

FoE 模型的训练方法研究

王国权, 刘 亮

(黑龙江科技学院 计算机与信息工程学院, 黑龙江 哈尔滨 150027)

摘 要: FoE 模型是 Roth 等于 2005 年提出的一种新的马尔可夫随机场模型, 其以优异的处理效果引起学界的瞩目。介绍了 FoE 模型及其训练的过程, 给出了应用 student-t 专家和 charbonnier 专家分别在几种不同训练设置下训练得到的 FoE 模型进行图像去噪的试验结果。并针对 FoE 模型训练时间较长的问题, 提出了一种改进的批训练方法应用于 FoE 模型的训练。这种方法通过逐步增大训练过程中所应用的批数量的方式, 将 FoE 模型的训练过程分为 4 个层次。图像去噪仿真试验的结果表明, 该方法可以在有效减少约 50% FoE 模型训练时间的同时, 取得与原训练方法相似的训练结果。

关键词: 马尔可夫随机场; FoE 模型; 图像去噪

中图分类号: TN911.73

文献标识码: A

文章编号: 1673-629X(2010)12-0086-04

A Study on Training Methods of FoE Model

WANG Guo-quan, LIU Liang

(Computer and Information Engineering College, Heilongjiang Institute of Science and Technology, Harbin 150027, China)

Abstract: FoE model is a new MRF model demonstrated by Roth et al in 2005. This model has attracted a lot of attention by its excellent performance. In this paper, described FoE model and its training progress, present image denoise experiment results with student-t expert and charbonnier expert under several training settings. To solve the time consuming training problem of FoE model, an improved batch training method is proposed. Through increasing batch number step by step, this method split the training progress of FoE model into 4 stages. Practical example of image denoise shows that the proposed method can effectively reduce approximate 50% training time of FoE model at the same acquire similar training result.

Key words: Markov random field; FoE model; image denoise

0 引 言

马尔可夫随机场模型应用于低层图像至今已有 30 多年的历史, 目前已经有应用于各种领域的众多模型。但是由于马尔可夫随机场模型一旦引入高阶邻域系统, 其在应用于解决低层视觉问题时, 对计算力的要求会大幅增加, 因而当前对应用于低层视觉领域的高阶马尔可夫随机场模型的研究还处于起步阶段, 其中典型的模型有 FRAME 模型^[1], VISTA 模型^[2]。

FoE (Fields of Experts) 模型是一种新型的高阶马尔可夫随机场模型, 自 Roth 等于 2005 年提出以来, 其优异的处理效果已经引起了学界的高度瞩目。目前 FoE 模型已经应用于一系列低层视觉处理, 如在灰度图像处理去噪^[3], 彩色图像去噪^[4], 光流 (Optical

Flow)^[5], 超分辨率^[6], 图像修补^[4], 将 FoE 模型的训练结果作为 NVS 的先验概率^[7]。以上应用证明 FoE 模型是一种多用途的马尔可夫随机场模型, 有着广泛的应用和研究前景。

1 马尔可夫随机场基础

设 S 为结点集, D 为 S 上的设置 (在底层视觉中指像素的灰度值), v 是 S 上的结点, 并设 $N = \{N_v, v \in S\}$ 为定义在 S 上的邻域系统, 其性质有: (1) $v \in N_v$; (2) 若 $v \in N_u$ 则 $u \in N_v$ 。邻域 N_v 中的结点与结点 v 的关系是相邻的。若 S 的一个子集 c 中结点间的关系是相邻的, 则该子集 c 是一个集团, 并设 C 为所有集团的集合。设 x 为随机场, $x(v)$ 为定义在 v 上的随机变量, 则一个定义在邻域系统 N 上的分布 $p(x)$ 满足:

$$p(x(v) | x(-v)) = p(x(v) | x(N_v)) \quad (1)$$

其中 $x(-v)$ 表示除 v 外所有取值, 且对于 $A \in S$, $x(A)$ 表示 A 中所有结点的取值, 其含义为 $x(v)$ 的取值仅与其相邻的结点有关, 则有 $p(x)$ 是该邻域

收稿日期: 2010-04-21; 修回日期: 2010-07-19

基金项目: 黑龙江省教育科研项目资助 (11511355)

作者简介: 王国权 (1968-), 男, 吉林榆树人, 博士, 教授, 硕士研究生导师, 研究方向为图形图像及多媒体技术、计算机辅助设计。

系统 N 上的一个 MRF 分布。

其形式可以记为:

$$p(x) = \frac{1}{Z} \prod_{c \in C} f_c(x_{(c)}) \quad (2)$$

其中 f_c 是势函数, c 是集团, Z 是归一化常数。

根据 Hamersley - Clifford 定理, 式(2) 可写为一个 Gibbs 分布的形式:

$$p(x) = \frac{1}{Z} \exp\{-E(x)\} \\ = \exp\{-\sum_{c \in C} U_c(x_{(c)})\} \quad (3)$$

其中 U_c 是势函数, Z 是归一化常数, 其定义为:

$$Z = \int_x \prod_{c \in C} f_c(x_{(c)}) dx \\ = \int_x \exp\{-\sum_{c \in C} f_c(x_{(c)})\} dx \quad (4)$$

2 FoE 模型

(1) 基本理论。FoE 模型是 Roth 等于 2005 年提出的一种新的用于低层图像处理的高阶马尔可夫随机场模型, 它采用专家乘积^[8]的方式来表示势函数, 在实际应用中取得了很好的效果。FoE 模型的邻域系统被定义为可以使 $m \times m$ 区域中的所有结点都是相邻结点, 并且一幅图像中每个可以使 $m \times m$ 区域成立的结点都用于定义一个集团。通过假设马尔可夫随机场模型是齐次的, 即所有集团上的势函数都是相同的, 其采用专家乘积方式的表示形式为^[6]:

$$f(x_{(k)}) = f_{\text{FoE}}(x_{(k)}; \Theta) = \prod_{i=1}^N \phi(J_i^T x_{(k)}; \alpha_i) \quad (5)$$

其中 $\phi(\cdot; \cdot)$ 是专家, J_i 是线性滤波器, α_i 是专家的参数, 若设 $\Theta = \{J_i, \alpha_i \mid i = 1, \dots, N\}$ 是所有模型参数的集合, 则 FoE 模型可以记为^[6]:

$$p_{\text{FoE}}(x; \Theta) = \frac{1}{Z(\Theta)} \prod_{k=1}^K f_{\text{FoE}}(x_{(k)}; \Theta) = \\ \frac{1}{Z(\Theta)} \prod_{k=1}^K \prod_{i=1}^N \phi(J_i^T x_{(k)}; \alpha_i) = \frac{1}{Z(\Theta)} \prod_{k=1}^K \prod_{i=1}^N \phi(J^{(i)} * x; \alpha_i) \quad (6)$$

其中 $*$ 表示卷积, 式(6) 的 Gibbs 等价形式为^[6]:

$$p_{\text{FoE}}(x; \Theta) = \frac{1}{Z(\Theta)} \exp\{-E_{\text{FoE}}(x; \Theta)\} = \\ \frac{1}{Z(\Theta)} \exp\left\{-\sum_{k=1}^K \sum_{i=1}^N \Psi(J_i^T x_{(k)}; \alpha_i)\right\} = \\ \frac{1}{Z(\Theta)} \exp\left\{-\sum_{k=1}^K \sum_{i=1}^N \Psi_k(J^{(i)} * x; \alpha_i)\right\} \quad (7)$$

其中 $\Psi((\cdot; \cdot)) = \log \phi((\cdot; \cdot))$ 是专家函数的对数形式 $Z(\cdot)$, 是归一化常数, 由式(7) 可知:

$$Z(\Theta) = \int_x \exp\left\{-\sum_{k=1}^K \sum_{i=1}^N \Psi(J_i^T x_{(k)}; \alpha_i)\right\} dx \quad (8)$$

目前在 FoE 模型中可以使用两种专家, 一种是

Student - t 专家, 其形式为^[6]:

$$\phi(y; \alpha) = (1 + \frac{1}{2} y^2)^{-\alpha} \quad (9)$$

其中 $y = J_i^T x_{(k)}$ 。

另一种是 charbonnier 专家, 其形式为^[6]:

$$\phi(y; \alpha) = \exp\{-\alpha \sqrt{1 + y^2}\} \quad (10)$$

(2) FoE 模型的训练。当应用极大似然 (Maximum Likelihood, ML) 学习方式在给定独立同分布的样本 $X = \{x_n \mid n = 1, \dots, N\}$ 时, 可以通过梯度上升法 (gradient ascent) 学习参数 W ^[10]:

$$W^{(i+1)} = W^{(i)} + \eta \frac{\partial L(W; X)}{\partial W} \Big|_{W^{(i)}} \quad (11)$$

其中 η 是学习率。平均对数似然为:

$$L(W; X) = \frac{1}{N} \sum_{n=1}^N \log p(x_n; W) = \langle \log p(x; W) \rangle_0 \\ >_0 = -\langle E(x; W) \rangle_0 - \log Z(W) \quad (12)$$

其中, $\langle \cdot \rangle_0$ 表示模型在数据分布 $p_0 = \frac{1}{N} \sum_{n=1}^N \delta(x - x_n)$ 上的期望。从而更新量可以表示为:

$$\frac{\partial L(W; X)}{\partial W} = -\langle \frac{\partial E(x; W)}{\partial W} \rangle_0 + \langle \frac{\partial E(x; W)}{\partial W} \rangle_{\infty} \quad (13)$$

其中, $\langle \cdot \rangle_{\infty}$ 表示模型在目标分布 $p_{\infty}(x; W) = p(x; W)$ 上的期望。期望 $\langle \cdot \rangle_0$ 可以通过使用样本数据 X 来计算得到, 而期望 $\langle \cdot \rangle_{\infty}$ 需要计算归一化常数 $Z(W)$, 这就需要使用马尔可夫链蒙特卡罗抽样 (Markov Chain Monte Carlo, MCMC), 通过抽样的方式来近似计算 $\langle \cdot \rangle_{\infty}$ 。由于计算 $\langle \cdot \rangle_{\infty}$ 需要运行 MCMC 抽样至平衡状态, 而达到这种状态需要众多的 MCMC 步, 为减少这种计算需求, Roth 等提出用对比散度 (Contrastive divergence, C.D.) 方式^[9] 来训练 FoE 模型。这种方式的好处有: 1) 不需要计算归一化常数 Z 。由式(8) 可以看出, 当集团的尺寸为 3×3 、灰度为 256 级时, 集团 $x_{(k)}$ 的组合为 256^9 , 因而实际上计算 $Z(\Theta)$ 是不可能的。2) C.D. 训练相比于极大似然方式不需要将马尔可夫链运行至平衡状态, 只需要运行 MCMC 抽样 1 步或几步, 因而可以节省大量的时间。C.D. 方式的原理是通过缩减二个 KL 散度之间的梯度的方式来进行训练^[10]:

$$CD_n = KL(p_0 \parallel p_{\infty}) - KL(p_n \parallel p_{\infty}) \quad (14)$$

其中 $KL(p_0 \parallel p_{\infty}) = \sum_x p_0(x) \log \frac{p_0(x)}{p(x; W)}$ 。

当应用于 FoE 模型时, 其参数更新如下^[6]:

$$\delta \Theta_j = \eta [-\langle \frac{\partial E_{\text{FoE}}}{\partial \Theta_j} \rangle_{p^0} + \langle \frac{\partial E_{\text{FoE}}}{\partial \Theta_j} \rangle_{p_{\Theta}^l}] \quad (15)$$

其中 p^0 指训练集的分布, p_{Θ}^l 指 MCMC 抽样 l 步

后的训练集分布。

3 图像去噪仿真实验

(1) FoE 模型参数的训练手段简介。训练 FoE 模型主要是通过使用若干 CD_1 步更新步使 FoE 模型的参数得到逐步的修正, 最终达到使 FoE 模型的分布与训练集图像的分布相近似的目的。在应用 C.D. 前需要准备训练集, 目前使用较多的训练集为 Berkeley 图像集^[11], 通过从图像集中随机截取若干较小的图像片来构造训练集, 在构造完成训练集后, 可以在去除训练集的均值后于应用 C.D. 来训练 FoE 模型。也可以如文献[4]中所使用的方法。即先对训练集进行白化处理, 使训练集具有单位协方差, 然后在变换了的坐标下训练 FoE 模型。

在构造完训练集后, 训练集可以在去除均值后用于 FoE 模型的训练, 也可以按 Roth 等在文献[4]中所使用的先对训练集进行白化操作, 再用于 FoE 模型的训练, 其流程为先求出训练集的协方差矩阵 Σ , 然后对 Σ 进行特征值分解, 得到特征值矩阵 Λ 和特征向量矩阵 U , 设 y 为将图像片向量化后的训练集, 再进行白化 $\tilde{y} = \Lambda^{-1/2} U^T y$, 则 \tilde{y} 就是即将用于 FoE 模型训练的训练集。

滤波器的训练目前有三种形式, 第一种是直接训练, 第二种是由 Roth 等在文献[4]中使用的, 在变换空间 A 基下学习滤波器 J , 其中 A 又有二种形式, 分别是白化的和反白化的。根据 Roth 等在文献[4]中的结论第二种反白化的基 A 是这三种滤波器学习形式中处理效果最好的; 第三种由 Woodford 等在文献[7]中使用的, 通过在向量形式的集团后加 1, 即 $\tilde{x}^T = [x^T, 1]$ 并在向量形式的滤波器 J_i 后加 0, 即 $\tilde{J}^T = [J_i^T, 0]$ 的方式学习 J_i 。文中只使用了第一种方式。

对权 α 的学习目前有二种手段, 一种是直接学习权 α , 另一种是在对数域学习 α 。直接学习 α 的方式经常会得到负数形态的 α , 于是会出现反概率的现象, 但是文中的试验结果表明这种现象的出现, 并不会使 FoE 模型失效。在对数域学习 α , 不会出现反概率的现象, 同时会得到较直接学习 α 更好的效果。

(2) 数值实验参数的具体设置。文中使用的 FoE 的滤波器尺寸是 3×3 像素, 共使用 8 个专家。使用的训练集是从 Berkeley 图像集中的训练集中总共 200 幅图像将其转化为灰度图像后随机截取的 3000 幅 9×9 像素的图像片。

文中专家参数的初始设置为设 α 为 0.01, 初始化滤波器为正态分布的随机数。使用批训练方式(batch training)学习, 即将 3000 幅图像片分割为 30 个子集,

每个子集由 100 幅图像片组成, 在 C.D. 学习中循环依次使用, 共执行 C.D. 步 3000 步。在 C.D. 学习中使用的抽样方式为 HMC (Hybrid Monte Carlo) 抽样^[12,13], 并以各个样本子集分布为初始分布。由于 HMC 抽样的步长对 HMC 抽样的成功率有重大的影响, 依据文献[14], 文中所采用的初始步长为: $\epsilon_0 = 2.38/\sqrt{d}$, 其中 d 为 FoE 模型的维数, 由式(6) 可知维度 $d = \text{专家数} \times \text{滤波器尺寸} = 8 \times 3 \times 3 = 72$ 。文中所使用的蛙跳(leap frog)步为 30 步。随着训练的进行, 为保证接受率的稳定, 采用的步长调整方式为设 ϵ_i 指第 i 次更新时 HMC 抽样所使用的步长, 若接受率少于 80%, 则 $\epsilon_{i+1} = \epsilon_i \times 0.9$; 若接受率为 100%, 则 $\epsilon_{i+1} = \epsilon_i \times 1.01$; 其它情况 $\epsilon_{i+1} = \epsilon_i$ 。在应用时, 步长的更新与参数相同, 为每批更新一次。采用这种更新方式可以使接受率在训练过程中维持在 90% 左右。文中对 FoE 参数的更新为每批更新一次, 更新的方法有二种, 一种是直接方式:

$$\Theta_{i+1} = \Theta_i + \eta \Delta \Theta_i \quad (16)$$

另外一种是在对数域对 α 更新^[6]:

$$\begin{cases} J_i^{(n)} = J_i^{(n-1)} + \eta \Delta J_i^{(n-1)} \\ \alpha_i^{(n)} = \exp\{\log(\alpha_i^{(n-1)}) + \eta \alpha_i^{(n-1)} \Delta \alpha_i^{(n-1)}\} \end{cases} \quad (17)$$

在文中设学习率 η 为 0.01, 对 $\Delta \Theta_i$ 的计算使用以下方式:

$$\Delta \Theta_{i+1} = 0.1 \times \Delta \Theta_i + 0.9 \times \Delta \Theta_{i-1} \quad (18)$$

在更新 J 的同时, 去掉了 J 的平均值。

在贝叶斯图像恢复中, 对目标图像 y 的恢复是指在知道目标图像 y 情况下, 通过最大后验概率 $p(x | y) \propto p(y, x) p(x)$ 的方式来找出原始图像 x 。文中假设噪声为 $N(0, \sigma)$ 。从而, 似然可以写为:

$$p(y | x) \propto \prod_{k=1}^K \exp\left(-\frac{1}{2\sigma^2}(y_k - x_k)^2\right) \quad (19)$$

其中 k 指图像中相应像素的位置。处理目标图像的手段为梯度上升法, 其方式为:

$$x_{i+1} = x_i + \mu \nabla_x \log p(x | y) \quad (20)$$

其中 $\nabla_x \log p(x | y) = \left[\sum_{i=1}^N J_-^{(i)} * \Psi'(J^{(i)} * x; \alpha_i) \right] + \frac{\omega}{\sigma^2}(y - x)$ 。式(20)中 μ 是步长, 文中设为 0.2, 由文献

[4] $w = \frac{\lambda}{1-\lambda}$, 且 λ 需要在使用前进行搜索以获得最好的处理效果, 其方法为从 Berkeley 图像集中的训练集中随机截取 10 个 200×200 的图像片, 构成一个图像片集, 对于不同的 σ 分别采用 $[0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9]$ 为 λ 值进行去噪处理, 并且利用三次样条插值的方法选取新的 λ 使 PSNR 值达到最大。为获得更优良的 λ 值还需要再次搜索, 即将所选择的

λ 值加上 $[-0.06, -0.04, -0.02, 0, 0.02, 0.04, 0.06]$ 作为 λ 值, 再次进行选取操作。文中所应用的梯度上升步总数为 3000 步。对测试图像^[15]的去噪结果见表 1。

表 1 应用训练得到的 FoE 模型对五幅测试图像进行去噪得到的结果(PSNR)

(a)					
σ	Boat	House	Lena	Peppers256	Barbara
10	30.72	32.62	32.19	31.28	30.49
25	25.43	26.75	26.92	25.39	24.56
50	15.38	15.45	15.44	15.38	15.30
(b)					
σ	Boat	House	Lena	Peppers256	Barbara
10	30.74	32.75	32.41	31.43	30.60
25	25.91	27.66	27.91	25.91	25.21
50	15.54	15.59	15.63	15.52	15.51
(c)					
σ	Boat	House	Lena	Peppers256	Barbara
10	31.99	33.84	33.88	33.08	31.45
25	27.47	29.78	29.44	27.85	25.91
50	17.84	17.97	18.02	17.78	17.59
(d)					
σ	Boat	House	Lena	Peppers256	Barbara
10	30.62	31.92	32.45	31.23	30.28
25	26.21	27.28	28.37	26.02	25.20
50	23.01	23.35	24.44	22.29	22.34
(e)					
σ	Boat	House	Lena	Peppers256	Barbara
10	32.79	34.09	34.34	33.32	30.98
25	27.91	30.10	30.08	28.17	24.53
50	24.29	26.31	26.60	23.85	22.78

(3)一种改进的批训练方法。即使使用了 C. D. 学习方式, FoE 模型学习所涉及的计算量仍然是十分巨大的, 为此提出一种改进的批训练方法, 具体为通过分次逐步增大每个批的数量, 一方面可以充分利用批训练的优势, 在训练前期减少计算量使模型加速学习; 另一方面通过增大每个批所包含的图像片的数量使模型更新量的数值更准确。在更换批数量的同时重置更新量 $\Delta\theta$, 其目的是使模型在更换批量之后可以重新适应, 也可以理解为, 应用上次批学习的结果再进行一次批训练。具体应用细节为: 设置批数量分别为 100, 200, 500, 1000 共 4 个, 依次应用。应用时, 在每个批数量下重复使用同一训练集 10 次, 以批训练的方式对 FoE 模型进行 C. D. 学习。在更换批数的同时将 $\Delta\theta$ 置零。本训练方法的全部更新步数为 540 轮, 所使用的时间约为未改进时的 50%, 所得到的图像去噪结果与未改进的方法相似, 具体图像去噪结果见表 2。

表 2 应用改进的批训练方法训练得到的 FoE 模型对五幅测试图像进行去噪得到的结果(PSNR)

(a)					
σ	Boat	House	Lena	Peppers256	Barbara
10	30.69	32.69	32.41	31.36	30.59
25	26.04	27.95	28.02	26.23	25.35
50	15.90	15.84	15.99	15.86	15.83
(b)					
σ	Boat	House	Lena	Peppers256	Barbara
10	30.72	32.66	32.27	31.39	30.53
25	26.05	27.88	27.87	26.07	25.17
50	15.96	16.01	16.06	15.86	15.91
(c)					
σ	Boat	House	Lena	Peppers256	Barbara
10	30.47	31.72	31.09	30.44	29.86
25	25.89	27.04	26.70	25.46	24.61
50	16.05	16.10	16.18	15.99	15.94
(d)					
σ	Boat	House	Lena	Peppers256	Barbara
10	30.35	31.28	31.77	30.78	30.28
25	25.63	26.39	27.76	25.18	24.90
50	21.81	21.89	22.75	21.05	21.48
(e)					
σ	Boat	House	Lena	Peppers256	Barbara
10	32.73	34.23	34.35	33.38	30.50
25	27.40	29.61	29.69	27.57	24.04
50	23.93	25.21	26.27	23.24	22.59

注: 表 1 和表 2 中(a), (b), (c)的训练设置为使用 student-t 专家分别在不白化训练集直接更新 α 、不白化训练集在对数域更新 α 、白化训练集在对数域更新 α ; (d)和(e)的训练设置为使用 charbonnier 专家分别在不白化训练集在对数域更新 α 和白化训练集在对数域更新 α 。

4 分析与展望

FoE 模型是一种新出现的模型, 它的处理效果是十分优良的。不过该模型也存在尚需解决的问题, 诸如 FoE 模型的训练时间相对较长, 训练的收敛问题尚没有明确的解决方案。

马尔可夫随机场的提出已经历了很长时间, 然而它的研究与探讨仍然处于创新发展阶段。虽然当前已经发展了出很好的研究内容和新型的相关应用途径, 未来可以探索的内容和应用的方式方法将是十分广泛的。对 FoE 模型的探讨对于引导更多应用问题的解决, 应用于更广泛的领域具有积极的意义。

参考文献:

[1] Zhu S C, Wu Y N, Mumford D B. FRAME: Filters, Random field and Maximum Entropy: Towards a Unified Theory for Texture Modeling[J]. International Journal of Computer

比较好的一种优化方法。

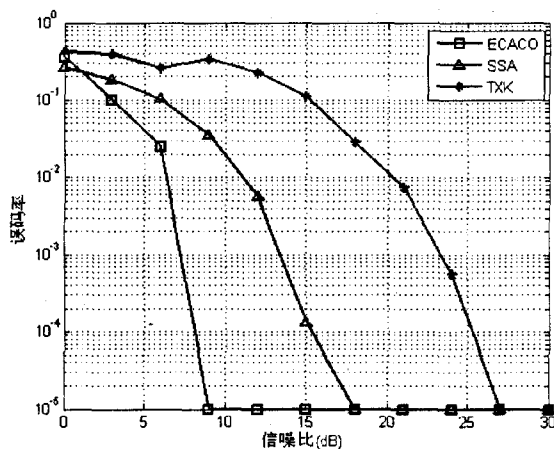


图 4 改进的蚁群优化算法与经典二阶盲均衡算法性能比较

参考文献:

- [1] 邱国霞, 张志涌. 基于离散微粒群优化的 MIMO 系统多用户盲检测[J]. 南京邮电大学学报, 2006, 26(6): 27-30.
- [2] Tong Lang, Xu Guanghan, Thomas K. Blind channel identification and equalization using second-order statistics: A time domain approach[J]. IEEE Transactions on Information Theory, 1994, 40(2): 340-349.
- [3] Shen Junqiang, Ding Zhi. Zero-forcing blind equalization based on subspace estimation for multiuser systems[J]. IEEE Transactions on Communications, 2001, 49(2): 262-271.
- [4] Maniezzo V, Colomi A, Dorigo M. The ant system applied to the quadratic assignment problem[R]. Belgium: Universit Libre de Bruxelles, 1994.
- [5] 张志涌, Er-Wei Bai. SIMO 含公零信道的直接盲序列检测[J]. 电子学报, 2005, 33(4): 671-675.
- [6] Zurada J M, Neural Networks: Binary Monotonic and Multiple-Valued[C]// Proc. of the 30th IEEE International Symposium on Multiple-Valued Logic. Portland, Oregon: [s. n.], 2000.
- [7] Dorigo M, Maniezzo V, Colomi A. Ant system: Optimization by a colony of cooperating agents[J]. IEEE Transactions on Systems, Man, and Cybernetics - part B Cybernetics, 1996, 26(1): 29-41.
- [8] 李士勇. 蚁群算法及其应用[M]. 哈尔滨: 哈尔滨工业大学出版社, 2004.
- [9] 吴启迪. 智能蚁群算法及其应用[M]. 上海: 上海科技教育出版社, 2004.
- [10] 徐精明, 曹先彬, 王煦法. 多态蚁群算法[J]. 中国科学技术大学学报, 2005(2): 2-4.
- [11] 刘玉霞, 王 萍, 修春波. 基于模拟退火策略的逆向蚁群算法[J]. 微计算机信息, 2006, 22(12/1): 265-267.
- [12] Ding Zhi. Blind equalization and identification[M]. [s. l.]: Marcel Dekker, Inc, 2000.
- [13] Vision, 1998, 27(2): 1-20.
- [14] Freeman W T, Pasztor E C, Carmichael O T. Learning Low-Level Vision[J]. International Journal of Computer Vision, 2000, 40(1): 25-47.
- [15] Roth S, Black M J. Fields of Experts[J]. International Journal of Computer Vision, 2009, 82(2): 205-229.
- [16] Mcauley J J, Caetano T S, Smola A J, et al. Learning High-Order MRF Priors of Color Images[C]// Proceedings of the 23rd International Conference on Machine Learning (ICML 2006). New York: ACM Press, 2006.
- [17] Roth S, Black M J. On the Spatial Statistics of Optical Flow[J]. International Journal of Computer Vision, 2007, 74(1): 33-50.
- [18] Roth S. High-Order Markov Random Fields for Low-Level Vision[D]. Stefan Roth Ph. D. Dissertation, Brown University, 2007.
- [19] Wooford O J, Reid I D, Torr P H S, et al. Fields of Experts for Image-based Rendering[C]// British Machine Vision Conference 2006 (BMVC 2006). Edinburgh: British Machine Vision Association, 2006.
- [20] Hinton G E. Products of experts[C]// Proceedings of the Ninth International Conference on Artificial Neural Networks. ICANN 99. London: The Institution of Electrical Engineers, 1999.
- [21] Hinton G. Training Products of Experts by Minimizing Contrastive Divergence[J]. Neural computation, 2002, 14(8): 1771-1800.
- [22] Carreira-Perpinán M Á, Hinton G E. On contrastive divergence learning[C]// 10th Int. Workshop on Artificial Intelligence and Statistics (AISTATS 2005). London: The Society for Artificial Intelligence and Statistics, 2005.
- [23] Arbelaez P, Fowlkes C, Martin D. The Berkeley Segmentation Dataset and Benchmark[DB/OL]. 2007-06. <http://www.eecs.berkeley.edu/Research/Projects/CS/vision/bsds/>.
- [24] Andrieu C, Freitas N, Doucet A, et al. An Introduction to MCMC for Machine Learning[J]. Machine Learning, 2003, 50(1): 5-43.
- [25] Neal R M. Probabilistic Inference Using Markov Chain Monte Carlo Methods, CRG-TR-93-1[R]. Toronto: Dept. of Computer Science, University of Toronto, 1993.
- [26] Gelman A, Roberts G O, Gilks W R. Efficient Metropolis jumping rules[C]. New York: Oxford Sci. Publ. Oxford Univ. Press, 1996.
- [27] Protilla J. Test Image[DB/OL]. 2005-03-14. <http://dec-sai.ugr.es/javier/denoise/test-images/index.htm>.

(上接第 89 页)