Ordinal Measures for Visual Correspondence

Dinkar N. Bhat and Shree K. Nayar

CUCS-009-96

bhat@cs.columbia.edu
nayar@cs.columbia.edu

Department of Computer Science
Columbia University
New York, N.Y. 10027

February 1996

Submitted to "IEEE Transactions on Pattern Analysis and Machine Intelligence".
A portion of this report will appear in "Proceedings of Computer Vision and Pattern Recognition", San Fransisco, June 1996.

Ordinal Measures for Visual Correspondence

Abstract

We present ordinal measures of correlation 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 defining a novel distance metric between the rank permutations of windows, 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. Since ordinal coefficients are non-parametric, confidence levels can be attached to them even in the absence of statistical structure to data, for instance, when a window straddles a depth discontinuity. We have developed a simple algorithm for efficient implementation of ordinal correlation coefficients. 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 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. It is a challenge for practical systems to produce accurate depth maps in such settings. The lack of robustness in dealing with depth discontinuities, occlusion, and specular reflection was clearly noted in [Bolles *et al.*-1993] while evaluating operational stereo systems. The same issues also arise in motion estimation, however, for the sake of clarity, we will uniformly discuss them in the context of robust stereo design.

Stereo matching algorithms have been classified into two categories: *area-based* and *feature-based* [Barnard and Fischler-1982]. Area-based schemes have been popular in photogrammetry [Panton-1978] and computer vision [Okutomi and Kanade-1992] since they produce dense depth maps without requiring explicit surface reconstruction. Feature-based methods, in contrast, obtain depth only at feature locations like edges and corners which are often sparse. The heart of any area-based method lies in the similarity criterion used that determines optimal statistical correlation between corresponding regions in stereo images. 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 two 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). [1]. Resorting to a large window to obtain higher data consistency is not a panacea since the recovered depth boundary would be blurred.

- **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) A stereo operator must correctly identify occlusion regions by reporting that no match can be found for points therein.

- **Noise:** Noise is caused during image sensing and digitization. Image pixel values can vary even within regions that are smooth and textureless in the original scene due to the limited counting statistics for the photons, or because of electronic imperfections. A reliable stereo measure must tolerate low signal-to-noise ratio which is common with low resolution sensors.

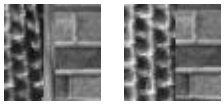
- **Specular reflection:** Corresponding point intensities are not identical in the presence of specular reflection, the specular intensity at any scene point being dependent

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

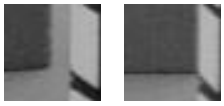
on the viewing direction [Torrance and Sparrow-1967]. As a result, corresponding windows can be quite dissimilar and their correlation can be low (see Figure 1c). Stereo algorithms must allow for deviations from the Lambertian model, specular reflection being the most dramatic of such aberrations.

- **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). While matching using a small window is less susceptible to the effects of window distortion, a low image signal-to-noise ratio would mandate the use of a larger window. Choosing the appropriate window size depends on local terrain curvature which is not known *apriori* and is difficult to model [Okutomi and Kanade-1992]. 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?

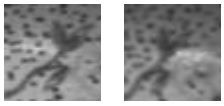
- **Camera Parameter Variations:** If two different cameras are used for binocular stereo, then the camera aperture, bias reference and gamma factors could be different between them. Therefore, sensor outputs could be unequal even in the ideal Lambertian case.



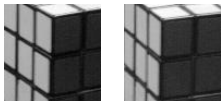
(a) Depth Discontinuity



(b) Occlusion



(c) Specular Reflection



(d) Projective Distortion

Figure 1: Illustration of different phenomena which affect window-based stereo matching. In pictures (a), (c), (d) identically sized windows are shown around corresponding pixels - the center pixels. (a) Since the center pixel is located on a depth discontiuity, the windows represent different surface locations, (b) Due to occlusion, pixels in the left window are not visible in the right, 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.

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 ([Critchlow-1985], [Alvo and Cabilio-1992], [Gideon and Hollister-1987]) 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. 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. Ordinal correlation measures are based on the rank permutations within windows rather than absolute intensity data. Well-known ordinal measures include the Kendall's τ and the Spearman's ρ [Kendall and Gibbons-1990]. 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, like bivariate normality. Thus, confidence thresholds for matching can be established to identify occlusion regions even in the presence of data contamination. We present a simple computationally economical algorithm to evaluate the measures. Experiments with real images and comparison with existing matching methods suggest their superiority.

2 Image Matching Approaches

Common window-based intensity matching approaches can be classified into categories listed below:

- **Linear correlation methods**: Most currently used stereo matching methods belong to this category; for instance those based on the sum of squared differences (*SSD*) [Okutomi and Kanade-1992] and cross correlation [Svedlow *et al.*-1978]. Let I_1 and I_2 represent intensities in two windows i.e we have n tuples $(I_1^1, I_2^1), \ldots, (I_1^n, I_2^n)$, n depending on the size of the window used. The quantity $SSD = \sum_{i=1}^{n}(I_1^i - I_2^i)^2$ 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:

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

  where $\overline{I}_1$ and $\overline{I}_2$ represent the corresponding sample means. Like *SSD* it appraises the degree of linearity between the samples being compared. It is related to the

least squares line of fit [Conover-1980] by:

$$NCC^2 = 1 - \frac{\sum_{i=1}^{n}(I_1^i - (aI_2^i + b))^2}{n\sigma^2}$$

where a, b are regression constants and σ^2 is the sample variance of I_2. When $a = 1, b = 1$, i.e there is no difference in scale or shift between the random variables, the relation between NCC and SSD follows immediately. Hence, NCC has similar properties to SSD. The absolute value of NCC lies between 0 and 1, and a value of 1 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, the following deficiencies restrict the use of both measures:

- In the presence of depth discontinuities, only part of the data is valid for cross correlation, and the rest can be regarded as outliers. The outliers should either be detected and discarded, or the similarity measure used should not be susceptible to their presence. While the former option is not often feasible, linear correlation and SSD are fragile in the presence of outliers [Black-1992]. In fact, the breakdown point of these measures is 0 since a single data value can distort them arbitrarily.
- As mentioned earlier, intensities in corresponding windows are not identical in the presence of specular reflection. Further, corresponding intensities are also not related linearly; thus NCC could be a poor estimator of correspondence. Bhat and Nayar [Bhat and Nayar-1995] formulated an approach to determine optimal stereo configurations – binocular and trinocular – such that intensity differences at corresponding points is limited (in at least one stereo pair, for the case of trinocular configurations), while depth resolution is maximized. The optimal configuration parameters were determined independent of surface normal and light source direction. Stereo images obtained using such a configuration could be reliably matched using correlation since the variation between corresponding windows was restricted. Nevertheless, the threshold which determines the allowable intensity difference at corresponding points had to be chosen empirically.
- Choosing a confidence threshold for matches (or mismatches) has to be empirical [Ching *et al.*-1993], [Ito and Ishii-1986][2]. The reason is that a significance level cannot be assigned to a correlation value without explicit distributional assumptions about the window data samples. Such statistical assumptions cannot be made especially when intensity distribution in a window results from different surfaces. For example, when a window corresponds to two surfaces with different albedo variations.

[2]A notable exception being [Huttenlocher and Jaquith-1995] who estimates probability of correct matches while using the Hausdorff distance.

– Unless image noise is additive and Gaussian, the least squares distance is not optimal for measuring linearity between random variables. Since image noise is often far more complicated than a Gaussian process, the line of regression obtained by *SSD* can be quite unsatisfactory.

- **Robust statistical methods**: Since the Euclidean norm is highly sensitive to outliers, robust statistical methods have been developed which seek to minimize their effects. The least squares criterion is replaced by:

$$E = \sum_{i=1}^{n} (\rho_\sigma(I_1^i - I_2^i))$$

where ρ_σ is a robust estimator, and the objective is to minimize E. The common characteristic of robust estimators is that they cause outliers to contribute less weight compared to inliers. The Lorentzian estimator ρ_σ:

$$\rho_\sigma(x) = log(1 + \frac{1}{2}(\frac{x}{\sigma})^2) \tag{1}$$

has been used for motion estimation [Black-1992]. Many other estimators have been defined, all of which have a common parameter that needs to be set beforehand: the point at which measurements must be considered outliers (for example, in the Lorentzian, σ determines the threshold). Too high a value for the parameter can cause the estimator to behave like *SSD* and too low a value can cause mismatches because the influence of valid data would reduce correspondingly. Setting this threshold parameter is not an easy problem and often has to be chosen based on experience. Further, it can vary with scenario depending on image contrast and noise level. It is desirable to have a universal measure of correlation independent of absolute intensity scale and experimental conditions.

- **Image transform-based methods**: These methods are based on comparing stereo images transformed using local window operators. Kories and Zimmerman's [Kories and Zimmerman-1986] *monotonicity operator*, and Zabih's [Zabih and Woodfill-1994] *rank transform* fall in this category. The latter defines rank (of the center pixel P in a window W) as:

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

where $|.|$ refers to cardinality. The stereo images are fully transformed using the above operator and the resulting mappings are compared using *SSD* or *NCC*. The advantage of these schemes are that correlation of rank transformed images is not dependent on absolute gray values, and hence relatively insensitive to data outliers. However, they depend quite heavily on the center pixel which is undesirable. The necessity to use linear correlation operators on the transformed images partially defeats the purpose of these schemes. We experimentally compare our approach with this scheme.

Methods have also been developed to match local intensity gradients instead of raw intensity values [Scharstein-1994]. However, their performance tends to be poor when gradient information is not reliable.

- **Filter-based methods**: Image correspondence proceeds in the frequency domain, after convolving the images with suitable band-pass filters like Gabor and hypergeometric filters[Sanger-1988], [Xiong and Shafer-1994]. These methods obtain subpixel disparity without resorting to *ad hoc* techniques like local interpolation of correlation values [Tian and Huhns-1986]. They are less susceptible to bandlimited noise and lighting differences between images. However, these algorithms work poorly at depth boundaries and occluded areas [Sanger-1988], and cannot deal with specular reflection satisfactorily.

In addition, some algorithms fall into two categories, for example, the method in [Xiong and Shafer-1994] typically uses *SSD* to find matches to pixel resolution and then uses hypergeometric filters to obtain subpixel accuracy. However, the original problem of locating matches with high confidence remains unresolved due to previous mentioned disadvantages of *SSD*.

3 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.

3.1 Motivation

As noted earlier, ordinal measures are more robust compared to linear correlation measures. To see this, consider the following example of a 3×3 reference window M with intensity I_1:

$$M$$

$$\begin{matrix} 10 & 30 & 70 \\ 20 & 50 & 80 \\ 40 & 60 & 100 \end{matrix}$$

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

$$\begin{array}{ccc} 1 & 3 & 7 \\ 2 & 5 & 8 \\ 4 & 6 & 9 \end{array} \qquad \begin{array}{ccc} 1 & 3 & 7 \\ 2 & 5 & 8 \\ 3 & 6 & 9 \end{array}$$

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 1. 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. We now formally introduce the concepts underlying ordinal measures 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 [Critchlow-1985]. Once a distance metric $d(\pi_1, \pi_2)$ is defined, a coefficient of correlation α can be obtained as:

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

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:

$$d_h(\pi_1, \pi_2) = \sum_i (|sgn(\pi_1^i - \pi_2^i)|) \tag{3}$$

where

$$sgn(x) = \begin{cases} 1 & \text{if } x > 0 \\ 0 & \text{if } x = 0 \\ -1 & \text{if } x < 0 \end{cases}$$

For the Hamming distance, $M = n$. The Kendall's τ and the Spearman's ρ can be expressed using distance metrics although it is seldom done. Let $d_s(\pi_1, \pi_2)$ and $d_k(\pi_1, \pi_2)$

denote distance metrics as defined below:

$$\begin{aligned} d_k(\pi_1, \pi_2) &= \sum_{i<j} \frac{1}{2}(1 - [sgn(\pi_1^i - \pi_1^j)][sgn(\pi_2^i - \pi_2^j)]) \\ d_s(\pi_1, \pi_2) &= \sum_i (\pi_1^i - \pi_2^i)^2 \end{aligned} \tag{4}$$

d_k estimates the number of discordant pairs, i.e pairs in one sample whose ordering do not agree with the corresponding pairs in the other. d_s is simply the Euclidean distance between the rankings. The maximum values of d_k and d_s are $M = \frac{n(n-1)}{2}$ and $M = \frac{n(n^2-1)}{3}$, respectively. By substituting d_k and d_s, and their corresponding maximum values in equation 2, we obtain Kendall's τ and Spearman's ρ, respectively. The above analysis strictly holds when ties are absent in each ranking. In the case of ties, some modification must be made in defining the distance metric; however, the approach for defining correlation remains unaltered.

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 3.1. As a result, ordinal measures like the Kendall's τ and Spearman's ρ are inadequate.

3.3 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:

$$\begin{aligned} s^i &= \pi_2^k \\ k &= (\pi_1^{-1})^i \end{aligned} \tag{5}$$

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 $u = (1, 2, \ldots, n)$.

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 $s^j, j = 1, \ldots, i$ greater than i. Formally,

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

$$= \quad i - \sum_{j=1}^{i} J(s^j \leq i) \tag{6}$$

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 $\mathbf{d_m}(s,u)$. 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 $\mathbf{d_m}(s,u) = (0,0,\ldots,0)$. 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 (see Appendix B). Now, a measure of correlation $\kappa = \kappa(I_1, I_2)$ is defined using equation 2 as:

$$\kappa(I_1, I_2) = 1 - \frac{2\, max_{i=1}^{n} d_m^i}{\lfloor \frac{n}{2} \rfloor} \tag{7}$$

If I_1 and I_2 are perfectly correlated ($s = u$), then $\kappa = 1$. It falls to -1 when (I_1, I_2) are uncorrelated. κ has desirable properties of a correlation coefficient, namely:

- 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 $\kappa(I_1, I_2) = \kappa(I_2, I_1)$ (see Appendix C). Therefore, either stereo image can be used as reference.
- $\kappa(f(I_1), h(I_2)) = \kappa(I_1, I_2)$ 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 two different cameras are used for binocular 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 two 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 two sensor outputs is lost. However, $|\kappa|$ remains at 1 because $(I_2)^t$ is a strictly increasing or decreasing function[3] of I_2 depending on t, and hence does not change its ranking.

[3]Note that this 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.

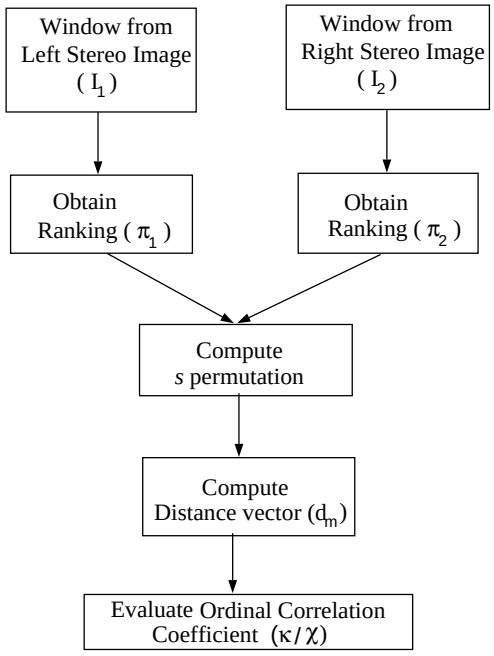


Figure 2: Flow-chart depicting the procedure involved in computing κ or χ

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

$$\chi(I_1, I_2) = 1 - \frac{2\, d_m^{mid}}{\lfloor \frac{n}{2} \rfloor} \tag{8}$$

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. Theoretically, deviation at any index of the distance vector could be used to define a correlation coefficient. However, the largest range of deviation – $[0, \lfloor \frac{n}{2} \rfloor]$ – occurs at the middle index position, the maximum occurring in the case of perfect negative correlation (see Appendix B). As a result, the discriminatory power of χ is higher, an useful property for stereo matching. Note, it does not imply that the maximum component value of each possible distance vector always occurs at the middle index position, in which case the definition of κ would be redundant. Before concluding this section, we summarize the procedure involved in computing our measures, with a flow-chart shown in figure 2.

3.4 An Example

We will illustrate the evaluation of the two coefficients κ, χ using the following example. Consider the following pair of 3×3 intensity windows which need to be compared:

10	30	75	15	30	60
20	50	85	20	50	90
45	60	95	45	70	85

The data in these windows are arbitrarily chosen. The permutation s is given by: $s = (1, 2, 3, 4, 5, 7, 6, 9, 8)$, and the distance vector $\mathbf{d_m}$ evaluates to $(0, 0, 0, 0, 0, 1, 0, 1, 0)$. Therefore, $max_{i=1}^{9}(d_m^i) = 2, \kappa = 0.5$, and $\chi = 1.0$. It can be seen that κ is more sensitive than χ for small deviations. For small samples, the measures can easily be computed by hand, as seen from the above example. For larger samples, obtaining $\mathbf{d_m}$ can be computationally expensive if the naive method of searching through s for each index, is used. We present a more efficient algorithm when computational issues are discussed.

3.5 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 3.1. κ remains at 1 when the 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 [Kendall and Gibbons-1990] which fall steeply to 0.556 and 0.4, respectively[4]. 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 – after all, specular reflection and discontinuities are "valid" physical phenomena) data. It is worth noting that the Hamming distance (equation 3) which can be used to define a measure of correlation, is not one of choice since it is sensitive to rank distortion.

3.6 Comparison

We compare our measures with SSD ,NCC and Zabih's rank transform. We have not tabulated the results based on the Lorentzian estimator (equation 1) because of the additional free parameter threshold σ. We use the [Aschwanden and Guggenbuhl-1993] test suite 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. Therefore, the matching location for any pixel remains the *same* between images, and there is no question of occlusion. Further, there are no depth discontinuities.

In Figure 3, salt and pepper noise[5] was added to the right image. Notice the significant degradation of image quality. 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. However, in Zabih's method, the rank transformed images are used for correlation instead of original intensity images; therefore, to keep the comparison fair, the test for sufficient texture is performed using the same variance threshold while transforming the images. As noted earlier, disparity evaluation is a non-issue. However, to simulate stereo matching, we use a search range R of ± 10 pixels. Matches are established for a region of size 100×100 $(= 10^4)$ pixels of the left reference image.

[4]It is futile to compare absolute values of correlation measures since each of them have different interpretations. Only if explicit distributional assumptions of the sample population can be made, like bivariate normality, can we relate different quantities.

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

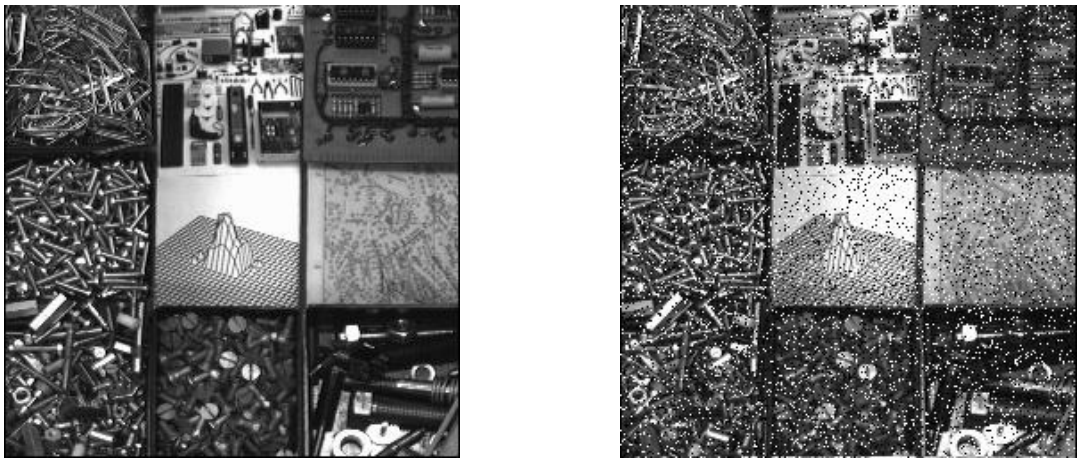


Figure 3: Image pair with salt and pepper noise added to the right image. This pair is used to test κ and χ with other measures. Locating right matches with such significant image degradation is an imposing computational challenge.

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. Each measure is denoted by the appropriate abbreviation. 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. While χ does not perform as well as κ, it is computationally less expensive which makes it attractive. All measures did better with increasing window size.

The two measures were next tested on a random dot stereogram (see Figure 4) and compared with the other methods. The random dot stereo pair, each image of size 64x64 pixels, depicts a square (size:20x20 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. Note that there is no window distortion since the surface is fronto-planar.

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

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.

than to use different thresholds for different similarity measures in order to determine mismatches. The results (number of mismatches) are shown in Table 2.

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*.

4 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. The question then is: What rank should be assigned to the set of tied data values? The use of *mid-ranking*[6], a

[6]All tied values are assigned ranks identical to the mean of the ranks that they would have had if their values had been slightly different. The mid-rank could be an integer or a fraction.

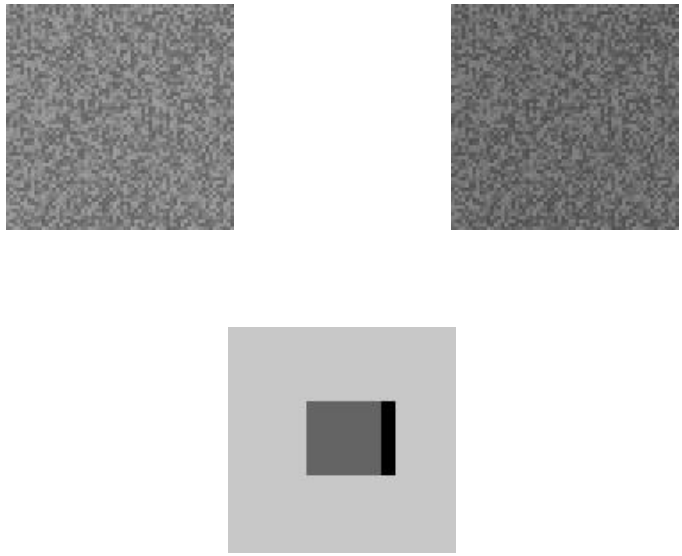


Figure 4: Left and right stereo images of a random dot stereogram are shown on the top row. The right image is 10% darker than the left. They represent a square moving against a stationary textured background. On the bottom row, true disparity levels with respect to the left image are shown in different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

popular technique [Kendall and Gibbons-1990] in rank tests, does not help in this case since it assigns identical ranks to all data values within a tied group. Hence, the definition of s in equation 5 remains unclear. In general, if there are g groups of tied data, each group containing $n_j, j = 1, ..., g$ elements, then the total number of rankings possible is $n_1!n_2!..,n_g!$. 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 as can be seen from the experiments.

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.

- *Confidence Thresholds*: As noted in section 2, the linear correlation coefficient has no clear meaning, i.e it is not possible to attach a significance level[7] to it without explicit distributional assumptions about the sample data. These assumptions cannot be made, especially for samples containing inhomegenous data. Therefore, an algorithm cannot reject or accept a match with any confidence. Typically an *ad hoc* threshold based on experience is chosen. This is true of distance measures like *SSD* and the Lorentzian estimator too.

  On the other hand, the significance level of our measures can be tested, as explained below. Ideally, we would like to know the distribution of κ (or χ) as a function of the sample size n, under a null hypothesis of there being no correlation between two data sets. In our problem, accepting the null hypothesis implies rejecting the match obtained, and *vice versa*. If the distribution was known, then we would be able to formulate a look-up table containing the threshold value of κ, for different window sizes, at which to reject a match with some chosen confidence level[8]. For example, a table for κ would look like:

[7] In this paper, the significance level denotes the maximum probability at which the 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 called the confidence level.

[8] This confidence level indicates how conservative or liberal is the user. A conservative would choose a high confidence level, like 99%, to reject the null hypothesis. It may be argued as to why choosing an arbitrary confidence level is better than choosing an arbitrary threshold for the correlation coefficient. The reason is that a confidence level is a universal estimate; it does not depend on image or sensor quality, and experimental conditions.

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

A match between two windows would be rejected, if the κ between them is below the appropriate threshold. Unfortunately, the exact distribution for κ as a function of sample size is not known.

However, we can obtain the above table using computer simulation [Gideon and Hollister-1987]. Under the null hypothesis of there being no correlation between the rankings, s can be any member of the universal permutation set S_n with equal probability of $\frac{1}{n!}$. For each s we evaluate κ, tabulate the frequency f of each occurrence of κ, the probability $\frac{f}{n!}$ of κ and its cumulative probability distribution P. Then we determine c in the equation: $P(\kappa \geq c) = 1 - t$, where c and t refer to the confidence threshold and confidence level in the above table. For large sample sizes, the number of s permutations can be prohibitively large to enumerate. Hence, we do not generate all s, but only a sufficiently large number.

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 obtained increased marginally from 54 to 57. This small increase could be attributed to the errors associated with hypothesis testing (see [Conover-1980]). We also used this threshold with a real stereo image pair which is discussed in section 6. Nevertheless, detailed experimental analysis has to performed before such thresholds can be used for practical stereo work.

- *Relationship to the Kolmogorov-Smirnov statistic*: κ bears a relationship to the Kolmogorov-Smirnov (K-S) statistic [Conover-1980] which is used for testing goodness of fit between two distribution functions. The problem of fitting can be stated as: Given a random sample $X_1, X_2, \ldots, X_n$ with some unknown distribution function $F(x)$, can we measure the "closeness" of $F(x)$ from a hypothesized distribution function $H(x)$? Since $F(x)$ is unknown, the *empirical distribution function* $S(x)$ is used for comparison with $H(x)$. $S(x)$ is defined at each $x = X$ as the fraction of X_is that are less than or equal to X. It has been shown that $S(x)$ converges to $F(x)$ as the sample gets larger. The K-S statistic T_n is defined by:

$$T_n = max_x|H(x) - S(x)|$$

In other words, T_n is the upper bound of point wise differences $|H(x) - S(x)|$. The exact distribution of T_n is known, and hence confidence levels can be established for testing goodness of fit.

T_n has a close functional resemblance to $max_{i=1}^{n} d_m^i$ as explained below. From equation 6,

$$\begin{aligned} max_{i=1}^{n} d_m^i &= max_{i=1}^{n}(i - \sum_{j=1}^{i} J(s^j \leq i)) \\ &= max_{i=1}^{n}(H_a(i) - S_a(i)) \end{aligned}$$

$H_a(i)$ and $S_a(i)$ are analogues of $H(x)$ and $S(x)$. Note the close similarity in definition between the $S(x)$ and $S_a(i)$. The difference is that in our case, we are not hypothesizing a distribution function for s, rather finding how close it is to the identity permutation.

5 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. In order to establish correspondence for a pixel in the master image, we search through a range R in the second image. Therefore, the cost of computing $\mathbf{d_m}$ could be $O(Rn^2)$. This is in addition to the sort operations to perform ranking. 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 $i = 1, \ldots, n$, and thus we obtain a $O(n)$ algorithm to calculate the distance vector. It is trivial to compute $max_{i=1}^{n}(d_m)$. An algebraic proof is also sketched in Appendix A. Mathematically, our algorithm states that,

$$d_m^{i+1} = d_m^i - \sum_{j=1}^{i} J(i < s^j \leq i+1) + \sum_{j=i}^{i+1} J(s^j > i+1) \tag{9}$$

The only 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

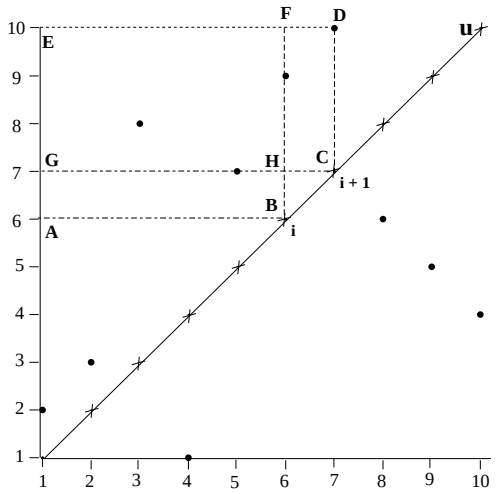


Figure 5: Scatter diagram of a bivariate rank distribution is shown. The large dots represent the ranking of I_2 with respect to I_1, i.e their ordinate values when read along the positive x axis give s as in equation 3. The crosses on the straight line at a slope of $45°$ refer to the identity permutation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

every time it slides across through one pixel distance within the search range R, if we use heap-sort in which data is maintained as a heap tree [Press *et al.*-1989]. 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, 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. Note that counting sort is effective with tied data values too. 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. When sensors of higher intensity resolution are used ($k > 10$), then the earlier algorithm would be of choice.

It should be noted that computation of κ and χ is much better than that of the Kendall's τ which is $O(n^2)$. It is asymptotically identical to 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.

6 Experiments

Ideally, we would like to compare our measures with others using dense ground truth, but unfortunately such data is lacking [Bolles *et al.*-1993]. In this paper, we present three experiments which qualitatively substantiate the results presented in section 3.6. In all cases, κ is used for matching.

The first is a stereo image pair in figure 6 from the Calibrated Imaging Laboratory at CMU [Maimone and Shafer-1996]. 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. 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. 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 stereo image pairs in the same sequence.

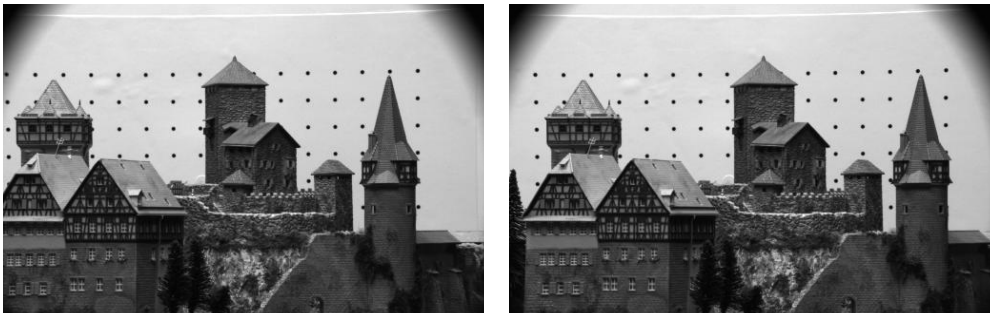
Next, we use a stereo pair of a densely textured cube[9](see Figure 7) with disparity variation in the range $[25-50]$ pixels. The issues are to obtain accurate disparity in spite of the significant projective distortion, and to match correctly at the edges. The window size is 9×9. The resulting dense disparity map is shown in figure 8 which is accurate. To verify, we compared the obtained disparity by the plane-fit error method [Xiong and Shafer-1994], and the result is nearly 100% accurate (upto pixel accuracy).

Finally, we use an outdoor stereo image pair (Figure 9) (from SRI) which was captured using a laterally moving camera. The disparities are in the range $[0-28]$ pixels with respect to the left image. A window size of 7×7 is used for matching. In this case, we do not use back-matching, and instead we use a confidence threshold of 0.33 for κ as explained in section 4. The resulting disparity map is shown in Figure 10 and is quite accurate. It must be noted that the result is qualitative since there is no ground truth.

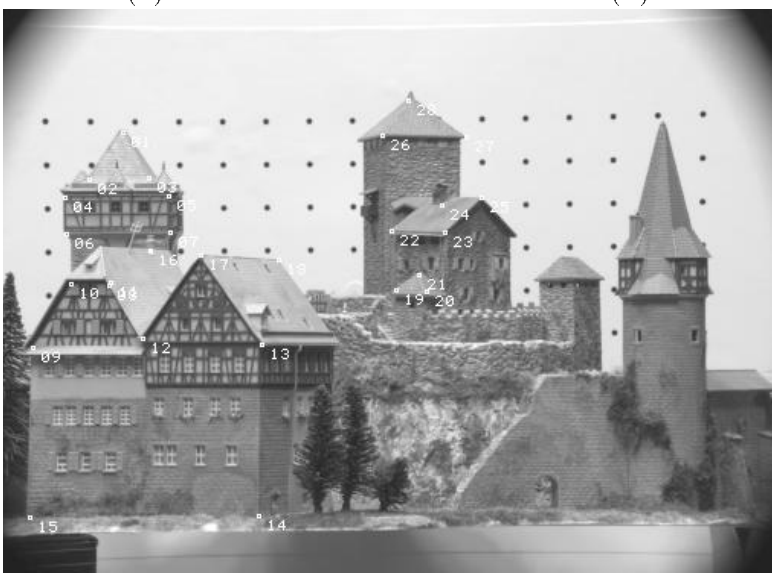
7 Discussion

We have 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

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



(a) (b)



(c)

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.

distribution. These textures may represent non-corresponding surface patches, therefore, an ordinal measure like ours will incorrectly report perfect 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. However, in practice, the above discussed case seems more pathological rather than the norm.

An issue which we did not deal with is that of obtaining subpixel disparity. Using ordinal measures, we could locate upto pixel accuracy. Two approaches could be adopted to achieve this: a) local interpolation of correlation values, and b) an iterative scheme based on non-parametric regression. Iterative approaches based on least squares regression between windows that correspond to pixel accuracy, are used often [Lucas

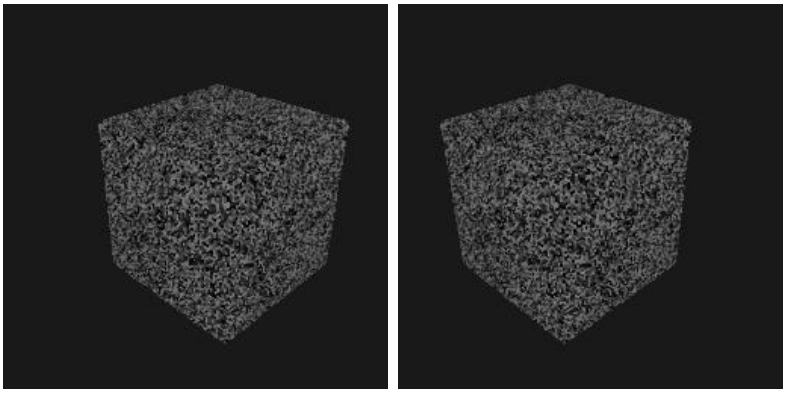


Figure 7: Stereo pair of a densely textured cube from the University of Illinois (by Bill Hoff)

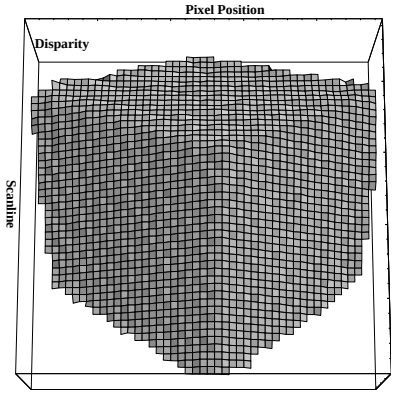


Figure 8: Dense disparity map corresponding to the cube stereo pair obtained using κ.

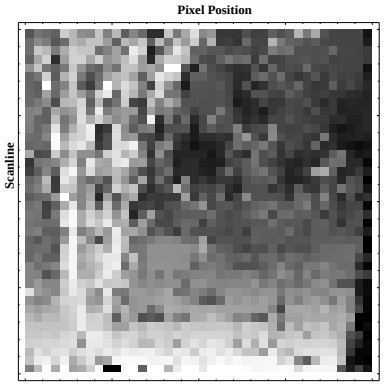
and Kanade-1981]. Non-parametric regression techniques may offer a good solution since they do not make strong model assumptions between the regression variables, namely, the intensities in windows.

A Computation of the Distance Vector

In section 5, we sketched a method to compute the distance vector using a geometrical construction. Here we will show the same algebraically and prove it is correct using induction. The distance vector $\mathbf{d_m}$ is given by equation 6. Our algorithm states that:

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

Figure 9: Stereo image pair of the SRI "tree" sequence.



(a) (b)

Figure 10: Density plot of disparity corresponding to the "tree" stereo pair, obtained using κ.

Since s^i, $i = 1, \ldots, n$ can only takes discrete values of $1, \ldots, n$, the above equation can be rewritten as:

$$d_m^{i+1} = d_m^i - (\sum_{j=1}^{i} J(s^j = i+1)) + J(s_{i+1} > i+1) \tag{10}$$

In computing d_m^{i+1}, only $J(s_{i+1} > i+1)$ has to be calculated, if $\sum_{j=1}^{i} J(s^j = i+1)$ was stored while evaluating d_m^i. Therefore, we have a $O(n)$ algorithm.

To prove that this algorithm correctly obtains the distance vector $\mathbf{d_m}$, we will use induction. The above equation is true for $i = 0$. If it is true for $i = k$ (hypothesis), i.e

$$d^{k+1} = d^k - (\sum_{j=1}^{k} J(s^j = k+1)) + J(s_{k+1} > k+1)$$

we will show it is true for $i = k + 1$. Consider the equation:

$$P = d^{k+1} - (\sum_{j=1}^{k+1} J(s^j = k + 2)) + J(s_{k+2} > k + 2)$$

From our hypothesis and equation 6, the above equation can be rewritten as:

$$P = \sum_{j=1}^{k} J(s^j > k) - (\sum_{j=1}^{k} J(s^j = k+1)) + J(s_{k+1} > k+1) - (\sum_{j=1}^{k+1} J(s^j = k+2)) + J(s_{k+2} > k+2)$$

which simplifies to $P = \sum_{j=1}^{k+2} J(s^j > k + 2)$. Thus $P = d^{k+2}$, hence proved by induction.

B Distance Vector under Negative Correlation

The maximum value that a distance vector component can take must occur when the two rankings being compared are negatively correlated (κ or χ should be equal to -1). When the rankings are negatively correlated, the s permutation (Equation 5) is identical to $[n, n-1, \ldots, 1]$, as shown in Figure 11.

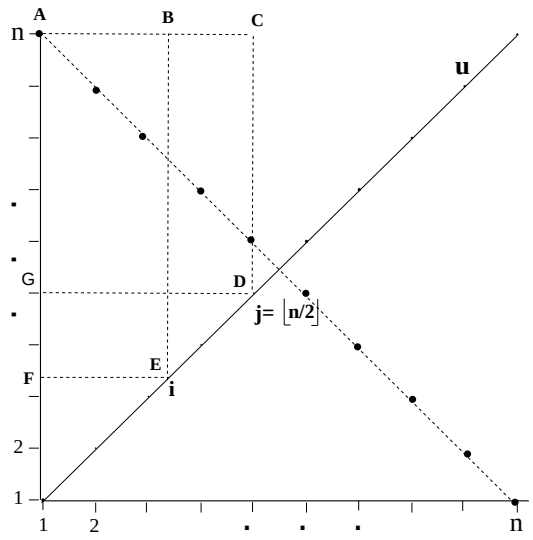


Figure 11: Scatter diagram of a perfect negatively correlated bivariate rank distribution. The negatively sloped dotted line indicates the s permutation. i and j correspond to the index of the distance vector.

The distance vector component d_m^i is equal to number of elements of the s vector lying in the area bounded by $(x = 0, x = i, y = i, y = n)$, i.e the region FEBA. This bounded area becomes a maximum when $i = j = \lfloor \frac{n}{2} \rfloor$ (the region GDCA). This region of maximum area includes the maximum number of s elements. Hence, the distance vector component d^j, $j = \lfloor \frac{n}{2} \rfloor$ is a maximum and is identical to $\lfloor \frac{n}{2} \rfloor$.

C Symmetric nature of κ

In this section, we will sketch a proof using induction for the symmetry of κ, i.e $\kappa(I_1, I_2) = \kappa(I_2, I_1)$. Equivalently, we have to show that $max_{i=1}^{n} d_m^i$ is invariant. We will prove a stronger result, namely, that the distance vector itself is invariant.

Let s_1^i and s_2^i denote the s permutations used to compute $\kappa(I_1, I_2)$ and $\kappa(I_2, I_1)$, respectively. Let $\mathbf{d_1} = \mathbf{d_m}(s_1, u)$ and $\mathbf{d_2} = \mathbf{d_m}(s_2, u)$ denote the corresponding distance vectors. Then from equation 6,

$$\begin{aligned} d_1^i &= \sum_{j=1}^{i} J(s_1^j > i) \\ d_2^i &= \sum_{j=1}^{i} J(s_2^j > i) \end{aligned} \tag{11}$$

We need to prove $P(i) : d_1^i = d_2^i, \forall i$. To do so, we use induction. For $i = n$, $d_1^i = d_2^i = 0$, hence $P(n)$ is true. Now we prove $P(k-1)$ holds when $P(k)$ is true. From the induction hypothesis,

$$\sum_{j=1}^{k} J(s_1^j > k) = \sum_{j=1}^{k} J(s_2^j > k)$$

To prove $P(k-1) : \sum_{j=1}^{k-1} J(s_1^j > k-1) = \sum_{j=1}^{k-1} J(s_2^j > k-1)$. Using the induction hypothesis, the LHS of $P(k-1)$ can be written as:

$$\begin{aligned} \sum_{j=1}^{k-1} J(s_1^j > k-1) &= \sum_{j=1}^{k} J(s_1^j > k) - J(s_1^k > k) + \sum_{j=1}^{k-1} J(s_1^j = k) \\ &= \sum_{j=1}^{k} J(s_2^j > k) - J(s_1^k > k) + \sum_{j=1}^{k-1} J(s_1^j = k) \end{aligned} \tag{12}$$

We show that $-J(s_1^k > k) + \sum_{j=1}^{k-1} J(s_1^j = k)$ is equal to $-J(s_2^k > k) + \sum_{j=1}^{k-1} J(s_2^j = k)$. Let Q_1, Q_2, Q_3, Q_4 represent Boolean expressions as below:

$$\begin{aligned} Q_1 &: s_1^k > k \\ Q_2(j) &: s_1^j = k,\ j \in [1, k-1] \\ Q_3 &: s_2^k > k \\ Q_4(j) &: s_2^j = k,\ j \in [1, k-1] \end{aligned}$$

Note that $\sum_{j=1}^{k-1} J(Q_2(j))$ and $\sum_{j=1}^{k-1} J(Q_4(j))$ can attain a maximum value of 1 (= True) since only one j can satisfy the the corresponding Boolean expression. Now, the following logic assertions can be easily deduced:

- If $\exists j, Q_2(j)$ is true, then Q_3 is false,
- If $\forall j, Q_2(j)$ is false, then Q_3 is true,
- If Q_1 is true, then $\forall j, Q_4(j)$ is false,
- If Q_1 is false, then $\exists j, Q_4(j)$ is true.

From symmetry, other assertions can be derived. Therefore, the following logic table is obtained:

	Q_1	Q_2	Q_3	Q_4
1	True	False	True	False
2	True	True	False	False
3	False	True	False	True
4	False	False	True	True

Using the above table, equation 12 can be expressed as:

$$\begin{aligned}\sum_{j=1}^{k-1} J(s_1^j > k-1) &= \sum_{j=1}^{k} J(s_2^j > k) - J(s_2^k > k) + \sum_{j=1}^{k-1} J(s_2^j = k) \\ &= \sum_{j=1}^{k-1} J(s_2^j > k-1)\end{aligned}$$

Therefore, $P(k-1)$ holds when $P(k)$ is true. Since $P(n)$ is true, the relation $d_1^i = d_2^i$ holds for all i. From the above proof, it follows that $\kappa(I_1, I_2) = \kappa(I_2, I_1)$.

References

[Alvo and Cabilio, 1992] 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.

[Aschwanden and Guggenbuhl, 1993] 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.

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

[Bhat and Nayar, 1995] D. N. Bhat and S. Nayar. Stereo in thre presence of specular reflection. *Proceedings of the IEEE Computer Society International Conference on Computer Vision*, pages 1086–1092, 1995.

[Bhat and Nayar, 1996] D. N. Bhat and S. K. Nayar. Ordinal measures for visual correspondence. Technical Report CUCS-009-96, Columbia University, Dept. of Computer Science, February 1996.

[Black, 1992] M. J. Black. *Robust Incremental Optical Flow.* PhD thesis, Yale University, September 1992.

[Bolles *et al.*, 1993] 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.

[Ching *et al.*, 1993] Wee-Soon Ching, Peng-Seng Toh, Kap-Luk, and Meng-Hwa Er. Robust vergence with concurrent detection of occlusion and specular highlights. *Proceedings of the IEEE Computer Society International Conference on Computer Vision*, pages 384–394, 1993.

[Conover, 1980] W. J. Conover. *Practical Nonparametric Statistics.* John Wiley, 1980.

[Critchlow, 1985] D. E. Critchlow. *Metric Methods for Analyzing Partially Ranked Data.* Springer-Verlag, 1985.

[Gideon and Hollister, 1987] R. A. Gideon and R. A. Hollister. A rank correlation coefficient. *Journal of the American Statistical Association*, 82(398):656–666, 1987.

[Huttenlocher and Jaquith, 1995] D. P. Huttenlocher and E. W. Jaquith. Computing visual correspondence: Incorporating the probability of a false match. *Proceedings of the IEEE Computer Society International Conference on Computer Vision*, pages 515–522, 1995.

[Ito and Ishii, 1986] M. Ito and A. Ishii. Range and shape measurement using three-view stereo analysis. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 9–14, 1986.

[Kendall and Gibbons, 1990] M. Kendall and J. D. Gibbons. *Rank Correlation Methods.* Edward Arnold, fifth edition, 1990.

[Kories and Zimmerman, 1986] R. Kories and G. Zimmerman. A versatile method for the estimation of displacement vector fields from image sequences. *Proceedings of the IEEE Workshop on Motion: Representation and Analysis*, pages 101–107, 1986.

[Lucas and Kanade, 1981] B. D. Lucas and T. Kanade. An iterative image registration technique with an application to stereo vision. *Proceedings of the Seventh International Joint Conference on Artificial Intelligence*, pages 674–679, 1981.

[Maimone and Shafer, 1996] M. Maimone and S. A. Shafer. A taxonomy for stereo computer vision. *Workshop on Performance Characterstics of Vision Algorithms*, April 1996.

[Okutomi and Kanade, 1992] M. Okutomi and T. Kanade. A locally adaptive window for signal matching. *International Journal of Computer Vision*, 7(2):1499–1512, 1992.

[Panton, 1978] D. A. Panton. A flexible approach to digital stereo mapping. *Photogrammetric Engg. and Remote Sensing*, 44(12):1499–1512, 1978.

[Press *et al.*, 1989] W. H. Press, B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling. *Numerical Recipes in C.* Cambridge University Press, 1989.

[Sanger, 1988] T. D. Sanger. Stereo disparity computation using gabor filters. *Biological Cybernetics*, 59:405–418, 1988.

[Scharstein, 1994] D. Scharstein. Matching images by comparing their gradient fields. *Proceedings of the Twelfth International Conference on Pattern Recognition*, 1994.

[Svedlow *et al.*, 1978] M. Svedlow, C. D. McGillem, and P. A. Anuta. Image registration: Similarity measure and preprocessing method comparisons. *IEEE Transactions on Aerospace and Electronic Systems*, AES-14(1):141–149, January 1978.

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

[Torrance and Sparrow, 1967] K. E. Torrance and E. M. Sparrow. Theory for off-specular reflection from roughened surfaces. *Journal of the Optical Society of America*, 57:1105–1114, 1967.

[Xiong and Shafer, 1994] 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.

[Zabih and Woodfill, 1994] 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.