Non-parametric Local Transforms for Computing Visual Correspondence

Ramin Zabih[1] and John Woodfill[2]

[1] Computer Science Department, Cornell University, Ithaca NY 14853-7501, USA
[2] Interval Research Corporation, 1801-C Page Mill Road, Palo Alto CA 94304, USA

Abstract. We propose a new approach to the correspondence problem that makes use of non-parametric local transforms as the basis for correlation. Non-parametric local transforms rely on the relative ordering of local intensity values, and not on the intensity values themselves. Correlation using such transforms can tolerate a significant number of outliers. This can result in improved performance near object boundaries when compared with conventional methods such as normalized correlation. We introduce two non-parametric local transforms: the *rank transform*, which measures local intensity, and the *census transform*, which summarizes local image structure. We describe some properties of these transforms, and demonstrate their utility on both synthetic and real data.

1 Introduction

The correspondence problem is a fundamental problem in vision, as it forms the basis for stereo depth computation and most optical flow algorithms. Given two images of the same scene, a pixel in one image corresponds to a pixel in the other if both pixels are projections along lines of sight of the same physical scene element. If the two images are temporally consecutive, then computing correspondence determines motion. If the two images are spatially separated but simultaneous, then computing correspondence determines stereo depth. *Area-based* approaches to the correspondence problem [4] find a dense solution, usually by relying on some kind of statistical correlation between local intensity regions.

In this paper we propose a new area-based approach to the correspondence problem, based on non-parametric local transforms followed by correlation. We begin by motivating our approach, then show how non-parametric local transforms can be used to determine correspondence. In section 3 we introduce the *rank* and *census* transforms, and describe their properties. We give empirical evidence of the performance of our methods in section 4, using both natural and synthetic images. Finally, in section 5 we survey related work and discuss some planned extensions.

2 Non-parametric local transforms

Our approach to the correspondence problem is first to apply a local transform to the image, and then to use correlation. In this respect, our work is similar

Lecture Notes in Computer Science, Vol. 801
Jan-Olof Eklundh (Ed.)
Computer Vision – ECCV '94

to that of Nishihara [12] and Seitz [14, 1]. Nishihara's transform is the sign bit of the image after convolution with a Laplacian, while Seitz's transform is the direction of the intensity gradient.

Most approaches to the correspondence problem have difficulty near discontinuities in disparity, which occur at the boundaries of objects. Near such a boundary, the pixels in a local region represent scene elements from two distinct instensity populations. Some of the pixels come from the object, and some from other parts of the scene. As a result, the local pixel distribution will in general be multimodal near a boundary. This poses a problem for many correspondence algorithms, such as normalized correlation [6].

Correspondence algorithms are usually based on standard statistical methods, which are best suited to a single population. Parametric measures, such as the mean or variance, do not behave well in the presence of distinct subpopulations, each with its own coherent parameters. This problem, which we will refer to as *factionalism*, is a major issue in computer vision, and has been addressed with a variety of methods, including robust statistics [2, 3], Markov Random Fields [5] and regularization [13].

The fundamental idea behind our approach is to define a local image transform that tolerates factionalism. Correspondence can be computed by transforming both images and then using correlation. For this approach to succeed, the transform must result in significant local variation within a given image; in addition, it must give similar results near corresponding points between the two images. (Marr and Nishihara [10] refer to these two properties as *sensitivity* and *stability*.) Finally, to handle stereo imagery, the transform should be invariant under changes in image gain and bias.

Our approach relies on local transforms based on non-parametric measures that are designed to tolerate factionalism. Non-parametric statistics [9] is distinguished by the use of ordering information among data, rather than the data values themselves. Non-parametric local transforms, which we introduced in [15], are local image transformations that rely on the relative ordering of intensities, and not on the intensity values themselves.

3 The rank transform and the census transform

We next describe two non-parametric local transforms. The first, called the *rank transform*, is a non-parametric measure of local intensity. The second, called the *census transform*, is a non-parametric summary of local spatial structure.

Let P be a pixel, $I(P)$ its intensity (usually an 8-bit integer), and $N(P)$ the set of pixels in some square neighborhood of diameter d surrounding P. All non-parametric transforms depend upon the comparative intensities of P versus the pixels in the neighborhood $N(P)$. The transforms we will discuss only depend on the sign of the comparison. Define $\xi(P, P')$ to be 1 if $I(P') < I(P)$ and 0 otherwise. The non-parametric local transforms depend solely on the set of pixel

comparisons, which is the set of ordered pairs

$$\Xi(P) = \bigcup_{P' \in N(P)} (P', \xi(P, P')).$$

They differ in terms of their exact reliance on Ξ.

The first non-parametric local transform is called the *rank transform*, and is defined as the number of pixels in the local region whose intensity is less than the intensity of the center pixel. Formally, the rank transform $R(P)$ is

$$R(P) = \|\{ P' \in N(P) \mid I(P') < I(P) \}\|.$$

Note that $R(P)$ is not an intensity at all, but rather an integer in the range $\{0, \ldots, d^2 - 1\}$. This distinguishes the rank transform from other attempts to use non-parametric measures such as median filters, mode filters or rank filters [7]. To compute correspondence, we have used L_1 correlation (minimizing the sum of absolute values of differences) on the rank-transformed images.

The second non-parametric transform is named the *census transform*. $R_\tau(P)$ maps the local neighborhood surrounding a pixel P to a bit string representing the set of neighboring pixels whose intensity is less than that of P. Let $N(P) = P \oplus D$, where $\oplus$ is the Minkowski sum and D is a set of displacements, and let $\otimes$ denote concatenation. The census transform can then be specified,

$$R_\tau(P) = \bigotimes_{[i,j] \in D} \xi(P, P + [i, j]).$$

Two pixels of census transformed images are compared for similarity using the Hamming distance, i.e. the number of bits that differ in the two bit strings. To compute correspondence, we have minimized the Hamming distance after applying the census transform.

These local transforms rely solely upon the set of comparisons Ξ, and are therefore invariant under changes in gain or bias. The tolerance of these transforms for factionalism also results from their reliance upon Ξ. If a minority of pixels in a local neighborhood has a very different intensity distribution than the majority, only comparisons involving a member of the minority are affected. Such pixels do not make a contribution proportional to their intensity, but proportional to their number. This limited dependence on the minority's intensity values is a major distinction between our approach and parametric measures.

To illustrate the manner in which these transforms tolerate factionalism, consider a three-by-three region of an image whose intensities are

$$\begin{array}{ccc} 127 & 127 & 129 \\ 126 & 128 & 129 \\ 127 & 131 & A \end{array}$$

for some value $0 \leq A < 256$. Consider the effect on various parametric and non-parametric measures, computed at the center of this region, as A varies over its

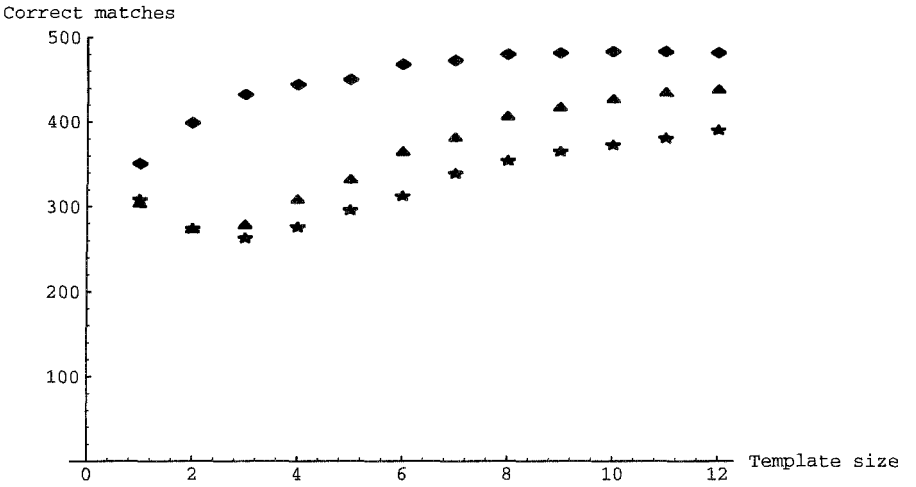

Fig. 1. Comparison of rank ($\diamond$), normalized ($\triangle$) and SSD ($\star$) correlation on Aschwanden data-set with salt-and-pepper noise

256 possible values. The mean[3] of this region varies from 114 to 142, while the variance ranges from 2 to 1823. These parametric measures exhibit continuous variation over a substantial range as A changes.

Non-parametric transforms are more stable, however. All the elements of Ξ except one will remain fixed as A changes. Ξ will be

$$\begin{matrix} 1\,1\,0 \\ 1\quad 0 \\ 1\,0\,a \end{matrix}$$

where a is 1 if $A < 128$, and otherwise 0. The census transform simply results in the bits of Ξ in some canonical ordering, such as $\{1,1,0,1,0,1,0,a\}$. The rank transform will give 5 if $A < 128$, and otherwise 4.

This comparison shows the tolerance that non-parametric measures have for factionalism. A minority of pixels can have a very different value, but the effect on the rank and census transforms is limited by the size of the minority.

4 Empirical results

We have implemented these non-parametric local transforms, and have explored their behavior on both real and synthetic imagery. The motivation for our approach was to obtain better results near the edges of objects. We have obtained comparative results on synthetic data which show that our methods can outperform normalized correlation.

In [1], Aschwanden and Guggenbühl have described the performance of a number of area-based stereo algorithms under several different noise models.

[3] For convenience, we are rounding the actual values

Figure 1 compares correlation with the rank transform against two standard stereo algorithms, namely normalized correlation and sum of squared differences (SSD) correlation. Performance is measured as function of template radius, as described in [1].

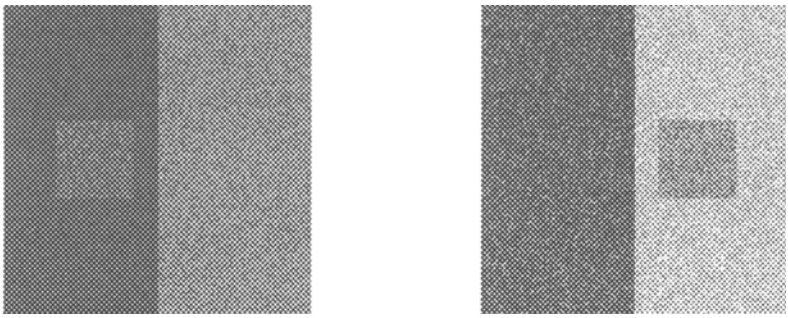

Fig. 2. Right and left random-dot stereograms

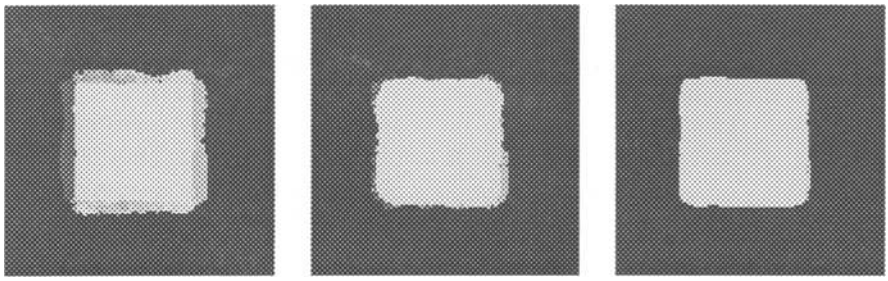

Fig. 3. Disparities from normalized correlation, rank and census transforms

Another way to compare correlation methods is with random dot imagery. Figure 2 shows a random dot stereogram of a square floating in front of a flat surface, on which there is a vertical intensity edge. The images are noise-free, but the intensities differ by fixed gain and bias.

Figure 3 shows the disparities computed from normalized correlation and from correlation with the rank and census transforms. There should only be 2 disparities in this scene: one for the background surface (which is at disparity 0), and one for the foreground square (which is at disparity 104). Notice the comparatively poor performance of normalized correlation near the edges, where it introduces spurious disparities. The performance of our approach can be seen by counting the pixels with incorrect disparities, as shown below.

Algorithm	Incorrect matches
Normalized	1385
Rank transform	609
Census transform	407

On this example, the non-parametric local transforms appear to exhibit better performance than normalized correlation.

The best evidence in favor of the non-parametric local transforms is their performance on real images. We have used the rank transform and the census transform on a number of different images to obtain stereo depth. Depth maps are shown with lighter shades indicating larger disparities and thus nearer scene elements. All the depth maps shown were generated with the same parameters (a transform radius of 7 pixels, and a correlation radius of 4 pixels).

Figure 4 shows a beam-splitter image of a puppet (Elmo from the television show "Sesame Street"). The depth results of the non-parametric local transforms are shown in figure 5. Figure 6 shows an image from a tree sequence[4] captured by moving a camera along a rail, and the depth results from the transforms.

5 Related work and planned extensions

The algorithms we describe are related to non-parametric measures of association, such as Spearman's correlation coefficient r_s or Kendall's τ. These are measures of association of paired data that are based upon comparisons. However, such measures are very expensive to compute, and do not capture the spatial structure of images.

Probably the most similar approach to ours is the work based on robust statistics [2, 11, 3]. Robust statistics differs from our approach in that they emphasize reducing the influence of outliers. Implicit in this work is the assumption that outliers are distributed randomly. However, at the edges of objects, factionalism produces outliers with consistent distributions. Our approach tolerates outliers with consistent distributions, and does not allow pixels from a small faction to contribute in a manner proportional to their intensity.

One limitation of the non-parametric transforms we have described is that the amount of information they associate with a pixel is not very large. We hope to address this shortcoming by combining a number of different non-parametric transforms into a vector of measures associated with a pixel. Ultimately, we would like to avoid the correlation phase altogether and simply match pixels according to a set of semi-independent measures, in a manner similar to that proposed by Kass [8].

Another limitation of our approach is that the local measures rely heavily upon the intensity of the center pixel. This has not been an issue in practice, but we propose to address it by doing comparisons from a local median intensity instead of $I(P)$. An additional idea we intend to pursue is to generalize Ξ, which currently uses the sign of the intensity differences. We plan to explore using higher-order differences, as well as the information contained in the total ordering of the local pixel intensities.

We are also interested in efficient algorithms for implementing such transforms. [15] describes a number of fast algorithms for computing the rank trans-

[4] The tree imagery appears courtesy of Harlyn Baker and Bob Bolles

form based on dynamic programming. We have recently implemented an approximation of the census transform on a Sun workstation, which produces stereo depth with 24 disparities on 640 by 240 images at 1–2 frames per second.

Acknowledgements

Portions of this work were done while the first author was at the Computer Science Department at Stanford University, supported by a fellowship from the Fannie and John Hertz Foundation. We wish to thank SRI for the use of their Connection Machine.

References

1. P. Aschwanden and W. Guggenbühl. Experimental results from a comparative study on correlation-type registration algorithms. In Förstner and Ruwiedel, editors, *Robust Computer Vision*, pages 268–289. Wichmann, 1993.
2. Paul Besl, Jeffrey Birch, and Layne Watson. Robust window operators. In *International Conference on Computer Vision*, pages 591–600, 1988.
3. Michael Black and P Anandan. A framework for the robust estimation of optical flow. In *International Conference on Computer Vision*, pages 231–236, 1993.
4. U. Dhond and J. Aggarwal. Structure from stereo — a review. *IEEE Transactions on Systems, Man and Cybernetics*, 19(6), 1989.
5. Stuart Geman and Donald Geman. Stochastic relaxation, gibbs distributions, and the bayesian restoration of images. *IEEE PAMI*, 6:721–741, 1984.
6. Marsha Jo Hanna. *Computer Matching of Areas in Stereo Images.* PhD thesis, Stanford, 1974.
7. R. Hodgson, D. Bailey, M. Naylor, A. Ng, and S. McNeill. Properties, implementations and applications of rank filters. *Journal of Image and Vision Computing*, 3(1):3–14, February 1985.
8. Michael Kass. Computing visual correspondence. *DARPA Image Understanding Proceedings*, pages 54–60, 1983.
9. E. L. Lehman. *Nonparametrics: statistical methods based on ranks.* Holden-Day, 1975.
10. David Marr and Keith Nishihara. Representation and recognition of the spatial organization of three-dimensional shapes. *Proceedings of the Royal Society of London B*, 200:269–294, 1978.
11. Peter Meer, Doron Mintz, Azriel Rosenfeld, and Dong Yoon Kim. Robust regression methods for computer vision: A review. *International Journal of Computer Vision*, 6(1):59–70, 1991.
12. H. Keith Nishihara. Practical real-time imaging stereo matcher. *Optical Engineering*, 23(5):536–545, Sept–Oct 1984.
13. Tomaso Poggio, Vincent Torre, and Christof Koch. Computational vision and regularization theory. *Nature*, 317:314–319, 1985.
14. Peter Seitz. Using local orientational information as image primitive for robust object recognition. *SPIE proceedings*, 1199:1630–1639, 1989.
15. Ramin Zabih. *Individuating Unknown Objects by Combining Motion and Stereo.* PhD thesis, Stanford University, 1994 (forthcoming).

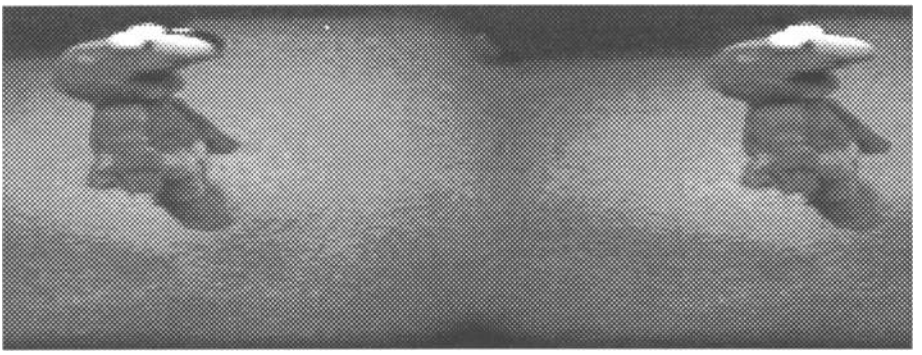

Fig. 4. Elmo stereo pair from beam-splitter

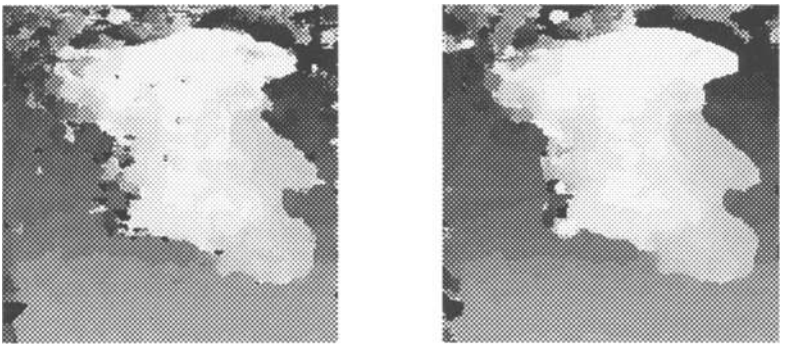

Fig. 5. Rank and census results on Elmo

Fig. 6. Tree image with rank and census correlation results