2005 年 2 月
第 7 卷第 2 期

中国工程科学
Engineering Science

Feb. 2005
Vol.7 No.2

学术论文

线性人脸对象类模型的匹配提升技术

付　昀，郑南宁
（西安交通大学人工智能与机器人研究所，西安　710049）

[摘要]　针对真实感人脸模型匹配的细节控制和稳健创建问题，提出了线性人脸对象类模型的匹配提升技术。基于非统一抽样（NUS）的动态高斯金字塔分析（DGPA）方法，结合不等概率抽样和整群抽样策略，自适应地动态调整每级高斯金字塔图像的抽样分布，利用最优化算法由粗到精的计算全局近似最优解，获得精确的模型匹配。动态调节整群区域边界并利用再抽样率调节抽样密度，可以有效控制人像模型的细节表达效果，提高模型创建的稳健性。随机梯度下降的线性相关性扰动（CD－SGD）和学习率自适应（ALR）技术，提高了模型匹配的准确性和收敛速度。以 MPI 和 AI&R 人像库为测试样本，主观与客观评价的实验结果验证了该模型匹配提升技术的有效性。

[关键词]　人脸建模；模型匹配；随机梯度下降；非统一抽样；多分辨率分析
[中图分类号]　TP391　　[文献标识码]　A　　[文章编号]　1009－1742（2005）02－0047－10

1　引言

人像处理和人脸属性（年龄、性别、人种等）变换是计算机视觉和图像图形学领域的重要研究内容[1~5]。在真实感人像处理技术中强调人像建模的准确性和细节化，从而捕捉并重构出逼真自然的人脸动作和属性变化效果。统计学模型[6~8]可以避免分析复杂的个体人脸结构和解剖信息，这类信息通过对大量的人像样本的综合分析获取。以人脸图像为特定对象的线性对象类模型（linear object class)[9~14]是建立在人脸图像线性空间假设的基础上的统计模型，即认为人脸空间中任意一幅人脸图像都可以通过空间基底的线性组合有效表达，空间的维数（秩）由组成人脸图像的像素点个数决定。在工程中通常应用 2 种空间基底：原型样本和 PCA 主元[15~18]。2 种技术在相应的人像建模过程中都没有应用整个人像空间进行人像表达，而是分别应用有限维的子空间进行表示。原型样本技术中的子空间维数就是样本个数；PCA 技术[18]中的子空间维数就是 PCA 主元的个数。这种利用子空间表达线性人脸对象类的思想是建立在实际工程应用基础上的，即尽量降低建模和计算开销，并获得足够主客观评价的人像处理质量。

利用线性人脸对象类模型建模的文献包括：Blanz 与 Vetter[6] 提出的 3D 渐变模型（morphable model），利用数据库中 200 幅激光扫描获取的人脸三维数据的线性组合自动建立，通过调节模型参数产生逼真的诸如表情、性别、脸型、姿态等属性变换效果；Vetter[7] 提出了大视点的人像变换，该方法基于线性对象类模型并结合 3D 模型，进行人脸图像单视点到多视点的变换处理，而且视点计算结果具有人像的个性特征（如斑点，皱纹等）；Jones 等[12] 提出的多维渐变模型（multidimensional morphable models），直接在高维空间进行模型匹配

[收稿日期]　2004－01－08；修回日期　2004－03－08
[基金项目]　国家自然科学创新研究群体基金资助项目（60024301）；国家自然科学基金资助项目（60205001）；河南省重大科技基金资助项目（0222020400）
[作者简介]　付　昀（1979－），男，西安市人，西安交通大学硕士研究生

和参数优化，以及利用原型样本线性组合表达新人像的建模和匹配技术[11, 16]。

建立线性人脸对象类模型的关键技术之一是模型匹配，即如何求得最佳的人像模型表达（近似全局最优）。有效的技术包括随机梯度下降方法[11, 16, 19, 20]和伪逆运算方法[15]等。前者对于人脸细节信息的模型表达能力强，而运算量大，迭代次数多；后者对人脸的整体信息表达能力强，运算量较小。笔者利用随机梯度下降算法进行模型匹配。模型匹配提升技术是为了进一步提高模型的细节表达和整体表达的折中效果，有效的方法包括，先进行人像五官分割，然后对分割后的图像区域进行线性对象类建模，最后融合区域模型为一个完整的人像模型[7]。该方法对人像的细节表达充分，但是为了消除区域的突变边界，需要增加滤波操作。高斯金字塔[21]分析可以提高模型匹配的鲁棒性[11~16]，抽样匹配可以减少运算量且不影响匹配质量[16]。笔者提出的线性人脸对象类模型的匹配提升技术，包括随机梯度下降的相关性扰动和学习率自适应，非统一抽样以及动态高斯金字塔分析等。

2 线性人脸对象类模型及模型匹配

2.1 人像对准模型

在建立精确的像素级人像对准前，需要先进行基于主要点特征的预对准[11~13]。这些主要点特征包括2个瞳孔中心，鼻尖和2个嘴角等。经过预对准实现特定人像在方位上和整体尺度上与参考人像或模型逼近，并且消除空间上的旋转现象。

定义人脸图像 I 为一个集合映射：$I:\mathbf{R}^2\rightarrow B$，其中 B 表示图像的灰度值集合，$\mathbf{R}$ 是实数集合。$I(x, y)$ 表示图像上某点的灰度值。设样本空间中的样本图像为 $I_r, I_1, I_2, \cdots, I_N$，其中 I_r 为参考图像（也可以定义某种参考模型 I_{model}），样本数目为 $(N+1)$。定义全局仿射变换 $A_j:\mathbf{R}^2\rightarrow\mathbf{R}^2$，可对图像进行平移、旋转和尺度变换，其中 $\boldsymbol{A}$ 满足

$$\boldsymbol{A}(x,y)=S\begin{pmatrix}\cos\theta & \sin\theta\\ -\sin\theta & \cos\theta\end{pmatrix}\begin{pmatrix}x\\ y\end{pmatrix}+\begin{bmatrix}t_x\\ t_y\end{bmatrix} \quad (1)$$

式中 S 为尺度，θ 为像平面旋转角度，t_x，t_y 为2D平移的参数。预对准前后的像素坐标对应关系为

$$I_j^{\mathrm{A}}(x, y)=I_j\circ A_j(x, y)=I_j\circ(x^{\mathrm{A}}, y^{\mathrm{A}})=I_j(A_j(x, y))=I_j(x^{\mathrm{A}}, y^{\mathrm{A}}) \quad (2)$$

其中运算符‘$\circ$’表示全局仿射变换，$A_j(x, y)=(x^{\mathrm{A}}, y^{\mathrm{A}})$，$(x^{\mathrm{A}}, y^{\mathrm{A}})$ 表示 I_j 中像素点 (x, y) 经过全局仿射变换后的坐标位置。对上述6个参数的确定需要已知至少3对相互对准的特征点坐标。

人脸形状包括图像中人脸的轮廓、大小和五官的位置分布信息。为了对形状进行有效标定，选择多条线段作为特征描述单位，将人脸对象按照生物外观形态特征分为脸庞、眼睛、眉毛、鼻子、嘴巴和耳朵6部分，分别利用多条线段将这几部分标定出来，得到6个线段集合，相应的定义为 S_1，S_2，S_3，S_4，S_5，S_6。这些线段集合中线段的数目分别定义为 n_1，n_2，n_3，n_4，n_5，n_6。特征线段的长度，数量和位置的变化会对最终的对准效果产生显著的影响。定义特征线段 l 为一个映射：$l:\mathbf{R}^4\rightarrow l$，其中 l 表示特征线段集合，$l_i(x_1, y_1, x_2, y_2)$ 表示特征线段集合中的一个特征，即一条线段特征由2个端点的坐标来决定，这样得到

$$S_i=\{l_{s1}, l_{s2}, \cdots, l_{sn_i}\}, i=1, 2, 3, 4, 5, 6 \quad (3)$$

特定人像与参考人像或参考模型的像素级对准，意味着主要特征点、特征线段的位置和数目应该是一致的。这样定义参考特征集合表示为 D_1，D_2，D_3，D_4，D_5，D_6，满足

$$\boldsymbol{D}_i=\{l_{d1}, l_{d2}, \cdots, l_{dn_i}\}, i=1, 2, 3, 4, 5, 6 \quad (4)$$

则人像对准的形状特征（线段集合）映射以及每个特征集合中的特征线段映射可以表示为

$$M_i^{\mathrm{SD}}:\boldsymbol{S}_i\rightarrow\boldsymbol{D}_i,\text{且 } M_i^l:\mathbf{R}^4\rightarrow\mathbf{R}^4 \quad (5)$$

人像纹理表示基于形状特征的图像灰度信息。人像纹理对准是耦合形状信息的像素级的全局性稠密对准。

每幅经过预对准和形状标定的样本图像 I_j^{A} 与参考图像 I_r 间的像素级对准，可以由

$$M_j:\mathbf{R}^2\rightarrow\mathbf{R}^2。 \quad (6)$$

的2D人像纹理映射表示。

式（6）表示由 I_r（或 I_{model}）到 I_j^{A} 的映射，例如 $M_j(x, y)=(\tilde{x}^{\mathrm{A}}, \tilde{y}^{\mathrm{A}})$，其中 $(\tilde{x}^{\mathrm{A}}, \tilde{y}^{\mathrm{A}})$ 表示 I_j^{A} 中与 I_r（或 I_{model}）中像素点 (x, y) 对准的点。这是后向映射，即逐一经过参考图像的每一个像素，在特定图像中取出对应像素。其优点是对准后的人脸图像中的每一个像素都有相应的像素填充，这就可以避免利用像素插值技术填补空缺位置时造成的失真[22]。由此可以定义像素级对准后的新样本图像

$\tilde{I}_j^{\mathrm{A}}$ 为

$$\tilde{I}_j^{\mathrm{A}}(x, y) = I_j \circ M_j(x^{\mathrm{A}}, y^{\mathrm{A}}) = I_j \circ (\tilde{x}^{\mathrm{A}}, \tilde{y}^{\mathrm{A}}) = I_j(M_j(x^{\mathrm{A}}, y^{\mathrm{A}})) = I_j(\tilde{x}^{\mathrm{A}}, \tilde{y}^{\mathrm{A}}) \quad (7)$$

运算符‘$\circ$’表示纹理映射变换。

为了实现精确的像素级对准，需要考虑在预对准的基础上进行纹理对准与形状对准的耦合。这样就得到新的人像纹理对准模型：

$$M_i^{\mathrm{SD}} : \boldsymbol{S}_i^{\mathrm{A}} \rightarrow \boldsymbol{D}_i^{\mathrm{A}}, i = 1, 2, 3, 4, 5, \text{且}$$

$$\tilde{I}_j^{\mathrm{A}}(x, y) = I_j(\tilde{x}^{\mathrm{A}}, \tilde{y}^{\mathrm{A}}), j = 1, 2, \cdots, N \quad (8)$$

人像纹理对准模型是建立在人像形状标定模型的基础上的，通过对人像形状标定模型中的 6 个特征集合的元素调整，可以控制每个表达区域产生不同的对准结果。当然，仅依靠主要特征的对准还无法实现完美的全局像素对准。但是通过特征集合间的合理配合可以达到理想效果。

2.2　线性对象类模型

线性对象类（linear object class）模型[9~14]的基本思想是，假设一类对象集合中的任何一个对象都可以由其他对象或对象元通过线性组合的方式表达。对象元可以是主分量分析（PCA）主元[18]，或者是未经过数据处理的样本元素（prototype）等[16, 17]。

假设 I^{object} 为一个对象元图像，并表示为一个映射：$I^{\mathrm{object}} : \mathbf{R}^2 \rightarrow B$，其中 $I^{\mathrm{object}}(x, y)$ 表示对象图像中一个像素的点的灰度值。计算对象集合中每一个对象元到参考对象元图像 I_{r} 的稠密对准，对准后的对象元图像表示为 $\tilde{I}_1^{\mathrm{object}}, \tilde{I}_2^{\mathrm{object}}, \cdots, \tilde{I}_N^{\mathrm{object}}$，其中 N 是集合中对象的个数。通过对象元图像的线性组合建立的对象类模型为

$$I_{\mathrm{model}} = I_{\mathrm{r}} + \sum_{i=1}^{N} p_i \tilde{I}_i^{\mathrm{object}}(x, y) \quad (9)$$

其中 $\boldsymbol{p} = [p_0, p_1, \cdots, p_N]$ 是线性模型的系数向量。结合基于对象形状的稠密对准，线性组合 $\sum_{i=1}^{N} p_i \tilde{I}_i^{\mathrm{object}}(x, y)$ 描述了一个由模型表达的对象元的纹理信息。对象集合外的同类新对象图像可以用已建立的对象类模型匹配表达，即 $I_{\mathrm{novel}} \approx I_{\mathrm{model}}$。取 I^{object} 为原型样本 I，经过对准后的图像纹理为 $\tilde{I}$，则 $I_{\mathrm{model}} = \sum_{i=0}^{N} p_i \tilde{I}_i(x, y)$，其中 $I_0 = I_{\mathrm{r}}$。

为了获得最佳的模型匹配和对象表达效果，定义新图像和模型重构图像间的误差能量函数为

$$E(p_0, p_1, \cdots, p_N) = \frac{1}{2} \sum_{x,y} [I_{\mathrm{novel}}(x, y) \circ M(x, y) - I_{\mathrm{model}}(x, y)]^2 = \frac{1}{2} \sum_{x,y} [I_{\mathrm{novel}}(x, y) \circ M(x, y) - \sum_{i=0}^{N} p_i \tilde{I}_i(x, y)]^2。$$

最佳的建模匹配求解过程是一个最优化问题。配合提出的新模型匹配技术（基于相关性扰动的随机梯度下降算法、学习率自适应、非统一抽样和动态高斯金字塔分析），求解误差能量函数 $E(\boldsymbol{p})$ 达到全局近似最优解时的线性系数 $\boldsymbol{p}$，就可以利用有限的样本集合描述最佳的模型匹配效果。计算出的局部最优线性系数解表示为 $\boldsymbol{p}^* = (p_0^*, p_1^*, \cdots, p_N^*) \in \mathbf{R}^{n+1}$，使误差 $E(\boldsymbol{p}^*)$ 满足最小值。

线性对象类模型的表达效果生动、自然且有相当高的真实效果，数据冗余少，通过建立不同的对象元数据库，可以获得不同特性的仿真效果。唯一的不足是该模型对于特定对象元的细节特征无法表达，特别是在包括人脸图像在内的特殊对象类的表达中尤为突出。但是从应用的角度看，该模型是相当有效和稳定的。加入一定的个体信息后建立的模型可以基本克服该模型的缺陷。

3　模型匹配提升

3.1　随机梯度下降及线性相关性扰动

设标量函数 $E(\boldsymbol{p})$ 是一个光滑且非负的能量函数，在其上任一点 $\boldsymbol{p}_k$ 对应的梯度是一个向量，其方向为此函数 $E(\boldsymbol{p})$ 增长最快的方向，那么负梯度方向就为函数 $E(\boldsymbol{p})$ 下降最快的方向[19]；因此要求目标函数的最小值，只要能从任意初始点 $\boldsymbol{p}_0$ 出发，沿着负梯度方向按照

$$\boldsymbol{p}_{k+1} = \boldsymbol{p}_k - \boldsymbol{a}(\nabla E(\boldsymbol{p}_k) + \varepsilon) \quad (10)$$

迭代来更新 $\boldsymbol{p}$ 值，就可以快速地找到函数的极小值[20,23,24]。式中 $\boldsymbol{a}$ 为随机梯度下降学习率，$\boldsymbol{a} = [a_0, a_1, \cdots, a_N]$，$\boldsymbol{\varepsilon}$ 为零均值固定概率分布的随机扰动，$\boldsymbol{\varepsilon} = [\varepsilon_0, \varepsilon_1, \cdots, \varepsilon_N]$，可以使 $E(\boldsymbol{p})$ 在局部最优点周围产生随机振荡，从而跨越局部最优解和全局极优解，尽量逼近全局最优解[20]。其中 $[p_0^{(k)}, p_1^{(k)}, \cdots, p_N^{(k)}]$，$\nabla E(\boldsymbol{p}_k) = \left(\frac{\partial E}{\partial p_0^{(k)}}, \frac{\partial E}{\partial p_1^{(k)}}, \cdots, \frac{\partial E}{\partial p_N^{(k)}}\right)$，控制随机梯度下降的迭代终止条件为

$$\| \nabla E(\boldsymbol{p}_{k+1}) - \nabla E(\boldsymbol{p}_k) \| \leqslant \eta \quad (11)$$

或

$$\| \boldsymbol{p}_{k+1} - \boldsymbol{p}_k \| \leqslant \zeta \quad (12)$$

实现该算法的关键问题是如何选择学习率 $\boldsymbol{a}$、随机扰动 $\boldsymbol{\varepsilon}$ 以及迭代终止阈值 η，ζ。为了使迭代收敛，学

习率 $\boldsymbol{a}$ 必须是随着每一步迭代逐次衰减的。通常学习率 $\boldsymbol{a}$ 满足

$$\boldsymbol{a}_n = \boldsymbol{a}_0\beta(n) \tag{13}$$

其中 $\boldsymbol{a}_0$ 是初始化学习率向量，$\beta(n)$ 是小于1的衰减函数，通常取负指数函数 e^{-xn}，整数 n 在区间 $[0,\infty)$ 逐次递增，则有 $\boldsymbol{a}_n = [a_0^{(n)}, a_1^{(n)}, \cdots, a_N^{(n)}] = [a_0^{(0)}, a_1^{(0)}, \cdots, a_N^{(0)}]\beta(n)$。通常衰减学习率应满足

$$\sum_{n=0}^{\infty}\boldsymbol{a}_n = \infty, \text{且} \sum_{n=0}^{\infty}\boldsymbol{a}_n^2 < \infty \tag{14}$$

学习率衰减过快或者过慢都会使 $\boldsymbol{p}$ 收敛到某一无效点，无法找到函数的极小值。此外，从算法优化角度来讲，为了提高随机梯度下降的学习效率和系统鲁棒性，利用高斯金字塔分解对每幅样本图像进行由粗到精的多分辨率处理，在每级金字塔中学习率 $\boldsymbol{a}$ 可选择不同的衰减规律。随机扰动 $\boldsymbol{\varepsilon}$ 的选择对于该算法的鲁棒性以及寻找极小值的准确性尤为重要。这里赋予 $\boldsymbol{\varepsilon}$ 向量与梯度值满足相关性，从而增加其扰动的针对性。定义具有与梯度值线性相关的随机扰动 $\boldsymbol{\varepsilon}$ 满足

$$\varepsilon_k = \varepsilon_0[\nabla E(\boldsymbol{p}_k)] \cdot \text{random}(\delta) \tag{15}$$

其中，ε_0 是随机扰动振幅，$\varepsilon_0 = [\varepsilon_0^{(0)}, \varepsilon_1^{(0)}, \cdots, \varepsilon_N^{(0)}]$ $\varepsilon_0 = [\varepsilon_0^{(0)}, \varepsilon_1^{(0)}, \cdots, \varepsilon_N^{(0)}]$，random (δ) 是在 $[-\delta, \delta]$ 范围内的随机数。式（10）可改写为

$$\boldsymbol{p}_{k+1} = \boldsymbol{p}_k - \boldsymbol{a}_0\beta(k)[1 + \varepsilon_0 \text{ random}(\delta)]\nabla E(\boldsymbol{p}_k),$$

$$p_i^{(k+1)} = p_i^k - a_i^{(0)}\beta(i)[1 + \varepsilon_i^{(0)} \cdot \text{random}(\delta)] \cdot \left(\frac{\partial E}{\partial p_i}\bigg|_{p_i = p_i^{(k)}}\right) \quad i = 1, 2, \cdots, N \tag{16}$$

利用相关性扰动的随机梯度下降求解能量函数 $E(\boldsymbol{p})$ 极小值甚至最小值，需要得到其每个参数的偏导数：

$$\frac{\partial E}{\partial p_i} = -\sum_{x,y}\{[I^{\text{novel}}(x,y) \circ M(x,y) - \sum_{i=0}^{N} p_i\tilde{I}_i(x,y)]\tilde{I}_i(x,y)\} \tag{17}$$

适用于这实验条件的经验结论：初始化学习率 $\boldsymbol{a}_0 = [1.0, 1.0, \cdots, 1.0]_{N+1}$，衰减因子 $\beta(n) = e^{-0.1n}$，随机扰动振幅 $\boldsymbol{\varepsilon}_0 = 3.2\,\boldsymbol{a}_0$，迭代终止阈值 $\eta, \zeta = 0^{-9}$，随机数幅度 $\delta = 0.1$。

3.2 学习率自适应

在3.1节中提到的随机梯度下降法是依靠固定的学习率衰减幅度，并结合相关性随机扰动实现全局极小值搜索的。这里还可以采用学习率自适应配合固定随机扰动的方法实现梯度下降法求解[5]。由式（10）开始，定义学习率满足

$$a_i^{(n+1)} = a_i^{(n)}\beta(n) = a_i^{(n)}\exp[\lambda q_i^{(n)}q_i^{(n+1)}/h_i^{(n)}] = \begin{cases} a_i^{(n)}u, & q_i^{(n)}q_i^{(n+1)} \geqslant 0 \\ a_i^{(n)}v, & q_i^{(n)}q_i^{(n+1)} < 0 \end{cases} \tag{18}$$

其中 $q_i^{(n)}$，$h_i^{(n)}$ 满足

$$q_i^{(n)} = \frac{\partial E}{\partial p_i}\bigg|_{p_i = p_i^{(n)}} + \varepsilon_i,$$

$$h_i^{(n)} = \gamma h_i^{(n-1)} + (1-\gamma)(q_i^{(n+1)})^2,$$

$$0 < \gamma < 1,\ i = 1, 2, \cdots, N。$$

可以看到第 $n+1$ 次迭代时的学习率是由第 n 次迭代时的学习率和一个幅度因子决定的，当相邻两次迭代时的偏导数异号时，幅度因子小于1，反之大于1。经验结论：$1 \leqslant u \leqslant 1.1$，$0.91 \leqslant v \leqslant 1$。

3.3 非统一抽样

抽样是一种非全面的信息获取，是指从研究对象的全体中抽取一部分单位作为样本，根据对所抽取的样本信息进行研究，获得有关总体目标量的了解。这是文献［25］中抽样的广义概念。非统一抽样（non-uniform sample）是一种非均匀的抽样统计分析方法，它是不等概率抽样和整群抽样的结合。不等概抽样（PS）即对于样本集合中的元素采用不等概率的随机抽取。赋予总体每个元素一个不同的入样概率，这样某类元素就比其他元素出现在样本中的机率大。这类抽样方法很多，按照样本元素是否放回分为放回不等概率抽样（PPS）和不放回不等概率抽样（πPS）[25]。整群抽样（cluster sampling）是将总体划分为若干群，然后以群为抽样单元，从总体中随机抽取一部分群，对选中群中的所有元素进行研究的一种抽样技术。群的规模是指组成群的元素数目。规模大，估计的精度差但运算量小；规模小，估计的精度高但运算量大。群划分可以根据元素在整体中的区域分布和密度分别来灵活划分。在群规模不等的整群抽样中，如果群规模差异较大，各个群对总体的影响是不同的。可以采用不等概率方式抽样群。这种结合不等概率抽样的整群抽样可以提高估计效果，具有容易定义和识别，以及操作稳定的优点。

在实际应用中，为了对整体和细节同时把握，采用多步抽样的方法来产生非统一抽样效果。首先对样本集合中所有的元素用一定数量的等概率抽样，然后

对所关注的细节特征（整群）进行再抽样，甚至对于更加重要的细节特征多次抽样。对于多次抽样的数据概率统称为再抽样率。定义元素总体为 n_{ALL}，再抽样元素总体为 n_{RE}，则再抽样率为

$$r_{rs} = n_{RE}/n_{ALL} \tag{19}$$

细节特征整群的定义以及位置的划分需要根据实际的研究对象来定。不同的再抽样率可以表达不同的估计效果，可操作性较强。

为了引入动态高斯金字塔分析，介绍分层随机抽样（stratified sampling）的概念[25]。首先将总体 N 个元素划分成 L 个子总体，每个子总体称为层，它们的大小分别为 N_1，N_2，…，N_L，这 L 个层合起来就是整个总体 $N = \sum_{i=1}^{L} N_i$；然后，在每个层中分别独立的进行抽样，所得到的样本称为分层样本。其中每层的抽样可以采用不等概率抽样或整群抽样等不同的抽样方式。分层抽样的估计精度和抽样效率都比较高，不仅可以对总体指标进行推算，也能对各层指标进行推算。

3.4　动态高斯金字塔分析

高斯金字塔（Gaussian pyramid）[21]分析也称作分层的多分辨率分析，是一种由低分辨率到高分辨率（由粗到精）逐步迭代求精的分析方法。原参考图像作为第一级金字塔；第二级金字塔是第一级图像通过一个 5×5 的低通滤波器产生的；第三级金字塔是第二级图像利用相同的方法产生，依次类推。对于 $C \times R$ 的参考图像，相邻级金字塔图像的关系满足

$$g_l(i,j) = \sum_{m=-2}^{2}\sum_{n=-2}^{2} w(m,n) g_{l-1}(2i+m, 2j+n),$$

其中 N 表示高斯金字塔的最大级数，C_l 和 R_l 表示第 l 级像素矩阵的列与行的维数，$1 < l < N+1$，$0 \leqslant i < C_l$，$0 \leqslant j < R_l$。由于滤波窗的大小为 5×5，所以最顶 2 级的图像尺寸不能都小于 5×5。原参考图像的维数与高斯金字塔级数的关系表示为：$C = M_C 2^N + 1$，$R = M_R 2^N + 1$，其中 M_C 和 M_R 是整数。一个 256×256 的 8 b 灰度图像最高可以建立一个 7 级的高斯金字塔，高一级图像是前一级图像维数的 1/4，图像的宽和高是前一级图像的一半。第 7 级图像尺寸是 4×4 像素单位。

这里采用的 5×5 滤波核是可分离的，即 $w(m,n) = \hat{w}(m)\hat{w}(n)$，在水平和垂直方向的滤波是独立的，长度为 5，滤波函数是归一化的并且是对称的。此外，每一级中所有的点还需满足向高一级提供相同的整体权重。则该滤波核满足 3 个约束条件：$\hat{w}(0) = a$，$\hat{w}(-1) = \hat{w}(1) = 1/4$，$\hat{w}(-2) = \hat{w}(2) = 1/4 - a/2$。通常工程中选择 $a = 0.4$，该滤波核如图 1 所示。

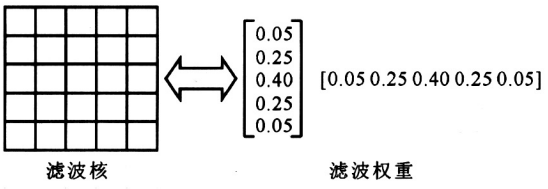

图 1　高斯金字塔滤波核

Fig.1　Gaussian pyramid kernel

动态高斯金字塔（dynamic Gaussian pyramid）分析是结合非统一抽样技术的分层多分辨率分析方法。在迭代求精的分析过程中，定义两种动态分析策略：

1）*建立满级高斯金字塔*　即最大限度的创建高斯金字塔图像级，例如对于 $N \times M$ 像素单位的参考图像可分的最大级数为 n_{max}，其中 $(\min\{N, M\})^{1/n_{max}} \geqslant 5$。整个多分辨率计算过程是从最高一级开始依次迭代到最低一级或者某一次低级终止。在每一级迭代前进行最优非统一抽样率的估计，然后在该抽样率下随机抽取样本点。这就意味着非统一抽样率在不同级的分析过程中是自适应的改变和调整的。通常情况下，由于高一级图像的分辨率低于低一级图像，高一级迭代时的主特征再抽样率预测值也会低于低一级图像的预测值。在该分析策略中需要强调的是，最终迭代终止的金字塔分级不一定是最高级，因为在某些场合次于最高一级的计算结果反而优于最高级的计算结果，产生这种差异的主要原因是观察者的主观评价与实际计算的客观评价具有非一致性。

2）*建立非满级高斯金字塔*　即创建有限的高斯金字塔图像级数，例如对于最大级数为 6 级的图像只创建 5 级或 4 级。整个多分辨率的计算过程是从最低一级开始迭代到最高一级终止。每一级的非统一抽样率依然是自适应的改变和调整。这种策略主要应用于图像分辨率过低以及重构图像过于光滑的情况。

4　实验与分析

4.1　人脸图像实验样本

实验中采用两种人像样本：德国 Max-Planck

研究所的200幅高加索人脸图像样本[6]，其中包括100幅男性和100幅女性256×256的24 bit-BMP正面灰度图像，利用激光扫描设备（Cyberware™）拍摄的结构光信息；西安交通大学人工智能与机器人研究所（AI&R）的80幅东方人脸图像样本[1,2]，其中包括40幅女性和40幅男性256×256的8 bit-BMP正面灰度图像，利用OLYMPUS C－5050ZOOM数码照相机拍摄的模拟光信息。

4.2 实验1 人像对准及线特征控制

图2描绘了从一幅MPI的256×256的8 b灰度男性人像（特定人像）到一幅女性人像（参考人像）的对准效果，3次实验中人脸特征线段数目与分布如表1所示。特定人像和对准人像相比，可以明显看出整体上以及特定人脸局部的对准变换效果。例如人脸轮廓的缩小，耳朵部分变窄，眉毛上翘等；特定人像矢量场描述对准过程中图像像素的位移矢量分布。在实验2和实验3中，采用了不同的特定人脸形状线段特征数目（参见表1）进行对准处理。实验2中，不对耳朵部分进行特征标定，并且在脸部轮廓靠近眼角周围去除两处线段特征的标定。实验3中，去除额头部分的7条线段特征，并且在耳朵轮廓处增加2条线段特征。实验显示标线部位像素位移量明显增大，未标线部位像素位移量明显降低，特征线段增减可有效控制人像对准效果。

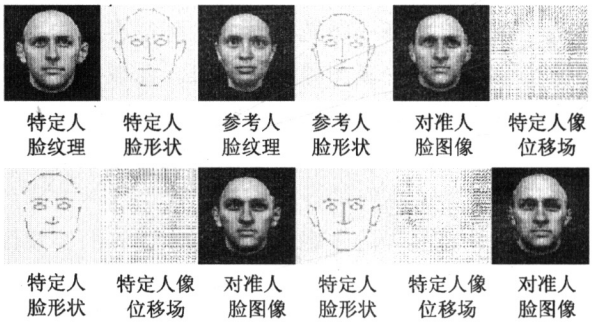

图2 不同特征线段数目表示形状的MPI人像对准效果

Fig.2 Correspondences of MPI image under varying line feature numbers

4.3 实验2 线性相关性扰动下的随机梯度下降

实验中对原图像进行高斯金字塔分层，对每级高斯金字塔人像随机抽样300个像素点，采用随机梯度下降，线性相关性扰动和学习率自适应技术进行人像重构。选择经验参数：初始化学习率 $\boldsymbol{a}_0 = [1.0, 1.0, \cdots, 1.0]_{N+1}$，衰减因子 $\beta(n) = e^{-0.1n}$，随机扰动振幅 $\boldsymbol{\varepsilon}_0 = 3.2\,\boldsymbol{a}_0$，迭代终止阈值 η，$\zeta = 10^{-9}$，随机数幅度 $\delta = 0.1$，学习率自适应参数 $1 \leqslant u \leqslant 1.1$，$0.91 \leqslant v \leqslant 1$

表1 实验中人脸形状线段特征数目

Table 1 Feature numbers of experimental face shape

部位名称	脸庞	眼睛	眉毛	鼻子	嘴巴	耳朵
特征集合	S_1	S_2	S_3	S_4	S_5	S_6
n_i(实验1)	14	8	4	3	2	6
n_i(实验2)	12	8	4	3	2	0
n_i(实验3)	7	8	4	3	2	8

图3显示了200幅MPI人像样本和80幅AI&R样本的库内重构效果。第一行图像是2幅MPI和2幅AI&R原图像，第二行图像是利用本文技术产生的重构图像，第三行图像是差图像，显示重构图像与实际图像的灰度值差异，差图像灰度颜色越重则差异越大。结果显示主要的重构误差集中于五官与脸庞的轮廓区域。由于线性对象模型本身是有损的，这种误差不可避免。但是，并不影响重构结果的主观相似度。

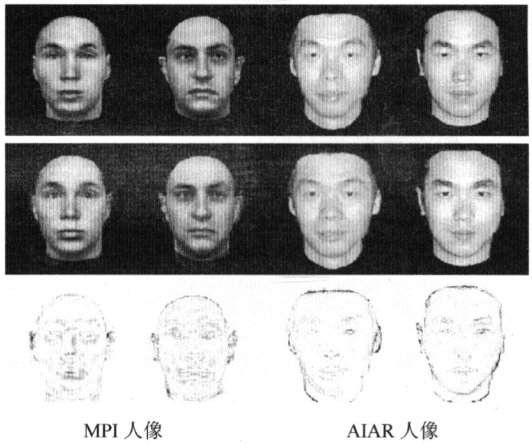

图3 人像库内的样本重构效果

Fig.3 Reconstruction of the face images in database

图4显示了人像库外新图像的重构效果。选择100幅MPI人像样本组成实验图像库1，其他100幅为库外人像。第一列图像是原图像，第二列图像是重构图像，第三列图像是差图像。比较图3与图4，结果显示库外人像重构误差大于库内人像重构。库外人像重构结果的重构误差仍然主要集中于五官与脸庞的轮廓区域。

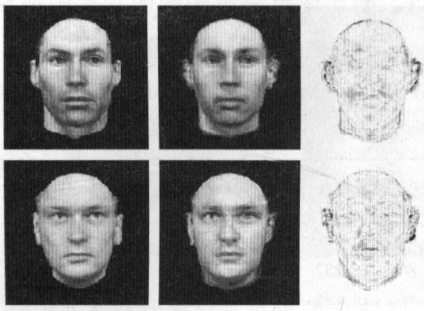

图 4　人像库外新图像的重建效果

Fig.4　Reconstruction of the new face images

4.4　实验 3 非统一抽样与动态高斯金字塔分析

在图 5 的实验中，选择 100 幅 MPI 人脸图像作为库内样本，输入一幅数据库外的新图像，分别利用统一抽样和非统一抽样方法对新图像进行重构，整群区域划分在眼睛，眉毛、嘴巴和鼻子的细节部分。将两种重构结果分别与原图像进行差运算获得差图像，结果显示非统一抽样在整群部位的确有明显的重构改善效果（差图像变淡）。这种改善也直接反映在正面的主观评价中。

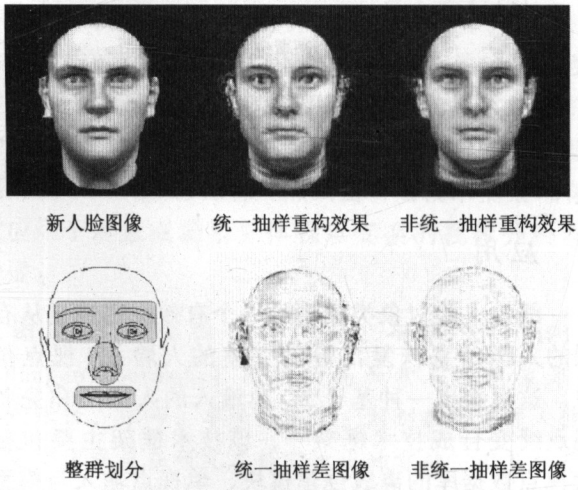

图 5　非统一抽样与统一抽样方法的重构效果比较

Fig.5　Comparison of the non-uniform and uniform sampling methods

迭代过程中对每一幅对准样本建立 6 级的动态高斯金字塔，从第 1 级开始动态选择 300 个抽样点作为估计样本，以 $r_{rs}=0.35$ 为估计值，并且动态的调整五官特征的再抽样率。每一级迭代结束后的准线性系数作为下一级的输入，同时调整再抽样率和整群区域的边界坐标。为了对重构图像效果进行客观评价，定义像素灰度值之间的平均距离差函数：

$$D_E=\sqrt{\frac{1}{WH}\sum(I_{novel}-I_{model})^2} \tag{20}$$

其中 W 表示图像宽度，H 表示图像高度。D_E 反映出特定图像与重构图像间的相似程度。D_E 越小则客观的重构效果越好。

图 6 显示了一幅 MPI 库外新图像与 200 幅平均脸图像。图 7 显示了对于图 6 人脸图像动态高斯金字塔分析的实验结果。其中每行结果显示一个特定再抽样率下的迭代计算过程。图 7a 为每一个特定再抽样率下的抽样模板，即随机抽样点的密度分布，图 7b 至图 7g 分别显示第 6 级到第 1 级的重建效果。迭代方向从高层到低层，上一级迭代终止时的线性系数作为下一级的输入参数。

图 6　库外新图像与平均脸图像

Fig.6　New face image and the average face image

在主观评价实验中，选择 20 位不同观察者来快速的评价出重构效果的最佳结果。如表 2 所示的主观评价结果，由于所有的观察者都只选择第 1 至第 3 级的结果，故表中只列出上述数据。其中，100 % 认为图 7f 的整体效果最佳。60 % 认为再抽样率为 0.6 的重构图像效果普遍较好，其他 40 % 认为再抽样率为 0.3 的重构图像效果普遍较好。

在客观评价中，利用公式（20）的平均距离差函数来计算图 7 中每幅重构图像与原图像（图 6）间的相似程度。表 3 显示图 7 中每幅重构图像与原图像间的平均距离差。$D_{E-3level}^{Average}$ 列表示第 1 级到第 3 级的距离差平均值，$D_E^{Average}$ 表示每级中不同再抽样率 r_{rs} 下的距离差平均值。图 8 是重构图像与原图像间的平均距离差曲线，其中唯一的一条实线代表平均曲线。结果显示，最小的 $D_E^{Average}$ 值是第 2 级的 14.440，这与主观评价的结果一致。

第 1 级到第 3 级的距离差平均值 $D_{E-3level}^{Average}$ 的最

小值是 14.304，即再抽样率 r_{rs}为 0.3 的重构结果。与如上的主观评价实验结论相比，仅有 40 %的主观评价结论与此吻合。存在差异的主要原因在于人类观察的选择性特征。通常情况下，当再抽样率 r_{rs}增大到一定程度时，重构对象的细节特征过于明显，并且其逼真效果与整体效果相比占有绝对优势。当再抽样率 r_{rs}为 0.6 时，观察者的整体真实感会被细节真实感所支配，所以有 60 % 认为再抽样率为 0.6 的重构图像效果普遍较好。

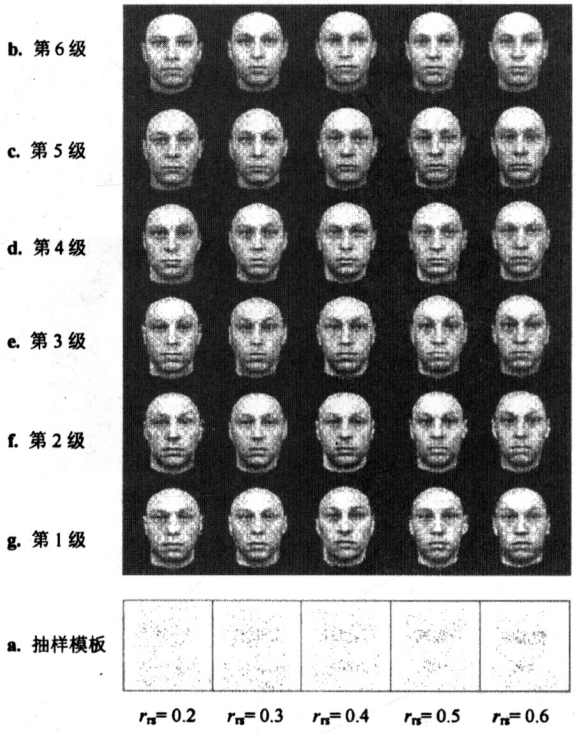

图 7 动态高斯金字塔分析实现的图 6 图像的重构效果比较

Fig.7 Comparison of the reconstructed face image in Fig.6 by dynamic Gaussian pyramid analysis

表 2 图 7 中每幅重构图像与原图像间相似度的主观评价

Table 2 Subjective quality evaluation of Fig.7

	第 3 级		第 2 级		第 1 级		最佳行
r_{rs}=0.2	1	5%	0	0	0	0	0
r_{rs}=0.3	3	15%	2	10%	8	40%	40%
r_{rs}=0.4	1	5%	11	55%	0	0	0
r_{rs}=0.5	11	55%	1	5%	0	0	0
r_{rs}=0.6	4	20%	6	30%	12	60%	60%
总和	20	100%	20	100%	20	100%	
最佳列	0		100%		0		

表 3 图 7 中每幅重构图像与原图像间的平均距离差

Table 3 Average distance of reconstructions in Fig.7

	第 6 级	第 5 级	第 4 级	第 3 级	第 2 级	第 1 级	$D_{E-3level}^{Average}$
r_{rs}=0.2	17.085	16.139	14.747	15.080	14.659	14.271	14.670
r_{rs}=0.3	16.867	16.287	14.126	14.565	14.242	14.106	14.304
r_{rs}=0.4	17.582	16.643	14.556	14.433	13.689	17.171	15.098
r_{rs}=0.5	16.979	16.165	15.155	14.120	14.398	16.591	15.036
r_{rs}=0.6	16.894	16.034	15.064	14.621	15.211	15.746	15.193
$D_{E-3level}^{Average}$	17.081	16.254	14.730	14.564	14.440	15.577	

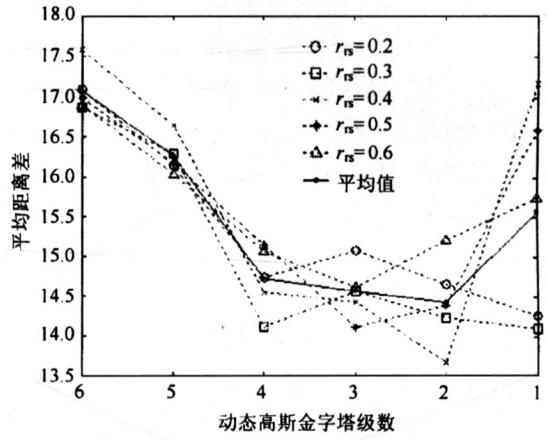

图 8 重构图像与原图像间的平均距离差曲线

Fig.8 Average distances between reconstructed image and the original image

5 应用

线性人脸对象类模型的一个有效应用就是从有限的人像信息恢复出近似完整的人脸 3D 视点信息。这里提出一种基于 2D 单输入的人像的非完整视点续变合成技术框架[2]，该技术框架主要包括大视点数据库的离线创建模块，单视点输入与模型匹配模块，以及大视点空间映射与连续视点重建模块。在离线学习阶段分别建立人脸多个离散视点下的大样本数据库以及相应的线性对象类模型；利用笔者提出的模型匹配提升技术，线性相关性扰动下的随机梯度下降，学习率自适应，非统一抽样和动态高斯金字塔分析求解模型参数；利用视点空间映射技术重构出输入人像的多个离散视点图像，利用图像视点前变换算法[2, 26]配合图像渐变算法[21, 27]实现从离散视点到连续视点图像序列的合成。图 9 给出了一个 AI&R 人像视点续变合成效果实例。从一幅正面人像合成连续多视点序列。

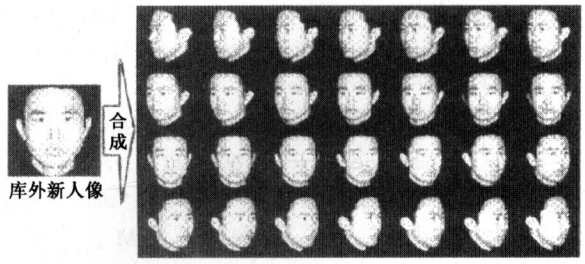

图 9　人像非完整视点续变合成实例

Fig.9　Invisible view reconstruction for an AI&R face example

6　结论

笔者研究了线性人脸对象类模型的稳健建模和细节匹配控制问题，提出了模型匹配的提升技术，利用线性相关性扰动下的随机梯度下降算法，配合学习率自适应方法计算全局近似最优解，结合不等概率抽样和整群抽样技术动态调整每级高斯金字塔图像的抽样分布（非统一抽样和动态高斯金字塔分析）提升模型匹配鲁棒性和准确性。实验结果表明，线性相关性扰动下的随机梯度下降算法和学习率自适应技术可以有效计算模型匹配；基于非统一抽样方法的动态高斯金字塔分析可以有效控制线性对象类细节的逼真重构效果，主观评价与客观评价的图像重建结果相吻合。最后提出的结合模型匹配提升技术的人像非完整视点续变合成技术框架可有效应用于对象或场景立体视点变换的视频分析与合成。

致谢　感谢德国 Max-Planck 研究所（MPI, Tübingen, Germany）提供的高加索人脸图像库；感谢张强，张婷，卓峰，王宏，刘剑毅等智能人像研究小组成员在研究工作中做出的贡献，以及给予的帮助；感谢所有配合 AI&R 东方人像库建库工作的研究人员，以及提供个人肖像的志愿人员。

参考文献

[1]　郑南宁，付　昀，张　婷，卓　峰. 人脸的表情与年龄变换和非完整信息的重构技术(上)[J]. 电子学报，2003，31(12A)：1955～1962

[2]　付　昀，郑南宁，张　婷. 人脸的表情与年龄变换和非完整信息的重构技术(下)[J]. 电子学报，2003，31(12A)：1963～1970

[3]　张　强. 基于稠密特征对应的人脸图像表达及人脸属性变换[D]. 西安：西安交通大学，2003

[4]　陈　洪. 人脸图像自动分析与绘制的统计学习方法[D]. 西安：西安交通大学，2002

[5]　卓　峰，徐维朴，张　强，等. 基于2D样本的人脸图像视点变换[J]. 计算机应用，2003，23(增刊)：74～76

[6]　Blanz V, Vetter T. A morphable model for the synthesis of 3D faces[A]. Computer Graphics Annual Conference Series, SIGGRAPH'99 [C]. ACM SIGGRAPH, Los Angeles, California, 1999. 187～194

[7]　Vetter T. Synthesis of novel views from a single face image [J]. International Journal of Computer Vision, 1998, 28(2): 103～116

[8]　Timothy F Cootes, Gareth J Edwards, Christopher J Taylor. Active appearance models [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2001, 23(6): 681～685

[9]　Ullman S, Basri R. Recognition by linear combinations of models [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1991, 13: 992～1006

[10]　Vetter T, Poggio T. Linear object classes and image synthesis from a single example image [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1997, 19(7): 733～742

[11]　Jones M J, Poggio T. Model-based matching of line drawings by linear combinations of examples [A]. IEEE International Conference on Computer Vision (ICCV'95) [C], Boston Massachusetts, 1995. 532～536

[12]　Jones M J, Poggio T. Multidimensional morphable models [A]. IEEE International Conference on Computer Vision (ICCV'98) [C], IEEE Computer Society, Bombay, 1998. 683～688

[13]　Vetter T, Jones M J, Poggio T. A bootstrapping algorithm for learning linear models of object classes [A]. IEEE International Conference on Computer Vision and Pattern Recognition (CVPR'97) [C], USA: IEEE Computer Society Press, Puerto Rico, 1997. 40～47

[14]　Beymer D, Poggio T. Image representation for visual learning [J]. Science, 1996, 272: 1905～1909

[15]　Beymer D. Vectorizing face images by interleaving shape and texture computations [R]. A. I. Memo 1537, Artificial Intelligence Laboratory, Massachusetts: MIT, 1995

[16]　Jones M J, Poggio T. Model-based matching by linear combinations of examples [R]. AI Memo No 1583, Artificial Intelligence Laboratory, Massachusetts: MIT, 1996

[17] Beymer D, Shashua A, Poggio T. Example based image analysis and synthesis [R]. AI Memo No 1431, Artificial Intelligence Laboratory, Massachusetts: MIT, 1993

[18] Torre F De la, Black M J. Robust principal component analysis for computer vision [A]. IEEE International Conference on Computer Vision (ICCV'01) [C], IEEE Computer Society, Vancouver, Canada, 2001. 362～369

[19] Leemon C Baird III. Reinforcement learning through gradient descent [D]. Pittsburgh: School of Computer Science at Carnegie Mellon University, 1999

[20] Gu M G, Kong F H. A stochastic approximation algorthm with Markov Chain Monte Carlo Method for incomplete data estimation problems [A]. Proceedings of National Academy of Sciences [C], 1998, 95: 7270～7274

[21] Peter J Burt, Edward II. Adelson. The Laplacian pyramid as a compact image code [J]. IEEE Transactions on Communications, 1983, COM－31 (4): 532～540

[22] Beier T, Neely S. Feature-based image metamorphosis [J]. Computer Graphics, 1992, 26(2): 35～42

[23] Richard O Duda, Peter E Hart, David G Stork. Pattern Classification, Second Edition [M]. Wiley-Interscience, 2000

[24] Tom M Mitchell. Machine Learning [M]. USA: MCGRWW-HILL, 1997

[25] 金勇进，蒋 妍，李序颖. 抽样技术 [M]. 北京：中国人民大学出版社，2002

[26] Seize S M. View morphing [A]. Computer Graphics Annual Conference Series, SIGGRAPH'96 [C], ACM SIGGRAPH, New Orleans, Louisiana, August, 1996. 21～30

[27] Wolberg G. Image morphing survey [J]. The Visual Computer, 1999, 14: 360～372

Improved Matching Algorithms for Linear Face Class Model

Fu Yun, Zheng Nanning

(*Institute of Artificial Intelligence and Robotics, Xi'an Jiaotong University, Xi'an* 710049, *China*)

[**Abstract**] An advanced matching technique for linear face class model is proposed, which can solve the problem of detailed controlling and robust iteration for the realistic facial modeling. A new method——Dynamic Gaussian Pyramid Analysis (DGPA), which combines Non-Uniform Sampling (NUS) method and Multi-Resolution Analysis, is presented. Integrating the PS Sampling and the Cluster Random Sampling, the distribution of the sampled points in each level images of the Gaussian pyramid is adjusted dynamically. In coarse-to-fine scheme, the minimization algorithm is used to compute the near global optimal solution that may fit to yield accurate model matching. Dynamic adjusting the boundary of the sampling cluster area and the re-sampling ratio, the detailed representations are effectively controlled, and the model creation is quite robust. An improved Stochastic Gradient Descent (SGD) algorithm based on the Correlative Disturbance(CD) and Adaptive Learning Rate (ALR) is exploited to accelerate iteration convergence and compute valid model parameters. With the examples of MPI Caucasian Face and AI&R Asian Face databases, experimental results in subjective evaluation and objective evaluation demonstrate the advanced model matching technique.

[**Key words**] facial modeling; model matching; stochastic gradient descent; non-uniform sampling; multi-resolution analysis