Super Resolution with Edge-Constrained Motion Estimation

Yue Zhuo, Jiaying Liu*, Mading Li and Zongming Guo
Institute of Computer Science and Technology, Peking University, Beijing, China

***Abstract*—Motion estimation is a critical step for most reconstruction-based super resolution methods. However, accurate motion estimation is difficult, and the unavoidable error degrades performance of super resolution rapidly. In this paper, we present a robust way to perform super resolution by improving motion estimation. Starting with feature points matching, we compute the local motion parameter of feature point correspondences by using the weighted Lucas-Kanade algorithm. Then accurate motion field is estimated by support region search, which refers to edge information and considers discontinuities of motion boundary and consistency of motion field. Experimental results validate the efficacy of each step in the proposed algorithm and show that it produces super resolved images with higher quality.**

I. Introduction

Reconstruction-based super resolution (SR) aims to reconstruct one high resolution (HR) image from multiple low resolution (LR) images. Since these observed LR images are blurry and noisy, what SR needs to do is not only to increase the resolution of images but also to restore them to better quality. A good method is supposed to be robust to the degrading processing. The key of multi-frame SR is to relate all the input images. Motion estimation based methods map each pixel in LR images to one position in the HR image. In most work, the underlying motion is represented by a simple parametric model, which assumes there is only global motion between input images. To handle more complex scene, optical flow is used to acquire motion vector of each pixel. The advantage of optical flow is that it does not need to assume the motion field satisfies some kind of parametric model. However, the unavoidable motion estimation error may lead to annoying artifacts in super resolved images. On the other hand, Potter *et al*. [1] proposed another motion-estimation-free method, which could be regarded as a probabilistic motion model. But the way it computed the motion probability still reflected its assumption on the motion field.

In the past decades, researchers have made great progress in both motion parametric model based methods and optical flow methods. Shen *et al*. [2] modeled the scene by multiple independently moving objects and approximated each object by a parametric model. Then they computed motion fields, segmentation fields and HR images in an alternate manner. Bruhn *et al*. [3] combined local method Lucas-Kanade and global method Horn/Schunck, and formulated them in one equation. Recently Su *et al*. [4] proposed a method only to extract reliable motion field rather than dense flow. Its motivation was that feature point correspondences are usually more precise and robust than dense flow field. Based on the local optical flow, they extracted a support region for each corresponding pair. Their methods gave us some useful information, and there are three points we consider:

First, a complex motion field can be decomposed into separate local motion, which can be represented using a simple model. Local motion estimation is more precise at feature points because feature points are stable and distinguishable in images. According to [3], the local method Lucas-Kanade is more robust to noise than global method, but there still might be some space to improve the performance of the local method at feature points. Second, motion boundaries are reference for local motion estimation. In [2], the motion boundaries were detected based on the motion field and then were used for next motion estimation. Instead, what we want is to compute the motion field and preserve the motion boundaries simultaneously. Finally, motion continuity is an important prior in optical flow, e.g. in Horn/Schunk. When there is no reliable flow, the smoothness prior can fill in motion from the neighborhood. While in local method, the smoothness of global motion field does not be involved. So we consider incorporating the smoothness constrain to the motion field obtained by local method.

In this paper, we try to find solutions to above problems. To make the local motion estimation at feature points more precise, we use the weighted Lucas-Kanade. For computing the motion field, we propose an edge-constrained motion estimation algorithm, which refers to edge information to preserve the motion boundaries. After motion estimation, we smooth the motion field by performing morphological opening.

The rest of the paper is organized as follows. In Section II, we formulate the SR problem. Section III proposes the edge-constrained motion estimation algorithm. Experimental results are presented in Section IV. Finally, concluding remarks are given in Section V.

II. Super Resolution problem formulation

The problem of SR is that we have a set of n observed LR images $L_1, \ldots, L_n$ and want to reconstruct HR image H with ideal quality. The imaging model is formulated as follows:

$$L_k = DB_kM_kH + N_k, \qquad k = 1, \ldots, n \tag{1}$$

*Corresponding author. This work was supported by National Natural Science Foundation of China under contract No.61101078, Beijing Municipal Natural Science Foundation under contract No.4102025, and Doctoral Fund of Ministry of Education of China under contract No.20110001120117.

where M_k, B_k, D denote warping matrix, blurring matrix and downsampling matrix respectively, and N_k is the additive noise. The super resolved image H can be solved by minimizing the following formula:

$$H = \arg\min_{H} \left(\sum_{k} \|DB_k M_k H - L_k\|^2 + \lambda\Gamma(H) \right), \quad (2)$$

where $\Gamma(H)$ is the regularization term and λ controls the effect of regularization.

Because of occlusion or motion estimation error, some pixels cannot find their corresponding motion. So we further use observable map to remove the outliers from the optimization formula, which can be written as follows:

$$H = \arg\min_{H} \left(\sum_{k} \|V_k^L (DB_k M_k H - L_k)\|^2 + \lambda\Gamma(H) \right), \quad (3)$$

where V_k^L is a 0/1 diagonal matrix.

Reference [4] had proved that under mild conditions, (3) can be reformatted as:

$$H = \arg\min_{H} \left(\sum_{k} \|DB_k V_k M_k H - L_k\|^2 + \lambda\Gamma(H) \right), \quad (4)$$

In this work, the matrix B_k and D are assumed known.

III. Edge-Constrained Motion Estimation

The framework of the proposed algorithm is showed in Fig. 1, in which we first detect the feature point correspondences. Then we compute the local parametric model for each pair by using the weighted Lucas-Kanade algorithm. Support regions are extracted for each local parametric model based on edge information (colored white in Fig. 1(c)). In Fig. 1(c), one color stands for one support region of a certain feature point. Finally, we obtain motion field and observable map. Unobservable area is colored black in Fig. 1(d).

A. *Improved local motion estimation*

Since the proposed motion estimation results are built on the parametric model of feature points, the accuracy of initial estimation of motion parameters at feature points is of great importance. We first use SIFT to extract feature points in the reference frame and candidate frame and then match these points. To make the feature points matching more robust, a double check is performed as [4] did.

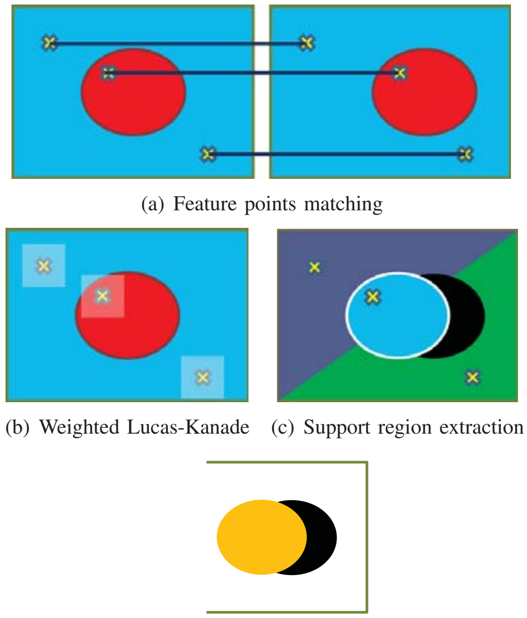
In this work, we use affine model to describe local optical flow. The affine model parameter T_i contains six unknown variables:

$$T_i = \begin{bmatrix} 1+p_1 & p_3 & p_5 \\ p_2 & 1+p_4 & p_6 \\ 0 & 0 & 1 \end{bmatrix}, \quad (5)$$

which minimizes the following formula:

$$\sum_{x} H^r(T_i x) - H^c(x), x \in N(f_i^c), \quad (6)$$

where H^r and H^c are interpolated reference frame and candidate frame, f_i^c is a matched feature point in candidate frame, and $N(f_i^c)$ is its neighborhood.



(a) Feature points matching

(b) Weighted Lucas-Kanade

(c) Support region extraction

(d) Motion field

Fig. 1. Framework of the proposed algorithm.

The above equation can be solved by Lucas-Kanade algorithm [5]. In traditional methods, the size of neighborhood $N(f_i^c)$ is fixed. Reference [4] proposed a way to choose the window size adaptively. However, their methods still used a uniform mask, which cannot detect motion boundaries. Ren [6] used intervening contour to compute the affinity to assign a weight to each point inside the window, which meant the probability that this point had the same motion parameter with the center point. Inspired by their work, we use Nonlocal means (NLM) to produce the mask. For each pixel (i, j) in the window, its weight $w(i, j, k, l)$ is measured by its similarity with the center pixel (k, l) as follow:

$$w(i, j, k, l) = \frac{1}{C(k, l)} \exp\left\{ -\frac{\|R_{i,j}Y - R_{k,l}Y\|_{2,a}^2}{2\sigma^2} \right\}, \quad (7)$$

where $R_{i,j}$ represents an operator which extracts a patch of a fixed and predetermined size $(q \times q)$ from an image and gets a vector of length q^2, σ acts as a smoothing parameter controlling the effect of the grey-level difference between these two image patches, Y is the image, a is the standard deviation of Gaussian kernel and $C(k, l)$ is the normalization constant defined as

$$C(k, l) = \sum_{(i,j) \in N(k,l)} \exp\left\{ -\frac{\|R_{i,j}Y - R_{k,l}Y\|_{2,a}^2}{2\sigma^2} \right\}. \quad (8)$$

With NLM mask and large window size, we can choose window shape adaptively. Since our goal is to detect motion boundaries, we reduce the influence of texture by decomposing image and compute NLM mask in structure layer. The weighted affine model is defined as follows:

$$\sum_{x} w(x)[H^r(T_i x) - H^c(x)], x \in N(f_i^c). \quad (9)$$

where $w(x)$ is the weight computed by NLM scheme.

B. Edge-constrained motion field search

Having the feature corresponding pair (f_i^r, f_i^c) and local affine parameter T_i, we extract support region for each pair, which means all the pixels in a support region share the same affine parameter. One simple way is to use confidence map C_i as follows:

$$C_i = \begin{cases} 1, & |B_k(H^r(T_i x) - H^c(x))| < \eta_c \\ 0, & |B_k(H^r(T_i x) - H^c(x))| \geq \eta_c \end{cases} \tag{10}$$

where η_c is a predetermined threshold. This method can make a correct decision in most time, but in motion boundaries it may include outliers due to the effect of blur operator. It leads to blurriness in super resolved results in motion boundaries, or even produces artifacts because of false information.

To improve the robustness of confidence map and preserve motion boundaries, we propose an edge-constrained algorithm for motion field search. In order to reduce outliers, we stop expanding support region at edge points which are found by Canny detector. Edge points provide useful information for support region searching, but there may contain gaps between points. To overcome this problem, we refer to the trapped ball method proposed in [7]. The basic idea of trapped ball method is that it uses a ball instead of a single pixel to perform searching, and the ball is trapped at small gaps between edge points. At first, the ball has a large radius, and then its radius becomes smaller when the ball cannot move any more. Finally, there are only regions closed by edge points left. Then we search these regions by priority order. The details of proposed algorithm are described as Algorithm 1.

With the affine parameter T_i and the corresponding support region indicated in matrix $label$, we can extract motion field and observable map. The process of motion field search is showed in Fig.2. Since Algorithm 1 does not consider the continuity of the motion field, we can see that there are still some holes in the motion field after searching as Fig. 2(d) shows. While in optical flow algorithm, the smoothness prior of motion field can fill in motion from the neighborhood when there is no reliable flow. To fill the holes of motion field without damaging motion boundaries, we perform morphological opening, which is erosion followed by dilation, on the motion field.

Algorithm 1: Edge-constrained Motion Field Search

Input:
$f_i^c (i = 1, \ldots, m)$: matched feature points in candidate frame;
$C_i (i = 1, \ldots, m)$: confidence map;
$radius$: initial size of searching ball;
Output:
$label$: $label(x) = i$ represents that x is in i-th support region;

Initialization:
Create a zero matrix $label$ of the same size as H^c;
Create a queue Q, if $C_i(f_i^c) = 1$ then Push f_i^c into Q;
Create a priority queue P;
Searching:

```
while radius ≥ 0 do
    while Q is not empty do
        Pop x from head of Q, i = label(x);
        for y ∈ 4-connected neighbors of x do
            if no edge points inside the ball or
            radius = 0 then
                if label(y) = 0 and C_i(y) = 1 then
                    label(y) = i;
                    if y is an edge point then
                        Push y into P;
                    else
                        Push y into Q;
                    end
                end
            end
        end
    end
    radius = radius − 1;
end
while P is not empty do
    Pop x from head of P, i = label(x);
    for y ∈ 4-connected neighbors of x do
        if label(y) = 0 and C_i(y) = 1 then
            label(y) = i;
            Push y into P;
        end
    end
end
```

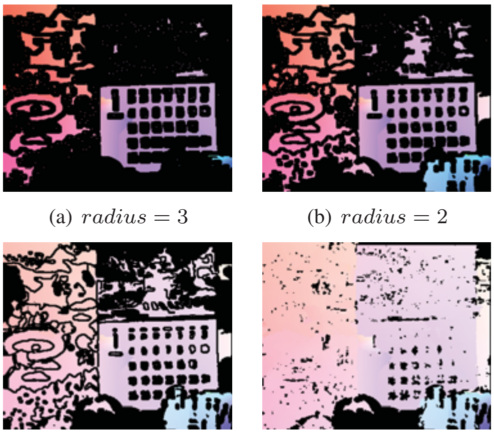
IV. Experimental Results

In the experiments, seven consecutive frames are extracted from each original HR video of size 352×288 and blurred using a 3×3 Gaussian filter with deviation 1, decimated by a factor of 2×2 and then contaminated by an additive noise with standard deviation 2. For each test sequence, the fourth frame is set as the reference frame and super resolved.

We set $radius = 3$, $\eta_c = 10$. The window size for estimating the local affine parametric model is 41×41. We use total variation (TV) as regularization and set $\lambda = 0.001$. To solve Eq.(4), we use gradient descend, and the number of iteration is 50.

To validate the efficacy of NLM mask, we compare the PSNR of SR results with and without using the NLM mask. For the NLM mask, we choose the σ from 20 to 40 and select an appropriate value in each test. The results are summarized in Table I, which shows that NLM mask can improve the SR performance in most cases, but its gain is small.

In the above experiments, we do not perform morphological opening. Then we perform the proposed algorithm with and without opening on motion field to see the difference. The PSNR gains over bicubic interpolation are showed in Fig.3, which proves the smoothness of motion field is an important prior. Finally, we compare different SR algorithms including SRWDF [4] and motion-estimation-free algorithm NLM SR [1]. We download code of SRWDF from the author's website



(a) $radius = 3$ (b) $radius = 2$

(c) $radius = 1$ (d) $radius = 0$

Fig. 2. The process of motion field search with the radius of searching ball decreasing.

TABLE I
PSNR OF SR RESULTS WITH AND WITHOUT NLM MASK.

Sequence	Uniform	NLM Mask
mobile	23.58	**23.59**
coastguard	28.57	**28.68**
ice	33.24	**33.25**
stefan	**28.37**	28.26
soccer	32.63	**32.76**

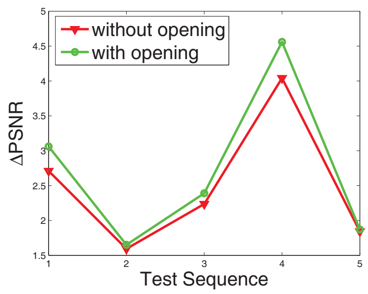


Fig. 3. The PSNR gains over Bicubic with and without morphological opening.

. For fair comparison, we perform support region algorithm by using the original code and reconstruct the high resolution image by using our TV regularization SR code. The method of NLM SR is implemented by ourselves according to [1] with 7×7 patch size, 41×41 window size and $\sigma = 2.2$. The PSNR results of all the tests are summarized in Table II. The proposed method generates the best results steadily. Fig.4 gives the visual comparison in stefan. We can see that the Bicubic and NLM SR lose some details compared to SRWDF and the proposed method. Furthermore, the proposed method produces more satisfactory result in edges than SRWDF because we consider smoothness of motion field.

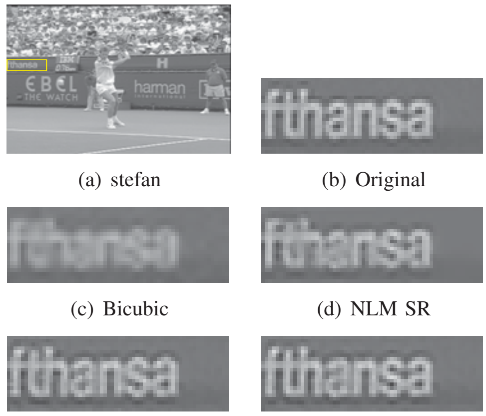
V. Conclusions

In this work, we have two main contributions. First, we use image decomposition and NLM mask to improve the estimation of local parametric model. Second, we propose an edge-constrained motion field search algorithm which maintains motion boundaries while computing motion field and considers smoothness of motion field. The experimental results have shown that the proposed algorithm gets better performance on SR.

http://www.hengsu.tk/research

TABLE II
PSNR RESULTS FOR THE TEST SEQUENCES

Sequence	Bicubic	NLM SR	SRWDF	Proposed
mobile	20.88	23.46	23.62	**23.94**
coastguard	27.09	27.60	27.66	**28.74**
ice	31.01	33.17	31.50	**33.40**
stefan	24.22	27.15	27.72	**28.78**
soccer	30.92	31.36	31.82	**32.79**



(a) stefan (b) Original

(c) Bicubic (d) NLM SR

(e) SRWDF (f) Proposed

Fig. 4. Visual comparison of the super resolved image in stefan.

References

[1] M. Potter, M. Elad, H. Takeda and P. Milanfar, "Generalizing the Nonlocal-Means to Super-Resolution Reconstruction," *IEEE Trans. on Image Processing,* vol. 19, no. 1, pp. 36-51, January 2009.

[2] H. Shen, L. Zhang, B. Huang and P. Li, "A MAP Approach for Joint Motion Estimation, Segmentation, and Super Resolution," *IEEE Trans. on Image Processing,* vol. 16, no. 2, pp. 479-490, February 2007.

[3] A. Bruhn, J. Weickert and C. Schnorr, "Lucas/Kanade meets Horn/Schunck: combining local and global optical flow methods," *International Journal of Computer Vision,* vol. 61, no. 3, pp. 211-231, 2005.

[4] H. Su, Y. Wu, and J. Zhou, "Super-Resolution Without Dense Flow," *IEEE Trans. on Image Processing,* vol. 21, no. 4, pp. 1782-1795, April 2012.

[5] B. D. Lucas and T. Kanade, "An iterative image registration technique with an application to stereo vision," *Proceedings of Imaging Understanding Workshop,* pp. 121-130, 1982.

[6] Xiaofeng Ren, "Local Grouping for Optical Flow," *Computer Vision and Pattern Recognition,* pp. 1-8, June 2008.

[7] S. Zhang, T. Chen, Y. Zhang, S. Hu and R. Martin, "Vectorizing Cartoon Animations," *IEEE Trans. on Visualization and Computer Graphics,* vol. 15, no. 4, pp. 618-629, July 2009.