Face image super-resolution from video data with non-uniform illumination

Andrey S. Krylov, Andrey V. Nasonov, Dmitry V. Sorokin
Faculty of Computational Mathematics and Cybernetics
Moscow Lomonosov State University, Moscow, Russia
kryl@cs.msu.su, nasonov@cs.msu.ru, sorokin_dm@bk.ru

Abstract

Tikhonov regularization approach and block motion model are used to solve super-resolution problem for face video data. Video is preprocessed by 2-D empirical mode decomposition method to suppress illumination artifacts for super-resolution.

Keywords: *face super-resolution, video, EMD.*

1. INTRODUCTION

The problem of super-resolution is to recover a high-resolution image from a set of several degraded low-resolution images. This problem is very helpful for face detection in human surveillance, biometrics, etc. because it can significantly improve image quality.

Face super-resolution algorithms can be divided into two groups: learning-based and reconstruction-based.

Learning-based algorithms collect the information about correspondence between low- and high-resolution images and use the gathered information for resolution enhancement. These methods are not actually super-resolution methods, because they operate with a single image. They do not reconstruct missed data, they only predict it using learning database. Several input images do not significantly improves the resolution and only help to reduce the probability of using incorrect information from the database. The most popular method is Baker method [1], [2] which decomposes the image into a Laplacian pyramid and predicts its values for high-resolution image. Patch-based methods are popular too. They divide low- and high-resolution images into a set of pairs of fixed size rectangles called patches and substitutes the most appropriate patches into high-resolution image. They vary by learning and substitution methods, for example, neural networks [3], locality preserving projections [4], asymmetric associative learning [5], locally linear embedding [6], etc. Principal component analysis is also used for learning-based super-resolution [7].

Reconstruction-based algorithms use only low-resolution images to construct high-resolution image. Most reconstruction-based algorithms use camera models [8] for downsampling the high-resolution image. The problem is formulated as error minimization problem

$$z = \arg\min_{z \in Z} \sum_k \left\| A_k z - v_k \right\| , \qquad (1)$$

where z is unknown high-resolution image, v_k is *k*-th low-resolution image, A_k is an operator which transforms high-resolution image into low-resolution. Various norms are used. The operator can be generally represented as $A_k z = DH_{cam} F_k H_{atm} z + n$, where H_{atm} is atmosphere turbulence effect which is often neglected, F_k is a warping operator like motion blur or image shift for *k*-th image, H_{cam} is camera lens blur which is usually modeled as Gauss filter, D is the downsampling operator, n is a noise, usually Gaussian. In many cases, only translation model is considered and noise is ignored, so F_k can be merged with H_{cam} and the transform operator is simplified as $A_k z = DH_k z$, where H_k is a shifted Gauss filter with the kernel

$$H_k(x, y) = \frac{1}{\sigma\sqrt{2\pi}} \exp(-\frac{(x - x_k)^2 + (y - y_k)^2}{2\sigma^2}) , \qquad (2)$$

where x_k and y_k are shifts of high-resolution image relatively to k-th low-resolution image along x and y axis respectively.

There are different methods to solve (1). The most widely-used methods are [9]: iterated error back-projection which minimizes the error functional using error upsampling and subtraction from high-resolution image [10], [11], stochastic reconstruction methods [12], projections onto convex set [13], [14], Tikhonov regularization [8] and single-pass filtering which approximates the solution of (1) [15], [16].

Linear translation model is usually insufficient for super-resolution problem, because the motion is non-linear. Different motion models are used [4], [15]. It is computational ineffective to calculate the motion for every pixel. The motion of adjacent pixels is usually similar, so, the motion of only several pixels is calculated. The motion of other pixels is interpolated. The simplest model is regular motion field [4]. For large images, it is effective to calculate the motion of pixels which belong to edges and corners [15].

2. OUR APPROACH

We consider the task of face image super-resolution from video data. We use Tikhonov regularization approach [8] and block motion model. The reason is that the problem (1) is ill-conditioned or either ill-posed. We use l_1 norm $\|z\|_1 = \sum_{i,j} | z_{i,j} |$ instead of standard Euclidian norm, because it has shown better results.

$$z = \arg\min_{z \in Z} \sum_k \left\| A_k z - v_k \right\|_1 + \alpha f(z) . \qquad (3)$$

We choose total variation functional $TV(z) = \sum_{i,j} | z_{i+1,j} - z_{i,j} | + \sum_{i,j} | z_{i,j+1} - z_{i,j} |$ and bilateral TV functional [i] $BTV(z) = \sum_{\substack{-p \le x \le p, \\ -p \le y \le p}} \gamma^{|x|+|y|} \left\| S_{x,y} z - z \right\|_1$ as a stabilizer

$f(z)$, $S_{x,y}$ is a shift operator along horizontal and vertical axis for *x* and *y* pixels respectively, $\gamma = 0.8$, $p = 1$ or 2. $TV(z)$ can be also represented as $TV(z) = \|S_{1,0}z - z\|_1 + \|S_{0,1}z - z\|_1$.

All the images are considered as the results of motion of the first image. Both first and target images are convolved with Gauss filter to suppress noise. For motion estimation, we calculate the motion on a regular grid G with a step within 8 to 16 pixels range. For every point from the grid G, we take a small square block (8–16 pixels width) from the first image centered in this point. Then we find the optimal shift of this block in target image with pixel accuracy using least mean square approach. To calculate the motion with subpixel accuracy, we convolve the first image with shifted Gauss filter (2).

The motion for other pixels is interpolated linearly (for example, using bilinear or Gauss filter). Then a set of matrices $T^{(k)}$ of 2-D points $T^{(k)}_{i,j} = (x_i, y_j)$ is constructed. The *k*-th matrix represents the correspondence between pixels from the *k*-th image and the first image. Next we multiply every element of these matrices by resampling scale factor, so the matrices represent the correspondence of pixels between low-resolution images and high-resolution image.

In this case, transform operator A_k looks as $A_k = T^{(k)}H$, where H is zero-mean Gauss filter and $u^{(k)} = T^{(k)}z$ is motion-compensated downsampling operator:

$$u^{(k)}_{i,j} = z_{x_i, y_j}, \text{ where } (x_i, y_j) = T^{(k)}_{i,j}.$$

If x_i or y_j is not integer value, then z_{x_i, y_j} is approximated using bilinear interpolation. Note: it is better to perform shifted Gauss filter to calculate z_{x_i, y_j} more precisely, but it would be very slow. Gauss filter H reduces high-band frequencies, so bilinear approximation is enough.

3. NUMERICAL METHOD

We use iterative subgradient method with non-constant step [17] for fast minimization of (3). The iterations look like

$$z^{(n+1)} = z^{(n)} - \beta_n g^{(n)}, \tag{4}$$

where $g^{(n)} \in \partial F(z)|_{z^{(n)}}$ is any subgradient of the object functional

$$F(z) = \sum_k \left\| T^{(k)}Hz - v_k \right\|_1 + \alpha f(z).$$

Vector $g^{(n)}$ is an element of subgradient set $\partial F(z)|_{z^{(n)}}$ of $F(z)$ at $z^{(n)}$ if it satisfies the condition $F(z) \geq F(z^{(n)}) + (g^{(n)}, z - z^{(n)})$ for all z. Only one subgradient exists and it is equal to normal gradient if $F(z)$ is differentiable at $z^{(n)}$.

The subgradient of $J(u) = \|u\|_1$ for the grid points is

$$g(u)_{i,j} \in \partial J(u)_{i,j} = \begin{cases} \{1\}, & u_{i,j} > 0, \\ \{-1\}, & u_{i,j} < 0, \\ [-1, 1], & u_{i,j} = 0, \text{ in this case we} \\ & \text{assume } g(u)_{i,j} = 0. \end{cases}$$

Thus $\partial J(u) = \operatorname{sign} u$ with sign function applied per each element of u. The subgradient of $F(z)$ can be written in the form

$$g^{(n)} = \sum_k H^* T^{(k)*} \operatorname{sign}(T^{(k)}Hz - v_k) + \alpha f'(z).$$

H^* and $T^{(k)*}$ are standard conjugate operators defined by Euclidian scalar product. For Gauss filter $H^* = H$. $z^{(k)} = T^{(k)*}u^{(k)}$ is constructed in the following way: first, $z^{(k)}$ is zero-filled, then for every pixel (i, j) from $u^{(k)}$ we obtain its coordinates in $z^{(k)}$: $(x_i, y_j) = T^{(k)}_{i,j}$ and add value $u^{(k)}_{i,j}$ to $z^{(k)}_{x_i, y_j}$. For non-integer coordinates, we add the value to the nearest pixels with coefficients obtained by bilinear interpolation.

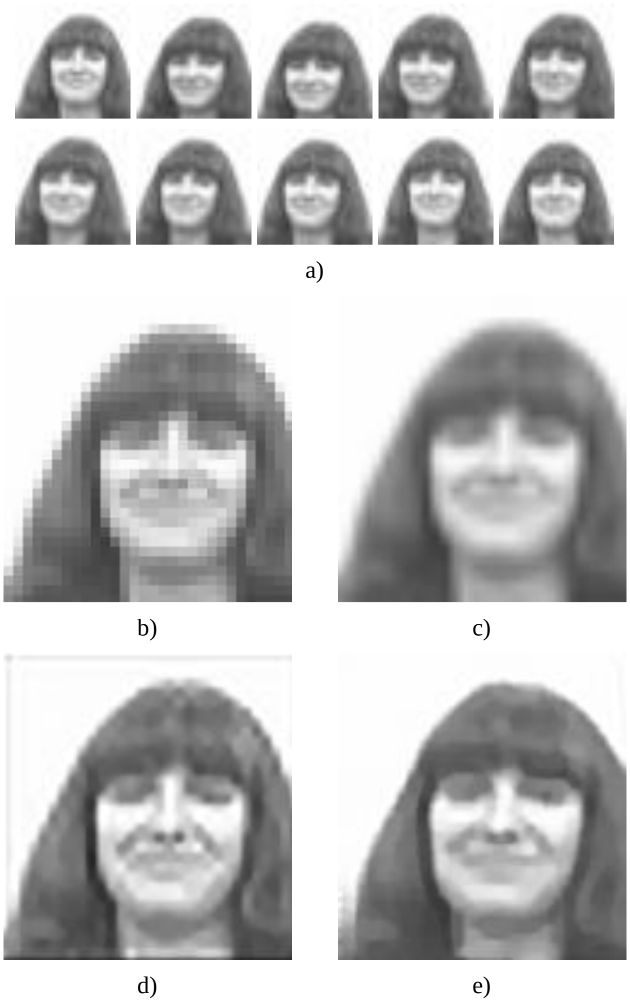
For the case of $f(z) = BTV(z)$, the subgradient looks as

$$f'(z) = \sum_{-p \leq x, y \leq p} \gamma^{|x|+|y|} (S_{-x,-y} - I) \operatorname{sign}(S_{x,y}z - z),$$

where I is unit operator. For $f(z) = TV(z)$ the subgradient is calculated the same way.

The coefficients β_n in (4) satisfy the condition for step lengths $\left\| \beta_n g^{(n)} \right\|_1 = s_n$, where step lengths s_n are chosen a priori in the form $s_n = s_0 q^n$, $0 < q < 1$. We use $s_0 = 50$ and choose q to obtain $s_{N-1} = 0.1$ for the last iteration.

The application of the proposed super-resolution method is shown in Figure 1. For sequent video data this method shows better results than any single image resampling method.



a)

b)

c)

d)

e)

Figure 1: Face super-resolution for the factor of 4 and 10 input images.
a) source low-resolution images;
b, c, d) single image interpolation using b) nearest neighbor;
c) bilinear interpolation; d) regularization-based method [26];
e) proposed super-resolution result.

4. EMD-BASED ILLUMINATION ARTIFACT REMOVAL

The initial super-resolution video data suffers from the illumination artifacts. To overcome this problem we use Empirical Mode Decomposition (EMD) method.

EMD is a multisolution decomposition technique which was first introduced by Huang et al. in [18]. This method is appropriate for non-linear, non-stationary signal analysis. The concept of EMD is to decompose the signal into a set of zero-mean functions called Intrinsic Mode Functions (IMF) and a residue. As the increasing of decomposition level, the complexion (frequency) of IMF decreases. In comparison to other time-frequency analysis tools such as Fourier analysis or wavelet analysis, EMD is fully data-driven i.e. there are no pre-determined basis functions.

At first we describe the algorithm for 1-D signals.

Huang et al. defined IMF as function that satisfies two conditions: a) the number of extrema equals the number of zero-crossing or differs at most by one; b) at any point, the mean value of upper envelope defined by local maxima and lower envelope defined by local minima is zero. Let $f(t)$ be the signal to be decomposed. Using this definition we can describe EMD algorithm as follows:

1. Identify all local extrema of $f(t)$.
2. Interpolate all local maxima to get upper-envelope $e_{\max}(t)$ and all local minima to get lower-envelope $e_{\min}(t)$.
3. Compute the local mean $m(t)=\dfrac{e_{\max}(t)+e_{\min}(t)}{2}$.
4. Compute $d(t)=f(t)-m(t)$. $d(t)$ is the candidate to be an IMF.
5. If $d(t)$ satisfies the definition of IMF, subtract it from the signal $r(t)=f(t)-d(t)$ and go to step 6.

   If $d(t)$ does not satisfy the definition of IMF, go to step 1 and use $d(t)$ instead of $f(t)$. Steps 1-5 are repeated until $d(t)$ satisfies the definition of IMF.
6. If residue $r(t)$ is a monotone function, the decomposition process is complete.

   If residue $r(t)$ is not a monotone function, go to step 1 and use $r(t)$ instead of $f(t)$.

The process of getting each IMF (steps 1-4) is called sifting process. When the decomposition is complete we can write $f(t)$ as follows:

$$f(t)=\sum_{k=1}^{N}\psi_k(t)+r(t),$$

where $\psi_k(t)$ is the *k*-th IMF and $r(t)$ is the residue.

There are several crucial points in the algorithm: the interpolation method for upper- and lower-envelopes calculation, boundary processing method, the stopping criterion and number of iterations in sifting process.

Huang et al. uses cubic spline interpolation to estimate the upper- and lower-envelopes [18]. Other methods for estimation are also used: B-splines [19], an optimization process based method [20], etc.

Several methods to process boundary points for interpolation of the envelopes were suggested. One of the ways to solve this problem is to consider the end points of the signal as the maximum and the minimum at the same time. Another way is to extend the signal, make envelopes for extended signal and then use only its original definition domain part [21].

In practice it is very difficult to get the physical meanfull IMF function that is strongly satisfies the definition. So different sifting process stopping criteria were introduced. Often the size of standard deviation SD computed from two consecutive sifting results [18, 22] is used as the criterion:

$$SD = \sum_{t=0}^{T} \frac{|d_{k-1}(t) - d_k(t)|^2}{d_{k-1}{}^2(t)}.$$

The typical used value of SD is between 0.2 and 0.3.

The limiting of the local mean value of sifting result $m(t)$ in each point is also used [23]. The number of iterations in sifting process can be restricted [22, 24].

2-D case of EMD is still an open problem but it has the same crucial points: extrema points locating process, the interpolation method for upper- and lower-envelopes estimation, boundary processing method, the stopping criterion and number of iterations in sifting process.

In our approach we locate local maxima and minima as follows: $f(i,j)$ is local maxima if $f(i,j) > f(k,l)$, where $i-1 \le k \le i+1, j-1 \le l \le j+1$, $f(i,j)$ is local minima if $f(i,j) < f(k,l)$ where $i-1 \le k \le i+1, j-1 \le l \le j+1$. We use Delaunay triangulation-based linear interpolation to estimate envelopes and even extension for boundary processing. This even extension for boundary processing is illustrated in Figure 2.

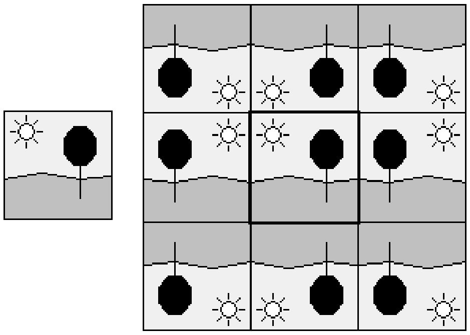


Figure 2: Boundary processing — a) original image; b) extended image for envelope construction.

As a stopping criterion we use the limitation of local mean in conjunction with restriction of the number of iterations in sifting process. An example of EMD applied to a face image from video is shown in Figure 3. The histogram of the IMF and residue images was adjusted to illustrate the behavior of these functions.

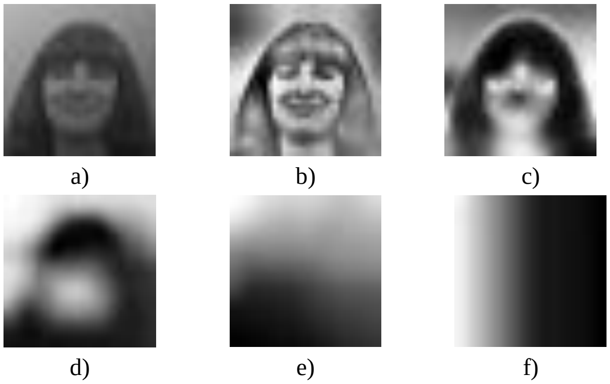


Figure 3: EMD example. a) original image; b) 1-st IMF; c) 2-nd IMF; d) 3-rd IMF; e) 4-th IMF; f) residue.

EMD method can be used for illumination artifact removal. The idea to remove illumination artifacts from image is based on the decomposition of the initial image using EMD $f(i,j) = \sum_{k=1}^{N} \psi_k(i,j) + r(i,j)$. Illumination artifacts are considered as low frequency information which can be eliminated from the image. We obtain the enhanced image using several first IMFs $f(i,j) = \sum_{k=1}^{M} \psi_k(i,j)$, where $M \le N$ which are the highest frequency components [25**]**.

In [25] the authors use 1-D EMD for illumination correction representing the image as 1-D signal. In our approach we use more effective 2-D EMD (see a result in Figure 4).

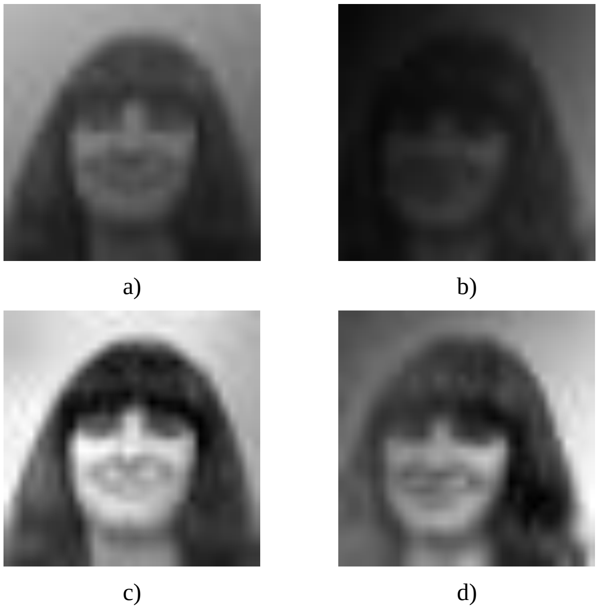
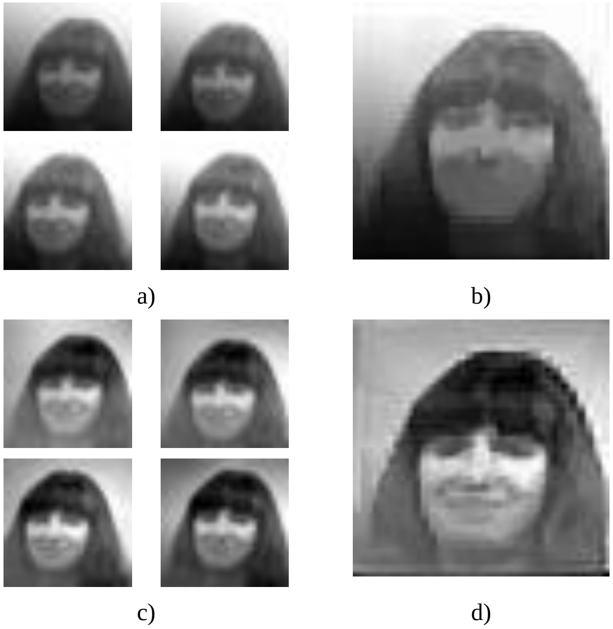


Figure 4: Illumination artifact removal — a,b) original images; c,d) processed images.

5. RESULTS

The results of super-resolution method depend drastically on the taken face video data. Serious enhancement of the tracked face by super-resolution method with EMD algorithm for illumination correction is typically obtained. To illustrate the general effect of the EMD enhancement we used a set of non-sequent images with artificially degraded illumination. This set is not typical for practical video data where the illumination change is continuous, but even in this case the EMD based result is reasonable (see Figure 5). Our tests show that the single image regularization resampling method [26] with EMD enhancement for the case of non-sequent images gives better result than the above super-resolution method.



a) b)

c) d)

Figure 5: An application of illumination artifact removal for super-resolution. a) original images; b) super-resolution result; c) EMD processed images; d) super-resolution result for EMD processed images.

6. CONCLUSION

Super-resolution method based on Tikhonov regularization approach and block motion model for face video data has been proposed. The approach was found promising to be used in real applications. The performance of the method has been improved by 2-D empirical mode decomposition method application to suppress illumination artifacts of video. The research on use of 2-D intrinsic mode functions inside super-resolution algorithm is under work.

This research was supported by RFBR grants 06-01-39006-ГФЕН_a and 06-01-00789-a.

7. REFERENCES

[1] S. Baker, T. Kanade "Hallucinating faces" // *In IEEE International Conference on Automatic Face and Gesture Recognition, March 2000.*

[2] S. Baker, T. Kanade "Limits on Super-Resolution and How to Break Them" // *IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 24, No. 9, Sep 2002, pp. 1167–1183.*

[3] Manuel Carcenac "A modular neural network for super-resolution of human faces" // *Applied Intelligence* http://www.springerlink.com/content/n7228127462q2457/ *to be published.*

[4] Sung Won Park, Marios Savvides "Breaking the limitation of manifold analysis for super-resolution of facial images" // *IEEE International Conference on Acoustics, Speech and Signal Processing, Vol. 1, April 2007, pp. 573–576*.

[5] Wei Liu, Dahua Lin, Xiaoou Tang "Face Hallucination Through Dual Associative Learning" // *IEEE International Conference on Image Processing, Vol. 1, Sept. 2005, pp. 873–876.*

[6] Hong Chang, Dit-Yan Yeung, Yimin Xiong "Super-Resolution Through Neighbor Embedding" // *Proceedings of the Computer Vision and Pattern Recognition, Vol. 1, 2004, pp. 275–282.*

[7] Wei Geng, Yunhong Wang "Aging Simulation of Face Images Based on Super-Resolution" // *Communications in Computer and Information Science, Vol. 2, Part 21, 2007, pp. 930–939.*

[8] S. Farsiu, D. Robinson, M. Elad, P. Milanfar "Fast and Robust Multi-Frame Super-Resolution" *// IEEE Trans. On Image Processing, Vol. 13, No. 10, pp. 1327-1344, October 2004*.

[9] S. Borman, Robert L. Stevenson "Super-Resolution from Image Sequences — A Review" // *Midwest Symposium on Circuits and Systems, 1998, pp. 374–378.*

[10] Jianguo Lu, Anni Cai, Fei Su "A New Algorithm for Extracting High-Resolution Face Image from Video Sequence" // *International Conference on Computational Intelligence and Security, Vol. 2, Nov. 2006, pp. 1689–1694.*

[11] Jiangang Yu, Bir Bhanu "Super-resolution Restoration of Facial Images in Video" // *18th International Conference on Pattern Recognition, Vol. 4, 2006, pp. 342–345.*

[12] R. R. Schultz, R. L. Stevenson "Extraction of highresolution frames from video sequences" // *IEEE Transactions on Image Processing, Vol. 5, No. 6, June 1996, pp. 996–1011*.

[13] A. J. Patti, M. I. Sezan, A. M. Tekalp "Superresolution video reconstruction with arbitrary sampling lattices and nonzero aperture time" // *IEEE Transactions on Image Processing, Vol. 6, No. 8, Aug. 1997, pp. 1064–1076.*

[14] P. E. Eren, M. I. Sezan, A. Tekalp "Robust, Object-Based High-Resolution Image Reconstruction from Low-Resolution Video" // *IEEE Transactions on Image Procession, Vol. 6, No. 10, 1997, pp. 1446–1451*.

[15] Ha V. Le, Guna Seetharaman "A Super-Resolution Imaging Method Based on Dense Subpixel-Accurate Motion Fields" // *Proceedings of the Third International Workshop on Digital and Computational Video, Nov. 2002, pp. 35–42.*

[16] F. Lin, J. Cook, V. Chandran, S. Sridharan "Face recognition from super-resolved images" // *Proceedings of the Eighth International Symposium on Signal Processing and Its Applications, Vol. 2, 2005, pp. 667–670.*

[17] S. Boyd, L. Xiao, A. Mutapcic "Subgradient methods" // *Lecture notes of EE392o, Stanford University, 2003.*

[18] N. E. Huang, Z. Shen, et al. “The empirical mode decomposition and the Hilbert spectrum for nonlinear and nonstationary time series analysis” // *Royal Society of London Proceedings Series A, Vol. 454, Issue 1971, 1998, pp. 903–1005.*

[19] Q. Chen, N. Huang, S. Riemenschneider, Y. Xu “A B-spline approach for empirical mode decompositions” // *Advances in Computational Mathematics, Vol. 24, 2006, pp. 171–195.*

[20] Yoshikazu Washizawa, Toshihisa Tanaka, Danilo P. Mandic, Andrzej Cichocki “A Flexible Method for Envelope Estimation in Empirical Mode Decomposition” // *Lecture Notes in Computer Science, Vol. 4253, 2006, pp. 1248–1255.*

[21] Kan Zeng, Ming-Xia He “A simple boundary process technique for empirical mode decomposition” // *IEEE International Proceedings of Geoscience and Remote Sensing Symposium, Vol. 6, Sept. 2004, pp. 4258–4261.*

[22] Liu Wei, Xu Weidong, Li Lihua “Medical Image Retrieval Based on Bidimensional Empirical Mode Decomposition” // *Proceedings of the 7th IEEE International Conference on Bioinformatics and Bioengineering, Oct. 2007, pp 641–646.*

[23] A. Linderhed “2D empirical mode decompositions in the spirit of image compression” // *Proceedings of the SPIE on Wavelet and Independent Component Analysis Applications IX, Vol. 4738, pp. 1–8.*

[24] Christophe Damerval, Sylvain Meignen, Valérie Perrier “A Fast Algorithm for Bidimensional EMD” // *IEEE Signal Processing Letters, Vol. 12, No. 10, Oct. 2005, pp. 701–704.*

[25] R. Bhagavatula, M. Savvides “Analyzing Facial Images using Empirical Mode Decomposition for Illumination Artifact Removal and Improved Face Recognition” *// Processing of IEEE International Conference on Acoustics, Speech and Signal, Vol. 1 April 2007, pp. 505–508.*

[26] Alexey Lukin, Andrey S. Krylov, Andrey Nasonov “Image Interpolation by Super-Resolution” // *Graphicon 2006 conference proceedings, Novosibirsk, Russia (2006), pp. 239–242.* http://imaging.cs.msu.ru/software/

About authors

Andrey S. Krylov is an associated professor, head of the Laboratory of Mathematical Methods of Image Processing, Faculty of Computational Mathematics and Cybernetics, Moscow Lomonosov State University.
Email: kryl@cs.msu.ru

Andrey V. Nasonov is a member of scientific staff of the Faculty of Computational Mathematics and Cybernetics, Moscow Lomonosov State University.
Email: nasonov@cs.msu.ru

Dmitry V. Sorokin is a student of the Faculty of Computational Mathematics and Cybernetics, Moscow Lomonosov State University.
Email: sorokin_dm@bk.ru