MAP with Total Variation Minimum and Sparse Representation Video Super-Resolution Algorithm and Smoothing in Time Domain

Weijie Gan
South China University of Technology
Guangzhou, Guangdong, China
Kevin.WJ.Gan@outlook.com

Abstract

Video super-resolution means to enhance solution of video by software technology. Because it is hard to do the same thing by hardware, it has great applications and demands. In this paper, the research of video super-resolution reconstruction mainly revolves around multi-frame-based method. We mainly implement maximum posterior probability(MAP) with prior knowledge: piecewise smooth, in each frame of video. This prior knowledge is combined according to two expression: total variation minimum and sparse representation. What's more, a time smoothing method is also introduced, with the aim to improve quality of whole video.

1. Introduction

Video super-resolution means enhance of resolution fo video by computer software technology. Why we need to research video super-resolution is because that, firstly, there is great demand on this technology in common life, industry and military. And it is hard for us to increase resolution by hardware, since it affects quality and sample speed of video and causes pressure in price[7].

For a given image, the process to get a new image have same content and higher resolution is called image super-resolution. Video super-resolution can be seen as a kind of extension of image super-resolution. The difference between super-resolution in video and in image is that[10]: (1)Super-resolution in each frame of video, can be done with near frames, which means more information can be used in this process, compared with singe image super-resolution. (2)Video super-resolution also should take care of construction of details between frames.

Main methods for video super-resolution can be classified into three major categories: interpolation-based method, multi-frames method and learning-based method[9].

Interpolation-based method is the most common and easiest way to enhance resolution of video. It implements a simple linear function to each pixel. Information from each pixel and its near pixel are used to reconstruct pixel in unknown location of image/video. The classical example is bilinear and bicubic method. The benefit of this method is easy to fulfill. However, since information it use is limited(only pixel itself and its near pixels), it can not enhance solution perfectly, especially details of image/video.

Shortcoming of interpolation-based method leads to the development of multi-frame method. Multi-frame method enhance single image or frame of video by many images or frames with very similar content. These extra images are usually obtained by multi-sample of same object or near frame from a video. It suppose a same down-quality model for all images used in this method and their similar and complementary in same location of each image[5]. Some reasons limit effect of multi-frame method, like difficulty of estimation of exact down-quality model, and inter-frame blurring in video super-resolution reconstruction.

To overcome these weak point, the learning-based method, which uses information from thousands of images in different content, is introduced. It suppose similarity of different images in high-frequency domain and reconstruct a single super-resolution image with the help of many low-high resolution training images[5]. Convolutional Neutral Network[2] and sparse transform dictionary[13] are common models to store information from these thousands of images.

Despite that, learning-based model needs huge of time to train certain model, like a very deep neutral network. And it perform better in certain field, like human face reconstruction. What's more, it can not combine prior knowledge, which is useful in many field.

Pointing to defect of multi-frame method, recent research points out that good and suitable prior knowledge can compensate inaccuracy of down-quality model[14] and a smoothing method in time domain can relieve blurring effect in video reconstruction[1]. So, multi-frame method is a more feasible choice for video super-resolution.

The main work of this paper include: (1)A multi-frame method with max posterior probability, total variation minimum restriction and sparse representation restriction is introduced. (2)A total variation minimum in time domain is used to relieve blurring of reconstructed video.

In section (2), we describe down-quality model and details of prior knowledge used in our model. Section (3) introduces concrete steps solution of our model. A experiment result and final conclusion are shown in section (4) and (5).

2. Background

2.1. Down-Quality Model and MAP

Down-quality model expresses how data in video produce. It suppose four different step in this process: motion transform, blurring, downsample and noising. A common model is shown as Figure (1)[7].

Motion transform process means device or object movement at the time while capturing. Blurring represents optical blur or motion blur. Downsample is a operation that device capture object with a fixed resolution. Imaging sensor noise is also taken into consideration.

A certain frame and its near frames in video can be seen as many results from same down-quality model with a same input and different parameters. Multi-frame method need to reconstruct this input high-resolution frame from many output low-resolution frames.

A mathematic expression of down-quality model is that:

$$y_k = D_i B_i M_i z + n_i, k = 1, 2, 3...$$

Where z is ground-truth image, y_k is No.i images got from the same object, D_i means downsample operation, B_i means blurring, M_i means motion transformation and n_i is additional noise.

Solving z from many y_i can be written as probability problem:

$$z = P(z|y_1, y_2, ..., y_i)$$

To apply Bayes Theory to above equation and convert it into maximum posterior probability:

$$z = arg\max_z \frac{P(y_1, y_2, ..., y_i|z)P(z)}{P(y_1, y_2, ..., y_i)}$$

Since each y_k is independent:

$$\begin{aligned} z &= arg\max_z \Big[P(y_1, y_2, ..., y_i|z)P(z)\Big] \\ &= arg\max_z \Big[In(P(y_1, y_2, ..., y_i|z)P(z))\Big] \\ &= arg\max_z \Big[In(P(y_1, y_2, ..., y_i|z)) + In(P(z))\Big] \\ &= arg\max_z \Big[\sum_{i=1}^{k} In(P(y_i|z)) + In(P(z))\Big] \end{aligned}$$

Usually, to suppose n_i matches gaussian distribution with mean equals to 0, variance equals to σ:

$$\begin{aligned} P(n_i) &= \frac{1}{\sigma^2\sqrt{2\pi}} exp\Big(\frac{-(n_i^2)}{2\sigma^2}\Big) \\ &= \frac{1}{\sigma^2\sqrt{2\pi}} exp\Big(\frac{-((D_iB_iM_iz - y_i)^2)}{2\sigma^2}\Big) \end{aligned}$$

Then:

$$\begin{aligned} z &= arg\max_z \Big[In\Big(\frac{1}{\sigma^2\sqrt{2\pi}} exp\Big(\frac{-(D_iB_iM_iz - y_i)^2}{2\sigma^2}\Big)\Big)\Big] \\ &= arg\max_z \Big[In\Big(\frac{1}{\sigma^2\sqrt{2\pi}}\Big) + In\Big(-(D_iB_iM_iz - y_i)^2\Big) - In\Big(2\sigma^2\Big)\Big] \\ &= arg\min_z \Big[(D_iB_iM_iz - y_i)^2\Big] \end{aligned}$$

However, lots of value of z fit above equation, which is a ill-problem. To add prior knowledge and convert it into a constrained optimization problem:

$$z = arg\min_z \Big[(D_iB_iM_iz - y_i)^2 + \lambda R(z)\Big]$$

Where $R(z)$ can be considered as constrained condition for z and λ means relaxation coefficient of this constriction.

2.2. Prior Knowledge

In this paper, we mainly use one prior knowledge in two presentation ways: piecewise smooth, which means a image is consisted of certain number of smooth areas. In smooth areas, value of pixels does not change much. And between the smooth areas, which means details or bound areas, value of pixels does change much. What's more, though details or bound areas exit, they are not the main composition in image. In general, change of value of pixels in whole image should be small. In

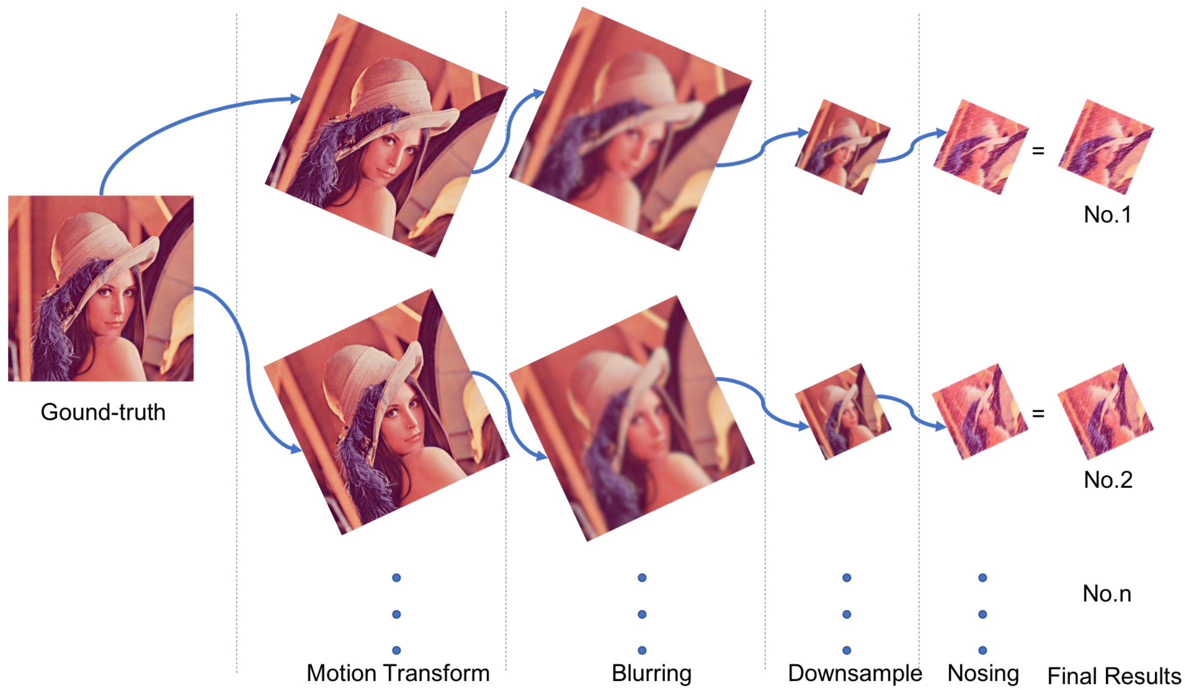


Figure 1. Down-Quality Model

mathematic language, change of value can be expressed as differential of value.

When it comes to a image in the space domain, the whole differential of value of the image is called total variation of the image. It is introduced by Fatemi[8]:

$$TV[u(x,y)] = \int\int |\nabla u(x,y)| dxdy$$

Where ∇ is two-dimensional gradient operator and $u(x,y)$ means a image in continuous space domain.

Under discrete space domain, total minimum can be expressed as two kinds: anisotropic total variation norm($l_1 - norm$):

$$TV_1 = \sum_{i=1}^{M}\sum_{j=1}^{N}\left|\frac{z_{i+1,j} - z_{i-1,j}}{2}\right| + \left|\frac{z_{i,j+1} - z_{i,j-1}}{2}\right|$$

and isotropic total variation norm($l_2 - norm$):

$$TV_2 = \sum_{i=1}^{M}\sum_{j=1}^{N}\sqrt{\frac{(z_{i+1,j} - z_{i-1,j})^2}{4} + \frac{(z_{i,j+1} - z_{i,j-1})^2}{4}}$$

Where z means discrete expression of u, M and N means width and height of image.

There is another way to express the same idea, which is sparse representation in certain transform domain. Compared with common way to express image in space domain, it aims to re-express a image to another transform domain, like Wavelet transform or Fourier transform. If number of non-zero value of image transformed into certain domain, is small(or sparse), it is said that this image can be sparse represented in that transform domain[6].

Researcher have proofed that sparse coefficient in certain domain also means sparse gradient in space domain, which have same conclusion with total variation minimum. So, sparse representation is another expression of our prior knowledge: piecewise smooth.

What sparse representation does is shown as Figure (2) and:

$$y = D\alpha$$

Where y is a vectorization of the image, D means transform matrix and α is sparse coefficient.

2.3. Relief of Inter-frames Blurring

Maximum posterior probability reconstruct single frame each time, and each output do not consider information between frames. That will lead to a blurring between frames. Traditional method to relieve it in image enhancement for each image.

Recent, the time domain information are taken into consideration[11]. It see video as a three-dimension data with width, height and time. Early researcher fulfill a Tikhonov regularization to video and keep con-

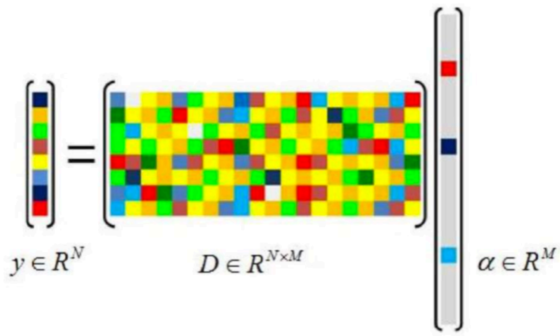


Figure 2. Sparse Representation

sistency of space-time data. However, it get unsatisfiable result on video inter-frames de-blurring. Based on this, a total variation in time domain is introduced[1]. Compared with traditional total variation in space domain, which means in width and height dimension, a total variation in time domain refers a gradient in time dimension of video. The mathematic expression can be shown with $l_1 - norm$ as:

$$TV_{time} = \sum_{i=1}^{M}\sum_{j=1}^{N}\sum_{t=1}^{L}\left|\frac{z_{i,j,t+1} - z_{i,j,t}}{2}\right|$$

Where M, N and t are width, height and time dimension data of video.

3. Proposed Method

Proposed method consists of two step. Firstly, we reconstruct every frame in the video with multi-frames super-resolution algorithm. In this algorithm, maximum posterior probability(MAP) with total variation minimum and sparse representation restriction is introduced. Secondly, for the video in step one, time domain smoothing method is fulfilled.

What we wants to do in step one, can be shown as Equation (1).

$$arg\min_{\alpha}\sum_{i=1}^{K}||D_iB_iM_i\Phi\alpha - y_i||_2 + \lambda(z)||\nabla z||_2 + \lambda_s||\alpha||_1 \tag{1}$$

Where α means a coefficient in the transform domain Φ, which is needed to be solved, y_i means i th frame used in this step, $z = \Phi\alpha$ means solution of this problem in time domain and ∇z means total variation of z.

However, it's hard to solve the first step problem directly and simply, since intermediate result should not fit two restriction at the same time. The solution process is divided into two-sub-problems. The first sub-problem is to solve MAP with total variation minimum restriction problem, which is shown as Equation (2). The second one is to make coefficient of result from above sub-problem sparse in a certain transform domain, which is shown as Equation (3).

$$arg\min_{z}\sum_{i=1}^{K}||D_iB_iM_iz - y_i||_2 + \lambda(z)||\nabla z||_2 \tag{2}$$

Where $\lambda(z)$ means a function. Output of the function relates to z and this self-adaptive coefficient decide extent of total minimum restriction.

$$arg\min_{\alpha}||\Phi\alpha - z||_2 + \lambda_s||\alpha||_1 \tag{3}$$

Where z means intermediate result obtained from first sub-problem, which means Equation (2), and λ_s is a constant to decide extent of sparse.

On the other hand, step two can be expressed as Equation (4).

$$arg\min_{z}\frac{\lambda_t}{2}||z - f||_2 + ||Dz||_1 \tag{4}$$

Where f means video obtained from step one, D mean forward forward difference operation, $Dz = \sum vec(f(x,y,t+1)) - vec(f(x,y,t))$ means expression of discrete total variation in time domain and λ_t controls degree of smoothing of video in time domain.

Following passage in this section is mainly aim to introduce how to solve above optimization problems. And Figure (3) concludes the whole process of our proposed method.

3.1. Total Variation minimum Restriction

In Equation (2), D_i, B_i and M_i means downsample matrix, blur matrix and motion compensation matrix, which change with y_i. Actually, we can suppose the same downsample matrix and blur matrix for every y_i. Then, Equation (2) can be expressed as Equation (5).

$$arg\min_{z}\sum_{i=1}^{K}||A_iz - y_i||_2 + \lambda(z)||\nabla z||_2 \tag{5}$$

Where $A_i = DBM_i$ is known before optimization operation.

To solve Equation (5) problem, we mainly implement variational method. So, Equation (5) can be re-expressed as Euler-Lagrange Equation (6).

$$\sum_{i=1}^{K}\frac{1}{2}A_i^T(A_iz - y_i) + \lambda(z)\partial(\nabla z) = 0 \tag{6}$$

Where ∂ is differentiation operation. Because A_i contain complex operation of image z and it is hard to

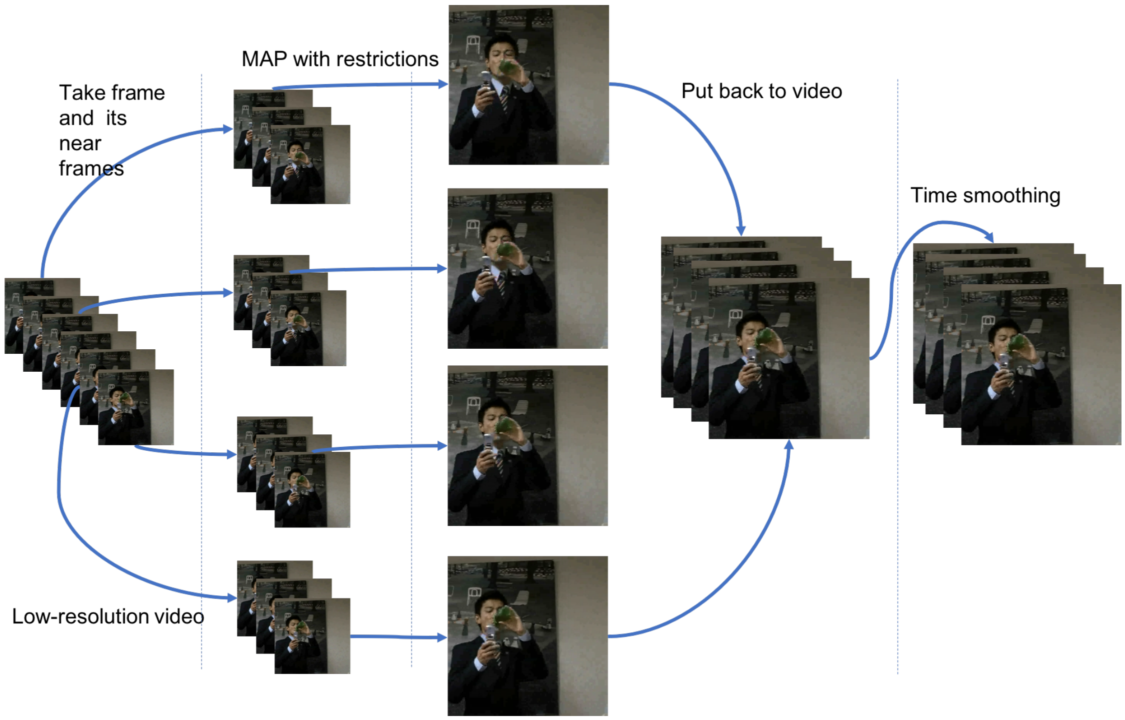


Figure 3. Process of Proposed Method

solve the problem directly, we usually solve it with iteration. At this paper, we mainly implement steepest descend method. Its iteration process can be shown as Equation (7) and Equation (8).

$$\hat{z}_{n+1} = \hat{z}_n - \alpha * G_n \tag{7}$$

Where α is a constant controls solution speed.

$$G_n = \sum_{i=1}^{K} \frac{1}{2} A_i^T (A_i z - y_i) + \lambda(z)\partial(\nabla z) \tag{8}$$

$\lambda(z)$ is a self-adaptive and important parameter in this problem, because it controls extent of total variation minimum restriction. The larger $\lambda(z)$ is, the smoother z is. If z was too smooth, some detail information of z would be loss. $\lambda(z)$ should be a function matched with z. When it comes to detail area in z, $\lambda(z)$ should be small. And in the smooth region, $\lambda(z)$ should be relatively large. This paper fulfills this relationship with barrier function[12]. With the aim to get $\lambda(z)$, we need to solve a given constant Th by 3-step scheme at the beginning:

Step1: compute the local variance of z

$$\sigma^2(x,y) = S \sum_{w=-W}^{w=W} \sum_{h=-H}^{h=H} [z(x+w, y+h) - \tilde{z}(x,y)]^2 \tag{9}$$

$$\tilde{z}(x,y) = S \sum_{w=-W}^{w=W} \sum_{h=-H}^{h=H} [z(x+w, y+h)] \tag{10}$$

Where W and H is width and height of z and $S = \frac{1}{(2W+1)(2H+1)}$.

Step2: computer corresponding binary image of z, defined as z_B. The background points set is shown as:

$$\Phi = \{(x,y) | z_B(x,y) = 0\} \tag{11}$$

Step3: to define a background region in $\sigma^2(x,y)$:

$$\Omega = \{\sigma^2_\Phi(x,y) | (x,y) \in \Phi\} \tag{12}$$

Mean value of Ω is the Th. When $\sigma^2(x,y) > Th$, point (x,y) is the detail region. And the value of $\sigma^2(x,y)$ in smooth region is smaller than Th. In conclusion, $\lambda(z)$ can be defined as:

$$\lambda_z(x,y) = \begin{cases} \lambda_{detail}(x,y) = \left\{ In[\frac{\sigma^2(x,y)}{Th}] + \frac{Th}{\sigma^2(x,y)} \right\} \\ \lambda_{smooth}(x,y) = 1.5 * \max\{\lambda_{detail}\} \end{cases} \tag{13}$$

Algorithm (1) concludes this whole process.

3.2. Sparse Representation Restriction

Since $l_1 - norm$ exits in Equation (3), optimization method in section (3.1) is not suitable here. A variable-split pre-process method is introduced in following passage.

Algorithm 1 Total Variation Minimum Restriction

Input K frames of data.
Input parameters D, B and M_i.
Initialize $z = y_{uint8(K/2)}$, $\alpha = 0.1$, $k = 0$
while not converge do
  1. For given z, compute specific $\lambda(z)$ using Equation (13).
  2. Compute gradient using Equation (8).
  3. Iteration of z using Equation (7).
  if $||z_{k+1} - z_k||_2/||z_k||_2 < tol_{map}$ then
    break.
  end if
end while

To express existed variable x as two positive vector subtraction: $x = u - v$, where $u = \max(0, x)$ and $v = \max(0, -x)$. Based on this idea, $l_1 - norm$ of x can be expressed as $|x|_1 = |u - v|_1 = 1_n^T u + 1_n^T v$. Similarly, Equation (3) can be shown as follow:

$$\begin{aligned} &arg \min_{u,v} ||\Phi\alpha - (u - v)||_2 + \lambda_s(1_n^T u + 1_n^T v) \equiv F(z) \\ &subjectto : u = \max(0, \alpha) \geq 0, v = \max(0, -\alpha) \geq 0 \end{aligned} \tag{14}$$

Equation (3) is a standard bound-constrained quadratic programming (BCQP) formulation and its gradient is shown as follow.

$$\begin{aligned} \nabla F(z) = &\left(\lambda_s 1_{2n} + \begin{bmatrix} -\Phi^T y \\ \Phi^T y \end{bmatrix}\right)^T z \\ &+ \frac{1}{2}\begin{bmatrix} -b \\ b \end{bmatrix}^T \begin{bmatrix} -\Phi^T\Phi & \Phi^T\Phi \\ \Phi^T\Phi & -\Phi^T\Phi \end{bmatrix} z \end{aligned} \tag{15}$$

Where $z = \begin{bmatrix} u \\ v \end{bmatrix}$.

This problem can be solved by gradient projection method[3], which means a additional projection operation in standard gradient descend method to make sure each intermediate result fits bound-constrained. Concretely:

$$W^k = max(z^k - \alpha\nabla F(z, 0) \tag{16}$$

$$z^{k+1} = z^k + \lambda(w^k - z^k) \tag{17}$$

Where α and λ are two important parameters control iteration speed and stability.

It is hard to choice a proper combination of these two parameter, since it has two many possibilities. Figueiredo indicates a good ways to overcome this challenge: to fix λ to 1 and performing a backtracking line search until a sufficient decrease is attained in $F(z)$[3]. It is a 3-step process shown as follow.

Step 1: To define a constrained set named $g^{(k)}$. It ensure z^k fix bound-constrained (non-negative) and existence of negative gradient of $F(z)$

$$g_i^{(k)} = \begin{cases} \nabla F(z^k)_i & z_i^{(k)} > 0 \; or \; F(z^k)_i < 0 \\ 0 & otherwise \end{cases} \tag{18}$$

Step 2: Based on step one, we hope to find a proper value of α(define as α_0) to make change of $F(z)$ max.

$$\begin{aligned} \alpha_0 &= arg \max_\alpha F(z^k - \alpha g^k) \\ &= \frac{(g^k)^T g^k}{(g^k)^T \begin{bmatrix} -\Phi^T\Phi & \Phi^T\Phi \\ \Phi^T\Phi & -\Phi^T\Phi \end{bmatrix} g^k} \end{aligned} \tag{19}$$

Step3: Though α_0 change $F(z)$ mostly, it may be not the best choice for iteration because it may be not stable. We then define a constant $\beta \in (0, 1)$ and build searching set of α: $\{\alpha_0\beta^k | k = 1, 2, 3..\}$. We should search the first element as final value of α in the set that can meets with following requirement:

$$F(\hat{z}^{k+1}) \leq F(z^k) - \mu\nabla F(z^k)^T \hat{z}^{k+1} \tag{20}$$

Where $\hat{z}^{k+1} = max(0, z^k - \alpha^k \nabla F(z^k))$ is also the Step4. Algorithm (2) shows all above solution.

Algorithm 2 Sparse Representation Restriction

Input transform domain matrix Φ.
Transform result of last step into a coefficient in Φ as α_0, which is also initialization of α.
while not converge do
  1. Compute gradient of Equation (14) using Equation (15).
  2. Choice the best parameter α using Equation (18), (19) and (20).
  3. Iteration of z using Equation (16) and (17).
  if $||z_{k+1} - z_k||_2/||z_k||_2 < tol_s$ then
    break.
  end if
end while

3.3. Time Domain Smoothing

When it comes to Equation (4), we optimize the whole video. Under the condition that video lasts a long time, the amount of data or length of $vec(f)$ in Equation (4) is so large that cause a huge time complexity. The way to solve the problem can not be similar to above two optimization method, since we need the higher efficiency.

In this paper, an Augmented-Lagrange-Based optimization method is implemented, which refers Chan's working[1]. We first define a intermediate variable u to convert this problem into a equivalent problem:

$$\begin{aligned} &arg\min_{z,u} \frac{\lambda_t}{2}||z-f||_2 + ||u||_1 \\ &subjectto : u = Dz \end{aligned} \tag{21}$$

Where $D(z)$ has same definition as Equation (4) and its corresponding Augmented-Lagrange-Equation is as follow:

$$L(u,v,z) = \frac{\lambda_t}{2}||z-f||_2 + ||u||_1 - y(u-Dz) + ||u-Dz||_2 \tag{22}$$

With Alternating Direction Method of Multipliers, above problem can be split into three-sub-problems.

$$z_{k+1} = arg\min_{z} \frac{\lambda_t}{2}||z-f||_2 - y_k(u_k - Dz) + ||u_k - Dz||_2 \tag{23}$$

$$u_{k+1} = arg\min_{u} ||u||_1 - y_k(u - Dz_{k+1}) + ||u - Dz_{k+1}||_2 \tag{24}$$

$$y_{k+1} = y_k - 2(u_{k+1} - Dz_{k+1}) \tag{25}$$

Z-sub-problem(Equation (23)) have its analytical solution[1]:

$$(\lambda_t + 2D^T D)z = \lambda_t f + 2D^T \lambda_t - D^T y \tag{26}$$

Analytical solution of u-sub-problem(Equation (24)) can obtained by shrinkage formula[4]:

$$u = \max\left\{|Df + \frac{1}{2}y| - \frac{1}{2},\ 0\right\} * sign(Df + \frac{1}{2}y) \tag{27}$$

Algorithm (3) expresses all the steps.

Algorithm 3 Total Variation minimum

Input video processed by method in section (3.1) and (3.2).
Express Equation (4) as three sub-problems.
while not converge **do**
  1. Solve Equation (23) z-sub-problem using Equation (26).
  2. Solve Equation (24) u-sub-problem using Equation (27).
  3. Update value of y using Equation (25).
  if $||z_{k+1} - z_k||_2/||z_k||_2 < tol_t$ **then**
    break.
  end if
end while

Parameter	Value
tol_{map}	10^{-1}
tol_s	10^{-2}
tol_t	10^{-3}
λ_s	5.325
λ_t	0.591

Table 1. Experiment Parameters

4. Experiment

4.1. Design

Videos used in this experiment come from two movies, one anime and one video recorded by myself. They belong to different content which are *Car*, *Anime*, *Person* and *Windows*. Each video last about 2 seconds, with the aim of computation speed. Frames in each video are about thirty. Especially, gaussian distribution noised videos are taken into consideration with mean equals to 0 and variation equals to 0.01.

The low-resolution videos are obtained from original high-resolution videos by downsample-operation. Size of low-resolution video is 150×150, while 600×600 for high-resolution at the same time. That means the enhance factor in this experiment is four.

Though methods discussed in this paper just include single-channel image/video, this experiment use three-channel image/video. We split channel to R, G, B at the beginning, deal with them separately and combine them at the end.

We make each frame and its near four frames as input of MAP with restrictions method. Other parameters are listed as Table (1).

The objective estimation mainly compared Peak Signal Noise Rate(PSNR) between constructed videos

4.2. Results

In this sub-section, results from four videos are introduced shown as Figure (4), Figure (5), Figure (6) and Figure (7). Each Figure is consisted of two part, sub-figure (a) to (e) belong to the whole part of certain frame in video and sub-figure (f) to (j) belong to the details corresponding to sub-figure (a) to (e).

4.3. Analysis

4.3.1 Subjective

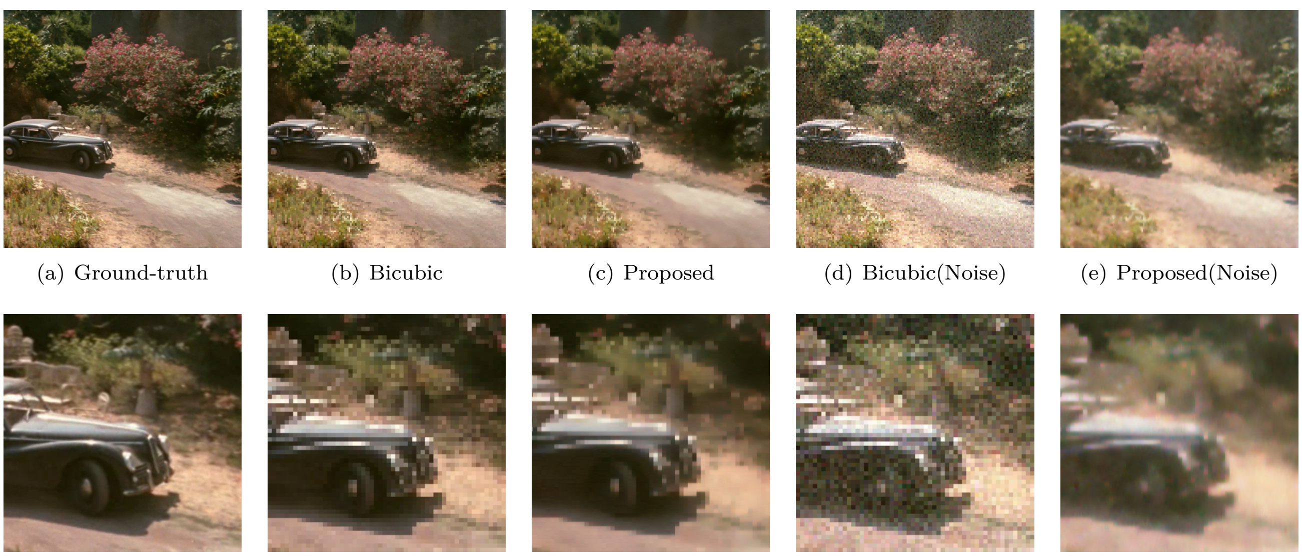
As shown in result figures, for one thing, proposed method can efficiently eliminates jaggies and mosaics. In the following description, I will mainly focus on frame's detail part.

The bicubic result of *Anime* video, which means sub-figure(b) and (g) in Figure (4), have obvious jag-

(a) Ground-truth (b) Bicubic (c) Proposed (d) Bicubic(Noise) (e) Proposed(Noise)

(f) Ground-truth (g) Bicubic (h) Proposed Method (i) Bicubic(Noise) (j) Proposed(Noise)

Figure 4. Result of *Anime* Video



(a) Ground-truth (b) Bicubic (c) Proposed (d) Bicubic(Noise) (e) Proposed(Noise)

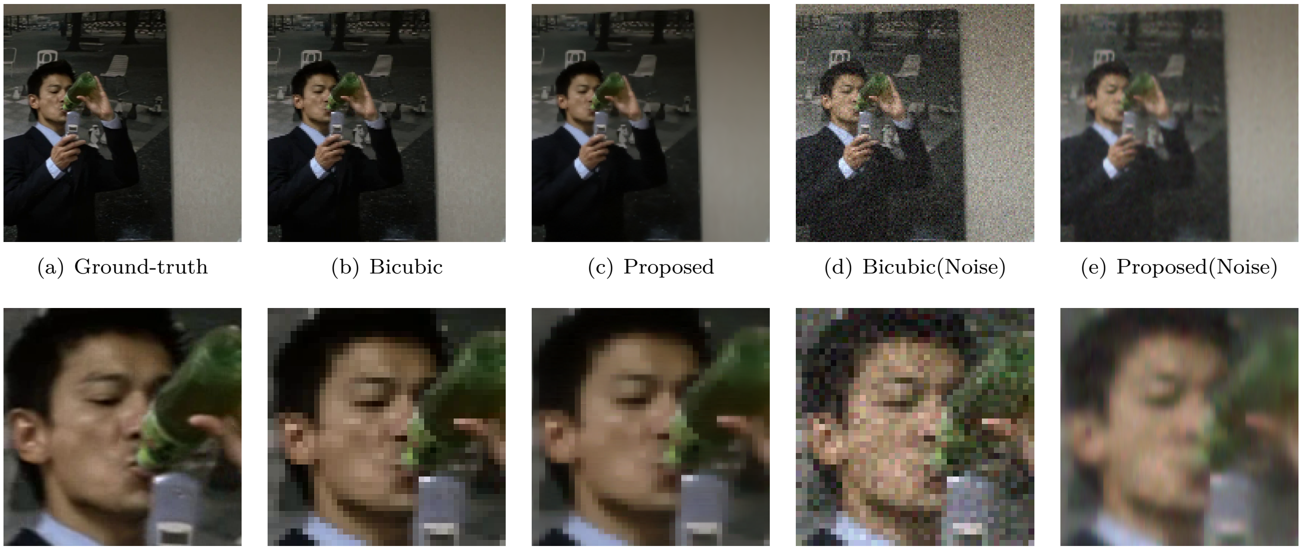
(f) Ground-truth (g) Bicubic (h) Proposed Method (i) Bicubic(Noise) (j) Proposed(Noise)

Figure 5. Result of *Car* Video

gies in character's hair and mosaics in plants of right part of image. In the final result of *Anime* video, which means sub-figure(c) and (h) in Figure (4), jaggies in hair disappears and character's hair become really smooth. And mosaics of plants also become blur.

The bicubic result of *Car* video, which is sub-figure(b) and (g) in Figure (5), have obvious jaggies in outline of the car and mosaics in plants of background. In the final result of *Car* video, which is sub-figure(c) and (h) in Figure (5), jaggies in outline of are relieved and plants of background become smooth.

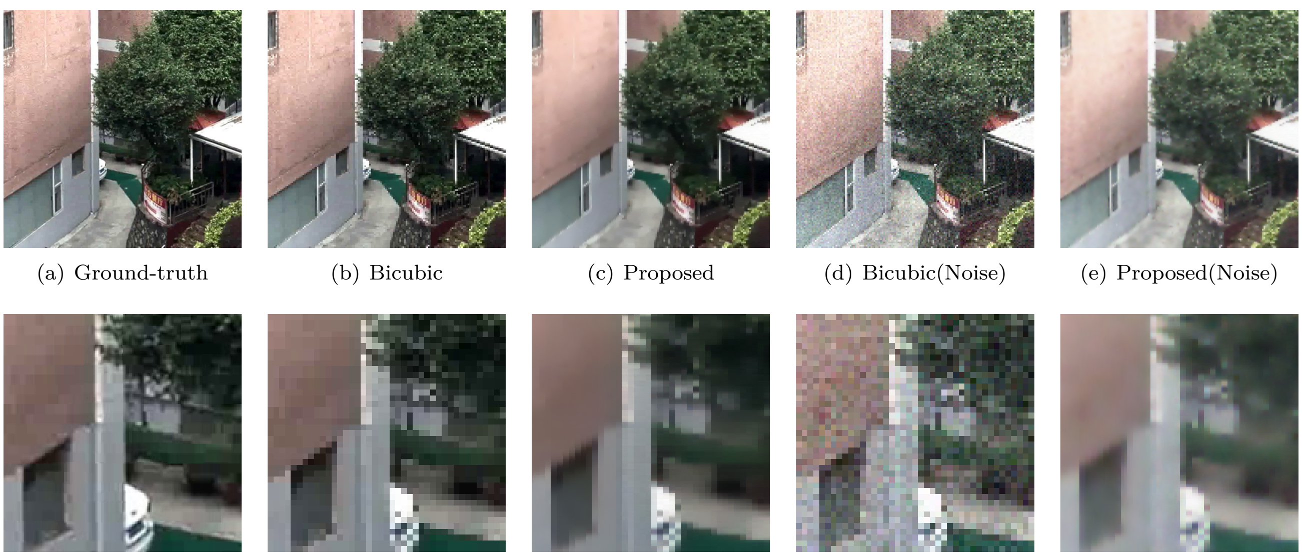
There is also jaggies and mosaics in *Person* video, like jaggies in the bound between human face and hair, and in human's finger(see sub-figure (g) in Figure (6)), and mosaics in human's face itself. Despite that, jaggies described above disappeared in sub-figure (h) in Figure (6) and mosaics in human's face also described above become smooth.



(a) Ground-truth (b) Bicubic (c) Proposed (d) Bicubic(Noise) (e) Proposed(Noise)

(f) Ground-truth (g) Bicubic (h) Proposed Method (i) Bicubic(Noise) (j) Proposed(Noise)

Figure 6. Result of *Person* Video



(a) Ground-truth (b) Bicubic (c) Proposed (d) Bicubic(Noise) (e) Proposed(Noise)

(f) Ground-truth (g) Bicubic (h) Proposed Method (i) Bicubic(Noise) (j) Proposed(Noise)

Figure 7. Result of *Windows* Video

Jaggies and mosaics appear in *Windows* video at the same time. Outline of windows and trees are example of jaggies and mosaics shown in sub-figure (g) in Figure (7). However, that jaggies in outline of windows disappear and mosaics in the tree become blur in sub-figure (h) in Figure (7), indicate efficiency of proposed method.

For another, proposed method cause some details information loss, since piecewise smooth suppose exists. When it comes to this aspect, we compare more about ground-truth frame and proposed method frame.

In *Anime* video, just like what we have taken in above passage, character's hair become smooth. But hair in ground-truth, a white detail exists and this detail in proposed method disappears. What finally leave in proposed method is just a blur bound.

In *Car* video, ground-truth frame have details about road, like variable textures. However, textures in proposed method do not exit. Roads are really smooth.

This phenomenon is not obvious in *Person* video. One the contrary, wall in *Windows* video is fairly smooth. Compared with ground-truth frame, proposed method frame loss lots of information about textures in the wall.

Though our proposed method cause some details loss, it does not means content of video is destroyed. Because what proposed method destroy is some sample noises and textures, not the main information of video. It does not interfere human's understand and feeling of the video.

What's more, proposed method have huge ability to video de-noise. For both four results of videos, the bicubic results are all so terrible that are highly affect human's understanding for the videos' content. In proposed method, noises data in bicubic results are clean up totally and some details information are recovered.

4.3.2 Objective

In this section, we mainly analyse results of value of psnr from four experiment videos.

Each value of PSNR of each frame in videos shown as Figure(8). It shows that in each frame of video, our proposed method have a better performance. The line of psnr is stable and no point is lower than result of bicubic, which means a great robust of proposed method.

Mean value of PSNR of all four videos are shown as Tale (2). The average increase of PSNR between bicubic result and proposed method is about 1dB. It proofs usefulness of our algorithm.

5. Conclusion

This paper introduces a novel video super-resolution method. It firstly implements MAP, total variation minimum restriction and sparse representation restriction to each frame of video. And then implement a total variation minimum in time domain for the whole video, to get better video quality and relieve inter-frame blurring.

Compared with common bicubic method ,experiment result reveals that proposed method can effectively relieve jaggies and mosaics. Though it cause some detail information loss, proposed method do not prevent human from understanding and having good a feeling for the video. What's more, experiment also demonstrates a surprising de-nosing performance.

References

[1] S. H. Chan, R. Khoshabeh, K. B. Gibson, P. E. Gill, and T. Q. Nguyen. An Augmented Lagrangian Method for Total Variation Video Restoration. IEEE Transactions on Image Processing, 20(11):3097–3111, 2011. 2, 4, 7

[2] C. Dong, C. C. Loy, K. He, and X. Tang. Image Super-Resolution Using Deep Convolutional Networks. IEEE Transactions on Pattern Analysis and Machine Intelligence, 38(2):295–307, Dec. 2015. 1

[3] M. A. T. Figueiredo, R. D. Nowak, and S. J. Wright. Gradient Projection for Sparse Reconstruction: Application to Compressed Sensing and Other Inverse Problems. IEEE Journal of Selected Topics in Signal Processing, 1(4):586–597, 2007. 6

[4] C. Li. An Efficient Algorithm For Total Variation Regularization with Applications to the Single Pixel Camera and Compressive Sensing. Dissertations & Theses - Gradworks, 2010. 7

[5] Z. Li, W. Suyu, and L. Xiaoguang. Image/Video Super Resolution. POST&TELECOM PRESS(Chiness), 2011. 1

[6] S. G. Mallat and Z. Zhang. Matching pursuits with time-frequency dictionaries. IEEE Transactions on Signal Processing, 41(12):3397–3415, Dec. 1993. 3

[7] S. C. Park, K. P. Min, and M. G. Kang. Super-resolution image reconstruction: a technical overview. IEEE Signal Processing Magazine, 20(3):21–36, 2003. 1, 2

[8] L. I. Rudin, S. Osher, and E. Fatemi. Nonlinear total variation based noise removal algorithms. Physica D: Nonlinear Phenomena, 60(1):259–268, Nov. 1992. 3

[9] W. C. Siu and K. W. Hung. Review of image interpolation and super-resolution. In Signal Information Processing Association Summit and Conference, pages 1–10, Oct. 2012. 1

[10] H. Su, J. Zhou, and Z. H. Zhang. Survey of super-resolution image reconstruction methods. Acta Automatica Sinica, 39(8):1202–1213, 2013. 1

[11] Y. Wexler, E. Shechtman, and M. Irani. Space-Time Completion of Video. IEEE Transactions on Pattern Analysis and Machine Intelligence, 29(3):463–476, 2007. 3

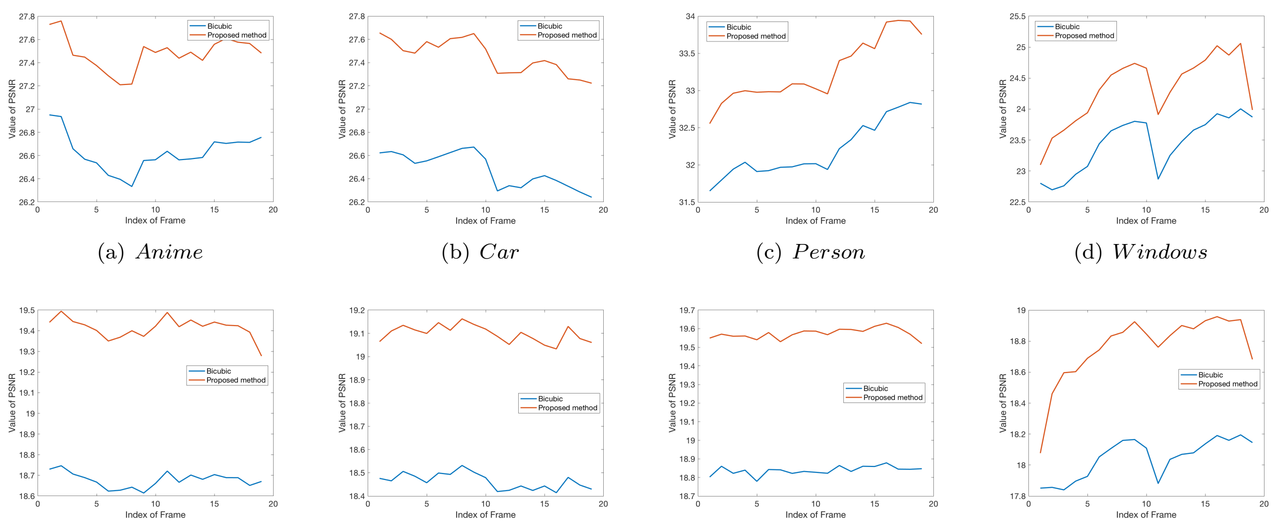
[12] Z. Wu, H. Gao, G. Ma, and L. Wu. An Improved Self-adaptive Regularization Method for Mixed Multiplicative and Additive Noise Reduction. In Pattern Recognition, pages 690–702. Springer Singapore, Singapore, Oct. 2016. 5

[13] J. Yang, J. Wright, T. S. Huang, and Y. Ma. Image Super-Resolution Via Sparse Representation. IEEE Transactions on Image Processing, 19(11):2861–2873, Oct. 2010. 1

[14] L. Yue, H. Shen, J. Li, Q. Yuan, H. Zhang, and L. Zhang. Image super-resolution: The techniques, applications and future. Signal Processing, 128(C):389–408, Nov. 2016. 2

Video Type	Mean PSNR(Bicubic)	Mean PSNR(Proposed)
$Anime$	26.63dB	27.48dB
$Anime(Noise)$	18.68dB	19.41dB
Car	26.48dB	27.45dB
$Car(Noise)$	18.46dB	19.10dB
$Person$	32.20dB	33.27dB
$Person(Noise)$	18.84dB	19.57dB
$Windows$	23.44dB	24.32dB
$Windows(Noise)$	18.05dB	18.76dB

Table 2. Experiment Parameters



(a) $Anime$ (b) Car (c) $Person$ (d) $Windows$

(e) $Anime - Noise$ (f) $Car - Noise$ (g) $Person - Noise$ (h) $Windows - Noise$

Figure 8. Result of Value of PSNR