

Motion Estimation using Ordinal Measures

Dinkar N. Bhat
Columbia University
450 Mudd Building, NY 10027
bhat@cs.columbia.edu

Shree K. Nayar
Columbia University
450 Mudd Building, NY 10027
nayar@cs.columbia.edu

Alok Gupta
Siemens Corporate Research
755 College Road East, Princeton, NJ 08540
alok@scr.siemens.com

Abstract

We present a method for motion estimation using ordinal measures. Ordinal measures are based on relative ordering of intensity values in a image region called rank permutation. While popular measures like the sum-of-squared-difference (SSD) and normalized correlation (NCC) rely on linearity between corresponding intensity values, ordinal measures only require them to be monotonically related so that rank permutations between corresponding regions are preserved. This property turns out to be useful for motion estimation in tagged Magnetic Resonance Images. We study the imaging equation involved in two methods of tagging and observe temporal monotonicity in intensity under certain conditions through the tags themselves fade. We compare our method to SSD and NCC in a rotating ring phantom image sequence. We present an experiment on a real heart image sequence which suggests the suitability of our method.

1. Introduction

In motion estimation from image sequences, an aim is to determine those displacements which best register pixels in one image frame with corresponding pixels in the next. In this paper, we suggest an approach using ordinal measures that possesses demonstrable robustness in scenes where pixel intensity varies temporally.

Extensive research has been reported in motion estimation and optical flow, and it is impossible to list all related work here for lack of space. A few surveys have appeared in the literature and an excellent recent one is [2]. Regardless of the approach used, almost all methods implicitly or explicitly assume that intensities at corresponding points are identical, which is sometimes called the Lambertian assumption. This assumption is not true either due to temporally changing characteristics of the imaging system, or due to effects such

as specular reflection. A strikingly illustrative example where the constant intensity assumption breaks is *Tagged Magnetic Resonance* imaging which is the main focus of the paper.

Standard cardiac MR imaging produces images in which the tissue has nearly constant brightness, making it impossible to determine image motion. However, by tagging [11], a prespecified spatially varying intensity pattern can be made to appear on the otherwise homogenous tissue. The pattern itself deforms according to the heart motion, and is then imaged by standard MR methods. But, the tag pattern *fades* with time as seen in figure 1. Consequently, algorithms that assume intensity constancy are deficient.

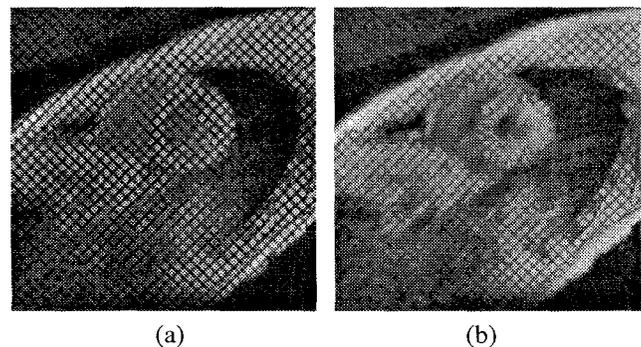


Figure 1. Tagged MR cardiac images at two different instances. Notice the fading of tags in (b) with respect to (a). Although the fading occurs in a non-linear fashion, intensities at corresponding points are monotonically related.

The use of optical flow techniques for cardiac motion analysis was pioneered by Prince and his colleagues. They have closely analyzed several issues including variants to standard

optical flow [11], [9], optimal intensity patterns for tagging [6], and the use of pre-filtering and multi-resolution methods [1]. Prince and McVeigh [11] developed the variable brightness flow algorithm which is a significant improvement over standard optical flow. It requires prior knowledge regarding spatial distribution of tissue and imaging parameters. Gupta and Prince [9] proposed an interesting alternative by adopting Gennert's [7] model of image intensity variation which relates intensity at corresponding points by a multiplier term that smoothly changes. This technique requires estimates, though only approximate, of certain tissue parameters. Therefore, the above schemes may be considered as parametric techniques for motion computation.

Consider the situation when no imaging parameters are available regarding the acquisition of the tagged MRI sequence. Here, recourse to a standard region matching approach or a gradient-based method becomes necessary. Our approach based on *ordinal measures* belongs to the former category. However, unlike the popular sum-of-squared-difference measure (*SSD*) and the normalized correlation coefficient (*NCC*) for region matching which rely on linearity between intensities at corresponding points, ordinal measures only require them to be related monotonically. We carefully analyze the tagged spin-echo MR imaging equation [11] and the echo-planar imaging equation which are used by medical methods for generating tags. From the analysis, we observe that intensity of a point in the tagged images at different times are related by a monotonic function, under certain assumptions. This forms the motivation for use of our measures. We compare our approach to *SSD* and *NCC* using an image sequence with ground truth. We present an experiment on heart images which suggests the suitability of our method.

2. An Ordinal Measure

The similarity criterion used for identifying corresponding windows is an ordinal measure, as in [4] where it was presented in the context of robust stereo matching. A brief review of the measure is given below. Methods based on local transformation of intensity values like those in [12] and [10] would also be interesting to consider.

Let I_1 and I_2 represent intensities in windows of successive images. For the 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. Below, we present a method for defining the distance between rank permutations similar to that reported by [8]. We assume that there are no ties in the data. The method to handle tied values is discussed in [4]. A composition permutation s can be defined as:

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

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 is the ranking of I_2 with respect to that of I_1 . Under perfect positive correlation, s should be identical to the *identity permutation* given by $u = (1, 2, \dots, n)$. By defining a distance measure between s and u , a notion of distance in turn is obtained between π_1 and π_2 . The deviation d_m^i at each s^i , $i = 1, \dots, n$ is defined as:

$$d_m^i = i - \sum_{j=1}^i J(s^j \leq i) \quad (2)$$

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 $\mathbf{d}_m(s, u)$. This measure has its analog in the definition of the Kolmogorov-Smirnov test statistic (see [4] for details). Each component of the distance vector, referenced by its positional index, estimates the number of preceeding elements in s that are out of position. If (I_1, I_2) were perfectly correlated, then $\mathbf{d}_m(s, u) = \mathbf{0}$. The maximum value that any component of the distance vector can take is $\lfloor \frac{n}{2} \rfloor$ which must occur under perfect negative correlation. Now, a measure of correlation $\kappa(I_1, I_2)$ is given by:

$$\kappa(I_1, I_2) = 1 - \frac{2 \max_{i=1}^n d_m^i}{\lfloor \frac{n}{2} \rfloor} \quad (3)$$

If I_1 and I_2 are perfectly correlated ($s = u$), then $\kappa = 1$. It falls to -1 when (I_1, I_2) are perfectly negatively correlated. Figure 2 describes a simple example which illustrates the procedure for computing κ . κ has the following desirable

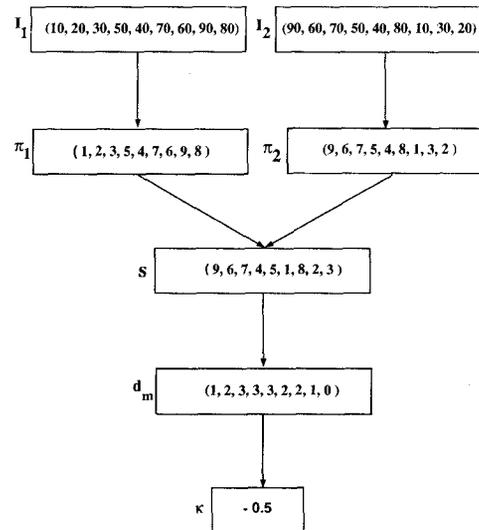


Figure 2. Example illustrating the procedure for computing κ .

properties of a correlation coefficient:

- It is independent of linear scaling and shift between I_1 and I_2 since π^1 and π^2 remain unchanged. This implies independence from camera gain and bias.
- It is symmetrical, i.e $\kappa(I_1, I_2) = \kappa(I_2, I_1)$. Hence, either image can be used as reference.
- $\kappa(f(I_1), h(I_2)) = \kappa(I_1, I_2)$ where f and h are monotonically increasing functions. This property turns out to be very useful in our application. To illustrate this property, consider the case when different cameras are used for stereo. Each sensor output I is related to image irradiance E as:

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

where g is the camera gain, m is the reference bias factor, and γ accounts for image contrast. Since gain and bias account for linear variations which, as noted earlier, do not affect κ , let us assume only gamma variation between the sensors. In other words, let the gains of the sensors be identically 1.0 and the bias factors be 0. Further, 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 as $I_1^i = (I_2^i)^t, \forall i$ where $t = \frac{\gamma_2}{\gamma_1}$. In general, $t \neq 1$, and hence the linearity between the sensor outputs is lost. If $t > 1$, then $I_2^i < I_1^i, \forall i$, and if $t < 1$, then $I_2^i > I_1^i, \forall i$. However, κ remains at 1 because π_2 remains the same as π_1 .

3. Tagged MRI Imaging

To study cardiac motion during systolic contraction, tags are generated by applying a magnetic field in the left-ventricle (LV) at the beginning of contraction (end-diastole). A two-dimensional cross-sectional image of the LV is taken at regular time intervals during motion. As noted earlier, the tags fade with time, i.e the image becomes progressively uniform in intensity. The reason is that the tissues return to their original magnetization state, i.e the state before tags were applied. Temporal tag fading depends upon the properties of the heart tissue being imaged, and the amount of motion of the heart itself. Under the assumption of zero heart motion, the intensity of a point $r = r(x, y)$ in the heart, at time t , is given by the tagged spin-echo MR imaging equation [11] as:

$$\begin{aligned} I(r, t) &= K(r) [\zeta(r) - 1] e^{-\frac{t}{T_1(r)}} \\ &\quad + K(r) [1 - \zeta(r) e^{-\frac{T_R}{T_1(r)}}] \\ K(r) &= D_0(r) e^{-\frac{T_R}{T_2(r)}} \end{aligned} \quad (4)$$

where D_0, T_1 and T_2 (called the spin density, longitudinal and transversal relaxation times, respectively) are determined by the properties of the heart tissue and vary throughout the image. T_E and T_R , (called the echo time and pulse repetition time, respectively) are fixed imaging parameters. $\zeta(r)$

Parameter	Value
D_0 (spin density)	300.0
T_E (echo time)	0.03 s
T_R (pulse repetition time)	10.0 s
T_1 (longitudinal relaxation time)	0.3 s
T_2 (transverse relaxation time)	0.1 s
k_x (frequency in radial x)	4.71 rad/cm
k_y (frequency in radial y)	4.71 rad/cm
θ (Tip angle)	45°

Table 1. Parameter values for the imaging equation, as described in Gupta and Prince.

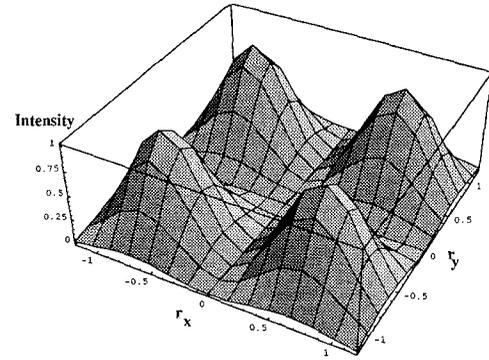


Figure 3. Illustration of a spatial tag pattern produced over a window.

is the spatially varying magnetization pattern applied at end-diastole, or $t = 0$, which causes the tags to appear and an example function (see [11] for more details) is given below:

$$\begin{aligned} \zeta(r) &= [\cos^2 \theta - \sin^2 \theta \cos(k_x r_x)] \\ &\quad \times [\cos^2 \theta - \sin^2 \theta \cos(k_y r_y)] \end{aligned} \quad (5)$$

where k_x, k_y and θ are prespecified constants, and r_x and r_y are components of r . Figure 3 illustrates the kind of tagging produced with equation 5, using constants specified in table 1. In this figure, D_0, T_1, T_2 are fixed and only $\zeta(r)$ varies.

To see how tags fade with time, consider figure 4. Each curve denotes intensity variation with time for a sample point in the heart defined by r_x, r_y . Notice that at time $t = 0$, intensity values of the three points differ much more than than at time $t = 1.0$. Hence, contrast in the image decreases with time, although not linearly. Using equation 5, we can relate the intensities of a point r at times t and at $t + \delta t$ as:

$$\begin{aligned} I(r, t + \delta t) &= e^{-\frac{\delta t}{T_1(r)}} [I(r, t) - F(r)] + F(r) \\ F(r) &= K(r) [1 - \zeta(r) e^{-\frac{T_R}{T_1(r)}}] \end{aligned} \quad (6)$$

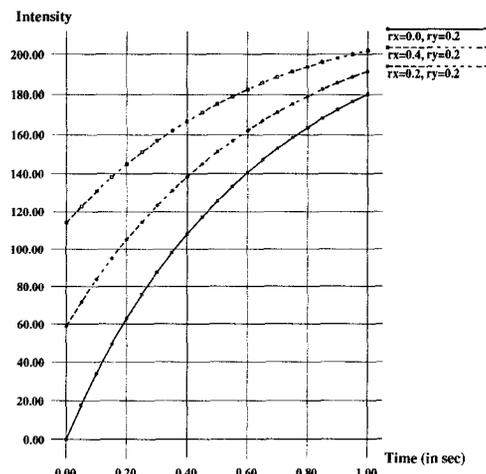


Figure 4. Illustration of temporal intensity variation of three different sample points in the heart. Notice the progressively narrowing difference between the intensity profiles which results in the image becoming more uniform.

It follows that the intensity at a point is a monotonically increasing function of time (also seen in figure 4). Now if we make the assumption that, over a window in the image which corresponds to a small region in the heart, the intensity profiles of different points do not cross each other, then the relative ranking of the intensities at time $t + \delta t$ is *preserved* with respect to that at time t . This is a valid assumption because D_0, T_1 and T_2 do not change dramatically in a small local region. Hence, the s permutation and κ value (in equations 1 and 3) between the two windows do not change. This is the key observation. As noted earlier, equation 4 is strictly true only under zero motion of the heart. However, if the motion is small between frames due to high temporal sampling, then the imaging equation is quite adequate.

There are other forms of tagging, and in one method called echo-planar imaging (EPI), there is a distinction between tagged and non-tagged regions in terms of applied magnetization. However, our method is equally applicable on images acquired using this method as discussed in [5].

4. Comparison

We now compare the performance of the ordinal measure κ against the sum-of-squared-differences SSD and normalized correlation NCC using a sequence (Figure 4). The sequence is that of a rotating ring phantom obtained using a Tesla whole-body scanner (see [11] for details). While the object undergoing motion is not the heart, an operational tagged MR imaging system was used to obtain the images which distinguishes this sequence from a synthetic one. The

parameters used in obtaining the images are given in table 1 and table 2. Using these parameters, it is possible to compute true motion at each pixel in the image. True motion is of subpixel resolution, but for comparison we check to see if correspondence has been established to the nearest pixel. The algorithm for all three measures is identical: For each pixel in the reference image (R) find the best matching window in the next image (S) in a pre-defined search range.

Image Parameter	Value
Image size	150 × 150
Pixel size (along x and y)	0.078125 cm
Period of rotation of ring	10 s
Time interval between images (δt)	0.05 s
Inner and outer radii of ring	2.5 cm, 4.5 cm

Table 2. Image and ring parameters used in the ring sequence (from Gupta and Prince).

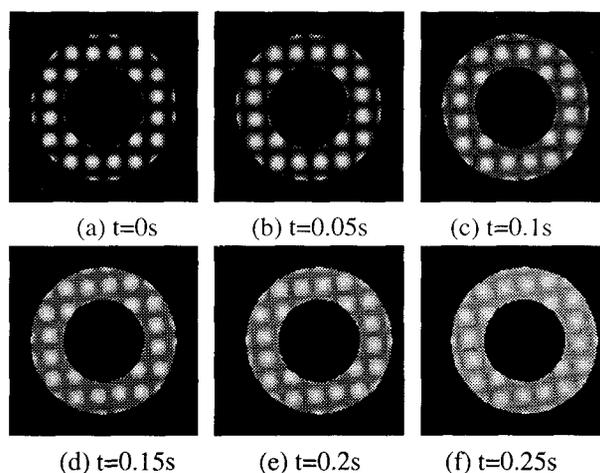


Figure 5. Six consecutive frames of a rotating ring phantom image sequence from Johns Hopkins University (courtesy Sandeep Gupta).

Table 3 summarizes the results of matching with each measure for four different pairs in the image sequence. It must be noted that for this sequence, there is no question of reporting false positives due to occlusion – each pixel in an image has a corresponding pixel in every other image. However, there can be false positives in the overlapping regions between the background and object which a window can straddle. All three methods did the best when band-pass filtered versions of the images were used, rather than the raw images themselves, hence the results are compiled on those filtered images. For all methods, the pre-defined search range

Matching using Frames (a) and (b)		
Measure	Mismatches (N_f)	Percentage Mismatches
κ	1309	18.18%
NCC	1560	21.67%
SSD	2202	30.58%

Matching using Frames (b) and (c)		
Measure	Mismatches (N_f)	Percentage Mismatches
κ	1244	17.27%
NCC	1577	21.90%
SSD	1820	25.28%

Matching using Frames (c) and (d)		
Measure	Mismatches (N_f)	Percentage Mismatches
κ	1009	14.01%
NCC	1810	25.14%
SSD	2222	30.86%

Matching using Frames (e) and (f)		
Measure	Mismatches (N_f)	Percentage Mismatches
κ	1172	16.27%
NCC	2498	34.69%
SSD	2692	37.39%

Table 3. Comparison of the three measures.

was ± 5 pixels in the horizontal and vertical directions. To keep the comparison fair, no thresholds were used on any measure for rejecting or accepting a match. Instead, in each case, every pixel correspondence (with respect to image R) was verified by using image S as reference and checking to see if the resulting match mapped back to the same pixel. If not, then the pixel correspondence was rejected. Also, we used a uniform threshold on the grey-level variance in a window around each pixel in image R , for attempting a match (for example, matching was not attempted by any method when a window completely occupies the black background region). Window size was 9×9 in all cases. If the number of mismatches is N_f , the number of false negatives for pixels in the object is N_{fn} and the number of false positives for pixels outside the object is N_{fp} , then $N_f = N_{fn} + N_{fp}$. Percentage mismatches is $\frac{N_f}{N_t} \times 100$ where N_t is the total number of pixels in the object, found equal to 7200.

Performance of κ was much superior to SSD , as predicted. Gratifying was the better performance of κ with respect to NCC , a more equitable comparison in a sense because both measures are normalized for intensity changes

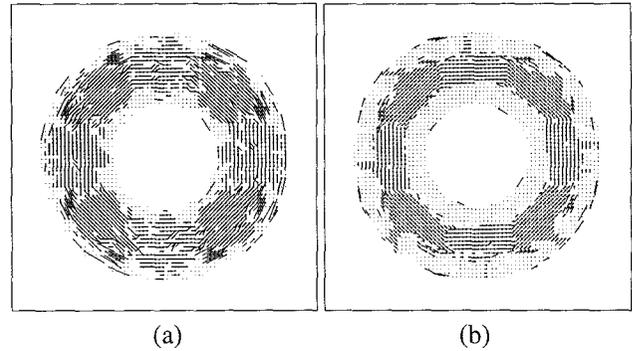


Figure 6. Subsampled needle diagram of flow (to nearest pixel) between frames (c) and (d) of the rotating ring phantom using a) κ , b) NCC .

between windows. A reason for better performance is that NCC compensates only for linear variations. As expected of any window measure, most errors occur on the boundary between the tagged region and background. We have tabulated N_{fn} and N_{fp} independently, in [5]. From a computational speed perspective, SSD was superior. Figure 6 shows a needle diagram corresponding to the flow obtained with κ (to the nearest pixel) with images (c) and (d). It would be illuminating to study matching when tag frequency is altered.

5. Experiment

We applied the motion estimation method as described in the previous sections, to a real heart image pair (see figure 5). The images are taken at consecutive time intervals. We chose

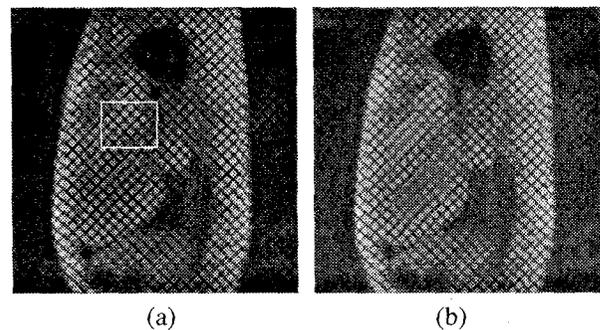


Figure 7. Cardiac images, taken at consecutive time instances, used for motion estimation. A portion of image (a) is marked for which flow is depicted separately in figure 9. This region was chosen because pixels therein undergo significant temporal intensity change.

a ± 3 pixel search range along each axis, and a window size of 9×9 . We also adopted an arbitrary confidence threshold of

0.6 on the correlation value to eliminate poor matches. The images are of size 256×256 . Motion is small in most regions, often much less than a pixel and hence subpixel motion estimation is required. The subpixel motion estimation method is discussed in [5]. Results of matching are shown in figures 8 and 9. The blank regions indicate portions where motion could not be computed due to lack of texture, or because the confidence in matching was low. It appears that the flow in the diagram is consistent with that of the heart.

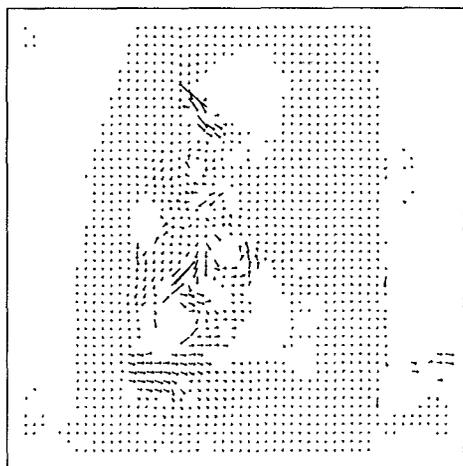


Figure 8. Complete needle diagram of the flow computed using the images in figure 7.

6. Discussion

We presented a method for motion estimation using ordinal measures in the context of tagged MR imaging where intensities at corresponding points, though not identical, are related monotonically. A restriction of our approach is that parameterized motion models like an affine model [3] cannot be embedded within the correlation measure, unlike *SSD* based schemes. In other words, our approach is not regression based where motion model parameters could be determined iteratively, possibly in a hierarchical fashion. However, for applications like cardiac motion estimation which involves deformable objects, such parameterized models are clearly insufficient in any case. Hence, our approach is non-parametric in intensity and motion modeling.

Acknowledgements

We are grateful to Sandeep Gupta at Johns Hopkins University for providing the rotating ring phantom images. This work was supported by ONR/ARPA MURI Grant No. N00014-95-1-0601.

References

- [1] S. Androutsellis-Theotokis and J. L. Prince. Experiments in multiresolution motion estimation for multifrequency tagged

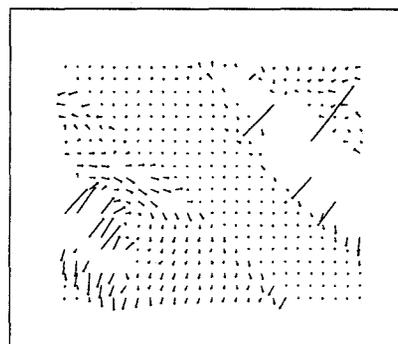


Figure 9. Flow computed for the marked region in figure 7. The flow has been subsampled and magnified. The blank region indicates pixels where motion could not be determined to the specified confidence.

- cardiac mr images. *International Conference on Image Processing*, Sept 1996.
- [2] J. L. Barron, D. J. Fleet, and S. S. Beauchemin. Performance of optical flow techniques. *International Journal of Computer Vision*, 12(1):29–56, 1994.
- [3] J. R. Bergen, P. Anandan, K. J. Hanna, and R. Hingorani. Hierarchical model-based motion estimation. *Proc. of the European Conference on Computer Vision*, pages 237–252, 1992.
- [4] D. N. Bhat and S. K. Nayar. Ordinal measures for visual correspondence. *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 351–357, 1996.
- [5] D. N. Bhat, S. K. Nayar, and A. Gupta. Motion estimation using ordinal measures. Technical report, Columbia University, 1997. (To appear).
- [6] T. S. Denney and J. L. Prince. Optimal brightness functions for optical flow estimation of deformable motion. *IEEE Transactions on Image Processing*, 3(2):178–191, March 1994.
- [7] M. A. Gennert and S. Negahdaripour. Relaxing the brightness constancy assumption in computing optical flow. Technical Report Memo no. 975, MIT, AI Lab, June 1987.
- [8] R. A. Gideon and R. A. Hollister. A rank correlation coefficient. *Journal of the American Statistical Association*, 82(398):656–666, 1987.
- [9] S. N. Gupta and J. L. Prince. On variable brightness optical flow for tagged mri. *International Conference on Information Processing in Medical Imaging*, June 1995.
- [10] R. Kories and G. Zimmerman. A versatile method for the estimation of displacement vector fields from image sequences. *Proc. of the IEEE Workshop on Motion: Representation and Analysis*, pages 101–107, 1986.
- [11] J. L. Prince and E. R. McVeigh. Motion estimation from tagged mr image sequences. *IEEE Transactions on Medical Imaging*, 11(2):238–249, June 1992.
- [12] R. Zabih. *Individuating Unknown Objects by Combining Motion and Stereo*. PhD thesis, Stanford Univ., August 1994.