MVA '96 IAPR Workshop on Machine Vision Applications, November. 12-14, 1996, Tokyo, Japan

A New Way to Visual Representation and Learning

Zhiyong YANG, Songde MA and Qifa KE
National Laboratory of Pattern Recognition
Institute of Automation, Chinese Academy of Sciences
P. O. Box 2728, Beijing 100080, P. R. China
Email: yangzy@prlsun6.ia.ac.cn masd@prlsun2.ia.ac.cn

Abstract

We explain some limitations of regularization theory of early vision and formulate visual representation and learning as statistical mechanics of surfaces with defects. In this new paradigm, pinning energy consists of a set of local oriented regularization fields. To reconstruct a D dimensional surface, known data, generally a set of patches of $0, 1, \ldots, D-1$ dimensions, are taken as defects to "pin" the surface. Each realization of the surface pinned by the known data contributes a Boltzman weight and ensemble average over all realizations gives the reconstructed surface. In 2D, we present a neural network dynamics to approximate the ensemble average. The dynamics displays recovering $1D$ shapes from local pinning bars, collinear grouping, edge and region filling-in, illusory figure perception and perception of apparent brightness as a kind of dynamic phase transitions.

1 Regularization Theory and Its Limitations

Visual representation and learning can be seen as hypersurface reconstruction[1]. Regularization theory of early vision and the induced architecture, radial basis function (RBF) networks have become an important paradigm. Due to its mathematical simplicity and biological supports, RBF networks have been suggested as building blocks of the brain and used for object recognition and motion control[2]. However, regularization theory and the induced RBF architecture suffer from the following shortcomings

1 *Quadratic functional are at best approximations to many cases,*

2 *The inputs are supposed to be un-correlated, sparse points. However, in early vision, the input data are usually correlated, for example, a segment of contrast edge, a surface patch of constant curvature or intensity. The usual functionals are not good approximations in these cases because they don't use the correlative information.*

3 *Linear superposition of RBFs contrasts vividly with massively connections in the brain cortex.*

Representing in a correlated way is the primary way used by the brain. For example, cells in the primary visual cortex respond most to simple bars of certain orientations and cells at higher hierarchy of visual cortex can represent more complex shapes through integrating inputs from lower hierarchy[3].

2 Representation and Learning as Statistical Mechanics of Surfaces with Defects

Visual representation and learning can be generally seen as statistical mechanics of surfaces with defects(SMSD)[4]. Take 2D surface reconstruction as an example. Continuous surfaces ($D = 2$) can be seen as manifolds embedded in $d = D + 1 = 3$ dimensional space. The known data can be seen as pinning centers used to 'pin' the wandering surfaces in $d = D + 1 = 3$ space. The effective energy functional is

$$H = \sum E_{pin} + \frac{K_1}{2} \times \int (\nabla f)^2 d\mu + \frac{K_2}{2} \times \int \left(\nabla^2 f\right)^2 d\mu \quad (1)$$

In (1), f is the surface. K_1, K_2 are regularization constants. ∇ is differential operator. $d\mu$ is integral measure. E_{pin} is pinning energy or kernel energy. The last two terms are from stretching and bending modes. In usual regularization theory, the pinning centers are

sparse, un-correlated points and $E_{pin} = (f_i - f)^2$. The partition function associated with (1) is

$$Z = \sum exp(-\beta H(f)) \tag{2}$$

Summation is over all possible surface configurations. β is a constant representing noise level. The reconstructed surface is then ensemble average

$$\langle f \rangle = \frac{1}{Z} \times \sum f \times exp(-\beta H(f)) \tag{3}$$

Mean field approximation recovers the solution in standard regularization theory

$$\frac{\delta H}{\delta f} = \frac{\delta}{\delta f}\left(\sum E_{pin} + \frac{K_1}{2}\int (\nabla f)^2 d\mu + \frac{K_2}{2}\int (\nabla^2 f)^2 d\mu \right) = 0 \tag{4}$$

Intuitively, the best pinning to pin a D dimensional surface is a set of D dimensional patches. Here, we suggest the pinning energy as

$$\begin{aligned} E_{pin} = {} & \int (f_i - f)^2 d\mu + \frac{\lambda_{11}(r)}{2}\int ((u_1 \bullet \nabla) f)^2 d\mu + \\ & \frac{\lambda_{12}(r)}{2}\int ((u_1 \bullet \nabla)^2 f)^2 d\mu + \frac{\lambda_{21}(r)}{2}\int ((u_2 \bullet \nabla) f)^2 d\mu \\ & + \frac{\lambda_{22}(r)}{2}\int ((u_2 \bullet \nabla)^2 f)^2 d\mu \end{aligned} \tag{5}$$

u_1 is the normal of the pinning patch. $u_2 \perp u_1$, and has two orthogonal unit vectors. r is the distance from the pinning center in the local coordinate (u_1, u_2). $\lambda(r) \geq 0$, $r \to \infty, \lambda(r) \to 0$ is local oriented regularization field (LORF). We choose the condition $\lambda_{11}(r) > \lambda_{21}(r)$, $\lambda_{12}(r) > \lambda_{22}(r)$ to penalize deviation from the tangential plane of the patch.

3 Local Orientedly Regularized Neural Networks (LORNN)

We do not have general solutions to Eq.(1)-(5). At present, we suggest to use some powerful algorithms in statistical physics to solve Eqs.(3), (4). Here, we design a neural network dynamics involving LORFs for shape representation and visual learning. The dynamics can be seen as approximation to the new paradigm at very low noise level. For illustration, we present the dynamics in 2D for 1D shape represen-tation, perceptual grouping, illusory figure perception and perception of apparent brightness. In these cases, the effective pinnings we are interested are simple bars.

Consider a 2D array of neurons, the system is described by the following equations

$$h = -\frac{MI}{2} + KI + g_1(\nabla I)^2 + g_2(\nabla^2 I)^2 + \sum R \tag{6}$$

$$\begin{aligned} R = {} & \lambda_{11}(r)((n_1 \bullet \nabla) I)^2 + \lambda_{21}(r)((n_2 \bullet \nabla) I)^2 \\ & + \lambda_{12}(r)((n_1 \bullet \nabla)^2 I)^2 + \lambda_{22}(r)((n_2 \bullet \nabla)^2 I)^2 \end{aligned} \tag{7}$$

$$M = \sum J \times I \tag{8}$$

$$I = F(M) \tag{9}$$

In (6), h is energy per site. $I \in [0, 1]$ is state variable. R is regularization term at each pinning bar and the summation is over all pinning bars. Each R contains four terms as in (7). K is applied field. ∇ is a differential operator. n_1 is the tangential direction of pinning bars. n_2 is the normal direction of pinning bars. Summation in (8) is over specific receptive fields. $F(\bullet)$ is characteristic function. In this paper, for simplicity, we select receptive field and characteristic function as

$$F(M) = \begin{cases} 1, & M \geq 1 \\ -1, & M \leq -1 \\ M, & -1 < M < 1 \end{cases} \tag{10}$$

$$J_{ij} = \begin{cases} J, & 0 < |i - j| < q \\ 1, & i = j \\ 0, & otherwise \end{cases} \tag{11}$$

q is the size of receptive fields. We adopt the following simple learning dynamics

$$J(t+1) = J(t) + l_c \times \left(I(t) - I(t-1) \right) \qquad (12)$$

By (12), all connection weights in each neuron's receptive field update in the same way.

Pinning bars induce an ensemble of oriented Gaussian(OG) or oriented exponential(OE) distribution. An OG is

$$G = exp\left(-1/2(x/\sigma_1)^2\right) exp(-1/2(y/\sigma_2)^2) \qquad (13)$$

x is along the tangential direction of pinning bars, y lies in the normal direction of pinning bars. $Z = \sigma_1/\sigma_2 > 1$ is orientation coefficient. At every site, the input is nonlinear superposition of an ensemble of OGs. We only consider nonlinear superposition of two OGs and suggest

$$I(I_1, I_2) = \begin{cases} I_1 + I_2, & I_1 < \theta_1 \text{ or } I_2 < \theta_1 \\ I_1 + I_2 + P(I_1 + I_2), & \theta_1 \le I_1 \le \theta_2,\ \theta_1 \le I_2 \le \theta_2 \\ I_1 + I_2 + N I_1 I_2, & otherwise \end{cases} \qquad (14)$$

I_1, I_2 are from two OGs at a site. P, N are two constants. θ_1, θ_2 are two thresholds. There is constraint between P, N, $2\theta_2 P = N\theta_2\theta_2$. Note that $I(I_1, I_2)$ is not Continuous across these regions. LORF can be of any form satisfying $\lambda(r) \ge 0;\ r \to \infty,\ \lambda(r) \to 0$. Here, we choose OG

$$\lambda_{ij}(r) = \lambda_{ij}(0) exp\left(-\frac{r_{II}^2}{2\sigma_{II}^2} \right) exp\left(-\frac{r_{\perp}^2}{2\sigma_{\perp}^2} \right) \qquad (15)$$

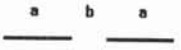
$\lambda_{ij}(0)\ (i, j = 1, 2)$ is amplitude of LORF. σ_{II}, $\sigma_{\perp}$ are the sizes of OG in the tangential and normal directions of pinning bars respectively. r_{II}, $r_{\perp}$ are the distances from the pinning center along the tangential and normal direction of pinning bars respectively.

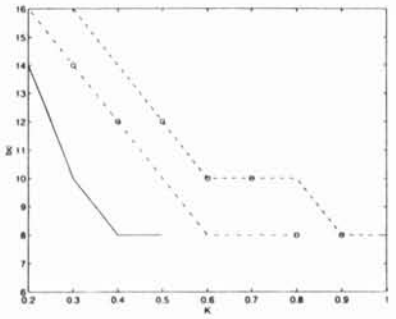
4 Collinear Grouping, Filling-in and Perceptual of Apparent Brightness as Dynamics Phase Transitions

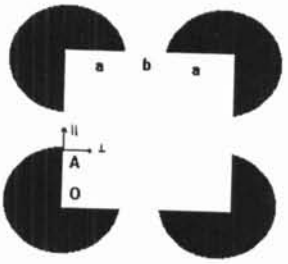
LORNN displays recovering $1D$ shapes from local pinning bars, collinear grouping, edge and region filling-in, illusory figure perception and perception of apparent brightness as a kind of dynamic phase transitions. When some parameters pass some critical values, $1D$ shapes are recovered, collinear grouping, edge and region filling-in, illusory figures and perception of apparent brightness emerge(See Table 1). There are finite energy jumps between up and below the critical values, so these phase transitions are of first order. These behaviors are definitely different from usual diffusion or reaction-diffusion, ART models or RBF models. Gs: Gaussians. Es: exponentials. CEs: Contrast Edges. b_c is critical length(distance, gap, or radius). J_c is critical initial connection weight. K_c is critical applied field. Fig.1 present some examples. In all these examples, we set $P = N = 0$.

Table 1

	Inputs	Critical parameters	phase transitions
Edge filling-in	OGs or OEs at end points	b_c, J_c, K_c	$b \le b_c, J \ge J_c, K \le K$
Collinear grouping	OGs or OEs at segment ends	b_c, J_c, K_c	$b \le b_c, J \ge J_c, K \le K$
$1D$ shape representation	OGs or OEs at pinning bar ends	b_c, J_c, K_c	$b \le b_c, J \ge J_c, K \le K$
Illusory figures	Gs or Es along CEs, OGs or OEs at CE ends	b_c, J_c, K_c	$b \le b_c, J \ge J_c, K \le K$
Apparent brightness	Gs or Es along CEs, OGs or OEs at CE ends	b_c, J_c, K_c	$b \le b_c, J \ge J_c, K \le K$







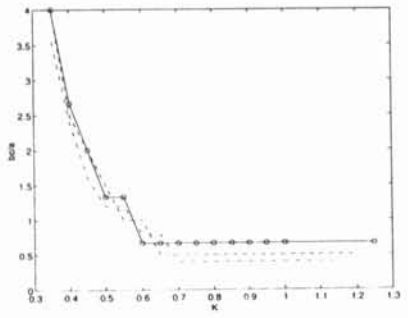


Fig.1 LORNN Examples. Top, Collinear grouping and edge filling-in. b_c--K plot. Input is two OGs at the near ends. When the gap b passes below some critical value b_c, collinear grouping and edge filling-in occur. $Z=4,\ \sigma_1=3,\ q=1,\ l_c=0.05,$ $g_1=0.02,\ g_2=0.05.\sigma_{II}=1.5\sigma_1,\ \sigma_{II}/\sigma_\perp=4$ $\lambda_{11}(0)=1.5g_1,\ \lambda_{12}(0)=1.5g_2,\ \lambda_{21}(0)=0.5g_1$ $\lambda_{22}(0)=0.5g_2$ $J_0=0.1$, '_'; $J_0=0.2$, ':'; $J_0=0.3$, '-.'; $J_0=0.4$, 'o'; $J_0=0.5$, '--'.

Bottom, Kanizsa map. b_c--K plot. When the gap b passes below some critical value b_c, a Kanizsa map emerges, a rectangle whiter than the background appears to occlude the four discs. Along OA is located a set of Gs of size σ_1. Located at A is an OE of size $\sigma_{II}=1.5\sigma_1, \sigma_{II}/\sigma_\perp=4$. Input is the superposition of all these Es and OEs.

$Z=3,\ q=1,\ l_c=0.05,\ g_1=0.02, g_2=0.05,$ $\lambda_{11}(0)=1.5g_1, \lambda_{12}(0)=1.5g_2, \lambda_{21}(0)=0.5g_1$ $\lambda_{22}(0)=0.5g_2, J_0=0.7$. $a=2, \sigma_1=1$, '_'; $a=3, \sigma_1=1.5$, 'o'; $a=4, \sigma_1=2$, '--'; $a=5, \sigma_1=3$; '-.'.

Refenences

[1] T Poggio & F Girosi, Regularization algorithms for learning that are equivalent to multilayer networks, *Sci* **24**, 978-982(1990).

[2] T Poggio, A theory of how the brain might work, *Sym on Qua. Biol* 899-910. (Cold Spring Harbor Lab Press 1990)

[3]S W Kuffler, J G Nicholls & A R Martin, From Neuron to Brain, (2nd Sinauer Asso Inc 1984).

[4] D R Nelson, T Piran and S Weinberg eds, Statistical Mechanics of Membranes and Surfaces, (*World Scientific* 1989).