저해상도 영상 얼굴인식을 위한 전처리 방법

이필규*, 김태윤**, 이다솔*, 김성재†
*(주)에이치시티, 부설연구소, 경기도 이천시
**청도대원학교
미시시피주립대학교 수학과†
e-mail : legendpig85@gmail.com

Preprocessing Methods for Low-Resolution Face Image Recognition

Philku Lee*, Tai Yoon Kim**, Dasol Lee*, Seongjai Kim†
*Affiliated Research Institutes, HCT Co., LTD.
**Qingdao Daewon International School
†Dept. of Mathematics and Statistics, Mississippi State University

Abstract

Face recognition systems are characterized by low invasiveness of acquisition, and increasingly better reliability. Such systems may not be applied effectively, when the images are in low resolution (LR) as in the case that photos are taken from long distances, typically public surveillance. In theory, the high resolution (HR) image reconstructed from an LR face image, applying a super resolution (SR) method, can be used for face recognition. However, existing face SR algorithms may not give satisfactory results in face recognition. This article investigates the very low resolution face recognition problem and introduces a partial differential equation (PDE)-based SR method for a face recognition system of convolutional neural network (CNN).

요 약

얼굴인식 시스템은 비접촉데이터 채집의 특성과 함께, 그 정확도가 점차 향상되고 있다. 공공 감시카메라와 같이 사진을 멀리서 찍는 상황에서는 저해상도의 얼굴 이미지로 인해 얼굴인식 시스템을 효과적으로 사용할 수 없는 경우가 있다. 이론적으로는 저해상도영상을 Super Resolution (SR) 방법으로 고해상도 영상으로 바꾸어 얼굴인식에 사용할 수 있지만, 기존의 SR 방법들은 얼굴인식에 만족할만한 결과를 내지 못할 수 있다. 이 논문은 극 저해상도 (very low resolution) 얼굴인식 문제를 살펴보고 편미분방정식 기반 SR 방법을 제안하고, CNN 기반 얼굴인식 시스템에 응용한다.

1. Introduction

During the last decade or so, face recognition (FR) has become a popular research area in computer vision and one of the most successful applications of image analysis and pattern recognition. An FR system is a computer program capable of identifying a person from a digital image or a set of video frames, by statistically comparing selected facial features from the image and a face database [5, 6, 9]. It has been typically used in security systems and it has recently become popular as a commercial identification and marketing tool [4].

Nowadays more-and-more automatic access systems are based on biometric techniques. Face recognition systems are characterized by low invasiveness of acquisition, and increasingly better reliability. The main problem that occurred in such systems is low resolution of details, for example when the photos are taken from long distances, typically public surveillance. The identification and recognition from this kind of pictures has become an important scientific issue to solve [8, 11]. Super-resolution (SR) is a method to construct a high resolution (HR) image from its low-resolution (LR) image; theoretically, the HR image reconstructed from an LR face image (or, a set of LR images) can be used for FR. However, existing face SR algorithms may not give satisfactory results in FR.

This article investigates the low-resolution FR problem and introduces a partial differential equation (PDE)-based SR algorithm in order to enhance edges in reconstructed HR images and in turn improve accuracy of a FR system. The next section describes details of the PDE-based SR method and its effectiveness is verified in Section 3.

2. Image Enhancement

Magnified images may involve interpolation artifacts, particularly in a vicinity of fine structures such as edges. It is quite natural to try to eliminate such artifacts, from a point of view that edges carry important information for imagery. The artifacts can be viewed as a sort of noise, which can be eliminated by applying methods of image enhancement or SR. This section begins with a PDE-based edge-forming method which has been considered as a SR algorithm [1, 2].

2.1 Preliminaries: PDE-based edge-forming

Let $\hat{v}$ be the magnified image from the low-resolution image v_{LR} by utilizing one of interpolation methods. Then we can write

$$\hat{v} = u + e,$$

where u is the desired image (hopefully, having sharp and clear edges) and e denotes the noise most of which is arisen during the interpolation. Consider the following nonlinear semi-discrete model of the form

$$\frac{du}{dt} + \mathcal{A}u = \beta(\hat{v} - u), \qquad (2.1)$$

where β denotes a constraint parameter and $\mathcal{A} = \mathcal{A}(u)$ is a *diffusion* operator in a matrix form. That is, for $\mathcal{A} = (a_{rs})$,

$$a_{rs} \le 0, r \ne s;\ a_{rr} = \sum_{s \ne r} |a_{rs}| > 0, \forall r.$$

The recovered image u becomes closer to $\hat{v}$ as β grows. Note that the matrix elements depend on the solution u. As a matter of fact, the diffusion operator can be formulated as a finite difference (FD) approximation of the following gradient weighted curvature.

$$\mathcal{A}_u \approx \mathcal{K}(u) \equiv -|\nabla u|^{1+\delta} \nabla \cdot \left(\frac{\nabla u}{|\nabla u|^{1+\delta}}\right), \qquad (2.2)$$

where $\delta \ge 0$ is a user parameter. When $\delta > 0$, the diffusion operator $\mathcal{A}$ incorporated with (2.2) can form and sharpen edges. In order to illustrate it, we will consider in the following an example of edge-forming formulated in one dimensional (1D) intervals.

Denote the time step size by Δt and define $t^n = n\Delta t$ and $u^n = u(\cdot, t^n)$ for $n \ge 0$. Consider the explicit time-stepping method for (2.1):

$$u^n = u^{n-1} + \Delta t[-\mathcal{A}^{n-1}u^{n-1} + \beta(\hat{v} - u)], \qquad (2.3)$$

where

$$\mathcal{A}^{n-1}u^{n-1} \approx -|u_x|_\varepsilon^{1+\delta} \left(\frac{u_x}{|u_x|_\varepsilon^{1+\delta}}\right) \qquad (2.4)$$

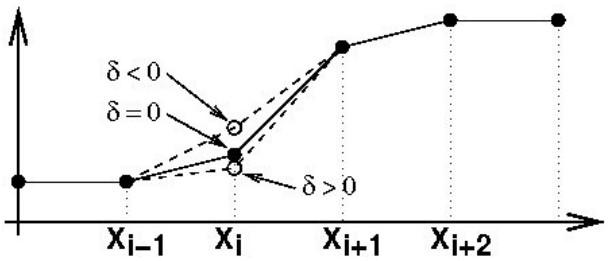
Here, in order to prevent the denominator from approaching zero, we have replaced $|u_x|$ with $|u_x| \equiv \sqrt{u_x^2 + \varepsilon^2}$, where ε is a small number.

Now, to formulate a spatial approximation, let x_i be the i-th pixel in the x-direction and $u_i^n = u(x_i, t^n)$. Then we construct the i-th row of $\mathcal{A}_1^{n-1}$ consisting of three consecutive non-zero elements which represent the connection of u_i^{n-1} to u_{i-1}^{n-1} and u_{i+1}^{n-1}:

$$[\mathcal{A}^{n-1}]_i = \left(-a_{i,W}^{n-1},\ 2,\ -a_{i,E}^{n-1}\right), \qquad (2.5)$$

where $a_{i,W}^{n-1} + a_{i,E}^{n-1} = 2$, of which coefficients are defined as

$$a_{i,W}^{n-1} = \frac{2d_{i,E}^{n-1}}{d_{i,W}^{n-1} + d_{i,E}^{n-1}}\ ,\ a_{i,E}^{n-1} = \frac{2d_{i,W}^{n-1}}{d_{i,W}^{n-1} + d_{i,E}^{n-1}}, \qquad (2.6)$$



(Figure 1) The edge-forming principle. The solid curve involving solid circles is the solution at the previous level u^{n-1} and the dashed curves indicate the solutions in the current level near the point x_i depending on δ.

with

$$d_{i,W}^{n-1} = [(u_i^{n-1} - u_{i-1}^{n-1})^2 + \varepsilon^2]^{\frac{1+\delta}{2}}, \qquad (2.7)$$

$$d_{i,E}^{n-1} = [(u_{i+1}^{n-1} - u_i^{n-1})^2 + \varepsilon^2]^{(1+\delta)/2}.$$

Note that $\mathcal{A}^{n-1}$ is a diffusion operator so that it can smears local extrema.

Let the solution u^{n-1} be concave at x_i and convex at x_{i+1}, as shown in Figure 1. When $\delta = -1$, for example, the scheme (2.5) turns out to be the central second-order approximation of $-\partial_{xx}$ (with the grid size equal to one). In this case, for the solution profile given as in Figure 1, we can see $u_i^n > u_i^{n-1}$ in (2.3), because $[\mathcal{A}_1^{n-1}u^{n-1}]_i < 0$. Furthermore, it is not difficult to check that $u_i^n = u_i^{n-1}$ for $\delta = 0$ and $u_i^n < u_i^{n-1}$ for $\delta = 1$, by setting $\beta = 0$ for simplicity and assuming that ε is sufficiently small. In general, we can see that at x_i in Figure 1,

$$u_i^n \begin{pmatrix} > \\ = \\ < \end{pmatrix} u_i^{n-1}, \text{ respectively for } \begin{cases} \delta < 0 \\ \delta = 0, \\ \delta > 0 \end{cases}$$

Exploiting the same arguments, we also can see that at x_{i+1} in Figure 1,

$$u_{i+1}^n \begin{pmatrix} > \\ = \\ < \end{pmatrix} u_{i+1}^{n-1}, \text{ respectively for } \begin{cases} \delta < 0 \\ \delta = 0, \\ \delta > 0 \end{cases}$$

We summarize the above observation as follows:

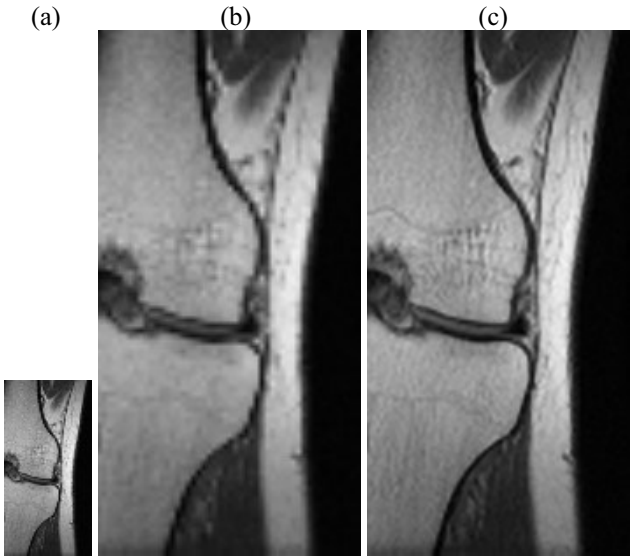
> *The algorithm* (2.3) *incorporated with the numerical schemes in* (2.5)-(2.7) *can not only reduce noise by smearing local extrema but also make edges in the image sharper for* $\delta > 0$.

Such a strategy will work for 1D problems and also for 2D images formulated with other time-stepping procedures such as the implicit and Crank-Nicolson methods.

2.2 The curvature interpolation method

Although the explicit procedure (2.3) can make edges sharper, it is a diffusion operator so that it may deteriorate image contents by trying to smear out local extrema. In order to reduce such artifacts, the algorithm can be reformulated

incorporating a curvature term estimated from the low



(Figure 2) Super resolution for MRI image (3-times magnification): (a) The original image, (b) the bilinearly interpolated image, and (c) the final image enhanced by the CIM. The original image is in 100×50 pixels and the magnified images are in 300 × 150 pixels.

resolution image.

$$u^n = u^{n-1} + \Delta t\left[-\mathcal{A}^{n-1}u^{n-1} + \beta(\hat{v} - u) + \widehat{\mathcal{K}}\right], \quad (2.8)$$

where $\widehat{\mathcal{K}}$ is a high-resolution interpolation of the gradient weighted curvature of the low-resolution image v_{LR}, $\mathcal{K}(v_{LR})$. The algorithm in (2.8) is called the curvature interpolation method (CIM). It has been numerically verified that the CIM is more effective than other state-of-the-art image zooming algorithms [3, 7].

2.3. Applications to medical image zooming

It is often the case that medical images acquired by CT and MRI scanners are saved in smaller sizes for a storage of long time. However, they must be resized (magnified) to be examined in detail later times. In this subsection, we will apply the CIM (2.8) for the magnification of medical images and show its effectiveness on edge-forming.

Figure 2 depicts MRI images magnified by 3-times in size; the original image (LR) is in 100 × 50 pixels and the magnified images are in 300 × 150 pixels. As one can see from the second subfigure, the HR image magnified by the bilinear interpolation method contains observable interpolation artifacts. The checkerboard effect is particularly vivid near around edges over whole image. The CIM can eliminate such artifacts in 28 steps of time-marching with δ = 0.1 and Δt = 0.25. The CIM has restored fine structures in the image as in Figure 2(c); edges are clear and sharp and texture components (as in the mid of figure) are well preserved.

In the last example, the original image is acquired by an MRI scanner for a diagnosis of osteonecrosis. Osteonecrosis (also known as Avascular necrosis) is a condition in which the bone dies due to a lack of blood flow. It occurs most often in the hip, but can develop in other joints in the body. If prompt treatment is not provided, osteonecrosis can cause the bone to completely degenerate and a serious form of arthritis may develop. Some of the tests that help with diagnosing osteonecrosis include MRI, X-rays, and CT scan. These tests allow doctors to check for damage in the bones and joints in various resolutions.

<Table 1> CNN Layer model from VGG16 [10].

layer	output shape	# of parameters
input(120 × 120 × 3)		
conv3-64	120 × 120 × 64	1792
conv3-64	120 × 120 × 64	36928
maxpool	60 × 60 × 64	
conv3-128	60 × 60 × 128	73856
conv3-128	60 × 60 × 128	147584
maxpool	30 × 30 × 128	
conv3-256	30 × 30 × 256	295168
conv3-256	30 × 30 × 256	590080
conv3-256	30 × 30 × 256	590080
maxpool	30 × 30 × 256	
conv3-512	15 × 15 × 512	1180160
conv3-512	15 × 15 × 512	2359808
conv3-512	15 × 15 × 512	2359808
maxpool	15 × 15 × 512	
conv3-512	7 × 7 × 512	2359808
conv3-512	7 × 7 × 512	2359808
conv3-512	7 × 7 × 512	2359808
maxpool	3 × 3 × 512	
FC-4096	1 × 1 × 4096	18878464
FC-4096	1 × 1 × 4096	16781312
FC-classes	1 × 1 × 25	258111
softmax		

3. Numerical Experiments with Face Recognition

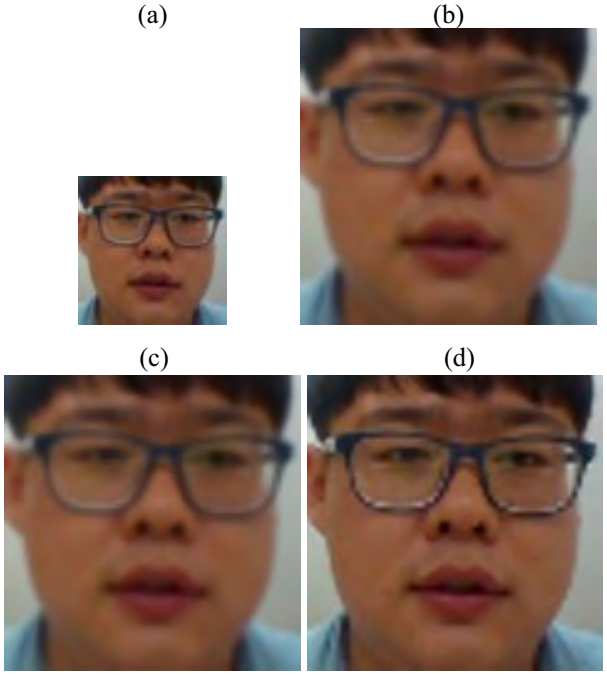
The CIM (2.8) written in C is incorporated with one of popular deep learning neural networks API written in Python, KERAS. For a comparison purpose, we also implement the median filter as another kind of image preprocessing algorithm.

We have produced our own face dataset for a CNN-based FR. A single face class is selected for simplicity of the prediction test and for training, 25000 train and 2500 validation images are utilized. In order to get a fine accuracy, we have constructed the CNN model from VGG16 [10] whose layer model is summarized in Table 1 and the batch size is set 200.

In Figure 3, we present a typical example of SR applied to the face data. When the LR image is bilinearly interpolated, the resulting image becomes blurry as shown in Figure 3(b). The median filter is also considered in order to reduce any noise involved in the HR image; its result is yet blurry as seen in Figure 3(c). On the other hand, the CIM can enhance the result utilizing the bilinearly interpolated image and resulting in a sharper image as depicted in Figure 3(d).

Table 2 gives the performance of our CNN model. The model is trained with 25 face classes in 25000 face images, each of which class is consist of 1000 images having 120 × 120 pixels as in Figure 3. For prediction test, 1000 LR single face images (in 60 × 60 pixels) are selected with no connection to train and validation images and magnified by a factor of 2 with three different ways: the bilinear interpolation, additional median filtering, and the CIM SR. As one can see from the first group of rows in Table 2, the CIM improves the prediction accuracy from 0.976 to 0.981, while the simple median filtering deteriorates it in the second

group of rows. In order to investigate the effect of SR to FR systems further, we downsize the LR images to 30 × 30 and 15 × 15 pixels and magnify them back to be tested. The three methods perform in a similar way; the lower the test image resolution is, the worse



(Figure 3) Face SR (2-times magnification): (a) The original LR image, (b) the bilinearly interpolated image, (c) the denoised image by the median filter, and (d) the enhanced image by the CIM. The original LR image is in 60 × 60 pixels and the HR images are in 120 × 120 pixels.

the accuracy is. However, the problem of very low resolution FR is not so much serious which is related to median filter and CIM SR methods can improve its accuracy somewhat.

4. Conclusions

Face recognition systems are becoming popular with increasingly better reliability. However, the main problem that occurred in such systems is low resolution of details as in the case that the photos are taken from long distances. In this article, we have investigated the very low resolution face recognition (FR) problem and introduced a partial differential equation (PDE)-based super resolution (SR) method. The problem of very low resolution FR is not so much serious in a CNN-based FR system and SR methods can improve its accuracy somewhat.

Reference

[1] Y. Cha and S. Kim, "Edge-forming methods for color image zooming," IEEE Trans. Image Process., vol. 15, no. 8, pp. 2315–2323, 2006.

[2] ——, "Edge-forming methods for image zooming," J. Math. Imaging and Vis., vol. 25, no. 3, pp. 353–364, 2006.

[3] Y. Cha, G. Y. Lee, and S. Kim, "Image zooming by curvature interpolation and iterative refinement," SIAM J. Imaging Sciences, vol. 7, no. 2, pp. 1284–1308, 2014.

[4] Consumer Reports, "Facial recognition: Who's tracking you in public?" 2015, (December 30, 2015).

<Table 2> Performance of the CNN model associated with the layer model (Table 1) and the face class (Figure 3): The total number of test images is 1000.

	Captured resolution	# of wrongs	accuracy
bilinear	60 × 60	24	0.976
median		22	0.978
CIM		19	0.981
bilinear	30 × 30	28	0.972
median		29	0.971
CIM		26	0.974
bilinear	15 × 15 (very low resolution)	44	0.956
median		44	0.956
CIM		41	0.959

[5] W. Deng, J. Hu, J. Guo, W. Cai, and D. Feng, "Robust, accurate and efficient face recognition from a single training image: A uniform pursuit approach," Pattern Recognition, vol. 43, no. 5, pp. 1748–1762, 2010.

[6] I. Kemelmacher-Shlizerman and R. Basri, "3d face reconstruction from a single image using a single reference face shape," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 33, no. 2, pp. 394– 405, 2011.

[7] H. Kim, Y. Cha, and S. Kim, "Curvature interpolation method for image zooming," IEEE Trans. Image Process., vol. 20, no. 7, pp. 1895–1903, 2011.

[8] G. Li, G. Sun, and X. Zhang, Robust Face Recognition in Low Resolution and Blurred Image Using Joint Information in Space and Frequency. Berlin, Heidelberg: Springer Berlin Heidelberg, 2012, pp. 616–624.

[9] Z. Liu, J. Yang, and C. Liu, "Extracting multiple features in the CID color space for face recognition," IEEE Transactions on Image Processing, vol. 19, no. 9, pp. 2502–2509, 2010.

[10] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," arXiv preprint arXiv:1409.1556, 2014.

[11] W. W. Zou and P. C. Yuen, "Very low resolution face recognition problem," IEEE Transactions on Image Processing, vol. 21, no. 1, pp. 327–340, 2012.