Proceedings of the Int. Conf. on Parallel Processing in Neural Systems and Computers (ICNC), Düsseldorf, (Elsevier, 1990) (in press)

Hierarchical Spin Model for Stereo Interpretation Using Phase Sensitive Detectors[†]

R. Divko and K. Schulten*
Physik-Department
Technische Universität München, D-8046 Garching

The property of physical spin systems to assume long range order is exploited to establish global depth interpretation of real scene stereo pairs. We choose spins to code for local disparity between corresponding pixels in both pictures [1]. The energy of the spin system entails a field contribution, which incorporates the information on the stereo images, and an attractive interaction between alligned neighboring feature spins, which implements the continuity propertiy of disparities. The energy function is chosen in a way that the ground state of the feature spin system – found by simulated annealing [2] – corresponds to the best global depth interpretation of the stereo pair. Disparity detectors, based on relative phase measurements between left and right picture, yield information on the local field.

1. Introduction

Our brain is effortlessly inferring depth from the two slightly different images of the world formed by our eyes. The difference of the position of objects in both images is called disparity and can be used to obtain a depth map of the observed scene. Current computer algorithms solving the stereo correspondence problem are based on two different approaches. Feature based models use specific features as matching primitives, e.g. zero crossings, proposed by Marr and Poggio [3]. Correspondence-less algorithms are not based on matching features but use cross-correlation between parts of each image. For example algorithms proposed by Jenkin [4] and Sanger [5] locally compute disparity from complex phase difference obtained from the Gabor transformed [6] images. Another classification scheme of algorithms evolves from the question, wether the algorithms are local [4, 5] or involve cooperation between neighboring features [3]. We showed previously [1], that with cooperative spin algorithms constraints obtained from 'apriori' knowledge on the images, e.g. continuity of depth, can easily be implemented. In the following we propose a hierarchical stochastic algorithm analog to physical spin systems with interaction between neighboring features and a field interaction based on the local measurements of phase detecting filters. The algorithm proposed belongs to a class of regularization methods [7] known as Markov random fields [8].

2. Feature Spins

In physics a spin system is characterized by spin variables s_{ij}, which can assume a certain set of values, and by the energy function of the system. In the two-dimensional Ising model the spin variables take the values ± 1 and the energy of the system is

$$E = -J \sum_{<(i,j),(k,l)>} s_{ij}\, s_{kl} - \sum_{(i,j)} h_{ij}\, s_{ij} \tag{1}$$

where the brackets indicate summation over next neighbors. In the ferromagnetic case ($J > 0$)

[†]This work has been supported by BMFT ITR-8800-G9

*Department of Physics and Beckman Institute, University of Illinois, Urbana, IL, 61801 USA

the first term gives a negative contribution for equal neighboring spins and, therefore, establishes a tendency for an alignment of the spins. The second term describes the interaction of the spins with a local magnetic field h_{ij} tending to align the spins locally with the field.

To use spin systems for pattern recognition, features, e.g. the disparity between corresponding pixels, are coded in spins. The pattern to be processed is incorporated in the field. Global 'apriori' knowledge, like continuity constraints, on the pattern to be interpreted is implemented in the interaction between the feature spins. The energy of the feature spin system is chosen such, that the ground state, i.e. the state of the feature spins with the lowest total energy, corresponds to the best global interpretation of the pattern. It is achieved by the standard method of Monte Carlo annealing [2].

3. Phase Detecting Filters

The external field serves to communicate the pattern to be processed to the system of feature spins. In our model we use the output of filters capable of computing local phase differences between both images of the stereo pair [4, 5]. These filters yield confident disparity information only on sparsely distributed regions of the image.

Disparity detecting filters are based on the property of the Fourier transform, that two functions globaly shifted against each other by the ammount $\delta(x)$ yield a phase factor of $e^{i\omega\,\delta(x)}$, where ω is the frequency. This property can be generalized to the case of a Gabor transform, i.e. a Fourier transform weighted with a Gaussian, if the width of the Gaussian is smaller than the relative shift. If $l(x)$ and $r(x)$ are the left and right image repectively, with a local relative disparity $\delta(x)$ between them, the Gabor transforms $L(x,\omega)$ and $R(x,\omega)$ of both pictures are locally related to each other:

$$R(x,\omega) \approx \exp^{i\omega\delta(x)} L(x,\omega).$$

Therefore $\delta(x) = \frac{1}{\omega} arg(\frac{R(x,\omega)}{L(x,\omega)})$ is a measure for the local disparity. The value of $c = min(|\frac{R}{L}|, |\frac{L}{R}|)$ is a confidence measure. Values not equal to one are indicating a false disparity value.

4. A Hierarchical Algorithm for Stereo Vision

Because the disparity filters yield confident results only if the maximum disparity to detect is smaller than the wavelength of the Gabor filter, a hierarchical process from larger to smaller wavelengths is used. The number of waves within the Gaussian is kept constant. Beginning with a sufficiently large wavelength the resulting disparities together with their confidence values are presented to a spin system to yield global disparity interpretation. At each wavelength the left image is shifted according to the disparity map, obtained by the spin system. This modified image is used to compute a disparity map at a smaller value of the wavelength.

5. The spin system

At each level of the hierarchy the following spin system is used. The disparity spins s_{ij} at lattice sites (i,j) can take integer values ($s_{ij} \in \{0, \pm 1, \ldots\}$). The energy function of the disparity spin system is $E = E_i + E_f$ with

$$
\begin{aligned}
E_i &= J_i \sum_{<(i,j),(k,l)>} (s_{ij} - s_{kl})^2 \\
E_f &= J_f \sum_{(i,j)} c_{ij}(s_{ij} - f_{ij})^2.
\end{aligned}
\tag{2}
$$

The interaction energy E_i between the disparity spins induces an alignment between neighboring spins. The field energy E_f incorporates the disparities f_{ij} from the detectors and their confidences c_{ij}. E_f yields a low energy contribution for spins, well aligned with the field f_{ij}. For high values of the confidences c_{ij}, the field contribution becomes more important.

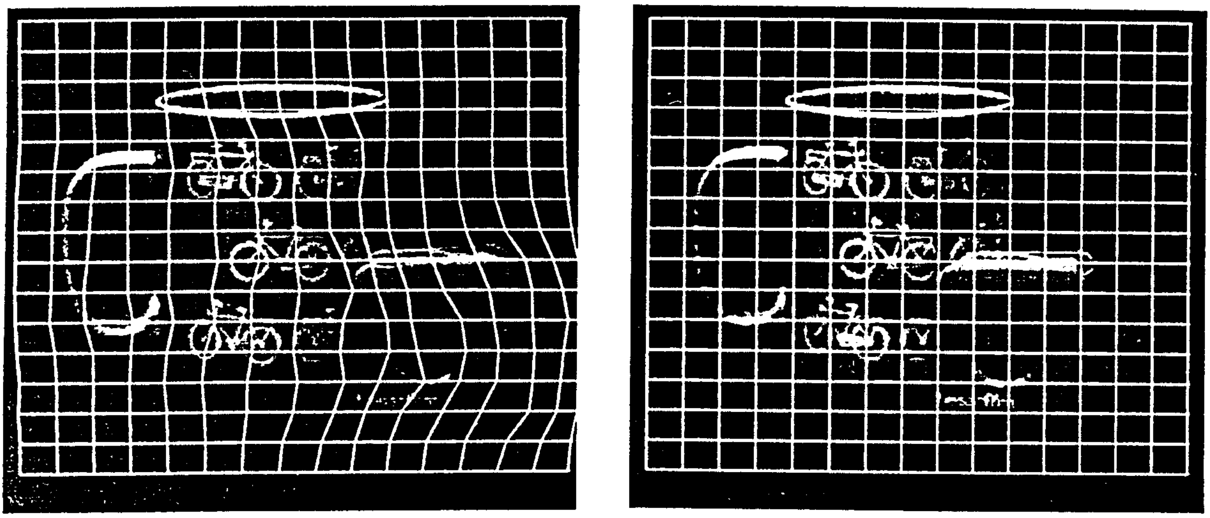

Fig. 1: Cup with tape holder - corresponding pixels in left and right image are indicated by the superimposed grid.

6. Results

The result for a 256 × 220 stereo image of a cup and a tape holder in front of it is presented in figure 1. The grid marks a subset of corresponding pixels in the left and in the right image as obtained from the algorithm. The energy weights used are $J_e = 0.25$ and $J_f = 1$ and the wavelengths have been lowered from 64 to 4 pixel in steps of a factor of 2. The disparity map coded as an intesity map is shown in Figure 2. The brighter the map the larger the disparity. Maximum disparity in the stereo image is about 23 pixels.

7. Conclusion

The hierarchical spin system proposed yields a disparity map for real stereo images. It can be extended to other features than the intensity of the images. With a combination of several features, e.g. intensities, edges and lines, as input to the disparity sensitive filters the algorithm can yield a better confidence of the disparity interpretation [9].

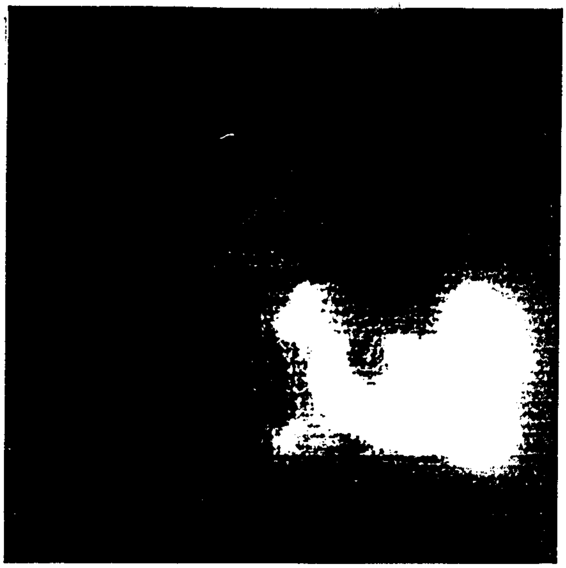

Fig. 2: Density coded disparity map for cup with tape holder.

REFERENCES

[1] Divko R, Schulten K (1986), *Stochastic Spin Models for Pattern Recognition*, in *Neural Networks for Computing*, Ed. Denker JS, 129-134, American Institute of Physics Publication

[2] Kirkpatrick J, Gelatt CD, Vecchi MP (1983), Science, 220:671

[3] Marr D, Poggio T (1979), *A computational theory of human stereo vision*, Proc R Soc Lond B 204:301-328

[4] Jenkin M, Jepson AD, (1988), *The measurement of binocular disparity*, In *Computational processes in human vision: An interdisciplinatory perspective*, Ed. Phylyshin Z, 69-98, Ablex Press, Norwood, NJ

[5] Sanger TD, (1988), *Stereo Disparity Computation Using Gabor Filters*, Biol Cybern 59:405-418

[6] Gabor D (1946) *Theory of communication*, J Inst Elec Eng 93 (III):429-457

[7] Poggio T, Torre V, Koch C (1985), *Computational Vision and Regularization Theory*, Nature 317:314-319

[8] Geman S, Geman S (1984), *Stochastic Relaxation, Gibbs Distribution, and the Bayesian Restoration of Images*, IEEE PAMI-6:721-741

[9] Divko R, Schulten K (1989), in preparation