A PRACTICAL AND ADAPTIVE FRAMEWORK FOR SUPER-RESOLUTION

Heng Su[*1], *Liang Tang*[2], *Daniel Tretter*[3] *and Jie Zhou*[4]

[1,4]Dept. of Automation, Tsinghua Univ., China; [2]HP Labs China; [3]HP Labs Palo Alto

ABSTRACT

In this paper a novel practical and adaptive framework for super-resolution is proposed. Existing super-resolution algorithms are limited by several assumptions, fit different imaging environments; hence they may fail when facing real complex scenes. We propose to divide the target image into adaptive-sized blocks, and apply different conventional algorithms to different parts of the image with different characteristics. The proposed framework is of high extensibility and generality; various super-resolution and single frame image enhancement technologies can be adaptively assembled to improve the robustness and result quality of SR operation. Experiments with real-life video indicate encouraging improvements.

Index Terms— Super-resolution, registration error, block based, de-block

1. INTRODUCTION

Super-resolution image reconstruction technology combines information from a set of successive low-resolution (LR) frames of the same scene to generate a relatively high-resolution (HR) image. Currently super-resolution technologies have been extensively studied, and widely used in video printing, satellite imaging, video compression, etc.

The basic idea of super-resolution is to exploit additional information from successive LR frames with subpixel displacements, and then synthesize a HR image or sequence. Early SR methods solve the problem in frequency domain which, however, are usually restricted to global translational motion and linear space invariant blur[1]. Most current active algorithms calculate SR problem in the spatial domain. IBP (Iteration Back Projection) algorithm estimates the HR image by iteratively back projecting the error between simulated LR images and the observed ones. MAP (Maximum *a Posteriori*)[2] approaches adopt the prior probability of target HR images to stabilize the solution space under a Bayesian framework and, POCS (Projection on Convex Sets)[3] tends to consider the solution as an element on a convex set defined by the input LR images. However, these approaches have higher computational costs. Every algorithm has its own assumptions and hence is restricted in different kinds of imaging environments. For example, MAP methods get better results where there are suitable prior knowledge (like face image SR); iterative methods fit image area with small registration error, otherwise error could be accumulated during iteration[4]. So, one may expect that better results could be obtained by combining different SR algorithms into a single framework and apply to a scene with different region characteristics, in order to improve the robustness and the quality of results. That is the main motivation of our work.

Most SR algorithms consist of two steps, the registration step and synthesis step. In the registration step, transformation parameters are estimated. In the synthesis step, registration results are used to estimate the target HR image. Many of the algorithms mentioned focus on the synthesis step. However, researchers are beginning to realize the significance of the registration step[5]. Unfortunately, limited to current computer vision and image processing technology, it is impossible to register LR images accurately in all cases, especially in the presence of large scale local motion.

A variety of work has been done to solve the registration problem. Milanfar[2] states that using first order norm instead of second order norm increases the robustness against registration errors. Joint estimation method[6] estimates motion vectors and HR image simultaneously, but can suffer from convergence problems. Another kind of algorithm, known as the confidence map[7] or channel adaptive regularization, tries to reduce the contribution of LR images or parts with large registration error in synthesis procedure.

On the other hand, single frame image enhancement is a conventional and mature topic in image processing, which is still active currently. [8] is a recent approach that introduces an edge-forming algorithm based on PDE model. The technology can also be applied in SR processing where there are no accurate corresponding image pieces.

Based on the prior work, a practical extendable super-resolution framework is proposed in this paper. The target image is divided into adaptive-sized blocks, which are classified into categories by their features, and then different algorithms are applied to blocks in different categories. We focus on the SR problem for large-scaled local motion images, and reduce computational cost as far as possible. A de-block process is also applied to reduce the block edge effects. As far as we know, such a framework making use of different algorithms

*This work was done when Heng Su was an intern student at HP labs China.

in a single SR procedure is completely novel.

In our framework, algorithms of SR, single image enhancement and interpolation can be combined together, and the output quality is improved by picking a suitable algorithm for each block. A significant advantage of our framework over algorithms like confidence map is that the proposed algorithm has more flexibility; we can analysis image blocks not only by local registration error, but also by the magnitude of motion fields, smoothness, texture, and all kinds of features of images.

In the next section, the framework details are presented. Experimental results of the proposed method are illustrated in Section 3. Finally, in Section 4 we provide conclusions and thoughts of possible future work.

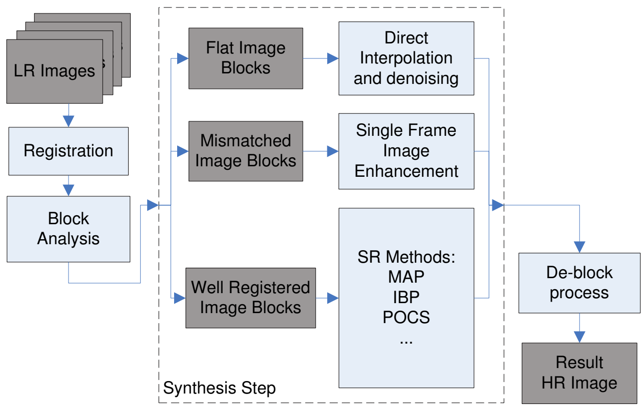

Fig. 1. Sample flow chart of the framework

2. BLOCK-BASED SUPER-RESOLUTION FRAMEWORK

Suppose there are altogether n LR images, the super resolution imaging model can be simply written as:

$$L_k = W_k H + N_k, k = 1 \ldots n \tag{1}$$

where L_k represents the kth observed LR images, H represents the original HR image, W_k is the transformation matrix including blur effect, motion effect and downsample effect, N_k denotes the noise image.

Noticing that there are different ways to classify the blocks and to select algorithm for each block category, the process of the framework is not fixed. A common sample of the framework flow chart is shown in Fig.1. In the registration step, $\{W_k\}$ are estimated, and then the whole image area is analyzed and divided into adaptive-sized blocks of different categories based on the content of the image and registration accuracy within the blocks. Then appropriate SR or single frame enhancement method is chosen and applied to each block. Finally before the HR image result is produced, a de-block process is introduced to reduce the block edge effect between neighboring blocks with different category labels.

2.1. Block analysis

There are reasons for dividing the target HR image into adaptive-sized blocks, rather than regions of arbitrary shapes, as most region-based SR algorithms do. First, a coarse-to-fine block analysis algorithm is computationally cheaper than a pixel-wise region segmentation algorithm for the same purpose. Second, our method is more efficient in removing spatial noise in the image, for small image areas or single pixels that are mis-labeled will be eliminated by the block in the proposed algorithm. Moreover, block-based super-resolution has natural facility for compressed video processing, for prevailing video compression algorithms are based on fixed- or adaptive-sized blocks.

First, the whole HR image area is divided into $p \times q$ square blocks of the same size as the initial block division set: $\mathcal{B} = \{b_i | i = 1 \ldots p \times q\}$. Then the features of each block b_i are analyzed to refine the block division set and to label the blocks.

A structure matrix is used to examine the smoothness around a pixel, where the structure matrix for a single pixel $I(x, y)$ in an image I is defined as:

$$S(x, y) = \nabla I(x, y) \cdot \nabla I(x, y)^T \tag{2}$$

Thus we can define the structure matrix of a block b_i:

$$S_i = \frac{1}{n_i} \sum_{(x,y) \in b_i} S(x, y) = \frac{1}{n_i} \sum_{(x,y) \in b_i} \nabla I(x, y) \cdot \nabla I(x, y)^T \tag{3}$$

where n_i represents the number of pixels in block b_i.

The eigenvalues $\lambda_1^{(i)}$ and $\lambda_2^{(i)}$ of the matrix S_i are a measurement of gradient strength in two perpendicular directions. The smoothness σ_i of a block b_i is defined as:

$$\sigma_i = \left|\lambda_1^{(i)}\right| + \left|\lambda_2^{(i)}\right| \tag{4}$$

The corresponding area of b_i in the kth LR image $b_i^{(k)}$ is defined to be *mismatched* when the correlation of the corresponding area is higher than a certain threshold l, that is, $b_i^{(k)}$ is *mismatched* only if

$$\| (W_k)_i b_i - b_i^{(k)} \| > l, k = 1 \ldots n \tag{5}$$

where $(\cdot)_i$ represents the transformation matrix limited to block b_i. Then the match measurement τ_i of a block b_i can be defined as:

$$\tau_i = \frac{1}{n} \left| \{b_i^{(k)} | b_i^{(k)} \text{ is mismatched}, \forall k\} \right| \tag{6}$$

where $|\mathcal{A}|$ denotes the number of elements in $\mathcal{A}$ when $\mathcal{A}$ is a set.

The blocks will be classified into three categories on the top level (see Fig.1): flat blocks, mismatched blocks and registered blocks. The algorithm below describes the details of the classification process.

Block Analysis Algorithm

Algorithm parameters: threshold σ_L, τ_L, τ_H, minimum block size m.

1. Initialize block division set $\mathcal{B}$ with square blocks of the same size;

2. Go to 6 if there are no unlabeled blocks in $\mathcal{B}$;

3. Pick an unlabeled block b_i from $\mathcal{B}$;

4. If $\sigma_i < \sigma_L$, label b_i as flat block; or if $\tau_i > \tau_H$, label b_i as mismatched block; or if $\tau_i < \tau_L$, or size of b_i is smaller than m, label b_i as registered block. If none of the conditions above are satisfied, eliminate b_i from $\mathcal{B}$, and break b_i into 2×2 square blocks of the same size, add all the 4 smaller blocks into $\mathcal{B}$;

5. Go to 2;

6. Output $\mathcal{B}$ with the labels of the blocks.

Flat blocks contain the least information in the image. Subpixel displacements provide no more information for flat region, except for noise statistics. Thus, only direct interpolation and denoising process is applied for those blocks to reduce the computational cost. The mismatched blocks contain meaningful information, but the information in the different input LR images does not seem to correspond to the same image data, so single frame image enhancement approach should be applied to increase the image quality. Only well-registered blocks deserve SR algorithms to exploit reliable additional information from input LR frames.

Moreover, different algorithms can be applied to different registered blocks based on their categories on the second level. [4] shows that an iterative method with relatively bad optical flow gets worse results than algorithms without iteration. On the other hand, with accurate enough optical flow results, iterative algorithms have advantages. Thus registered blocks can be thresholded by flow consistency into two second-level categories, and iterative and non-iterative approaches can be applied respectively.

2.2. De-block process

Visual discordance may appear at the edge area between different categories of blocks enhanced by different algorithms. A de-block process is necessary to obtain smooth transition across the edge after the synthesis step of our framework.

To accomplish this, the blocks along the edge area are dilated, resulting in an overlapping region along the edge. Suppose the width of the overlapped area is $2r_0$, and $I_1(x,y)$, $I_2(x,y)$ denote the result image of the two blocks in the overlapped area. Thus the final result image $I(x,y)$ in the latter block is obtained by:

$$I(x,y) = I_1(x,y)f(r) + I_2(x,y)f(-r) \qquad (7)$$

where r represents the distance between (x,y) and the edge, and $f(r)$ represents the combine function and satisfies: $f(-r_0) = 1$,$f(r_0) = 0$, and $f(r) = 1 - f(-r)$.

Our goal is to find the function $f^*(r)$ that gets the best combined result, which means that, $f^*(r)$ should not vary severely by r. So we get

$$f^*(r) = \arg\min \int_{-r_0}^{r_0} |f'(r)|^2 \, dr = \frac{r_0 - r}{2r_0} \qquad (8)$$

3. EXPERIMENTAL RESULTS

Fig. 2. Middle frame of the 3 fragments, respectively

Several experiments are carried out to verify the effectiveness of our algorithm. The experimental video "bicycler", which contains large scale motion, both global and local, is captured by a hand-held camera, with resolution of 640×480. The sequence is downsampled into 320×240 as LR images, while the original sequence is considered as target HR images for image quality measurement (such as PSNR). 3 fragments of the LR video, each of which contains 7 images, are picked as the input LR images, see Fig.2.

A 2-by-2 super-resolution process is applied on each of the fragments. σ_L, τ_L, τ_H are set 150.0, 0.7, 1.0 respectively, and initial block size is 24, $m = 6$. Algorithm [2] based on MAP is applied as SR algorithm in registered blocks, and single color image enhancement algorithm [8] is selected as single frame enhancement algorithm in mismatched area. We don't divide registered block into second-level block categories in the experiments for it's not needed in such SR tasks as without many blocks or a large canvas. The results of selected patch in each fragment are shown in Fig. 3, with comparative results of [2], [8], and confidence map algorithm [7].

Notice that the proposed algorithm gets apparently better results overall, especially in mis-matched area like the forest background in experiment 1 and 3. Also, flat blocks like on the hat (experiment 1) are directly interpolated in the proposed algorithm, which doesn't result in any reduction of the image quality. Table 1 shows the PSNR results, as well as structural similarity measurement (SSIM)[9] results.

The percentage of image as flat blocks in the experiments are 12.83%, 1.24%, and 10.70%, which illustrates the reduction of computational cost of our framework. The percentage of registered blocks are 63.34%, 95.17%, and 45.16%.

4. CONCLUSION

A practical framework for super-resolution image reconstruction is proposed in this paper. The algorithm divides the target

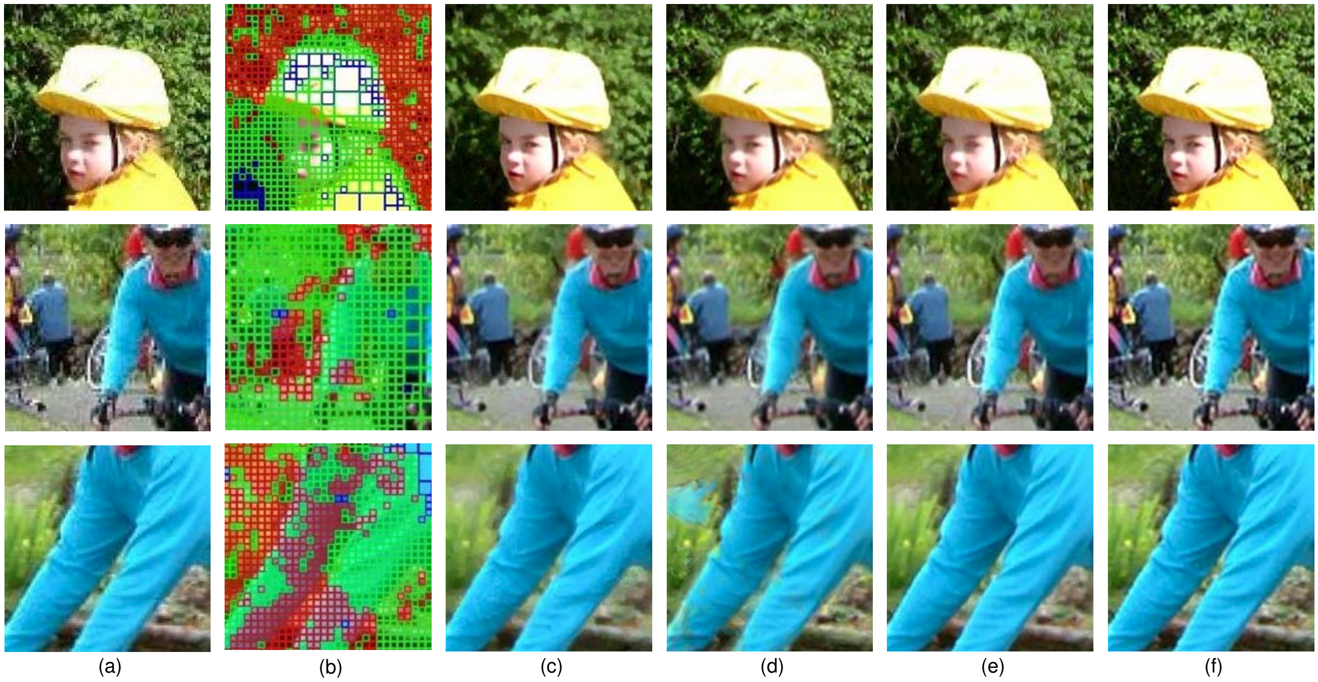

Fig. 3. Experimental results: Column (a) Selected LR patch as the red boxes show in Fig.2; (b) Block analysis result, blue blocks represents flat ones, red for mismatched ones, green for registered ones; (c) Results of [2]; (d) Results of [7]; (e) Results of [8]; (f) Results of proposed algorithm.

Table 1. PSNR(dB) and SSIM[9] Results

Exp.#	1		2		3	
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
MAP[2]	65.26	0.757	65.37	0.694	66.15	0.813
C-map[7]	65.07	0.841	65.86	0.811	64.97	0.853
Edge[8]	66.00	0.894	66.40	0.845	67.23	0.932
Proposed	66.45	0.920	67.05	0.887	67.62	0.953

image into different categories of blocks and applies different algorithms to them. In future work, efforts can be made in selecting different SR algorithms for registered blocks, such as MAP, IBP, POCS, etc. via machine learning approaches to get better results. Self-adaptive algorithm parameters can be used for blocks of the same category, too. We can also incorporate the framework with video compression and decompression algorithms to get video streams with better quality.

5. REFERENCES

[1] S.Park M.Park and M.Kang, "Super-resolution image reconstruction: A technical overview," *IEEE Signal Process. Mag.*, vol. 20, pp. 21–36, 2003.

[2] S.Farsiu M.D.Robinson M.Elad and P.Milanfar, "Fast and robust multiframe super-resolution," *IEEE Trans. on Image Processing*, vol. 13, pp. 1327–1344, 2004.

[3] C.Fan J.Zhu J.Gong and C.Kuang, "Pocs super-resolution sequence image reconstruction based on improvement approach of keren registration method," *Intelligent Systems Design and Applications*, pp. 333–337, 2006.

[4] W.Zhao and H.S.Sawhney, "Is super-resolution with optical flow feasible?," *European Conf. on Computer Vision*, 2002.

[5] Y.He and K.H.Yap et al, "A nonlinear least square technique for simultaneous image registration and super-resolution," *IEEE Trans. on Image Processing*, vol. 16, pp. 2830–2841, 2007.

[6] B.Huang H.Shen, L.Zhang and P.Li, "A map approach for joint motion estimation, segmentation, and super resolution," *IEEE Trans. on Image Processing*, vol. 16, pp. 479–490, 2007.

[7] M.Chen, "Dynamic content adaptive super-resolution," *International Conference on Image Analysis and Recognition*, 2007.

[8] Y.Cha and S.Kim, "Edge-forming methods for color image zooming," *IEEE Trans. on image processing*, vol. 15, pp. 2315–2323, 2006.

[9] Z.Wang and A.C.Bovik et al, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. on image processing*, vol. 13, pp. 600–612, 2004.