

文章编号:1674-2869(2017)03-0267-06

稀疏表示中字典学习的影响因子研究

赵娜^{1,2}, 赵彤洲^{1,2*}, 邹冲^{1,2}, 刘莹^{1,2}, 蔡敦波^{1,2}

1. 武汉工程大学计算机科学与工程学院, 湖北 武汉 430205;

2. 智能机器人湖北省重点实验室, 湖北 武汉 430205

摘要:研究了稀疏表示中影响字典矩阵构建质量的关键因素,并实现了关键因子定量化表示.分别对图像数量、取块大小、字典列数和取块步长等因子进行参数调整并生成字典矩阵,结合系数矩阵对原始图像重构,以峰值信噪比和结构相似性索引测量这两种质量评价指标作为字典质量的评估依据.实验以CMU_PIE_Face数据库为数据源,结果表明当图像数量为500张、取块大小为4个像素点、字典列数为512维、取块步长为2个像素点时,所得到的字典具备对原始图像的最佳表示能力.因此,稀疏表示中关键因子的定量化表示可加速字典学习过程且简化模型复杂度,提高字典抽象层质量,具备更强的图像表现力.

关键词:稀疏表示;字典学习;字典精度;图像质量评价指标

中图分类号:TP391 文献标识码:A doi:10.3969/j.issn.1674-2869.2017.03.011

Influence Factors of Dictionary Learning in Sparse Representation

ZHAO Na^{1,2}, ZHAO Tongzhou^{1,2*}, ZOU Chong^{1,2}, LIU Ying^{1,2}, CAI Dunbo^{1,2}

1. School of Computer Science and Engineering, Wuhan Institute of Technology, Wuhan 430205, China;

2. Hubei Key Laboratory of Intelligent Robot(Wuhan Institute of Technology), Wuhan 430205, China

Abstract: We studied the key factors influencing the construction quality of dictionary matrix in sparse representation, and represented them quantitatively. The factors such as the number of images, patch size, dictionary columns and patch step were adjusted as parameters and the dictionary matrix was generated. Combined with the coefficient matrix, the original image was reconstructed, and the dictionary quality was evaluated by using the image quality assessment indices of peak signal to noise ratio and structural similarity index metric. Experiments on CMU_PIE_Face database demonstrate that the resulting dictionary has the best ability to represent the original image at image numbers of 500, patch size of 4 px, dictionary columns of 512 and patch step of 2 px. We found that the quantitative representation of key factors in sparse representation can accelerate the dictionary learning process, simplify the complexity of the model, improve the quality of the dictionary abstraction layer, and show stronger image expression.

Keywords: sparse representation; dictionary learning; dictionary accuracy; image quality assessment index

收稿日期:2016-12-06

基金项目:国家自然科学基金(61103136);武汉工程大学创新基金(CX2015057);武汉工程大学创新基金(CX2016070)

作者简介:赵娜,硕士研究生. E-mail:zhaona_wit@163.com

*通讯作者:赵彤洲,硕士,副教授. E-mail:zhao_tongzhou@126.com

引文格式:赵娜,赵彤洲,邹冲,等.稀疏表示中字典学习的影响因子研究[J].武汉工程大学学报,2017,39(3):267-272.

ZHAO N, ZHAO T Z, ZOU C, et al. Influence factors of dictionary learning in sparse representation [J]. Journal of Wuhan Institute of Technology, 2017, 39(3): 267-272.

人类视觉系统中仅用少量视觉神经元就能捕获自然场景中的关键信息^[1],即场景主要信息可以通过稀疏表示充分表达.现在稀疏表示已经在信号和图像处理领域得到了广泛应用,如图像去噪^[2]、图像恢复^[3]、人脸识别^[4-5]、物体检测^[6]等方面.稀疏表示理论也得到了深入的研究,其中最早由Engan提出了最优方向算法(method of optimal directions, MOD),该算法字典学习方式简单,但是收敛速度很慢.在此基础上,Micheal Elad^[7]于2006年提出了K-SVD算法.该算法在收敛速度上有了很大的提升,但是对噪声很敏感,噪声加大时,用该算法降低噪声会丢失图像的纹理细节而产生模糊的效果.Mairal^[8]于2010年提出了一种在线字典学习算法(online dictionary learning, ODL).与传统算法相比,该算法能大幅降低计算量,提高字典更新速度,能够较好地恢复图像边缘锐度和纹理细节,使字典学习不再局限于小规模、确定的训练样本.本文在介绍基本稀疏表示模型^[9]的基础上,讨论了ODL算法,并使用稀疏建模工具箱(sparse modeling software, SPAMS)对图像不同参数分别训练,从而得到学习字典.通过字典进行图像的重构^[10],并通过峰值信噪比(peak signal to noise ratio, PSNR)和结构相似性索引测量(structural similarity index metric, SSIM)质量评价准则对字典质量进行评估分析.

1 稀疏表示模型的构建

稀疏表示是将自然信号压缩表示为一组基向量的线性组合.假设 $X \in R^m$ 是一个信号, $D = [d_1, d_2, \dots, d_n] \in R^{m \times n}$ 是一个基向量的集合,称为字典,其中 $m \ll n$,稀疏表示就是要找到一个 n 维的稀疏向量 α ,使得

$$X = D \cdot \alpha. \quad (1)$$

由于 $m \ll n$,即 D 是过完备的,则能满足式(1)的 α 有无数个.稀疏表示的关键就是找到非零元个数尽可能小的 α ,因此稀疏表示模型可以表示为:

$$\min \|\alpha\|_0 \text{ s.t. } D \cdot \alpha = X. \quad (2)$$

Donoho 和 Elad 证明了当字典 D 满足 $\sigma(D) \geq 2\|\alpha\|_0$ 时,式(2)有唯一解,但是求解式(2)仍然是 NP(Non-deterministic polynomial)问题.文献^[11]证明了在有限等距性质(restricted isometry property, RIP)条件下, L_0 范数优化问题与 L_1 范数优化问题具有相同的解,即式(2)等同于:

$$\min \|\alpha\|_1 \text{ s.t. } D \cdot \alpha = X. \quad (3)$$

其中 RIP 条件为存在满足某种条件的常数 μ_N ,有:

$$(1 - \mu_N) \|\alpha\|_2^2 \leq \|D\alpha\|_2^2 \leq (1 + \mu_N) \|\alpha\|_2^2, \forall \alpha, \|\alpha\|_0 \leq N. \quad (4)$$

考虑到信号含噪声的情况,模型可以表示成:

$$\min \|\alpha\|_1 \text{ s.t. } \|D \cdot \alpha - X\|_2^2 \leq \varepsilon. \quad (5)$$

为了实现信号的稀疏表示,就必须求解上述的稀疏表示模型问题.因此,需要解决两个基本问题^[12]:一是稀疏分解问题,即如何获取信号在字典 D 下最稀疏的分解系数 α ;二是稀疏字典学习问题^[13],即如何设计与构造有效的稀疏表示字典 D .

2 ODL 算法

2.1 算法原理

构造合适有效的字典是稀疏表示的核心问题之一,其中字典的质量不仅仅取决于算法的优化,而且还与一些影响字典的因素相关,例如:训练图像数量、图像取块大小、字典列数、图像取块步长.本文在 ODL 算法^[14]的基础上,对这些相关因素进行了研究.

假设 $X \in R^{m \times n}$ 是原始信号, $D \in R^{m \times K}$ ($m \ll K$) 为稀疏字典, λ 为正则化因子, t 为迭代次数,则 ODL 算法的步骤为:

步骤1:初始化.使用固定字典初始化 D_0 , $A_0 \leftarrow 0$, $B_0 \leftarrow 0$, A_0 , B_0 为中间变量.

步骤2:稀疏编码.从原始信号中随机选取子集 $x_t = [x_{t,1}, x_{t,2}, \dots, x_{t,\eta}] \in R^{m \times \eta}$,使用 LARS 算法计算字典 D_{t-1} 的稀疏系数 $\alpha_t = [\alpha_{t,1}, \alpha_{t,2}, \dots, \alpha_{t,\eta}]$,使用式(6)更新 A_t 和 B_t ,

$$\begin{cases} A_t \leftarrow A_{t-1} + \alpha_t \alpha_t^T \\ B_t \leftarrow B_{t-1} + x_t \alpha_t^T \end{cases}. \quad (6)$$

其中 $A_t = [a_1, \dots, a_k]$, $B_t = [b_1, \dots, b_k]$, η 为 x_t 的数量.

步骤3:字典更新.使用式(7)依次更新字典 D_{t-1} 的第 j 列 $d_{t-1,j}$,直到 RMSE 满足要求.

$$\begin{cases} \mu_j \leftarrow \frac{1}{A_t(j,j)}(b_j - D_{t-1}a_j) + d_{t-1,j} \\ d_{t-1,j} \leftarrow \frac{1}{\max(\|\mu_j\|_2, 1)} \mu_j \end{cases}. \quad (7)$$

重复步骤2和步骤3,直到满足迭代次数.

2.2 图像重构

为了衡量字典质量,对原始图像进行重构,并以重构图像质量作为字典质量的评估标准,重构

过程如下:

1)对原始图像进行不重叠分块,视每个图像块为一个样本,将图像块按列展开成一个列向量 y_i ,所有样本的列向量并列组合成联合矩阵 $Y=[y_1, y_2, \dots, y_n]$;

2)求得联合矩阵 Y 的稀疏系数矩阵 $\alpha=[\alpha_1, \alpha_2, \dots, \alpha_n]$,使 α_i 为满足 $\min_{\alpha_i} \|y_i - D \cdot \alpha_i\|^2$ $i=1, 2, \dots, n$ 的最优解;

3)将每个样本 y_i 重构为 \hat{y}_i ,使 $\hat{y}_i = D \cdot \alpha_i$;

4)将重构所得的列向量 \hat{y}_i 重新按列组合成图像块,再按分块索引将图像块还原成图像,得到重构图像.

2.3 图像质量评价

对重构图像的质量评价可分为客观评价和主观评价,客观评价是以符合人眼视觉特性的计算模型给出量化指标来进行评价,主观评价是以人的主观感知来打分.本文分别采用 PSNR 和 SSIM 两种客观评价准则来评估字典的重建能力.

1)峰值信噪比

PSNR 是使用最广泛的质量评价方法,其计算公式为:

$$Q_{\text{PSNR}} = 10 \log_{10}((2^n - 1)^2 / S_{\text{MSE}}). \quad (8)$$

其中 n 为每像素的比特数,一般取 8, S_{MSE} 为原始图像与重构图像的均方误差. Q_{PSNR} 值越大表明重建质量越高.

2)结构相似性索引测量

SSIM 表示了原始图像与重建图像的结构相似程度,其计算公式为:

$$Q_{\text{SSIM}}(X, Y) = \frac{2\mu_X\mu_Y + C_1}{\mu_X^2 + \mu_Y^2 + C_1} \cdot \frac{2\sigma_X\sigma_Y + C_2}{\sigma_X^2 + \sigma_Y^2 + C_2} \cdot \frac{\sigma_{XY} + C_3}{\sigma_X\sigma_Y + C_3}. \quad (9)$$

式(9)中 μ_X 、 μ_Y 、 σ_X 、 σ_Y 分别为图像矩阵 X 和 Y 的均值和方差, σ_{XY} 为图像矩阵 X 和矩阵 Y 的协方差, C_1 、 C_2 、 C_3 为常数. Q_{SSIM} 取值介于 0 和 1 之间,值越大表明重建质量越高.

3 实验过程与结果分析

3.1 实验过程

SPAMS 是 Julien Mairal (INRIA) 开发的稀疏建模工具箱,可以解决各类稀疏优化问题,如矩阵分解和字典学习等.

使用该工具箱中的 ODL 算法来研究字典 D 的影响因素(训练字典图像数量、取块大小^[15]、字典列数、取块步长).实验数据采用的是

CMU_PIE_Face 数据库中大小为 32×32 的人脸图像,其中该数据库包括 68 位志愿者在不同光照、姿态、表情的 11 560 张面部图像.假设 X 为训练集图像按取块大小重新排列后的联合矩阵, X_1 为原始图像, X_2 为稀疏表示所得图像,实验步骤如下:

1)确定实验所需取块大小,按取块大小获得训练集图像的联合矩阵 X ;

2)设置字典学习的参数值,包含字典列数、迭代次数等;

3)求得联合矩阵 X 的稀疏字典 D ;

4)求得原始图像矩阵 X_1 的稀疏系数 α ;

5)通过字典 D 和稀疏系数 α ,重构得到稀疏表示图像矩阵 X_2 ;

6)计算 X_1 与 X_2 的 Q_{PSNR} 和 Q_{SSIM} 值,以此来评估字典 D 的质量.

其中步骤 3 中的稀疏字典 D 是通过对联合矩阵 X 中图像块用 ODL 算法直接训练所得,并没有对其先进行特征提取;步骤 4 中求取稀疏系数是用 SPAMS 中的 Lasso 函数;步骤 5 中稀疏表示图像 X_2 的重构是按 2.2 节中的重构方法所得.

本实验分四个阶段对字典 D 的影响因素进行研究:第一阶段是对图像数量的研究,图像数量以步长为 10 从 10 递增至 2 000,然后每组数量的取块大小分别采用 6、8,并固定字典 D 的列数为 512,通过两段曲线来研究图像数量对字典的影响;第二阶段是对取块大小的研究,取块大小以步长为 1 从 3 递增至 8,然后每组取块大小的字典列数分别采用 128、256、512,并固定训练图像数量为第一阶段的最佳图像数量,通过三段曲线来研究取块大小对字典的影响;第三阶段是对字典 D 列数的研究,列数以步长为 8 从 128 递增至 1 024,然后固定图像数量为第一阶段的最佳值,通过取块大小分别为 6、8 的两段曲线来研究字典列数对字典的影响;第四阶段是对取块步长的研究,取块步长以步长为 1 从 1 递增至 8,然后每组的取块大小分别采用 6、8,并固定训练图像数量和字典列数为前几阶段的最佳值,通过两段曲线来研究取块步长对字典的影响.

3.2 实验结果分析

3.2.1 图像数量对字典的影响分析 第一阶段的实验结果如图 1 所示,图 1(a)中随着训练图像数量的增加, Q_{PSNR} 值也逐渐增大,并在数量为 500 后趋于平稳,图 1(b)中的 Q_{SSIM} 值也有相同的趋势,不同的是在数量为 100 时就开始趋于平稳.综合两个质量评价准则,选取本实验的最佳训练图像数量为 500.

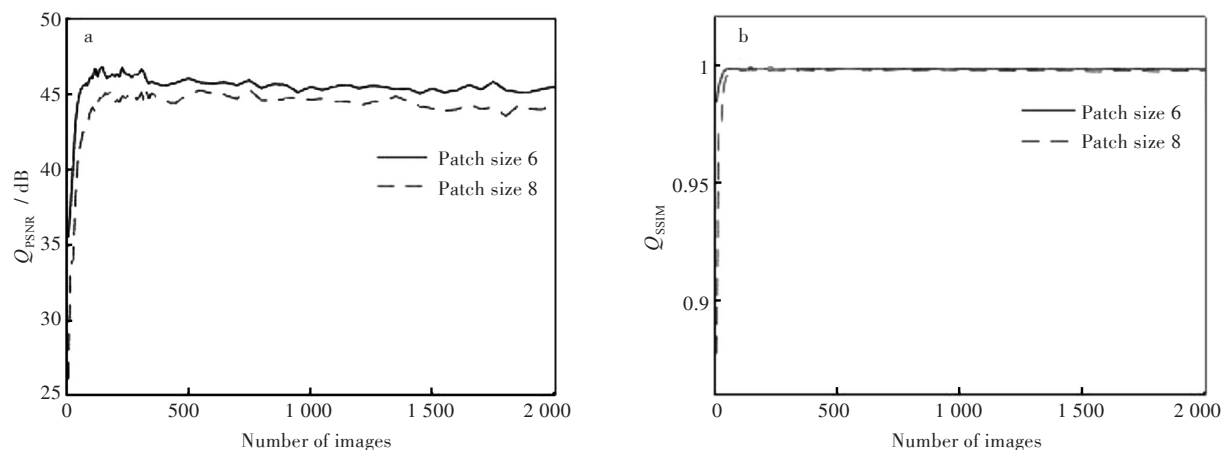


图1 图像数量与(a) Q_{PSNR} 和(b) Q_{SSIM} 的关系

Fig.1 Relationship between the number of images and (a) Q_{PSNR} and (b) Q_{SSIM}

3.2.2 图像取块大小对字典的影响分析 第二阶段的实验结果如图2所示,由第一阶段所得的最佳训练图像数量500作为本阶段的训练数量,选取了字典列数分别为128、256和512的三条曲线来研究取块大小与误差的关系,由图2可以看出,随着

取块大小的递增,质量评价价值不断在下降,三条曲线均在取块大小为3时, Q_{PSNR} 和 Q_{SSIM} 处于最高值,但取块大小为3时,所得图像块的信息太少,笔者认为不适合做取块大小,因此,选取结果次优的取块大小为4作为本实验的最佳值。

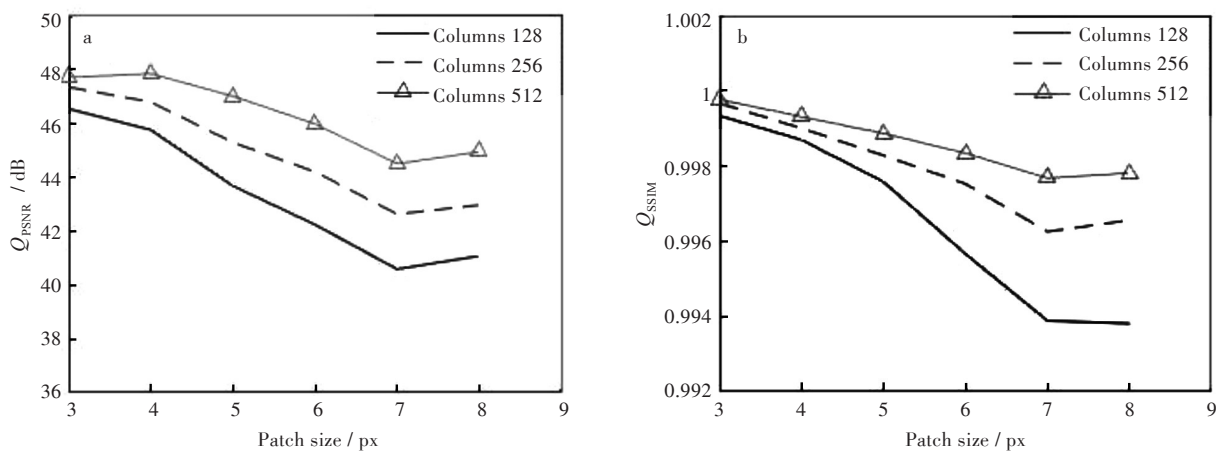
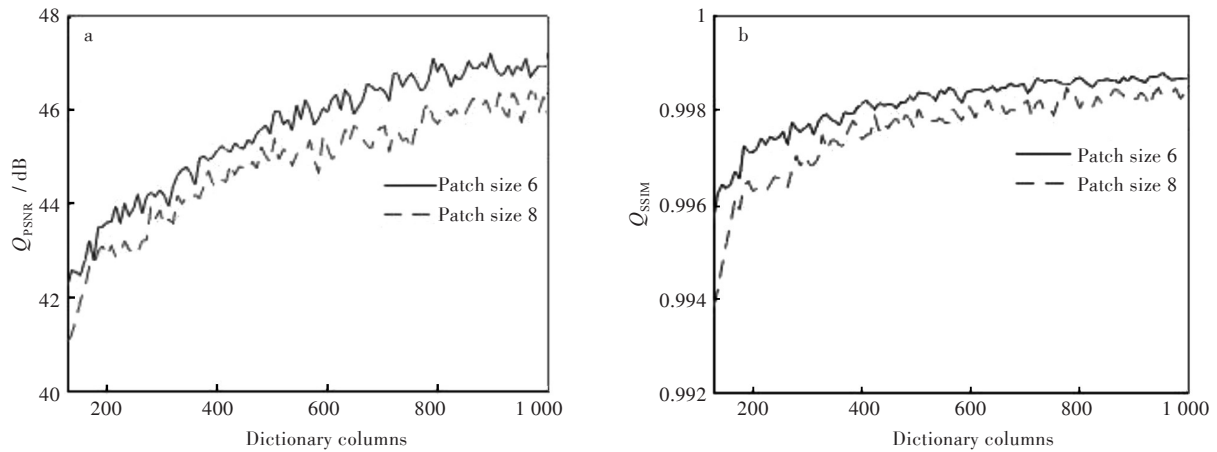
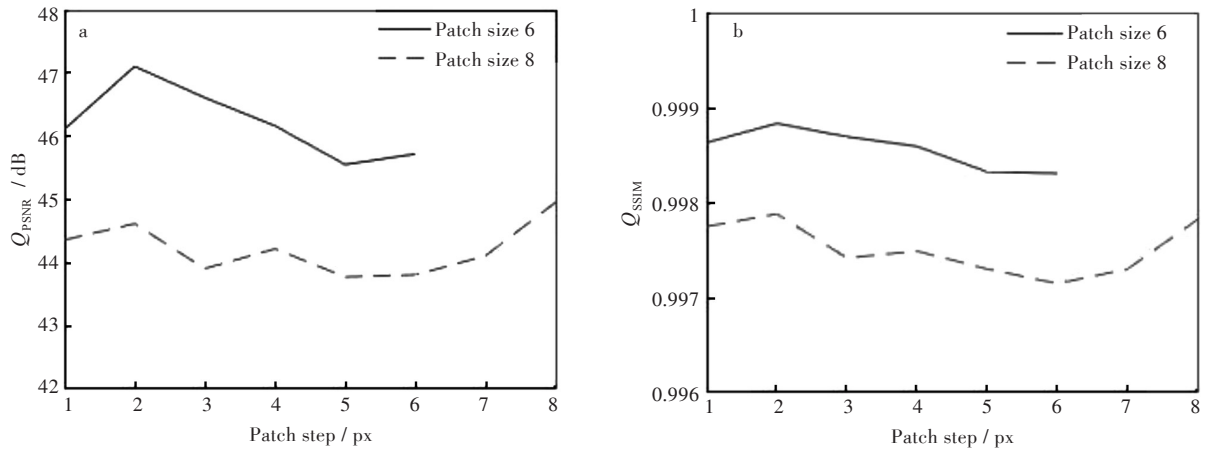


图2 取块大小与(a) Q_{PSNR} 和(b) Q_{SSIM} 的关系

Fig. 2 Relationship between patch size and (a) Q_{PSNR} and (b) Q_{SSIM}

3.2.3 字典列数取值对字典的影响分析 第三阶段的实验结果如图3所示,由图3可知,随着字典列数的不断增加,图3(a)中的 Q_{PSNR} 和图3(b)中的 Q_{SSIM} 值都在不断增大,最后几乎趋于平稳。但是字典列数越大,字典学习时间就越长且计算越复杂,综合图3中两个准则的趋势变化和时间考虑,当字典列数为512时,所得字典能更好地表示图像。

3.2.4 图像取块步长对字典的影响分析 第四阶段的实验结果如图4所示,由前三阶段所得的实验结果分析,本阶段选取图像数量500,字典列数为512来进行实验,并设置取块大小为6、8来研究取块步长与误差的关系,分别由图4(a)和图4(b)可以得出,两条曲线均在取块步长为2时,质量评价指标达到最大,因此,取块步长为2时,训练所得的字典效果最佳。

图3 字典列数与(a) Q_{PSNR} 和(b) Q_{SSIM} 的关系Fig. 3 Relationship between dictionary columns and (a) Q_{PSNR} and (b) Q_{SSIM} 图4 取块步长与(a) Q_{PSNR} 和(b) Q_{SSIM} 的关系Fig. 4 Relationship between patch step and (a) Q_{PSNR} and (b) Q_{SSIM}

为了进一步说明各影响因子对字典的影响,训练图像、取块大小、字典列数这3种因子采取不同参数值时的 Q_{PSNR} 和 Q_{SSIM} 对比如表1所示。由表1可知,随着训练图像数量的增加,在其他影响因子相同的情况下,质量评价指标大致呈逐渐上升

然后平稳的趋势,并在图像数量为500后保持稳定;在相同训练图像数量和字典列数下,图像取块大小为4时图像重构质量最佳;随着字典列数不断增加,重构质量越来越好。

表1 取不同参数时的 Q_{PSNR} 和 Q_{SSIM} 对比Tab. 1 Comparison of Q_{PSNR} and Q_{SSIM} with different parameters

图像数量 the number of images	字典列数 dictionary columns	取块大小 patch size/px					
		4		6		8	
		Q_{PSNR}	Q_{SSIM}	Q_{PSNR}	Q_{SSIM}	Q_{PSNR}	Q_{SSIM}
50	128	47.209 5	0.999 1	42.433 7	0.995 7	41.093 0	0.994 4
	256	48.001 3	0.999 2	44.270 1	0.997 6	41.628 9	0.995 5
	512	48.361 6	0.999 4	44.486 3	0.997 9	40.614 5	0.994 7
100	128	47.348 9	0.999 1	43.195 9	0.996 6	41.695 9	0.995 2
	256	48.059 3	0.999 3	44.541 6	0.997 6	42.974 3	0.996 8
	512	48.118 1	0.999 4	46.111 9	0.998 3	43.774 6	0.997 5
300	128	46.006 6	0.998 7	42.442 8	0.996 3	41.549 9	0.994 9
	256	47.165 1	0.999 1	44.273 6	0.997 5	43.277 0	0.996 9
	512	48.118 1	0.999 3	46.322 1	0.998 5	44.760 5	0.997 6
500	128	47.776 2	0.998 7	42.252 5	0.995 7	41.083 0	0.993 8
	256	48.109 9	0.999 0	44.214 6	0.997 5	42.978 2	0.996 6
	512	48.207 5	0.999 3	45.998 8	0.998 3	44.956 1	0.997 8

4 结 语

以上分别研究了影响稀疏字典的4种因素,研究结果表明在本实验所使用的数据库中,当训练图像数量为500张、取块大小为4个像素点、字典列数为512维、取块步长为2个像素点时,训练得到的稀疏字典能很好地表示该数据库中的图像。当对上述因素定量化时,字典的训练时间维持在几十秒内,因此,本算法无论是训练速度还是字典质量都可达到还原度高的重构图像,满足应用层面的需求。综上所述,本文的研究与图像的尺寸有密切关系,不同的尺寸的图像所需的参数也不尽相同,因此图像大小对字典的影响将是今后的研究方向。

参考文献:

- [1] OLSHAUSEN B A, FIELD D J. Emergence of simple-cell receptive field properties by learning a sparse code for natural images [J]. *Nature*, 1996, 381 (6583): 607-609.
- [2] LI S, YIN H, FANG L. Group-sparse representation with dictionary learning for medical image denoising and fusion [J]. *IEEE Transactions on Biomedical Engineering*, 2012, 59(12): 3450-3459.
- [3] DONG W, ZHANG L, SHI G, et al. Nonlocally centralized sparse representation for image restoration [J]. *IEEE Transactions on Image Processing*, 2013, 22 (4): 1620-1630.
- [4] WAGNER A, WRIGHT J, GANESH A, et al. Toward a practical face recognition system: robust alignment and illumination by sparse representation [J]. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, 2012, 34(2): 372-386.
- [5] 朱杰, 杨万扣, 唐振民. 基于字典学习的核稀疏表示人脸识别方法[J]. *模式识别与人工智能*, 2012, 25 (5): 859-864.
ZHU J, YANG W K, TANG Z M. A dictionary learning based kernel sparse representation method for face recognition [J]. *Pattern Recognition and Artificial Intelligence*, 2012, 25(5): 859-864.
- [6] SHEKHAR S, PATEL V M, NASRABADI N M, et al. Joint sparse representation for robust multimodal biometrics recognition [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2014, 36 (1): 113-126.
- [7] 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.
- [8] MAIRAL J, BACH F, PONCE J, et al. Online learning for matrix factorization and sparse coding [J]. *Journal of Machine Learning Research*, 2010, 11(1): 19-60.
- [9] 练秋生, 石保顺, 陈书贞. 字典学习模型、算法及其应用研究进展[J]. *自动化学报*, 2015, 41(2): 240-260.
LIAN Q S, SHI B S, CHEN S Z. Research advances on dictionary learning models, algorithms and applications [J]. *Acta Automatica Sinica*, 2015, 41(2): 240-260.
- [10] 韦仙, 康睿丹. 基于降维压缩法的图像重构[J]. *武汉工程大学学报*, 2015, 37(12): 69-74.
WEI X, KANG R D. Image reconstruction based on dimension reduction and compression technology [J]. *Journal of Wuhan Institute of Technology*, 2015, 37 (12): 69-74.
- [11] CANDES E J, TAO T. Decoding by linear programming [J]. *IEEE Transactions on Information Theory*, 2005, 51(12): 4203-4215.
- [12] 欧卫华. 基于稀疏表示和非负矩阵分解的部分遮挡人脸识别研究[D]. 武汉: 华中科技大学, 2014.
- [13] 李洪均, 谢正光, 胡伟, 等. 字典原子优化的图像稀疏表示及其应用[J]. *东南大学学报(自然科学版)*, 2014 (1): 116-122.
LI H J, XIE Z G, HU W, et al. Optimization of dictionary atoms in image sparse representations and its application [J]. *Journal of Southeast University (Natural Science)*, 2014 (1): 116-122.
- [14] 吴双, 邱天爽, 高珊. 基于在线字典学习的医学图像特征提取与融合[J]. *中国生物医学工程学报*, 2014, 33(3): 283-288.
WU S, QIU T S, GAO S. Medical image features extraction and fusion based on online dictionary learning[J]. *Chinese Journal of Biomedical Engineering*, 2014, 33(3): 283-288.
- [15] 霍雷刚. 图像处理中的块先验理论及应用研究[D]. 西安: 西安电子科技大学, 2015.

本文编辑: 陈小平