Fast Super-Resolution from video data using optical flow estimation

Andrey Krylov
Faculty of Computational Mathematics and Cybernetics, Moscow Lomonosov State University, Moscow, Russia
kryl@cs.msu.ru

Andrey Nasonov
Faculty of Computational Mathematics and Cybernetics, Moscow Lomonosov State University, Moscow, Russia
nasonov@cs.msu.ru

Abstract

Regularization-based and a fast non-iterative methods using optical flow estimation are suggested for video data super-resolution with correction of non-uniform illumination.

1. Introduction

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

There are two groups of video SR algorithms: learning- and reconstruction-based. Learning-based algorithms enhance the resolution of a single image using information on the correspondence of sample low- and high-resolution images. Reconstruction-based algorithms use only a set of low-resolution images to construct high-resolution image. More detailed introduction into video SR problems is given in [1], [2].

The majority of reconstruction-based algorithms use camera models [3] for downsampling the high-resolution image. The problem is posed as error minimization problem

$$z_R = \arg\min_z \sum_k \left\| A_k z - w_k \right\|, \qquad (1)$$

where z is reconstructed high-resolution image, w_k is k-th low-resolution image, A_k is a downsampling operator which transforms high-resolution image into k-th low-resolution image. Different 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 motion deformation, H_{cam} is camera lens blur which is usually modeled by Gauss filter, D is a decimation operator, n is a noise which is usually ignored.

Various models of warping operator F_k are used. The simplest model is a translation model. In this case, k-th frame is considered as a shifted first image.

Translation model is not appropriate for SR problem, when the motion is not constant. Different motion models are used [4], [5]. 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 more effective to calculate the motion of pixels which belong to edges and corners [5].

Optical flow estimation algorithms are used in the case of small motion vectors. They produce accurate enough result. The key idea of flow estimation is the following representation of the consecutive frames:

$$I(x + u(x, y), y + v(x, y)) = J(x, y), \qquad (2)$$

where $u(x, y)$ and $v(x, y)$ are components of motion vector field $\bar{v} = (u, v)$. Frames are considered to be smooth enough and differentiable, so $I(x+u, y+v)$ can be approximated as

$$I(x+u, y+v) \approx I(x, y) + \frac{\partial I(x, y)}{\partial x} u + \frac{\partial I(x, y)}{\partial y} v. \qquad (3)$$

Under assumption (3), equation (2) takes the form

$$I_x(x, y)u(x, y) + I_y(x, y)v(x, y) = I_t(x, y), \qquad (4)$$

where $I_t(x, y) = J(x, y) - I(x, y)$. Since motion vectors of adjacent pixels are close to each other, various additional constraints are used. For example, in [6] partial derivatives I_x, I_y, I_t are smoothed. Regularization algorithms are also used [7], [8]. To improve the accuracy for video sequences with non-uniform illumination, image gradient is used instead of image intensity in (2) [9].

2. Our approach

We propose a reconstruction-based algorithm with optical flow estimation model incorporating both translation model and variational approach for optical flow estimation.

2.1. Optical flow estimation

The proposed optical flow estimation algorithm consists of two steps:

1. At the first step, we find the best shift between two frames to reduce average length of motion vectors to achieve better results at the second step. For every frame, we find 10 most significant points obtained from Harris detector [10]. Then we find the best shift vector so that the points from the first frame should fit the points from the second frame. On practice, for consecutive frames, about 7–8 points from both frames correspond each other, so we assume that at least 5 points correspond each other. We minimize the following functional

$$F(\bar{v}) = \sum_{k=1}^{5} \left\| P_k + \bar{v} - Q_k \right\|^2 , \tag{5}$$

where P_k are key points from the first frame, Q_k — from the second, and values $\left\| P_k + \bar{v} - Q_k \right\|$ are the first five minimal penalty values for the given $\bar{v}$. To minimize the functional (5), we calculate its value for every pair $\bar{v} = (P_i - Q_j), \quad i, j = 1,2,...,10$.

2. After the rough motion estimation, we apply a modification of Kanade-Lucas method [6]. This modification adds image gradient conditions:

$$\begin{cases} I(x+u, y+v) = J(x, y), \\ I_x(x+u, y+v) = J_x(x, y), \\ I_y(x+u, y+v) = J_y(x, y). \end{cases} \tag{6}$$

For every point (x, y) we minimize the following functional:

$$\begin{aligned} F(u,v) &= \alpha_1 \mid I_x u + I_y v - I_t \mid + \\ &+ \alpha_2 (\mid I_{xx} u + I_{xy} v - I_{xt} \mid + \mid I_{xy} u + I_{yy} v - I_{yt} \mid) . \end{aligned} \tag{7}$$

Here weights α_1 and α_2 indicate the importance of the conditions (6). To minimize (7), we solve Euler equation

$$\begin{cases} a_{11} u + a_{12} v = a_{13}, \\ a_{21} u + a_{22} v = a_{23}, \end{cases} \tag{8}$$

where

$$a_{11} = \alpha_1 I_x^2 + \alpha_2 (I_{xx}^2 + I_{xy}^2) ,$$

$$a_{12} = \alpha_1 I_x I_y + \alpha_2 (I_{xx} I_{xy} + I_{xy} I_{yy}) ,$$

$$a_{13} = \alpha_1 I_x I_t + \alpha_2 (I_{xx} I_{xt} + I_{xy} I_{yt}) ,$$

$$a_{21} = \alpha_1 I_x I_y + \alpha_2 (I_{xx} I_{xy} + I_{xy} I_{yy}) ,$$

$$a_{22} = \alpha_1 I_y^2 + \alpha_2 (I_{xy}^2 + I_{yy}^2) ,$$

$$a_{23} = \alpha_1 I_y I_t + \alpha_2 (I_{xy} I_{xt} + I_{yy} I_{yt}) .$$

The conditions (8) are independent for different points (x, y) , so the motion vector field is not accurate. To add the condition of similarity of the motion of close pixels, we use the approach used in Kanade-Lucas method [6]. We spatially smooth all the coefficients in (8) using Gauss filter. We use Gauss filter with a radius of 5.

The proposed flow estimation forms warping operator F_k used in (1). A result is shown in Fig. 1.

2.2. Regularization

The SR problem (1) is ill-conditioned, so we use Tikhonov regularization approach [11]. We use l_1 norm $\|z\|_1 = \sum_{i,j} | z_{i,j} |$.

$$z_R = \arg\min_z \left(\sum_k \left\| A_k z - v_k \right\|_1 + \alpha f(z) \right). \tag{9}$$

We choose bilateral total variation functional [3] $f(z) = \sum_{-p \le x,y \le p} \gamma^{|x|+|y|} \left\| S_{x,y} z - z \right\|_1$ as a stabilizer, $S_{x,y}$ is a shift operator along horizontal and vertical axis by x and y pixels respectively, $\gamma = 0.8$, $p = 1$.

The functional (9) is minimized by subgradient method [2], [12]. A result is shown in Fig. 2.

2.3. Fast super-resolution

Solving of (9) is time-consuming and it is often important to get a fast approximation of SR problem. Our approach is close to [13]. The algorithm looks as follows:

1. Fix the first frame w_1 and calculate the flow between w_1 and w_k, $k = 2,3,...,n$, n is the number of input frames. The flow between w_1 and w_1 is zero-filled.

2. Upsample every frame w_k taking into account optical flow estimation for the frame to compensate the motion and to make frame close to the first frame:

$$W_k = F_k U w_k , \tag{10}$$

where U is Gauss upsampling operator. Fast implementation of the procedure to calculate $F_k U$ is described in Section 3.

3. Calculate an average image

$$z = \frac{1}{n}\sum_{k=1}^{n} W_k \,. \quad (11)$$

4. Deblur the resulting image.
A result is shown in Fig. 3.

2.4. Illumination correction

We have applied the proposed SR methods for input frames with non-uniform illumination. In this case equation (4) gives bad results. So, to estimate the flow, we use only information about image gradient, i.e. $\alpha_1 = 0$ and $\alpha_2 = 1$ in (7). Using this estimation fast approximation has successfully processed the frame, while regularization-based method (9) did not show good results. A result is shown in Fig. 4. Another approach to illumination correction by Empirical Mode Decomposition for regularization-based method was suggested by us in [2].

3. Numerical methods

To upsample the image for fast SR, we use modified Gauss resampling to calculate (10). Original Gauss method for scale factor p looks as follows:

$$W_k(px, py) = \frac{\sum_{(x_i, y_j)\in\Omega} e^{\frac{-(x-x_i)^2-(y-y_i)^2}{2\sigma^2}} w_k(x_i, y_i)}{\sum_{(x_i, y_j)\in\Omega} e^{\frac{-(x-x_i)^2-(y-y_i)^2}{2\sigma^2}}}, \quad (12)$$

where w_k is the low-resolution image, W_k is the high-resolution image, σ is a Gauss filter radius (we choose $\sigma = 0.4$), (x_i, y_j) are grid points where w_k is known.

We apply the warping operator F_k directly to (12) by shifting the grid points by motion vectors

$$W_k(px, py) = \frac{\sum_{\Omega} e^{\frac{-(x-x_i+u_{i,j})^2-(y-y_i+v_{i,j})^2}{2\sigma^2}} w_k(x_i, y_i)}{\sum_{\Omega} e^{\frac{-(x-x_i+u_{i,j})^2-(y-y_i+v_{i,j})^2}{2\sigma^2}}}. \quad (13)$$

To perform fast computation of (13), we represent both numerator and denominator as convolutions of delta functions with Gauss filter. In discrete form, we form two images (W_k^* and W_k^{**}), initially zero-filled. Next for every point (x_i, y_j) from w_k, we calculate its coordinates on upsampled image W_k:

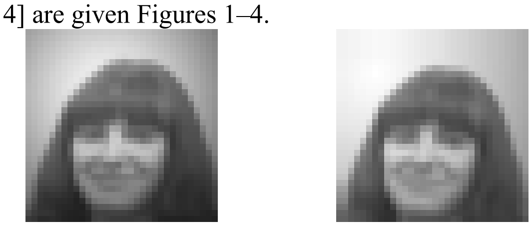
$$(x_i, y_j) \rightarrow (p(x_i - u_{i,j}), p(y_i - v_{i,j})) = (x_{i,j}^*, y_{i,j}^*)\,.$$

Then we add the value $w_k(x_i, y_j)$ to $W_k^*(x_{i,j}^*, y_{i,j}^*)$ and 1 to $W_k^{**}(x_{i,j}^*, y_{i,j}^*)$. If coordinates $(x_{i,j}^*, y_{i,j}^*)$ are not integer, then we approximate the convolution with a single delta function as a convolution with a sum of delta functions defined at integer coordinates using bilinear interpolation.

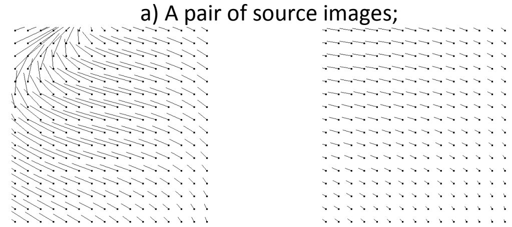
After the images W_k^* and W_k^{**} formed, we apply Gauss filter to both, then divide W_k^* on W_k^{**} elementwise $W_k = W_k^* / W_k^{**}$.

4. Results

Results of the comparison of the proposed SR method with single-frame linear methods and regularization-based non-linear interpolation method [14] are given Figures 1–4.



a) A pair of source images;



b) Kanade-Lucas method; c) The proposed method.

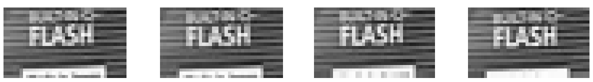
Figure 1. Optical flow estimation for images with non-uniform illumination.

a) Source image examples;

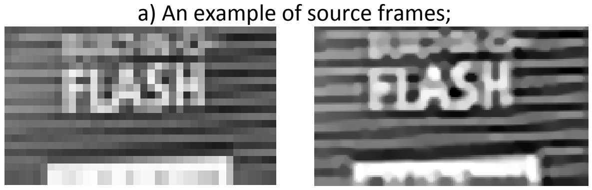
b) Box filter interpolation; c) Bilinear interpolation;

d) Non-linear interpolation; e) Super-resolution

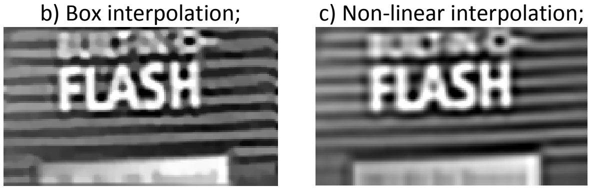
Figure 2. Super-resolution by a factor of 4 for 10 images with uniform illumination.



a) An example of source frames;



b) Box interpolation; c) Non-linear interpolation;



d) Super-resolution; e) Fast super-resolution.

Figure 3. Real-time super-resolution by a factor of 4 for 12 images with uniform illumination obtained from a camera without pre-processing.

a) Source images; b) Super-resolution.

Figure 4. Fast super-resolution by a factor of 2 for 8 images with non-uniform illumination.

5. Conclusion

Tikhonov regularization-based and a fast non-iterative methods using optical flow estimation for video data SR have been suggested. Fast method is less time-consuming than non-linear resampling method and of the same SR quality. Tikhonov regularization-based method gives the best quality but it is slower.

The work was supported by grant RFBR 06-01-39006-ГФЕН.

6. References

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

[2] A.S. Krylov, A.V. Nasonov, D.V. Sorokin, "Face image super-resolution from video data with non-uniform illumination", *Proc. Int. Conf. Graphicon*2008, pp. 150–155.

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

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

[5] Ha V. Le, and 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.

[6] B.D. Lucas, T. Kanade, "An iterative image registration technique with an application to stereo vision", *Proc. of Imaging understanding workshop*, 1981, pp. 121–130.

[7] A. Bruhn, J. Weickert, and C. Shnorr, "Lucas/Kanade Meets Horn/Schunck: Combining Local and Global Optic Flow Methods", *International Journal of Computer Vision*, Vol. 61, No. 3, 2005, pp. 211–231.

[8] J. Weickert, C. Shnorr, "Variational Optic Flow Computation with a Spatio-Temporal Smoothness Constraint", *J. Math. Im. & Vis.*, Vol. 14, 2001, pp. 245–255.

[9] T. Brox, A. Bruhn, N. Papenberg, and J. Weickert, "High Accuracy Optical Flow Estimation Based on a Theory for Warping", *Proc. 8th European Conf. on Comp. Vis.*, Vol. 4, 2004, pp. 25–36.

[10] C. Harris, and M.J. Stephens, "A combined corner and edge detector", *In Alvey Vision Conf.*, 1998, pp. 147–152.

[11] A.N. Tikhonov, and V.Y. Arsenin, *Solutions of Ill Posed Problems*, WH Winston, Washington DC, 1977.

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

[13] F. Lin, C. Fookes, V. Chandran, and S. Sridharan, "Investigation into Optical Flow Super-Resolution for Surveillance Applications", *Proceedings of APRS Workshop on Digital Image Computing*, February 2005, pp. 73–78.

[14] A.S. Lukin, A.S. Krylov, A.V. Nasonov, "Image Interpolation by Super-Resolution", *Proc. Int. Conf. Graphicon2006*, pp. 239–242.