Journal of Al-Nahrain University Vol.15 (1), March, 2012, pp.170-174 Science

Enhancement of Video Image Resolution Based on POCS (projection onto convex set)

Maisaa Husam Al_anizy
Department of Computer Science, College of Medicine, Al-Mustansria University.

Abstract

Our approach based on enhancing the resolution of video (data set of image which captured) through the addition of perceptually plausible high frequency information. By introducing an appropriate prior distribution over such data set we can ensure consistency of static image regions across successive frames of the video, and also take account of object motion. A key concept is the use of the previously enhanced frame to provide part of the training set for super-resolution enhancement of the current frame and that by adopted the local variance of reference frame as the foundation to establish the operator for projection onto convex set (POCS) and local threshold value for pixel-repair. And the results show that an improvement in video quality can be achieved.

Keywords: Super-Resolution, POCS, LR, HR, Deblurring.

1. Introduction

Video sequences usually contain a large overlap between successive frames, and regions in the scene are sampled in several images. This multiple sampling can sometime be used to achieve images with a higher spatial resolution. The process of reconstructing a high resolution image from several images covering the same region in the world is called Super Resolution. Additional tasks may include the reconstruction of a high resolution video sequence [1].

2. Review of Related Work

The problem of recovering a high resolution image from a sequence of low resolution compressed images considered is proposed by (Liyakathunisa; Kumar, C.N.R.; Ananthashayana, V.K.;2009) used a lifting schemes for intentionally introducing down sampling of the high resolution image sequence before compression and then utilize super resolution techniques for generating a high resolution image at the decoder.

Other proposed deals with high resolution video reconstruction from low resolution video by using an algorithm for enhancing the resolution of video through histogram based segmentation and frequency domain registration is proposed by (Madhusudhan, T.; Pais, Alwyn Roshan;2007).

The low resolution input images are the result of projection of a high resolution image onto the image plane, followed by sampling. The goal is to find the high resolution image which fits this model.

Formulating it in mathematical language [2]: Given K images $\{X_L^n\}_{n=1}^K$ $of\ size\ M_1 x\ m_2$ find the image X_H of size$N_1 x\ N_2$, which minimizes the error function:

$$E(X_H) = \sum_{n=1}^{K} \|P_n(X_H) - X_L^n\|^2 \quad \ldots\ldots\ldots\ldots\ldots\ldots(1)$$

3.Super-Resolution Reconstruction

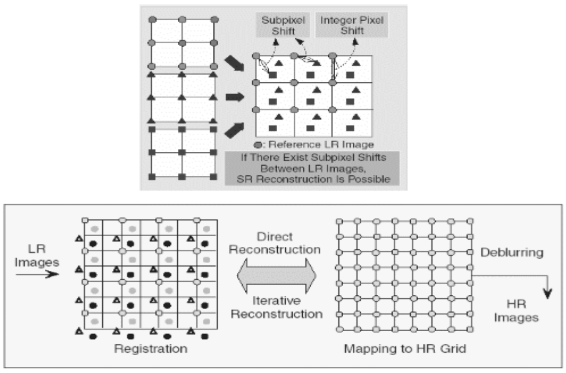


Fig.(1) Super Resolution Reconstruction.

The reconstruction process as shown in Fig.(1) of the reconstruction-based super-resolution can be formulated as an inverse problem that estimates the source of the high resolution image from observed low-resolution images assuming an image generative model. Several image generative models have been proposed for SR reconstruction. A widely used image generative model is:

$$y_i = DH_iF_ix + v_i \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(2)$$

Where y_i is the vectorized i-th observed image, x is the vectorized high resolution image, F_i is a matrix representing the motion of the i-th image, H_i is a matrix representing the point-spread function (PSF) of the i-th image, D is a matrix representing the down sampling, and vi represents the noise of the i-th image.

A *Maximum a Posteriori* (MAP) estimation is the popular solution of the inverse problem associated with the generative model in Eq. (3). The MAP estimation is performed by minimizing the cost function:

$$E = \sum_{i=1}^{N} \|y_i - D\, H_i F_i x\|_2^2 + \alpha\, \lambda(x) \ldots\ldots\ldots (3)$$

where $\lambda(x)$ is a constraining function derived from the prior distribution of the high-resolution image, α is a hyper-parameter, and N is the number of observed images. The cost function is the negative logarithm of the posterior distribution of the high-resolution image when the observed low-resolution images are given, wherein the noise model is assumed to be independent Gaussian distribution[4].

4. Robust SR Reconstruction

First, we show the pixel-wise cost function to express a general robust cost function. The pixel-wise generative model can be derived from Eq. (3) as

$$y_{i,j} = b_{i,j}^T .x + v_{i,j} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (4)$$

where yi,j is the j-th pixel value of the i-th observed image, vi,j represents the noise of the j-th pixel value of the i-th observed image, and $bT\ i,j$ is defined as

$$DH_iF_i = \begin{pmatrix} b_{i,1}^T \\ b_{i,2}^T \\ \ldots. \end{pmatrix} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (5)$$

The cost function of the MAP estimation in Eq. 8.2 can also be rewritten as a pixel-wise expression:

$$E = \sum_{i=1}^{N} \sum_{j=1}^{M_i} \left[y_{i,j} - b_{i,j}^T.x\right]^2 + \alpha\lambda(x) \ldots\ldots\ldots (6)$$

where Mi is the pixel number of the i-th observed image. The occlusion and/or the registration error are not involved in the pixel wise generative model. Therefore, we must reject the occluded pixels and the pixels with registration error as outliers for SR reconstruction. This is the general concept of robust SR reconstruction.

The mathematical expression to reject the certain pixel is to assign it zero weight. The other approach is to use a robust error function or M-estimator instead of the $L2$ norm. The general expression of the cost function of the robust SR reconstruction is:

$$\boldsymbol{I} = \sum_{i=1}^{N} \sum_{j=1}^{M_i} \boldsymbol{w_{i,j}\, \rho\left(b_{i,j}^T.x - y_{i,j}\right) + \alpha\lambda(x)} \ldots\ldots \mathbf{(7)}$$

Where, wi,j is a real value weight or a binary mask for the j-th pixel of the i-th observed image, and function $\rho(x)$ is a robust error function. The key of robust super-resolution is how to design the weight and robust error function adaptively.We overview the pixel-weight approach. In the pixel-weight approach, zero weights or small weights are used to reject the outlier pixels, where the error function is the $L2$ norm.

4.Deblurring

Since we did not include the effect of sensor blur in the data model of the KR framework, deblurring is necessary as a post processing step to improve the outputs by 3-D is KR further. Defining the estimated frame at time t as[4]

$$\boldsymbol{\hat{z}(t)} = \left[\ldots, \boldsymbol{\hat{z}(x_j)}, \ldots\right]^{\boldsymbol{T}} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (8)$$

Where j is the index of the spatial pixel array and $u(t)$ as the unknown image of interest, we deblur the frame $z(t)$ by a regularization approach:

$$\hat{u}(t) = \arg \min_{u} \left\| u(t) - G\hat{z}(t \quad \_ \right\|_2^2 + \lambda C_R\big(\hat{u}(t)\big) \ldots\ldots\ldots\ldots\ldots\ldots. (9)$$

where G is the blur matrix, $\lambda(\geq 0)$ is the regularization parameter, and CR(u) is the regularization term. More specifically, we rely on our earlier work and employ the *bilateral total variation* (BTV) framework:

$$C_R\big(u(t)\big) = \sum_{v_1=-v}^{v} \sum_{v_2=-v}^{v} \eta^{|v_1|+|v_2|} \left\| u(t) - F_{x_1}^{v_1} F_{x_2}^{v_2} u(t) \right\|_1 \ldots\ldots\ldots (10)$$

where η is the smoothing parameter, v is the window size, and $Fv1\ x1$ is the shift matrix that shifts $u(t)$ $v1$-pixels along $x1$-axis.

In the present work, we use the above *Bilateral Total Variation* (BTV) regularization framework to deblur the up scaled sequences frame-by-frame, which is admittedly suboptimal[4].

In our work we have introduced a different regularization function called *adaptive kernel total variation* (AKTV). This framework can be extended to derive an algorithm that can simultaneously interpolate and deblur in one integrated step.

5. Projection onto Convex Set

According to the method of POCS for super-resolution reconstruction, the space of estimated high-resolution solutions is restricted by a set of constraints (closed convex sets) which characterize desirable properties, such as fidelity to data, smoothness, sharpness *etc.*, to be consistent in the final solution. For each set of convex constraints C_i, a projection operator T_i is defined. The problem is then reduced to iteratively locate, given a point in the high-resolution image space, the closest solution which intersects with all the given convex constraints, C_i. The convergence can be given as[5]:

$$X^{n+1} = T_i X^n \Rightarrow X^{n+1} = T_m T_{m-1} \ldots T_2 T_1 X^n \quad \text{for } n = 0,1,2 \ldots\ldots\ldots\ldots (11)$$

6. Motion Estimation Error

Motion estimation error in a super-resolution image reconstruction algorithm is an important part and it relates to the adjacent sub-pixel information between frames can be effectively utilized. The use of the image intensity of the inherent two-dimensional motion estimation (pore size, cover, reveal effects, etc.), the result of motion estimation error can be seen as noise. To effectively control not accurate motion estimation errors, need to judge these areas. This approach will be classified as POCS, can create noise on the error set of constraints as follows [6]:

$$C_M^k[n_1, n_2] = \{x[n_1, n_2]: |D(n_1, n_2)| \le \sigma_k\} \ldots\ldots\ldots\ldots\ldots (12)$$

There for:

$$D(n_1, n_2) = \sum_{p=-1}^{1} \sum_{q=-1}^{1} \big| X\big(a(n_1 + p) + MV_y, b(n_2 + q) + MV_x\big) - y(n_1 + p, n_2 + q) \big| \ldots\ldots\ldots\ldots (13)$$

σk is a reference frame fixed in the center of that neighborhood where the standard deviation, and its value reflects the size of the local characteristics of image; when σk is large the texture regions of image be detailed, and when σk is small the area will be smoothed.

POCS established operator as:

$$P_{d(n_1,n_2)} = \begin{cases} \text{internal repair}, & |D(n_1 n_2)| < \sigma_k(n_1, n_2) \\ \text{directional interpolation}, & |D(n_1, n_2)| > \sigma_k(n_1, n_2) \end{cases} \ldots\ldots\ldots\ldots\ldots (14)$$

When the motion estimation of the projection error D (n1, n2)> σk (n1, n2), indicating the motion estimation error is larger at this time, while maintaining image quality as much as possible, the region-based direction of interpolation will repair.

7. Experimental Result and Analysis

First, translation motion for each low resolution frames and set resolution factor, and implemented them by HPSF(Hi function to all frames and space invariant).

Estimates (Robustly) the blurred high resolution as a median of the LR images after up sampling and shifting to correct position.

An efficient implementation is performed here in which we note that the high resolution image consists of the median of all the LR images which have the same displacement value. That is, the LR images are partitioned to unique groups, each group have equal displacement values in the HR image. The median of each group is then up sampled and shifted into the HR image.

In some cases, not all displacements exists. This leaves us with undetermined cases. In this case some other interpolation method needs to be implemented. In this implementation we "fill" the holes using a spatial median filter. Computes the gradient back-projection for the fast deblurring and interpolation method. This function implements the gradient of the level one norm between the blurred version of the current deblurred HR estimate and the original blurred HR estimate created by the median and shift method.

HR reconstruction of the i-th frame given the whole image sequence. Eventually, we evaluate the conditional expectation of $SR(v_i)(x)$ for any x $2\in \psi$, given the information of all of the frames $\{u_j\}$ for $1< j<k$:

$$E\left[SR(v_i)(x)\{u_j\}_{1\le j\le k}\right] = \frac{1}{G(i)}\sum_{i\le j\le k} g(|i-j|) \times E\left[SR(v_i)(x)\middle|u \times_j\right] \ldots\ldots\ldots\ldots\ldots\ldots..(15)$$

$$G(i) = \sum_{i\le j\le k} g(|i-j|) \ldots\ldots\ldots\ldots..\ldots(16)$$

where g is a decaying function of |i-j|, and G is a normalization factor. The expression g(|i-j|) in this equation represents the temporal confidence on the expectations computed for each of the various frames, j, which has been taken into account in reconstructing the HR image $SR(v_i)$. In the experiments reported below, we have assumed that each of the frames in hand are equally likely useful in producing the HR details of the i-th frame. Hence, we have taken g to be a box-function of large enough support which yields g(|i-j|)= 1. As a result,

$$E\left[SR(v_i)(x)\{u_j\}_{1\le j\le k}\right] = \frac{1}{k}\sum_{i\le j\le k} E\left[SR(v_i)(x)\middle|u_j\right] \ldots\ldots\ldots\ldots..\ldots(17)$$

By taken a frame from motion sequence of video, we've set resolution factor and implemented with HPSF function, as in equation blow:

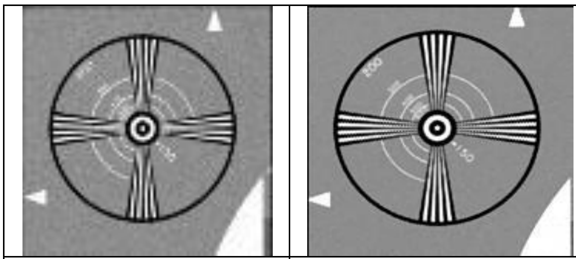
$$\forall_{\omega}\, 0 < \left|1 - \frac{HPSF(\omega)HAUX(\omega)}{C}\right| < 1 \ldots(18)$$

And deblurred ads in *eq.(9, 10)* the LR frames, we got the results as in Fig.(2) below which is a sequences of image.

LR Object Image	HR Object Image

Fig.(2) Sequences of Image.

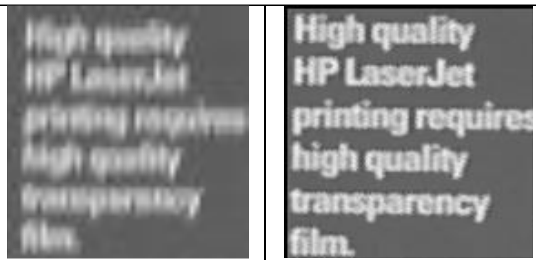
In same way, we got the results for other kind of sequenced image as circle rotation in Fig.(3).



LR circle motioned image	HR Circle motioned image

Fig.(3) Circle Rotation.

Other results of sequences text image in Fig.(4).



LR Text motioned Image	HR Text motioned Image

Fig.(4) Text Image.

Also, we test a human motion, as shown in Fig.(5).

	
LR Human motioned Image	HR Human motioned Image

Fig.(5) Human Motion.

References

[1] Bir Bhanu, Ju Han; "**Human Recognition at a distance in video"**; Springer-Verlag London Limited; pp 277; **2011.**

[2] Peyman Milanfar; "**Super Resolution Imaging**"; Taylor and Francis Group, LLC; pp 492; **2011.**

[3] Subhasis Chaudhuri; **"Super Resolution Imaging"**; Kulwer Acadmic Press,Boston ; pp 271; **2001.**

[4] Subhasis Chaudhuri, Manjunath V. Joshi; **"Motion Free Super Resolution"**; Springer Science + Business Media, Inc. ; **2005**.

[5] Vivek Bannore; **"Iterative-Interpolation Super-Resolution Image Reconstruction";** Springer-Verlag Berlin Heidelberg; pp 120; **2009**.

[6] Aggelos K. Katsaggelos, Rafael Molina and Javier Mateos; "**Super Resolution of Image and Video**"; by Morgan & Claypool; pp 150; **2007**.

الخلاصة

النهج الذي سنتبعه لتحسين مستوى الصور المستلمة من مجموعة صور فديوية هو بالاعتماد على خزين من البيانات المتراكمة لأفلام مماثلة لغرض الوصول إلى دقة عالية في تحسين اللقطة المشوشة الحالية , حيث يتم مقارنة بيانات اللقطة المشوشة مع مجموعة كبيرة من البيانات الخاصة بلقطات واضحة مماثلة ذات وضوح عالي لغرض الوصول إلى أقرب حالة تطابق مع القطة المشوشة والتي سيتم استبدالها باللقطة الواضحة المقاربة لها ، على أن يتم قبل ذلك التأكد من الانسيابية في اللقطات من خلال مقارنة اللقطة السابقة واللاحقة مع اللقطات المناظرة لها في الفلم الواضح ، وكذلك بالاستفادة من حركة الجسم الموجود في اللقطة المشوشة . هذه التقنية المتبعة في أسلوب مقارنة اللقطات المشوشة مع اللقطات المماثلة الواضحة التي تلتقطها العدسة المحدبة وتعرف بـ(POCS) لغرض التحديد الدقيق لقيمة جميع النقاط المكونة للصورة .وقد أظهرت النتائج بأن تحسن في نوعية الفيديو يمكن تحقيقه .