

基于组稀疏表示的在线单帧图像超分辨率算法

李键红¹ 吴亚榕² 吕巨建³

(广东外语外贸大学语言工程与计算实验室 广州 510006)¹

(仲恺农业工程学院科学技术处 广州 510225)² (广东技术师范学院计算机学院 广州 510665)³

摘要 基于稀疏表示的图像超分辨率重建算法以近似随机抽取的方式选取字典中的原子来拟合图像片,而实际中的字典原子的选择体现出了很强的结构稀疏性,从而导致算法计算复杂且引入了大量的误差,影响重建图像的质量。针对该问题,提出了一种基于组稀疏表示的在线图像超分辨率重建算法。该方法引入组稀疏理论,仅利用输入的低分辨率图像作为样本来构建组稀疏字典,通过结合组稀疏性和几何对偶性来构建超分辨率图像算法的成本函数,并使用提出的一种迭代的方法进行求解。实验表明,该算法在视觉观察和参数比较上都优于当前主流的超分辨率算法。

关键词 组稀疏表示,单帧图像超分辨率,正交匹配追踪,字典学习,迭代优化

中图分类号 TP391.41 文献标识码 A DOI 10.11896/j.issn.1002-137X.2018.04.053

Online Single Image Super-resolution Algorithm Based on Group Sparse Representation

LI Jian-hong¹ WU Ya-rong² LV Ju-jian³

(Laboratory of Language Engineering and Computing,Guangdong University of Foreign Studies,Guangzhou 510006,China)¹

(Department of Science and Technology,Zhongkai University of Agriculture and Technology,Guangzhou 510225,China)²

(School of Computer Science,Guangdong Polytechnic Normal University,Guangzhou 510665,China)³

Abstract The super-resolution algorithm based on sparse representation selects atoms in the dictionary by approximately random style to fit the specified image patch, but the selected atoms show strong structure sparsity in practice, which leads to the complexity of calculation and a great deal of errors, affecting the quality of the reconstructed images. For conquering the problem, a online single image super-resolution algorithm based on group sparse representation was proposed. The proposed algorithm takes advantage of the inputted low-resolution image to construct the group sparse dictionary by introducing the group sparse theory, and then incorporates the group sparse prior and geometric duality prior to design the cost function of the algorithm, which is solved by a proposed iterative optimization method. The experiments demonstrate that the proposed algorithm is superior to the main stream algorithms subjectively and objectively.

Keywords Group sparse representation, Single image super-resolution, Orthogonal matching pursuit, Dictionary learning, Iterative optimization

1 引言

图像超分辨率指的是基于已有的一幅或多幅同一场景的低分辨率图像,通过软件计算的方式获取其对应的高分辨率图像的方法。其所估计的高分辨率图像不仅增加了像素点,同时比原始的低分辨率图像更加清晰,能展现原低分辨率图像中没有的细节,因而备受人们的追捧。单帧图像的超分辨率在实际问题中更为普遍,因此本文主要研究单帧图像的超分辨率算法。

主流的单帧图像超分辨率算法大致可分为 3 类:基于差值的方法、基于重建的方法和基于学习的方法。

基于差值的方法通常假定图像在局部结构上是分片平滑的^[1]。从这一假设出发,这类方法通过使用一个插值核或基函数来逼近高分辨率网格中的像素点,使得离散的数字图像

变为连续的曲面,从而估计出指定位置未知点的像素值。这类方法无须对图像的内容做任何判定,因此算法简单、高效,能够实现实时处理。如当前的 Windows 操作系统和 Photoshop 软件就是采用这类方法中的双三次插值作为默认的图像放大方法。这类方法的缺点在于不能恢复图像在降质过程中丢失的高频信息,在图像的边缘位置容易引起模糊,影响图像观感。

图像超分辨率重建问题是一个严重的病态,通常在求解超分辨率重建的方程中,方程的解并不唯一。基于重建的算法要求得到的高分辨率图像在通过图像降质模型的滤波和下采样处理后,与输入的低分辨率图像尽可能相似;同时要求得到的高分辨率图像满足指定的先验知识。先验知识的作用在于压缩超分辨率重建方程的解空间,使得到的解更加精确。Shan 等^[2]指出同一场景的高分辨率图像和低分辨率图像的

到稿日期:2017-05-11 返修日期:2017-08-14 本文受广东省科技计划项目(2016A020210131,2016A070712020),语言工程与计算实验室项目(LEC2016ZBK T004)资助。

李键红(1981—),男,博士,助理研究员,主要研究方向为图像处理、语音处理,E-mail:lijianhonghappy@163.com;吴亚榕(1982—),女,硕士,助理研究员,主要研究方向为图像处理;吕巨建(1983—),男,博士,讲师,主要研究方向为机器学习。

梯度直方图基本稳定,进而用曲线拟合低分辨率图像梯度直方图作为高分辨率图像的约束条件。由于直方图只代表图像的像素值分布信息,对图像的描述并不充分,因此在他们的结果中边缘和纹理等高频位置易出现噪声和伪影。另外,基于边缘特征的先验相继被提出,如边先验^[3]、相似性冗余先验^[4]、梯度轮廓先验^[5]等。这一类先验知识通过统计高分辨率图像的边缘而来,因此对于边缘位置有很好的重建效果,能有效抑制边缘附近的噪声和伪影,但对高频纹理和细节等位置的处理仍不尽如人意。基于重建的方法得到了非常广泛的研究,很多学者提出了有效的算法。但当放大倍数增大时,重建约束所能提供的有效信息越来越少,致使这类方法恢复高频的能力受到了很大的制约。

基于学习的超分辨率算法通过外界高分辨率图像来建立学习集,并利用学习集探索低分辨率图像和高分辨率图像的映射关系^[6-13],使得先验知识以一种隐含的方式包含在这一映射中,然后利用这种映射关系来求解对应的高分辨率图像。Yang等^[7-9]提出了稀疏表示的超分辨率算法。稀疏表示指的是图像片在指定的过完备字典下,能由字典中尽可能少的图像片来加权表示。过完备字典的训练对于稀疏表示的性能起着至关重要的作用。基于稀疏表示的超分辨率重建算法将训练集合中的高分辨率图像片和低分辨率图像片联合训练,得到两个对应不同分辨率的字典。对于输入的低分辨率图像片,求解其在对应的低分辨率字典下的稀疏系数,然后将稀疏系数传递到高分辨率字典中,通过相乘的方式得到对应的高分辨率图像片。在求解稀疏系数的过程中,这种方法以一种近似随机的方式来选择字典中的原子,这种方式往往会使用过完备字典中与输入图像片差异很大的原子参与到拟合中来,从而引入拟合误差,使重建结果中显著边缘附近存在可观察到的噪声,降低了重建图像的质量。总体来说,这类方法对图像的高频恢复展现出了巨大潜力,得到了研究人员的极大关注,但其对外界的图像样本集合过于依赖,随着算法的日益复杂,对所需样本的数量也呈现出指数级的增长趋势,这对普通用户而言是一个极大的挑战。同时,人们对样本图像本身的研究少之又少,只是通过大量堆叠来扩充样本容量。对于一幅指定的待放大的图像而言,样本中哪一幅图像起着怎样的作用根本无从得知,只知道若其与待求的高分辨率图像的内容越相近,它所能提供的正确的高频越多;就图像的内容而言,输入图像本身就与其对应的高分辨率图像的内容最为接近。因此,一种特殊的样本学习算法即基于自相似性质的超分辨率算法日益引起了人们的关注^[14-16]。图像的自相似性质是指对于一幅图像,当从小的局部即图像片进行考查时,对于每一个小图像片,在这幅图像自身的其他位置或在这幅图像的其他尺度中都极易找到与之相似的图像片。基于自相似性质的超分辨率算法是一类特殊的基于样本学习的超分辨率算法,它是利用图像的自相似性质把输入图像自身作为训练本来求取对应的高分辨率图像的一种算法。

近年来,不同类别的超分辨率算法表现出了相互融合的趋势,综合几类算法的特点,实现优势互补是图像超分辨率重建的又一发展趋势。

2 图像的稀疏与组稀疏表示

在现有文献中^[7-12],图像的稀疏表示是通过逐个处理从

图像中抽取的图像片这一方式实现的。给定用向量形式表示的图像 $X \in R^N$ 和在图像 X 中任意抽取的尺寸为 $a \times a$ 的图像片 $x_k \in R^{a^2}$ 。其中 N 和 a^2 表示图像和图像片中拥有像素点的个数; k 代表图像片 x_k 在图像 X 中的位置,取值范围为 $k \in [1, \dots, n]$; n 表示在 X 中抽取的图像片的总数量。 x_k 的抽取操作可表示为:

$$x_k = R_k(X) \quad (1)$$

其中, $R_k(X)$ 表示在图像 X 中抽取图像片 x_k , $R_k^T x_k$ 能够将图像片 x_k 放置到 X 的重构图像的 k 位置处,图像片以外的位置为 0。通常图像片抽取的方式是有重叠的,因此从 $\{x_k\}_{k=1}^n$ 重构 X 可记作:

$$X = \sum_{k=1}^n R_k^T(x_k) \cdot \sum_{k=1}^n R_k^T(I_{a^2}) \quad (2)$$

其中, $\cdot /$ 表示两个向量点对点地做除法操作,其结果仍然为一个向量; I_{a^2} 表示一个维度为 a^2 、元素全为 1 的向量。式(2)在重构 X 的过程中在图像片的重叠区域选择取平均值的策略,在实际操作过程中也可以用其他策略替代。

对于一个给定的字典 $D \in R^{a^2 \times M}$ (M 是 D 中列向量的个数,即原子的个数, $M > a^2$), 每一个图像片 x_k 在字典 D 下的稀疏编码过程指的是:通过找到一个稀疏向量 $\alpha_k \in R^M$ (稀疏指的是 α_k 中绝大多数系数是 0,或接近于 0),使得 $x_k \approx D\alpha_k$,那么整幅图像就可以用稀疏编码系数集合 $\{\alpha_k\}_{k=1}^n$ 进行表示。在实际应用中,图像片 x_k 在字典 D 下的稀疏编码问题通常可以表示为:

$$\arg \min_{\alpha_k} \frac{1}{2} \|x_k - D\alpha_k\|_2^2 + \lambda \|\alpha_k\|_p \quad (3)$$

其中, λ 是平衡第一项(保真项)和第二项(稀疏项)的权重系数。 p 通常取值为 0 或 1,当 $p=0$ 时表示稀疏项是求取 α_k 的 l_0 范数, α_k 的 l_0 范数指的是计算 α_k 中非 0 元素的个数。然而 α_k 的 l_0 范数问题通常是非凸且 NP 难问题,一般使用贪婪算法求解近似解,如正交匹配追踪(Orthogonal Matching Pursuit, OMP)算法^[17],或用 l_1 范数来近似 l_0 范数。 l_1 范数问题能够通过近年来发展的大尺度工具快速求解。

与式(2)相似,由稀疏编码系数 $\{\alpha_k\}_{k=1}^n$ 重构 X 可以表示为:

$$X = D \circ \alpha = \sum_{k=1}^n R_k^T(D\alpha_k) \cdot \sum_{k=1}^n R_k^T(I_{a^2}) \quad (4)$$

其中, $\alpha = [\alpha_1^T, \alpha_2^T, \dots, \alpha_n^T]^T$ 是由所有稀疏系数向量链接而成的向量。引入 \circ 符号的目的是利用 $D \circ \alpha$ 使得表达式 $\sum_{k=1}^n R_k^T(D\alpha_k) \cdot \sum_{k=1}^n R_k^T(I_{a^2})$ 更加简洁。

在图像的降质模型中,观察图像 Y 可以表示为 $Y = AX + n$ 。其中 X 是降质前的清晰图像; n 通常假定为高斯白噪声; A 可以根据降质的形式设定,若 A 为单位矩阵时,对应的是图像的去噪模型, A 为滤波下采样矩阵时对应的是图像的超分辨率模型。在稀疏先验下使用图像片系数表示模型来恢复 X 的方案可以构造为:

$$\hat{\alpha} = \arg \min_{\alpha} \frac{1}{2} \|y - A D \circ \alpha\|_2^2 + \lambda \|\alpha\|_p \quad (5)$$

其中, λ 是两项间的权重系数, p 取 0 或 1。在求解出 $\hat{\alpha}$ 后,重构的图像 X 可以表示为 $\hat{X} = D \hat{\alpha}$ 。

稀疏表示最核心的任务是字典 D 的选取,即怎样确定一个最适合的空间来稀疏化一幅给定的图像。大多数稀疏表示

领域的工作都集中在给定图像片训练集样本 $\{x_i\}_{i=1}^n$, 如何从样本集合中学习冗余的字典 D . 本文要求学习出的字典 D 既能通过稀疏系数矩阵 $\Lambda = [\alpha_1, \alpha_2, \dots, \alpha_n]$ 重构图像 X , 即对每一个图像片 x_k , 都有 $x_k \approx D\alpha_k$, 又能使得每一个系数向量具备稀疏的性质, 即 $\|\alpha_k\|_p \leq L$. 这一目标可以通过构造下列的最小化方程实现.

$$\begin{aligned} (\hat{D}, \hat{\Lambda}) &= \arg \min_{D, \Lambda} \sum_{k=1}^n \|x_k - D\alpha_k\|_2^2 \\ \text{s. t. } &\|\alpha_k\|_p \leq L, \forall k \end{aligned} \quad (6)$$

由于式(6)中的最小化问题规模庞大且非凸, 因此对其直接求解变得极为困难. 近年来, 一些近似求解的方法(如 MOD, K-SVD^[18])相继被提出, 这些方法通过交替最优求解 D 和 Λ , 在图像处理领域取得了目前最优秀的结果. 然而这些方法都不可避免地面临时间复杂度过高这一问题. 另外, 通过分析式(3)和式(6)不难发现, 每一个图像片在稀疏编码的过程中都是被独立考虑的, 从而忽略了相似图像片间的关系, 导致了稀疏编码系数的不准确.

稀疏表示通常假定稀疏系数向量中非零元素出现的位置是随机的. 然而, 稀疏系数通常以聚类形式呈现出内在的结构. 因此, 近年来人们提出了组稀疏表示的概念. 组稀疏表示假定图像片能够被字典中少数的几组原子线性表示, 即稀疏系数向量中的非零元素只出现在字典中少数的几组里. 假定字典中有 m 个原子, 序号集合为 $\{1, 2, \dots, m\}$, $G = \{G_1, G_2, \dots, G_g\}$ 是序号集合的一种划分, g 是该组划分中分组的数量. 给定字典 $D \in R^{a^2 \times m}$, D_{G_i} 是字典中的第 i 组子集合, 那么图像片 $x_k \in R^{a^2}$ 能够表示为:

$$x_k = D\alpha_k = \underbrace{[D_{G_1}, D_{G_2}, \dots, D_{G_g}]}_D \underbrace{[\alpha_{kG_1}^T, \alpha_{kG_2}^T, \dots, \alpha_{kG_g}^T]}_{\alpha_k} \quad (7)$$

其中, T 是向量或矩阵的转置操作, $\alpha_k \in R^m$ 是图像片 $\|x_k\|_{2,1} = \sum_{i=1}^g \|\alpha_k G_i\|_2$ 的组稀疏系数向量. 在组稀疏的约束作用下, 求解 α_k 的方程可以表示为:

$$\arg \min_{\alpha_k} \|\alpha_k\|_{2,0}, \text{ s. t. } x_k = D\alpha_k \quad (8)$$

其中, $\|\alpha_k\|_{2,0} = \sum_{i=1}^g b(\|\alpha_{G_i}\|_2)$. 函数 $b(\cdot)$ 是一个二值函数, 定义为:

$$b(\|\alpha_{G_i}\|_2) = \begin{cases} 1, & \|\alpha_{G_i}\|_2 \geq 0 \\ 0, & \text{otherwise} \end{cases} \quad (9)$$

在图像片 x_k 包含噪声的情况下, x_k 可以稀疏表示为 $x_k = D\alpha_k + \nu$, ν 代表高斯白噪声. 那么组稀疏系数相应地改写为:

$$\arg \min_{\alpha_k} \|\alpha_k\|_{2,0}, \text{ s. t. } \|x_k - D\alpha_k\|_2^2 \leq \epsilon \quad (10)$$

其中, $\epsilon > 0$ 是一个很小的固定常数, 用来保证 x_k 和 $D\alpha_k$ 间的差异不会过大. 然而, 式(10)仍然是非凸问题, 人们通常用 $l_{2,1}$ 范数来近似替代 $l_{2,0}$, 使问题凸化为:

$$\arg \min_{\alpha_k} \|\alpha_k\|_{2,1}, \text{ s. t. } \|x_k - D\alpha_k\|_2^2 \leq \epsilon \quad (11)$$

其中, $l_{2,1}$ 范数定义为 $\|\alpha_k\|_{2,1} = \sum_{i=1}^g \|\alpha_k G_i\|_2$. 式(11)可以进一步转换为:

$$\arg \min_{\alpha_k} \frac{1}{2} \|x_k - D\alpha_k\|_2^2 + \lambda \|\alpha_k\|_{2,1} \quad (12)$$

式(12)可以通过组正交匹配追踪(Group Orthogonal Matching Pursuit, GOMP)算法^[19]求解, 该算法的本质是贪婪算法, 它由正交匹配追踪算法发展而来, 但与正交匹配追踪算

法每次选取一个原子参与计算不同, 组正交匹配追踪算法每次选取最相关的一组原子参与计算.

另外, 基于边缘导向的图像插值算法(New Edge Directed Interpolation, NEDI)^[21]是一种重要的图像超分辨率重建算法. 它利用图像的几何对偶性质计算待插入像素点邻域的加权值, 通过加权求和的方法计算待插入像素点的亮度值. 几何对偶性质指的是不同分辨率间拥有相同方向走向的像素点协方差间的对应关系. NEDI 算法^[21]将以待插入像素点为中心的 $M \times M$ 小窗口中的低分辨率像素点排列成列向量 y , 然后构造矩阵 C , C 中的第 i 个行向量是 y 中第 i 个元素的四邻域, 那么低分辨率图像的协方差所构成的向量 R 和 r 可以表示为:

$$R = \frac{1}{M^2} C^T C, r = \frac{1}{M^2} C^T y \quad (13)$$

进而待插入点四邻域的加权系数向量可以表示为:

$$\alpha = R^{-1} r = (C^T C)^{-1} C^T y \quad (14)$$

分析式(14)可以发现, 向量 y 和其对应的邻域矩阵 C 满足关系 $C\alpha = y$, 即小窗口中的所有低分辨率像素点和它们各自对应的邻域有着相同的加权关系. 我们将这一关系进一步拓展到重建图像的所有像素点中, 可以得到:

$$X = WX \quad (15)$$

其中, X 是以列向量形式表示的重建图像, W 是一个稀疏的权重系数矩阵. 第 i 行中非零元素出现的位置对应着 X 中第 i 个像素点在二维图像中邻域的位置, 非零元素的值对应的是该像素点邻域的权重.

3 组稀疏框架下的超分辨率

对于图像的超分辨率重建问题, 由高分辨率图像生成低分辨率图像模型可以表示为: $Y = SHX + n$, 其中 X 和 Y 分别表示原始的高分辨率图像和其对应生成的低分辨率图像. 为了避免符号重复, 这里用 S 和 H 分别表示生成过程中的下采样和滤波操作; n 表示生成过程中引入的 0 均值加性高斯白噪声. 超分辨率问题是低分辨率图像生成过程的逆过程, 它的目标是根据给定的低分辨率图像 Y 来估计原始的超分辨率图像 X . 由于低分辨率图像的生成过程受到下采样、滤波和噪声等影响, 一幅低分辨率图像可能对应着多幅不同的高分辨率图像, 因此图像的超分辨率问题是一个典型的逆问题. 为了求解该逆问题, 需要引入一些关于高分辨率图像的先验知识. 在引入先验知识后, 图像超分辨率问题的最大后验估计(Maximum A Posterior estimation, MAP)^[13]可以表示为:

$$X = \arg \min_X \left\{ \|Y - SHX\|_2^2 + \sum_{i=1}^{N_p} \lambda_i r_i(X) \right\} \quad (16)$$

其中, λ_i 是第 i 个先验约束项的权重系数, N_p 是算法中引入先验约束项的数量, $r_i(X)$ 是引入的第 i 项先验约束项. 首先假定重构图像中的图像片满足组稀疏的性质, 它对应的约束项可以写成:

$$r_1(X) = \sum_{i=1}^N \|R_i X - D\alpha_i\|_2^2, \text{ s. t. } \|\alpha_i\|_{2,0} \leq \epsilon \quad (17)$$

然后引入几何对偶性质, 作为另一个重构图像的先验知识, 它对应的约束项可以写成:

$$r_2(X) = \|(I - W)X\|_2^2 \quad (18)$$

其中, I 为单位矩阵, W 为权重系数矩阵.

最后, 给出本文的基于组稀疏表示的图像超分辨率重建

算法的成本函数的完整形式:

$$\begin{aligned} \arg \min_X & \|Y - SHX\|_2^2 + \lambda \sum_{i=1}^N \|R_i X - D\alpha_i\|_2^2 + \\ & \beta \| (I - W) X \|_2^2 \\ \text{s. t. } & \|\alpha_i\|_{2,0} < \varepsilon, \forall i \end{aligned} \quad (19)$$

3.1 字典学习

字典学习对基于组稀疏表示的图像超分辨率算法的性能有极为重要的作用。从理想的重构高分辨率图像中抽取的图像片和从低分辨率输入图像中抽取的图像片能够在同一个字典下进行稀疏表示,根据图像自相似的性质可知,这一假设是成立的。同时,因为高分辨率图像和它对应的低分辨率图像有着极高的相似度,为了避免对外界数据的依赖,我们只用输入图像来构建字典的训练数据集合,无需通过收集外界的高分辨率图像来建立庞大的训练集合,从而省去了复杂耗时的字典训练步骤。又因为输入图像和它的高分辨率图像的相似度极高,所以用输入图像建立的字典质量是有保证的。另外,图像除了有极强的自相似性质外,从小的局部观察可知,它还存在大量的对称、旋转和镜像等结构。如图 1 所示,在蝴蝶的图像中用不同的颜色框标出了对称结构的图像片,其实,蝴蝶身体的左右两部分几乎是完全对称的。那么如何利用这些信息来提高图像超分辨率重建的质量呢?

为了能够充分利用输入的低分辨率图像自身的一些结构



图 1 包含对称结构的图像“Butterfly”

Fig. 1 “Butterfly” image including symmetrical structures

图 2 “闹钟”图像通过旋转、镜像等操作构造的训练集合

Fig. 2 Training set of image “alarm clock” constructed by rotation, mirror and other operations

在构造完成训练样本图像集合后,在这些图像中抽取尺寸为 $a \times a$ 的图像片。设定阈值 σ , 通过比较方差的方式剔除掉内容平滑、信息量较少的图像片,最后建立图像片的训练集合,并以矩阵的形式表示为: $Z = [z_1, z_2, \dots, z_N] \in R^{a^2 \times N}$, 其中 z_i 是以列向量的形式表示的第 i 个图像片, N 为图像片集合中样本的总数量。整个过程的目标是训练一个字典 $D = [d_1, d_2, \dots, d_m] \in R^{a^2 \times m}$ 和一个稀疏系数矩阵 $\Delta = [\alpha_1, \alpha_2, \dots, \alpha_N]$, 使得训练图像片矩阵 Z 能够在字典 D 下由稀疏系数矩阵 Δ 表示, 即 $Z = D\Delta$, 同时稀疏系数矩阵 Δ 的每一个列向量 α_i 满足组稀疏的性质。

首先,定义一个二值矩阵 $M \in R^{N \times N}$ 来标记字典内原子间的关系,其中第 i 行,第 j 列元素 m_{ij} 的值定义为:

$$m_{ij} = \begin{cases} 1, & \|d_i - d_j\| \leq \theta \\ 0, & \text{otherwise} \end{cases} \quad (20)$$

其中, θ 为用户指定的阈值。当字典中的原子 d_i 和 d_j 的内容相似且距离很小时,就会把它们分配到同一组,并用 $m_{ij} = 1$ 来记录;反之,当字典中的原子 d_i 和 d_j 的差异很大时,原子 d_i 和 d_j 未分配在同一组内,用 $m_{ij} = 0$ 来标记。显然,矩阵 M 的主要意义在于记录字典 D 的分组关系。在得到矩阵 M 后,可以做如下推导:

信息,对输入的低分辨率图像进行旋转、镜像等操作,并把得到的结果也作为图像的训练样本图像。另外,Zontak 等^[16]指出图像在单独进行水平或垂直方向的下采样操作后,得到的结果仍然有着很强的自相似性质,即输入图像中的图像片在输入图像水平或垂直方向下采样的结果中仍然存在着大量的重现。因此本文将训练集合中的图像做水平和垂直的下采样操作,并把得到的结果同样并入到训练集合中。图 2 是由一幅闹钟图像构成的训练图像集合。为了操作简单,同时避免引入过多的噪声,输入图像的旋转只考虑 $90^\circ, 180^\circ$ 和 270° , 其他角度的旋转不做考虑。



图 1 包含对称结构的图像“Butterfly”

Fig. 1 “Butterfly” image including symmetrical structures

$$\begin{aligned} & \frac{1}{2} \sum_{i=1}^m \sum_{j=1}^m m_{ij} \|d_i - d_j\|_2^2 \\ & = \sum_{i=1}^m d_i^T m_{ii} d_i - \sum_{i=1}^m \sum_{j=1}^m d_i^T m_{ij} d_j \\ & = \text{tr}(DCD^T) - \text{tr}(DMD^T) \\ & = \text{tr}(DLD^T) \end{aligned} \quad (21)$$

其中, C 是一个对角矩阵, 对角线上第 i 个元素 $c_{ii} = \sum_{j=1}^m m_{ij}$, $L = C - M$ 是一个拉普拉斯矩阵, $\text{tr}(\cdot)$ 是对矩阵的迹操作。式(21)用于计算字典中组内原子差异的总和, 当其取得最小值时,可以保证字典中同一组内的原子尽可能相似,从而使得组间的差异尽可能地被拉开。这一点对于组稀疏编码中字典的学习是极为重要的。本文将式(21)中的 $\text{tr}(DLD^T)$ 和组稀疏的思想相结合,提出一种新的字典学习的方法,即:

$$\begin{aligned} \arg \min_{D, \{\alpha_i\}_{i=1}^N} & \sum_{i=1}^N \|z_i - D\alpha_i\|_2^2 + \lambda \text{tr}(DLD^T) \\ \text{s. t. } & \|\alpha_i\|_{2,0} \leq \tau, \forall i \end{aligned} \quad (22)$$

其中, $\sum_{i=1}^N \|z_i - D\alpha_i\|_2^2$ 用于保证训练图像片集合中所有的图像片 $\{z_i\}_{i=1}^N$ 都能在字典 D 下由对应的系数 $\{\alpha_i\}_{i=1}^N$ 表示; $\text{tr}(DLD^T)$ 能够保证字典中的原子在组内尽可能相似,组间差异尽可能大; λ 是其对应的权重系数; $\|\alpha_i\|_{2,0}$ 用来保证系数

向量具有组稀疏的性质; τ 是一个小的正数,用来控制组稀疏的程度。将式(22)写成矩阵的形式:

$$\begin{aligned} \arg \min_{D, A} \| Z - DA \|_F^2 + \lambda \text{tr}(DLD^T) \\ \text{s. t. } \| \alpha_i \|_{2,0} \leq \tau, \forall i \end{aligned} \quad (23)$$

其中, $Z = [z_1, z_2, \dots, z_N]$, $\Delta = [\alpha_1, \alpha_2, \dots, \alpha_N]$ 。

对于式(23)的求解,本文采用迭代的策略,分为组稀疏编码和字典更新两个步骤。在组稀疏编码阶段,固定字典 D 不变,组稀疏编码系数矩阵的求解公式为:

$$\arg \min_A \| Z - DA \|_F^2, \text{ s. t. } \| \alpha_i \|_{2,0} \leq \tau \quad (24)$$

采用组正交匹配追踪算法来求解式(24)。在字典更新步骤,固定组稀疏编码系数矩阵,更新字典 D 的问题写成:

$$\arg \min_A \| Z - DA \|_F^2 + \lambda \text{tr}(DLD^T) \quad (25)$$

对于式(25)的求解,我们对 D 求导,并令导数为 0,可以得到 D 的求解表达式为:

$$D = ZA^T(\Delta\Delta^T + \lambda I)^{-1} \quad (26)$$

3.2 迭代求解

在求解出字典后,开始求解算法的成本函数,即式(19)。直接求解式(19)并不容易,因为 $\| Y - SHX \|_2^2$ 和 $\| (I - W)X \|_2^2$ 两项是以像素点为单位进行计算的,而 $\sum_{i=1}^N \| R_i X - D\alpha_i \|_2^2$ 是以图像片为单位进行计算的。为了使问题能够求解,采用两步迭代策略。首先将 X 初始化为输入图像 Y 的双三次上采样版本 X_b ,然后固定 X 不变,求解组稀疏系数集合 $\{\alpha_i\}_{i=1}^N$,此时的方程变为:

$$\hat{\alpha}_i = \arg \min_{\alpha_i} \| R_i X - D\alpha_i \|, \text{ s. t. } \| \alpha_i \| \leq \epsilon \quad (27)$$

式(27)可以转化为式(10)的形式,用正交匹配追踪算法进行求解。在得到 $\{\alpha_i\}_{i=1}^N$ 后,用式(4)重构出 X 的估计版本,记作 \tilde{X} 。在第二步中,先用 \tilde{X} 替换掉成本函数中的组稀疏系数集合 $\{\alpha_i\}_{i=1}^N$,然后由式(14)计算、构造并更新 \tilde{X} 的权重系数矩阵 W ,使得成本函数转化为:

$$\arg \min_X \| Y - SHX \|_2^2 + \lambda \| X - \tilde{X} \|_2^2 + \beta \| (I - W)X \|_2^2 \quad (28)$$

式(28)可直接采用梯度下降法来求解,其中第 $(t+1)$ 次的求解结果为:

$$X^{t+1} = X^t + \tau \{ H^T S^T (Y - DHX^t) - \lambda (X^t - \tilde{X}) - \beta (I - W)^T (I - W) X^t \} \quad (29)$$

字典学习的伪代码如算法 1 所示。

算法 1 在线组稀疏表示的超分辨率字典学习

Input: small image Y , group sparsity level τ , iterative number $Iter$, parameter λ , group partition $G = \{G_1, G_2, \dots, G_g\}$

Output: Dictionary D

1. Construct the training set using Y ;
2. Formulate the matrix W according to G ;
3. Calculate the laplacian matrix L ;
4. for $i = 1: Iter$ do
5. Update the group sparse coefficients matrix;
6. Update the dictionary;
7. end for
8. return D

4 实验与分析

为了证明组稀疏框架下图像的超分辨率重建算法的有效性,本文将在一些指定的数据集上进行实验,并与一些经典、

前沿的算法作比较,以 PSNR 和 SSIM 为客观评价参数来评价重建图像的质量。

4.1 实验设置

在该算法的实验中,首先对高分辨率图像进行滤波、下采样操作,滤波器选择尺寸为 7×7 、标准差 $\sigma = 1.6$ 的高斯低通滤波器,下采样的倍数为 s (s 设定为与超分辨率放大的目标倍数相同即可),然后把得到的结果作为输入的低分辨率图像,用于算法的处理。对于彩色图像,首先将其转换到 YUV 空间;然后对人眼较为敏感的 Y 通道采用相对复杂的超分辨率算法,对人眼不敏感的颜色通道 U 和 V 直接采用高效的双三次上采样的办法放大到目标倍数;最后,合并 3 个通道的结果,并将其转换到 RGB 空间,作为输入彩色图像的最终结果。对于需要放大的低分辨率图像,先对这幅图像进行旋转、镜像、水平下采样 2 倍、垂直下采样 2 倍等操作来构造组稀疏字典的训练样本集合,下采样的方法选择双三次插值。然后有重叠地将这些图像分割成 7×7 的图像片,相邻两个片有 5 行像素重叠,并设定阈值为 $\sigma = 10^{-5}$,通过与这些图像片的方差做比较,滤除掉过于平滑、信息量较少的图像片,将剩余的图像片作为组稀疏字典的训练样本集合。在训练组稀疏字典的过程中,首先假定各组中的原子数量是相同的,每组子字典中的原子个数设定为 5;然后设定算法的迭代次数为 40,设定式(19)中的权重系数 $\alpha = 0.5, \beta = 0.3$,设定式(22)中的控制组稀疏程度的参数 $\tau = 3$ 。对于要求放大结果超过 2 倍的图像,先对其放大 2 倍,然后把放大 2 倍的结果作为输入图像并用该算法对其做进一步放大,直到达到目标倍数。整个算法的流程如图 3 所示。

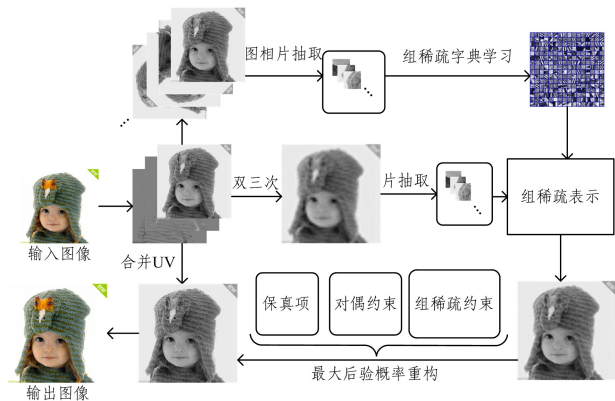


图 3 基于组稀疏表示的图像超分辨率重建算法的流程

Fig. 3 Flowchart of image super-resolution algorithm based on group sparse representation

4.2 实验结果与分析

本节先对组稀疏方法建立字典的性能进行讨论。为了验证这种方法建立的字典的性能,以经典的 512×512 的“lena”图像为例进行字典训练。初始字典采用离散余弦变换基生成,训练字典包括 49×100 和 49×200 两种大小,两种字典中的分组分别为 20 组和 40 组,即默认每组子字典中包含 5 个原子。采用均方根误差(Root Mean Square Error, RMSE)作为评价指标,并用经典的字典学习方法 K-SVD^[18]与之做比较。如图 4 所示,图 4(a)是尺寸为 49×100 的字典的比较结果,图 4(b)是尺寸为 49×200 的字典的比较结果。从中可以看出,本文方法相较于 K-SVD 算法^[18]具有更快的收敛速度,且对数据有更准确的描述能力。

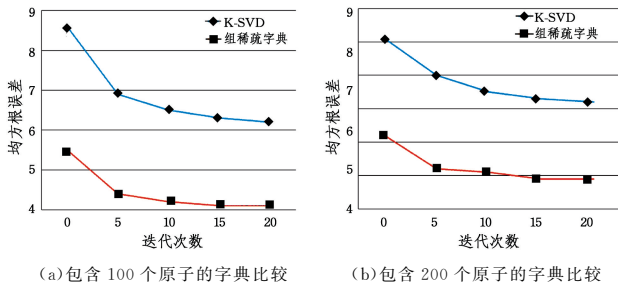


图 4 包含 100 及 200 个原子的字典性能比较

Fig. 4 Performance comparison between two dictionaries including 100 atoms and 200 atoms

从该算法的成本函数可知,基于组稀疏表示的图像超分辨率重建算法同时使用了组稀疏约束和几何对偶性约束。为了检验这两个约束项,首先设定 $\beta=0$,以检测组稀疏项对超分辨率图像的放大能力,继而再重新设定 $\lambda=0$,以检测几何对偶性先验的约束能力。根据前文的说明,这两种方式都容易求解,此处不再说明。本文用图像处理领域常见的 6 幅图像(如图 5 所示)来完成此实验,其中放大倍数为 2。

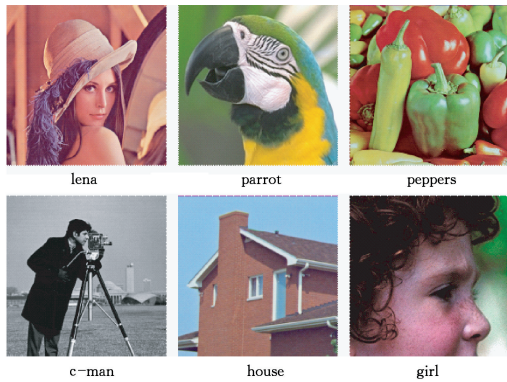


图 5 用于测试约束项超分辨率性能的 6 幅图像

Fig. 5 Six images for testing super-resolution performance of constrained term

实验中 PSNR 和 SSIM 结果如表 1 所列,其中每个单元格中上面的值为 PSNR,下面的值为 SSIM。从表 1 可知,分别使用两种约束得到的结果仍好于双三次算法。因为该算法迭代过程的初值选定为双三次的结果,所以这两种方法都能够对图像的求解起到积极的作用。同时使用这两个先验约束项后,图像的质量得到进一步提升,说明这两个先验约束项之间并不矛盾,能够共同提高超分辨率图像的质量。

表 1 不同约束项放大 2 倍的结果比较

Table 1 Comparison of 2X image results with different constrained terms sparately

图像	lena	parrot	peppers	c-man	house	girl
双三次	32.11	23.07	34.26	24.78	25.04	31.37
	0.805	0.512	0.893	0.756	0.687	0.851
组稀疏	32.32	23.23	34.67	25.06	25.23	31.92
	0.883	0.569	0.916	0.783	0.702	0.911
对偶性	32.91	23.26	34.57	24.96	25.27	31.50
	0.881	0.558	0.914	0.770	0.703	0.892
两者同时	32.64	23.58	36.13	25.23	25.65	31.50
	0.897	0.603	0.928	0.788	0.730	0.916

本文的组稀疏超分辨率重建算法在组稀疏理论的基础上利用输入图像构造训练集合,以在线的方式训练字典。Yang

等^[7-9]通过稀疏表示重建超分辨率图像(Sparse coding Super-Resolution, ScSR)的方法同样建立了字典的步骤,只是它以大量外界数据集合作为训练集合,以一种更复杂的方式来完成。以“lena”图像为例,分别用 ScSR^[9]和提出的组稀疏方法将“lena”图像放大 3 倍,如图 6(a)、图 6(b)所示。观察发现,ScSR^[9]的结果较组稀疏的结果更模糊;同时从框出的对比显著的帽子边缘和纹理丰富的头发两个小局部发现,在 ScSR^[9]的结果中,这些人眼敏感的位置都出现了视觉能够观察到的噪声,而在组稀疏算法的结果中观察不到明显的噪声。

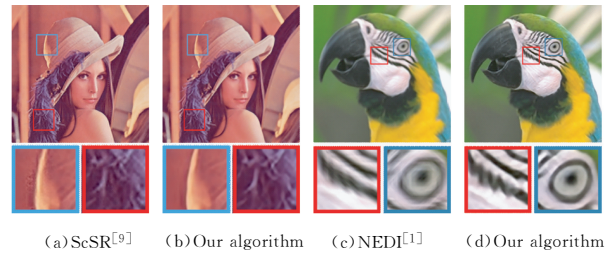


图 6 ScSR, NEDI 及本文方法的结果比较

Fig. 6 Results comparison of ScSR, NEDI and the proposed algorithm

产生这种现象的原因可以解释为:Yang 等^[7-9]的基于稀疏表示的超分辨率方法在求解稀疏系数的过程中,以一种近似随机抽取的方式在字典中选择尽可能少的原子来拟合待超分辨率的图像片,这一过程引入的原子中很可能存在与之差异很大的原子,从而导致在拟合的过程中产生噪声。而基于组稀疏的方法在训练字典的阶段对字典进行了分组,相似的原子处在同一组内。对于待超分辨率的图像片,在求解系数的过程中只“激活”少数与之结构相似的几组原子,即这几组原子的系数可以不为零,其他位置的原子所对应的系数都为零,这样可以保证参与拟合的原子都与待超分辨率的图像片结构相似,可以在很大程度上避免引入噪声。

另外,本文方法中使用的几何对偶性先验来自于 NEDI 算法^[1]的扩展。NEDI 算法^[1]假定图像满足静态高斯过程,即图像中的每一个像素点都能由其周围的邻域通过线性加权组合的方式来表示。同时,它还假定图像在一个小的局部范围内所有像素点都具有相似的结构,即它们都有相同的加权系数,可以表示为 $C\alpha=y$,其中 y 是这个局部中的所有像素点构成的列向量, C 中的第 i 行元素是 y 中第 i 个元素的邻域集合,那么 α 就是这个局部内所有像素点所共有的加权系数。这样的假设本身是符合图像的特性的,然而这一想法只考虑了图像的局部性质,忽略了小区域外部有着相似性质的信息。若直接通过扩大小区域面积的方式来扩展计算样本,则会引入诸多不满足上述性质的样本点,使得系数的求解不准确,从而引入噪声。图 6(c)是用 NEDI 算法^[1]将图像“parrot”放大 4 倍的结果,图 6(d)是用提出的组稀疏算法将“parrot”放大 4 倍的结果。观察图 6(c)可以发现,NEDI 算法基本保持了图像的结构,但其边缘位置产生了明显的模糊和噪声,而提出的方法由于在组稀疏项使用输入图像的全局自相似、对称、镜像等结构信息,使得图像的边缘仍然锐利、自然。

最后,在图像集 BSD500 中随机抽取 10 幅图像,按照实验参数说明设置参数值来进行实验,并选择 LLE^[6], ScSR^[9], GPR^[14], Glasner^[15]和 ASDS^[10]进行比较。在放大 3 倍后,记录超分辨率结果的 PSNR 和 SSIM 值,每个单元格上面的值

为 PSNR, 下面的值为 SSIM, 如表 2 所列。观察表 2 可以发现, 本文算法在大多数的图像上都取得了优于其它算法的结果。

表 2 BSD500 部分图像放大 3 倍的结果比较

Table 2 Comparison of the 3X results selected from training set BSD500 partialy

图像名称	LLE ^[6]	ScSR ^[9]	GPR ^[14]	Glasner ^[15]	ASDS ^[10]	Our algorithm
flower	29.41 0.884	30.16 0.872	28.29 0.849	29.38 0.876	30.91 0.891	30.96 0.887
hat	29.79 0.843	30.15 0.854	28.19 0.829	29.93 0.864	30.93 0.871	31.51 0.880
girl	32.03 0.772	33.41 0.831	31.51 0.792	32.47 0.821	33.54 0.824	34.17 0.852
butterfly	25.02 0.862	25.15 0.853	23.42 0.844	26.39 0.890	27.34 0.905	27.94 0.912
leaf	23.89 0.837	24.62 0.846	21.66 0.812	25.94 0.918	26.78 0.905	27.21 0.913
building	24.55 0.678	24.88 0.710	22.94 0.671	24.66 0.728	25.57 0.745	25.99 0.754
woman	29.57 0.884	30.72 0.896	28.20 0.887	30.37 0.915	31.63 0.922	31.83 0.925
wing	27.12 0.820	27.49 0.838	25.52 0.815	27.83 0.866	28.81 0.873	29.15 0.876
elephant	30.75 0.773	30.72 0.786	29.11 0.754	30.32 0.782	31.18 0.803	31.61 0.811
plant	31.37 0.864	31.94 0.885	29.66 0.827	31.86 0.896	33.47 0.910	33.72 0.909

这 10 幅图像运行的平均时间如图 7 所示, 基于组稀疏表示的图像超分辨率重建算法的平均运行时间低于 LLE^[6] 和 GPR^[14], 高于 ScSR^[9] 和 ASDS^[10], 与 Glasner^[15] 的运行时间相当。同时应注意的是, ScSR^[9] 和 ASDS^[10] 需要耗费大量的时间来计算一个通用的字典, 而在运行时间的比较实验中, 这两种算法训练字典步骤所消耗的时间并未计算在内。

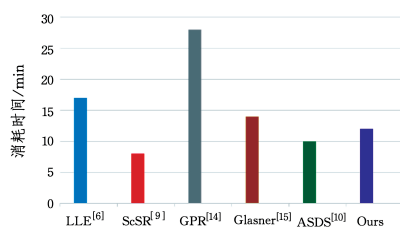


图 7 各算法平均 CPU 消耗时间的比较

Fig. 7 Average CPU consuming time of state-of-the-art algorithms

结束语 本文提出了一种基于组稀疏表示的在线图像超分辨率重建算法。它通过输入图像以在线的方式建立字典, 以输入图像的双三次结果作为超分辨率图像的初值, 结合以图像片为约束对象的组稀疏先验和以像素点为单位的几何对偶性先验, 以迭代的方式求解最终的超分辨率图像。实验表明, 这种方法的放大结果不会在边缘或纹理区域产生显著的模糊和噪声, 其得到的评价参数 PSNR 和 SSIM 也高于参与比较的基于字典的或基于自相似性质的超分辨率算法。

参考文献

- [1] LI X, ORCHARD M T. New edge-directed interpolation[J]. IEEE Transactions on Image Processing, 2001, 10(10): 1521-1527.
- [2] SHAN Q, LI Z R, JIA J Y, et al. Fast image/video upsampling [C]//ACM Transactions on Graphics, 2008.
- [3] TAI Y W, LIU S, BROWN M S, et al. Super resolution using edge prior and single image detail synthesis[C]//Computer Vision and Pattern Recognition. 2010:2400-2407.
- [4] ZHANG K, GAO X, TAO D, et al. Single image super-resolution with multiscale similarity learning[J]. IEEE Transactions on Neural Networks & Learning Systems, 2013, 24(10): 1648-1659.
- [5] SUN J, XU Z, et al. Gradient profile prior and its applications in image super-resolution and enhancement[J]. IEEE Transactions on Image Processing, 2011, 20(6): 1529-1542.
- [6] CHANG H, YEUNG D Y, XIONG Y. Super-Resolution through Neighbor Embedding[C]//Computer Vision and Pattern Recognition. 2004:275-282.
- [7] YANG J, WRIGHT J, HUANG T, et al. Image super-resolution as sparse representation of raw image patches[C]//IEEE Computer Society Conference on Computer Vision and Pattern Recognition. 2008:1-8.
- [8] YANG J, WRIGHT J, HUANG T S, et al. Image super-resolution via sparse representation[J]. IEEE Transactions on Image Processing, 2010, 19(11): 2861-2873.
- [9] YANG J, WANG Z, LIN Z, et al. Coupled Dictionary Training for Image Super-Resolution[J]. IEEE Transactions on Image Processing, 2012, 21(8): 3467-3478.
- [10] DONG W, ZHANG L, SHI G, et al. Image deblurring and super-resolution by adaptive sparse domain selection and adaptive regularization[J]. IEEE Transactions on Image Processing, 2010, 20(7): 1838-1857.
- [11] LU J, SUN Y. Context-aware single image super-resolution using sparse representation and cross-scale similarity[J]. Signal Processing Image Communication, 2015, 32(C): 40-53.
- [12] LI J, GONG W, LI W. Dual-sparsity regularized sparse representation for single image super-resolution[J]. Information Sciences, 2015, 298(C): 257-273.
- [13] YU J, GAO X, TAO D, et al. A unified learning framework for single image super-resolution[J]. IEEE Transactions on Neural Networks & Learning Systems, 2014, 25(4): 780-792.
- [14] HE H, SIU W C. Single image super-resolution using Gaussian process regression[C]//Computer Vision and Pattern Recognition. 2011:449-456.
- [15] GLASNER D, BAGON S, IRANI M. Super-resolution from a single image[C]//International Conference on Computer Vision. 2009:349-356.
- [16] ZONTAK M, MOSSERI I, IRANI M. Separating Signal from Noise Using Patch Recurrence across Scales[C]//Computer Vision and Pattern Recognition. 2013:1195-1202.
- [17] CAI T T, WANG L. Orthogonal Matching Pursuit for Sparse Signal Recovery with Noise[J]. IEEE Transactions on Information Theory, 2011, 57(7): 4680-4688.
- [18] AHARON M, ELAD M, BRUCKSTEIN A. K-SVD: An Algorithm for Designing Overcomplete Dictionaries for Sparse Representation[J]. IEEE Transactions on Signal Processing, 2006, 54(11): 4311-4322.
- [19] MAJUMDAR A, WARD R K. Fast Group Sparse Classification [C]//IEEE Pacific Rim Conference on Communications, Computers and Signal Processing. 2009:11-16.