

# Text images superresolution and enhancement

Andrey V. Nasonov and Andrey S. Krylov

Laboratory of Mathematical Methods of Image Processing  
Faculty of Computational Mathematics and Cybernetics  
Lomonosov Moscow State University

**Abstract**—Two problems of text images enhancement are considered in the paper: superresolution and deringing. The problem of superresolution is the reconstruction of a high resolution image from several low resolution observations. Regularization method with bilateral total variation stabilizer and bimodal penalty function are used to perform the superresolution. Enhancement of text images which includes deringing and denoising is performed using projection onto convex sets with the control of noise level in the basic edges neighborhood — the area where ringing and noise effects are the most noticeable.

## I. INTRODUCTION

The problems of image upsampling and enhancement are ones of the most important problems in image processing. Despite the increase of quality and resolution of modern camera sensors, it is still actual for upsampling and enhancement of small objects in surveillance applications.

In this paper, we consider the problem of text images enhancement in video sequences. This is important for a variety of applications like the recognition of car plates. We consider two problems: the problem of multi-frame text image upsampling and the problem of single-frame text image deringing.

Single-frame interpolation uses only prior knowledge about the interpolated image to reconstruct missed pixels. As an example, image self-similarity at different resolutions is used in NEDI algorithm [1]. Nevertheless this approach improves the image quality only if a priori information is true.

Multi-frame superresolution is an alternative method to obtain a high resolution image [2]. Superresolution algorithms use several low resolution images to restore pixels of the high-resolution image while a priori information is used for image refinement. Image pixels correspond to camera sensors which have non-zero size, and observed pixel intensities are approximations of continuous image intensity in a certain area. If the object motion and the approximation function are known, then the information from all frames can be used to construct a single high-resolution image. This process is illustrated by fig. 1.

Bayesian approaches are popular for text superresolution [3], [4], [5]. In [4] a maximum a posteriori (MAP) estimator based on a Huber prior and an estimator regularized using the Total Variation norm are used. The paper [5] introduces text-specific bimodal prior that enhances sharpness and contrast. The Teager filter, a quadratic unsharp masking filter, is used to highlight high frequencies which are then combined with the warped and interpolated image sequence following

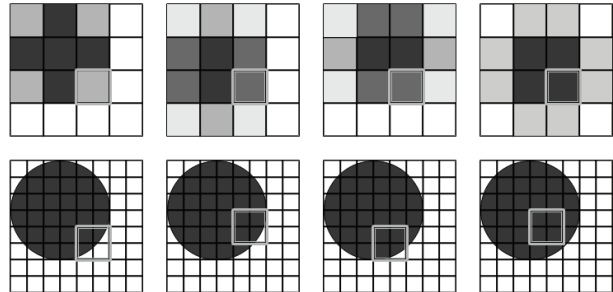


Fig. 1. The correspondence between pixels of low-resolution images (top row) and pixels of high-resolution image (bottom row).

motion estimation using Taylor series decomposition in [6]. Super-resolved text binarization is proposed in [7]. Markov random fields are used for text superresolution in [8].

The paper presents the text images superresolution and enhancement regularization based algorithm which uses bilateral total variation functional and bimodal penalty as stabilizers.

## II. TEXT SUPERRESOLUTION

### A. Mathematical model

We use the approach [9] to pose text superresolution problem. In this approach the superresolution problem is posed as an inverse problem. The corresponding direct problem includes a set of downsampling procedures. It produces the low-resolution images  $u_k$  after motion transformation and downscaling from the high-resolution image  $z$  as:

$$A_k z = u_k, \quad k = 1, 2, \dots, N.$$

The operator  $A_k$  in general case is represented as  $A_k = D H_{cam} F_k H_{atm} z + n$ , where  $H_{atm}$  is the atmosphere blur,  $F_k$  is the motion operator,  $H_{cam}$  is the camera lens blur,  $D$  is a downscaling operator,  $n$  is noise.

The atmosphere blur and the camera lens blur are modeled by a single Gauss filter  $H$ , and the system of equations takes the following form

$$A_k z = D F_k H z = u_k, \quad k = 1, 2, \dots, N. \quad (1)$$

Motion estimation algorithms [10], [11] are used to calculate motion operators  $F_k$ . Many of them use optical flow method for accurate motion estimation [10]. In our work, we consider the motion operators  $F_k$  known.

## B. Regularization

The super-resolution problem described by the inverse problem for the system of equations (1) is an ill-posed problem. We use iterative regularization methods based on Tikhonov regularization [12] to solve it.

The idea of regularization is to introduce an additional restriction that makes the problem well posed. This is typically achieved by adding a stabilizer term  $\Omega[z]$ :

$$z_\alpha = \arg \min_z \left( \sum_{k=1}^N \|A_k z - u_k\|_2^2 + \alpha \Omega[z] \right), \quad (2)$$

where the regularization parameter  $\alpha > 0$  controls importance of the stabilizer term. Iterative gradient and subgradient methods are used to minimize (2).

We use a combination of the bilateral total variation stabilizer [9] and the penalty function based on bimodal prior [5] as the stabilizer:

$$\Omega[z] = BTV[z] + \lambda \sum_{x,y} P_b(z_{x,y}).$$

The bilateral total variation operator  $BTV[z]$  [9] is the adaptation of continuous total variation operator for discrete case. When the bilateral total variation is used as the stabilizer, the resulting image tends to be piecewise constant with sharp edges. This effect is good for text enhancement.  $BTV$  is formulated as follows:

$$BTV[z] = \sum_{s,t=-p}^p \gamma^{|s|+|t|} \|S_x^s S_y^t z - z\|_1,$$

where  $S_x$  and  $S_y$  are shift operators along  $x$  and  $y$  axes by  $s$  and  $t$  pixels respectively. The parameter  $\gamma$  is equal to 0.8. The parameter  $p$  specifies the number of directions. Higher values of  $p$  result in slightly better quality but greater computation time. We use  $p = 1$ .

The bimodal penalty function  $P_b(v)$  looks as:

$$P_b(v) = \frac{(v - v_1)^2 (v - v_2)^2}{\sigma_b^4}.$$

The idea of this function was suggested in [5]. This function has two minima at  $v_1$  and  $v_2$ . These values are intensities of text background and text foreground respectively. Therefore, if the bimodal penalty function is used in the stabilizer term, pixel intensities of the reconstructed image tend to be close to either background or foreground intensity. Parameter  $\sigma_b$  is the normalization coefficient. We use

$$\sigma_b = v_2 - v_1.$$

The weight of the bimodal prior can be set using the parameter  $\lambda$ .

## C. Results

The proposed text superresolution method was tested on both synthetic and real images.

The results obtained for synthetic data are shown in fig. 2. 16 low-resolution images were generated from the given

reference high-resolution image by random shifts followed by downsampling with a factor of 2. Uniform noise in  $[-15, 15]$  range was added to low-resolution images (for the intensity range  $[0, 255]$ ). Then the proposed superresolution method was applied to low-resolution images. The following images were compared: high-quality single-frame image resampling [13], superresolution without bimodal prior ( $\lambda = 0$ ) and superresolution with bimodal prior. In order to estimate the quality of obtained images, we calculated the mean square error (MSE) between binarized ground truth image and obtained images.

An example of superresolution for real text images obtained from a video camera is shown in fig. 3.

In both cases the proposed superresolution method showed better visual and objective results.

## III. ENHANCEMENT

Ringling artifact is an annoying rippling effect which appears near strong edges. This effect is caused by damage or loss of high-frequency information. An example of image with ringling artifact is shown in fig. 4, b). Poor quality text images are usually suffered from ringling artifact and noise.

In [14] the connection between image total variation and ringling effect was shown. Appearing of ringling artifact and increasing noise level results in rise of image total variation. Projection of image containing ringling artifact onto the set of images of bounded total variation reduces ringling artifact [15]. We use the same approach as for superresolution to perform text images enhancement:

$$z_\alpha = \arg \min_z \left( \|z - z_0\|_2^2 + \alpha \left( BTZ[z] + \lambda \sum_{x,y} P_b(z_{x,y}) \right) \right),$$

where  $z_0$  is the corrupted image.

Efficient automatic text image enhancement is not possible without proper quality estimation. To estimate the image quality, we analyze the total variation value of text background. In the ideal case, when the background has constant intensity value, the total variation of the background equals to zero. Therefore, the normalized total variation of text background can be a measure of text image quality.

Now all the background of the text image is suitable for quality estimation. We use the basic edges algorithm [16], [17] to find the areas suitable for quality estimation. Basic edges are sharp edges distant from other edges. We use the basic edges neighborhood area — the area of pixels with the nearest edge being the basic edge and the distance to it greater than  $r$  and less than  $R$ . Ringling and noise artifacts usually are the most noticeable in these areas.

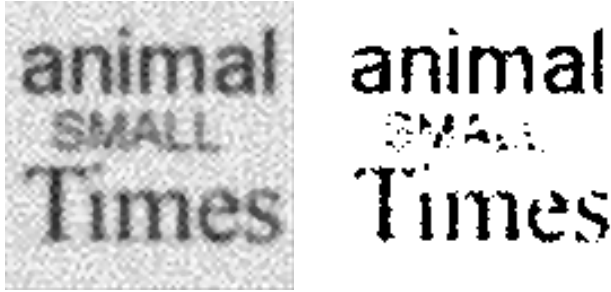
### A. Results

The fig. 4 illustrates the suggested text images enhancement algorithm. The corrupted image was modeled by truncation the 3/4 part of high frequency coefficient after Fourier transform followed by adding the Gaussian noise with a standard deviation of 5. The fig. 4, c) is the result of basic edges and basic edges areas detection. Yellow areas are basic edges, gray areas are edges that are not basic. The green area is the basic edges

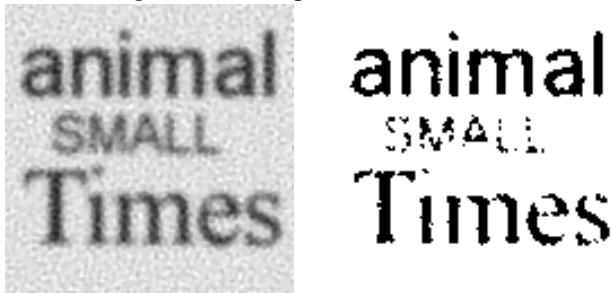


Ground truth image

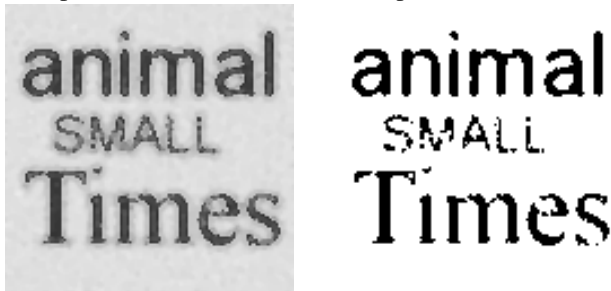
One of noisy low-resolution images, 2x magnified using nearest neighbor method



Single-frame interpolation, MSE = 3202



Superresolution without bimodal prior, MSE = 2752



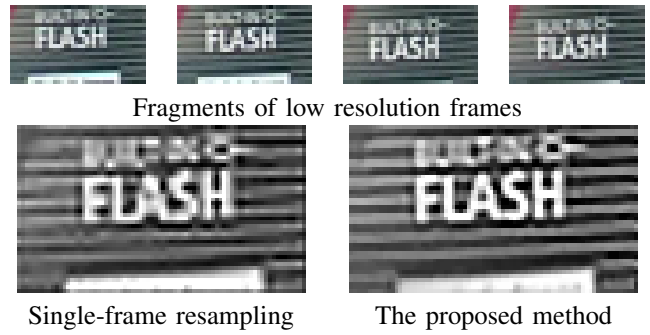
Superresolution with bimodal prior, MSE = 1917

Fig. 2. Superresolution results for synthetic image. 16 low-resolution images were used to reconstruct a high-resolution image with 2x upsampling ratio.

neighborhood where the quality level is estimated. We used  $r = 2$  and  $R = 6$ .

#### IV. CONCLUSION

It was found that regularization algorithms based on joint minimization of total variation and bimodal penalty function are very effective in text enhancement tasks like resampling and noise and ringing reduction. As an example, use of the bimodal prior showed the increase of the quality of the



Fragments of low resolution frames



Single-frame resampling



The proposed method

Fig. 3. Superresolution results for real image. Fragments of 16 low-resolution video frames were used to reconstruct a high-resolution image with 2x upsampling ratio. The left column contains the results of superresolution, the right one — thresholded superresolution results.

Euclidean distance tra  
gion of interest. Right  
he nearest black pixel

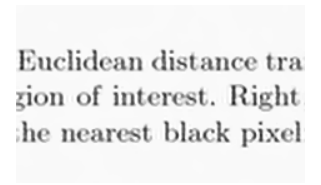
a) Original image

Euclidean distance tra  
gion of interest. Right  
he nearest black pixel

b) Corrupted image,  
MSE = 68.81, TV = 20.49



c) Basic edges areas



d) Result,  
MSE = 47.43, TV = 5.12

Fig. 4. The results for the proposed text image enhancement algorithm.

standard total variation based superresolution method. Automatic control of ringing and suppression by this method were achieved using previously suggested basic edges technique.

#### ACKNOWLEDGMENT

The work was supported by Federal Targeted Programme "R&D in Priority Fields of the S&T Complex of Russia 2007–2013" and by RFBR grant 10-01-00535.

#### REFERENCES

- [1] Xin Li and M.T. Orchard, "New edge-directed interpolation," *IEEE transactions on image processing*, vol. 10, no. 10, pp. 1521–1527, 2001.
- [2] S. Borman and R.L. Stevenson, "Super-resolution from image sequences - a review," *Midwest Symposium on Circuits and Systems*, pp. 374–378, 1998.
- [3] F.J. Cortijo, "Bayesian super-resolution of text image sequences from low resolution observations," *Seventh International Symposium on Signal Processing and Its Applications*, vol. 1, pp. 421–424, 2003.
- [4] David Capel and Andrew Zisserman, "Super-resolution enhancement of text image sequences," *Proceedings of ICPR'2000*, pp. 600–605, 2000.
- [5] Katherine Donaldson and Gregory K. Myers, "Bayesian super-resolution of text in video with a text-specific bimodal prior," *International journal on document analysis and recognition*, vol. 7, no. 2-3, pp. 159–167, 2005.

- [6] Celine Mancas-Thillou and Majid Mirmehdi, "Super-resolution text using the teager filter," *First International Workshop on Camera-Based Document Analysis and Recognition*, pp. 10–16, 2005.
- [7] Thibault Lelore and Frederic Bouchara, "Super-resolved binarization of text based on the fair algorithm," *2011 International Conference on Document Analysis and Recognition (ICDAR)*, pp. 839–843, 2011.
- [8] K.V. Suresh, G. Mahesh Kumar, and A. N. Rajagopalan, "Superresolution of license plates in real traffic videos," *IEEE Transactions on Intelligent Transportation Systems*, vol. 8, no. 2, pp. 321–331, 2007.
- [9] S. Farsiu, D. Robinson, M. Elad, and P. Milanfar, "Fast and robust multi-frame super-resolution," *IEEE Transactions on Image Processing*, vol. 13, no. 10, pp. 1327–1344, 2004.
- [10] A. Bruhn, J. Weickert, and C. Schnorr, "Lucas/Kanade Meets Horn/Schunck: Combining Local and Global Optic Flow Methods," *International Journal on Computer Vision*, vol. 61, no. 3, pp. 211–231, 2005.
- [11] H.V. Le and G. Seetharaman, "A super-resolution imaging method based on dense subpixel-accurate motion fields," *Proc. of the Third International Workshop on Digital and Computational Video*, pp. 35–42, 2002.
- [12] A. N. Tikhonov and V. Y. Arsenin, *Solutions of Ill-Posed Problems*, WH Winston, Washington DC, 1977.
- [13] A. Lukin, A. Krylov, and A. Nasonov, "Image interpolation by super-resolution," *16th International Conference Graphicon'2006*, pp. 239–242, 2006.
- [14] S. Mallat, *A Wavelet Tour of Signal Processing*, Academic Press, 1999.
- [15] A. Krylov and A. Nasonov, "Adaptive total variation deringing method for image interpolation," *Proceedings of International Conference on Image Processing (ICIP'08)*, pp. 2608–2611, 2008.
- [16] A. V. Nasonov and A. S. Krylov, "Image enhancement quality metrics," *21-th International Conference on Computer Graphics GraphiCon'2011*, pp. 128–131, 2011.
- [17] A. V. Nasonov and A. S. Krylov, "Finding areas of typical artifacts of image enhancement methods," *Pattern Recognition and Image Analysis*, vol. 21, no. 2, pp. 316–318, 2011.