This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2018.2799861, IEEE Access

IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Long-distance object recognition with image super resolution: A comparative study

Xiaomin Yang,[1] Wei Wu,[1] Kai Liu,[2] *Senior Member, IEEE*, Pyoung Won Kim,[3] Arun Kumar Sangaiah,[4] and Gwanggil Jeon,[5] *Member, IEEE*

[1]College of Electronics and Information Engineering, University of Sichuan, Chengdu, Sichuan 610064
[2]School of Electrical Engineering and Information, University of Sichuan, Chengdu, Sichuan 610064, China
[3]Department of Korean Language Education, Incheon National University, Incheon, 22012, Korea
[4]School of Computing Science and Engineering, Tamil Nadu, 632014, India
[5]Department of Embedded Systems Engineering, Incheon National University, Incheon, 22012, Korea

Corresponding author: Wei Wu (e-mail: wuwei@ scu.edu.cn).

This research is supported by the National Natural Science Foundation of China (#61701327, #61711540303, and #61473198), National Research Foundation of Korea (#NRF-2017K2A9A2A06013711).

ABSTRACT Monitor systems are ubiquitously deployed in public areas. However, monitor systems face a major challenge regarding long-distance object recognition. Super-resolution constitutes a popular choice to address this challenge. Since super-resolution methods are used in many applications, it is necessary to understand these methods and make a comparative study of them. In this paper, we perform a comparative study on six super-resolution methods over two recognition algorithms. The paper evaluates super-resolution performance based on recognition accuracy, and serves as a summary assessment of image super-resolution algorithms.

INDEX TERMS: super-resolution; sparse representation; deep learning; convolutional neural networks.

I. INTRODUCTION

Visual surveillance is the latest paradigm of monitor public security through machine intelligence. It comprises the use of visual data captured by various cameras placed in vehicles, corridors, traffic signals, etc. Visual surveillance facilitates the recognition of human behavior, faces, vehicles, crowd activities, and gestures to achieve application-specific objectives. However, in surveillance applications, objects are often far away from cameras, leading to low resolution, as shown in Fig. 1. The low resolution object contain little information for recognition leading to poor recognition accuracy [1, 2] for most classification algorithms. Improving image resolution is a promising way to increase recognition accuracy.

Super-resolution (SR) technology constitutes a popular method to improve image resolution. At present, there are various SR methods. Generally, SR methods can be divided into three categories: (1) interpolation-based methods; (2) multi-frame-based methods; and (3) example-based methods. Fig. 2 presents a summary of different image super-resolution methods.

Interpolation-based methods [3, 4, 5] estimate high resolution (HR) pixels through information from neighboring pixels. However, these methods suffer from producing blurred edges and artifacts since they cannot introduce new details.

FIGURE 1. **An image obtained from a monitor system.**

Multi-frame-based methods [6, 7, 8] improve resolution by combining information of serious low resolution (LR) images obtained from the same scene with slightly-moved viewpoints, as shown in Fig. 3. However, information from the frames of the same scene is limited. These methods fail to achieve satisfactory resultant images for large magnification factors.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2018.2799861, IEEE Access

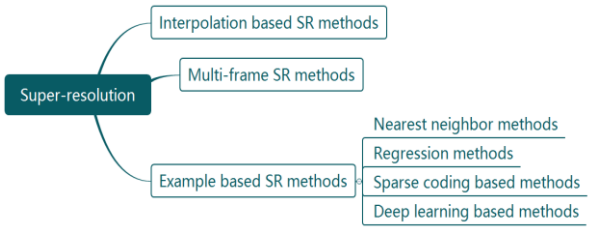


FIGURE 2. Summary of image super-resolution

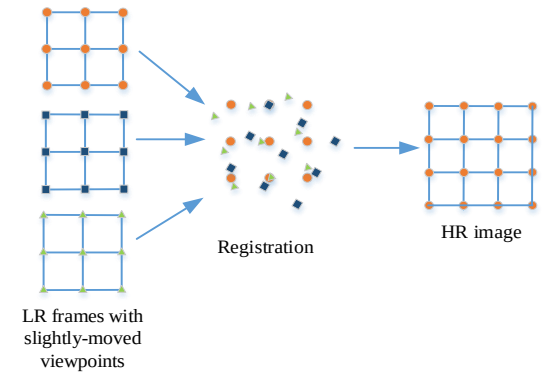


FIGURE 3. Implementation of multi-frame SR methods.

Example-based methods usually utilize a database containing many LR–HR patch pairs to establish the mapping relationship of the corresponding LR and HR patches (see Fig. 4). Generally, there are four types of example-based methods: (1) nearest neighbor (NN) -based methods [9, 10]; (2) regression-based methods [11, 12, 13, 14]; (3) sparse-coding-based methods [15, 16, 17, 18, 19, 24]; and (4) deep-learning-based methods [20, 21]. Figs. 5–8 illustrate the four types of methods. NN-based methods assume that the corresponding LR and HR patches possess locally similar geometry at certain manifolds (see Fig. 5). Regression-based methods build a regression model estimating the relationship of the corresponding LR and HR patches (see Fig. 6). Sparse-coding-based methods compute sparse coefficients of the input LR patch, and use them to estimate an HR patch (see Fig. 7). Deep-learning-based methods employ deep convolutional neural networks (CNN) to estimate the mapping relationship of corresponding LR and HR images (see Fig. 8).

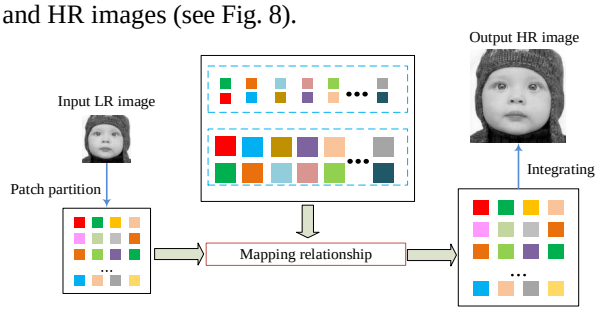


FIGURE 4. Implementation of example-based SR methods.

SR methods have been widely applied to improve image quality. However, precisely how different SR methods effect recognition performance is still not well understood.

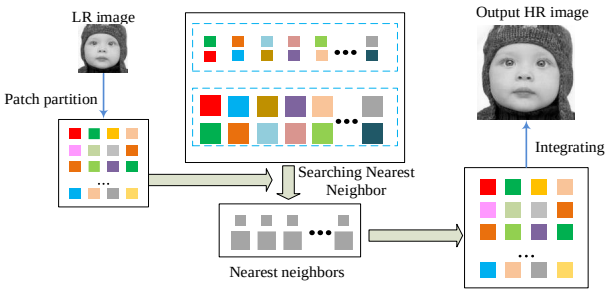


FIGURE 5. Implementation of the NN-based method.

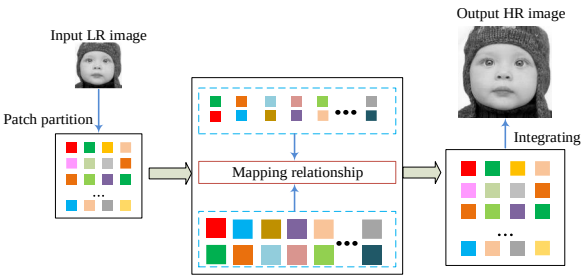


FIGURE 6. Implementation of regression-based methods.

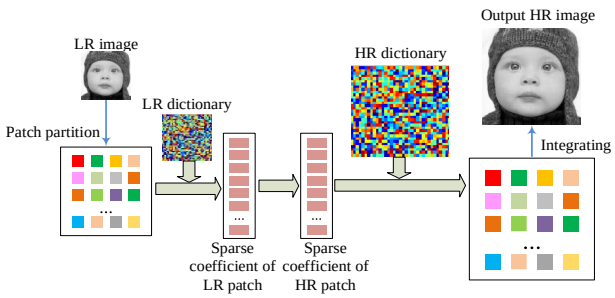


FIGURE 7. Implementation of sparse-coding-based methods.

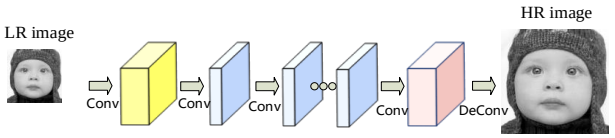


FIGURE 8. Implementation of deep-learning based-methods.

The objective of this paper is to provide a comparative study of SR methods and elucidate their effectiveness for recognition in low-resolution images for monitor systems. Our contributions include the following:

1) Understanding the differences between various SR methods;

2) Demonstrating the effectiveness of SR methods in recognition accuracy of LR images;

3) Learning the relationship between accuracy performance with upscale factors.

The remainder of this paper is as follows. The details of different SR methods are discussed in Section 2. Experimental results are presented in Section 3. Conclusions are given in Section 4.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2018.2799861, IEEE Access

II. SUPER-RESOLUTION METHOD

The objective of SR is to improve the resolution of LR image Y to reconstruct its corresponding HR image X:

$$Y = F(X) = SHX \tag{1}$$

where H denotes a low-pass filter; and S is a down-sampling operator. The problem of super-resolution is, given an LR image Y, finding an image $\hat{X}$ with $\hat{X} \approx X$.

Generally, example-based methods successfully surpass other methods, such as interpolation-based methods or multi-frame-based methods in perceptual quality. Furthermore, due to their good performance, sparse-coding-based methods and deep-learning-based methods have widely utilized SR methods in recent years. This study focuses on the performance of sparse-coding-based methods and deep-learning-based methods.

A. Sparse-coding-based methods

Sparse-coding-based super-resolution methods work on the basis that natural image patches could be sparsely represented by an over-complete dictionary. Let x and y denote the HR and LR patches, respectively. We use D_h and D_l to denote the HR and LR dictionaries, respectively. The HR patch x and LR patch y can be sparsely represented as $x=D_h\alpha_h$ and $y=D_l\alpha_l$, respectively.

1) TRADITIONAL SPARSE-CODING-BASED METHOD

The traditional sparse-coding based (TSC) SR method [16] assumes that the corresponding y and x share the same sparse coefficients, according to the D_l and corresponding D_h, respectively, i.e., $\alpha_l=\alpha_h$. Thus, the sparse coefficients of a low-resolution patch in terms of D_l will be used to reconstruct the corresponding high-resolution patch from D_h.

To ensure that x and y share the same sparse coefficients with respect to D_l and D_h, respectively, the two dictionaries are learned simultaneously. Let $X=\{x_1,x_2,\ldots,x_n\}$ be the set of HR patches and $Y=\{y_1, y_2,\ldots, y_n\}$be the set of LR patches. $A_h=\{\alpha_{h1}, \alpha_{h2,\ldots,} \alpha_{hn}\}$ and $A_l=\{\alpha_{l1}, \alpha_{l2,\ldots,} \alpha_{ln}\}$are sparse coefficients of X and Y respectively. HR and LR dictionaries learning can be approached through solving the following minimization problem:

$$\min_{D_h,D_l,\mathrm{A}_h,\mathrm{A}_l} \left\|\mathrm{X}-D_h\mathrm{A}_h\right\|^2+\left\|\mathrm{Y}-D_l\mathrm{A}_l\right\|^2 +\lambda\{\left\|\mathrm{A}_h\right\|_1+\left\|\mathrm{A}_l\right\|_1\}+\gamma\left\|\mathrm{A}_h-\mathrm{A}_l\right\|^2 \tag{2}$$

λ and γ are regularization parameters to balance the terms in the objective function. The dictionaries D_l and D_h can be easily generated from a set of samples by the methods such as KSVD.

For reconstruction, the input LR image is divided into overlapping patches. It is just necessary to calculate the sparse coefficients of LR patch y_{input} according to LR dictionary D_l, with the following optimization problem:

$$\min_{\{\alpha_l\}} \left\|y_{input}-D_l\alpha_l\right\|_2 \tag{3}$$

The corresponding HR patch x_{out} can then be obtained by the product of HR dictionary D_h and coefficients α_l, as $x_{out}=D_h\alpha_l$. Finally, the SR output is recovered by averaging all of the overlapping HR patches. An implementation of the TSC method is presented in Fig. 9.

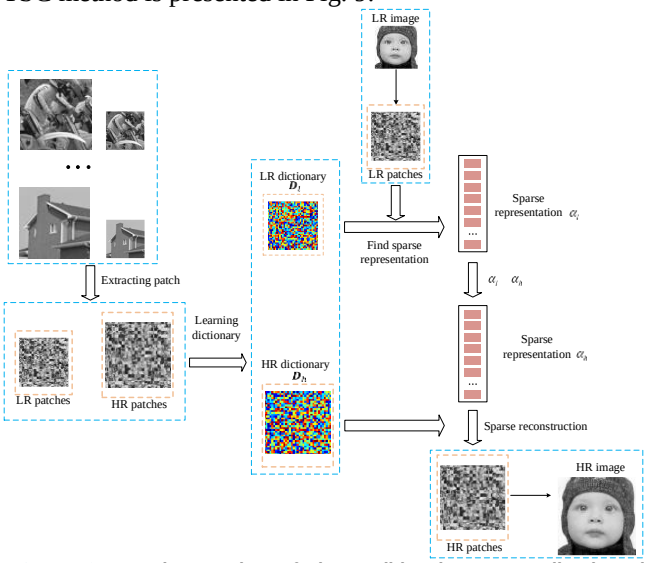


FIGURE 9. Implementation of the traditional sparse-coding-based method.

2) SEMI-COUPLE DICTIONARY LEARNING-BASED METHOD

The semi-couple dictionary learning (SCDL) based method [18] assumes that the sparse coefficients of corresponding LR and HR patches are linearly related, i.e., $\alpha_l = W\alpha_h$. So, W, HR and LR dictionaries can be approached through solving the following minimization problem:

$$\min_{D_h,D_l,\mathrm{A}_h,\mathrm{A}_l,W} \left\|\mathrm{X}-D_h\mathrm{A}_h\right\|^2+\left\|\mathrm{Y}-D_l\mathrm{A}_l\right\|^2 +\lambda\{\left\|\mathrm{A}_h\right\|_1+\left\|\mathrm{A}_l\right\|_1\}+\gamma\left\|\mathrm{A}_h-W\mathrm{A}_l\right\|^2 \tag{4}$$

For reconstruction, the sparse coefficient α_l of the LR patch can be obtained by Eq. (3). Then, the sparse coefficient of HR can be obtained as $\alpha_h=W\alpha_l$. HR patch x_{out} can be produced by the product of HR dictionary D_h and coefficients α_h, as $x_{out}=D_h\alpha_h$. Finally, the SR output is recovered by averaging all of the overlapping HR patches. The implementation of the SCDL-based method is illustrated in Fig. 10.

3) COUPLE DICTIONARY LEARNING-BASED METHOD

The couple dictionary learning (CDL) based method [19] assumes that the sparse coefficients of the corresponding LR patch y and HR patch x are the same in a certain space, i.e., $U\alpha_l = V\alpha_h$. So, U , V, HR and LR dictionaries can be approached through solving the following minimization problem:

$$\min_{D_h,D_l,\mathrm{A}_h,\mathrm{A}_l,V,U} \left\|\mathrm{X}-D_h\mathrm{A}_h\right\|^2+\left\|\mathrm{Y}-D_l\mathrm{A}_l\right\|^2 +\lambda\{\left\|\mathrm{A}_h\right\|_1+\left\|\mathrm{A}_l\right\|_1\}+\gamma\left\|V\mathrm{A}_h-U\mathrm{A}_l\right\|^2 \tag{5}$$

For reconstruction, the sparse coefficient α_l of the LR patch can be obtained by Eq. (3). The sparse coefficient α_h

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2018.2799861, IEEE Access

of HR can then be obtained as $\alpha_h = V^{-1}U\alpha_l$. HR patch x_{out} can be produced by the product of HR dictionary D_h and coefficients α, as $x_{out} = D_h\alpha_h$. Finally, the SR output is recovered by averaging all of the overlapping HR patches. The implementation of the CDL-based super-resolution method is presented in Fig. 11.

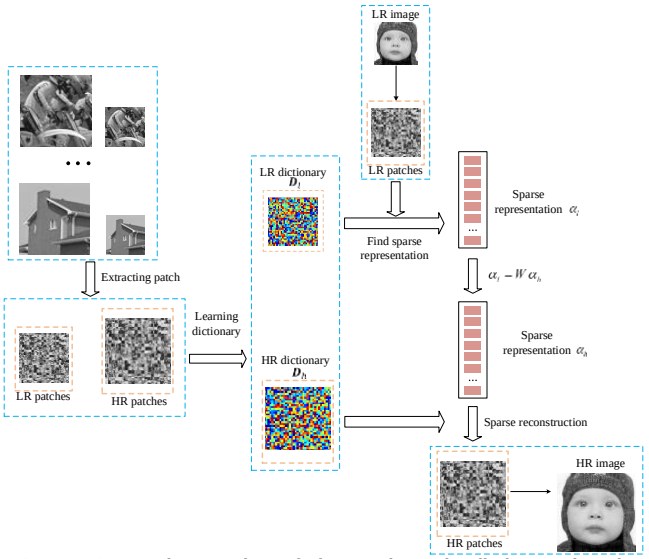


FIGURE 10. Implementation of the semi-couple dictionary learning based method.

B. Deep-learning-based methods

Deep-learning-based methods use convolutional neural networks (CNN) to estimate the mapping relationship of the corresponding LR and HR images, demonstrating promising SR performance. The SRCNN method [20] and VDSR method [21] are state-of-the-art deep-learning-based methods.

1) SRCNN METHOD

The SRCNN method [20] works on the assumption that mapping from LR to HR patches could be learned in an end-to-end manner by CNN.

The SRCNN method comprises three layers: (1) patch extraction and representation; (2) non-linear mapping; and (3) reconstruction.

In the patch extraction and representation layer, the image is divided into overlapping patches and represented as a high-dimensional vector. This layer can be interpreted as an operation that convolves the input image with a set of filters, and obtains an n_1-dimensional feature for each patch.

In the non-linear mapping layer, the n_1-dimensional feature is nonlinearly mapped to the n_2-dimensional vector that represents the HR patch. This layer can be interpreted as a nonlinear map between LR and HR domains.

In the reconstruction layer, HR patches are combined to generate the final reconstructed HR image through averaging. This HR image is expected to be similar to the ground truth $\boldsymbol{X}$.

Denote the convolution layer as Conv(f_i, n_i, c_i), where the variables f_i, n_i, and c_i represent filter size, number of filters and number of channels, respectively. The SRCNN method uses three layers, including Conv(9,64,1), Conv(5,32,64), and Conv(5,1,32) as the network structure. The CNN-based super-resolution method is presented in Fig. 12.

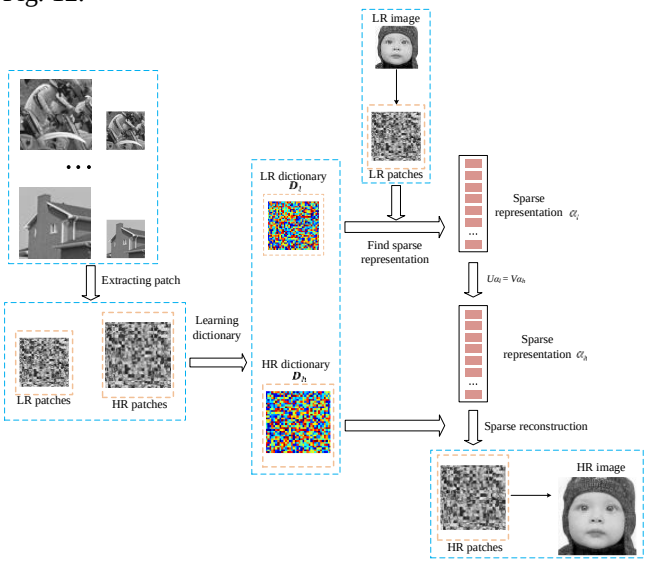


FIGURE 11. Implementation of the couple dictionary learning-based method.

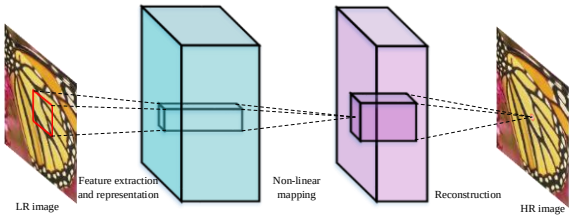


FIGURE 12. Implementation of the SRCNN method [20].

1) VDSR METHOD

The VDSR method [21] utilizes deep convolutional networks for SR. VDSR also uses multiple layers. The first layer operates on the input image. The last layer is used for image reconstruction. The remaining layers are of the same type, i.e., 64 filter of the size 3×3×64. These filters operate on a 3×3 spatial region across 64 channels (feature maps).

The network structure consists of 20 convolution layers, including Conv(3,64,1), 18×Conv(3,64,64), and Conv(3,1,64). For the training process, residual learning and adjustable gradient clipping were employed for fast training. Regarding the scales, it uses a single model for multiple SR scales by embedding training samples with different scales. An implementation of the VDSR method is shown in Fig. 13.

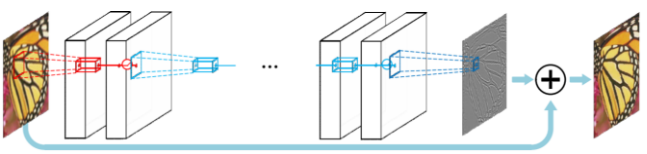


FIGURE 13. Implementation of the VDSR method [21].

III. EXPERIMENT

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2018.2799861, IEEE Access

The objective of this work is to perform a comprehensive analysis for how SR could improve recognition performance in low-resolution images. The super-resolution-based recognition scheme combines super-resolution and recognition. For a comprehensive analysis, we are only concerned with face recognition applications. The scheme is shown in Fig. 14.

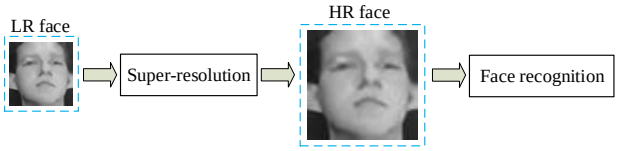


FIGURE 14. **The super-resolution-based recognition scheme.**

The experiments investigate the following issues:
1) The impact of image resolution on recognition accuracy;
2) Comparison of SR methods' performance on improving recognition accuracy;
3) The relation of upscale factors with the improvement of recognition accuracy.

Experiments are conducted to evaluate the scheme on the ORL database (http://www.cl.cam.ac.uk/research/dtg/attarchive/facedatabase.html). The ORL database contains approximately 400 face images of 40 individuals.

FIGURE 15. **Samples of the ORL database.**

A. Recognition method

To understand the performance of different SR methods, the combination of SR methods and recognition algorithms are tested in the experiments. The following SR methods are considered: Bicubic [2], TSC [16], SCDL [18], CDL [19], SRCNN [20], and VDSR [21]. For recognition, it is generally difficult to determine which recognition algorithms are optimal. So, we take two popular face algorithms, i.e., LBP and VGG Face, to evaluate the improvement performance of the SR method. A brief description of each face recognition algorithm is given in Table I. The implementation of each algorithm is presented in Figs. 16-17

TABLE I
PROPERTY DESCRIPTION OF METHODS

Method	Brief descriptor
LBP [22]	The LBP method forms labels of each pixel in an image by thresholding the neighborhood of the pixel, and considers the result as a binary number.
VGG Face [23]	The VGG Face method is based on the CNN architecture. The network is composed of a sequence of convolutional, pool, and fully-connected (FC) layers. The VGG-Face network has a deep architecture composed of 3×3 convolution layers and 2×2 pooling layers, and three fully-connected layers.

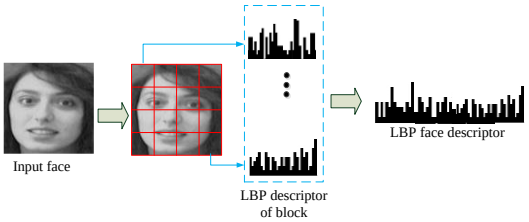


FIGURE 16. **Implementation of LBP Face descriptor.**

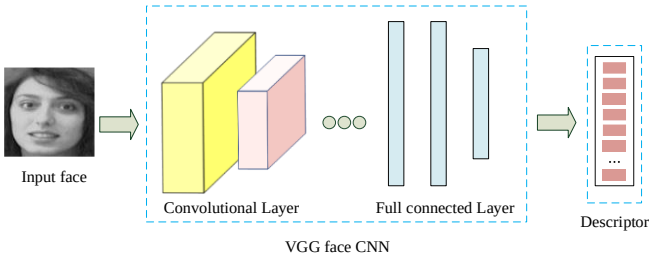


FIGURE 17. **Implementation of VGG Face descriptor.**

A. Experiment results

1) IMPACT OF IMAGE RESOLUTION ON RECOGNITION ACCURACY

We conduct several experiments to evaluate the impact of resolution on recognition accuracy (see Fig. 18). In the experiments, the face images in the ORL database are resized to different sizes, i.e., 10×10, 20×20, 40×40, 60×60, and 80×80. We conduct LBP and VGG Face algorithms on the faces. Figure 19 shows the recognition accuracy of the faces with different resolutions. From Fig. 19, it is observed that face recognition accuracy is correlated with face resolution. All of the face recognition methods achieve poor recognition accuracy for faces with a resolution of 10×10, due to the less information contained for these faces. Overall, recognition accuracy increases as resolution increases. The cause of this phenomenon is that high resolution faces could provide more information than low resolution faces. However, when the resolution of a face is higher than 40×40, recognition accuracy is stable. We can conclude that, for better face recognition accuracy, face resolution should be more than 40×40.

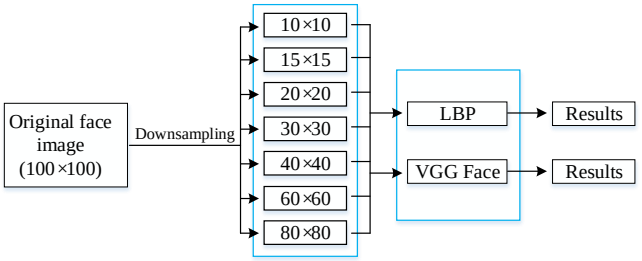


FIGURE 18. **Illustration of the experiment to evaluate the impact of resolution on recognition accuracy.**

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2018.2799861, IEEE Access

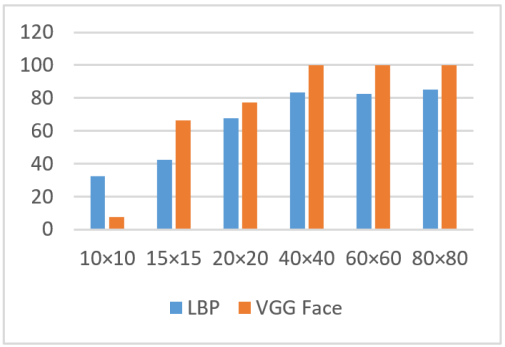


FIGURE 19. Recognition accuracy for faces at different resolutions.

2) COMPARISON OF SR METHODS' PERFORMANCE ON IMPROVING RECOGNITION ACCURACY

To determine the effectiveness of the SR method on face accuracy, we compare face recognition results by using different SR methods, i.e., Bicubic, TSC, SCDL, CDL, SRCNN, and VDSR (see Fig. 20). We resize the images to size 20×20 as LR images. The upscale factor is set to ×3. Visual outputs of the different SR methods and the original images are depicted in Fig. 21. From Fig. 21, it can be seen that, even though there are some artifacts in the face region for the results of the SR methods, the overall quality of the image has improved. In addition, the reconstructed images of the VDSR methods are very close to the original HR images.

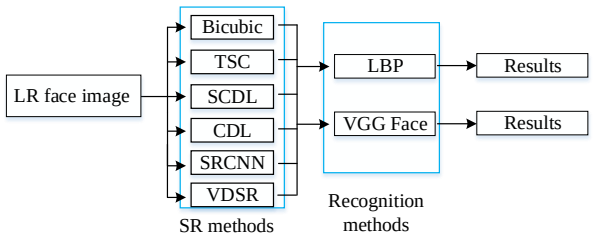


FIGURE 20. Illustration of the experiment for comparison of SR methods' performance on improving recognition accuracy.

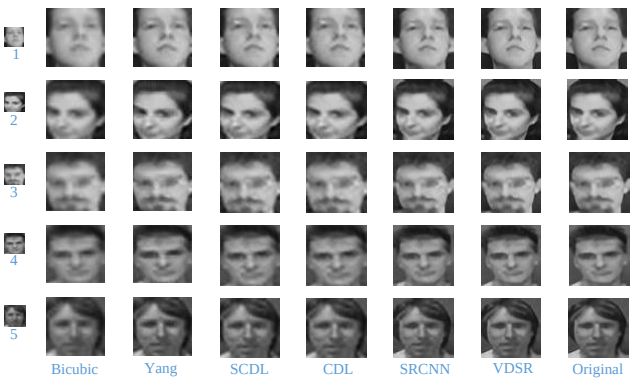


FIGURE 21. Visual comparison of different SR methods.

Face recognition results of different SR methods and the original HR images are shown in Fig. 22. The experimental results are presented in terms of recognition accuracy. It is observed that deep-learning-based methods outperform other methods. In addition, the VDSR method has the highest recognition accuracy and consistently outperforms the other SR methods. The VDSR method improves performance by 8.75% and 7.5% in accuracy compared with that of the Bicubic method by LBP and VGG Face algorithms, respectively. In fact, the recognition accuracy of reconstructed faces by using the VDSR method is almost as good as that of the original faces. We can thus conclude that reconstructed HR images are more similar to the original images, and the recognition accuracy of the reconstructed HR images is closer to that of the original images. Consequently, in monitor systems, high recognition accuracy can be obtained by employing SR methods to improve face resolution to solve the challenge of monitor system face recognition.

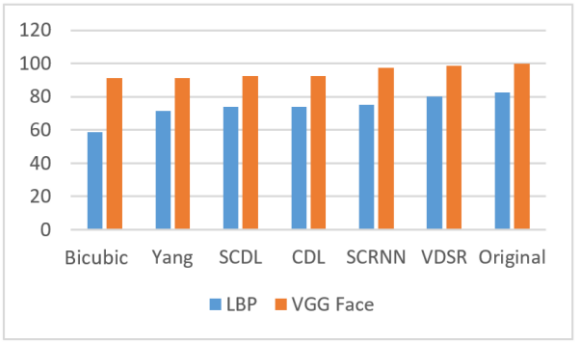


FIGURE 22. Recognition accuracy with different SR methods.

3) RELATION OF UPSCALE FACTOR WITH IMPROVEMENT IN RECOGNITION ACCURACY

We subsample the original face images to different sizes, i.e., 10×10, 15×15 and 20×20, to obtain low resolution face images. Then, we upscale the LR face images with different upscale factors, i.e., ×2, ×3, and ×4 (see Fig. 23). Visual outputs are shown in Figs. 24-26. SR results for LR images with a size of 20×20 have the best visual quality, and consistently compare well with the other SR results.

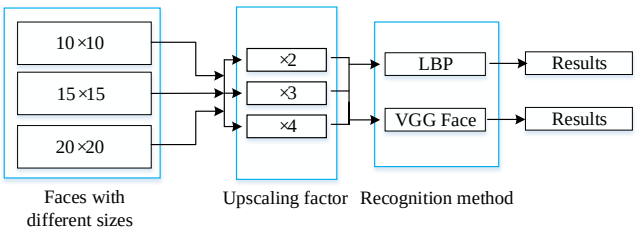


FIGURE 23. Illustration of the experiment to evaluate the relation of upscale factors with improvement in recognition accuracy.

Figs. 27-38 show the recognition accuracy of HR images from LR images with a size of 10×10 and 15×15. We find that accuracy of HR images is poorer that of the LR images. Consequently, we can conclude that, if resolution is too small, SR methods will fail to improve quality. This is because LR images with a size of 10×10 and 15×15 contain too little information. The SR method

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2018.2799861, IEEE Access

could include some incorrect information, which leads to poor performance.

Figs. 39-44 show the recognition accuracy of HR images from the LR image with a size of 20×20. We find that the accuracy of HR images is poorer than that of the LR images for the Bicubic method and sparse-coding-based methods. While accuracy of HR images is better than that of the LR images for deep- learning-based methods. Furthermore, we find that when we upscale up to three, the improvement of the recognition accuracy is stable for deep-learning methods. Deep-learning-based methods could improve recognition accuracy effectively and robustly.

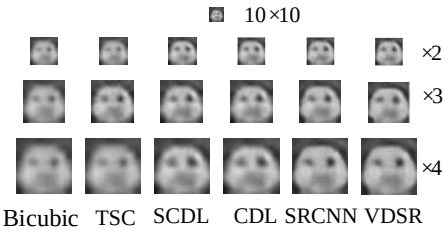


FIGURE 24. SR results of different upscale factors for an LR image with a size of 10×10.

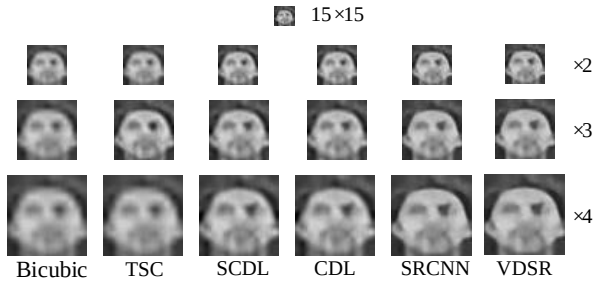


FIGURE 25. SR results of different upscale factors for LR images with a size of 15×15.

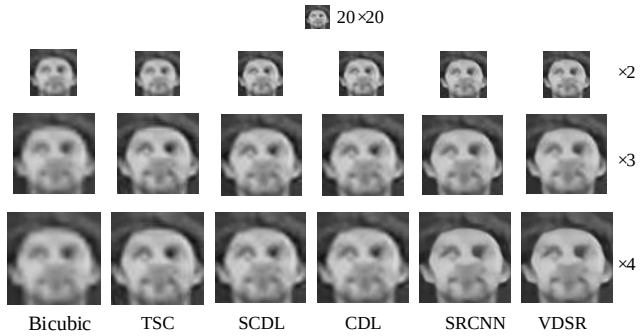


FIGURE 26. SR results of different upscale factors for an LR image with a size of 20×20.

IV. CONCLUSION

The proposed scheme uses super-resolution methods to improve recognition accuracy for monitor systems. Recognition is performed after super-resolution. We conduct experiments with six super-resolution methods and two recognition methods. We can conclude that, when the resolution is too low, no recognition method is successful. Reconstructed HR images are more similar to original images, and the recognition accuracy of reconstructed HR images is closer to that of the original images. Therefore, in monitor systems, we can increase recognition accuracy by employing the SR method to improve face resolution.

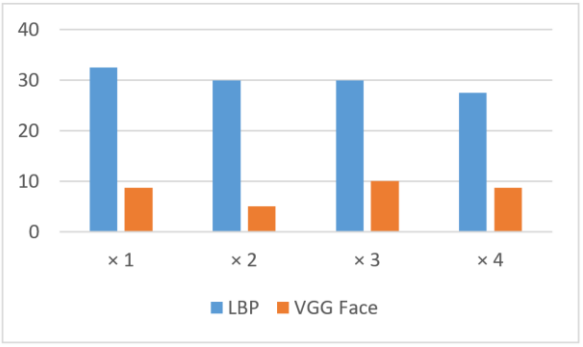


FIGURE 27. Recognition accuracy with different scales for faces of 10×10 by the bicubic method.

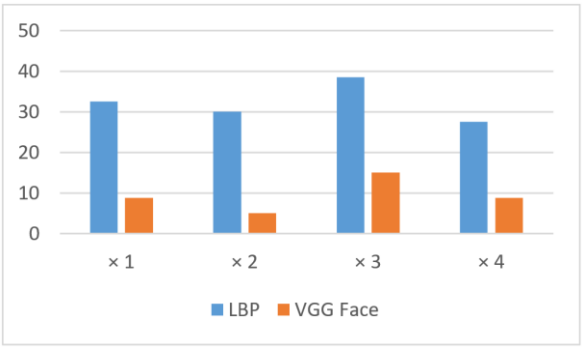


FIGURE 28. Recognition accuracy with different scales for faces of 10×10 by the TSC method.

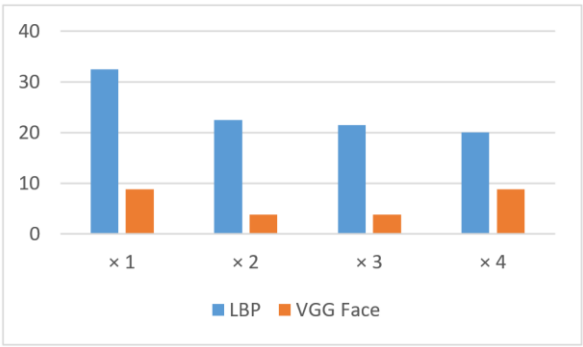


FIGURE 29. Recognition accuracy with different scales for faces of 10×10 by the SCDL method.

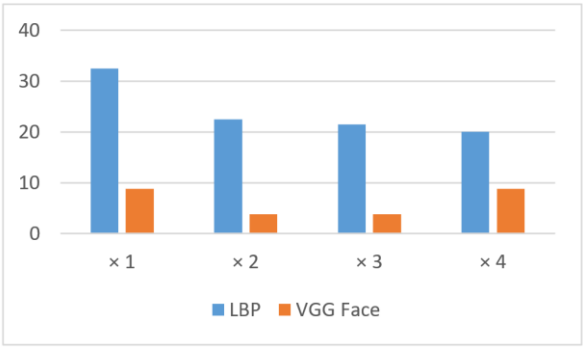


FIGURE 30. Recognition accuracy with different scales for faces of 10×10 by the CDL method.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2018.2799861, IEEE Access

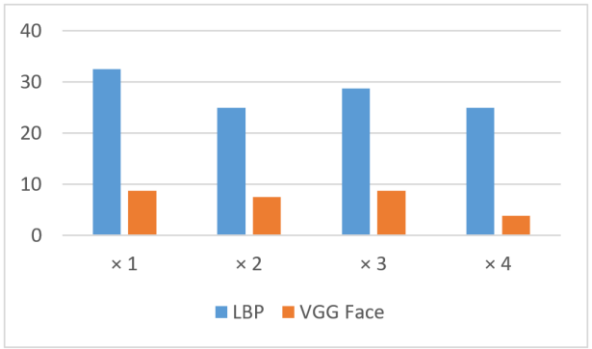


FIGURE 31. Recognition accuracy with different scales for faces of 10×10 by the SRCNN method.

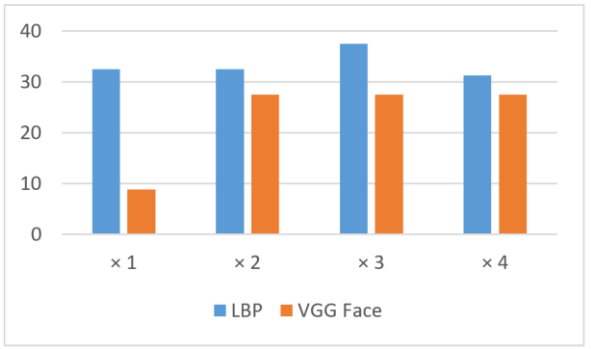


FIGURE 2. Recognition accuracy with different scales for faces of 10×10 by the VDSR method.

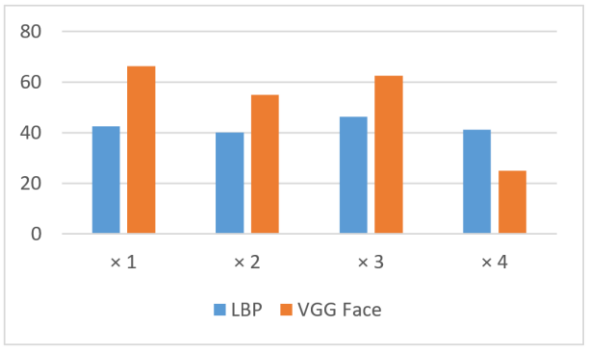


FIGURE 33. Recognition accuracy with different scales for faces of 15×15 by the bicubic method.

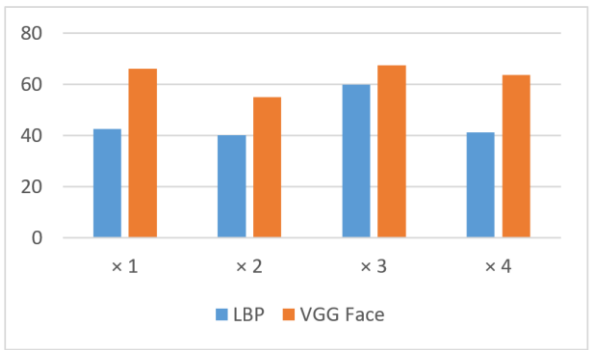


FIGURE 34. Recognition accuracy with different scales for faces of 15×15 by the TSC method.

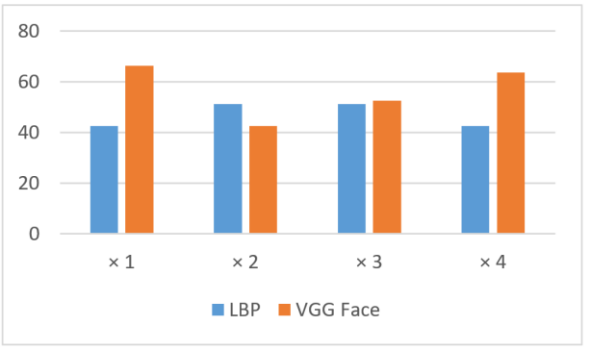


FIGURE 35. Recognition accuracy with different scales for faces of 15×15 by the SCDL method.

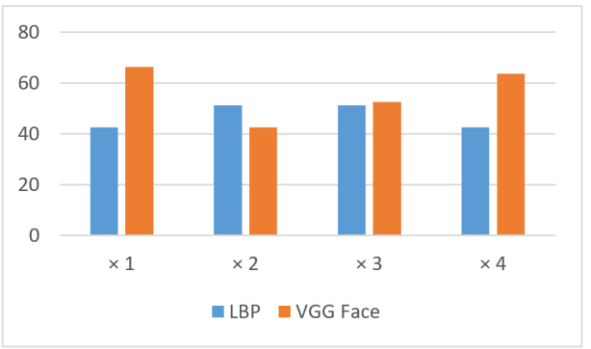


FIGURE 36. Recognition accuracy with different scales for faces of 15×15 by the CDL method.

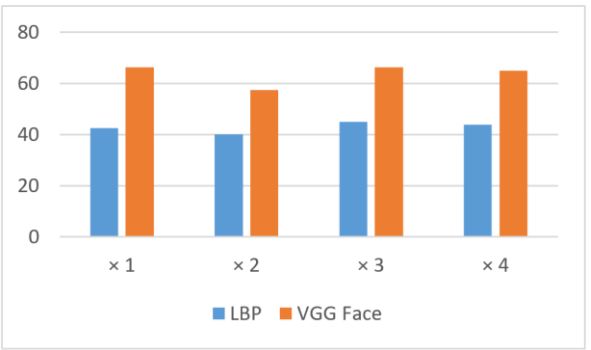


FIGURE 37. Recognition accuracy with different scales for faces of 15×15 by the SRCNN method.

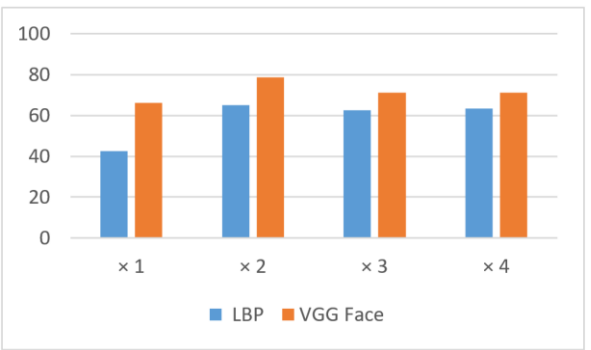


FIGURE 38. Recognition accuracy with different scales for faces of 15×15 by the VDSR method.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2018.2799861, IEEE Access

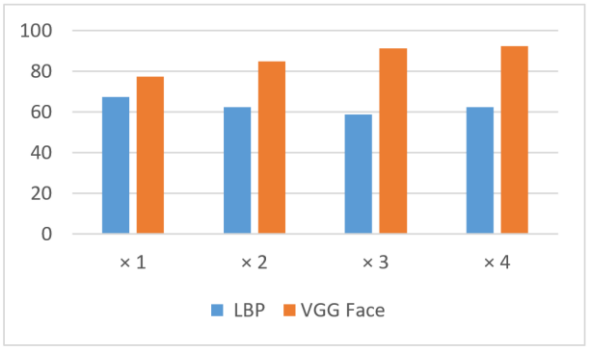


FIGURE 39. Recognition accuracy with different scales for faces of 20×20 by the bicubic method.

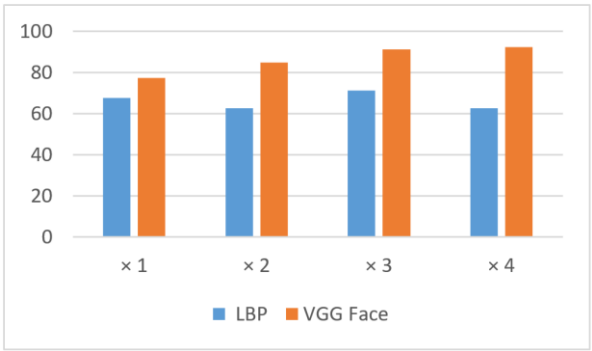


FIGURE 40. Recognition accuracy with different scales for faces of 20×20 by the TSC method.

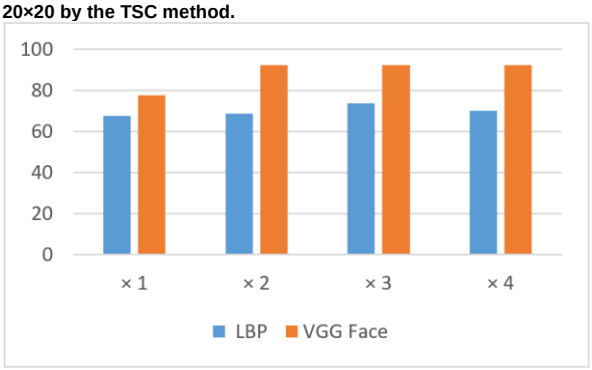


FIGURE 41. Recognition accuracy with different scales for faces of 20×20 by the SCDL method.

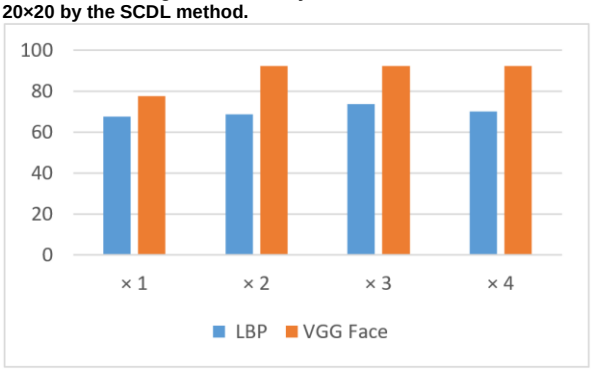


FIGURE 42. Recognition accuracy with different scales for faces of 20×20 by the CDL method.

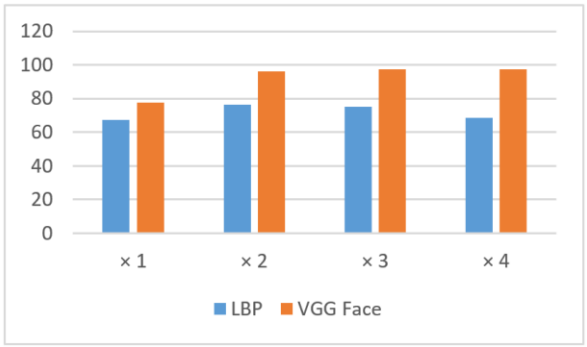


FIGURE 43. Recognition accuracy with different scales for faces of 20×20 by the SRCNN method.

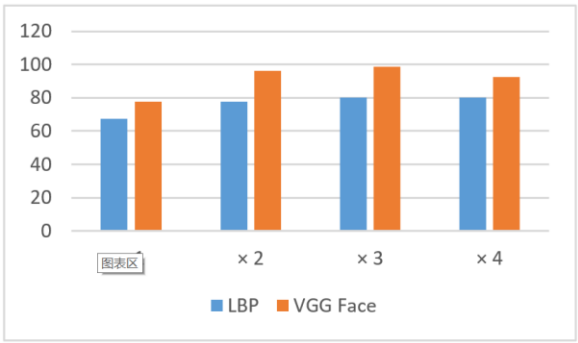


FIGURE 44. Recognition accuracy with different scales for faces of 20×20 by the VDSR method.

REFERENCES

[1] Cong, Dung Nghi Truong, C. Achard, and L. Khoudour. “People re-identification by classification of silhouettes based on sparse representation,” in Proc. IPTA, Paris, France, 2010, pp. 60-65.

[2] P.E. Forssen, “Maximally Stable Colour Regions for Recognition and Matching,” in Proc. IEEE CVPR, Minneapolis, MN, 2007,pp. 1-8

[3] X. Li, M. Orchard, “New edge-directed interpolation, ” *IEEE* Trans. Image Process. vol. 10, no .10, pp. 1521-1527.2001.

[4] L. Zhang, X. Wu, “An edge-guided image interpolation algorithm via directional filtering and data fusion,” *IEEE* Trans. Image Process. vol. 15, no.8, pp. 2226–2238. 2006.

[5] M. Li, T. Nguyen, “Markov random field model-based edge-directed image interpolation,” *IEEE* Trans. Image Process. vol. 17, no. 7, pp. 1121–1128. 2008.

[6] X. Li, Y. Hu, X. Gao, D. Tao, B. Ning, “A multi-frame image super-resolution method,” Signal Process. vol. 90, no. 2， pp. 405–414. 2010.

[7] Hirsch M , Harmeling S , Sra S , Sch ölkopf B . “Online multi-frame blind deconvolution with super-resolution and saturation correction,” Astron Astrophys. vol. 531, no. 3, p. A9, May 2011.

[8] S. Kim, W. Su, “Recursive high-resolution reconstruction of blurred multiframe images,” IEEE Trans. Image Process. vol. 2. no. 10. pp. 534–539. 1993.

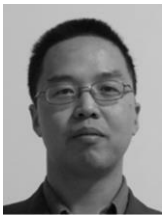
[9] H. Chang, D.Y. Yeung, Y. Xiong, Super-resolution through neighbor embedding, in:*Proc*. IEEE Conf. Comput. Vis. Pattern Recognit., 2004, pp. 275–282.

[10] K. Zhang, X. Gao, D. Tao, X. Li, “Single image super-resolution with sparse neighbor embedding,” *IEEE* Trans. Image Process. vol. 21, no. 7, pp. 3194–3205. 2012.

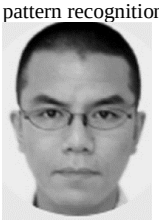
[11] C.-Y. Yang, M.-H. Yang, “Fast direct super-resolution by simple functions, “in: *Proc*. IEEE Int. Conf. Comput. Vis., 2013, pp. 561–568.

[12] J. Yang, Z. Lin, S. Cohen, "Fast image super-resolution based on in-place example regression," in: *Proc*. IEEE Int. Conf. Comput. Vis., 2013.
[13] K. Zhang, B. Wang, W. Zuo, H. Zhang, "Joint learning of multiple regressors for single image super-resolution, " *IEEE* Signal Process. Lett. vol. 23, no. 1, pp. 102–106. 2016.
[14] K. Zhang, D. Tao, X. Gao, X. Li, Z. Xiong, "Learning multiple linear mappings for efficient single image super-resolution," *IEEE* Trans. Image Process. vol. 24, no. 3, pp. 846–861. 2015.
[15] Yang J, Wright J, Huang T, Ma Y. "Image super-resolution as sparse representation of raw image patches," in: *Proc*. IEEE CVPR; 2008. p. 1–8.
[16] Yang J , Wright J , Huang TS , Ma Y . "Image super-resolution via sparse representation," *IEEE* Trans Image Process.vol. 19, pp.2861–73. 2010.
[17] R. Zeyde, M. Elad, and M. Protter." On single image scale-up using sparse-representations， " In Curves and Surfaces, pp.711–730. Springer, 2012.
[18] Wang S , Zhang L , Yan L , Quan P ." Semi-coupled dictionary learning with applications to image super-resolution and photosketch synthesis" in: Proc. IEEE CVPR; Providence, RI. 2012. pp. 2216–23 .
[19] Huang D A, Wang Y C F." Coupled Dictionary and Feature Space Learning with Applications to Cross-Domain Image Synthesis and Recognition，"in: Proc. IEEE ICCV, Sydney, Australia, 2013:2496-2503.
[20] C. Dong, C. Loy, K. He, X. Tang, "Learning a deep convolutional network for image super-resolution," in: *Proc*. European Conf. Comput. Vis., 2014. pp. 184–199.
[21] J. Kim, J. K. Lee, K. M. Lee, Accurate image super-resolution using very deep convolutional networks, in: Proc. IEEE CVPR , Las Vegas, NV, USA, 2016, pp.1646-1654.
[22] Baker, S. "Simultaneous super-resolution and feature extraction for recognition of low resolution faces," in: *Proc*. IEEE CVPR, Anchorage, AK, USA, 2008.pp.1-8.
[23] Parkhi, O. M. and Vedaldi, A. and Zisserman, A. "Deep Face Recognition," in: *Proc*. BMVC, Swansea, UK,2015
[24] Yang, X., Wu, W., Liu, K. et al. "Multi-sensor image super-resolution with fuzzy cluster by using multi-scale and multi-view sparse coding for infrared image,"*Multimed Tools Appl.* to be published. DOI: https://doi.org/10.1007/s11042-017-4639-4.
[25] Yang, X., Wu, W., Liu, K. et al. "Multiple dictionary pairs learning and sparse representation-based infrared image super-resolution with improved fuzzy clustering, "Soft Comput. to be published. DOI: https://doi.org/10.1007/s00500-017-2812-3

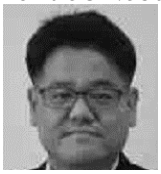
Xiaomin Yang is currently an Associate Professor in the College of Electronics and Information Engineering, Sichuan University. She received her B.S. degree from Sichuan University in 2002, and received her Ph.D. degree in communication and information systems from Sichuan University in 2007. She worked at the University of Adelaide as a post-doctorate for one year. Her research interests are image processing and pattern recognition.



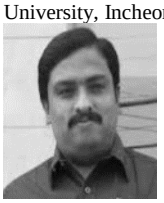
Wei Wu is currently a Professor in the College of Electronics and Information Engineering, Sichuan University. He received his B.S. degree from Tianjin University, and received his M.S. and Ph.D. degrees in communication and information systems from Sichuan University. He worked at the National Research Council Canada as a post-doctorate for one year. His research interests are image processing and pattern recognition.



Kai Liu is currently a Professor in the College of Electronics and Engineering Information, Sichuan University. He received his B.S. degree from Sichuan University, and received his Ph.D. degree from the University of Kentucky.



Pyoung Won Kim is currently an Associate Professor in the Department of Korea Language Education, Incheon National University, Incheon, Korea.



Arun Kumar Sangaiah is currently a Professor in the School of Computing Science and Engineering, VIT University, India.

Gwanggil Jeon received the B.S., M.S., and Ph.D. degrees from the Department of Electronics and Computer Engineering, Hanyang University, Seoul, Korea, in 2003, 2005, and 2008, respectively. From 2008 to 2009, he was with the Department of Electronics and Computer Engineering, Hanyang University, from 2009 to 2011. He was with the School of Information Technology and Engineering (SITE), University of Ottawa, as a post-doctoral fellow, and from 2011 to 2012. He was with the Graduate School of Science and Technology, Niigata University, as an Assistant Professor. He is currently an Assistant Professor with the Department of Embedded Systems Engineering, Incheon National University, Incheon, Korea. His research interests comprise image processing, particularly image compression, motion estimation, demosaicking, image enhancement, and computational intelligence, such as fuzzy and rough sets theories. He was the recipient of the IEEE Chester Sall Award in 2007 and the 2008 ETRI Journal Paper Award.