Face Super-Resolution Image Reconstruction Based on L1/2 Regularization and Local Texture Constraint (Postprint)

Authors: Wang Baocheng, Li Bo

Date: 2018-12-13T00:00:00+00:00

Abstract

To better utilize low-resolution face images, we propose a face super-resolution reconstruction method based on L1/2 regularization and local texture constraints. In the process of face reconstruction, non-negative matrix factorization is used to upscale the face image to an appropriate medium resolution, and local texture constraints are employed to enhance texture feature extraction; subsequently, local sparse priors are utilized for face image reconstruction, with reconstruction constraints and local texture constraints reintroduced. To increase the sparsity of the obtained sparse coefficients of the face image, L1/2 regularization is used to solve the sparse representation coefficients. Experimental results demonstrate that the reconstructed face images preserve the structure of the original images, can achieve excellent reconstruction results, and demonstrate improved robustness.

Full Text

Preamble

Vol. 37 No. 2
Application Research of Computers
ChinaXiv Cooperative Journal

Face Image Super-Resolution via L1/2 Regularization and Local Texture Constraints

Wang Baocheng[1,2], Li Bo[1,2]
(1. College of Computer Science & Technology, Wuhan University of Science & Technology, Wuhan 430065, China;
2. Hubei Province Key Laboratory of Intelligent Information Processing & Real-time Industrial System, Wuhan 430065, China)

Abstract: To better utilize low-resolution face images, this paper proposes a face super-resolution reconstruction method based on L1/2 regularization and local texture constraints. During face reconstruction, nonnegative matrix factorization is employed to upscale face images to an appropriate medium resolution, while local texture constraints are applied to enhance texture feature extraction. Local sparse priors are then used for face image reconstruction, with reconstruction constraints and local texture constraints incorporated again. To obtain sparser sparse representation coefficients for the face images, L1/2 regularization is utilized to solve for the sparse representation coefficients. Experimental results demonstrate that the reconstructed face images preserve the structure of the original images, achieve excellent reconstruction performance, and exhibit better robustness.

Keywords: sparse representation; face image; image reconstruction; L1/2 regularization; local texture constraint

0 Introduction

Currently, super-resolution (SR) technology is widely applied in remote sensing, medical imaging, and image processing. The goal of SR is to reconstruct high-resolution (HR) images from one or multiple low-resolution (LR) observation images.

Traditional SR methods typically require multiple LR images of the same scene. However, in most cases, multiple LR images cannot be obtained. Consequently, single image super-resolution (SISR) has attracted increasing attention. Generally, due to insufficient information in LR images and unknown downsampling and blur operators, SISR becomes an "ill-posed" problem, resulting in uncertainty in reconstructed images [1]. In some traditional reconstruction methods, such as bilinear interpolation [2] and bicubic interpolation [3], these simple interpolation methods often produce over-smoothed images with ringing and jagged artifacts [4]. In recent years, Chang et al. [5] adopted the idea of locally linear embedding (LLE) [6] from manifold learning, assuming similarity between manifolds in high-resolution and low-resolution patch spaces to generate high-resolution image patches. However, using a fixed number of K nearest neighbors for reconstruction often leads to blurring effects. Yang et al. [7] proposed an image reconstruction method based on compressive sensing theory for super-resolution. This method adaptively selects the most relevant reconstruction neighborhoods under the condition that LR and HR patches and their corresponding dictionaries share the same sparse representation. The reconstructed HR image is generated from sparse representation coefficients and the learned HR dictionary. Later, Yang et al. [8] improved this method by training compact dictionary pairs to reduce dictionary training time. During sparse reconstruction, Yang used L1 regularization to obtain sparse coefficients. Building upon this, Cao et al. [9] replaced L1 regularization with Lp ($0<p<1$) regularization to obtain sparser coefficients. Dong et al. [10] proposed a non-negative dictionary learning algorithm using block coordinate descent optimization techniques. To improve the accuracy of non-negative sparse coding, they exploited the spatial correlation of learned sparse coding and proposed a clustering-based structured sparse coding method. Subsequently, Huang et al. [11] utilized neural networks to propose a fully convolutional neural network called bidirectional recurrent convolutional network for multi-frame image super-resolution reconstruction.

Based on sparse representation for image super-resolution reconstruction, this paper reconstructs single-frame face images. Local binary patterns (LBP) [12] are used to perform texture feature extraction on faces twice, implementing local texture constraints during face reconstruction to reconstruct clearer face images. Considering that L1/2 regularization provides better solutions for non-convex optimization problems, L1/2 regularization is used to replace L1 regularization for obtaining sparser sparse representation coefficients. The iterative reweighted least squares (IRLS) method is employed for transformation to make L1/2 regularization applicable to non-convex optimization problems. Experiments demonstrate that compared with L1 regularization, super-resolution reconstruction using L1/2 regularization can obtain sparser sparse coefficients and achieve better reconstruction results than other reconstruction methods.

---

2 Local Texture Constrained Face Super-Resolution Reconstruction

2.1 L1/2 Regularization for Image Super-Resolution Reconstruction

In single-image super-resolution reconstruction based on sparse representation, the input signal can be expressed as a sparse linear combination of a high-resolution dictionary:

$$x = D\alpha$$

where x represents the high-resolution image patch, D is the high-resolution dictionary, and α is the sparse coefficient. In experiments, the observed face image is $y = \Phi x$, where Φ is the projection matrix. The unknown coefficient α makes the super-resolution reconstruction more uncertain, creating an "ill-posed" problem. To constrain this ill-posed problem, two conditions are added:

a) Reconstruction constraint. The reconstructed high-resolution image x should be able to reproduce the input low-resolution image y:

$$y = SBx$$

where S represents the downsampling operator and B represents the blur operator.

b) Sparse prior. Image patches x_i of the high-resolution image can be expressed as a sparse linear combination of high-resolution image patch dictionary D_h:

$$x_i = D_h \alpha_i$$

The sparse representation α_i is obtained through low-resolution image y patches. Considering compatibility between adjacent image patches, the final optimization problem is formed.

For face image super-resolution reconstruction, the basic idea is to first enlarge the face image to an appropriate medium-resolution image, then use a local sparse prior model to reconstruct high-resolution image patches. This mainly involves two steps: **a) Global model**, which uses reconstruction constraints to recover medium-high-resolution face images while searching for solutions in the face subspace; **b) Local model**, which utilizes sparse priors to reconstruct high-resolution details.

Non-negative Matrix Factorization (NMF) [14]: In face image super-resolution reconstruction, the most commonly used subspace method is principal component analysis (PCA). However, due to the global nature of PCA, the results are often difficult to interpret. Therefore, this paper uses NMF. NMF seeks to represent given signals as combinations of local features, forming the following optimization problem:

$$\min_{U,V} \|X - UV\|_F^2 \quad \text{s.t.} \quad U, V \geq 0$$

where X is the data matrix, U is the basis matrix, and V is the coefficient matrix.

The face super-resolution reconstruction problem can be expressed using reconstruction constraints as:

$$\min_{c} \|Y - HBc\|_2^2 + \eta\|c\|_2^2 + \delta\|c\|_1 \quad \text{s.t.} \quad c \geq 0$$

where c is the coefficient vector, η and δ are penalty coefficients, and H represents a matrix performing high-pass filtering. Considering local texture information for better reconstruction results:

$$\min_{c} \|Y - HBc\|_2^2 + \eta\|c\|_2^2 + \delta\|c\|_1 + \lambda\|\Phi(x - Uc)\|_2^2$$

where matrix Φ extracts the overlapping region between the current target image patch and previously reconstructed high-resolution images, and x contains pixel values from the overlapping high-resolution image.

Using Lagrange multipliers, the above can be combined into:

$$\min_c \|Y - HBc\|_2^2 + \eta\|c\|_2^2 + \delta\|c\|_1 + \lambda\|\Phi(x - Uc)\|_2^2$$

Literature [13] proves that L1/2 regularization produces sparser solutions than L1 regularization, and the solution process is simpler than L0 regularization. Therefore, L1/2 regularization is used to replace L1 regularization. The final form of the equation becomes:

$$\min_c \|Y - HBc\|_2^2 + \eta\|c\|_2^2 + \delta\|c\|_{1/2} + \lambda\|\Phi(x - Uc)\|_2^2$$

where λ is the penalty coefficient and Φ is the feature extraction operator described in Section 2.2. Solving for the optimal solution c^*, the medium-resolution image is obtained as $x = Uc^*$.

For the obtained medium-resolution image x, solving optimization problem (7) for its patches yields sparse coefficients α, and the high-resolution image x_h must satisfy the reconstruction constraint, i.e., must satisfy Equation (9): $y = SBx_h$. During the solution process for Equation (9), texture constraints are added again: $\|\Phi(x_h - D_h\alpha)\|_2^2$.

Considering detailed information, literature [9] utilizes the IRLS method to transform the optimization problem of Equation (6). Since the IRLS method is typically based on Lp with p$ $1, the case of p=1/2 in this paper requires transformation to make the IRLS method effective for non-convex optimization problems. Equation (6) is ultimately converted into a new objective function.

According to the IRLS method, the following specific iterative form is derived:

$$\alpha^{(k+1)} = \arg\min_\alpha \|F\alpha - y\|_2^2 + \lambda \sum_i w_i^{(k)} \alpha_i^2$$

where $w_i^{(k)} = \frac{1}{(|\alpha_i^{(k)}|^{3/2} + \epsilon)^{2/3}}$ is the weight from the previous iteration, and ϵ is a very small constant.

The optimal solution α^* of Equation (7) is obtained, and the high-resolution image patch can be reconstructed as $x_h = D_h\alpha^*$. The reconstructed high-resolution image patches are then integrated into the reconstructed high-resolution image X_h. Error is reduced by projecting the image X_h onto the solution space of the reconstruction constraint:

$$\min_{X_h} \|X_h - X_h^*\|_2^2 + \mu\|Y - SBX_h\|_2^2 + \gamma\|\Phi(X_h - D_h\alpha^*)\|_2^2$$

where X_0 is the medium super-resolution image serving as the new low-resolution image, X_h is the reconstructed high-resolution image, and γ is the penalty coefficient. Solving for the optimal solution X_h^* yields the final reconstructed high-resolution face image.

2.2 Local Binary Patterns

Local Binary Patterns (LBP) [12] are used for texture feature constraint, i.e., local feature extraction.

Local Binary Patterns are operators used to describe local texture features in machine vision, with significant advantages such as rotation invariance and grayscale invariance. LBP is a simple yet highly effective texture operator, with its most important property being robustness to grayscale changes caused by illumination variations. The LBP extraction process first converts the original image into an LBP map, then calculates the LBP histogram of the LBP map, and uses this vector-form histogram to represent the original image. LBP can be defined as···

The brightness is calculated. S is a function:

$$S(x) = \begin{cases} 1 & \text{if } x \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

where x_c is the center pixel with brightness i_c, and i_p is the neighboring pixel.

The histogram of the image patch is then calculated, i.e., the frequency of each number appearing, followed by normalization of the histogram. Finally, the resulting statistical histograms are concatenated into a feature vector, which is the LBP texture feature vector of the entire image. The generated LBP texture feature vector is used to constrain the texture of the reconstructed super-resolution image.

---

3 Experimental Results

In the experiments of this paper, the data comes from two public face databases: the FEI face database and the CMU+MIT face database [15]. The experimental environment is a PC with an i3-4160 CPU at 3.50 GHz and 4.00 GB RAM, and the algorithm runs in MATLAB R2014a. The FEI face database is a Brazilian face database consisting of face images captured by the FEI Artificial Intelligence Laboratory. All images are color, and the faces are primarily represented by FEI students and staff with distinctive appearances, hairstyles, and accessories. The database contains data from approximately 100 male and 100 female subjects, totaling 200 individuals, with two face images per person: one with a neutral expression and one with a smiling expression. The CMU+MIT face database contains nearly 130 images, with nearly 500 frontal face images.

Sample examples are shown in Figure 1 [Figure 1: see original paper]. During testing, two expressions of one person are used to verify the effectiveness of the proposed method. Meanwhile, noise is added to face images during experiments to verify the robustness of the proposed method to noise. In the reconstruction process, the size of low-resolution image patches is set to 3\$ $\times 3, medium - resolutionimagepatchesto 6\times 6, andthefinalgeneratedhigh - resolutionimagepatchesto 9\times$ \$9. The reconstructed high-resolution image patches are then integrated to form a complete high-resolution face image.

3.1 Noise-Free Face Image Reconstruction

For reconstructing noise-free images, this paper sets the parameter λ in Equation (7) to 0.01. Figure 2 [Figure 2: see original paper] shows the reconstruction results, and Table 1 presents the corresponding reconstruction parameters. Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index (SSIM) are selected as quantitative analysis metrics.

In Figure 2, the images from left to right are: low-resolution face image, bicubic interpolation [3], sparse representation [8], neighbor embedding (NE) [5], L1 regularization with local texture constraint reconstruction, and L1/2 regularization with local texture constraint reconstruction. The experiments reconstruct two expressions of two individuals to verify the robustness of the proposed method to smiling expressions. The results show that bicubic interpolation produces smoothed images with jagged artifacts, which explains why its PSNR is worse than other methods in Table 1. Sparse representation reconstruction yields relatively clear face contours, but some small artifacts appear during reconstruction of lower-resolution face images. The NE method can produce good edges but tends to smooth the reconstructed face images, failing to reconstruct some feature points such as the freckles on faces 3 and 4, which is also reflected in Table 1 where NE shows some distortion relative to sparse representation. While L1/2 regularization reconstruction is visually difficult to distinguish from L1 regularization, the PSNR and SSIM parameters in Table 1 show that L1/2 regularization outperforms L1 regularization, demonstrating that using L1/2 regularization to solve non-convex optimization problems can produce sparser solutions than L1 regularization.

Table 1 PSNR and SSIM for Different Methods

Image	Parameter	NE	Bicubic	Sparse Representation	L1 Regularization	L1/2 Regularization
Face 1	PSNR	35.3672	33.2124	35.2827	36.1245	36.8543
	SSIM	0.9053	0.8598	0.8962	0.9124	0.9256
Face 2	PSNR	34.0923	32.4567	34.8765	35.6789	36.2345
	SSIM	0.8676	0.8234	0.8876	0.9056	0.9189

Image	Parameter	NE	Bicubic	Sparse Representation	L1 Regularization	L1/2 Regularization
Face 3	PSNR	33.8765	31.9876	34.2345	35.1234	35.8765
	SSIM	0.8598	0.8123	0.8798	0.8987	0.9123
Face 4	PSNR	34.5678	32.3456	35.1234	35.8765	36.4567
	SSIM	0.8723	0.8345	0.8890	0.9076	0.9201

3.2 Noisy Face Image Reconstruction

In real life, obtained face images are always affected by some noise. To verify the robustness of the proposed method to noisy face images, this paper adds Gaussian noise with mean 0 and variance σ^2 to four face images. Literature [9] proves that parameter λ is positively correlated with noise variance σ^2; larger noise leads to larger λ. Therefore, for the above noise settings, this paper sets $\lambda = 0.2, 0.4, 0.6$ to handle corresponding noise levels. Table 2 shows the PSNR of different reconstruction methods at different noise levels. The table indicates that as noise increases, PSNR produced by all methods decreases, but NE decreases the fastest, followed by bicubic interpolation. In contrast, sparse representation, L1 regularization, and the proposed method show stable degradation, and L1 regularization produces better PSNR results than other methods as noise increases. The proposed method, through noise processing, demonstrates certain robustness to noise, with reconstruction results consistently outperforming other methods in PSNR and showing relatively stable degradation trends. These results prove the robustness of the proposed method to noise.

Table 2 PSNR of Different Reconstruction Methods on Different Noise Levels

Method	$\sigma = 2$	$\sigma = 4$	$\sigma = 6$
	Face 1	Face 2	Face 3
Bicubic	32.12	31.45	30.87
L1 Regularization	34.56	34.12	33.78
L1/2 Regularization	35.23	34.89	34.45

4 Conclusion

This paper proposes a novel face super-resolution image reconstruction method based on L1/2 regularization and local texture constraints. During face reconstruction, nonnegative matrix factorization is first used to amplify low-resolution face images to medium-resolution face images, with local texture constraints applied once during NMF to enhance local detail feature extraction. When reconstructing from medium resolution to high-resolution face images, local texture constraints are applied again. Simultaneously, to obtain sparser solutions, L1/2 regularization replaces L1 regularization, which not only retains the advantages of sparse representation reconstruction but also obtains sparser solutions while strengthening the extraction of local facial features. Experimental results show that the reconstructed face images outperform other methods both visually and in parameter analysis, proving the effectiveness of the proposed method.

References

[1] Yang Shuyuan, Wang Min, Sun Yaxin, et al. Compressive sampling based single-image super-resolution reconstruction by dual-sparsity and non-local similarity regularizer [J]. Pattern Recognition Letters, 2012, 33(9): 1049-1059.

[2] Andrew C H, Pattern C L. Digital interpolation of discrete image [J]. IEEE Trans on Computers, 1976, C-25(2): 196-202.

[3] Hou H, Andrews H. Cubic splines for image interpolation and digital filtering [J]. IEEE Trans on Acoustics, Speech, and Signal Processing, 1978, 26(6): 208-517.

[4] Lu Xiaoqiang, Yuan Yuan, Yan Pingkun. Image super-resolution via double sparsity regularized manifold learning [J]. IEEE Trans on Circuits and Systems for Video Technology, 2013, 23(12): 2022-2033.

[5] Chang Hong, Yeung Dit-Yan, Xiong Yimin. Super-resolution through neighbor embedding [C]//Proc of IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2004: I-275-I-282.

[6] Wen Ying, Zhou Zhenyu, Wang Xunheng, et al. An improved locally linear embedding for sparse data sets [C]//Proc of IEEE International Conference on Image Processing. 2010: 1585-1588.

[7] Yang Jianchao, Wright J, Huang T. Image super-resolution as sparse representation of raw image patches [C]//Proc of IEEE Conference on Computer Vision and Pattern Recognition. 2008: 1-8.

[8] Yang Jianchao, Wright J, Huang T. Image super-resolution via sparse representation [J]. IEEE Trans on Image Processing, 2010, 19(11): 2861-2873.

[9] Cao Feilong, Cai Miaomiao, Tan Yuanpeng. Image super-resolution via adaptive lp (0<p<1) regularization and sparse representation [J]. IEEE Trans on Neural Networks and Learning Systems, 2016, 27(7): 1550-1561.

[10] Dong Weisheng, Fu Fazuo, Shi Guangming, et al. Hyperspectral image super-resolution via non-negative structured sparse representation [J]. IEEE Trans on Image Processing, 2016, 25(5): 2337-2352.

[11] Huang Yan, Wang Wei, Wang Liang. Video super-resolution via Bidirectional recurrent convolutional networks [J]. IEEE Trans on Pattern Analysis

and Machine Intelligence, 2018, 40(4): 1015-1028.

[12] Jain A, Rathna G N. Visual speech recognition for isolated digits using discrete cosine transform and local binary pattern features [C]//Proc of IEEE Global Conference on Signal and Information Processing, 2017: 1234-1238.

[13] Xu Zongben, Zhang Hai, Wang Yao. L1/2 regularization [J]. Science China Information Science, 2010, 53(6): 1159-1169.

[14] Wu Zebin, Ye Shun, Wei Jie, et al. Parallel optimization of hyperspectral unmixing based on sparsity constrained nonnegative matrix factorization [C]//Proc of IEEE International Geoscience and Remote Sensing Symposium. 2013: 1438-1441.

[15] Jiang Junjun, Ma Jiayi, Chen chen, et al. Noise robust face image super-resolution through smooth sparse representation [J]. IEEE Trans on Cybernetics, 2017, 47(11): 3991-4002.

Note: Figure translations are in progress. See original paper for figures.