Ordinal Measures for Image Correspondence

Dinkar N. Bhat and Shree K. Nayar

Abstract—We present ordinal measures of association for image correspondence in the context of stereo. Linear correspondence measures like correlation and the sum of squared difference between intensity distributions are known to be fragile. Ordinal measures which are based on relative ordering of intensity values in windows—rank permutations—have demonstrable robustness. By using distance metrics between two rank permutations, ordinal measures are defined. These measures are independent of absolute intensity scale and invariant to monotone transformations of intensity values like gamma variation between images. We have developed simple algorithms for their efficient implementation. Experiments suggest the superiority of ordinal measures over existing techniques under nonideal conditions. These measures serve as a general tool for image matching that are applicable to other vision problems such as motion estimation and texture-based image retrieval.

Index Terms—Image matching, stereo, ordinal measures, correlation, correspondence.

1 Introduction

STEREO systems for depth estimation work reasonably well with smooth surfaces that are mostly Lambertian in reflectance. However, many surfaces in real scenes exhibit sharp discontinuities with non-Lambertian reflectance. It is a challenge for practical systems to produce accurate depth maps in such settings. In this paper, we present ordinal measures for image correspondence, in the context of stereo matching.

The heart of any window-based method for stereo matching lies in the underlying similarity criterion that determines optimal statistical correlation between windows around corresponding points. The basic assumption used is that these windows represent the same location in the scene and have identical intensity distributions. However, the assumption is violated due a number of physical phenomena whence intensity data in windows around corresponding points can be inconsistent. When the compared windows straddle a depth discontinuity, they represent projections of different surface regions. The same is the result of projective distortion caused by varying viewpoint in acquiring the images. The presence of depth discontinuities also causes occlusion due to which scene points are visible in only one of the two images, and they must correctly identified. Additionally, in the presence of noise, specular reflection, and possibly varying camera parameters, intensities at corresponding points may not be identical or even linearly related. A larger window is not a cure-all, since it can result in a greater number of false positives in occlusion zones and increased smoothing of disparity across discontinuities, although the number of false negatives due to noise and outliers may decrease. Different approaches have been developed to tackle individual issues (for example, [9], [10]) within the framework of linear correlation measures. For instance, Quam [10] addresses the

- *D.N. Bhat is with LG Electronics Research Center of America, 40 Washington Road, Princeton, NJ 08550. E-mail: dbhat@lgerca.com.*
- *S.K. Nayar is with the Department of Computer Science, Columbia University, New York, NY 10027. E-mail: nayar@cs.columbia.edu.*

Manuscript received 12 Mar. 1996; revised 24 Feb. 1998. Recommended for acceptance by G. Medioni.
For information on obtaining reprints of this article, please send e-mail to: tpami@computer.org, and reference IEEECS Log Number 106435.

problem of projective distortion in a hierarchical correlation-based framework wherein the disparity estimates from the prior resolution level of matching are used to geometrically warp the second image towards the first at the current resolution level. Here, the problems are addressed from a different basis, namely, the use of an ordinal scale for intensity.

We introduce *ordinal* measures of association [6], [7] that are robust to a high degree. An ordinal variable implies one drawn from a discrete ordered set like the grades in school. The ratio between two measurements is not of consequence; only their relative ordering is relevant. The relative ordering between measurements is expressed by their *ranks*. A rank permutation is obtained by sorting the sample in ascending order and labeling them using integers $[1, 2, \ldots, n]$, n being the size of the sample. In our application, intensity is viewed as an ordinal variable. Consequently, ordinal correlation measures are based on rank permutations within windows rather than absolute intensity data. Two well-known ordinal measures include the Kendall's τ and the Spearman's ρ [7]. Both coefficients are relatively unaffected by the presence of random data outliers, like noise, in comparison to direct image correlation. However, if the ranks within each window are distorted, like in the presence of projective distortion, they are not satisfactory. By rank distortion, we mean rank permutations that bear strong structural relationship, but there can be significant difference between some corresponding elements; for instance, a cyclic shift of one permutation with respect to another. Simple computationally economical algorithms are presented to evaluate the measures. Experiments with real images and comparison with existing matching methods suggest the suitability of ordinal measures for applications.

2 Correlation-Based Image Matching

In a correlation-based framework, correspondence for a pixel in a reference image is obtained by searching in a predefined region of the second image. Correlation could be performed in a hierarchical framework or at one single resolution. Most currently used stereo methods belong to the category of linear correlation methods, which include those based on the sum of squared differences (SSD) and cross correlation. Let I_1 and I_2 represent intensities in two windows, i.e., there exist n tuples $\left(I_1^1, I_2^1\right), \ldots, \left(I_1^n, I_2^n\right)$, n depending on the size of the window used. The quantity $SSD = \sum_{i=1}^{n}\left(I_1^i - I_2^i\right)^2$ measures the squared Euclidean distance between (I_1, I_2), and a value close to zero indicates a strong match. The *normalized correlation coefficient* (NCC) is given by:

$$NCC = \frac{\sum_{i=1}^{n}\left(I_1^i - \bar{I}_1\right)\left(I_2^i - \bar{I}_2\right)}{\sqrt{\sum_{i=1}^{n}\left(I_1^i - \bar{I}_1\right)^2}\sqrt{\sum_{i=1}^{n}\left(I_2^i - \bar{I}_2\right)^2}},$$

where $\bar{I}_1$ and $\bar{I}_2$ represent the corresponding sample means. Like SSD, it appraises the degree of linearity between the samples being compared. It is closely related to the least-squares line of fit [4], hence NCC has similar properties to SSD. The absolute value of NCC lies between zero and one, and a value of one indicates perfect matching windows. While NCC is preferable since it is invariant to linear brightness and contrast variations between perfect matching windows, SSD is computationally more attractive. However, both measures are nonrobust in that a single outlying pixel can distort them arbitrarily. Further, by definition, they are not suitable in the presence of nonlinear intensity variation at corresponding pixels.

Since the Euclidean norm (SSD) is sensitive to outliers, robust M-estimators have been used that are more resilient (for instance, the Lorentzian in motion estimation [3]). The characteristic of robust estimators is that they cause outliers to contribute less weight compared to inliers. Most M-estimators include a parameter that needs to be set beforehand: the point at which measurements must be considered outliers. The threshold can vary depending on image contrast and noise level, and setting it in a correlation-based framework is often done empirically. It is desirable to have a universal measure of correlation independent of absolute intensity scale and experimental conditions.

Image transform methods are based on comparing stereo images transformed using local window measures. Kories and Zimmerman's [8] *monotonicity operator*, Zabih and Woodfill [11] *rank transform* and *census transform* fall in this category. The rank transform defines rank (of the center pixel P in a window W) as:

$$R(P, W) = \bullet P' \in W \mid I(P') < I(P) \bullet$$

where $\bullet.\bullet$ refers to cardinality. In the case of the rank transform method, the stereo images are fully transformed using the above operator and the resulting mappings are compared using linear correspondence measures. The advantage of the above schemes is that correlation of transformed images is not dependent on absolute gray values, and hence relatively insensitive to data outliers. However, one drawback of the approaches is that they depend quite heavily on the center pixel. Other methods have been developed to match local intensity gradients instead of raw intensity values. However, their performance can be poor when gradient information is not reliable. Finally, an entire class of methods has been developed in frequency domain where the issue to estimate the change in phase between Fourier or similar representations of the original intensity signal. The implicit assumption regarding surfaces being smooth and Lambertian remains. These techniques are not considered since they belong to a different mould from correlation-based methods.

3 Ordinal Measures

In this section, ordinal measures of association are presented after a brief review of the concept of correlation based on distance metrics. Sensitivity of the measures with respect to outliers and rank distortion is discussed and compared to other correlation methods.

3.1 Motivation

To illustrate the robustness of ordinal measures, consider the following example of a 3×3 reference window R with intensity I_1:

	R	
10	30	70
20	50	80
40	60	100

Under ideal conditions, the corresponding window S with intensity I_2 is identical, and so are their rank matrices:

	R			S	
1	3	7	1	3	7
2	5	8	2	5	8
4	6	9	4	6	9

Recall, that an ordinal measure of association is based on ranks rather than intensity values themselves. Let us modify one pixel A in S, say the one with intensity value 100, through a range of different values between $[0, 255]$. This simulates the effect of a random outlier. Clearly, in the range $(80, 255]$, ranks of the intensity values in S are not modified, and hence any ordinal measure of

correlation remains at one. This is unlike the linear correlation coefficient which can *substantially deviate*. For example, when the pixel takes a value of 255, *NCC* = 0.645. This attractive property of ordinal measures motivates us to apply them for stereo matching. The same robustness can also be observed of local transform techniques like that of Zabih and Woodfill [11]. The concepts underlying ordinal measures are now introduced using distance metrics.

3.2 Review

A ranking which represents the relative ordering between values of an ordinal variable is simply a permutation of integers. More precisely, if S_n denotes the set of all permutations of integers $[1, 2, \ldots, n]$, then any ranking is an element of this set. To define correlation between two rankings π_1, π_2, we require a measure of closeness—a *distance metric*—between them. Once a distance metric $d(\pi_1, \pi_2)$ is defined, a coefficient of correlation α can be obtained as:

$$\alpha = 1 - \frac{2d(\pi_1, \pi_2)}{M} \tag{1}$$

where M is the maximum value of $d(\pi_1, \pi_2)$, $\forall(\pi_1, \pi_2) \in S_n$. α lies in the range $[-1, 1]$. M is attained when the two permutations are reverses of each other, and hence $\alpha = -1$. Different distance metrics are possible, an example being the *Hamming distance* d_h:

$$d_h(\pi_1, \pi_2) = \sum_i \left(\left| \mathrm{sgn}\left(\pi_1^i - \pi_2^i\right) \right| \right) \tag{2}$$

where $sgn(x) = x/|x|$ if $x \neq 0$, and 0 if $x = 0$. For the Hamming distance, $M = n$. The Kendall's τ and the Spearman's ρ can also be expressed using distance metrics (see [2]). Kendall's τ computes the number of discordant pairs between samples, and the Spearman's ρ estimates the Euclidean distance between permutations.

Data inconsistency can occur between corresponding windows due to the presence of specular reflection and discontinuities. This could result in corresponding rank matrices being distorted unlike the example discussed in Section 3.1. As a result, ordinal measures like the Kendall's τ and Spearman's ρ are inadequate. Therefore, measures of association are defined that are unaffected by maverick data items but capture the overall sense of correlation between permutations.

3.3 Ordinal Measures

Let I_1 be a window in one image, and I_2 be a window in the next image of the sequence. For the set of window intensity values $\left(I_1^i, I_2^i\right)_{i=1}^n$, let π_1^i be the rank of I_1^i among the I_1 data, and π_2^i be the rank of I_2^i among the I_2 data. Below, we present a method for defining the distance between rank permutations along the lines reported by [6]. However, our measure is motivated by the definition of the Kolmogorov-Smirnov test statistic (see [2]). Here we assume that there are no ties in the data. The method to handle tied values is discussed in [2]. A composition permutation s can be defined as follows:

$$s^i = \pi_2^k, \quad k = \left(\pi_1^{-1}\right)^i \tag{3}$$

where π_1^{-1} denotes the inverse permutation of π_1. The inverse permutation is defined as follows: If $\pi_1^i = j$, then $\left(\pi_1^{-1}\right)^j = i$. Informally, s is the ranking of I_2 with respect to that of I_1. Under perfect positive correlation, s should be identical to the *identity permutation* given by $u = (1, 2, \ldots, n)$. By defining a distance measure between s and u, a notion of distance is obtained between π_1 and π_2. The deviation d_m^i for $i = 1, \ldots, n$ is defined as:

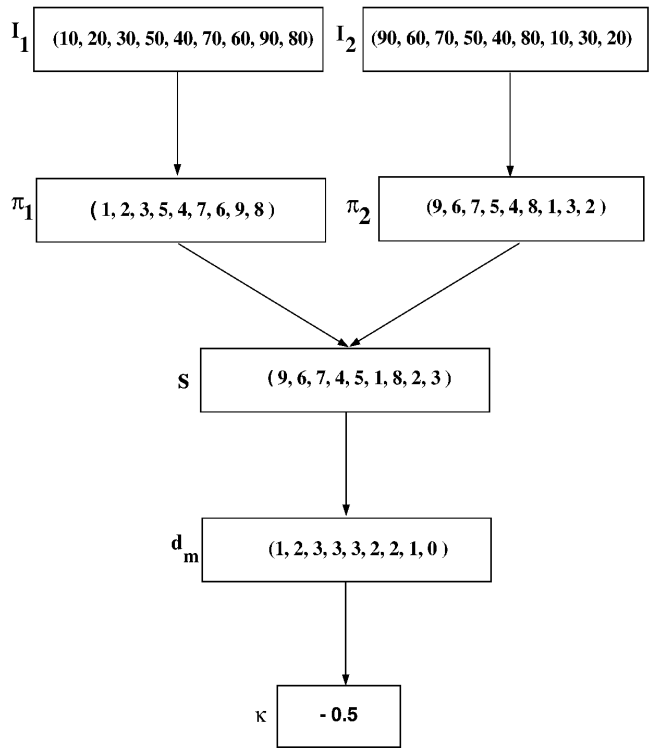


Fig. 1. Example illustrating the procedure for computing κ.

$$d_m^i = i - \sum_{j=1}^{i} J\left(s^j \leq i\right) \tag{4}$$

where $J(B)$ is an indicator function of event B, i.e., $J(B) = 1$ when B is true, and $B = 0$ otherwise. The vector of d_m^i values is termed as the distance vector $\mathbf{d_m}(s, u)$. Each component of the distance vector, referenced by its positional index, estimates the number of predecessing elements in s that are out of position. If (I_1, I_2) were perfectly correlated, then $\mathbf{d_m}(s, u) = (0, 0, \ldots, 0)$. The maximum value that any component of the distance vector can take is $\left\lfloor \frac{n}{2} \right\rfloor$ which must occur in the case of perfect negative correlation (see [2]). Now, a measure of correlation $\kappa(I_1, I_2)$ is defined as:

$$\kappa(I_1, I_2) = 1 - \frac{2 \max_{i=1}^{n} d_m^i}{\left\lfloor \frac{n}{2} \right\rfloor} \tag{5}$$

Unlike [6], we do not consider the deviation from the negative identity permutation since that is computationally more expensive. If I_1 and I_2 are perfectly correlated ($s = u$), then $\kappa = 1$. It falls to -1 when (I_1, I_2) are perfectly negatively correlated. Fig. 1 describes a simple example which illustrates the procedure for computing κ. κ has the following desirable properties of a correlation coefficient:

- It is independent of linear scaling and shift between I_1 and I_2 since π_1 and π_2 remain unchanged. This implies independence from camera gain and bias.
- It is symmetrical, i.e., $\kappa(I_1, I_2) = \kappa(I_2, I_1)$. Hence, either image can be used as reference (see [2]).
- $\kappa(f(I_1), h(I_2)) = \kappa(I_1, I_2)$, where f and h are monotonically increasing functions.

  To illustrate this property, consider the case when different cameras are used for stereo and have different responses to image irradiance. Each sensor output I is related to image irradiance E as:

$$I = gE^{\frac{1}{\gamma}} + m$$

where g is the camera gain, m is the reference bias factor, and γ accounts for image contrast. Since the gain and bias account for linear variations which, as noted earlier, do not affect κ, let us assume there is only gamma variation between the sensors. In other words, let the gains of the sensors be identically 1.0 and the bias factors be zero. Further, let the imaged surface be Lambertian, i.e., the image irradiance from any point is identical for both sensors. The sensor outputs are related by the equation $I_1^i = \left(I_2^i\right)^t$, $\forall i$ where $t = \frac{\gamma_2}{\gamma_1}$. In general, $t \neq 1$, and hence the linearity between the sensor outputs is lost. If $t > 1$, then $I_2^i < I_1^i$, $\forall i$, and if $t < 1$, then $I_2^i > I_1^i$, $\forall i$. However, κ remains at one, because π_2 remains the same as π_1. The usefulness of this property for motion estimation in tagged Magnetic Resonance Images (MRI) was shown elsewhere (see [2]).

- The distance vector is not affected by arbitrary relabeling of data items. Equivalently, the distance measure is right-invariant, i.e.:

$$d_m(s(\pi_1\tau, \pi_2\tau), u) = d_m(s, u)$$

where $\tau \in S_n$ is an arbitrary relabeling of the data items. The reason for right-invariance is that the s permutation is not affected by the relabeling. This property is useful when the method of ranking the data samples may change, for instance, one can choose the descending order for generating permutations.

Another measure of correlation $\chi(I_1, I_2)$ which is computationally less expensive is defined as:

$$\chi\left(I_1, I_2\right) = 1 - \frac{2d_m^{mid}}{\left\lfloor \frac{n}{2} \right\rfloor} \tag{6}$$

Here d_m^{mid} refers to the deviation at the $\left\lfloor \frac{n}{2} \right\rfloor$ index of the distance vector. It has the same properties as κ, but in practice is somewhat less robust. Theoretically, deviation at any index of the distance vector could be used, however, the largest range of deviation, $\left[0, \left\lfloor \frac{n}{2} \right\rfloor\right]$, occurs at the middle index position, the maximum occurring in the case of perfect negative correlation, which means higher discriminatory power.

3.4 Sensitivity

There are two issues concerning a stereo measure: its robustness and discriminatory power. The first determines the amount of data inconsistency that can be withstood by the measure at corresponding windows before mismatches begin to occur. The second is concerned with its ability to reject noncorresponding windows. These are conflicting requirements but crucial for stereo application.

A useful quality of our measures is their insensitivity to random noise and rank distortion. Consider the example of Section 3.1. κ remains at one when the intensity of pixel A is modified to a value in the range (80–255]. The reason is that corresponding rank matrices remain unchanged. Now let pixel A in window S first change to 75, and then to zero. The rank matrices of S corresponding to the two cases are:

	S			S	
1	3	7	2	4	8
2	5	9	3	6	9
4	6	8	5	7	1

Note that in case 2, the rank permutations of R and S are strongly related, i.e., all rank values of R except one has shifted by unit value. A correlation measure should be relatively unaffected by such changes, instead it should capture the overall relationship between permutations. For instance, the Euclidean distance between permutations, the one used in the Spearman's ρ is not attractive because it does not capture the underlying structure between the rankings. In the first case, the Spearman's ρ, the Kendall's τ, and *NCC* take values 0.98, 0.94, and 0.6, respectively. They change significantly to 0.40, 0.56, and 0.31, respectively, in the second case. On the other hand, κ remains fixed at 0.5 in *both cases*. It is futile to compare absolute values of correlation measures since each of them have different interpretations and are on different scales. Only their change with varying input is important since it represents sensitivity. Therefore, our measures capture the general relationship between data without being unduly influenced by unusual yet accurate data—after all, specular reflection and discontinuities are physical phenomena. It is worth noting that the Hamming distance (2) which can be used to define a measure of correlation, is not one of choice since it is sensitive to rank distortion.

Now we present the other side of the coin, namely, discriminatory power of κ. We concentrated on robustness in the presence of rank distortion in corresponding windows. But this robustness could turn into a liability when comparing windows which do not correspond. When window intensity values are replaced by their corresponding ranks, there is a loss of information, because the ratio between different measurement values is no longer used. The loss of information due to the choice of an ordinal scale of measurement is the price one pays for robustness, and is a well-known trade-off in nonparametric statistics [4]. One factor that affects discriminatory power is the window size. The window size determines the amount of sample data that will be used for comparison. When the window size is small, say 3×3, only five values $\left(\left\lfloor \frac{n}{2} \right\rfloor + 1\right)$ are possible for κ. Hence, the discriminatory power of the coefficients is low, and mismatches could result with high probability. As the window size increases, the discriminatory power of the coefficients increases. On real images, typically, window sizes of 7×7 or 9×9 perform well, as can be seen from experiments in the later sections. However, as with any window-based measure, continually increasing the window size causes performance to degrade because of increased false positives in the occlusion regions and smoothing of disparity values across depth boundaries. Choosing the appropriate window size is an issue for which we *do not* have a solution. In [2], we discuss the discriminatory power of κ in more detail.

3.5 Experiments With Synthetic Images

We compared the ordinal measures with *SSD*, *NCC*, and Zabih and Woodfill rank transform. We used the test suite [1] consisting of four sequences of images generated as benchmarks for matching algorithms. In each sequence, one parameter is varied; we will use sequences in which the noise level is varied. No pair of images in a sequence are stereoscopic since viewpoint between them remains unchanged. Therefore, the matching location for any pixel is known exactly. Further, there is no question of occlusion.

In Fig. 2, salt and pepper noise was added to the left image to generate the right image. Notice the significant degradation of image quality. The intensity variance in the window is used to estimate the amount of texture around the center pixel. If the variance is below a threshold, then we do not consider that point for matching. However, in the Zabih and Woodfill method, the rank transformed images are used for correlation instead of original intensity images; therefore, to keep the comparison fair, the test for

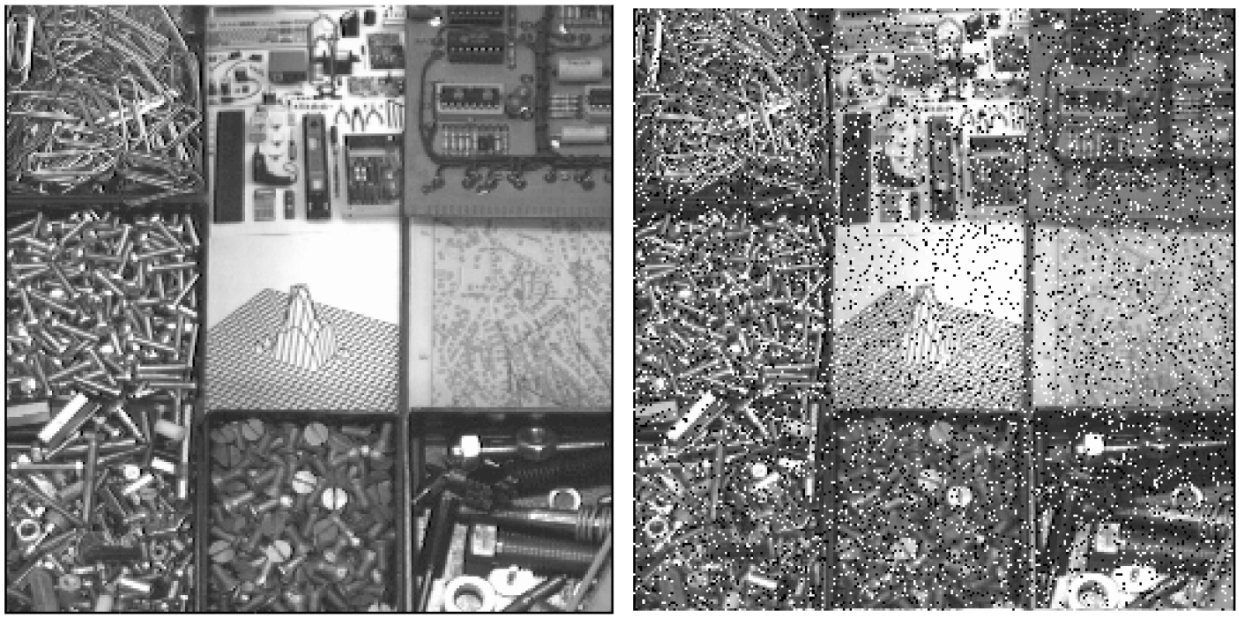


Fig. 2. Image pair with the right image generated by adding salt and pepper noise to the left image. This pair is used to test κ with other measures. Locating right matches with such significant image degradation is the computational challenge.

sufficient texture is performed using the same variance threshold while transforming the images. To simulate stereo matching, we use a search range of ±10 pixels. Matches are established for a region of size 100×100 ($= 10^4$) pixels of the left reference image.

The results of matching are shown in Table 1 which tabulates the number of matches incorrectly identified by each measure, i.e., the number of *false negatives* are reported. Each measure is denoted by the appropriate abbreviation. Note, since there is no occlusion in this example, the issue of false positives *does not arise*. It can be seen that κ gives the best results of all. χ does better than *NCC* and *SSD* but not as well as Zabih's method. All measures did better with increasing window size which is expected since there are no depth discontinuities and hence no occlusion. The graph in Fig. 3a represents the performance of *SSD*, *NCC*, and κ on a different image set (from the Aschwanden and Guggenbuhl test suite) with increasing amounts of shot noise. The performance of κ is much better than *SSD* or *NCC*. The rank-transform does better than *SSD* or *NCC* but not as well as κ. It may be argued that rather than using κ, one might first design a filter for the images to remove noise and then use regular correlation, but the problem of designing an optimal filter is hard. Next, the measures were compared (see Fig. 3b) with respect to increasing Gaussian noise level. In this case, κ does not do as well as *SSD* or *NCC* which are optimal estimators of linear regression parameters. Zabih's method also does better than κ because it uses linear correspondence on the mappings obtained through the rank transform.

The two measures were then tested on a random dot image pair (see Fig. 4) and compared with the other methods. The random dot image pair, each image of size 64×64 pixels, depicts a square (size: 20×20 pixels) moving four pixels to the right in front of a stationary textured background. Note that this is strictly not a stereo pair, but a motion sequence. Gaussian noise of variance 5.0 is added to both images, and there is a difference in intensity scale of 10 percent between the images. The computational problems are:

- to obtain correct correspondence including those at depth boundaries between the background and the moving plane and
- to correctly report that no matches can be found in the occlusion region—the region of size 4×20 to the right of the moving square with respect to the reference image.

The search range is fixed at ±10 pixels on a scanline for all methods. All methods also incorporate a *back matching strategy* wherein each match is verified independently by matching patches from the left image in the right image, and vice versa. If the match for a window from the left image is not mapped back to within a pixel of its location in the left image, it is not considered valid. This is a more uniform way to compare measures than to use different thresholds for different similarity measures in order to determine mismatches. The results (number of mismatches) are shown in Table 2.

TABLE 1
COMPARISON OF DIFFERENT MEASURES
USING THE IMAGES SHOWN IN FIG. 2

Measure	Mismatches		
	7×7 Window	9×9 Window	11×11 Window
κ	**1,324**	**923**	**791**
Zabih	1,752	1,171	809
χ	**1,856**	**1,270**	**1,001**
Norm. Corr.	4,128	2,991	2,245
SSD	4,567	3,469	2,645

Number of incorrect matches identified by each measure at different window sizes is tabulated.

TABLE 2
COMPARISON OF DIFFERENT MEASURES
USING THE RANDOM DOT IMAGES SHOWN IN FIG. 4

Measure	Mismatches		
	7×7 Window	9×9 Window	11×11 Window
κ	**51 (11)**	**69 (16)**	**103 (28)**
χ	**87 (14)**	**79 (16)**	**110 (30)**
Norm. Corr.	72 (15)	95 (22)	108 (30)
Zabih	124 (21)	100 (22)	112 (32)
SSD	211 (12)	141 (20)	134 (34)

The total number of mismatches which is the sum of the false negatives and the false positives is shown for different window sizes. The number of false positives which are the incorrectly reported matches in the occlusion zone is shown separately in brackets.

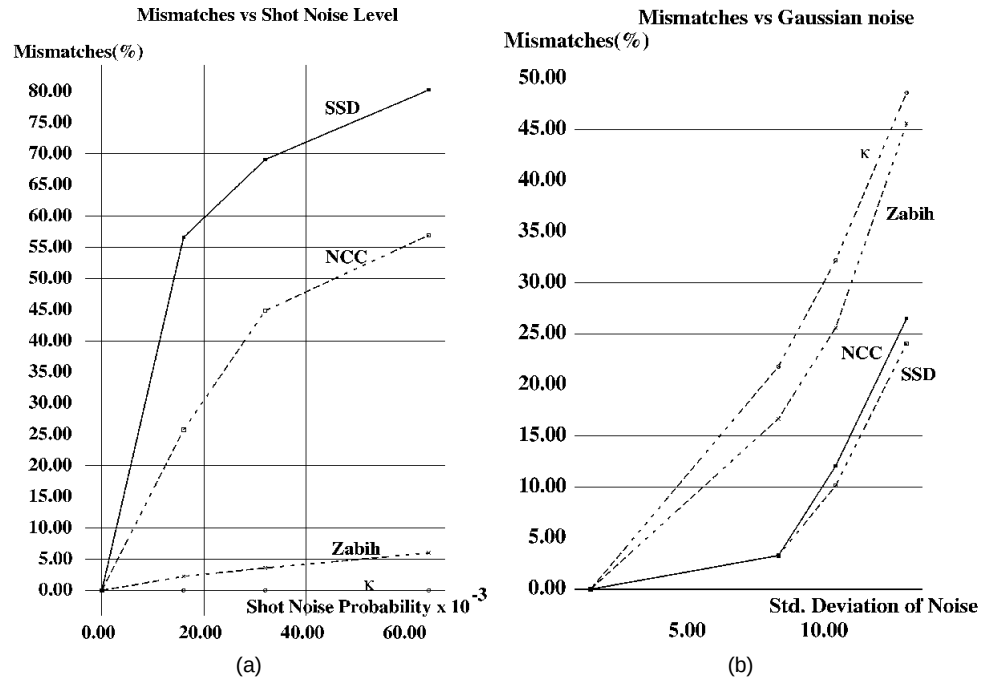


Fig. 3. Performance of the measures with (a) increasing shot noise level and (b) increasing Gaussian noise level.

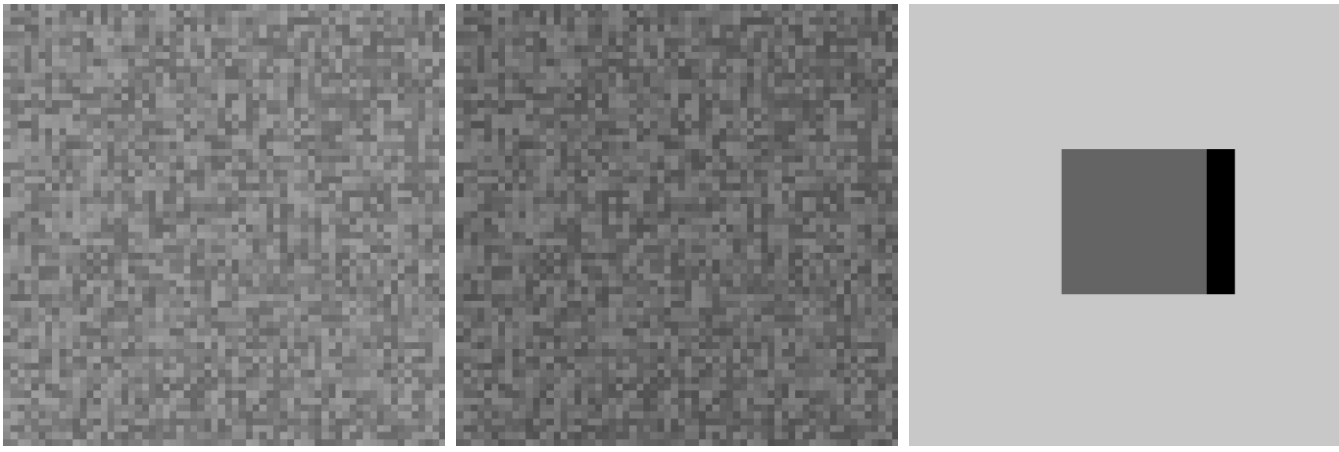


Fig. 4. Image pair (a)-(b) representing a square moving against a stationary textured background. In (c), true disparity levels with respect to the left image are shown in different shades. The darkest region indicates the occlusion region.

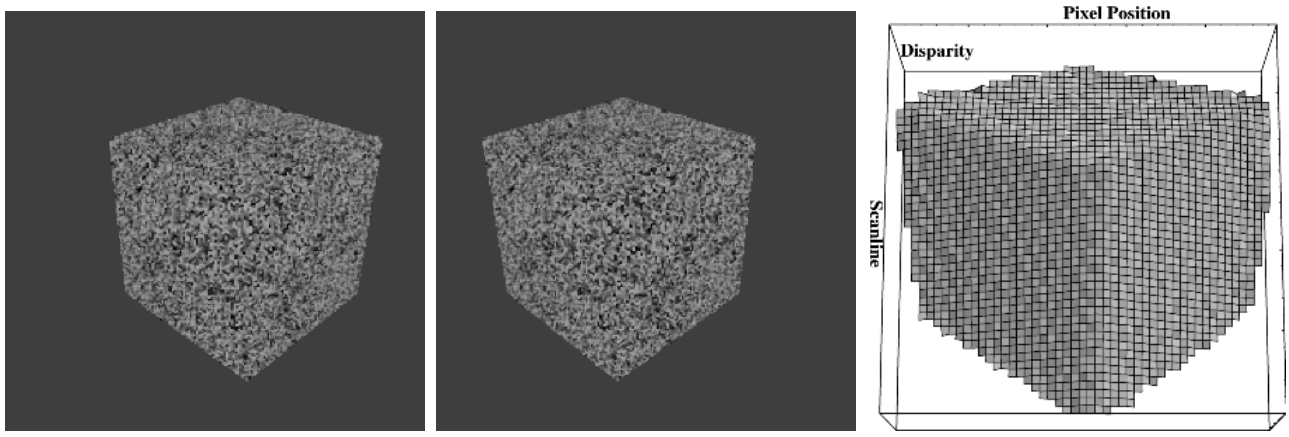
κ again does the best in comparison to the other measures. The improvement may not seem as drastic as in the earlier example, because the number of pixels on discontinuities and in the occlusion region is small. Observe that the number of false positives in the occlusion zone increases with window size for all measures, as expected. Also note that the total number of mismatches decreases with increasing window size in the case of *SSD*, because the number of false negatives decreases faster than the increase in false positives. This can be explained by the fact that at smaller window sizes the effect of intensity difference between images and noise is more pronounced on *SSD* than on the others, and as window size increases, the performance of *SSD* with respect to false negatives increases fast. Presumably, there would be a window size beyond which all measures would do equally badly because of false positives and disparity smoothing.

Table 3 shows the number of false negatives obtained using each measure. The results were obtained by retracting the backmatching strategy as explained earlier. Also, no thresholds were used for any measure, and hence all measures were forced to report incorrect matches in the occlusion region (in truth, there should be no match for any pixel in that region). The aim was to

TABLE 3
COMPARISON OF THE MEASURES
USING THE RANDOM DOT IMAGES OF FIG. 4

Measure	Mismatches		
	7×7 Window	9×9 Window	11×11 Window
κ	**35**	**43**	**59**
χ	**47**	**50**	**62**
Norm. Corr.	47	54	60
Zabih	56	65	54
SSD	136	97	84

In this case, only the false negatives are reported (see text for details).



(a) (b) (c)

Fig. 5. Stereo pair (a)-(b) of a textured cube (by Bill Hoff) and disparity plot (c) obtained using κ.

compare the performance with respect to false negatives alone. Thus, the mismatches reported in the table are those due to disparity smoothing across boundaries, noise, and intensity bias between the images. The results are consistent with the earlier table. Particularly, note that with increasing window size, the number of false negatives increases (except in the case of *SSD*), which can be attributed to disparity smoothing across the discontinuities. The reason for diminishing mismatches in the case of *SSD* is explained earlier.

Next, we use a stereo pair of a densely textured cube[1] (see Fig. 5) with disparity variation in the range [25–50] pixels. The issues were to obtain accurate disparity in the presence of significant projective distortion, and to match correctly at the depth discontinuities between the cube and the black background. The window size used was 9×9. The resulting dense disparity map is shown in Fig. 5 which is accurate. To verify, we compared the obtained disparity by the plane-fit error method as follows. The disparity is obtained manually (i.e., by visually matching) at the corners of the cube. We then fit planes to each of the three visible faces of the cube using the pixel positions of the corners and the manually calculated disparity. Finally, we computed the squared difference of the expected disparity from the plane fit and the computed disparity using κ, at all pixels on each of the three faces. We observed that, up to pixel accuracy, the result is nearly 100 percent accurate. While this is not the best way to compare results, it may provide qualitative justification.

4 Computational Issues

The naive algorithm for computing the distance vector by searching linearly through s is an $O(n^2)$ method. Hence, if D is the disparity search range, then the cost of computing $\mathbf{d_m}$ at every pixel would be $O(Dn^2)$. This is in addition to the sort operations to perform ranking. We developed a simple $O(n)$ algorithm for building $\mathbf{d_m}$ while simultaneously evaluating $\max_{i=1}^{n} d_m^i$ (see [2] for details). Mathematically, our algorithm states that,

$$d_m^{i+1} = d_m^i - \sum_{j=1}^{i} J\left(s^j = i+1\right) + J\left(s^{i+1} > i+1\right) \tag{7}$$

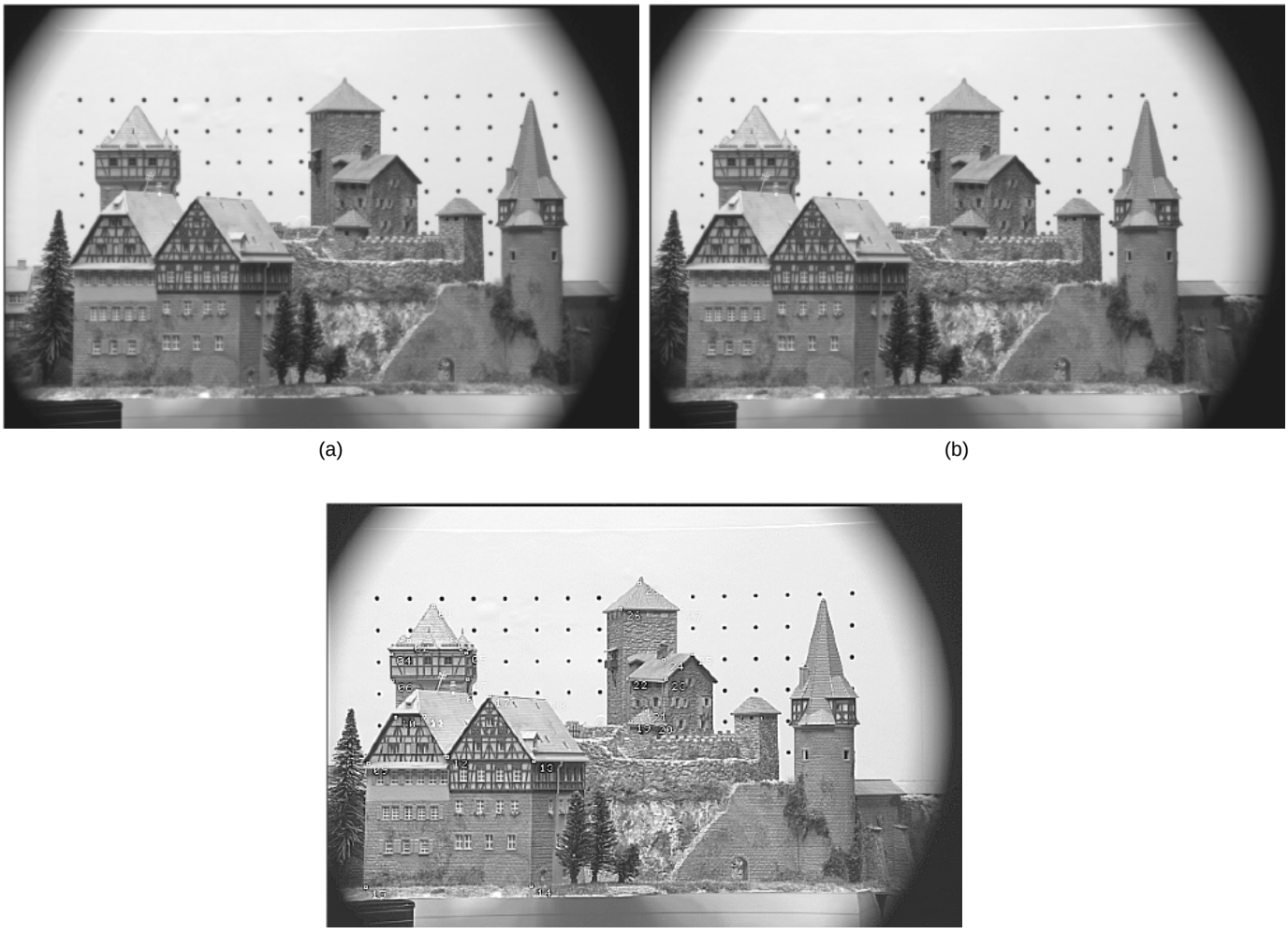
The costly operations are therefore those of sorting window data, which is $O(n \log n)$. Note, however, that we do not have to sort a window in the second image every time it slides across through one pixel distance within the search range D, if we use heap-sort in which data is maintained as a heap tree. Only delete and insert operations corresponding to difference between the old and the new window have to be performed. Since each operation is of the order $O(\log n)$ and the total number of operations is less than n, this scheme is more economical than sorting anew.

A preferable alternate scheme to avoid comparison sorting (heapsort, quicksort) is as follows. Note that intensity values are integers and lie in the range $[0, 2^k - 1]$, where k represents the number of bits of intensity resolution. We can now use *counting sort* which is $O(n + 2^k)$. Currently, eight-bit sensors are the norm, which implies intensity values must lie in the range [0–255]. Hence, sorting in a window is $O(n + 256)$, *linear* in n. Counting sort is effective with tied data values too since it is a stable sorting algorithm, i.e., the relative ordering of tied data values is preserved in the generated permutation. To find that value of n when counting sort begins to perform better than comparison sort, the following inequality must be satisfied: $c_1 n \log n > c_2(n + 256)$, where c_1, c_2 are constants of the algorithm. If $c_1 = c_2$, then for $n \geq 64$ (or equivalently, a window of size 8×8), counting sort is better. Further improvement in computational complexity is gained if the images are preprocessed to determine the maximum and minimum intensity values in windows of each image. Consequently, the complexity of sorting each sample is $O(n + B)$ where B is often much less than 256. When sensors of higher intensity resolution are used ($k > 10$), then a comparison sort algorithm would be of choice. Alternatively, the images can be subsampled in intensity resolution to eight bits. For ordinal measures, higher intensity resolution is not highly beneficial, since actual intensity values are not used.

5 Experiments

The first is a stereo image pair in Fig. 6 from the Calibrated Imaging Laboratory at Carnegie Mellon University. A sequence of images was obtained by moving the camera horizontally. Precise disparity was tabulated at 28 points (shown in the figure) using an active range-sensing method. We used the third and fourth images in the sequence. Note that many points are located on depth discontinuities which pose a serious problem for stereo matching. The disparity range is [20–35] pixels with respect to the left stereo image which is used as the reference. Since we do not attempt subpixel matching, the nearest integer coordinates for each reference point is used as its position in the image. A window size of 9×9 is used for matching. Point 18 was matched inaccurately, the possible reason being that the window which straddles it is largely untextured. Also, correspondence at point eight was off by two pixels. All other points (26 of them) were matched correctly to the nearest pixel. The complete results of matching are tabulated in [2].

1. This stereo pair was developed at the University of Illinois by Bill Hoff. Its title is "Synthetic Image of a Cube With Gray Random-Dot Texture."



(a) (b)

(c)

Fig. 6. (a) and (b) Stereo image pair obtained in the Calibrated Imaging Lab at Carnegie Mellon University (by Mark Maimone). (c) The 28 points for which precise disparity has been obtained using an active range method.

Fig. 7. Stereo images of the Pentagon.

Next, we present results with the rather popular stereo image pair of the Pentagon (see Fig. 7). The images are of resolution 512×512. Disparity is horizontal in the range [–10, 10] pixels. We matched the images using two different window sizes: 9×9 and 11×11, and the respective results are shown in Fig. 8. We also used a threshold of 0.5 on κ in each case. It may be observed that the disparity map at the larger window size is smoother at the edges as expected. Note that the results obtained were using pure image matching, i.e., without any preprocessing or by using any complex heuristics. Presumably, the results would improve further if an approach specifically for handling occlusion like in [5] was used in tandem.

(a) (b) (c)

Fig. 8. (a) and (b) Density plots of the disparity maps representing the Pentagon pair obtained at two different window sizes. (c) A surface plot.

6 Conclusion

We have presented ordinal measures for stereo correspondence and have shown it to be robust under nonideal conditions. We also developed computationally efficient algorithms for evaluating them. Although the issues were presented in the context of stereo matching, they are equally applicable in motion estimation. The main feature of this work is the reinterpretation of intensity on an ordinal scale, although measurements are on a ratio scale. One possible drawback of using an ordinal scale is that effect of common image error models cannot be easily incorporated. In other words, it not clear how the parameters of an error model would translate to parameters in a ranking model; our current work seeks to address this issue. The use of ordinal measures can be extended to applications like image retrieval based on texture where the problem is to match texture descriptors of two images. Texture descriptors could represent outputs of an image convolved with, say, Gabor filters of different parameterizations. Hence, the use of an ordinal scale for measurements may have wider scope.

Acknowledgments

We are grateful to Dr. Michael Oren at MIT for his detailed comments on the paper. We thank Dr. Harlyn Baker of SRI for kindly providing the Aschwanden test images. We thank Prof. Ramin Zabih for providing his thesis.

This research was conducted at the Center for Research on Intelligent Systems at Columbia University. It was supported in parts by ARPA Contract DACA-76-92-C-007, DOD/ONR MURI Grant N00014-95-1-0601, and a U.S. National Science Foundation National Young Investigator Award.

References

[1] P. Aschwanden and W. Guggenbuhl, "Experimental Results From a Comparative Study on Correlation-Type Registration Algorithms," Forstner and Ruweidel, eds., *Robust Computer Vision*, pp. 268-289. Wichmann, 1993.

[2] D.N. Bhat and S.K. Nayar, "Ordinal Measures for Visual Correspondence," Columbia Univ., Computer Science, tech. rep. CUCS-009-96, Feb. 1996.

[3] M.J. Black, "Robust Incremental Optical Flow," PhD thesis, Yale Univ., Sept. 1992.

[4] W.J. Conover, *Practical Nonparametric Statistics*. New York: John Wiley, 1980.

[5] D. Geiger, B. Ladendorf, and A. Yuille, "Binocular Stereo With Occlusion," *Proc. European Conf. Computer Vision*, pp. 423-433, 1992.

[6] R.A. Gideon and R.A. Hollister, "A Rank Correlation Coefficient," *J. Am. Statistical Assoc.*, vol. 82, no. 398, pp. 656-666, 1987.

[7] M. Kendall and J.D. Gibbons, *Rank Correlation Methods*, fifth ed. New York: Edward Arnold, 1990.

[8] R. Kories and G. Zimmerman, "A Versatile Method for the Estimation of Displacement Vector Fields From Image Sequences," *Proc. IEEE Workshop Motion: Representation and Analysis*, pp. 101-107, 1986.

[9] M. Okutomi and T. Kanade, "A Locally Adaptive Window for Signal Matching," *Int'l J. Computer Vision*, vol. 7, no. 2, pp. 1,499-1,512, 1992.

[10] L.H. Quam, "Hierarchical Warp Stereo," *Proc. ARPA Image Understanding Workshop*, pp. 149-155, 1984.

[11] R. Zabih and J. Woodfill, "Non-Parametric Local Transforms for Computing Visual Correspondence," *Proc. European Conf. Computer Vision*, pp. 151-158, 1994.