A new approach to qualitative stereo

Yacov Hel-Or Shimon Edelman*

Dept. of Applied Mathematics and Computer Science
The Weizmann Institute of Science
Rehovot 76100, ISRAEL
$[toky, edelman]@wisdom.weizmann.ac.il$

Abstract

Nonmetric multidimensional scaling (MDS) allows one to derive a quantitative representation from a set of qualitative data which satisfy certain simple constraints. As a tool for vision, MDS combines the advantages of both qualitative and classical approaches, by relying, on the one hand, on an ordinal-scale input representation, and by supporting, on the other hand, the extraction of metric information. The proposed approach is illustrated on the example of stereopsis, although it is applicable also to the processing of other visual cues, as well as to the integration of several cues within a common computational framework.

1 Motivation: quantitative information from qualitative measurements

A growing body of work in computer vision and visual perception is aimed at exploiting the advantages of the qualitative approach to the representation and processing of visual information [1]. According to the proponents of qualitative vision, trading precision for robustness can yield stable and noise-free representations without sacrificing utility for real-world tasks.

A typical qualitative method starts with a set of measurements (which themselves are, most frequently, quantitative), and derives from them a representation in which only certain qualitative features pertinent to the task are retained. In binocular stereopsis, for example, one may process disparity data to compute qualitative characteristics of the surface such as the sign of its Gaussian curvature [16], instead of trying to recover surface depth.

In the present work, we show that it is possible to derive a quantitative representation from an intermediate set of qualitative data which satisfy certain simple constraints. By relying, on the one hand, on ordinal intermediate representations, and by supporting, on the other hand, the extraction of metric information, the proposed method combines the advantages of both qualitative and classical approaches. The method is illustrated on the example of stereopsis, although it is applicable also to the processing of other visual cues, as well as to the integration of several cues within a common computational framework.

*To whom correspondence should be addressed.

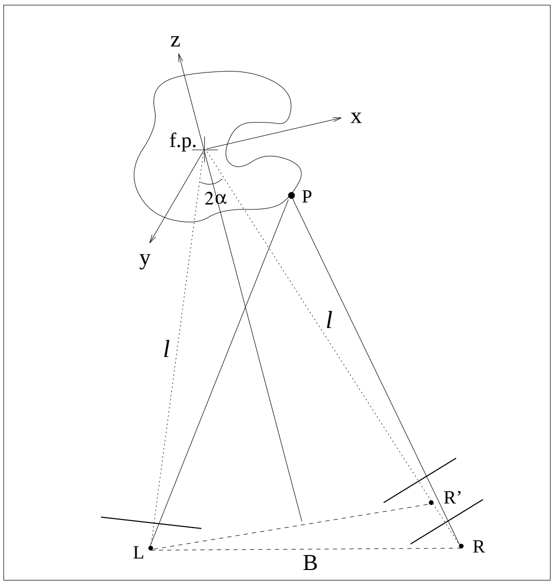

Figure 1: The viewing geometry assumed in the derivation of eq. 4.

2 Depth from pairwise disparity differences

2.1 Binocular viewing geometry

Consider two images of the same object, taken from viewpoints L and R, with the two cameras or eyes aimed at a common fixation point fp (see Figure 1). Assuming that the baseline B is small compared to the distance to the object ($B << l$), we can treat the right image as if it is taken from R' instead of R. We place the origin of the coordinate system in which the depth will be reconstructed at the fixation point fp, so that the X-axis is parallel to the baseline LR' and the Z-axis coincides with the normal to LR' through fp. In this coordinate system, a point $P = (P_x, P_y, P_z)^T$ is projected onto the left image as:

$$P_x^l = \frac{P_x \cos\alpha + P_z \sin\alpha}{-P_x \sin\alpha + P_z \cos\alpha + l} \tag{1}$$

and to the right image as:

$$P_x^r = \frac{P_x \cos\alpha - P_z \sin\alpha}{P_x \sin\alpha + P_z \cos\alpha + l} \tag{2}$$

where the angle α is as depicted in Figure 1. Consider now two points, P and Q, and define disparity differences

$$\begin{aligned}
(\Delta l - \Delta r)_{P,Q} &\doteq \left(P_x^l - Q_x^l\right) - \left(P_x^r - Q_x^r\right) \\
&= \left(P_x^l - P_x^r\right) - \left(Q_x^l - Q_x^r\right) \\
&= \left(\frac{P_x \cos\alpha + P_z \sin\alpha}{-P_x \sin\alpha + P_z \cos\alpha + l} - \frac{Q_x \cos\alpha + Q_z \sin\alpha}{-Q_x \sin\alpha + Q_z \cos\alpha + l}\right) -
\end{aligned}$$

$$-\left(\frac{P_x \cos\alpha - P_z \sin\alpha}{P_x \sin\alpha + P_z \cos\alpha + l} - \frac{Q_x \cos\alpha - Q_z \sin\alpha}{Q_x \sin\alpha + Q_z \cos\alpha + l}\right) \tag{3}$$

If the points P, Q are close enough to the fixation point so that $P_x, P_z, Q_x, Q_z << l$, equation 3 yields:

$$(\Delta l - \Delta r)_{P,Q} = d_P - d_Q \approx \frac{2\sin\alpha}{l}(P_z - Q_z) \tag{4}$$

where d_P and d_Q are the horizontal disparities of points P and Q respectively. For two pairs of points (P, Q) and (U, V) we can define a quantity $\delta(P, Q; U, V)$ as

$$\delta(P, Q; U, V) \doteq (\Delta l - \Delta r)_{P,Q} - (\Delta l - \Delta r)_{U,V} \tag{5}$$

The sign of $\delta(P, Q; U, V)$ then indicates whether the difference between the depth values of P, Q is greater or smaller than the difference between the depth values of U, V. We proceed to show how the information provided by the δ's, or even merely by the sign of the δ's, can be used to recover depth.

2.2 Multidimensional scaling (MDS)

Our method is based on multidimensional scaling (MDS) — a family of algorithms that allow the recovery of the geometric structure of a set of points from information related to their pairwise distances (for a concise and readily accessible review, see [13]). If the distances themselves are given, the problem is called metric MDS. In this case, the coordinates of the points can be recovered essentially by an eigenvalue decomposition of the distance matrix, the rank of which, in the absence of errors, must be equal to the dimensionality of the underlying space [18].

In many applications, the exact distances may be unknown, but a set of quantities monotonically related to the true distances may be available. The problem in this case is known as nonmetric MDS. Beals et al. [2] have identified a set of simple conditions that must be fulfilled by distance ranks if the points that gave rise to the distances are to be embeddable in a metric space. As we shall see below, both the metric and the nonmetric formulations of MDS are relevant to the problem of recovering depth from binocular disparities.

2.3 Metric MDS

Suppose that we are given the disparities d_i of a set of points (x, y, z) in a stereo pair, where z_i denotes the unknown depth value at the image point $(x, y)_i$. As indicated by equation 4, the pairwise differences between disparity values are monotonic with the differences between the corresponding depth values:

$$\Delta d_{ij} = d_i - d_j = C(z_i - z_j) \doteq C\Delta z_{ij} \tag{6}$$

where C is a constant for a fixed choice of α, l (for large enough l's; see Figures 2 and 3). Because of this monotonicity, a matrix of measurements of Δd_{ij} can be used in a metric MDS procedure to recover the unknown z_i's.

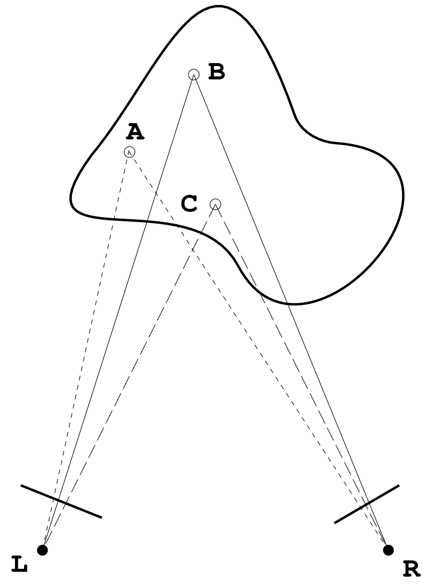

Figure 2: An intuitive illustration of the monotonicity of Δd_{ij} (the difference of disparities of pairs of points) in Δz_{ij} (their difference in depth), as expressed by eq. 6: $\angle ARB > \angle ALB$ because B is farther away than A; $\angle ARC < \angle ALC$ because C is closer than A.

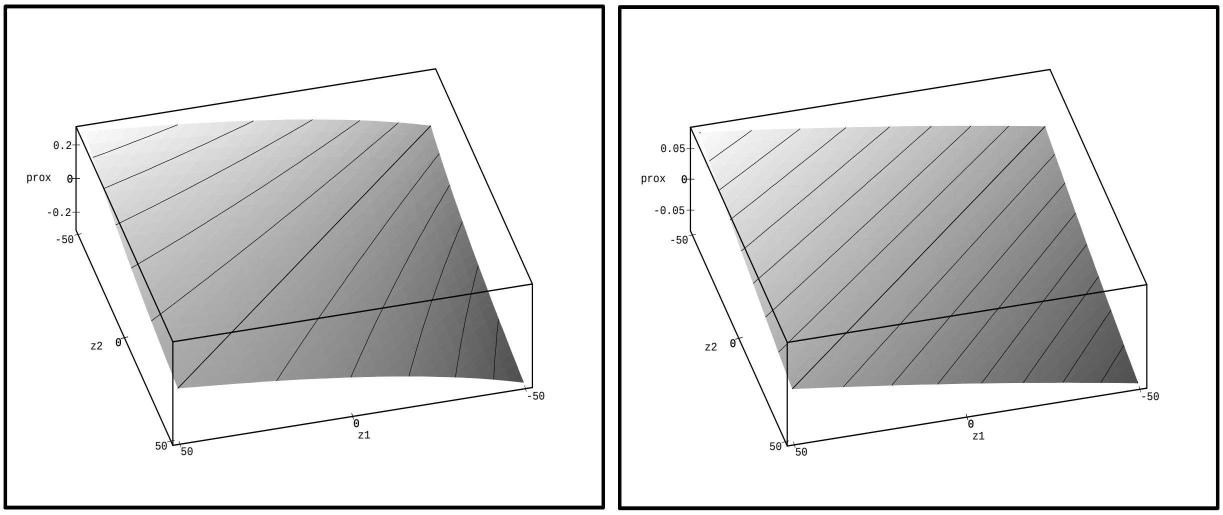

Figure 3: Plots of the iso-Δd_{ij} contours vs. z_i and z_j (see eq. 6), for two values of absolute depth l. In the left plot the ratio between the object size and the distance was 1 : 1.5; in the right plot it was 1 : 5. It can be seen that for large enough values of l (right panel), the contours are, for all practical purposes, parallel, and the approximation expressed by eq. 6 holds.

Even though this procedure seems to end up with the same information that is available in the initial data (d_i),[1] it actually allows one to circumvent the geometric distortion introduced into the

[1]In fact, the values are recovered up to a relief transformation. This ambiguity may be removed by combining results from two pairs of images, taken from different viewpoints.

disparity values by oblique viewing [10]. Mitchison and Westheimer [8] pointed out that the linear disparity gradient introduced by oblique viewing can be eliminated by basing the computation of depth on differences of disparities instead of the disparities themselves. This is precisely what the proposed method does.

2.4 Nonmetric MDS

Suppose now that instead of having the exact values of the disparity differences Δd_{ij} we only know their ranks. The imposition of this constraint may be motivated by a number of considerations:

- *Neurobiological data.* Binocular cells in the primary visual cortex in the monkey[2] seem to be better suited to signal the sign of disparity difference rather than its exact value.
- *Functional considerations.* Implementational factors such as imprecise binocular control of gaze [9] may create a bias in favor of the use of disparity ranks.
- *Psychophysical observations.* Experimental data on shape perception indicate that human subjects are better at perceiving quantities related to the rank order of points in depth (e.g., a global characterization of their distribution) than at perceiving the metrics of the viewed shapes (this result has been established for the perception of shape from shading [14], although it seems to hold also for binocular stereo [4]).

Surprisingly, retaining just the qualitative information (e.g., the ranks of Δd_{ij}) in the data submitted to MDS does not preclude the possibility of recovering the geometry of the configuration that gave rise to the data, provided that enough points are available. Extensive experience with nonmetric MDS shows that rank data involving a few dozen points suffice for an accurate recovery of their configuration in a low-dimensional space (see Figure 4 for demonstration). The only condition imposed on the inter-point difference data that serves as input to MDS is that they covary monotonically with the "real" distances to be recovered by the procedure [13]. Clearly, this condition is satisfied in the present case. In the rest of this section, we describe an application of nonmetric MDS to the recovery of depth information from qualitative disparity data computed over synthetic and natural images.

2.5 Synthetic data

Practical MDS algorithms usually work by starting with a random initial configuration of points, and by iteratively adjusting point coordinates so as to minimize a sum of squared differences between the computed and the given inter-point distances. In nonmetric MDS, the cost function penalizes instead deviations from the given rank order of the pairwise distances. In our implementation, the algorithm followed the monotone regression approach described in [7].

We first tested the nonmetric MDS stereo algorithm (NMDS) on artificial data sets generated by hand (see Figure 4). The results of these tests showed that NMDS is a viable approach to qualitative stereo. Specifically, the configuration recovered by the algorithm was very close to the true one with as few as eight points for which only the rank order of the disparities was assumed to be known.

[2]Of these there are two main varieties: cells tuned to zero disparity, and cells with a highly nonlinear sigmoidal response which prefer either crossed or uncrossed disparities [11].

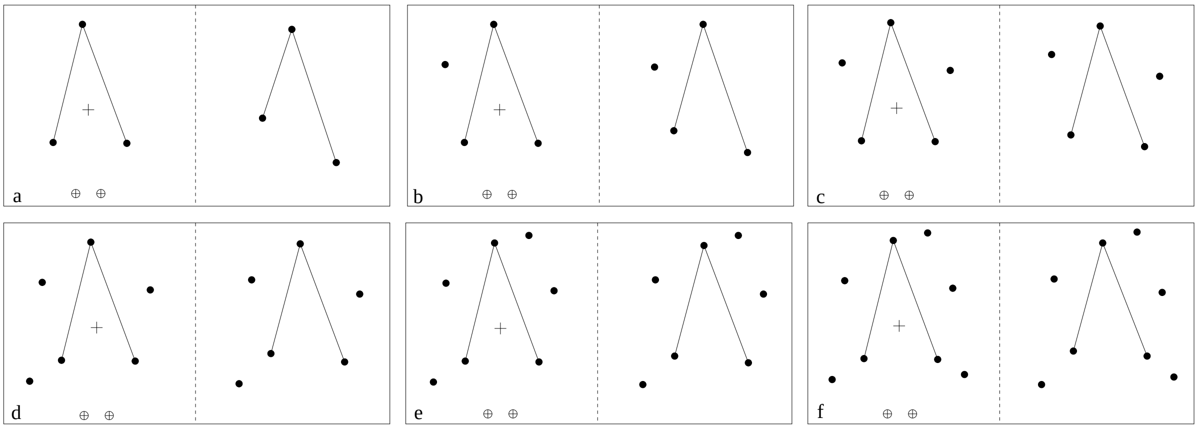

Figure 4: An illustration of the MDS-based stereo algorithm applied to synthetic disparity data. Each of the panels (*a* through *f*) shows the real configuration of a triplet of points (on the left) alongside with the configuration recovered from the rank order of Δd_{ij} (on the right). The total number of points varies between 3 and 8, at which stage the recovered configuration becomes difficult to distinguish from the true one. The eye separation is marked by $\oplus$ $\oplus$, and the fixation point by a $+$. The viewing distance (400 units) is not shown to scale.

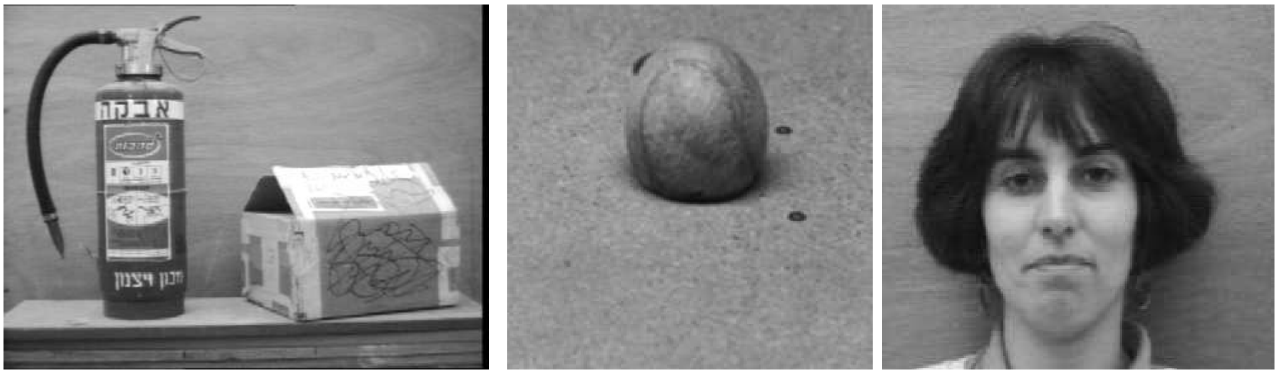

Figure 5: The left images of three of the natural image pairs used for testing the NMDS algorithm. Left: a fire extinguisher scene (original resolution 384×288 pixels). Middle: a ball on a slanting surface (256×256; courtesy of the JISCT database [3]). Right: a face (234×218 pixels).

2.6 Real stereo pairs: NEC matcher

We then tested the NMDS method on a few natural images using two different matching algorithms.[3] The first one is the intensity-based matcher described in [5].[4] This matcher produces a dense disparity field, in which the value of pixels designated as occluded is set to 0. To facilitate further processing and the display of the results, we have subjected this field to median filtering, with the zero-valued pixels excluded from the computation of the median. The filtered images were submitted to the MDS procedure.

[3] The choice of the matching algorithm is of secondary importance to the present work, which concentrates on what should be done with disparity information rather than on how to obtain that information.

[4] An implementation of this algorithm is publicly available via anonymous ftp from the vision archives at teleos.com

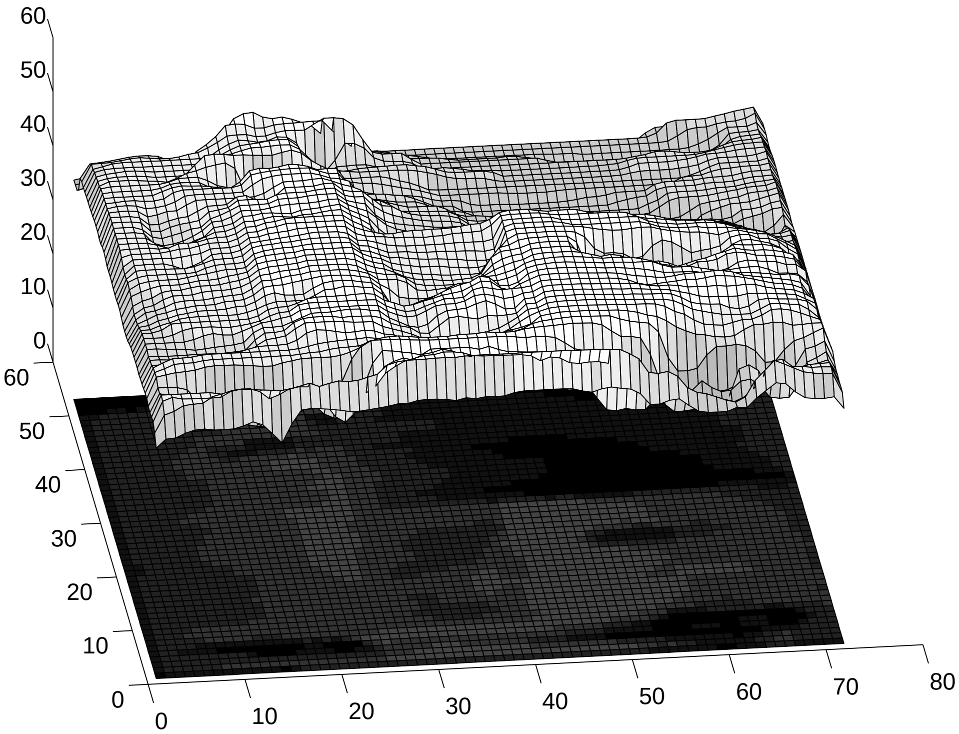

Figure 6: Fire extinguisher image: the output of the NEC matcher, followed by nonmetric MDS.

The results obtained with the NEC matcher are shown in Figures 6 through 8. The output of the matcher — a dense field of disparities in a 20×20 window — was converted into a 400×400 matrix of pairwise disparity differences (one for each pair of pixels in the window; only half of the matrix had to be filled). The differences were then ranked, and the ranks (**not** the differences themselves) were submitted to NMDS. The resulting depth map was computed for overlapping windows covering the entire image, and the values in the areas of overlap were used to obtain a globally consistent depth map.

The performance of the NMDS stage implementing the above procedure appears to be completely satisfactory. Despite being based on seemingly impoverished information — disparity difference ranks — the depth map recovered by the NMDS algorithm agrees closely with the true depth as it appears in Figure 5.

2.7 Real stereo pairs: RF-based matcher

In addition to the experiments with the NEC matcher, we implemented and tested a receptive-field (RF) matching algorithm whose main advantages are simplicity and relevance to biological vision. Each image was represented by a large number (typically, 20,000) of linear RFs, each of which consisted of several Gabor filters of different orientation (the response of each RF was therefore a vector; cf. [6]). To add biological realism, the RFs were highly overlapping and were placed at random locations chosen independently (for each RF and each "eye") from a uniform distribution. Each left-image RF was matched to that right-image RF which resided in a 40×6 window centered on the exactly corresponding location and which had the most similar response vector. Consistent

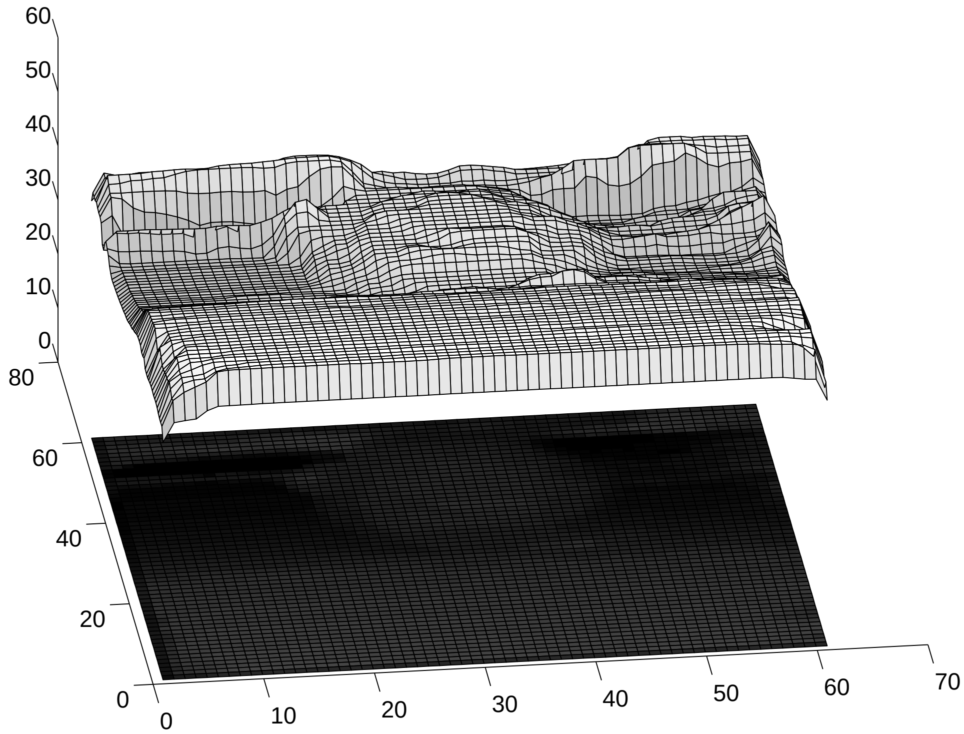

Figure 7: Ball image: the output of the NEC matcher, followed by nonmetric MDS.

L-R and R-L matches were combined into a relatively sparse field of estimated disparity values, and were passed through the median filter described above.

The results of running nonmetric MDS on two of the test pairs processed with the RF matcher appear in Figure 9. The performance of the algorithm on the fire extinguisher pair is similar to the results obtained with the NEC matcher. In the ball pair, the outline of the ball appears distorted, due to excessive smoothing introduced by the matcher, For the face image pair (not shown) the performance was poor, due to the high false match rate of the RF stage (combining matches across several spatial scales could remedy this problem).

To summarize, tests conducted on synthetic and natural images indicate that the NMDS algorithm is capable of recovering depth from qualitative information in the form of the rank order of pairwise disparity differences taken over all the points of interest in the input. The performance of the algorithm appears to be comparable to that of state of the art stereo algorithms, and, in the present experiments, is limited essentially by the quality of the disparity input to the NMDS stage. Obviously, the NMDS approach can be fully exploited only in conjunction with a scheme for qualitative measurement of disparity (which should be less error-prone than the standard quantitative matchers). Other ways to make the most out of NMDS in the recovery of depth are mentioned below.

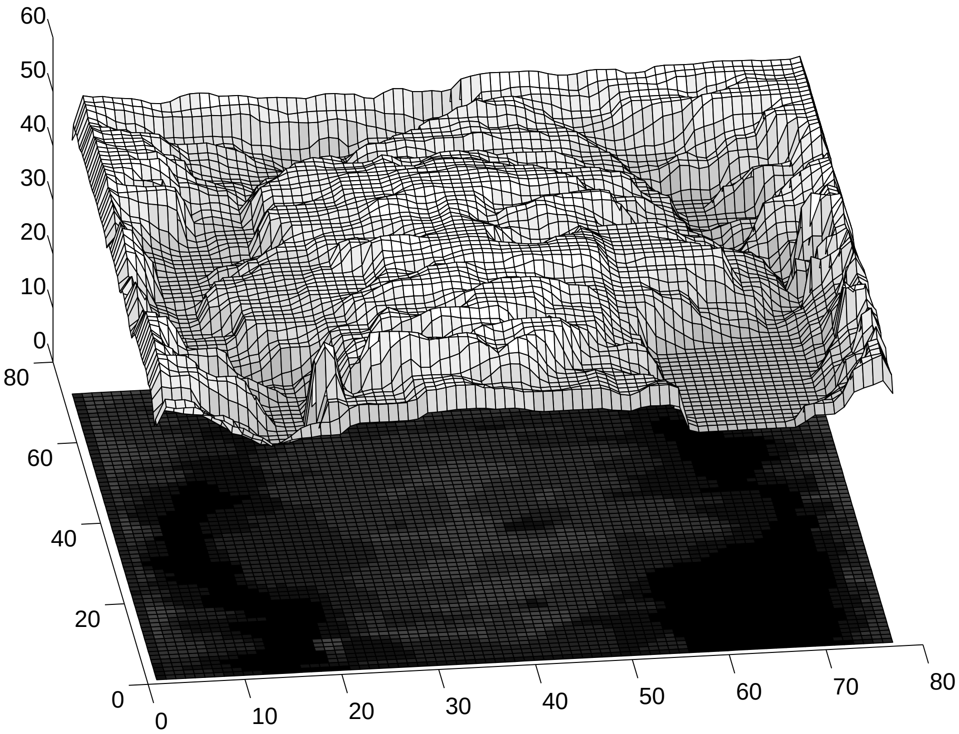

Figure 8: Face image: the output of the NEC matcher, followed by nonmetric MDS.

3 Discussion

The application of multidimensional scaling to the recovery of visual depth opens up a new avenue of research which may be called quantitative qualitative vision. The main computational idea behind this approach is that combination of qualitative measurements from a number of mutually related sources can support the recovery of the underlying quantitative information. In the present case, the measurements are used to form the rank order of depth differences, the multiple interrelated sources are simply the various point pairings for which the measurements are made, and the final result of the computation is the values of depth at those points.

The above formulation of the present approach in terms of combining measurements from multiple sources hints at the possibility of using MDS to integrate the recovery of depth from a number of visual cues within the same computational framework. This common framework is provided by the iterative MDS algorithm, which strives to minimize the discrepancy between measured depth (actually, disparity or disparity rank) differences and those derived from the current configuration. For example, it should be possible to integrate shading with stereo simply by including input from shading as an additional term in the MDS cost function. Integration of motion information may also be possible (note that Ullman's incremental rigidity algorithm for the recovery of structure from motion [15] minimizes a cost function expressed in terms of inter-point distances).

The proposed approach to the recovery of depth from stereo has interesting biological connotations. Psychophysicists in the past have advanced computational arguments in favor of the use of disparity differences instead of raw disparities, based on observations of human performance in stereopsis [17, 8, 12]. The present work indicates that multidimensional scaling (both metric and

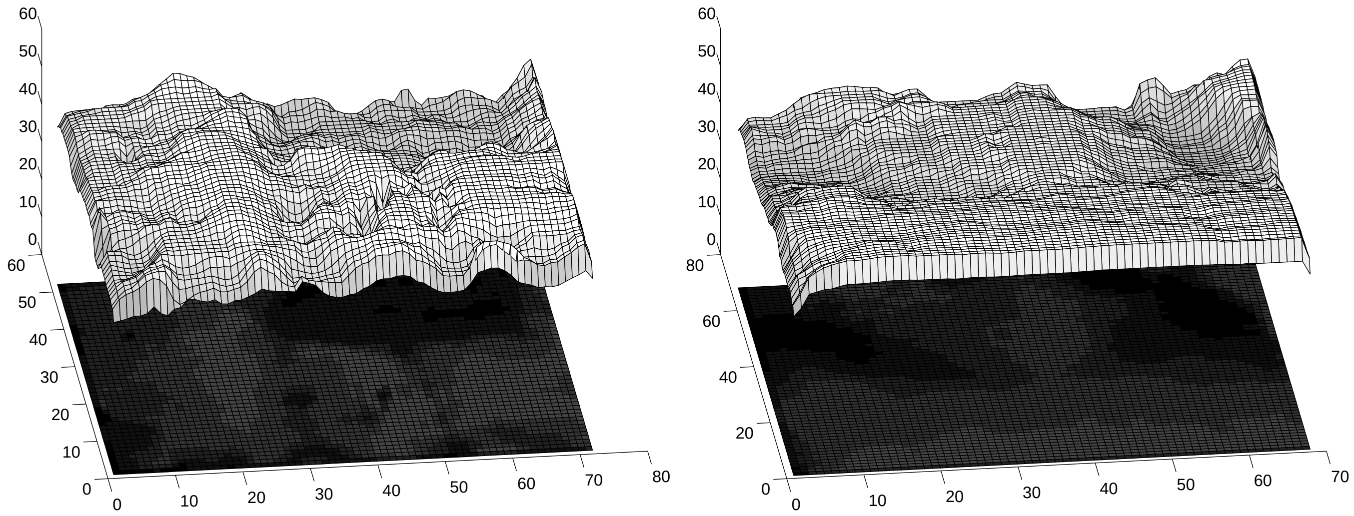

Figure 9: Left: fire extinguisher image, the output of the RF-based matcher, followed by nonmetric MDS. Matching was done with $20,000$ RFs, each of which was composed of 5 Gabor filters at equally spaced orientations, with $\sigma = 8$ pixels and period$= 4$ pixels. Right: ball image, the output of the RF-based matcher, followed by nonmetric MDS (same RF parameters as before).

nonmetric) constitutes a promising approach to the integration of biological and computational insights into the problem of 3D vision.

Acknowledgements

We thank Shimon Ullman for useful discussions, and the NEC vision research group for making their stereo matching software publicly available.

References

[1] J. Y. Aloimonos. Purposive and qualitative vision. In *Proc. AAAI-90 Workshop on Qualitative Vision*, pages 1–5, San Mateo, CA, 1990. Morgan Kaufmann.

[2] R. Beals, D. H. Krantz, and A. Tversky. The foundations of multidimensional scaling. *Psychological Review*, 75:127–142, 1968.

[3] R. C. Bolles, H. H. Baker, and M. J. Hannah. The JISCT stereo evaluation. In *ARPA Image Understanding Workshop*, pages 263–274, 1993.

[4] H. H. Bülthoff and H. A. Mallot. Integration of stereo, shading and texture. In A. Blake and T. Troscianko, editors, *AI and the Eye*. Wiley, London, UK, 1990.

[5] I. J. Cox, S. Hingorani, B. M. Maggs, and S. B. Rao. Stereo without disparity gradient smoothing: a Bayesian sensor fusion solution. In *British Machine Vision Conf.*, pages 337–346, Berlin, 1992. Springer-Verlag.

[6] D. G. Jones and J. Malik. A computational framework for determining stereo correspondence from a set of linear spatial filters. In G. Sandini, editor, *Proc. ECCV-92*, pages 395–410, Berlin, 1992. Springer.

[7] J. B. Kruskal. Non-metric multidimensional scaling: a numerical method. *Psychometrika*, 29:115–129, 1964.

[8] G. J. Mitchison and G. Westheimer. Viewing geometry and gradients of horizontal disparity. In C. Blakemore, editor, *Vision: coding and efficiency*, chapter 28, pages 302–309. Cambridge University Press, 1990.

[9] B. C. Motter and G. F. Poggio. Binocular fixation in the Rhesus monkey: spatial and temporal characteristics. *Exp. Brain Res.*, 54:304–314, 1984.

[10] K. N. Ogle. *Researches in Binocular Vision.* Hafner, New York, 1950.

[11] G. F. Poggio and T. Poggio. The analysis of stereopsis. *Ann. Rev. Neurosci.*, 7:379–412, 1984.

[12] B. Rogers and R. Cagenello. Disparity curvature and the perception of three-dimensional surfaces. *Nature*, 339:135–137, 1989.

[13] R. N. Shepard. Multidimensional scaling, tree-fitting, and clustering. *Science*, 210:390–397, 1980.

[14] J. T. Todd and F. Reichel. Perception of ordinal depth relations from patterns of shading. *J. Exp. Psychol.: HPP*, 16:583–595, 1990.

[15] S. Ullman. Maximizing rigidity: the incremental recovery of 3D structure from rigid and rubbery motion. *Perception*, 13:255–274, 1984.

[16] D. Weinshall. Qualitative depth from stereo, with applications. *Computer Vision, Graphics, and Image Processing*, 49:222–241, 1990.

[17] G. Westheimer. Cooperative neural processes involved in stereoscopic acuity. *Exp. Brain Res.*, 36:585–597, 1979.

[18] G. Young and A. S. Householder. Discussion of a set of points in terms of their mutual distances. *Psychometrika*, 3:19–22, 1938.