Using Markov Random Field to Integrate Stereo Modules

Kok-Guan Lim and Richard Prager

CUED/F-INFENG/TR 109

July 30, 1992

Cambridge University Engineering Department
Trumpington Street
Cambridge CB2 1PZ
England

Email : kgl/rwp@eng.cam.ac.uk

Using Markov Random Field to Integrate Stereo Modules

Kok-Guan Lim and Richard Prager
Email : kgl/rwp@eng.cam.ac.uk

Cambridge University Engineering Department
Trumpington Street, Cambridge CB2 1PZ

July 30, 1992

Abstract

We present a new method using a Markov Random Field (MRF) to integrate edge and intensity based stereo algorithms. First, we derive the intensity based stereo algorithm under the MRF framework. The integration is then performed by coupling the disparity estimates from an independent edge based stereo module to the energy functional of the MRF, associated with the intensity based stereo algorithm. The maximum *a posteriori* estimate of the resulting MRF is obtained using the mean field annealing algorithm. Results from real and artificial images show a consistent improvement in the accuracy under this scheme of integration.

1 Introduction

Psychophysical experiments like [5] have shown strong evidence that the human visual system is able to integrate information from several sources for its visual processing tasks, such as 3D surface reconstruction. Integration of information from several sources allows the results from various visual cues to inter-validate each other, hence giving more reliable and accurate aggregate results.

Encouraged by the same argument, workers in computational vision are trying to integrated various low-level visual modules in building a more robust vision system [2]. More importantly, a unified framework based on the Bayes' theorem and MRF model has been used successfully in a few examples of integration. Under the MRF framework, Poggio *et al.* [13] integrated several low level vision modules to compute the discontinuity map of a 3D scene. In the context of stereo vision, Nasrabadi *et al.* [12] integrated intensity and optical flow information, while Yuille *et al.* [15] integrated intensity and feature based stereo. In this paper, we describe a new method using a MRF to integrate the results of edge based stereo with an intensity based stereo algorithm.

In our scheme of integration, a sparse disparity map is first obtained from an independent edge based stereo module. Next, we formulate our intensity based stereo algorithm using the MRF framework. Finally, in order to integrate the edge based results, we couple the sparse disparity map to the energy functional of the intensity based MRF as an external field.

Conventional methods to find the maximum *a posteriori* (MAP) estimate of a MRF are based on stochastic relaxation methods like simulated annealing [10] [9]. These techniques are able to guarantee convergence close to a global minimum of the energy functional associated with the MRF[1] [9]. They are however very slow. More recently, several new methods based on the mean field approximation have been developed to obtain an approximate solution [4] [8]. With these methods, the solution can usually converge towards the ground state more than 50 times faster than simulated annealing. In particular, we use the method of mean field annealing described by Bilbro *et al.* [4], which can be seen as a deterministic form of simulated annealing.

We tested our implementations with computer generated stereo images and an outdoor stereo scene. Comparison of results in Section 3 shows a consistent improvement after the integration of edge and intensity based modules. Furthermore, this work has demonstrated an effective scheme using a MRF model to integrate two low level visual modules.

[1]finding MAP estimate of a MRF is equivalent to minimizing its energy functional

2 Theory and Implementation

2.1 Markov Random Field for Intensity based Stereo

Like most low level vision problems, stereo correspondence may be expressed as an energy function minimization. Given left and right images data $\mathbf{g}$, if we need to estimate the disparity field $\mathbf{d}$, the energy functional for intensity based stereo may take the form [14] :

$$E(\mathbf{d} \mid \mathbf{g}) = \sum_{i,j} \left[F(\mathbf{g}, d_{ij}) + \alpha \sum_{n \in N} (d_{ij} - d_n)^2 \right] \tag{1}$$

$F(\mathbf{g}, d_{ij})$ represents the matching cost based on the intensity value of a small matching window around image coordinate (i, j). It is usually a correlation measure or sum of squared difference (SSD). The second expression in the energy functional is the cost associated with the smoothness constraint. It is expressed in terms of the disparity values of the neighbouring points d_n, N indicates the 8 neighbourhood. α is the parameter controlling the degree of smoothness of solution. It has a similar function to the regularization constant used in [14].

Since the disparity field d_{ij} at any site (i, j) depends solely on its neighbouring disparity values d_n, the disparity field $\mathbf{d}$ is a MRF. Its corresponding Gibbs distribution is given by :

$$P(\mathbf{d} \mid \mathbf{g}) = \frac{1}{Z} \exp\left(-\frac{E(\mathbf{d} \mid \mathbf{g})}{T}\right) = \frac{1}{Z} \exp\left(-\frac{1}{T} \sum_{i,j} \left[F(\mathbf{g}, d_{ij}) + \alpha \sum_{n \in N} (d_{ij} - d_n)^2 \right]\right) \tag{2}$$

where Z is the normalization constant called the partition function. Under Bayesian interpretation, the Gibbs distribution is equivalent to the *a posteriori* distribution of disparity field $\mathbf{d}$ given image data $\mathbf{g}$:

$$P(\mathbf{d} \mid \mathbf{g}) = \frac{P(\mathbf{g} \mid \mathbf{d}) P(\mathbf{d})}{\sum_{\mathbf{d}} P(\mathbf{g} \mid \mathbf{d}) P(\mathbf{d})} \tag{3}$$

where $P(\mathbf{g} \mid \mathbf{d})$ is the likelihood of $\mathbf{d}$ given data $\mathbf{g}$, and is associated with the matching cost $F(\mathbf{g}, \mathbf{d})$. $P(\mathbf{d})$ is the *a priori* distribution of the disparity field $\mathbf{d}$, and is corresponding to the smoothness term in the energy functional domain which represents the *prior* assumption of the solution $\mathbf{d}$.

Finding the MAP estimate of the Gibbs distribution is equivalent to minimizing its energy functional. Kirkpatrick *et al.* [10] implemented the Metropolis procedures as a simulated annealing algorithm to search for the global minimum of the energy functional of a system. A variation of this is the algorithm based on a Gibbs Sampler by Geman and Geman [9]. They have proved that the solution converges asymtoptically to the Gibbs distribution if a sufficiently large number of iterations are performed [9].

While the stochastic relaxation methods are able to escape from the system local minima, they are very slow to converge. In the next section, we will describe a deterministic form of annealing scheme known as mean field annealing. It was first described in detail by Bilbro *et al.* [4] and has since been applied to numerous image processing tasks like image restoration [3] and stereo [6].

2.2 Mean Field Annealing

Just like the simulated annealing algorithm, mean field annealing consists of a relaxation operation which iteratively finds the system equilibrium states at a sequence of temperatures, decreasing over time under a specific cooling schedule. However, the mean field annealing algorithm attains the equilibrium states using a deterministic approach, in contrast with the stochastic nature of simulated annealing.

In simulated annealing, the ground state of the system is achieved through many probabilistic transitions of every site according to the local characteristic of the site (described by the Gibbs density function). In mean field annealing, probabilistic transitions of a Markov field are replaced by direct evaluation of the average field at each site, given the local probability density function. In this way, the system is able to attain its equilibrium state at a much faster rate leading to a more efficient algorithm, typically 50 times faster [4]. Using our example of intensity based stereo in Section 2.1, the mean field annealing algorithm consists of the following steps [6]:

1. Start with a high temperature (5.0 or higher) and initialize the mean disparity field $\bar{\mathbf{d}}$ with random values taken from the disparity range of R_d.

2. Visit a site (i, j) and calculated its marginal *a posteriori* density function. Using the energy functional from (1), we obtain :

$$P(d_{ij} \mid d_n, \mathbf{g}) = \frac{1}{Z_{ij}} \exp\Big(-\frac{E(d_{ij} \mid d_n, \mathbf{g})}{T}\Big) = \frac{1}{Z_{ij}} \exp\Big(-\frac{1}{T}\big[F(\mathbf{g}, d_{ij}) + \alpha \sum_{n \in N} (d_{ij} - d_n)^2\big]\Big) \tag{4}$$

where Z_{ij} is the localized partition function :

$$Z_{ij} = \sum_{d_{ij} \in R_d} \exp\Big(-\frac{E(d_{ij} \mid d_n, \mathbf{g})}{T}\Big) \tag{5}$$

3. Calculate the mean disparity field $\bar{d_{ij}}$ as :

$$\bar{d_{ij}} = \sum_{d_{ij} \in R_d} d_{ij} P(d_{ij} \mid d_n, \mathbf{g}) = \sum_{d_{ij} \in R_d} d_{ij} \frac{1}{Z_{ij}} \exp\Big(-\frac{1}{T}\big[F(\mathbf{g}, d_{ij}) + \alpha \sum_{n \in N} (d_{ij} - d_n)^2\big]\Big) \tag{6}$$

4. Iterate steps 2 and 3 for the whole image until equilibrium at the current temperature is attained. (see text for determination of an equilibrium state)

5. Lowering the current temperature and repeat step 2 to 4 until a sufficiently low temperature is reached (usually 0.001 or lower). The temperature may be reduced according to the cooling schedule of :

$$T(k+1) = 0.7 \times T(k) \tag{7}$$

There are various methods to detect the state of equilibrium for the relaxation. We do this by measuring the average number of level of transition between two subsequent iterations. When the average transition falls below a certain small value Δ, an equilibrium state is assumed and iteration at the current temperature is terminated. For an image size of $n_i \times n_j$ and if the number of possible disparity levels is n_d, the equilibrium state is detected by checking the following condition :

$$\mathbf{d}^{\sim} = \sum_{i,j} \frac{d_{i,j}(k+1) - d_{i,j}(k)}{n_i \times n_j \times n_d} \quad < \quad \Delta \tag{8}$$

If a very small value of Δ is used (0.0001 or less), the value of $\mathbf{d}^{\sim}$ may oscillate near freezing temperature. When this happen, the iteration should be terminated after a fixed number of iterations (about 10).

2.3 The Edge Based Stereo

In our implementation of an edge based stereo module, we use a Deriche edge detector to extract the gradient and direction of edges in left and right images. Edgel[2] points with the same direction (to within 22.5°) and similar gradient are considered as potential matches. The best match is then selected from these potential matches using the matching algorithm described by Drumheller

[2] an edgel is an edge having the extent of one pixel.

and Poggio [7]. Only a sparse disparity map is obtained from this method as no interpolation is performed.

Next, the results from the edge based matching are corrected by imposing figural continuity constraint [11]. Figural continuity constraints require disparity along an edge contour[3] to vary smoothly. When the disparity values from the same edge contour are plotted against the length of the contour, we derive a best-fitted straight line. Using this line, any wrong disparity value far away from the line is treated as an outlier and is replaced by a value interpolated from the line.

2.4 Integration of Results from Edge Based Stereo using a MRF

In section 2.1, we have shown how to formulate an intensity based stereo algorithm using a MRF model. With this model, together with the disparity estimates given by the edge based stereo, we are now ready to integrate the edge based results with the intensity based module. Our strategy is to couple the sparse disparity map from the edge based stereo as an external field to the energy functional of the original MRF (see Figure 1). From Equation (1), the energy functional of intensity based matching is repeated here as :

$$E(\mathbf{d} \mid \mathbf{g}) = \sum_{i,j} \Big[F(\mathbf{g}, d_{ij}) \quad + \quad \alpha \sum_{n \in N} (d_{ij} - d_n)^2 \Big]$$

At the sites where the edge based results d_e are available, the energy functional is modified as :

$$E(\mathbf{d} \mid \mathbf{d}_e, \mathbf{g}) = \sum_{i,j} \Big[F(\mathbf{g}, d_{ij}) \quad + \quad F(\mathbf{g}, d_{ij}) \cdot \psi \cdot \mid d_{ij} - d_e \mid \quad + \quad 2\alpha \sum_{n \in N} (d_{ij} - d_n)^2 \Big] \qquad (9)$$

where ψ is a normalization parameter for the value $\mid d_{ij} - d_e \mid$ which measures the deviation of current disparity estimate from edge based results d_e. The cost for this deviation is weighted by $F(\mathbf{g}, d_{ij})$, the matching cost of intensity based stereo. Therefore, the resultant cost is high only when both $\mid d_{ij} - d_e \mid$ and $F(\mathbf{g}, d_{ij})$ are high. Intuitively, the purpose of this weighting is to *cushion* the effect of any false match caused by the edge based algorithm.

To gain further insight into our integration scheme, we show how the Bayes' theorem can be used to justify the energy functional in (9). Given the image data $\mathbf{g}$ and edge based result $\mathbf{d}_e$, the *a posteriori* distribution of $\mathbf{d}$ becomes[4] :

$$P(\mathbf{d} \mid \mathbf{d}_e, \mathbf{g}) = \frac{P(\mathbf{d}_e, \mathbf{g} \mid \mathbf{d}) P(\mathbf{d})}{\sum_{\mathbf{d}}} \qquad (10)$$

If we treat $\mathbf{d}_e$ as an external field, then the equivalent interpretation in probabilistic terms implies that $\mathbf{d}_e$ and $\mathbf{g}$ are independent. Thus we may write :

$$\frac{P(\mathbf{d}_e, \mathbf{g} \mid \mathbf{d}) P(\mathbf{d})}{\sum_{\mathbf{d}}} = \frac{P(\mathbf{d}_e \mid \mathbf{d}) P(\mathbf{d}) \cdot P(\mathbf{g} \mid \mathbf{d}) P(\mathbf{d})}{\sum_{\mathbf{d}}} \qquad (11)$$

Expressing this result in the domain of energy functional, equation (9) can be rewritten as :

$$\begin{aligned} E(\mathbf{d} \mid \mathbf{d}_e, \mathbf{g}) &= E(\mathbf{d} \mid \mathbf{d}_e) + E(\mathbf{d} \mid \mathbf{g}) \\ &= \sum_{i,j} \Big[F(\mathbf{g}, d_{ij}) \cdot \psi \cdot \mid d_{ij} - d_e \mid + \alpha \sum_{n \in N} (d_{ij} - d_n)^2 \Big] + \sum_{i,j} \Big[F(\mathbf{g}, d_{ij}) + \alpha \sum_{n \in N} (d_{ij} - d_n)^2 \Big] \end{aligned} \qquad (12)$$

[3] contour is defined herein as a segment of connected edge bounded by corners or the end of the curve.

[4] $\sum_{\mathbf{d}}$ denotes the normalization constant of the density function.

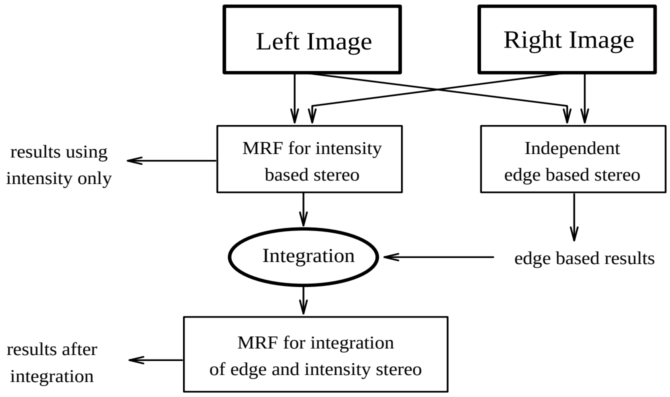

Figure 1: Scheme of integration of edge and intensity based stereo modules

3 Results

We tested our implementations with a computer generated stereo pair and the 256×256 Pentagon stereo pair. For the computer generated stereo pair, the right image is derived from the intensity values of its left image, added with Gaussian noise. Next, the pixels in the image are shifted to the right or left, giving rise to the effect of a disparity map as shown in Figure 3**c**. Given the true disparity map, we can then measure the percentage of the error of the results from each algorithm. The error of the disparity estimate given by the intensity based stereo (Figure 3**e**) is found to be 16% , compared to 11% after the integration of edge based stereo results (Figure 3**f**).

For the Pentagon stereo pair, results from the intensity based stereo (Figure 2**d** and **f**) is similar to the results obtained by Chang and Chatterjee [6]. However, the results obtained with the integration of edge and intensity based stereo (Figure 2**e** and **g**) are shown to be consistently better.

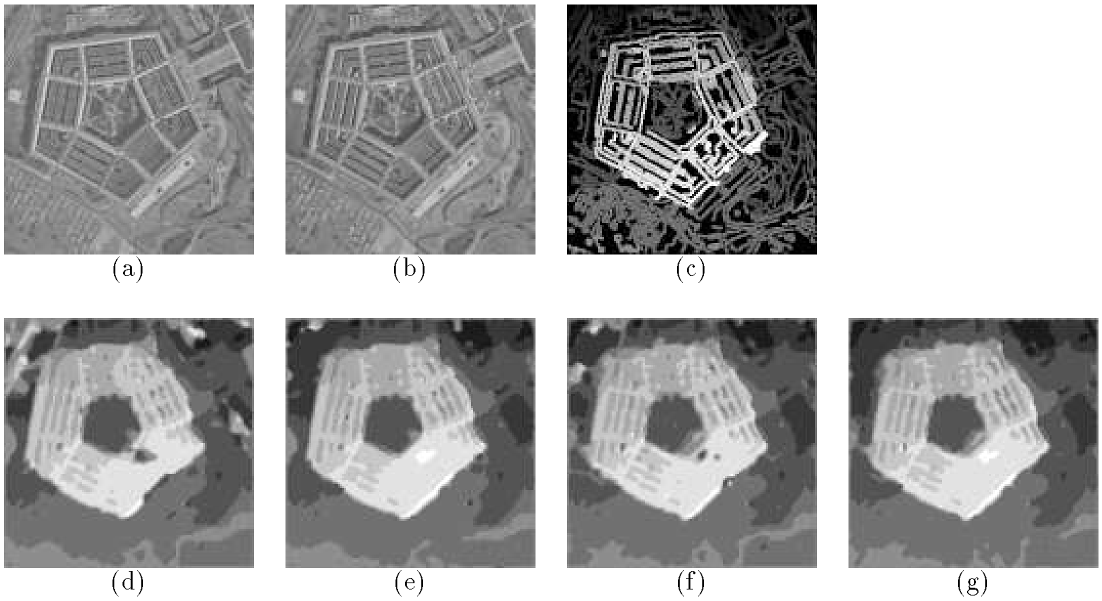

Figure 2: Results of The Pentagon Stereo Pair **a**. left image **b**. right image **c**. results of edge based stereo. **d**. results of intensity based stereo, using correlation window as matching primitive. **e**. result with integration of edge and intensity modules, using correlation window. **f**. results of intensity based stereo, using SSD as matching primitive. **g**. result with integration of edge and intensity modules, using SSD.

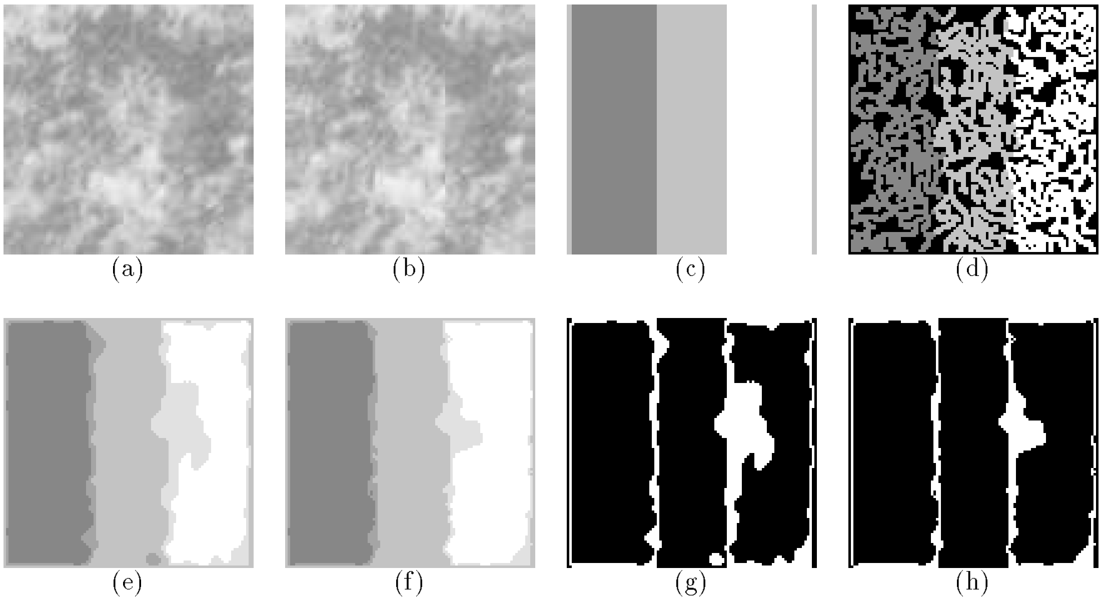

Figure 3: Results of computer generated stereo pair **a**. left image **b**. right image **c**. disparity used to generate the stereo pair (disparity used: left region 2, middle 0 and right -2). **d**. results of edge based stereo. **e**. results of intensity based stereo, using SSD as matching primitive. **f**. result with integration of edge and intensity modules, using SSD. **g**. error of the results in **e** (shown as white). **h**. error of the results in **f**.

4 Conclusions and Future Work

The MRF model is shown to be an effective framework for the integration of low level visual modules such as stereo. Using this framework, we have derived a new scheme of integrating edge based and intensity based stereo. Results from our implementations have demonstrated a consistent increase in matching accuracy after the integration. The success of this scheme stem from the fact that edge based stereo is more robust to the noise and imperfect epipolar alignment in stereo images. On the other hand, results from intensity based matching degrade rapidly with these two factors. By integrating the edge based results, we can reduce the number of false matches, hence increase the system accuracy.

One unique feature of our scheme is that the edge based results are generated by an independent algorithm. This allows one to impose higher level physical constraints like the figural continuity constraint. Our scheme works best on images with rich edge information, and when a fairly reliable edge based stereo algorithm is used. When the amount of edge information in the images is too low, other visual cues like shading become more important. In future, we plan to extend this framework to include shading information and other monocular cues.

Interestingly, the computational scheme of simulated annealing leads naturally to neural network processing models such as the Hopfield network and Boltzmann machine [4] [1]. Using MRF framework and Bayes' theorem, it is very likely that low level vision algorithms and their integration may be mapped onto a neural network model leading to useful theoretical insights.

References

[1] Aarts, E. H. L. and Korst, J. "Simulated annealing and Boltzmann machines" Chichester: Wiley, 1989.

[2] Aloimonos, J. and Brown, C. M. "Robust Computation of Intrinsic Images from multiple cues" in *Advances in Computer Vision* (Brown, C. M. ed.), ch. 2, Hillsdale NJ: Lawrence Erlbaum Associates, 1987.

[3] Bilbro, G. and Snyder, W. E. "Range image restoration using mean field annealing" in *Advances in Neural Information Processing Systems I* (Touretzky, D. S. ed.), p54–61, Morgan Kaufman, 1989.

[4] Bilbro, G., Snyder, W. E., Mann, R., van den Bout, D. E., Miller, T. K. and White, M. "Optimization by mean field annealing" in *Advances in Neural Information Processing Systems I* (Touretzky, D. S. ed.), p91–8, Morgan Kaufman, 1989.

[5] Bulthoff, H. and Mallot, H. "Integration of depth modules: stereo and shading" *J. Optical Society of America*, **5**, p1749–58, 1988.

[6] Chang, C. and Chatterjee, S. "A Deterministic Approach for stereo disparity calculation" in *Proc. European Conf. on Computer Vision*, 1992.

[7] Drumheller, M. and Poggio, T. "On Parallel Stereo" in *IEEE Intl. Conf. Robotics and Automation*, p1439–48, April 1987.

[8] Geiger, D. and Girosi, F. "Parallel and deterministic algorithms from MRF's: surface reconstruction" *IEEE Trans. on Pattern Analysis and Machine Intelligence*, **13(5)**, p401–12, 1991.

[9] Geman, S. and Geman, D. "Stochastic relaxations, Gibbs distributions and the Bayesian restoration of images" *IEEE Trans. on Pattern Analysis and Machine Intelligence*, **6(6)**, p721–41, 1984.

[10] Kirkpatrick, S., Gelatt, C. and Vecchi, M. "Optimization by simulated annealing" *Science*, **220**, p219–27, 1983.

[11] Mohan, R., Medioni, G. and Nevatia, R. "Stereo error detection, correction and evaluation" in *Proc. 1st International Conf. on Computer Vision*, p315–23, 1987.

[12] Nasrabadi, N. M., Clifford, S. P. and Liu, Y. "Integration of stereo vision and optical flow by using an energy-minimization approach" *J. Optical Society of America*, **6(6)**, p900–7, 1989.

[13] Poggio, T., Gamble, E. and Little, J.J. "Parallel integration of visual module" *Science*, **242**, p436–439, 1988.

[14] Poggio, T., Torre, V. and Koch, C. "Computational vision and regularization theory" *Nature*, **317**, p314–319, 1985.

[15] Yuille, A. L., Geiger, D. and Bulthoff, H. "Stereo Integration, Mean Field Theory and Psychophysics" in *Proc. European Conference on Computer Vision*, p87–98, 1990.