

ランダムドットステレオグラムを用いた立体視機構解明の研究

平井 有三

筑波大学
電子・情報工学系 教授

1 まえがき

人の立体視を成立させる手がかりには、物体の形、陰影、テクスチャ勾配など単眼視的なものもあるが、鮮明な立体感を与える手がかりは両眼視差である。Julesz が考案したランダムドットステレオグラムは、単眼視による手がかりが全くなく、両眼視差のみの手がかりから立体視を知覚させる視覚刺激として多くの研究者に利用されてきた [1]。左右の画像から両眼視差を検出するためには、両眼像間の対応問題を解く必要がある。これまでランダムドットステレオグラムをベンチマークとして、対応問題を解くための多くの両眼視差検出モデルが提案されてきた [2, 3, 4, 5, 6]。

対応問題を解くためには、3次元世界の見え方に関する制約を利用することが必要である。これまで提案されてきた制約には、

1. 一対一の制約：左眼の一つの刺激は、右眼の高々一つの刺激とのみ対応する。
2. 滑らかさの制約：3次元物体の表面の奥行きは滑らかに変化する。
3. 順序保存の制約：対応の結果は左眼と右眼の刺激の順序を変えない。

などがある。ほとんどのモデル [2, 4, 5, 6] は、一対一の制約と滑らかさの制約を用いている。しかしながら、両眼視差から復元できる表面の3次元的な曲率は、滑らかさの制約そのものが制約している曲率に大きく影響される。また、モデルの性能は、滑らかさの制約を実現している局所的な興奮性結合領域の広さに依存する。

本研究の目的は、以前提案した順序保存の制約 [3] を、両眼融合のデータ [7] から導き出された3次元的な forbidden cone に拡張し、その有効性を、

1. 静的ランダムドットステレオグラム
2. 動的ランダムドットステレオグラム

を用いて検証することにある。Julesz [1] によれば、3次元構造を保ったままフレーム毎にドットの配置が変化する動的ランダムドットステレオグラムは、静的ランダムドットステレオグラムより知覚しやすいことが報告されている。これまで、動的ランダムドットステレオグラムをベンチマークとして用いた研究はなかった。

2 モデルの構造と動作方程式

モデルの構造を図1に示した。図(a)は、左右のランダムドットステレオグラムのk番目の行間、すなわちエピポーラ線上の点 l_i^k と r_j^k の対応をとる、k番目の disparity field を示している。素子 $b_{x,z}^y$ は

$$x = \frac{i+j}{2}, \quad y = k, \quad z = \frac{i-j}{2}$$

の関係を満たす入力間の対応をとる照合素子を表す。素子 $b_{x,z}^y$ と $b_{\xi,\zeta}^{\eta}$ 間の視差 d と距離 D を

$$d = 2|\zeta - z|, \quad D = |\xi - x|$$

で定義する。

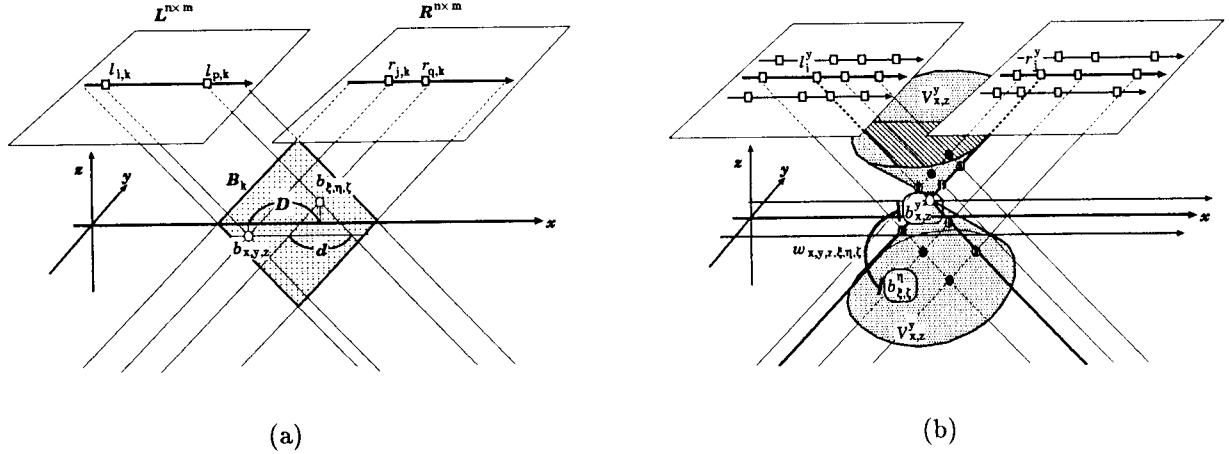


図 1: (a) 入力画像 $L^{n \times m}(t), R^{n \times m}(t)$ と、disparity field (B^k) の定義。2つの照合素子 $b_{x,z}^y, b_{\xi,\zeta}^\eta$ 間の視差 d と距離 D を図のように定義した。(b) 素子 $b_{x,z}^y$ がもつ forbidden cone による抑制領域の定義。上下の cone 内のすべての素子との相互抑制結合により、制約を表現している。黒丸で示された素子は順序保存の制約を、灰色の素子は一対一の制約を破っている素子である。

図 (b) は、本研究で用いた forbidden cone の定義を示している。素子 $b_{x,z}^y$ がもっている抑制領域 $V_{x,z}^y$ は

$$(\zeta - z)^2 \geq (\xi - x)^2 + \left(\frac{\eta - y}{2}\right)^2$$

の関係を満たす上下の二つの cone 状の領域で定義される。両眼融合のデータ [7] から導かれた forbidden cone は等方的であるが、本研究では y 方向に広がった異方性の cone を用いている。この理由は、 x 方向の素子の密度に比較して y 方向の素子の密度が $\frac{1}{2}$ となることを考慮したからである。実験の結果、ドット密度が高い場合には異方性の cone の方が精度がよく、また他の心理データによっても異方的な処理の存在が示唆されている [8]。

disparity field 内の各素子は、以下の非線形一次微分方程式にしたがって動作する。

$$\tau \frac{db_{x,z}^y(t)}{dt} = -b_{x,z}^y(t) + e_{x,z}^y(t) - \sum_{b_{\xi,\zeta}^\eta \in V_{x,z}^y} w_{x,y,z,\xi,\eta,\zeta} \cdot \varphi(b_{\xi,\zeta}^\eta(t)) \quad (1)$$

$$\varphi(u) = \begin{cases} u & \text{if } u > 0 \\ 0 & \text{otherwise} \end{cases}$$

τ は素子の時定数、 $e_{x,z}^y(t)$ は両眼からの入力で、

$$e_{x,z}^y(t) = \min[l_{x-z}^y(t), r_{x+z}^y(t)]$$

で定義される。各画素は、ドットがある場合 1、ない場合 0 をとるものとする。結合係数 $w_{x,y,z,\xi,\eta,\zeta}$ は、素子 $b_{x,z}^y$ への $V_{x,z}^y$ 内の素子 $b_{\xi,\zeta}^\eta$ からの抑制性結合であり、

$$w_{x,y,z,\xi,\eta,\zeta} = \begin{cases} W \cdot \left(\frac{2}{3}\right)^{|\eta-y|} & \text{if } b_{\xi,\zeta}^\eta \in V_{x,z}^y \\ 0 & \text{otherwise} \end{cases}$$

と定義した。すなわち、行が離れるにしたがって両眼視差の相関が低くなることを表現している。以下の実験では、 $W = 0.6$ とした。また、対応を計算した範囲は視差が ± 36 まで、すなわち $|z| \leq 18$ までである。また、微分方程式はオイラー法で解いている。

表 1: 提案したモデルの性能と、Marr-Poggio のアルゴリズム [2] の比較。

	Surfaces portrayed in random-dot stereograms				
	square	semisphere	semisphere with hollows	crossed semicylinders	saddle
present network	92.6%	99.5%	96.3%	95.6%	97.0%
<i>unmatches</i>	7.0%	0.5%	2.7%	4.2%	2.9%
<i>incorrect matches</i>	0.4%	0%	1.0%	0.2%	0.1%
Marr-Poggio	98.8%	86.8%	88.1%	86.0%	92.0%
<i>unmatches</i>	0.1%	11.2%	8.9%	12.0%	6.3%
<i>incorrect matches</i>	1.1%	2.0%	3.0%	2.0%	1.7%

3 実験結果

3.1 静的ランダムドットステレオグラム

図 2 の左側に示した 2 枚の 2 次元ランダムドットステレオグラムを入力画像として、式 (1) を用いて対応問題を解いた結果を図の右側に示した。画像の大きさは 256×256 であり、ドット密度は 20% である。これらのステレオグラムの奥行きのエッジ部分の最大視差は 12 ピクセルである。図の結果は、平衡状態に達した照合素子の出力値が 1 に近いものを用いて、検出された形状を表示したものである。

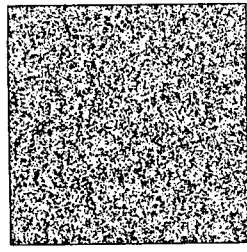
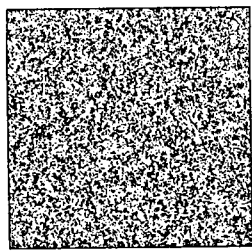
各ステレオグラムに対する結果を、Marr-Poggio のアルゴリズム [2] と比較して表 1 に示した。Marr-Poggio のアルゴリズムは、いずれも最大の正解率が得られるようにパラメータを調節して得られた結果であり、ドット密度が変わればパラメータも調節しなおす必要がある。表から明らかなように、四角形を除いて提案したモデルの方が高い正解率を示している。四角形の場合でも誤対応率は Marr-Poggio のアルゴリズムより低い値となっている。これらの結果から、提案モデルの性能は、ドット密度など画像の性質に依存せず、同じ回路パラメータで広範囲な対応問題を解くことができることが分かった。

3.2 動的ランダムドットステレオグラム

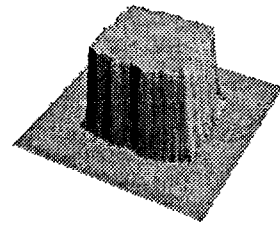
Julesz によれば、3 次元構造を保ったままフレーム毎にドットの配置が変化する動的ランダムドットステレオグラムは、静的ランダムドットステレオグラムより知覚しやすいことが報告されている [1]。

動的ランダムドットステレオグラムを構成するために、知覚される 3 次元形状は同一で、ドットの配置が異なる 50 組のステレオグラムを用意し、数値計算の 1 ステップ毎にランダムに一組を入力し、解を求めた。図 3(a) に一組のランダムドットステレオグラムを示した。画像の大きさは 128×128 であり、ドット密度は 20% である。隠されている形状は図 2 と同じくサドルであり、左右の奥行きのエッジ部の最大視差は 12 ピクセルである。

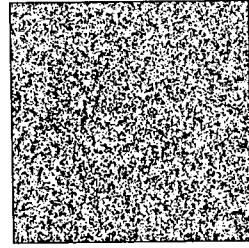
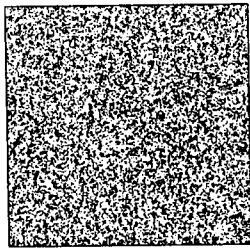
入力画像のドット配置が時間的に変化するため、静的ランダムドットステレオグラムの場合のように、正しい対応を表す照合素子に常に 1 の入力を与えられ続けることはない。したがって、これらの素子は一度発火しても、入力がなくなると次第に出力が減少する。この減少速度は、動作方程式 (1) の時定数 τ に依存する。時定数の効果を見るため、図 (b) に示したように 5 つの時定数による正解率 (正対応の割合) の時間変化を測定した。横軸は計算の繰り返し回



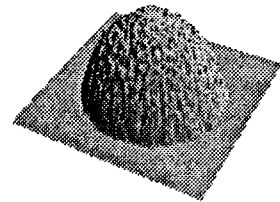
(a)



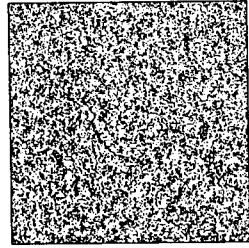
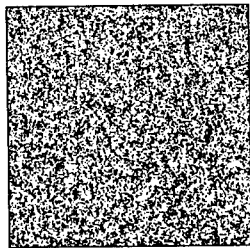
(b)



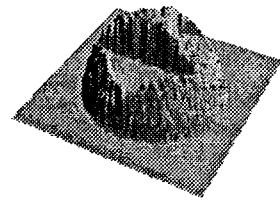
(c)



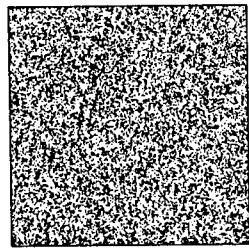
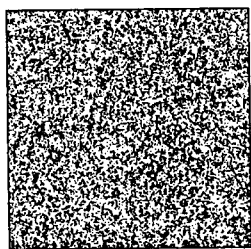
(d)



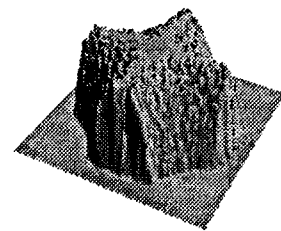
(e)



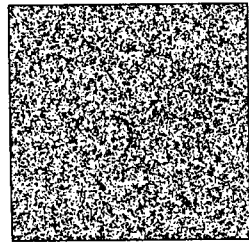
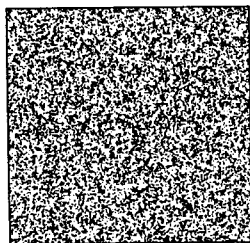
(f)



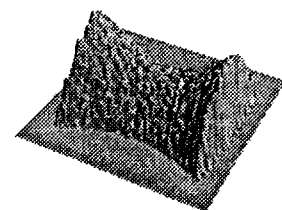
(g)



(h)

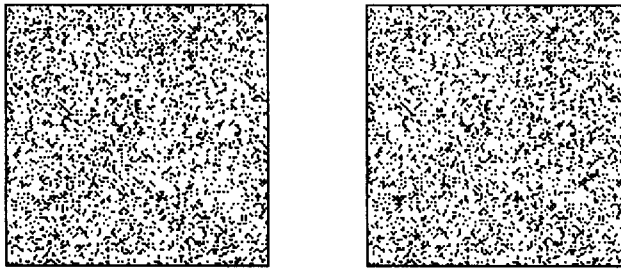


(i)

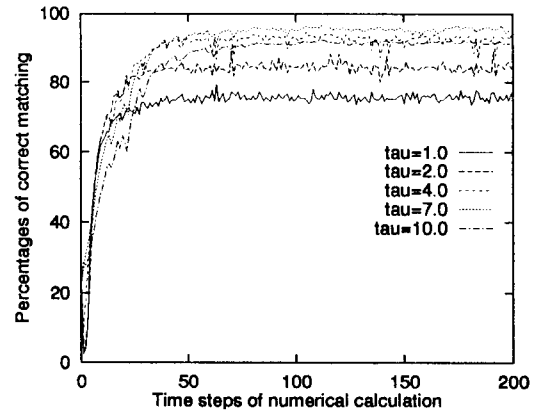


(j)

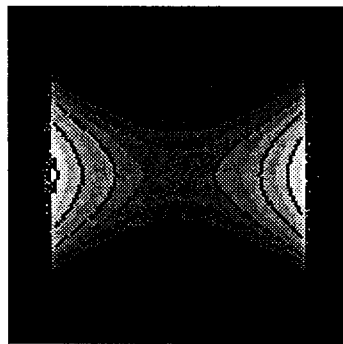
図 2: ランダムドットステレオグラムの例。平行法で観察すると、右に示したような曲面が観測できる。これらはモデルで再構成した曲面である。



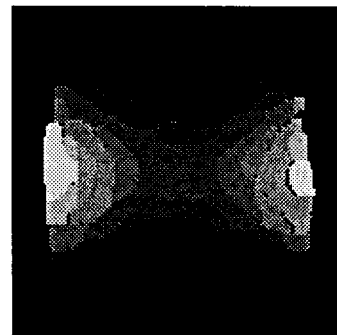
(a) 動的ランダムドットステレオグラムの一組



(b) 5つの時定数に対するモデルの動作。横軸は時間、縦軸は正しい対応が検出できた割合。



(c) 提案したモデルによる検出結果。手前ほど明るく表示している。



(d) Marr-Poggio のアルゴリズムによる結果。

図 3: 動的ランダムドットステレオグラムによる実験結果

数で、縦軸が正解率 (correct) である。平衡状態へは、約 50 回の繰り返しまでで収束していることが分かる。静的ランダムドットステレオグラムの場合は、500 回以上の繰り返し回数が必要であった。モデルでも、人の知覚と同様に、動的ランダムドットステレオグラムの方が容易であることが分かる。この理由は、動的ランダムドットステレオグラムでは入力画像が変化するため、付随するエネルギー関数が時間的に平均化され、局所解の谷が平坦化するからであると考えられる。

図から、時定数が 4 以上で性能が飽和しており、時定数を 10 にしたときの正解率は 96% 以上であった。図 (c) に検出した視差を示した。手前ほど明るく表示している。比較のために、図 (d) に Marr-Poggio のアルゴリズム [2] を用いた結果を示した。正解率は 43%、誤対応は約 57% であった。この結果はパラメータを最適に調整して得た結果である。

4 まとめ

異方性の forbidden cone を対応点検出の制約として用いた、両眼視差検出機構のモデルを提案し、従来のモデルより優れていることを示した。滑らかさの制約を用いたモデルの性能はパラメータに強く依存するのに対して、提案モデルは依存しない。また、人の知覚現象と同様

に、動的ランダムドットステレオグラムの方が静的ランダムドットステレオグラムより検出が容易であることを示した。

本研究の詳細は、文献 [9, 10, 11] で報告している。文献 [10] では、モデルの各素子が、生理実験により猿の脳視覚領で発見された 4 種類の両眼視差検出細胞 [12] と類似した反応特性を示すことが示されている。また、文献 [11] では、透明な面や曖昧な面の検出について議論している。

今後、奥行きのエッジ検出にとって重要な単眼視領域の取り扱い、自然画像への適用などについて研究を進展させる予定である。

本研究の遂行にあたり多大な援助を賜った、(財)高柳記念電子科学技術振興財団ならびに関係者各位に深く感謝致します。

参考文献

- [1] B.Julesz, "Foundations of Cyclopean Perception," University of Chicago Press, Chicago, 1971.
- [2] D.Marr, and T.Poggio, "A cooperative computation of stereo disparity," *Science*, **194**, 283-287, 1976.
- [3] Y.Hirai, and K.Fukushima, "An inference upon the neural network finding binocular correspondence," *Biological Cybernetics*, **31**, 209-217, 1978.
- [4] K.Prazdny, "Detection of binocular disparities," *Biological Cybernetics*, **52**, 93-99, 1985.
- [5] S.B.Pollard, J.E.W.Mayhew and J.P.Frisby, "PMF: A stereo correspondence algorithm using a disparity gradient limit," *Perception*, **14**, pp.449-470, 1985.
- [6] J.P.Frisby, and S.B.Pollard, "Computational issues in solving the stereo correspondence problem," In: Landy MS and Movshon JA (eds) *Computational models of visual processing*. Cambridge, MA: MIT Press, 1991.
- [7] P.Burt, and B.Julesz, "A disparity gradient limit for binocular fusion," *Science*, **208**, 615-617, 1980.
- [8] H.Bülthoff, and M.Fahle, "Disparity gradients and depth scaling," *Massachusetts Institute of Technology*, A.I.Memo No. 1175, C.B.I.P Memo No.39, September 1989.
- [9] 塚原朋哉, 平井有三, "ランダムドットステレオグラムからの曲面再構成," 信学論 (J76-D-II), No.8, pp.1676-1683, 1993.
- [10] T.Tsukahara and Y.Hirai, "A stereo matching network and its biological implications," *Proceedings of ICONIP'96 Hong Kong*, Vol.1, 356-361, 1996.
- [11] 塚原朋哉, 平井有三, "透明な面と曖昧な面からの両眼視差検出モデル," 電子情報通信学会、ニューロコンピューティング研究会、1996年2月
- [12] G.F.Poggio, F.Gonzalez, and F.Krause, "Stereoscopic mechanisms in monkey visual cortex: Binocular correlation and disparity selectivity," *The Journal of Neuroscience*, **8**, 4531-4550, 1988.