

Ordinal Measures for Visual Correspondence *

Dinkar N. Bhat and Shree K. Nayar

Department of Computer Science, Columbia University, NY 10027

Abstract

We present ordinal measures for establishing image correspondence. Linear correspondence measures like correlation and the sum of squared differences are known to be fragile. Ordinal measures, which are based on relative ordering of intensity values in windows, have demonstrable robustness to depth discontinuities, occlusion and noise. The relative ordering of intensity values in each window is represented by a rank permutation which is obtained by sorting the corresponding intensity data. By using a novel distance metric between the rank permutations, we arrive at ordinal correlation coefficients. These coefficients are independent of absolute intensity scale, i.e they are normalized measures. Further, since rank permutations are invariant to monotone transformations of the intensity values, the coefficients are unaffected by nonlinear effects like gamma variation between images. We have developed a simple algorithm for their efficient implementation. Experiments suggest the superiority of ordinal measures over existing techniques under non-ideal conditions. Though we present ordinal measures in the context of stereo, they serve as a general tool for image matching that is applicable to other vision problems such as motion estimation and image registration.

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. The lack of robustness in such settings was clearly noted in [4] while evaluating operational stereo systems. These issues also arise in motion estimation, however, for the sake of clarity, we will uniformly discuss them in the context of robust stereo.

Area-based stereo methods have been popular since they produce dense depth maps without requiring explicit surface reconstruction. The heart of any such method lies in the similarity criterion used that determines optimal statistical correlation between corresponding regions. A similarity measure has to satisfactorily deal with the following issues:

- **Depth discontinuities:** A window located on a depth discontinuity will represent scene points at different depths. Further, windows around corresponding points in the stereo images do not represent the same surface regions. The issue then is to identify correspondence even in the presence of such inconsistent intensity data (see Figure 1a).

*This research was conducted at the Center for Research on Intelligent Systems, 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 NSF National Young Investigator Award.

- **Occlusion:** A problem related to depth discontinuities is that of occlusion. Due to occlusion, portions of a scene are visible only in one of the two images (see Figure 1b)¹. A stereo operator must correctly identify occlusion regions by reporting that no match can be found for points therein.
- **Noise:** Noise is caused due to statistical limitations in the sensing and digitization process, and electronic imperfections. A reliable stereo measure must tolerate low signal-to-noise ratio which is common with low resolution sensors.
- **Specular reflection:** Corresponding point intensities are not identical in the presence of specular reflection, the specular intensity at any scene point being dependent on the viewing direction (see Figure 1c). Stereo algorithms must allow for deviations from the Lambertian model, specular reflection being the most dramatic of such aberrations.
- **Window Distortion:** Due to projection from different viewpoints, corresponding windows do not represent the same surface patch in the scene, except when the surface is fronto-planar (see Figure 1d). The computational challenge is essentially similar to that with discontinuities, i.e can the stereo measure endure certain degree of data inconsistency in corresponding windows?
- **Camera Parameter Variations:** If different sensors are used for stereo, then the camera aperture, bias reference and gamma factors could vary between them. Therefore, sensor outputs could be unequal even in the ideal Lambertian case.

Stereo methods must be robust to the above vagaries, i.e we require stereo operators that: a) are insensitive to outliers to a high degree, b) can reliably locate matches even with inconsistent or inhomogeneous intensity in corresponding windows, c) can identify mismatches (or matches) with prescribed confidence, d) are insensitive to deviations from the Lambertian model, and e) are independent of sensor gain and bias. In this paper, we present *ordinal* measures of association ([5], [1], [6]) which possess the above desirable qualities to a high degree. An ordinal variable implies one drawn from a discrete ordered set like the grade in school. The ratio between two measurements is not of consequence, only their relative ordering is relevant which is expressed by their *ranks*. A rank permutation is obtained by sorting the sample in ascending order and labeling them using

¹Figure 1b, 1d are clipped from a image pair captured by Steve Cochran, USC Institute of Robotics and Intelligent Systems. Its description: "Partial view of a Rubik's cube occluding a block".

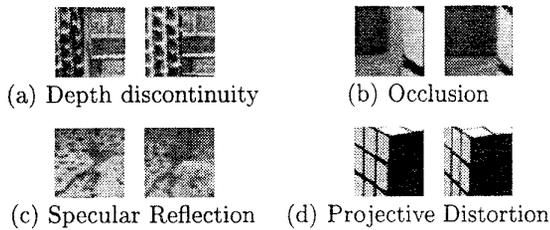


Figure 1: Illustration of different phenomena which affect window-based stereo matching. In (a), (c), (d) identically sized windows are shown around corresponding pixels - the center pixels. (a) Since the center pixel is located on a depth discontinuity, the windows represent different surface locations. (b) Due to occlusion, pixels in the left window are not visible in the right one, therefore, correspondence cannot be achieved at those left pixels. (c) Specular reflection causes intensities in corresponding windows to differ. Notice the varying location of the highlights with respect to the texture. (d) Projective distortion results in windows being different which can be seen from the unequal texture frequency.

integers $[1, 2, \dots, n]$, n being the size of the sample. In our application, intensity is viewed as an ordinal variable. Consequently, ordinal correlation measures are based on the rank permutations rather than absolute intensity data. 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 significantly distorted like in the presence of specular reflection or discontinuities, they are not satisfactory. This is in contrast to the measures described in this paper which are robust to rank distortion. They are non-parametric, which means, they can be interpreted even in the absence of strong structural assumptions about the data in windows. Thus, confidence thresholds for matching can be established to identify occlusion regions. We present a simple and computationally economical algorithm to evaluate the measures. Experiments with real images and comparison with existing matching methods suggest their superiority.

2 Ordinal Measures

In this section, we present ordinal measures of association after a brief review of the concept of correlation based on distance metrics. We discuss the sensitivity of the measures with respect to outliers and rank distortion, and compare them to other correlation methods.

2.1 Motivation

Consider the following example of a 3×3 reference window M (master) with intensity I_1 :

M		
10	30	70
20	50	80
40	60	100

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

1	3	7	1	3	7
2	5	8	2	5	8
4	6	9	3	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. 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 1. This is unlike the linear correlation coefficient which can *substantially deviate*. For example, when the pixel takes a value of 255, the *normalized correlation coefficient* $NCC = 0.645$. This attractive property of ordinal measures motivates us to apply them for stereo matching. We now formally introduce the concepts underlying ordinal measures using distance metrics.

2.2 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 $[1, 2, \dots, n]$, 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 [5]. Once a distance metric $d(\pi_1, \pi_2)$ is defined, a coefficient of correlation α can be obtained as:

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

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 permutations are reverses of each other, and hence $\alpha = -1$. Different distance metrics are possible, an example being the *Hamming distance* d_h :

$$d_h(\pi_1, \pi_2) = \sum_i (|\text{sgn}(\pi_1^i - \pi_2^i)|) \quad (2)$$

where $\text{sgn}(x) = x/|x|$, if $x \neq 0$, and 0 otherwise. For the Hamming distance, $M = n$. The Kendall's τ and the Spearman's ρ too can be expressed using distance metrics although it is seldom done. The reader is referred to [3] for details.

We noted earlier that data inconsistency between corresponding windows can occur due to the presence of specular reflection and discontinuities. This could result in corresponding rank matrices being distorted unlike in the case of the example discussed in section 2.1. As a result, ordinal measures like the Kendall's τ and Spearman's ρ are inadequate.

2.3 Proposed Measures

For a set of window intensity values $(I_1^i, I_2^i)_{i=1}^n$, 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. Let us assume that the ranks are unique for the time being; we will discuss tied ranks at a later juncture. We define a composition permutation s as follows:

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

where π_1^{-1} denotes the inverse permutation of π_1 . The inverse permutation is defined as follows: If $\pi_1^i = j$, then $(\pi_1^{-1})^j = i$. Informally, s^i is the rank of the pixel in I_2 that corresponds to the pixel with rank i in I_1 . Under perfect positive correlation, s should be identical to the *identity permutation* given by $u = (1, 2, \dots, n)$.

By defining a distance measure between s and u , we in turn obtain a notion of distance between π_1 and π_2 . The deviation d_m^i at each s^i is defined as the number of $s^j, j = 1, \dots, i$ greater than i . Formally,

$$d_m^i = \sum_{j=1}^i J(s^j > i) \quad (4)$$

where $J(B)$ is an indicator function of event B , i.e. $J(B)$ is 1 when B is true and 0 otherwise. The vector of d_m^i values is termed as the distance vector $\mathbf{d}_m(s, u)$ [6]. Informally, the distance vector estimates the number of s elements that are out of position, similar to the Hamming distance. However, it does not penalize out of position elements as severely as the Hamming distance, which makes our distance vector relatively less sensitive to rank distortion effects observed between corresponding windows during specular reflection and depth discontinuities. If (I_1, I_2) were perfectly correlated, then $\mathbf{d}_m(s, u) = (0, 0, \dots, 0)$. The maximum value that any component of the distance vector can take is $\lfloor \frac{n}{2} \rfloor$ which must occur in the case of perfect negative correlation [3]. Now, a measure of correlation $\kappa = \kappa(I_1, I_2)$ is defined using equation 1 as:

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

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 negative correlated. κ has the following desirable properties of a correlation coefficient:

- it is independent of scaling and shift of the intensity values. For our application, it implies independence from camera aperture settings and bias,
- it is a normalized measure, i.e. $-1 \leq \kappa \leq 1$,
- it is symmetrical, i.e. $\kappa(I_1, I_2) = \kappa(I_2, I_1)$. Hence, either stereo image can be used as reference [3],
- $\kappa(f(I_1), h(I_2)) = \kappa(I_1, I_2)$ where f and h are strictly monotonically increasing or decreasing functions of I_1 and I_2 , respectively. If one of f, h is increasing and the other is decreasing, then the value of κ simply has its sign switched. This property comes useful when different cameras are used

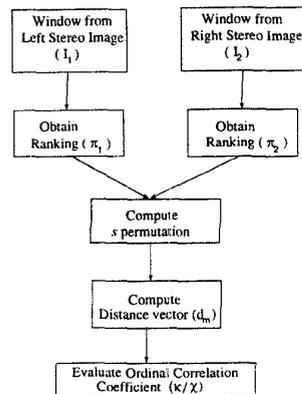


Figure 2: Flow-chart depicting the procedure involved in computing κ or χ .

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. For illustration, let the gains of the cameras be identically 1.0 and the bias of the cameras be 0. Let the imaged surface be Lambertian, i.e. the image irradiance from any point is identical for both sensors. Then, the sensor outputs are related by the equation $I_1 = (I_2)^t$ where $t = \frac{\gamma_2}{\gamma_1}$. In general, $t \neq 1$, and hence the linearity between the sensor outputs is lost. However, $|\kappa|$ remains at 1 because $(I_2)^t$ is a strictly increasing or decreasing function² of I_2 depending on t , and hence its ranking remains unchanged.

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

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

Here d_m^{mid} refers to the deviation at the $\lfloor \frac{n}{2} \rfloor$ index of the distance vector. It has the same properties as κ , but in practice is somewhat less robust. We conclude by summarizing the procedure involved in computing our measures using the flow-chart shown in figure 2.

2.4 Sensitivity

The most useful quality of the measures are their insensitivity to random noise and rank distortion which can occur due to specular reflection and discontinuities. Our discussion of these effects will be in the context of κ but could be easily extended to χ too. Consider the example of section 2.1. κ remains at 1 when the

²This property of the correlation coefficient does not help to deal with specular reflection since no monotonic relationship between the variables I_1 and I_2 can be established.

intensity of pixel A is modified to a value in the range $(80 - 255]$. The reason is that the corresponding rank matrices remain unchanged. Now let the value of pixel A in window S be changed to 75. Then, the rank matrices representing π_1 and π_2 are:

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

Note the modification of ranks in S . As might be expected, κ decreases and acquires a value of 0.8. This is in fact quite comforting since it shows that κ is sensitive to changing data. On the other hand, NCC changes from 1.0 to 0.6. Now let A take a value between 0 and 10, in which case the rank matrix of S is significantly modified as shown below:

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

However, the value of κ remains at 0.8. This behaviour is in sharp contrast to the Kendall's τ and Spearman ρ [7] which fall steeply to 0.556 and 0.4, respectively. If pixel A takes a value of 0, then the linear correlation coefficient NCC drops to 0.311.

The above example, albeit contrived, serves to illustrate the robustness of the measures we propose. In reality, the manifestation of specular reflection and discontinuities can distort ranks between corresponding windows more drastically, i.e more than one data value in S may differ from the corresponding value in M . However, by choosing a sufficiently large window we achieve similar insensitivity which is demonstrated in the experiments. In summary, our measures capture the general relationship between data without being unduly influenced by unusual yet accurate data.

2.5 Comparison

We compare our measures with the sum of squared differences measure SSD , the normalized correlation coefficient NCC , and Zabih's rank transform [10]. We use the test suite developed by [2] 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 (see Figure 3). None of the pair of images in a sequence are stereoscopic since viewpoint between them remains unchanged.

In Figure 3, salt and pepper noise³ was added to the right image. Notice the significant degradation in quality of the right image. We use the intensity variance in the window 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. To simulate stereo matching, we use a search range R of ± 10 pixels. Matches are established for a region of size

³This is used to model electronic noise. Pixels are randomly chosen and set to black("pepper") or white("salt").

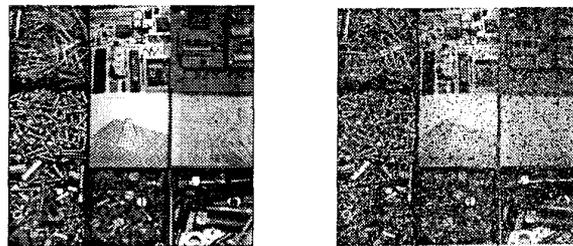


Figure 3: Image pair with salt and pepper noise added to the right image. This pair is used to test κ and χ with other measures.

$100 \times 100 (= 10^4)$ 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 positives are reported. It can be seen that κ gives the best results.

Measure	Mismatches		
	7 × 7 Window	9 × 9 Window	11 × 11 Window
κ	1324	923	791
Zabih	1752	1171	809
χ	1856	1270	1001
Norm. Corr.	4128	2991	2245
SSD	4567	3469	2645

Table 1: Comparison of different measures using the images shown in Figure 3. The number of incorrect matches identified by each measure at different window sizes is tabulated.

χ does better than NCC and SSD but not as well as Zabih's method. While χ does not perform as well as κ , it is computationally less expensive, and is comparable to Zabih's rank transform. All measures did better with increasing window size.

The two measures were tested next on a random dot stereogram (see Figure 4) and compared with the other methods. The random dot stereo pair, each image of size 64×64 pixels, depicts a square (size: 20×20 pixels) moving 4 pixels to the right in front of a stationary textured background. Gaussian noise of variance 5.0 is added to both images, and there is a difference in intensity scale of 10% between the images. The computational problems are: a) To obtain correct disparity at all corresponding points including those at depth boundaries between the background and the moving plane, and b) 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 scan-

line 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. The results (number of mismatches) are shown in Table 2.

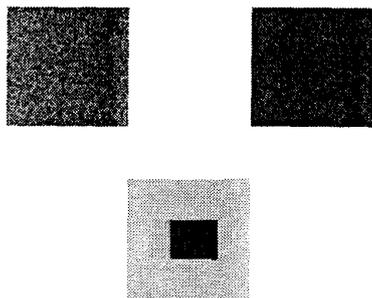


Figure 4: Left and right stereo images of a random dot stereogram are shown on the top row. The right image is 10% darker than the left. They represent a square moving against a stationary textured background. On the bottom row, true disparity levels with respect to the left image are shown in different shades. The darkest region (to the right of the inner square) indicates the occlusion region.

Measure	Mismatches		
	7 × 7 Window	9 × 9 Window	11 × 11 Window
κ	54	75	98
χ	87	79	110
Norm. Corr.	72	95	108
Zabih	124	100	112
SSD	211	141	134

Table 2: Comparison of different measures using the random dot stereo images shown in Figure 4. The number of matches incorrectly identified by a measure at different window sizes is shown.

Once again κ does the best in comparison to the other measures. The improvement may not seem as drastic as in the earlier example. The reason is that the number of pixels on discontinuities and in the occlusion region is small - only 140. On the average, κ performs about 21% better than *NCC*, 49% better than *SSD*, and about 31% better than the rank transform. χ does almost as well as normalized correlation but better than *SSD* and Zabih's rank transform. The number of mismatches obtained by *SSD* in regions not corresponding

to depth discontinuities or the occlusion zone decreases progressively with window size. However, in all other cases, the corresponding figure remains nearly constant at 0. Hence, mismatches are all on the depth boundaries and in the occlusion region. With increasing window size, the probability of smoothing of disparity values across depth boundaries and the occlusion region increases, which explains the observed degradation in performance of all measures except *SSD*.

3 Statistical Issues

In this section, we will discuss statistical issues relevant to the proposed measures.

- *Tied data*: In practice, tied ranks within a window are highly possible, i.e. two or more pixels can have identical intensity values. Hence, the definition of s in equation 3 is unclear. The question then is: What rank should be assigned to the set of tied data values? In general, if there are g groups of tied data, each group containing $n_j, j = 1, \dots, g$ elements, then the total number of rankings possible is $n_1!n_2!\dots n_g!$. We rank tied values such that the relative spatial ordering between them is preserved. This method of breaking ties ensures that when sample windows correspond, the two rankings are consistent. In other words, our ranking method most favors positive correlation.
- *Sample 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 5 values ($\lfloor \frac{n}{2} \rfloor + 1$) are possible for κ (and χ). Hence, the discriminability 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.
- *Confidence Thresholds*: The linear correlation coefficient has no clear meaning, i.e. it is not possible to attach a significance level⁴ to it without explicit distributional assumptions about the sample data. These assumptions cannot be made, especially for samples containing inhomogeneous data. Therefore, an algorithm cannot reject or accept a match with any confidence. Typically an *ad hoc* threshold based on experience is chosen.

On the other hand, the significance level of our measures can be tested using hypothesis testing methods. The reason is that although the distribution of the sample data is not known *a priori*, the s permutation is uniformly distributed under the null hypothesis of there being no association between the samples. We do not explain the entire procedure here due to lack of space (see [3]), but we obtain a table for κ as below:

Specified Confidence level = $t\%$	
Window size	Computed Threshold
3×3	c_1
5×5	c_2
...	...

⁴Significance level denotes the maximum probability at which a true null hypothesis is rejected. The corresponding value of the test statistic (κ or χ) is the confidence threshold, and (1-significance level) expressed in percentage is the confidence level.

A match between windows would be rejected at a specified confidence level, if κ between them is below the appropriate threshold.

As an example, for a window size of 7×7 and $t = 99\%$, we would not accept a match unless κ exceeds $c = 0.33$. To verify this threshold, we used it for matching in the random dot stereogram example. We replaced the back matching strategy by the threshold to evaluate a match. The number of mismatches increased marginally from 54 to 57. This small increase could be attributed to the errors associated with hypothesis testing. Nevertheless, detailed experimental analysis has to be performed before such thresholds can be used for practical stereo work.

4 Computational Issues

The naive algorithm for computing every distance vector component d_m^i by searching linearly through s is an $O(n^2)$ method. Since, for every pixel in the reference image, we search in a range R in the second image, the cost of computing \mathbf{d}_m would be $O(Rn^2)$. In this section, we sketch a simple $O(n)$ algorithm for building \mathbf{d}_m while simultaneously evaluating $\max_{i=1}^n d_m^i$. This is explained below using a geometrical construction (see Figure 5).

The large dots represent the elements of the s permutation. The distance vector component d_m^{i+1} is equal to the number of s elements in the rectangular region GCDE. This is identical to the total number of s elements in rectangular regions HCDF and GHFE. If, while computing d_m^i , the total number of s elements equal to $[i, i + 1]$ was recorded as say m (equal to s elements in the region ABHG), then the number of s elements within the area GHFE = $d_m^i - m$. Therefore, while computing d_m^{i+1} , only s elements in the area HCDF need to be determined. This is true for $i = 1, \dots, n$, and thus we obtain a $O(n)$ algorithm to calculate the distance vector. It is trivial to compute $\max_{i=1}^n (d_m^i)$. An algebraic proof is also sketched in [3]. Mathematically,

$$d_m^{i+1} = d_m^i - \sum_{j=1}^i J(i < s^j \leq i+1) + \sum_{j=i}^{i+1} J(s^j > i+1) \quad (7)$$

We use the following method to obtain the rank permutations instead of comparison sorting schemes which are $O(n \log n)$. 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, 8-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 . Furthermore, counting sort is effective with tied data values too. To find the value of n when counting sort out-performs comparison sort, the following inequality must be satisfied:

$$c_1 n \log n > c_2 (n + 256)$$

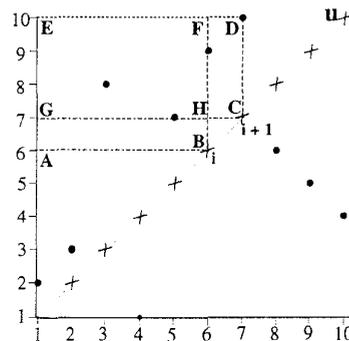


Figure 5: Scatter diagram of a bivariate rank distribution is shown. The large dots represent the ranking of I_2 with respect to I_1 , i.e their ordinate values when read along the positive x axis give s as in equation 3. The crosses on the straight line at a slope of 45° refer to the identity permutation u . i and $i + 1$ refer to indices of the distance vector.

where c_1, c_2 are constants of the algorithms. If $c_1 = c_2$, then for $n \geq 64$ (or equivalently, a window of size 8×8) counting sort is better.

The computational cost of κ and χ is much lower than that of the Kendall's τ which is $O(n^2)$. It is asymptotically identical to that of Spearman's ρ , and in practice slightly better. It is less economical than NCC and much costlier than SSD . However, the increased reliability over SSD and NCC may compensate for the decreased economy in certain applications.

5 Experiments

Ideally, we would like to compare our measures with others using dense ground truth, but unfortunately such data is lacking [4]. In this paper, we present two of the conducted experiments (see [3] for other experiments) which qualitatively substantiate the results presented in section 2.5. In both cases, κ is used for matching.

The first is a stereo image pair in figure 6 from the Calibrated Imaging Laboratory at CMU [8]. A sequence of images was obtained by moving the camera laterally. Precise disparity was tabulated at 28 points (shown in the figure) using an active range sensing method. Note that many points are located on depth discontinuities which pose a serious problem for stereo matching. The disparity range is $[20 - 35]$ pixels, and a window size of 9×9 is used for matching. Except for point 14 located at the bottom left in figure 6c, all others were matched accurately upto pixel accuracy. This result was consistent with two other image pairs in the same sequence.

Next, we use a stereo pair of a densely textured cube⁵ (see Figure 7) with disparity variation in the range $[25 - 50]$ pixels. The issues are to obtain accurate disparity in spite of the significant projective distortion, and to match correctly at the edges. The window size

⁵This stereo pair was developed at the University of Illinois by Bill Hoff. Its title, "Synthetic image of a cube with gray random-dot texture".

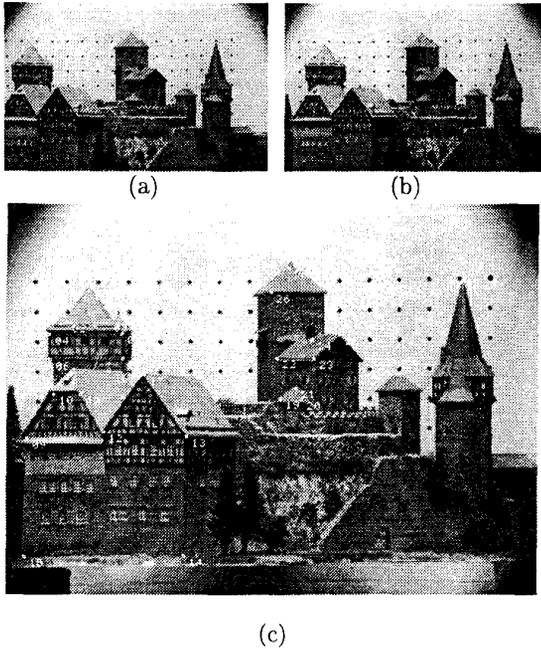


Figure 6: Stereo image pair obtained by the Calibrated Imaging Lab in CMU (by Mark Maimone) is shown on the top row. The bottom picture shows 28 points for which precise disparity has been obtained using an active range method.

is 9×9 . The resulting dense disparity map is shown in figure 8 which is accurate. To verify, we compared the obtained disparity by the plane-fit error method [9], and the result is nearly 100% accurate (upto pixel accuracy).

6 Discussion

We presented ordinal measures for visual correspondence and have shown them to be robust in the presence of depth discontinuities, occlusion, and non-linear reflectance. We also developed a computationally efficient algorithm for evaluating these measures. We have concentrated on robustness in the presence of rank distortion and outliers 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, namely, the ratio between different measurement values. As a result, textures that are different may have the same rank distribution and hence an ordinal measure would report a good match between them. On the other hand, normalized correlation between the intensity values will be low which enables disambiguating correspondence. The loss of discriminability due to the choice of an ordinal scale of measurement is the price one pays for robustness. Equivalently, in statistical terms, ordinal measures are much more robust but less efficient than their counterparts.

Acknowledgements

We are very grateful to Michael Oren of MIT for his extremely detailed comments on the paper. We thank

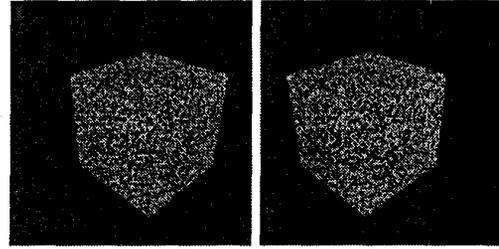


Figure 7: Stereo pair of a textured cube (by Bill Hoff at the University of Illinois).

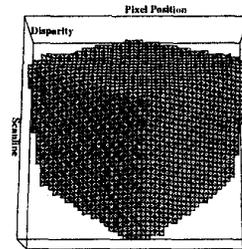


Figure 8: Dense disparity map corresponding to the cube stereo pair obtained using κ .

Simon Baker and Kristin Dana for reading an initial version of this paper.

References

- [1] M. Alvo and P. Cabilio. Rank correlations and the analysis of rank-based experimental designs. In M. A. Fligner and J. S. Verducci, editors, *Probability Models and Statistical Analyses for Ranking Data*, pages 141–156. Springer-Verlag, 1992.
- [2] P. Aschwanden and W. Guggenbuhl. Experimental results from a comparative study on correlation-type registration algorithms. In Forstner and Ruweidel, editors, *Robust Computer Vision*, pages 268–289. Wichmann, 1993.
- [3] D. N. Bhat and S. K. Nayar. Ordinal measures for visual correspondence. Technical Report CUCS-009-96, Columbia Univ. Center for Research in Intelligent Systems, 1996.
- [4] R. C. Bolles, H. H. Baker, and M. J. Hannah. The JISCT stereo evaluation. *Proceedings of the ARPA Image Understanding Workshop*, pages 263–274, 1993.
- [5] D. E. Critchlow. *Metric Methods for Analyzing Partially Ranked Data*. Springer-Verlag, 1985.
- [6] R. A. Gideon and R. A. Hollister. A rank correlation coefficient. *Journal of the American Statistical Association*, 82(398):656–666, 1987.
- [7] M. Kendall and J. D. Gibbons. *Rank Correlation Methods*. Edward Arnold, fifth edition, 1990.
- [8] M. Maimone and S. A. Shafer. A taxonomy for stereo computer vision. *Workshop on Performance Characteristics of Vision Algorithms*, April 1996.
- [9] Y. Xiong and S. A. Shafer. Recursive filters for high precision computation of focus, stereo, and optical flow. *Proceedings of the ARPA Image Understanding Workshop*, 1994.
- [10] R. Zabih and J. Woodfill. Non-parametric local transforms for computing visual correspondence. *Proceedings of the European Conference on Computer Vision*, pages 151–158, 1994.