Liu YL, Xu XG, Guo YW *et al.* Pores-preserving face cleaning based on improved empirical mode decomposition. JOURNAL OF COMPUTER SCIENCE AND TECHNOLOGY 24(3): 557–567 May 2009

Pores-Preserving Face Cleaning Based on Improved Empirical Mode Decomposition

Yan-Li Liu[1,2] (刘艳丽), Xiao-Gang Xu (徐晓刚)[2], Yan-Wen Guo[3] (郭延文), Jin Wang[1] (王 进)
Xin Duan[1] (段 鑫), Xi Chen[1] (陈 曦), and Qun-Sheng Peng[1,2] (彭群生), *Senior Member, CCF*

[1] *State Key Lab of CAD & CG, Zhejiang University, Hangzhou 310027, China*
[2] *Department of Mathematics, Zhejiang University, Hangzhou 310027, China*
[2] *Department of Equipment Automatization, Dalian Naval Academy, Dalian 116026, China*
[3] *State Key Lab of Novel Software Technology, Nanjing University, Nanjing 210000, China*

E-mail: liuyanli@cad.zju.edu.cn; ywguo@nju.edu.cn; jwang@cad.zju.edu.cn

Received May 29, 2008; revised February 28, 2009.

Abstract In this paper, we propose a novel method of cleaning up facial imperfections such as bumps and blemishes that may detract from a pleasing digital portrait. Contrasting with traditional methods which tend to blur facial details, our method fully retains fine scale skin textures (pores etc.) of the subject. Our key idea is to find a quantity, namely normalized local energy, to capture different characteristics of fine scale details and distractions, based on empirical mode decomposition, and then build a quantitative measurement of facial skin appearance which characterizes both imperfections and facial details in a unified framework. Finally, we use the quantitative measurement as a guide to enhance facial skin. We also introduce a few high-level, intuitive parameters for controlling the amount of enhancement. In addition, an adaptive local mean and neighborhood limited empirical mode decomposition algorithm is also developed to improve in two respects the performance of empirical mode decomposition. It can effectively avoid the gray spots effect commonly associated with traditional empirical mode decomposition when dealing with high-nonstationary images.

Keywords image enhancement, empirical mode decomposition, normalized local energy

1 Introduction

Facial skin not only exhibits various fine scale details, e.g., pores, but also sometimes exhibits distractions such as blemishes, bumps, acne scarring etc. From many existing photo enhancement softwares that can be used to remove facial distractions, a common characteristic is observed: the resulting face is often over-smoothed, of which the natural pores disappear completely. However, human skin is perception sensitive, in which fine scale texture accounts for an important part in overall appearance. Over-smoothed face not only looks unnatural but also dissolves the individuality of a person. In this paper, we focus our attention on retaining fine skin texture features, e.g., pores structure of the skin, while simultaneously removing facial imperfections such as scars, bumps and blemishes.

Since the pores and some distractions of imperfect faces all belong to high frequency component of images, to preserve the former and remove the latter make face cleaning a challenging problem. For approaches to eliminating distractions with denoising techniques[1−3], which are widely adopted by commercial software, facial details such as pores are often accounted as noise and thereby removed.

In this paper, based on empirical mode decomposition[4] (EMD), we set a quantity, i.e., normalized local energy (NLE), to reveal the characteristics of fine scale and distractions. Building upon NLE, we further develop a quantitative characterization of facial skin appearance, namely imperfect degree, to depict the visual perception of facial skin. Finally, we employ imperfect degree as a guide to enhance facial skin. In addition, we also propose an algorithm called adaptive local mean and neighborhood limited empirical mode decomposition (ALNEMD) to reduce the gray spots effect commonly associated with traditional EMD when dealing with high nonstationary images.

Regular Paper

This work is supported by the National Natural Science Foundation of China under Grant Nos. 60403038 and 60703084, the Natural Science Foundation of Jiangsu Province under Grant No. BK2007571, and the Natural Science Foundation of Liaoning Province under Grant No. 20082176.

In the process of facial image cleaning, our algorithm only requests the user to tune a few high level, intuitive parameters to interactively control the amount of enhancement. The advantage of our technique is that while effectively removing facial imperfections, it does not blur fine scale facial details. Another important feature of our technique is: it is a general model which characterizes both imperfections and facial details in a unified framework, thereby it does not require user to interactively mark imperfections on facial images.

The rest of this paper is structured as follows. Section 2 reviews some related work about face cleaning and EMD. Section 3 describes our improved EMD algorithm. The algorithm of face cleaning is introduced in detail in Section 4. Experimental results are demonstrated in Section 5. Finally, Section 6 concludes the whole paper and highlights future work.

2 Related Work and Background

2.1 Related Work

Face Cleaning. The problem of facial image editing has received much attention both from the computer graphics and image processing communities. As a result, many techniques have been developed to achieve various facial appearance effects. Leyvand *et al.*[5] proposed an approach to enhanceing the facial attractiveness by adjusting facial features. Nguyen *et al.*[6] presented a layer extraction method to remove and synthesize beard. Using reflectance transfer, Peers *et al.*[7] produced face relighting in the post-production process. Most recently, Bitouk[8] introduced an algorithm that can automatically replace faces in photographs. Moreover, the algorithm also allows the user to interactively edit the illumination and colors of face.

However, little research work has been published on face cleaning in literature. Nevertheless, in practice, several methods have been employed to clean or enhance facial images. While cleaning face with denoising techniques is popular, an alternative approach to serving this problem is the so-called interactive cut-and-paste method, of which poison image editing[9] and healing brush in Adobe Photoshop are well-known. This approach repairs imperfections on the facial images by seamless cloning, with little effects on preserving skin details. Moreover, for facial images with many imperfections, it would incur a lot of user interactions to mark out source and destination areas. Since a great deal of skill is demanded to achieve satisfactory results, it is a suitable tool only for trained designers. Our goal is to develop a technique with a few high level, intuitive parameters that can be easily adopted by naive users.

In the field of face recognition, Lin *et al.*[10] used scale-invariant feature transform (SIFT) framework to detect irregular skin details. Later, Pierrard *et al.*[11] employed 3D morphable model reconstruction to recognize facial moles. Their techniques are effective in detecting moles with relatively fixed pattern, for example, circular shape. However, as these methods detect and localize facial moles in spatial domain, they are not suitable for detecting and localizing general skin irregularities, due to a large number of possible spatial variations of distractions. Instead of explicitly detecting and localizing distractions, our new approach suggests a quantitative characterization of both pores and distractions, and uses it to enhance facial skin.

Among the current image filtering techniques, bilateral filtering[1] is one of the most powerful filters and has been used in various fields. For the typical case in which the spatial and intensity weighting functions are Gaussian, there are two important parameters, namely geometric spread and photometric spread. While small geometric spread corresponds to filtering small intensity changes, large photometric spread would preserve edges with large discontinuity. For face cleaning, of which the aim is to smooth large discontinuity and preserve small scale texture, the settings for the two parameters are difficult: when the photometric spread is set to a very large value in order to smooth imperfections, the bilateral filter nearly degenerates into a Gaussian filter. In such a case, the smoothing effect of bilateral filter only relies on the geometric spread. Apparently, the geometric spread cannot be set to a large value, otherwise the pores of skin will be removed. On the other hand, the edges of bumps or scars in the skin are left aside if it is set to a small value.

To the best of our knowledge, the most related work to ours is developed by Matsui *et al.*[12] This work also aims at removing spots and at the same time preserving skin natural roughness. The method uses ε-filters to decompose image into several different frequency components. It assumes spots are of medium frequency and pores of high frequency, and then discards medium frequency component and retains the high frequency component. The major drawback of this approach is: since spots may exist also in low frequency and high frequency components, it cannot remove spots completely. Moreover, the discarded layers are determined by several parameters, which are not intuitive, hence requiring heavy user interactions. The users also need to use unsharp mask to avoid image blurring, which might magnify image noise.

EMD. Traditional energy-frequency analysis are based on Fourier transformation and wavelet

transformation. The Fourier transform is designed to work with liner and stationary signals. The wavelet transform, on the other hand, is well-suited to handle non-stationary data, but is poor at processing nonlinear data. However, few energy-frequency data are truly linear and stationary. To address this problem, the Hilbert-Huang transform (HHT) has been recently developed[4]. The HHT comprises two steps: EMD generates a finite number of intrinsic mode functions (IMF), and then the Hilbert transform is applied to IMF. The combination of IMF and its Hilbert transform forms an analytic signal, which can be used to generate a "time-frequency-energy" representation of the data.

In this paper, we choose EMD and HHT rather than wavelet transform due to the following considerations. First, imperfect facial images are in general nonlinear and nonstationary in nature, EMD is suitable to process these data. Second, based on the local time scale of the data, EMD decomposes a signal into IMFs adaptively without using a prior basis which do not necessarily match the varying nature of the signals. Without the need of pre-specifying a decomposition level as requested by wavelet decomposition to extract out all the image's oscillatory modes, EMD allows us to use all the oscillatory modes to further define imperfect degree. Thus, by exploiting EMD, we can obtain adaptive multi-resolution representation of image. Due to unpredictable appearances of various kinds of distractions, such an adaptive multi-resolution representation is favorable. Third, the IMF is generally in good agreement with intuitive and physical signal interpretations, and has well-defined instantaneous frequency, allowing us to define local energy to capture the characteristics of distractions and pores. On the contrary, the local energy extracted by Fourier or wavelet[13−15] suffers from the problem of energy leakage, due to the infinite or limited finite length of the basis. The energy leakage makes the quantitative definition of the energy-frequency-space distribution of image difficult.

Due to the nature of image data, when processing images with idea of EMD, bidimensional empirical mode decomposition (BEMD) is necessary. Accordingly, to perform 2D space-frequency analysis of IMFs, Riesz transform, an isotropic generalization of the Hilbert transform to multiple dimensions is also needed[16]. For BEMD, existing methods can be divided into two categories. The first category of the approaches divide an image into one-dimensional data. Then the 1D EMD is applied to a limited number of orientations: two (horizontal and vertical) or more[17]. For simplicity, we call it directional BEMD. The methods of the second category are genuine sense 2D EMD which use radial basis functions (RBF) or cubic spline to interpolate envelopes[18−20]. Although they have been successfully used in some cases, these algorithms suffer from the problem of overshoot or undershoot because of their interpolation scheme. In this paper, we propose two solutions, namely limiting minimum frequency and adaptive local mean, to improve the performance of EMD.

2.2 Background of EMD

The principle of EMD is to identify the intrinsic oscillatory modes embedded in the signals based on their characteristic time scales, and then to decompose the signals into a sum of IMFs. In order to permit physically meaningful instantaneous frequency to be defined over it, an IMF must satisfy the following two conditions[4]: first, the numbers of extrema and zero-crossings must differ by at most one; second, it is symmetric with respect to the local mean of the data. Ideally, the requirement of the second condition should be "the local mean of the data is zero". Since it is difficult to define a local averaging time scale in general case, in practice, the local mean of the data is usually replaced by the mean of the envelopes defined by interpolating the local maxima and local minima.

EMD is an iterative process. To illustrate its concept, here we describe the algorithm of EMD in 1D case. Given a signal $x(t)$, the algorithm of EMD can be summarized as follows.

1) Identify all the local extrema of $x(t)$.

2) Interpolate between local minima (resp. local maxima), ending up with the envelope $e_{\min}(t)$ (resp. $e_{\max}(t)$).

3) Compute the mean $m(t) = (e_{\min}(t) + e_{\max}(t))/2$.

4) Extract the detail $d(t) = x(t) - m(t)$. Check property of $d(t)$:

- if it meets the above two conditions, an IMF is derived, and replace $x(t)$ with the residual $r(t) = x(t) - d(t)$;
- if it is not an IMF, replace $x(t)$ with $d(t)$.

5) Repeat Steps 1)∼4) until the residual satisfies some stopping criterion.

Finally, $x(t)$ is represented as the sum of a finite number of IMFs (both amplitude and frequency modulated) and a residual trend. The process of iteratively extracting an IMF is also called sifting process. One can observe from the above formulation that there is no prefixed basis involving in the sifting process, instead, the decomposition proceeds depending on the data itself.

3 Improved EMD

Theoretically, there are two factors incurring the drawback of classic EMD mentioned in the above section. First, since there is no bandwidth constraint during the decomposition process in classical EMD, over-iterating on the whole signal for a better local approximation has the drawback of contaminating other parts of the signal, in particular, in uniformizing the amplitude and in over-decomposing it by spreading out its components over adjacent modes[21]. Second, classical EMD uses the mean of envelopes interpolated by the local maxima and minima respectively as a substitute for the local mean of data. Unfortunately, the interpolation method, for example spline fitting, often overshoots or undershoots the strongly nonlinear and nonstationary data. Such overshoots and undershoots would be amplified through many times of spline fitting in the sifting process, and eventually result in divergence of IMF. Therefore, unless a better approximation to local mean is conducted, EMD cannot be improved significantly.

For each of the above two problems, we propose a solution to get around it. The first solution is related with time-frequency uncertainty principle. As it is known, uncertainty principle formulated as

$$\nabla f \times \nabla t \geqslant 1/2 \tag{1}$$

defines the relation between the width of frequency window ∇f and that of the time window ∇t. According to (1), if we set $\nabla t \leqslant a$ $(a > 0)$, then it must have $\nabla f \geqslant b$ $(b = 1/2a)$. That is, if we limit the maximal temporal resolution, a minimal frequency resolution can be confined correspondingly. In the paper, we specify a maximum neighborhood in sifting process, as a result, the minimal frequency of the resulting IMF is defined. Our solution to the second problem is: instead of endeavoring to develop a fitting method which always cause overshoot or undershoot, we directly use the local mean of the image data to ascertain symmetry. To do so, the only issue is to define a local space scale. Fortunately, the above specified maximum neighborhood is a good candidate for defining such a local space scale. Taking computation cost and locality into account, an adaptive local mean algorithm is developed.

Integrating the above two strategies into one framework, for a given image I with size of $M \times N$ in pixels, our ALNEMD algorithm is formulated as the following steps.

1) Initialize $r_0(p) = I(p)$.

2) Extract k-th IMF imf_k.

• Set the size of maximum neighborhood $S_k \times S_k$, $S_k = 2^{k+1} + 1$.

• Initialize $h_{k,0}(p) = r_{k-1}(p)$, $l = 1$.

• Detect all local extreme points of $h_{k,l-1}$. Establish the local maximum point set $\mathscr{A}$ and local minimum point set $\mathscr{B}$ respectively. For every pixel p of $h_{k,l-1}$, if $I(p)$ is strictly higher or lower than all its 8-connected neighbors, it is a local extreme point. Otherwise, it is not a local extreme point.

• Compute the geometric distance (Euclidean distance) between every two adjacent local maximum points row-wisely and then column-wisely in $\mathscr{A}$. For example, suppose two adjacent local maximum points in the same column are (x_1, y), (x_2, y) and $x_1 < x_2$, then the distance S between them is $x_2 - x_1$. If $S > S_l$, then put the pixels $(x_1 + i \times S_l, y)$, $i = 1, 2, \ldots, \lfloor (S/S_l) \rfloor$ into $\mathscr{A}$. The purpose of supplementing $\mathscr{A}$ is to make the distance between every two adjacent local maximum points not more than S_l, ensuring the distribution of extreme points is not too sparse. Furthermore, to reduce the boundary effects, we add extreme surrounding the original image boundaries by mirror symmetry. The same operations are also performed on $\mathscr{B}$.

• For every pixel p, compute its local mean $e_{\text{mean},l-1}(p)$ using the *adaptive local mean algorithm* described below.

• $h_{k,l}(p) = h_{k,l-1}(p) - e_{\text{mean},l-1}(p)$.

• If the stop criterion is met[4], $imf_k(p) = h_{k,l}(p)$, and stop the sifting process; else $l = l + 1$.

3) $r_k(p) = r_{k-1}(p) - imf_k(p)$.

4) If the $r_k(p)$ is a monotonic or constant signal, terminate the decomposition process. Otherwise go to Step 2).

The image I is finally expressed as the sum of finite IMFs and a residue, as shown in the following equation:

$$I(p) = \sum_{k=1}^{K} imf_k(p) + r_K(p). \tag{2}$$

The adaptive local mean algorithm is shown in Table 1: the function *NumExtrema*(p) returns the number of local extreme points in $\Omega_{p,T}$. *Symmetry*(p) tests whether the local extreme points in $\Omega_{p,T}$ are spatially symmetric, and is defined as follows. The column and row of p equally divide $\Omega_{p,T}$ into four quadrants. *Symmetry*(p) is true if the following two conditions are both satisfied: first, in each of the four quadrants one local extreme point can be found; second, among these four local extreme points, two of them are local maximum points and the other two are local minimum points. Otherwise *Symmetry*(p) is false. The threshold of μ_T is used to guarantee that the number of local extreme points is not too few and typically set

as $(T^2-1)/3$.

Compared with interpolation-based EMD, in our adaptive local mean algorithm, the approximation to the local mean of image data is much more accurate, therefore the problem of overshoot or undershoot of interpolation-based EMD is reduced. On the other hand, by setting a maximum neighborhood in sifting process, the minimal frequency of IMF is kept under control and hence image overdecomposition is prohibited. Integrating both adaptive local mean algorithm and neighborhood limitation schemes, more accurate frequency and amplitude can be achieved by our method, laying a foundation to make quantitative energy-space analysis of facial images.

Table 1. Adaptive Local Mean Algorithm

The l-th Decomposition in Sifting Process
for every pixel p of $h_{k,l-1}$
  set $T=3$, *flag* = *false*, μ_T;
  let $\Omega_{p,T}$ be the $T\times T$ neighborhood of p;
  while $T<S_l$
    search for all local extreme points in $\Omega_{p,T}$;
    if *NumExtreme*$(p)>\mu_T$ and *Symmetry*(p) == *ture*
      $e_{\text{mean}}(p)=\frac{1}{T\times T}\sum_{q\in\Omega_{p,T}}h_{k,l-1}(q)$;
      flag = *true*;
      break;
    else
      $T=T+2$;
    end if
  end while
  if *flag* == *false*
    $e_{\text{mean}}(p)=\frac{1}{S_l\times S_l}\sum_{q\in\Omega_{p,T}}h_{k,l-1}(q)$;
  end if
end for

4 Algorithm of Face Cleaning

We will present in this section our face cleaning algorithm. First, we give an overview of the algorithm. Formally, for the facial image I to be processed, we first use ALNEMD to decompose it into a series of both frequency and amplitude modulated IMFs. Next, Riesz transform is applied to every IMF, generating a corresponding monogenic signal[22] which is the extension of the analytical signal from 1D to 2D. Built upon monogenic signals, the meaningful local energy of every pixel is extracted. By analyzing normalized local energy, we propose a function to quantitatively measure the imperfect degree of image. The cleaning result is finally produced after the coefficients of IMFs are adaptively adjusted under the guide of imperfect degree.

4.1 Extract the Local Property

Since our characterization and cleaning operation are only performed on skin parts of face, for a given image, we first need to know which regions on the image are skin parts. This information can be obtained either by requiring users to indicate the skin regions, or by using facial feature detection algorithms to automatize the process. In the paper, we manually mark out the contour of the face, eyes and lips that are essential to the character of individuals which we do not intend to adjust. In the following, the operations are all performed on the skin parts. Assuming that we have decomposed I into K IMFs with the ALNEMD algorithm described in the above section, now we apply Riesz transform to each IMF.

Defining $x=(x_1,x_2)^{\mathrm{T}}$, $x\in\Omega$, Ω is the pixel set of I, the Riesz kernel in the spatial domain is given as $R(x)=x/2\pi|x|^3$. We can see from the kernel that the integration in the Reisz transform is extremely local, which is a crucial requirement for analyzing non-stationary signals. The combination of a signal and its Riesz transformed is called the monogenic signal. Let $f_k(x)$ be the k-th IMF of the image, the corresponding monogenic signal thus takes the following form:

$$\begin{aligned} f_{k,M}(x) &= f_k(x)+(R*f_k)(x)=f_k(x)+f_{k,R}(x)\\ &=\int_{\Omega}(\delta(t)+t/2\pi|t|^3)f_k(x-t)\mathrm{d}t. \end{aligned} \tag{3}$$

From the monogenic signal, meaningful local amplitude and local phase can be extracted. The local amplitude of $f_{k,M}(x)$ is:

$$\begin{aligned} |f_{k,M}(x)| &= \sqrt{f_{k,M}^2(x)}\\ &=\sqrt{f_k^2(x)+|f_{k,R}(x)|^2}. \end{aligned} \tag{4}$$

As local amplitude offers image's energetic information, we use the square of local amplitude, namely local energy, to characterize the oscillatory degree of every pixel x at level k.

$$LE_k(x)=|f_{k,M}(x)|^2. \tag{5}$$

According to the theory of EMD, with the increase of k, $LE_k(x)$ decreases. In order to study the behavior of $LE_k(x)$, for a given pixel x, across different levels k in a unified fashion, we normalize $LE_k(x)$, $x\in\Omega$ at every level, getting the normalized local energy $NLE_k(x)$. Being a relative value, $NLE_k(x)$ reflects the relative weight of local energy of pixel x at level k. Figs. 1(b), 1(c), 1(d) show the first three normalized local energy maps of Fig.1(f), a selected part of Fig.1(a).

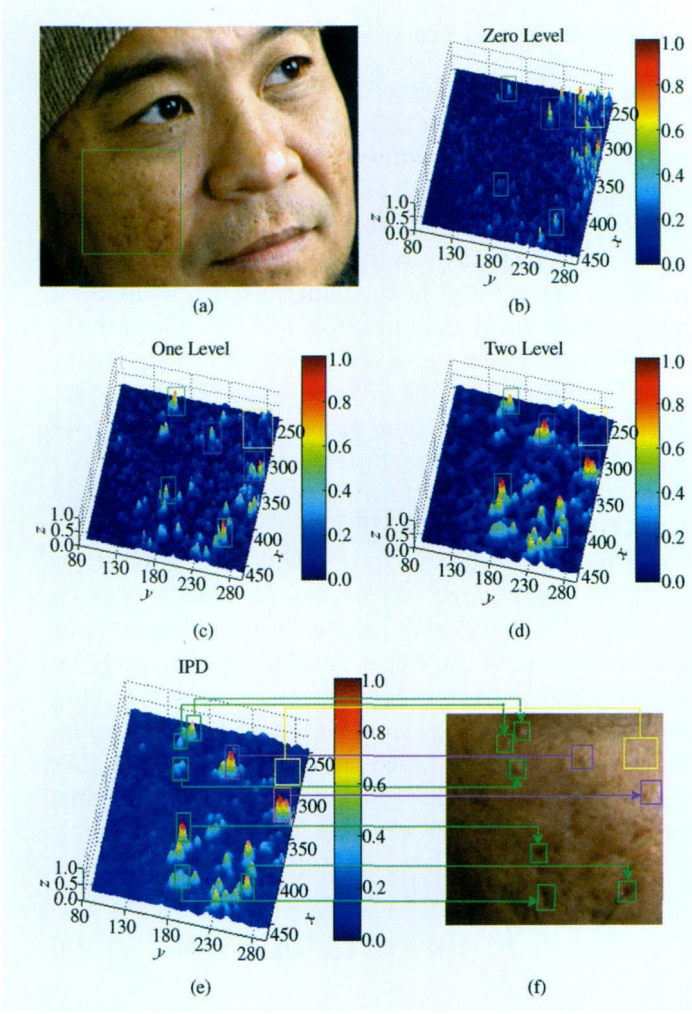


Fig.1. (f) is an image patch cropped from (a) (green rectangle). (b)~(d) are the first, second and third NLE ($k = 0, 1, 2$) of (f). Pixels in purple, green and yellow rectangle correspond to cases 1), 2), 3) respectively. (e) IPD of (f). In (e) and (f), the correspondences between pixels in IPD and the image are marked out with different colors.

4.2 Analyze the Normalized Local Property

Having extracted the NLE of every pixel at different levels, in this subsection, we make an analysis of NLE, and then propose a function of every pixel to characterize the energy distribution of the facial image.

From an overall view of Figs. 1(b)~1(d), we can clearly see: the pores and imperfections (bumps, scars) exhibit different characteristics from the first level to the third level. For the former, NLE decreases, and for the latter, NLE increases. More specifically, in Fig.1(b), the NLE of the first level, both the pores and bumps all have high levels of energy, especially the pores and some distractions. Note that, the part with most highest NLE is on the upper right in Fig.1(f) where the pores can be clearly seen. In Fig.1(c), the NLE of distractions goes up sharply compared with Fig.1(b), and that of pores begins to attenuate. Finally, in Fig.1(d), while the NLE of pores drastically decreases, the NLE of the distractions continues to grow. At this level, the differentiation between pores and distractions are justified. According to NLE behaviors from the first to third level, all pixels in the image fall into three categories.

1) Pixels with NLE relatively large at all levels, mainly relate to some sharp imperfections, e.g., pixels in purple rectangles in Figs. 1(b)~1(d).

2) Pixels with NLE from small to large mainly relate to gently rolling distractions, e.g., pixels in green rectangles in Figs. 1(b)~1(d).

3) Pixels with NLE from large to small mainly relate to skin details, e.g., pixels in yellow rectangles in Figs. 1(b)~1(d).

As the residual image of EMD is nearly constant, meaning that all the oscillatory modes of original image have been picked out, it is reasonable to use NLE of all IMFs to define pixels' imperfect degree. We define an imperfect degree for every pixel x of I as the weighted sum of NLE of all IMF levels and their differences between adjacent levels:

$$\alpha = \sum_{k=0}^{K} NLE_k(x),$$
$$\beta = \sum_{k=0}^{K-1} (NLE_k(x) - NLE_{k+1}(x)),$$
$$IPD(x) = \lambda e^{\alpha} + (1-\lambda)e^{-\beta}. \quad (6)$$

In the above formula, the first term rewards pixels in case 1), and the second term punishes those pixels in case 3) and rewards pixels in case 2). λ is a weight to balance two terms and usually set between 0.2~0.5.

The IPD defined as above quantitatively measures the imperfect degree of every pixel. The larger the IPD is, the more imperfect the facial skin is. As seen from Figs. 1(e), 1(f), while pixels with large IPD exactly correspond to distractions as we have marked with green and purple rectangles, pixels with small IPD directly relate to image regions with fine scale details, see yellow rectangles. The accurate correspondences show that our IPD indeed delivers an intuitive measure of degree of visual imperfection.

4.3 Adjust the Coefficient

Once the IPD for every pixel has been calculated, the next step is to adjust the coefficients of every IMF to meet our goal based on the hints of IPD. It is recognized that pixels with large IPD generally correspond to imperfect skin such as large bumps which we intend

to smooth, while pixels with small IPD relate to skin details such as pores which we need to preserve. In the following, we present a flexible filtering approach for modulating the pixels so as to remove facial imperfections while preserving details.

We take Gaussian filtering to smooth IMF. Normally, the larger standard deviation is estimated, the larger smooth window should be employed. By experiment, we found that for most imperfect images a 11×11 pixels window of Gaussian filter is sufficient. By introducing a parameter h to control the degree of filtering, the standard deviation of Gaussian filtering for x at level k is formulated as

$$\sigma_k(x) = h \times (k+1) \times IPD(x), \tag{7}$$

where $IPD(x)$ is the imperfect degree of x, $k+1$ is a level-related quantity, meaning that standard deviation increases at the high level. h is the filtering parameter, large h implies high degree of smoothing, and vice versa. According to our experiments, a value of h ranging from 2 to 8 is suitable to achieve visually satisfactory results for most images. Typically, the default setting of h is 3.

The resultant facial image is efficiently reconstructed by adding modified IMFs and the residual image together.

5 Experiments and Discussion

In this section, we show some experimental results of our ALNEMD and face cleaning algorithm in comparisons with other relevant algorithms. We first compare ALNEMD with several other BEMD methods including directional BEMD, triangle-based cubic interpolation BEMD and RBF interpolation BEMD. The test images are "Barbara" (the left image in the first row of Fig.2), "Lena" (the middle one) and "Girl" (the right one). In these results, when performing ALNEMD the threshold μ_T is set to $(T^2-1)/3$ where T^2 is the window size.

Fig.2 shows the visual comparison between our ALNEMD algorithm, directional BEMD and RBF interpolation BEMD. The second row shows the first three IMFs of the "Barbara" obtained by directional BEMD. The third row is the first three IMFs by RBF interpolation BEMD. The last row is our results. Unlike the results obtained by other two algorithms, no obvious overdark or overbright spot is present on IMFs produced by our algorithm.

We further adopt root mean square error (RMSE) between the local mean $p(i,j)$, $i = 1,\ldots,M$, $j = 1,\ldots,N$ computed by different approaches and the true local mean of the image (computed in a 5×5 neighborhood) $q(i,j)$, $i = 1,\ldots,M$, $j = 1,\ldots,N$ to compare

Fig.2. IMFs obtained by decomposing "Barbara" with directional BEMD (the second row), RBF-based interpolation (the third row) and our ALNEMD (the fourth row).

the performance of different approaches. Specially, the RMSE is defined as

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{i=1}^{M}\sum_{j=1}^{N}(p(i,j)-q(i,j))^2}{MN}}. \tag{8}$$

Table 2. RMSEs of Three Approaches

Image	Cubic	RBF	Our Algor.
Barbara	0.0712	0.0539	0.0176
Lena	0.1015	0.0960	0.0180
Girl	0.0840	0.0703	0.0114

Since the local mean of directional BEMD is not a truly 2D local mean, directional BEMD is not to be compared in terms of RMSE. Table 2 shows the RMSEs of the three images by the three approaches, namely, triangle-based cubic interpolation, RBF-based interpolation and our adaptive local mean algorithm. From the table, we can see that the results of our algorithm are much closer to the true local mean of image.

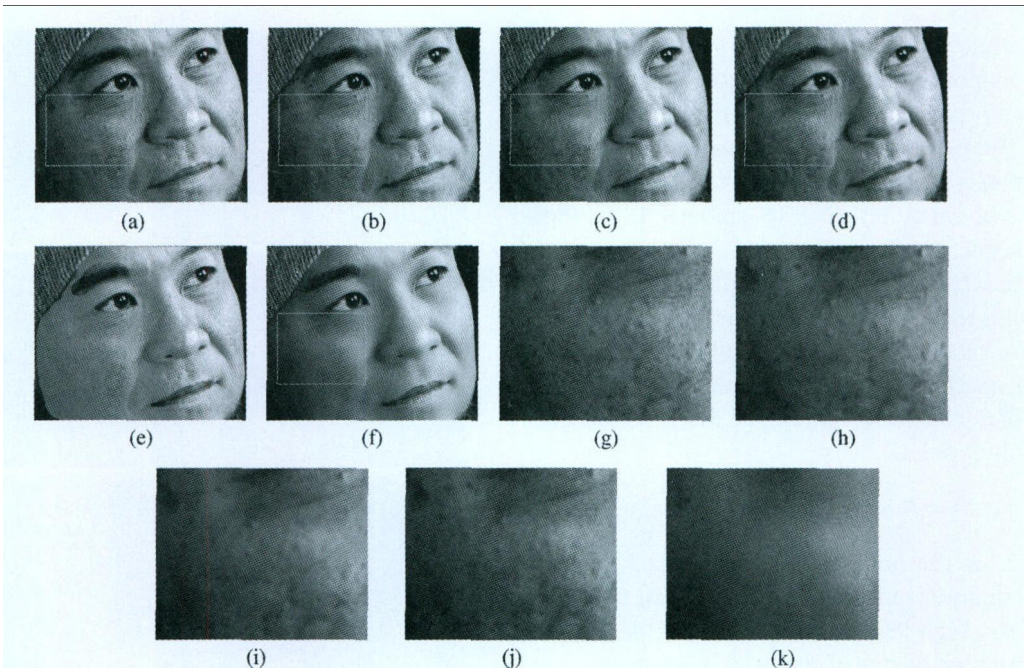


Fig.3. Cleaning effects comparison for the facial image. (a) Image with size 528×644 in pixels. (b) and (c) Results obtained by bilateral with different parameters. (d) Result of Matsui *et al.*[12] (e) Mask painted by the user to specify the region of cleaning. (f) Result of our method. Apparently, the bumps as well freckles are removed thoroughly, while the pores on the face are retained. (g)~(k) Five zoomed in skin patches selected from the same position (the rectangle) of Figs. 3(a)~3(d) and 3(f) respectively.

Subsequently, we tested our face cleaning approach with a set of facial images. For color image, we first convert it from RGB space to YCrCb space, and then perform cleaning operation on its Y component. Finally, we invert the YCrCb space to RGB space to produce the resulting image. In the following experimental results, when computing IPD with our algorithm the λ is set as 0.4.

Fig.3 shows a comparison of our cleaning method, bilateral filtering and the technique of Matsui *et al.*[12] The original image to be processed is shown in Fig.3(a), and the mask image we used is shown in Fig.3(e). Our ALNEMD decomposes the original image into three IMFs ($K = 2$) together with a residual. In the original image, there are a number of pronounced scars and bumps, especially the scars on the left part of the face, making the task of cleaning challenging. However, as can be seen from our processed image (Fig.3(f) with parameter $h = 5$), the pores on the face are well preserved, while the large rough part of the face (at bottom left) has been repaired. Figs. 3(b) and 3(c) show the results of bilateral filtering with different parameters. For Fig.3(b), the geometric spread $\sigma_d = 1$ and the photometric spread $\sigma_r = 1000$. For Fig.3(c), $\sigma_d = 2$ and $\sigma_r = 3000$. It should be noted that in Fig.3(b) small σ_d is adopted with an attempt to preserve pores and large σ_r is adopted intending to smooth the edges of bumps and scars, but the small bumps (at top left) are not removed completely in the result. In another experiment of bilateral filtering, we increase the σ_d to remove bumps (at top left), it appears that pores are removed and the scars (at bottom left) are still present, as shown in Fig.3(c). Since large σ_r would blur the edges of face, when performing bilateral filtering for Fig.3(a), we also use the mask image, see Fig.3(e), to keep the eyes and mouth out of filtering in Figs. 3(b) and 3(c). Fig.3(d) is the result of Matsui *et al.*[12] with $\varepsilon_1 = 31$, $\varepsilon_2 = 29$, $\varepsilon_3 = 10$, $a = 3$. As can be seen, part of skin pores are retained, but the spots and the scares on the left face are left behind. Figs. 3(g)~3(k) are five zoomed in skin patches selected from the same position (the rectangle) of Figs. 3(a)~3(d) and 3(f) respectively.

Since the IPD is derived from the normalized local energy, it reflects the relative energy of image pixels and work in the case of noisy, blurry and different resolution images. Fig.4(a) is the original image part and Fig.4(b) is the corresponding IPD. Fig.4(c) is a noisy version of Fig.4(a) with additive Gaussian white noise of variance 25, and Fig.4(d) is the IPD of Fig.4(c). As can be seen, the IPD can still represent the visual perception of face images. Fig.4(e) is a motion blurred version of Fig.4(a) with translation 20 pixels and rotation 10 degrees. Fig.4(f) is the corresponding IPD, we can see that in Fig.4(e) pores are blurred, leaving behind some scars which have been represented with high IPD values in Fig.4(f). Fig.4(g) is an image with a

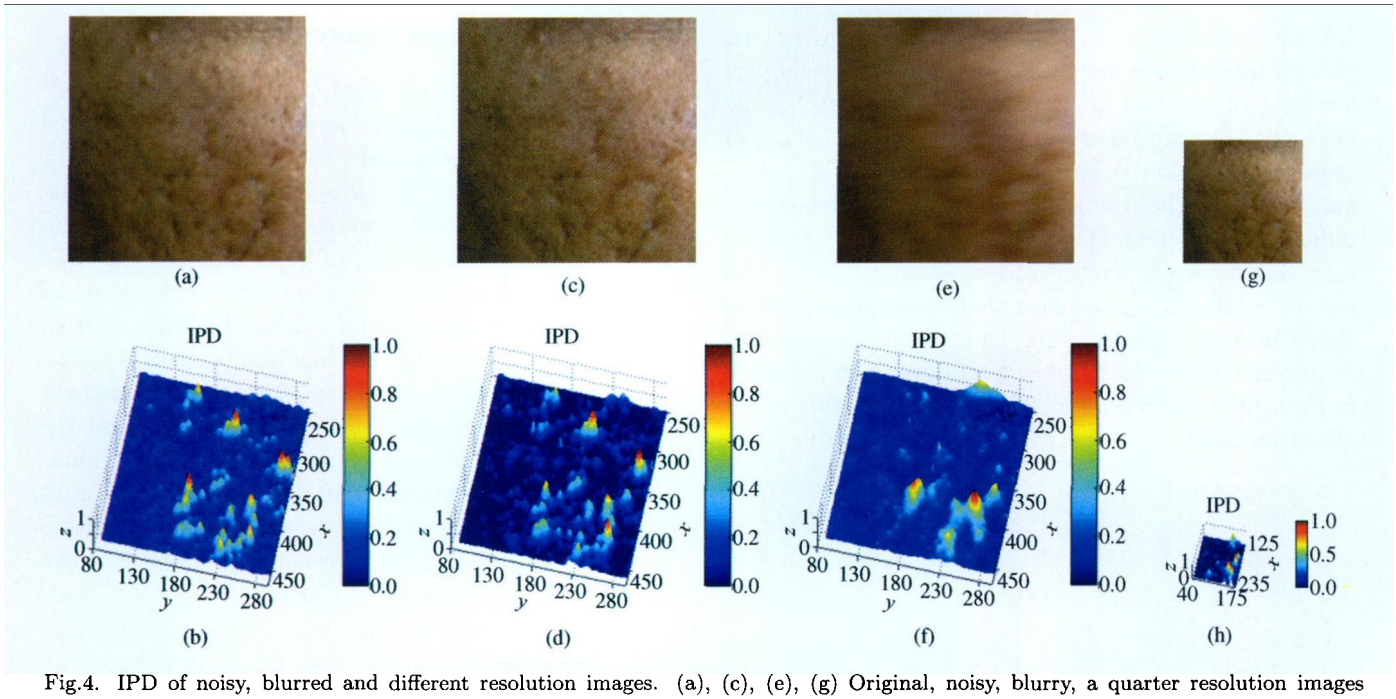


Fig.4. IPD of noisy, blurred and different resolution images. (a), (c), (e), (g) Original, noisy, blurry, a quarter resolution images respectively. (b), (d), (f), (h) Corresponding IPD.

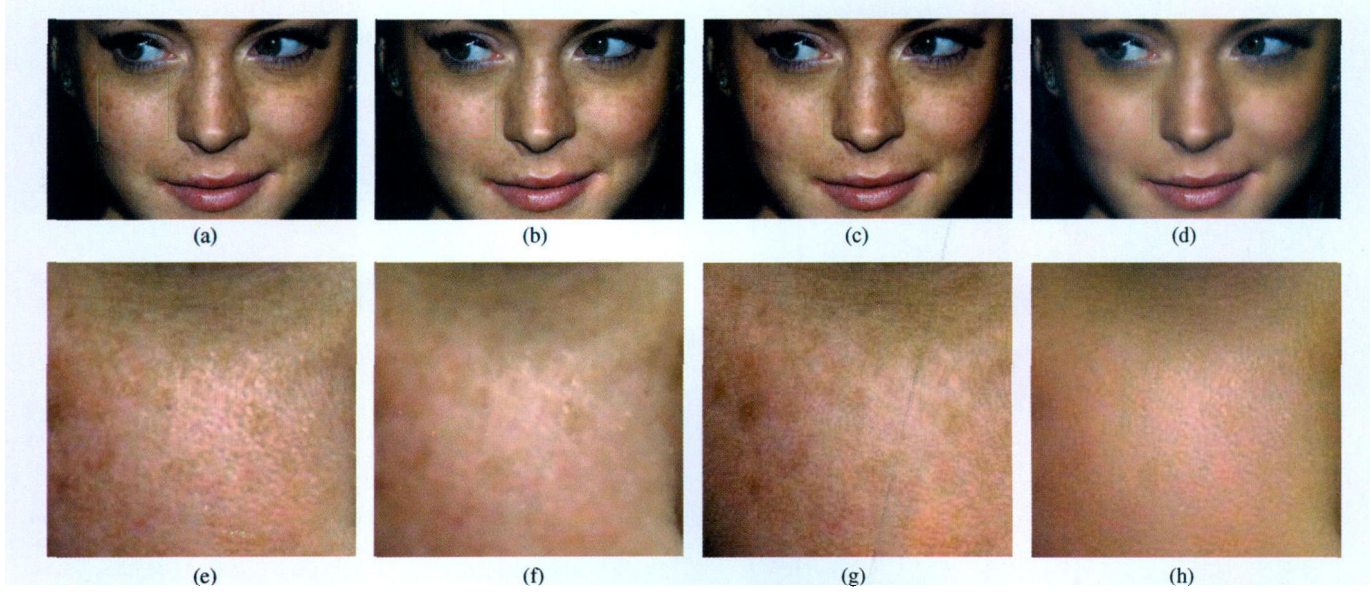


Fig.5. Comparison of the three approaches. (a) Girl's face with image size 944 × 592 in pixels. (b) Result of bilateral filtering. (c) Result of Matsui *et al.*[12] (d) Our result. (e)~(h) Four zoomed in skin patches selected from the same position (the green rectangle) in (a)~(d).

quarter of resolution of Fig.4(a), and Fig.4(h) is the corresponding IPD.

Fig.5 provides another comparison for Fig.5(a). Fig.5(b) (parameters $\sigma_d = 5$ and $\sigma_r = 10$) is produced by bilateral filtering and Fig.5(c) ($h = 3$) by our algorithm. We can see that the young girl's pores are retained and meanwhile the blemishes are diminished in Fig.5(c). By contrast, small distractions are still present in Fig.5(b). Fig.5(d) is the result of Matsui *et al.*[12], the pores of girls are retained. However, the freckles are present. Moreover, since the method uses unsharp mask to enhance the pores, the noise in dark regions, for example, the left face and right face near the hairs, have been amplified.

Fig.6 demonstrates another result of our algorithm (Fig.6(d)) for a patch of a young man's facial skin

(Fig.6(a)) with $h = 3$. Because there is no rugged part on the face, the smoothing operation is evenly performed. In contrast with original skin texture of the man, the enlarged pores of skin have been refined, and more delicate skin is obtained (Fig.6(d)). Figs. 6(b) and 6(c) show the results of the bilateral filtering with different parameters, for Fig.6(b), the geometric spread $\sigma_d = 3$ and photometric spread $\sigma_r = 10$; for (c), $\sigma_d = 5$ and $\sigma_r = 20$. It can be seen that some pores (at top right) in Figs. 6(b) and 6(c) have been smeared out, meanwhile others (at bottom left) have been left aside in Figs. 6(b) and 6(c). Fig.7 ($h = 3$) shows other result of our method.

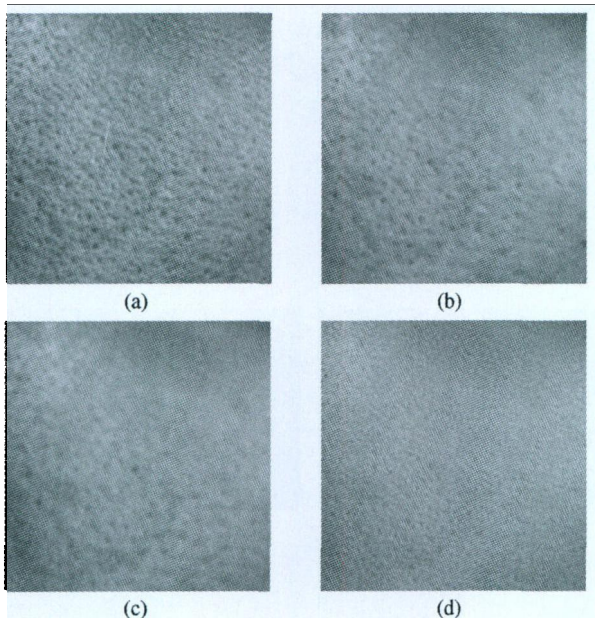


Fig.6. Result comparison for a facial image patch. (a) Image patch with size 266 × 299 in pixels. (b) and (c) Results obtained by bilateral filtering with different parameters. (d) Result produced by our method.

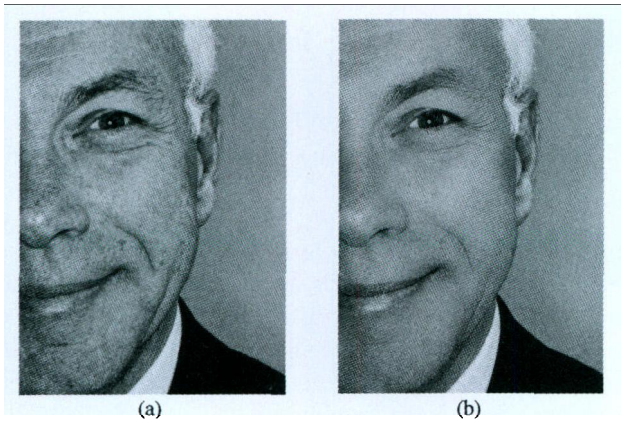


Fig.7. Face cleaning effect (b) for a gentleman (a) with image size 352 × 488 in pixels.

6 Conclusions and Future Work

We have presented a novel method to deal with imperfect facial images. It is found that normalized local energy extracted from IMFs can effectively reveal different characteristics of fine scale details and distractions. After analysis of normalized local energy, we propose a quantitative measure of facial skin, namely IPD. Experiments show our IPD is an intuitive measure of degree of visual imperfection. Under the guide of IPD, we can adaptively adjust the coefficients of IMFs to improve the appearance of a facial skin. Our method not only removes distractions such as bumps and blemishes effectively, but also successfully maintains facial pores. Furthermore, our method does not request much interaction and professional skills. In addition, an algorithm named as ALNEMD is also developed to improve in two respects the performance of EMD.

There are some possible extensions to this work. Combining facial feature detection algorithms, we would like to extend our approach to enhance facial skins in video sequence which might be interesting. Being a function of space, our IPD gives sharp identifications of imbedded structures in facial images. Hence, it can be further used in other face-related work. For example, for those work which need to explicitly detect facial imperfections, e.g., for identification[5], IPD can aid in localizing skin irregularities.

References

[1] Tomasi C, Manduchi R. Bilateral filtering for gray and color images. In *Proc. the Sixth International Conference on Computer Vision*, Bombay, India, Jan. 4–7, 1998, pp.839–846.

[2] Weiss B. Fast median and bilateral filtering. *ACM Trans. Graphics*, 2006, 25(3): 519–526.

[3] Buades A, Coll B, Morel J M. A non-local algorithm for image denoising. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, San Diego, USA, June 20–25, 2005, Vol.2, pp.60–65.

[4] Huang N E *et al.* The empirical mode decomposition and Hilbert spectrum for nonlinear and nonstationary time series analysis. In *Proc. the Royal Society A: Mathematical, Physical and Engineering Sciences*, March 8, 1998, Vol.454, pp.903–995.

[5] Leyvand T, Cohen-Or D, Dror G, Lischinski D. Data-driven enhancement of facial attractiveness. *ACM Trans. Graphics*, 2008, 27(3): 38:1–38:9.

[6] Nguyen M H, Lalonde J F, Efros A A, Fernando De la Torre. Image-based shaving. *Computer Graphics Forum Journal*, 2008, 27(2): 627–635.

[7] Peers P, Tamura N, Matusik W, Debevec P. Post-production facial performance relighting using reflectance transfer. *ACM Trans. Graphics*, 2007, 26(3): 52:1–52:10.

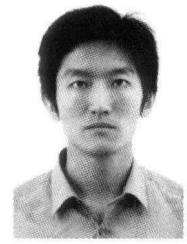
[8] Bitouk D, Kumar N, Dhillon S, Belhumeur P, Nayar S K. Face swapping: Automatically replacing faces in photographs. *ACM Trans. Graphics*, 2008, 27(3): 39:1–39:8.

[9] Perez P, Gangnet M, Blake A. Poisson image editing. *ACM Trans. Graphics*, 2003, 22(3): 313–318.

[10] Lin D, Tang X. Recognize high resolution faces: From macrocosm to microcosm. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, New York, USA, June 17–22, 2006, pp.1355–1362.
[11] Pierrard J S, Vetter T. Skin detail analysis for face recognition. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, Minneapolis, USA, June 18–23, 2007, pp.1–8.
[12] Matsui T, Arakawa K, Nomoto K. A nonlinear filter system for beautifying face images with enhancement using interactive evolutionary computing. In *Proc. International Symposium on Intelligent Signal Processing and Communications*, Yonago, Japan, Dec. 12–15, 2006, pp.534–537.
[13] Kovesi P. Invariant measures of image features from phase information [Ph.D. Dissertation]. University of Western Australia, Australia, 1996.
[14] Granlund G H, Knutsson H. Signal Processing for Computer Vision. Kluwer Academic Publishers, 1995.
[15] Robbins B, Owens R. 2D feature detection via local energy. *Image and Vision Computing*, 1997, 15(5): 353–368.
[16] Larkin K G. Uniform estimation of orientation using local and nonlocal 2-D energy operators. *Optics Express*, 2005, 13(20): 8097–8121.
[17] Han C M, Guo H D, Wang C L. A novel method to reduce speckle in SAR images. *International Journal of Remote Sensing*, 2002, 23(23): 5095–5101.
[18] Nunes J C, Guyot S, Del echelle E. Texture analysis based on local analysis of the bidimensional empirical mode decomposition. *Machine Vision and Applications*, 2005, 16(3): 177–188.
[19] He L, Wang H. Spatial-variant image filtering based on bidimensional empirical mode decomposition. In *Proc. International Conference on Pattern Recognition*, Hong Kong, China, Aug. 20–24, 2006, Vol.2, pp.1196–1199.
[20] Sinclair S, Pegram G S. Empirical mode decomposition in 2-D space and time: A tool for space-time rainfall analysis and nowcasting. *Hydrology and Earth System Sciences Discussions*, 2005, 2(1): 289–318.
[21] Rilling G, Flandrin P, Goncalves P. On empirical mode decomposition and its algorithms. In *IEEE-EURASIP Workshop on Nonlinear Signal and Image Processing*, Grado, Italy, June 8–11, 2003.
[22] Felsberg M, Sommer G. The monogenic scale-space: A unifying approach to phase-based image processing in scale-space. *Journal of Mathematical Imaging and Vision*, 2004, 21(1): 5–26.

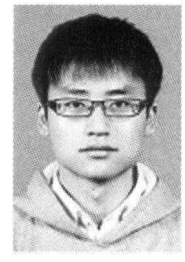
Yan-Li Liu received her B.S. degree in applied mathematics from Henan University in 2004. Currently, she is a Ph.D. candidate in State Key Lab of CAD & CG, Zhejiang University, majoring in applied mathematics. Her research interests include computer graphics and image processing.

Xiao-Gang Xu received his Ph.D. degree from Dalian University of Technology in 1999. Currently he is a professor of the Department of Navigation at Dalian Naval Academy. His main research includes signal analysis, image processing and virtual simulation.



Yan-Wen Guo received his Ph.D. degree in applied mathematics from State Key Lab of CAD & CG, Zhejiang University in 2006. He is currently an associate professor at the National Laboratory of Novel Software Technology, Nanjing University. His main research interests include image and video processing, geometry processing, and face related applications.

Jin Wang received his Ph.D. degree in computer science and technology from Zhejiang University, China in 2007. Currently he is a researcher in State Key Lab of CAD & CG, Zhejiang University. His research interests focus on photo/video enhancement, face relevant research, computer vision & computer graphics.



Xin Duan now is an undergraduate student in Mixed-Class, Chu-Kochen Honors College of Zhejiang University. His main research interests are image and video processing.

Xi Chen received her B.S. degree in computer science and technology from Zhejiang University, China in 2007. She is currently pursuing her M.S. degree in Department of Computer Science, University of British Columbia, Canada. Her research interests include image and video processing.

Qun-Sheng Peng graduated from Beijing Mechanical College in 1970 and received a Ph.D. from the Department of Computing Studies, University of East Anglia, UK in 1983. Now he is a professor of computer graphics at Zhejiang University. He serves currently as a member of the editorial boards of several international and Chinese journals. His research interests include realistic image synthesis, computer animation, scientific data visualization, virtual reality, bio-molecule modeling. He is a senior member of CCF.