Signal Processing: *Image Communication* 29 (2014) 875–886

Contents lists available at ScienceDirect

Signal Processing: *Image Communication*

journal homepage: www.elsevier.com/locate/image

Video super-resolution based on automatic key-frame selection and feature-guided variational optical flow

Yanming Zhu [a], Kun Li [a,*], Jianmin Jiang [b]

[a] Tianjin Key Lab. of Cognitive Computing and Application, School of Computer Science and Technology, Tianjin University, Tianjin 300072, China

[b] Research Institute of Future Media Computing, Shenzhen University, Shenzhen 518060, China

ARTICLE INFO

Article history:
Received 5 June 2013
Received in revised form
13 June 2014
Accepted 15 June 2014
Available online 17 July 2014

Keywords:
Video super-resolution
Feature-guided variational optical flow
Key-frames
PSNR

ABSTRACT

This paper proposes a new video super-resolution method based on feature-guided variational optical flow. The key-frames are automatically selected and super-resolved using a method based on sparse regression. To overcome the blocking artifacts and deal with the case of small structures with large displacement, an efficient method based on feature-guided variational optical flow is used to super-resolve the non-key-frames. Experimental results show that the proposed method outperforms the existing benchmark in terms of both subjective visual quality and objective peak signal-to-noise ratio (PSNR). The average PSNR improvement from the bi-cubic interpolation is 7.15 dB for four datasets. Thanks to the flexibility of designed automatic key-frame selection and the validness of feature-guided variational optical flow, the proposed method is applicable to various practical video sequences.

1. Introduction

Video super-resolution is a hot topic in computer vision and video processing [1–4]. In the case of video super-resolution, almost all of the approaches are based on image super-resolution methods. For the last few decades, many image super-resolution methods have been developed to display high quality images and provided a remarkable progress.

With the development of image super-resolution methods, video super-resolution approaches are basically classified into three categories: (1) using the single image super-resolution approaches to super-resolve each video frame independently; (2) constructing the high-resolution frames using motion compensation techniques and the information extracted from multiple views of the same object along frames (usually known as reconstruction-based methods) [5–9]; and (3) generating the high-resolution frames using codebooks derived from key-frames in mixed-resolution-video (usually known as learning-based or example-based methods) [4,10,11].

The methods of super-resolving each frame independently can use the newest and most effective single image super-resolution approaches to get high quality results. But these methods do not take advantage of the great similarity and interrelationship between adjacent frames. In addition, they usually take a very long time when the video contains a large number of frames. Introducing motion compensation techniques to video super-resolution is more popular in the last few decades. It constructs the high-resolution frame by searching the best matching patches from adjacent key-frames. Recently, Song et al. [5] propose an improved algorithm to get higher quality results, which uses bi-directional overlapped block motion compensation and on-the-fly dictionary training. However, the essence of these methods is

* Corresponding author.
E-mail addresses: zhuym@tju.edu.cn (Y. Zhu), lik@tju.edu.cn (K. Li), jianmin.jiang@szu.edu.cn (J. Jiang).

http://dx.doi.org/10.1016/j.image.2014.06.005
0923-5965/

block matching, which may incur blocking artifacts and information missing. Moreover, constructing patch-pairs increases the complexity and the patch matching is not always accurate. To avoid the above-mentioned problems, a method using codebooks derived from key-frames is proposed by Hung et al. [4] and achieves high quality results. It combines low-resolution information with high-resolution information that is extracted from codebooks to generate high-resolution frame. However, this method usually requires to train multiple large codebooks, in the order of hundreds of thousands or even millions. Furthermore, the high-resolution results largely depend on the similarity between the input frame and the samples in the codebooks, and new noise may be introduced from the training set.

In this paper, we propose a new video super-resolution method to overcome the above-mentioned problems. The key-frames are automatically selected and super-resolved by a method based on sparse regression and natural image prior (SRNIP). Considering the case of small structures with large displacement, the non-key-frames are super-resolved by a method based on feature-guided variational optical flow (FVOF). The missing part is effectively recovered by an inpainting method. Experimental results prove that our method obtains better visual quality and higher peak signal-to-noise ratio (PSNR) than the existing methods.

The rest of this paper is organized as follows: Section 2 reviews related work on video super-resolution and optical flow. The proposed video super-resolution method is described in Section 3. Validation experiments and results are presented in Section 4. Finally, we conclude this paper in Section 5.

2. Related work

2.1. Video super-resolution

Video super-resolution based on image super-resolution approaches has been developed to display high quality scaled images in the last few decades [12–14]. In traditional interpolation methods, such as bilinear interpolation, bi-cubic [15], and cubic convolution [16], the information is local and local structures dictate how the missing information is filled. This kind of interpolation methods usually suffers from visual degradation, e.g., jagging and stair-case artifacts. Although many improved interpolation methods [17,18] appear lately to overcome this problem, in essence, interpolation methods rarely introduce any new high frequency information to increase the image resolution.

Many example-based or learning-based super-resolution methods have been developed to avoid this problem in the past few years. The example-based methods [10,19] construct the codebooks composed of a huge number of patch pairs, and recover the high-resolution image by finding the best matched patch pairs in the codebooks. Therefore, the computation complexity of these methods is very high due to the complex patch matching in a huge codebook. Furthermore, wrongly matched patches usually lead to incorrect reconstructions. Learning-based methods [20–22] are proposed assuming that the super-resolved image is a sparse representation of raw patches, which achieve significant improvements over other methods. In this model, each patch of the image to be super-resolved is represented by a linear combination of a very few dictionary elements. The method proposed by Yang et al. [21] is a typical method based on this model, which achieves significant improvements for both visual quality and PSNR. However, the running time of learning-based methods is long because of the time-consuming dictionary-training.

Recently, Kim et al. [23] generalize a super-resolution method using sparse regression and natural image prior. The underlying idea is to learn a map from input low-resolution images to target high-resolution images based on the pairs of input and output images. By using a sparse solution for kernel ridge regression and a prior natural image model, it reduces the computation complexity, and shows the effectiveness for image super-resolution compared with existing algorithms.

Block-based motion compensation is the most widely accepted approach used in standard video coding algorithms [24,25]. In case of video super-resolution, motion compensation based on block matching is also a popular method to generate a high-resolution frame by finding the best matching blocks from its adjacent frames [26–28]. However, this motion compensation method uses an implicit assumption that each block of pixels moves with uniform translational motion. Since this assumption is usually invalid, the method is well-known to produce blocking artifacts. Recently, Song et al. [5] have proposed a new video super-resolution algorithm using overlapped block motion compensation (OBMC) to reduce the blocking artifacts, and get better results. However, this method is also poor for de-blocking and the quality of the interpolated frame may be degraded because the overlapped block motion compensation is applied to all blocks uniformly.

Using codebooks derived from key-frames is another basic approach to super-resolve a video sequence [29,30]. It constructs high-resolution frame by using codebooks derived from key-frames in mixed-resolution video. Recently, Hung et al. [4] propose a method based on this kind of approaches. By using multiple overlapped variable-block-size codebooks instead of fixed-codebooks, this method gets better results than other codebooks-based super-resolution methods. However, the high-resolution results of this method largely depend on the similarity between the input frame and the patches in the codebooks, and may introduce new noise from the training set.

2.2. Optical flow

Optical flow estimation is one of the key problems in computer vision. It estimates the displacement field between two images and can be applied to motion estimation, 3D reconstruction, and image registration. Since the optical flow method can provide motion and structure information, it has two significant advantages: (1) its calculation accuracy is higher and can detect

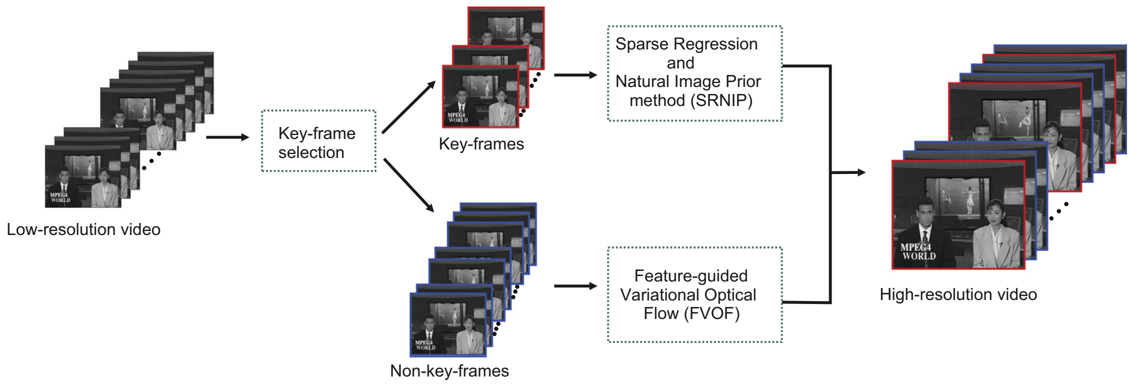

Fig. 1. The framework of the proposed method.

sub-pixel displacements and (2) it can be applied to relatively complex movement. In the last two decades, the quality of optical flow estimation methods has increased dramatically [31–33]. One of the predominant methods to estimate optical flow is proposed by Horn et al. [32] based on variational methods. It combines a gradient-based matching of pixel gray values with a global smoothness assumption. However, this method shows many limitations in practice. Later, a lot of modified and extended methods are proposed to get higher quality results. Memin et al. [34] employ non-quadratic penalizers in the smoothness term and the data term to estimate discontinuous and occlusive motion. Some methods use photometric invariant constraints to solve the violation of constant brightness assumption problem, such as Papenberg's method [35] using higher order derivatives and Zimmer's method [36] using color models with photometric invariant channels. But these methods cannot deal with large displacements and introduce global smoothness. Brox et al. [37] propose a method without linearizing the constancy constraints and design a model to deal with large displacements. By combining the non-linearized models with a continuation method, it leads to coarse-to-fine warping schemes. However, coarse-to-fine warping schemes have somehow relaxed the constraint mentioned above, and there is an inherent dependency between the scale of structures and the velocity that can be estimated. This particularly affects the estimation quality, since the displacements of the small structures are usually large in most images. Recently, Brox et al. [38] propose a large displacement optical flow to solve this problem by integrating rich descriptors into the variational optical flow framework. It estimates a dense optical flow field and achieves new domains of motion analysis. By using this method, it can estimate the motion between two images with high accuracy.

In this paper, we propose a new automatic video super-resolution method. Instead of using motion compensation or codebooks, we formulate the non-key-frame super-resolution problem as a warping system, in which a feature-guided variational optical flow method is used to ensure high accuracy of the warping process. In addition, a method using sparse regression and natural image prior is employed to super-resolve the key-frames. Our method can be straightforwardly applied to the video in which all frames are low-resolution.

3. The proposed method

In this section, we present the details of our method as depicted in Fig. 1. For an input low-resolution video, the key-frames (frames with the red border in the figure) are first automatically selected and super-resolved with SRNIP (Section 3.1), and then the non-key-frames (frames with the blue border in the figure) are super-resolved based on feature-guided variational optical flow (FVOF) (Section 3.2).

3.1. The key-frame selection and super-resolution

The key-frames are automatically selected according to the motion errors with respect to the previous nearest key-frame. For a low-resolution video, we set the first frame as the initial key-frame, and super-resolve it using a method based on sparse regression and natural image prior (SRNIP) [23]. Then, the motion error is calculated for each low-resolution frame. If the motion error is larger than a threshold T, this frame is selected as a new key-frame and super-resolved by the SRNIP method. Otherwise, it will be treated as a non-key frame and super-resolved by the FVOF method (Section 3.2).

To calculate the motion error, we first up-scale the current frame by a bicubic method [15] and subtract this frame from the previous nearest key-frame. Then, we get a difference matrix which represents the difference between the two frames. The average of the non-zero entries in the difference matrix is calculated as the motion error, which is formulated as

$$e = \frac{f_{NZ}(P(y) - P(y_{pre}))}{k}, \tag{1}$$

where y and y_{pre} are the current frame and the previous nearest key-frame, respectively. $P(\cdot)$ is an operator that converts an RGB image into a gray image. f_{NZ} is a function which selects non-zero entries from the matrix, and k is the number of non-zero entries in the difference matrix.

We use an iterative method to select the threshold T automatically. First, we define a constant C, which represents the lower bound of the threshold T. The ratio of non-key-frames in the video sequence is represented by constant R_0. The mean value $\overline{e}$ of motion errors between adjacent frames can be calculated by

$$\bar{e} = \left(\sum_{i=1}^{N-1} e_{i,i+1} \right) \Big/ (N-1), \tag{2}$$

where $e_{i,i+1}$ is the motion error between frame i and frame $i+1$, and N is the number of frames in the video. The iterative process is run in the following manner. If the mean error $\bar{e} < C$, we set $T = \bar{e}$ as an initial value of the threshold T and execute the key-frames selection process based on this threshold. Then, we examine the number of key-frames n and calculate the ratio r of the non-key-frames by $r = (N-n)/N$. If r is smaller (or larger) than the initial setting value R_0, we set the threshold $T = T+ (or-)0.1$, and repeat the above process until $r \geq$ (or $\leq$)R_0. Finally, we get the ultimate threshold T subject to $r \approx R_0$ by this iterative process. Otherwise, if $\bar{e} \geq C$, we set the ultimate threshold $T = \bar{e}$. The algorithm for the automatic threshold selection is detailed in Algorithm 1. The initial definition of the parameter C uses a statistical approach. We set C to be 3.0 based on large amounts of experimental data analysis and statistics. The definition of parameter R_0 is dependent on the request of super-resolution quality and time cost. Obviously, when R_0 becomes smaller, the video super-resolution quality becomes higher and the running time becomes longer.

Algorithm 1. Algorithm for automatic threshold selection.

```
Require: lower bound C, the ratio of non-key-frames R0,
   and video mean motion error ē.
  if ē < = C then
    T = ē.
    Count the number of key – frames n by executing key–
 frame selection process based on threshold T.
    Calculate the ratio of non – key – frames r by r = (N – n)/N.
    if r < R0 then
      while r < R0 do
        T = T+0.1
        Calculate a new r based on the new threshold T.
      end while
    else if r > R0 then
      while r > R0 do
        T = T–0.1
        Calculate a new r based on the new threshold R.
      end while
      Output the ultimate threshold T.
    end if
  else if ē > C then
    T = ē.
    Output the ultimate threshold T.
  end if
```

The super-resolution of the selected key-frame is in essence a single image super-resolution problem. As described in Section 2.1, many approaches have been developed to generate high quality scaled images in recent years. The image super-resolution that is applied to the video super-resolution, need have two main advantages: (1) low computation complexity and short running time and (2) easy to accomplish and can generate high quality result. Therefore, example-based super-resolution methods are the most appropriate. In this paper, we adopt SRNIP [23] to super-resolve the key-frames. The reasons are that (1) it uses sparse regression to reduce the computation complexity which satisfies the time requirement of the video super-resolution and (2) it uses a natural image prior model to deal with the discontinuity of images which satisfies the quality requirement of video super-resolution. The simpleness and flexibility make it suitable for various applications.

3.2. The non-key-frame super-resolution

For the input low-resolution non-key-frame, we first interpolate it into the desired scale using the bi-cubic interpolation method [15]. Assume that K and I are the key-frame and the non-key-frame to be aligned, respectively, and $p = (x, y)^T$ denotes a pixel in the frame. Then, the non-linear optical flow model can be presented as

$$\begin{aligned} E(O) = & \int \Phi(|I(p+O(p)) - K(p)|^2)\, dx \\ & + \int \Phi(|\nabla I(p+O(p)) - \nabla K(p)|^2)\, dx \\ & + \int \Phi(|\nabla u(p)|^2 + |\nabla v(p)|^2)\, dx, \end{aligned} \tag{3}$$

where O represents the optical flow. $\nabla K(p) = (\partial K/\partial x, \partial K/\partial y)$ and $\nabla I(p+O(p)) = (\partial I/\partial(x+O(x)), \partial I/\partial(y+O(y)))$ are the gradients of the key-frame and the non-key-frame at pixel p, respectively. $\nabla u(p)$ and $\nabla v(p)$ are the gradients of the horizontal and perpendicular components of optical flow at pixel p, respectively. The function $\Phi(t^2) = \sqrt{t^2+\varepsilon^2}, \varepsilon = 0.001$, which yields a TV regularization. This regularization corresponds to an ℓ_1 norm minimization, but is still differentiable everywhere. As we can see in the equation, in contrast to the predominant linear model, there is no linearization here. Combining this non-linear model with a continuation method leads to a coarse-to-fine warping scheme, which works well in almost all the large displacement case [37]. Considering the case of small structure with large displacement, a feature-guided variational model is adopted, which fully exploits the temporal correlations between frames and can be expressed as [38]

$$\begin{aligned} E'(O) = & E(O) \\ & + \int \theta(p)\omega(p)\Phi(|O(p) - O_1(p)|^2)\, dx \\ & + \int \theta(p)|F_{non}(p+O_1(p)) - F_{key}(p)|^2\, dx, \end{aligned} \tag{4}$$

where $O_1(p)$ denotes the correspondence vectors obtained by feature matching at pixel p. $\theta(p)$ denotes whether there is a feature available in the key-frame K at point p, whose value is 0 or 1. $\omega(p)$ is the weight function of correspondence optical flow. $F_{key}(p)$ and $F_{non}(p)$ are the feature vectors in the key-frame K and the non-key-frame I, respectively. By integrating the correspondences from feature matching into the variational optical flow model, the solution is guided towards large displacements of small structures.

The operation of warping the previous nearest high-resolution key-frame with the optical flow O to generate high-resolution non-key-frame is formulated as

$$x_{kl} = y_{ij}, (i \in (1, m), j \in (1, n)), \tag{5}$$

Fig. 2. Results of continuous non-key-frame super-resolution for *Container*, *Foreman* and *Big Buck Bunny*: (a) high-resolution key-frames, (b) low-resolution non-key-frames, (c) images warped by the feature-guided variational optical flow, (d) images corrected by inpainting, and (e) original high-resolution non-key-frames.

with

$$k = i + O^1_{ij}$$
$$l = j + O^2_{ij}, \tag{6}$$

where y is the previous nearest high-resolution frame, m and n are the height and width of y, x is the high-resolution version of the current non-key frame, O is the optical flow, and O^1, O^2 are the first and the second dimension of the optical flow O, respectively. We ignore the pixel values of x, when its subscripts k and l are beyond the image boundary.

Although we adopt feature-guided variational optical flow to enhance the quality of warping scheme, some information may be lost in the generated high-resolution frame. We formulate the information repairing problem into an image inpainting problem. Assume that D is a region needed to be repaired in a frame L, and $L \backslash D$ is the region without information missing. For a point $d \in D$, the inpainting of d is determined by the values of the known points close to d. We consider a first order approximation $L_f(d)$ of the frame in point d as

$$L_f(d) = L(f) + \nabla L(f)(d-f), \tag{7}$$

where $f \in L \backslash D$ is a point close to d, $L(f)$ is the pixel value of f and $\nabla L(f)$ is the gradient of point f. We estimate the pixel value of d as a weighted average over a small neighborhood Ω of d, which can be formulated as [39]

$$L(d) = \frac{\sum_{f \in \Omega} W(d,f)[L(f) + \nabla L(f)(d-f)]}{\sum_{f \in \Omega} W(d,f)}, \tag{8}$$

where $W(d,f)$ is the normalized weighting function, which is used to define the contribution of pixel f in the neighborhood. The weighting function is defined as

$$W(d,f) = P(d,f) \cdot Q(d,f) \cdot R(d,f), \tag{9}$$

with

$$P(d,f) = \frac{d-f}{\| d-f \|} \cdot N(d)$$
$$Q(d,f) = \frac{d_0^2}{\| d-f \|^2}$$
$$R(d,f) = \frac{S_0}{1 + |S(d) - S(f)|}, \tag{10}$$

where $S(d)$ is the level set and $N(d) = \nabla S(d)$ is the normal direction of d. d_0 and S_0 are the distance parameter and the level-set parameter, respectively, both of which are generally set to be 1. Directional factor $P(d,f)$ ensures that the contribution of the pixels close to normal direction N is higher than those farther from N. Distance factor $Q(d,f)$ guarantees that the contribution of the pixels closer to d is higher. Level set distance factor $R(d,f)$ ensures that the contribution of pixels close to the contour through d is higher.

Fig. 2 shows the intermediate and final results of continuous non-key-frame super-resolution for *Container*, *Foreman* and *Big Buck Bunny*. We only give results of a few non-key-frames nearest to the key-frame as a schematic. The top four rows are experimental results for *Container* video and the ratio r is set to be 0.1; the middle three rows are experimental results for *Foreman* video and the ratio r is set to be 0.2; the bottom two rows are experimental results for *Big Buck Bunny* video and the ratio r is set to be 0.3. It can be seen that the proposed method can accurately predict the non-key-frame with respect to the key-frame, even if for *Foreman* and *Big Buck Bunny* with large displacements.

4. Experimental results

In this section, we evaluate the performance of the proposed video super-resolution method with publicly available datasets (in Section 4.1) and datasets acquired from internet or captured by the author (in Section 4.2).

4.1. Evaluation on publicly available datasets

For evaluating the performance and comparing with other methods, we use four publicly available YUV video sequences: two CIF sequences (*Container* and *News*) and two 1280 $\times$ 720 sequences (*Mobcal* and *Shields*). In the experiment, we first down-sample the original high-resolution non-key-frame by the factor of 2 and then super-resolve the low-resolution version using the proposed method. In order to compare with the other method, we also assume that the key-frame is originally high-resolution and set the ratio r to be 0.5.

We compare the performance of the proposed method with five existing methods: the bi-cubic method [15], Fan's method [40], Brandi's method [41], Hung's method [4] and three versions of the method in [5]: learning-based super-resolution (LSR) version, motion-compensated super-resolution (MSR) version and hybrid super-resolution (HSR) version. In the LSR method, the cluster number is set to be 512. In the MSR method, the motion search range is set to be 64, and the matching block for motion vector is set to be 16×16. In Fan's method [40], the number of nearest neighbor is set to be 5 and the patch size is set to be 7×7. For Brandi's method [41], the matching block size is set to be 16×16, which provides the best performance among all the sizes. The reference frames are interpolated by bi-cubic method. In Hung's method [4], the number of patch-pairs is set to be 1000, and the block size is set to be 2×2. Table 1 shows the PSNR improvement of the 15th frame from the bi-cubic interpolation, where the values in red are the highest of all methods for each sequence and the values in blue indicate the second highest. As shown in Table 1, our method has obviously better performance than all the methods for the sequences *Container* and *News*, and the PSNR value is even 11.2 dB higher than the bi-cubic method. For the sequences *Mobcal* and *Shields*, the

Table 1
Quantitative evaluation of different algorithms: improvement of PSNR values from the bi-cubic interpolation.

Sequence	Fan [40]	Brandi [41]	MSR [5]	LSR [5]	HSR [5]	Hung [4]	Our method
Container	−0.1	0.9	4.0	2.7	5.3	8.1	11.2
News	1.0	1.0	2.5	4.5	6.7	9.4	9.6
Mobcal	−0.8	0.1	3.2	1.1	3.3	7.3	5.9
Shields	−1.1	−0.2	0.3	1.2	1.6	4.9	1.9

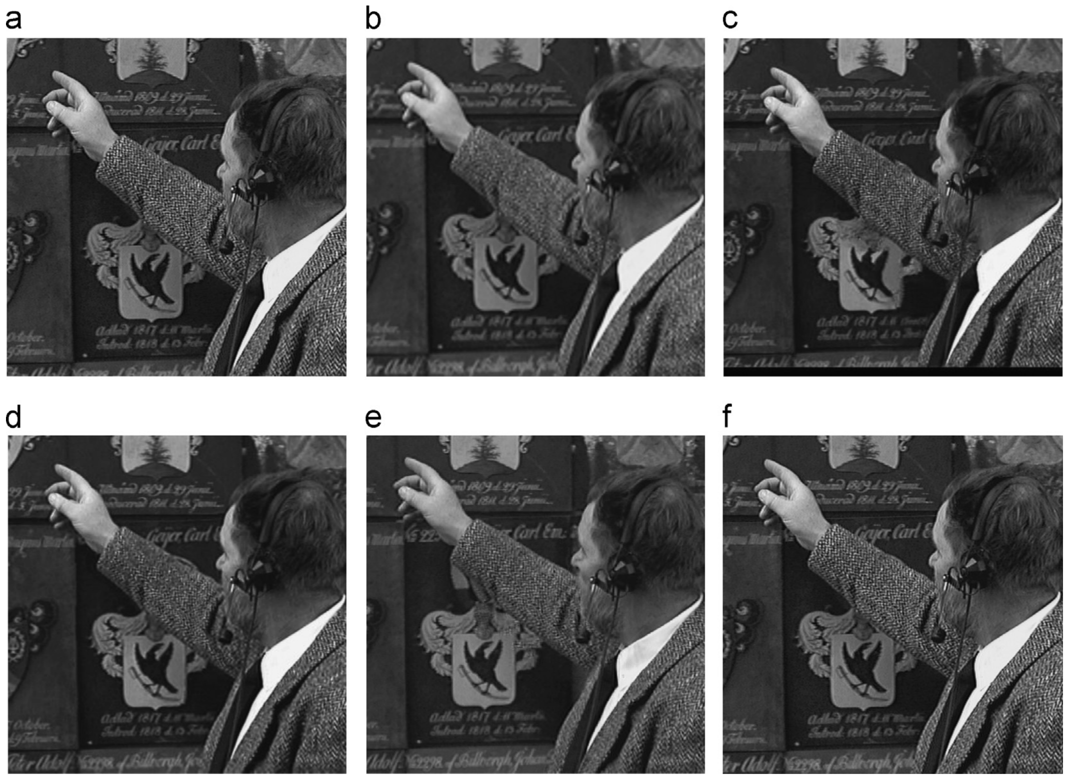

Fig. 3. Super-resolution results for a center region of *Shields*: (a) Ground truth, (b) bi-cubic, (c) MSR in [5], (d) HSR in [5], (e) Hung's method [4], and (f) our method.

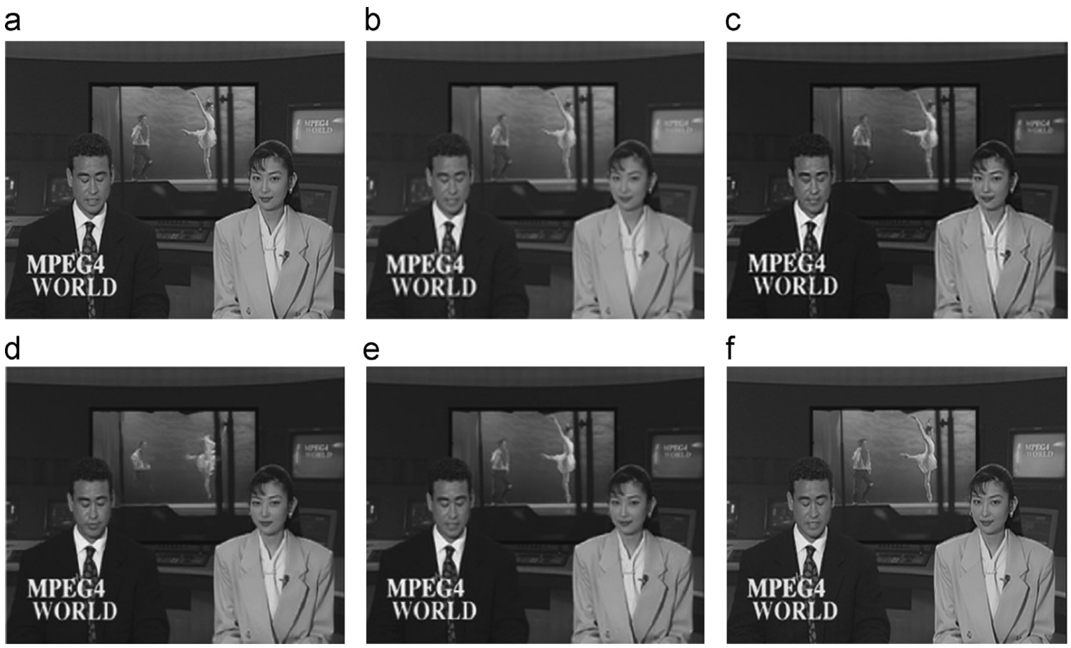

Fig. 4. Super-resolution results for *News*: (a) Ground truth, (b) bi-cubic, (c) Brandi's method in [41], (d) MSR in [5], (e) HSR in [5], (f) Our Method.

performance of our method is better than other methods except Hung's method. Although the PSNR value is lower than Hung's result, the visual quality of our results is significantly better than all the other's results.

Fig. 3 gives the qualitative evaluation of the 15th frame for the *Shields* sequence. We compare our method with the bi-cubic method, the MSR method, the HSR method and Hung's method. As shown in Fig. 3, the result of bi-cubic

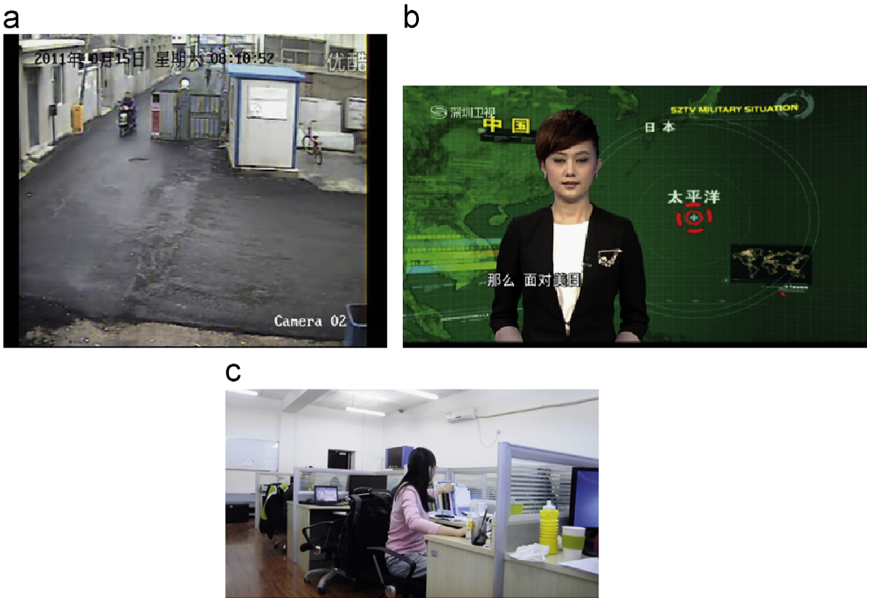

Fig. 5. Three video sequences. (a) *Surveillance*, (b) *TV*, (c) *Lab*.

method is very blurred in the whole image, even though its PSNR value is not low. Obvious errors such as the logo in the middle of the image and the information missing such as the bottom of the image can be seen in the result of MSR method. The HSR method performs slightly better than MSR method, but the blocking artifacts are still apparent, and the edges and cloth textures are also blurred. The result of Hung's method also has error regions, such as the logo in the middle of the image. On the contrary, our method provides the best visual quality with sharper edges and richer textures.

For more comprehensive comparison of the performances of our method and other methods, we provide the super-resolution results of the 15th frame for the *News* sequence with large displacement in Fig. 4. We compare our method with the bi-cubic method, Brandi's method, the MSR method and the HSR method. As shown in the figure, the results of bi-cubic and Brandi's method are very blurred. In the result of the MSR method, the information of dancers in the background is wrong and missing, and the hair details of the female anchor's forehead are also missing. The result of the HSR method has the same problem in the hair area, and the words on the background LED screen are blurred. Our method overcomes all the above-mentioned problems and gets the significantly better visual quality than all the other methods.

4.2. Evaluation on more datasets

To prove a good practical application, we super-resolve three real video sequences: *Surveillance* (352×288, 120 frames), *TV* (426×240, 360 frames) and *Lab* (640×360, 240 frames). *Surveillance* and *TV* are two popular video sequences on the internet, and the *Lab* sequence is captured using a Nikon D90 camera, as shown in Fig. 5.

For the *Surveillance* sequence, we super-resolve it to verify the performance of our method when the proportion of non-key-frames is large. Therefore, the threshold T is set to be 3.0 and the ratio r is set to be 0.7. Fig. 6 gives the $3\times$ super-resolution results of the 1st, 11th, and 21st frames in *Surveillance* sequence. These frames are super-resolved by the proposed FVOF method, and selected to show randomly. In order to demonstrate the performance of our method more comprehensively, we select different amplification regions for different frames. The regions highlighted by rectangles in Fig. 6 are enlarged and shown in the associated images for closer observation. For the low-resolution input frame, it is stretched to the same size as the high-resolution frame. As shown in the figure, for all the regions, our method provides obvious good subjective visual quality with sharp edges and rich textures.

In order to demonstrate the performance of our method in displaying high quality scaled images on cutting-edge digital consumer application such as high-definition television, we super-resolve the *TV* sequence by factor 3 with threshold T equal to 2.8 and ratio r equal to 0.5. To display the continuous super-resolution performance of our method, the $3\times$ super-resolution results of three continuous frames are shown in Fig. 7. For different frames, we select different amplification regions to comprehensively show the super-resolution quality of our method. By inspecting the regions highlighted with rectangles, it can be seen that our method provides obviously sharp edges and rich details.

Another video sequence *Lab* is captured by the author. We set threshold T to be 3.8 and ratio r to be 0.5. Fig. 8 gives the super-resolution results of the 63rd, 65th and 67th frames selected randomly from the result set. The regions highlighted by rectangles are enlarged and shown in the associated images for closer observation. In order to show the super-resolution quality of our

Fig. 6. 3 × Super-resolution results of *Surveillance*. Top row: the 1st frame result. Middle row: the 11th frame result. Bottom row: the 21st frame result.

method comprehensively, we select different magnification regions for different frames. As shown in the figure, our method provides obviously good subjective visual quality with sharp edges and rich textures.

5. Conclusion

This paper proposes a new video super-resolution method to magnify a low-resolution video. The key-frames

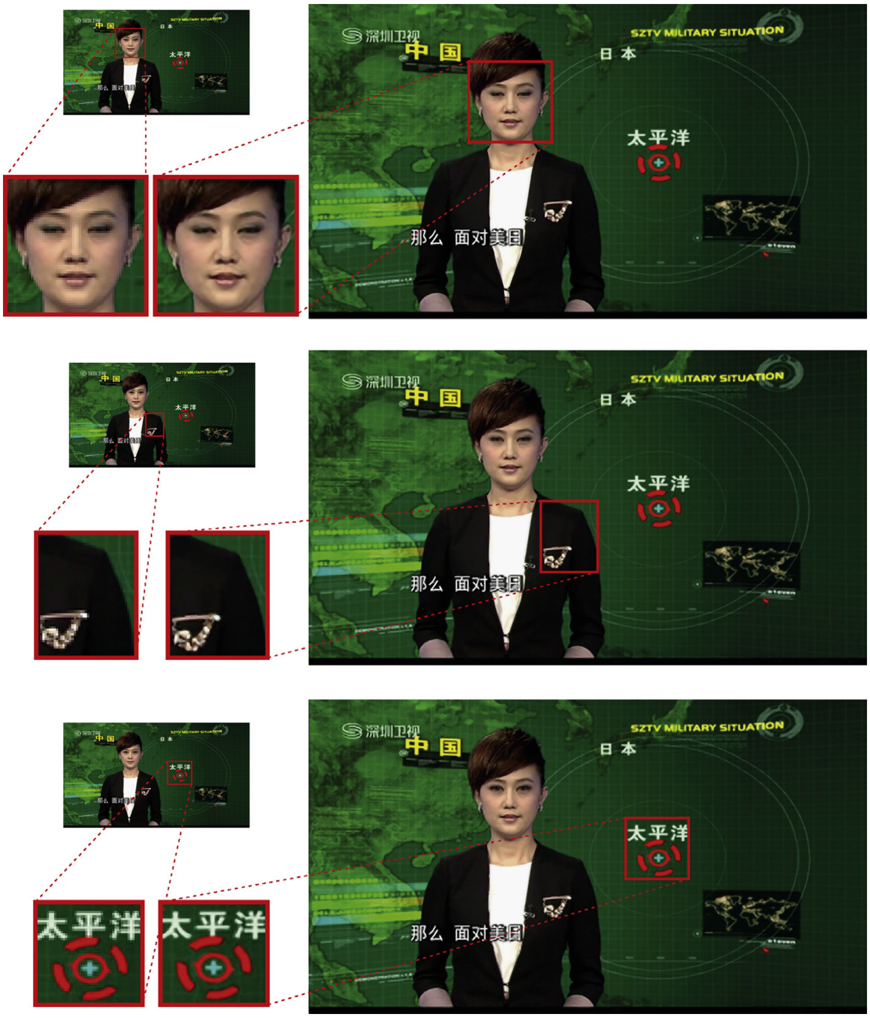

Fig. 7. Super-resolution results of *TV* sequence by magnifying 3 ×. Top row: the 2nd frame result. Middle row: the 3rd frame result. Bottom row: the 4th frame result.

Fig. 8. Super-resolution results of *Lab*. From left to right: the 63rd frame result, the 65th frame result and the 67th frame result.

are automatically selected based on motion errors and super-resolved by a method based on SRNIP. The non-key-frames are super-resolved by a method based on FVOF. By using an accurate warping method based on feature-guided variational optical flow, the proposed approach overcomes blocking artifacts and achieves promising results even for small structures with large displacement. Experimental results prove that the proposed method provides significantly better visual quality as well as higher PSNR value in comparison with the state-of-the-art methods. Moreover, the proposed method is applicable to various practical video sequences due to the flexibility of the designed automatic key-frame selection and the validness of the feature-guided variational optical flow. The applications include but not limited to the following areas:

(1) *Surveillance video*. Most surveillance cameras are low-resolution due to the cost, while our method can provide high-resolution videos for this application.
(2) *Video compression*. Video compression usually involves quantization and sub-sampling to control the balance between image quality and bit rate. The sub-sampled frames have to be up-sampled to achieve full resolution. Our method can be used to increase the reconstructed image quality.
(3) *Remote sensing*. Currently, the resolution of the remote sensing video is low, and our method can achieve the high-resolution remote sensing video. This is important for the object recognition.

Although feature-guided variational optical flow is robust for most scenes, it is difficult to handle situations with fast scene changes or lots of noises. As a future work, we will explore the pre-processing such as sparse feature tracking or denoising before super-resolution to improve the performance of this case.

Acknowledgments

This work was supported by the NSF of China (61302059, 61373103), Ph.D. Programs Foundation of Ministry of Education of China (20120032120040), and the Tianjin Research Program of Application Foundation and Advanced Technology (12JCYBJC10300, 13JCQNJC03900).

References

[1] P. Kramer, O. Hada, J. Benois-Pineau, J.-P. Domenger, Super-resolution mosaicing from MPEG compressed video, Signal Process. Image Commun. 22 (10) (2007) 845–865.
[2] S. Knorr, M. Kunter, T. Sikora, Stereoscopic 3D from 2D video with super-resolution capability, Signal Process. Image Commun. 23 (19) (2008) 665–676.
[3] Y. Liang, P.C. Yuen, J.-H. Lai, Image super-resolution by textural context constrained visual vocabulary, Signal Process. Image Commun. 27 (10) (2012) 1096–1108.
[4] E.M. Hung, R.L. de Queiroz, F. Brandi, Video super-resolution using codebooks derived from key-frames, IEEE Trans. Circuits Syst. Video Technol. 22 (9) (2012) 1321–1331.
[5] B.C. Song, S.-C. Jeong, Y. Choi, Video super-resolution algorithm using bi-directional overlapped block motion compensation and on-the-fly dictionary training, IEEE Trans. Circuits Syst. Video Technol. 21 (3) (2011) 274–285.
[6] J. Zhang, M.O. Ahmad, M.N.S. Swamy, Overlapped variable size block motion compensation, in: Proceedings of the IEEE International Conference on Image Processing, vol. 3, 1997, pp. 164–167.
[7] E.M. Hung, R.L. de Queiroz, Blocking-effect reduction in a reversed-complexity video codec based on a mixed-quality framework, in: Proceedings of the IEEE International Symposium on Telecommunications, 2010, pp. 1–6.
[8] S. Borman, R.L. Stevenson, Simultaneous multi-frame MAP super-resolution video enhancement using spatio-temporal priors, in: Proceedings of the International Conference on Image Processing, vol. 3, IEEE, 1999, pp. 469–473.
[9] S.P. Belekos, N.P. Galatsanos, A.K. Katsaggelos, Maximum a posteriori video super-resolution using a new multichannel image prior, IEEE Trans. Image Process. 19 (6) (2010) 1451–1464.
[10] W.T. Freeman, T.R. Jones, E.C. Pasztor, Example-based super-resolution, IEEE Trans. Comput. Graphics Appl. 22 (2) (2002) 56–65.
[11] F. Brandi, R. de Queiroz, D. Mukherjee, Super-resolution of video using key frames, in: IEEE International Symposium on Circuits and Systems, Seattle, WA, 2008, pp. 1608–1611.
[12] S.C. Park, M.K. Park, M.G. Kang, Super-resolution image reconstruction: a technical overview, IEEE Signal Process. Mag. 20 (3) (2003) 21–36.
[13] A.K. Katsaggelos, R. Molina, J. Mateos, Super-resolution of image and video, Synthesis Lectures on Image, Video, and Multimedia Processing, San Mateo, CA, 2007.
[14] S. Farsiu, M.D. Robinson, M. Elad, P. Milanfar, Fast and robust multiframe super-resolution, IEEE Trans. Image Process. 13 (10) (2004) 132–1344.
[15] H.S. How, H.C. Andrews, Cubic spline for image interpolation and digital filtering, IEEE Trans. Acoust. Speech Signal Proc. 26 (5) (1978) 508–517.
[16] R.G. Keys, Cubic convolution interpolation for digital image processing, IEEE Trans. Acoust. Speech Signal Proc. 29 (6) (1981) 1153–1160.
[17] X. Li, M.T. Orchard, New edge-directed interpolation, IEEE Trans. Image Process. 10 (10) (2001) 1521–1527.
[18] Q. Wang, R.K. Ward, A new orientation-adaptive interpolation method, IEEE Trans. Image Process. 16 (4) (2007) 889–900.
[19] X. Li, K.M. Lam, G. Qiu, L. Shen, S. Wang, Example-based image super-resolution with class-specific predictors, J. Vis. Commun. Image Represent. 20 (2009) 312–322.
[20] R. Zeyde, M. Elad, M. Protter, On single image scale-up using sparse-representations, Curves and Surfaces Lecture Notes in Computer Science, 2010, pp. 711–730.
[21] J. Yang, J. Wright, T. Huang, Y. Ma, Image super-resolution via sparse representation, IEEE Trans. Image Process. 19 (11) (2010) 2861–2873.
[22] G. Mu, X. Gao, K. Zhang, X. Li, D. Tao, Single image super-resolution with high resolution dictionary, in: Proceedings of the IEEE International Conference on Image Processing, 2011, pp. 1141–1144.
[23] K.I. Kim, Y. Kwon, Single-image super-resolution using sparse regression and natural image prior, IEEE Trans. Pattern Anal. Mach. Intell. 32 (6) (2010) 1127–1133.
[24] ISO, Information Technology-Coding of Moving Pictures and Associated Audio for Digital Storage Media at up to About 1.5 mbit/s, Technical Report, Motion Pictures Experts Group, 1993.
[25] I.T.T.C. Community, Video Codec for Audiovisual Services at px64 kb/s, Technical Report, ITU-T Recommendation H.261, 1993.
[26] J.Y. Tham, S. Ranganath, M. Ranganath, A.A. Kassim, A novel unrestricted center-biased diamond search algorithm for block motion estimation, IEEE Trans. Circuits Syst. Video Technol. 12 (10) (2002) 934–947.
[27] F. Brandi, R. de Queiroz, D. Mukherjee, Super-resolution of video using key frames and motion estimation, in: Proceedings of the IEEE Internatinal Conference on Image Processing, San Diego, CA, 2008, pp. 321–324.
[28] B.D. Choi, J.W. Han, C.S. Kim, S.J. Ko, Motion-compensated frame interpolation using bilateral motion estimation and adaptive overlapped block motion compensation, IEEE Trans. Circuits Syst. Video Technol. 17 (4) (2007) 407–416.
[29] K.F. de Oliveira, F. Brandi, E.M. Hung, R.L. de Queiroz, D. Mukherjee, Bipredictive video super-resolution using key-frames, in: Proceedings of the SPIE Symposium on Electronic Imaging, Visual Information Processing and Communication, 2010, pp. 1–5.
[30] H. Kawano, N. Suetake, B. Cha, T. Aso, Super-resolution via matching from self-decomposed codebook with local distance measure incorporating pixel correlation, in: Proceedings of the Biennial Conference of the Australian Pattern Recognition Society on Digital Image Computing Techniques and Applications, 2007, pp. 317–323.
[31] L. Alvarez, J. Esclarin, M. Lefebure, J. Sanchez., A PDE model for computing the optical flow, in: Proceedings of the XVI Congreso de Ecuaciones Diferenciales y Aplicaciones, 1999, pp. 1349–1356.
[32] B. Horn, B. Schunck, Determining optical flow, Artif. Intell. 7 (1981) 185–203.
[33] R. Deriche, P. Kornprobst, G. Aubert, Optical-flow estimation while preserving its discontinuities: a variational approach, in: Recent Developments in Computer Vision, 1996, pp. 69–80.
[34] E. Memin, P. Perez, Dense estimation and object-based segmentation of the optical flow with robust techniques, IEEE Trans. Image Process. 7 (5) (1998) 703–719.

[35] N. Papenberg, A. Bruhn, T. Brox, S. Didas, J. Weickert, Highly accurate optic flow computation with theoretically justified warping, Int. J. Comput. Vis. 67 (2) (2006) 141–158.
[36] H. Zimmer, A. Bruhn, J. Weickert, L. Valgaerts, A. Salgado, B. Rosenhahn, H.P. Seidel, Complementary optic flow, in: Energy Minimization Methods in Computer Vision and Pattern Recognition, 2009, pp. 207–220.
[37] T. Brox, A. Bruhn, N. Papenberg, J. Weickert, High accuracy optical flow estimation based on a theory for warping, in: Proceedings of the European Conference on Computer Vision, 2004, pp. 25–36.
[38] T. Brox, J. Malik, Large displacement optical flow: descriptor matching in variational motion estimation, IEEE Trans. Pattern Anal. Mach. Intell. 33 (3) (2011) 500–513.
[39] A. Telea, An image inpainting technique based on the fast marching method, J. Graphics Tools 9 (1) (2004) 23–34.
[40] W. Fan, D.Y. Yeung, Image hallucination using neighbor embedding over visual primitive manifolds, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2007, pp. 1–7.
[41] F. Brandi, R. Queiroz, D. Mukherjee, Super-resolution of video using key frames and motion estimation, in: Proceedings of the IEEE International Conference on Image Processing, 2008, pp. 321–324.