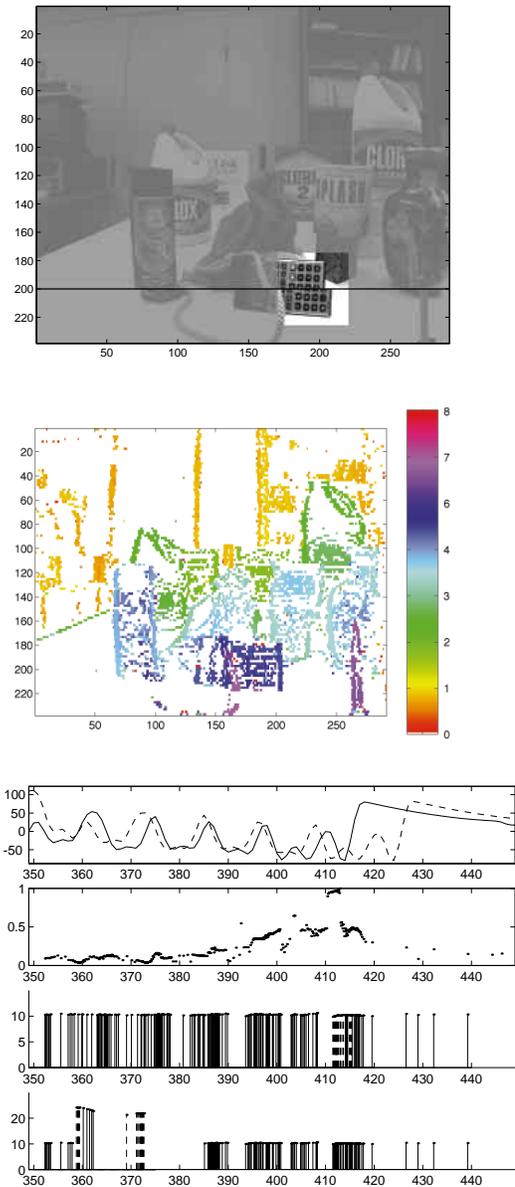


Figure 3. Matching experiments with the test stereo pair “Clorox”. The images have been previously subsampled by two along the horizontal and vertical axes. The left image is shown above the dark scanline, the right image is shown below. The computed disparity field is represented with pseudocolors. The first plot shows the full-rate intensity profile in the scanline corresponding to the highlighted area (solid line: left image, dashed line: right image.) The second plot shows the normalized distinctiveness function relative to the scanline in the left image. The third plot shows the disparity estimates obtained with the hierarchical stereo algorithm; the values computed in the first stage are depicted with dashed line. The fourth plot shows the results in the case the most ambiguous points are matched first.

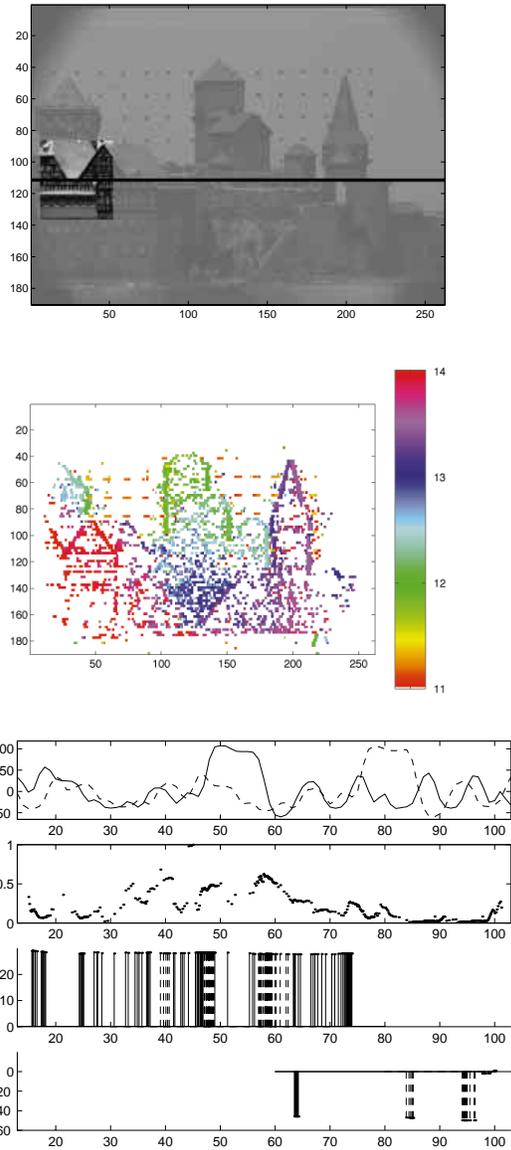


Figure 4. Matching experiments with the test stereo pair "Castle" (see caption of figure 3.)