Mutual Information as a Stereo Correspondence Measure

Geoffrey Egnal
GRASP Laboratory
University of Pennsylvania
Email: gegnal@grasp.cis.upenn.edu

Abstract

Traditional stereo systems often falter over changes in lighting between their two views. Unfortunately, such changes often occur when using stereo with a wide baseline or between images from different spectra. In this paper, we propose a new dense stereo correspondence similarity metric, mutual information, which has the potential to overcome such adverse conditions. We explore the strengths and weaknesses of this metric, both quantitatively and qualitatively, under a variety of conditions. Throughout the exploration, we compare mutual information to a more traditional cross-correlation stereo system. We show that mutual information performs under conditions in which traditional dense stereo fails.

1 Introduction

Traditional stereo systems work reasonably well under ideal conditions. However, many stereo correspondence similarity metrics cannot tolerate even minor changes in lighting. Such variation is not rare; a large-baseline stereo often has unworkable discrepancies in lighting. At the extreme, stereo vision between two cameras with different spectral response, such as an infrared (IR) and visible camera, presents an insurmountable task for many current similarity metrics. As such multi-spectral systems grow in popularity, we will want to augment their abilities with stereo.

In this paper, we introduce a new stereo similarity metric, mutual information (MI). Mathematically, this metric relies on the entropy of the images' underlying probability densities. This reliance on the probability densities of the two images enables the metric to transcend many constraints that bind other systems. For example, mutual information handles matching sample A with the negative of sample B as easily as simply matching A and B. Also, MI can often handle non-functional relationships between A and B, as occurs when the same object is viewed under two different spectra. In later sections, we will show that mutual information can better cope with such a lighting change than a correlation-based system.

The origination of mutual information is largely credited to the 1948 work by Shannon [12]. Since then, there have been many uses of mutual information, including basic statistics, communication theory, and complexity analysis [5]. The first use of mutual information to measure pixel correspondence was in 1995, when Viola and separately, Collignon et al., employed mutual information to register multi-modal medical images, such as CAT and MR images [4, 13]. In subsequent comparisons, mutual information has performed on a level with manually assisted methods in registering multi-modal images [15]. In 1998, Chrastek and Jan give a preliminary exploration of mutual information as a stereo matching metric, with disappointing results. However, they only test their algorithm on one difficult stereo pair without any rectification procedures [3].

The correspondence problem is as old as stereo itself. The specific portion of the problem that this paper addresses, dense similarity measures using windows, has also been researched extensively [7]. Rather than correlation, Cox et al. use a maximum likelihood estimator on individual pixels, and Belhumeur uses a Bayesian framework in his matching [1, 6]. Like many who use energy functionals, both assume that the underlying data or features in the two images would be identical without noise, and occlusion. In a slight relaxation of that assumption, Zabih and Woodfill and later, Bhat and Nayar, successfully use ordinal measures to replace correlation, yet they assume that the intensity in the two images falls in the same order within a window [16, 2]. Other approaches pre-process the images and use a different matching primitive in an effort to increase robustness [9, 14, 8]. If one views MI stereo as using a probability density primitive and a MI similarity metric, then MI uses a different primitive and a different similarity metric to previous projects. We categorize MI as having a pixel-based primitive, rather than a preprocessed primitive, because the joint entropy term requires direct pixel comparison; thus, MI falls in with other pixel-based dense similarity metrics.

In light of previous work, our main contributions in this paper are (i) applying information theory, in the form of mutual information, to the problem of stereo, (ii) exploring the strengths and weaknesses of MI as a stereo similarity metric, and (iii) quantitatively and qualitatively comparing traditional normalized correlation with the new MI metric. Section 2 of this paper describes mutual information, correlation, and our implementation of these metrics in a stereo system. Section 3 shows results for MI, and compares them with modified normalized cross-correlation (MNCC), and Section 4 offers an analysis of the results.

2 Mutual Information Stereo

Mutual information depends upon the entropy and joint entropy of two random variables. In the case of stereo, the random variables are the image pixels we take from each image in a stereo pair. If we assume the pixel values X are discrete random variables (RVs) with discrete density P, then we can define the entropy H:

Definition 1 (Entropy)

$$\mathrm{H}(X) \triangleq -E_X[\log(P(X))]$$

An intuitive description of entropy is that it measures the randomness of a RV. A low entropy means that the average probability over the support set for a given RV is low. For example, a constant region in an image has a lower entropy than a highly textured region. The joint entropy is defined similarly for two random variables X and Y, replacing the univariate P with the joint probability function $P(X,Y)$. Joint entropy can be used to measure alignment, or similarity, because it describes the 'crispness' of a joint probability function. Two identical samples will have a lower joint entropy when aligned than when these same samples are misaligned. However, two constant regions will have a low joint entropy as well. To avoid such a spurious match, we want to maximize the entropy in the individual samples that we are comparing. For this reason, we use mutual information.

Using the definition of entropy, we define mutual information:

Definition 2 (Mutual Information)

$$\begin{aligned}\mathrm{MI}(X,Y) &\triangleq E_{X,Y}\left[\log\left(\frac{P(X,Y)}{P(X)P(Y)}\right)\right] \\ &\triangleq \mathrm{H}(X) + \mathrm{H}(Y) - \mathrm{H}(X,Y)\end{aligned}$$

Mutual information is bounded, so that $0 \leq \mathrm{MI}(X,Y) \leq \min(\mathrm{MI}(X,X),\mathrm{MI}(Y,Y))$. The minimum value occurs when X and Y are completely independent, and the maximum value occurs when X and Y are identical or there is a $1-1$ mapping T between the two, since $\mathrm{MI}(X,T(X)) = \mathrm{MI}(X,X)$. These last points deserve special attention: they help to justify the performance of mutual information as a similarity metric. Mutual information measures similarity, but it is also invariant to $1-1$ transformations of the data. This invariance enables MI to measure similarity in more situations than many traditional similarity metrics. It also explains MI's limitations: when a transformation is not $1-1$, as in adding extreme Gaussian noise, MI has difficulty measuring similarity. In these cases, a large sample size often alleviates the problem.

In order to calculate MI for two image samples, one needs to estimate the probability functions in the underlying images. Perhaps the easiest and fastest method uses a simple histogram [11]. The most difficult aspect in using this method is the decision of how many bins to use. We find that a 20-bin histogram performs well in our experiments. An alternate method uses kernel density estimation, or Parzen windowing [10]. We have implemented this method as well, using Gaussian kernels. This method offers a smooth density estimate, but requires many more sample points because it split the original sample into a density estimate sample and an entropy estimate sample. Combined with the additional time for calculation, we find kernel estimation's additional needs prohibitive.

We install this metric in a classical scanline-search stereo algorithm which obtains a dense disparity map from the left image to the right image, comparing a single left window to all right image windows within a given search range. For each shift at each window, we calculate the mutual information using histograms and maximize the MI over the possible matches. We obtain subpixel accuracy in our curve maximization by fitting a three point quadratic curve. However, we do not embellish the algorithm using any smoothness enforcement, occlusion avoidance, foreshortening accomodation, or scanline consistency methods. We measure confidence as the curvature of the similarity (MI or MNCC) curve around the maximal similarity score, S_{Max}:

$$\text{Conf} = 2S_{Max} - S_{Left} - S_{Right}$$

where S_{Left} and S_{Right} are the correlation scores to the left and right of S_{Max}.

As with other correlation algorithms, MI works better with a larger sample size, or window size. Unfortunately, in a stereo system, larger window sizes take time to calculate as well as blurring sharp boundaries that normally occur in the real world. We present some sensitivity results in the next section for both MI and MNCC around the customary 9 and 11 pixel square windows, as well as an even larger 15 pixel window.

The MI algorithm is relatively computationally expensive because density estimation costs more labor than a simple correlation calculation. A rough estimation reveals that a simple entropy calculation takes $histWidth + windowSize^2$ operations, and a joint entropy calculation takes $histWidth^2 + 2 * windowSize^2$ operations. Since we perform these calculations for each window at each shift, we have $imgHeight * imgWidth * numShifts * (2(histWidth + windowSize^2) + (histWidth^2 + 2 * windowSize^2))$ operations. Our runs typically take about 10 minutes on a Sun Ultra-4 workstation. However, memoization in calculating the individual entropies would save significant time.

In our comparisons, we use modified normalized cross-correlation (MNCC) as the traditional correlation metric. We choose this over the more typical sum-of-squared-differences or sum-of-absolute-differences because it is more robust, and provides a higher hurdle of comparison.

Definition 3 (MNCC)

$$\text{MNCC}(X, Y) \triangleq \frac{2\ \text{Cov}(X, Y)}{\text{Var}(X) + \text{Var}(Y)}$$

This metric resists blowup near constant areas by adding the two denominator variances rather than multiplying them. Like regular NCC, it measures the linear relation between the two image samples, normalizing for any overall intensity changes. However, the metric is unable to measure similarity in complex relationships, such as those found between non-linearly transformed images.

3 Results

We first show the qualitative results of the MI algorithm on the famous Pentagon stereo pair and compare the results with those of MNCC (see Figure 1). In this example and the next, our search range is ± 17 pixels, and our window size is 15. Although both algorithms perform well, the MI results appear slightly rougher. Some of the near-constant patches of the image create problems for MI, mostly because of the histogram's discretization process. Small portions of each window cross a histogram bucket limit in one image but not the other, fooling MI into believing the two samples have different densities.

Our second general test involves random dot stereograms, which were formed by taking a 300x300 image of random dots with intensity distributed uniformly over $[0, 255]$. We then displaced a 100x100 square within the right image by 16 pixels rightwards and filled the resulting gap with additional random dots to make the left image. Subsequently, Gaussian noise of variance 5 was added. At this low level of noise, both algorithms perform similarly, with very little error.

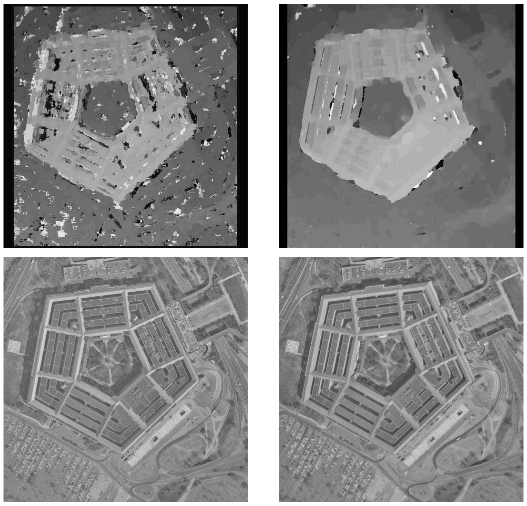

Figure 1: MI and MNCC Results For the Pentagon Stereogram. The top row shows the disparity maps for the MI (left) and MNCC (right) algorithms. The bottom row displays the stereo pair. Although the results appear roughly the same, the MI results are more patchy than the MNCC results around the less textured portions of the image. Both results are hampered by the ± 17 pixel search range.

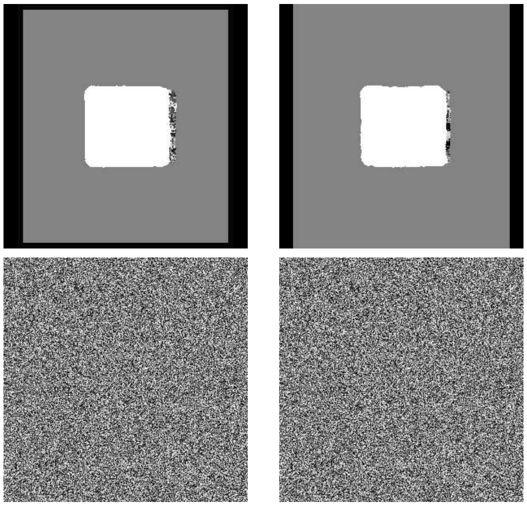

Figure 2: Random Dot Stereogram Results. The layout mirrors that of Figure 1, with the disparity maps (MI left) on top of the stereo pair. Both results are almost perfect, with error stemming from window effects.

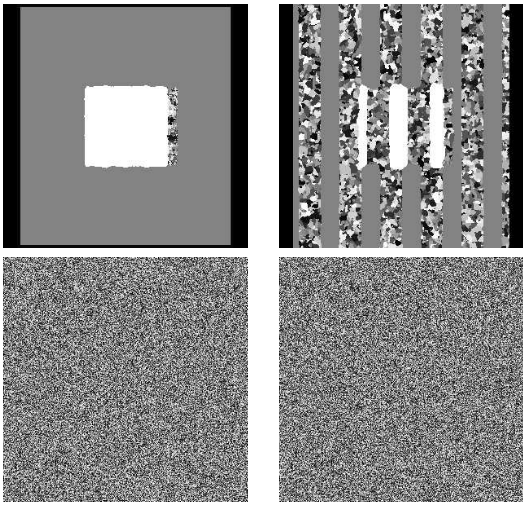

Figure 3: Flipped Random Dot Stereogram Results. The layout is the same as that of Figure 2. In the right image, we negated values in alternate bands of 25 pixels. As we can see, MNCC cannot tolerate this transformation, while MI can.

3.1 The MI Advantage

MI has the advantage of being able to detect similarity between samples with a complex relationship. As a simple demonstration, we negate the right random dot image in bands of 25 pixels before shifting the central square (see Figure 3). The results clearly demonstrate that MI tolerates this relationship, while MNCC cannot. If we had known beforehand what the relationship was, we could have prepared the data, and MNCC would perform almost perfectly. However, when actually imaging the same object under two different imaging modalities, we do not know the relationship between the two views. For instance, under IR and visible stereo, a black object could be hot or cold, appearing white or black in the far-IR image.

To demonstrate MI stereo in real world situations, yet to still have a solid ground truth, we test the algorithm on pairs of images from a single viewpoint. Since the view remains the same, the stereo disparity is exactly zero. In other words, we take two images from the same perspective, but change one parameter - the lighting. These images have the same CCD noise and lighting variation of an equivalent stereo pair, but they have no occlusion. Since we are only testing stereo matching metrics, this setup is appropriate. A similar setup has been used previously by [2]. To simulate stereo matching, our search range for these images is ± 10 pixels.

In calculating our statistics, we count a hit as any match falling within one pixel of the truth, which we know to be zero. Often, we separate the hit scores by high/low confidence or entropy values. A low label indicates that we have taken the hits corresponding to the top 96% of possible confidence or entropy values, while a high label indicates the top 60% of values. The actual participation rates vary and are displayed where significant.

Using this setup, we first view blue paper shapes on a red paper background alternately through a red and blue filter (see Figure 4). In this case, MNCC fails in the high entropy areas of the picture because of the distinct negation near the shapes' edges. Note that low entropy includes roughly 80% of possible points, while high entropy includes only 30%. In the center of the shapes, MNCC performs well because the lighting variation is directionally similar in the two images. Interestingly, MNCC does not improve using a larger window because the larger windows include the inverted edges. We believe that one could see a similar effect in visible/far-IR stereo. Unfortunately, high-quality far-IR cameras are not available to us at this time.

To view the potential for near-IR/visible stereo, we view a truck in two ways: under both visible and near-IR and then with the visible light filtered out (see Figure 5). In this case, the distinction between the two test images is not as severe as in the last example. Nevertheless, the two images have an overall intensity difference, and certain parts appear differently as a result of the different spectra involved. Overall, the two perform similarly, with errors caused mostly by constant regions. At large window sizes, MI gets more hits than MNCC, while at high confidence, MI clearly outperforms MNCC. In near-constant regions, the histogram buckets have helped MI, smoothing out the small variations and indicating low confidence. In the same areas, MNCC not only sees many false matches, but unfortunately, also has high confidence in them. For this reason, MI has a more reliable confidence measure than that of

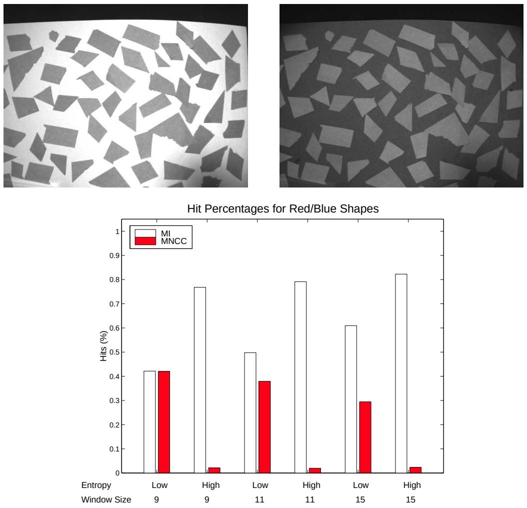

Figure 4: Blue Paper Shapes on a Red Paper Background. The left image comes from a camera with a red filter, while the right image comes from a camera with a blue filter. Near the blue shapes' edges, the relationship of blue to red inverts between the images, confusing MNCC, but not MI. Also, note that MI profits more from a larger window than MNCC does.

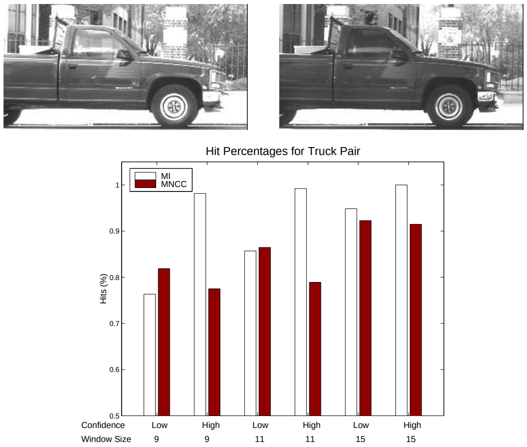

Figure 5: Truck Viewed Under Only Near-IR (left) and Visible/Near-IR (right). At high confidence levels, MI is superior to MNCC because MNCC's confidence metric reports artificially high confidence in near-constant regions. Moreover, MI capitalizes on a larger window size more than MNCC.

MNCC. Note that low confidence includes near 80% of possible matches, while high confidence includes about 13%.

3.2 Differences Between MI/MNCC

We analyze the sensitivity of MI and MNCC to additive Gaussian noise using random dot stereograms at various noise levels (see Figure 6). Unfortunately, MI is less robust to this type of noise than MNCC due to our density estimation technique. Noisy samples often fall outside the histogram bin in which the data should be, forming a different density estimate and making it hard for MI to match. This is precisely where larger sample sizes and more continuous density estimates, such as Parzen windowing, would help. In contrast, MNCC uses the data without any additional discretization, making it more robust to noise. However, at a window size of 15 pixels, MI survives extreme noise, probably more severe than in most imaging situations. MI handles the noise levels of the imagery in this paper because the noise is neither Gaussian nor as severe as the noise in this test.

Our last test involves simply changing the lighting

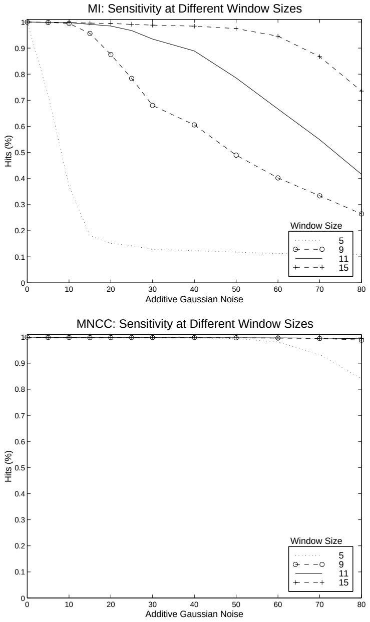

Figure 6: MI and MNCC Performance Under Additive Gaussian Noise. This test was based on random dot stereograms with various levels of additive Gaussian Noise. MI is less able to handle noise than MNCC because the density estimate is delicate to its fixed histogram bucket limits. However, at a window size of 15 pixels, MI's noise tolerance appears acceptable for most applications.

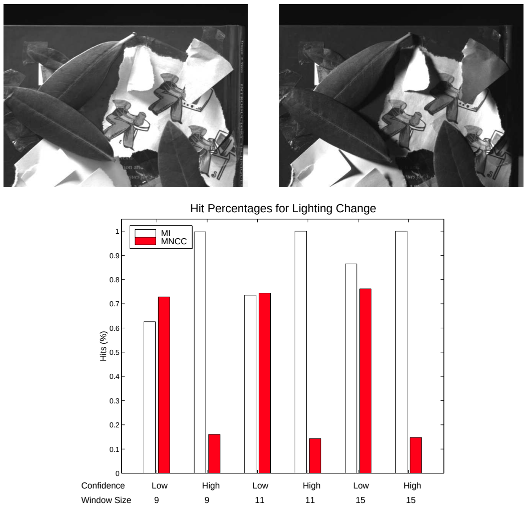

Figure 7: Diffuse Lighting vs Directional Lighting. MI outperforms MNCC at high confidence matches and when using large windows.

source (see Figure 7). The image shows various paper objects and leaves attached to a flat surface. The diffuse, overhead lighting in one image contrasts sharply with the strong, directional light in the other. A large baseline stereo might undergo a similar lighting variation. Although MI does not consistently outperform MNCC, at high confidence levels it is extremely accurate, far above MNCC. Here, low confidence includes 50%, and high confidence includes 10% of all possible matches. As with the truck example, MI's confidence metric has proved more useful than MNCC's because it better indicates low confidence in near-constant regions.

4 Discussion

Based on the results, we can see the usefulness of MI as a stereo similarity metric. Conceptually, mutual information can measure similarity or alignment in more circumstances than MNCC. When applied to stereo, the MI measure can handle different lighting conditions between the two views, or at the extreme, multi-spectral stereo. MNCC, like other current stereo similarity metrics, will not behave well under such adversity. Real world situations, such as a large baseline stereo or a robot with multispectral cameras, demand this kind of robustness. Even as a part of a trinocular, visible/visible/IR stereo system, inter-spectral stereo could add a verification layer that previously didn't exist.

In addition, MI produces more accurate confidence scores than the MNCC algorithm. This knowledge is extremely important in any real-world application; it gives the stereo system a method to decide which matches to trust. Having a few trustworthy points is often better than having many more matches, but many of which are wrong.

Despite these advantages, MI does have limitations as a measure. In its current implementation, it is more susceptible to noise than MNCC, largely because the histogram buckets' inflexibility cannot tolerate large noise levels. However, a large window size seems sufficient to tolerate ordinary camera noise. In addition, like other metrics, we can not yet predict which relationships MI can measure, and which kinds it cannot. We have seen that MI can handle images within a larger class than MNCC, but we would like to understand more.

For the future, methods of making MI more robust to noise are needed. Although a faster Parzen windowing scheme may help, we believe that true innovation will be required for MI to overcome this hurdle using the small sample sizes that fine resolution stereo requires. We would also like to test the MI on extremely high quality far-IR/visible systems that were simply not available to us. However, the tests in the paper sufficiently show that MI has the ability to measure stereo similarity under what is now considered extreme circumstances.

References

[1] P. N. Belhumeur. A Bayesian approach to binocular stereopsis. *IJCV*, 19(3):237–260, 1996.

[2] D. N. Bhat and S. K. Nayar. Ordinal measures for image correspondence. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(4):415–423, 1998.

[3] R. Chrastek and J. Jan. Mutual information as a matching criterion for stereo pairs of images. *Analysis of Biomedical Signals and Images*, 14:101–103, 1998. VUTIUM Press, ISBN 80-214-1169-4.

[4] A. Collignon, F. Maes, D. Delaere, D. Vandermeulen, P. Suetens, and G. Marchal. Automated multi-modality image registration based on information

theory. In *Information Processing in Medical Imaging*, pages 263–274, 1995.

[5] T. M. Cover and J. A. Thomas. *Elements of Information Theory.* John Wiley & Sons, USA, 1991.

[6] I. J. Cox, S. L. Hingorani, and S. B. Rao. A maximum likelihood stereo algorithm. *Computer Vision and Image Understanding*, 63(3):542–567, 1996.

[7] U. R. Dhond and J. Aggarwal. Structure from stereo - a review. *Systems, Man, and Cybernetics*, 19(6):1489–1510, 1989.

[8] D. G. Jones and J. Malik. A computational framework for determining stereo correspondence from a set of linear filters. In *Proceedings of the European Conference on Computer Vision*, pages 395–410, 1992.

[9] M. Kass. Computing visual correspondence. In *DARPA Image Understanding Workshop*, volume 14, pages 54–60, June 1983.

[10] E. Parzen. On estimation of a probability density function and mode. *Annals of Mathematical Statistics*, 33:1065–1074, September 1962.

[11] D. W. Scott. *Multivariate Density Estimation: Theory, Practice and Visualization.* John Wiley & Sons, USA, 1992.

[12] C. Shannon. A mathematical theory of communication. *Bell Systems Technical Journal*, 27:379–423, 623–656, 1948.

[13] P. A. Viola and W. M. Wells III. Alignment by maximization of mutual information. *International Journal of Computer Vision*, 24(2):137–154, 1997.

[14] J. Weng. A theory of image matching. In *Proceedings of the International Conference on Computer Vision*, pages 200–209, 1990.

[15] J. West, J. Fitzpatrick, M. Wang, B. Dawant, C. Maurer Jr., R. Kessler, R. Maciunas, C. Barillot, D. Lemoine, A. Collignon, F. Maes, P. Suetens, D. Vandermeulen, P. van den Elsen, S. Napel, T. Sumanaweera, B. Harkness, P. Hemler, D. Hill, D. Hawkes, C. Studholme, J. Maintz, M. Viergever, G. Malandain, X. Pennec, M. Noz, G. Maguire Jr., M. Pollack, C. Pellizzari, R. Robb, D. Hanson, and R. Woods. Comparison and evaluation of retrospective intermodality brain image registration techniques. *Journal of Computer Assisted Tomography*, 21(4):554–566, 1997.

[16] R. Zabih and J. Woodfill. Non-parametric local transforms for computing visual correspondence. In *Proceedings of the European Conference on Computer Vision*, pages 151–158, 1994.