2009 International Conference on Advances in Recent Technologies in Communication and Computing

Super Resolution of Video with sharpened edges using Multiple Frames – A Novel Approach

Alwyn Roshan Pais, Department of Computer Engineering. National Institute of Technology Karnataka, Surathkal, INDIA alwyn.pais@gmail.com

John D'Souza Department of E&C Engineering. National Institute of Technology Karnataka, Surathkal, INDIA john_krec@yahoo.com

G Ram Mohan Reddy Department of Information Technology, National Institute of Technology Karnataka, Surathkal, INDIA profgrmreddy@gmail.com

Sandeep B Patil Department of Computer Engineering. National Institute of Technology Karnataka, Surathkal, INDIA patil.sandip@gmail.com

Abstract

In most electronic video applications, video with Higher Resolution(HR) are often required. HR video can offer more details that may be critical in various applications like Video Surveillance. Super resolution technique refers to generation of high resolution image from low resolution image by adding some extra information. Super resolution allows us to reduce the need of extra hardware to obtain HR image. This paper proposes a new method for super resolution video generation with sharpened edges using multiple frames. Unique information present in subpixel shifted frames is extracted and used for frequency domain registration process.

Key words— **Super resolution, frequency domain registration.**

1. Introduction.

The number of pixels per unit area (i.e. the spatial resolution) in an image is the principal factor in determining the quality of an image. Since the HR image not only gives the viewer a good quality picture but also gives details that are important for the analysis in many applications. So there is a big demand for high-resolution (HR) images.

Edges in images contribute largely to the visual perception quality. So it is necessary to sharpen or at least retain edges during the process of super resolution.

In our work, we have designed and implemented a system for super resolution of a color video with sharpening of edges. Unique information due to subpixel shifting is extracted using frequency domain approach.

Rest of the paper is organized as follows. Related work is explained in section 2. Section 3 elaborates on the proposed solution. Experimental results are given in section 4. Section 5 gives the quantitative analysis of the results obtained. We conclude the paper in Section 6.

2. Related Work.

Generation of high resolution (HR) images from low resolution (LR) images is achieved through reconstruction based and learning based approaches.

Any low resolution image has certain amount of information from a scene. In reconstruction based methods multiple low resolution images of the same scene are used to gather the extra information needed for the HR. Its objective is to extract the independent information from each image and combine the information into a single high resolution (HR) image. It is mandatory that each LR image must contain some unique information which is not present in any other image. This means that when these LR images are mapped onto a common reference plane their samples must be sub-pixel shifted from samples of other images, otherwise the images would contain only redundant information and SR reconstruction would not be possible. Frequency domain approach is one of the methods used for reconstruction [1].

Learning based approaches build a relation between LR and HR images, based on the imaging process and/or description of corresponding edges between LR and HR. The learning-based methods rely on the learning of characteristics of a specific image set to inject the extra information for HR generation [1].

978-0-7695-3845-7/09 $26.00 © 2009 IEEE
DOI 10.1109/ARTCom.2009.153

There are many super-resolution techniques and algorithms. Some uses kernel based interpolation techniques like bilinear and bicubic [6]. Another approach that is explicitly directed at maintaining sharp edges is the use of Sub-pixel edge localization. This approach detects the most prominent edge in the local window with sub-pixel precision and uses the resulting edge template to obtain sharper edges in super-resolved images. Use of Neural network to train the algorithm to achieve better visual quality for a specific image set is also explored [7].

Proposed solution for super resolution works on reconstruction based approach. Frequency domain based method is used to estimate the rotation and planner shifts.

3. Methodology.

Digital video consists of sequence of multiple frames which are similar to each other but not identical. In video, there will be frames which are shifted by sub pixel distances relative to the previous (reference) image. In this case these frames will be having some unique information which is present in that image only. This information will be used collectively to improve the quality of super resolved video. Major focus is also on sharpening the edges in video.

The design of solution for the problem is explained in the following algorithm-

3.1 Algorithm:

1. Histogram based segmentation is used to group the similar frames together i.e. segmentation in temporal domain. Here similar frames mean frame related to same scene.
2. The above video segments are processed one by one. Two frames from a video segment are taken at a time - current frame f_c and previous (reference) frame f_p.
3. Rotation angle between current frame f_c and previous frame f_p is calculated by Planner Motion Estimation. Let's say angle is φ_m.
4. Rotate the image f_c by - φ_m to cancel the rotation.
5. Calculate Shift Estimation i.e. horizontal (X) and vertical (Y) shifts between current frame f_c and previous frame f_p.
6. Reconstruct the high resolution image (f_h) using current frame f_c , previous frame f_p , φ_m , X and Y.
   a. Map the co-ordinates of each pixel of the current frame f_c to the previous frame f_p using φ_m , X and Y.
   b. Using these samples, interpolate the pixel values on High resolution grid.
7. Compute the edge representation (L_{-1}) of the reference image f_p by SR method.
8. Apply sharpness enhancement on the edge representation (L_{-1}) by Unsharp Mask method.
9. Inject the sharpened edge representation in the high resolution image (f_h).
10. Repeat Steps 2-9 with $f_p = f_c$ and f_c = next frame in video.
11. Repeat Steps 2-10 for all video segments generated by histogram based segmentation.

3.2 Temporal video segmentation.

It is the first step towards Super resolution data generation of digital video sequences [9]. Its goal is to divide the video stream into a set of meaningful and manageable segments that are used as basic elements for SR algorithm. So to divide the given input video into meaningful segments we can choose histogram segmentation, which is having the advantage of reducing sensitivity to camera and object movements by comparing the histograms of successive images [9].

3.2.1 Histogram based video segmentation

In the histogram based video segmentation, differences of the number of pixels of each gray level of consecutive frames are calculated. Let's say histogram of a frame f_p is an n-dimensional vector $H_p(j)$, j=1,2,.. n, where n is the number of gray levels, $H_p(j)$ is the number of pixels from the frame f_p with gray level j and $H_c(j)$ is the number of pixels from the frame f_c with gray level j. If the absolute sum of histogram differences between two successive frames D is greater than a threshold T as given in [2], a cut is declared.

$$D = \sum_{j=1}^{n} |Hp(j) - Hc(j)|$$

3.3 Registration

The motion can be described as a function of three parameters - horizontal shift X and vertical shift Y, and a planar rotation angle φ. The frequency domain approach allows estimation of the horizontal and vertical shift and the (planar) rotation separately. Frequency domain methods mentioned in [10] are used to calculate Rotation motion estimation, Planner motion estimation.

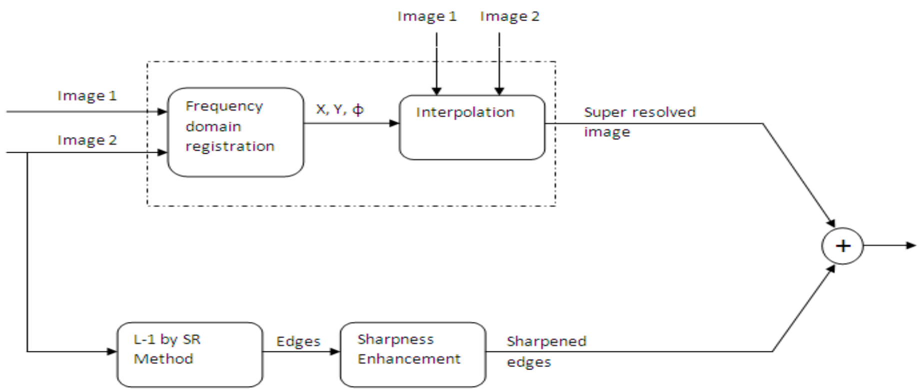

Figure 1. Block diagram of process of generating super-resolved image and injecting sharpened edges.

3.3.1 Super-resolved image reconstruction.

It involves generation of high resolution image by using multiple low resolution image and relation between them. Here relation means Rotation and Shift estimation between low resolution images. High resolution image is reconstructed from two consecutive Low resolution images in video sequence at a time.

Figure 1. gives the Block diagram of an algorithm.

3.4 L-1 generation.

For generating an HR video frame, the sharp edges must be identified in the LR and their sharpness must be restored in the HR image. The current work makes use of recent results related to Laplacian pyramid to identify edges. The Laplacian pyramid, like all other multiresolution representations, creates a hierarchy of subbands encoding edges of decreasing sharpness [13].
As shown in the figure 2 the Laplacian image is the result of the low pass filtered image subtracted from the original image.

Let G_0 be the original image and G_1 be the 1st level Gaussian pyramid image. G_1 is a result of applying a low-pass filter to G_0.

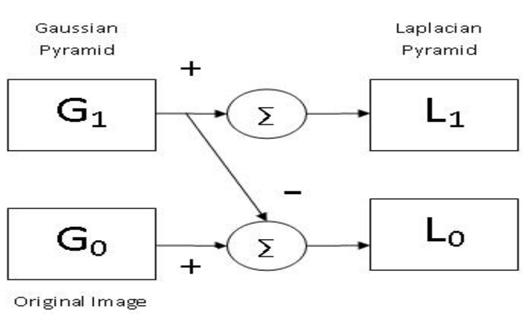

Figure 2. Shows Laplacian Generation Method

The error L_0 is then given by

$$L_0 = G_0 - G_1$$

Another method - Symmetric Residue Pyramids [4] was proposed as an extension to Burt Laplacian pyramids. This extension of the Laplacian pyramid exploits the redundancy present in the Laplacian pyramid to achieve better signal compaction and non redundant edge representation. The process of generating a *Li* given a *Gi* starts with an initial guess, which may even be a blank (zero) image. An iterative process is deployed to get one of the acceptable *Li*: (*exp* is the expand/interpolation operation, *ss* is subsampling and *lpf* is low-pass filtering)

Algorithm

1. $Li[0]$ = Initial guess (may even be 0 image)
2. $Gi[k] = exp(Gi{+}1) + Li[k]$ (usual pyramid reconstruction)
3. $Li[k+1] = Gi[k] - exp(ss(lpf(Gi[k])))$

The output $Li[k + 1]$ will contain the detected edges. Here L-1 is nothing but a zoomed edge representation.

3.4 Sharpness enhancement.

The visual appearance of an image may be significantly improved by emphasizing its high frequency contents to enhance the edge and detail information in it. The classic unsharp masking (UM) technique is often employed for this purpose.

The edges L_{-1} calculated in section 3.4 are sharpened by using unsharp masking (UM) technique. Let's say output of Unsharp masking is $\hat{L}$*-1*.

3.5 Injecting edge details in image.

Now sharpened edges $\hat{L}$-1 needs to be added to the zoomed image I'. As I' can be a color image, sharpened edges $\hat{L}$-1 can't be added directly to image. So first I' is converted to YCbCr color space from RGB color space. Y channel of YCbCr color space is used to add edges. After adding edges image is converted back to RGB color space.

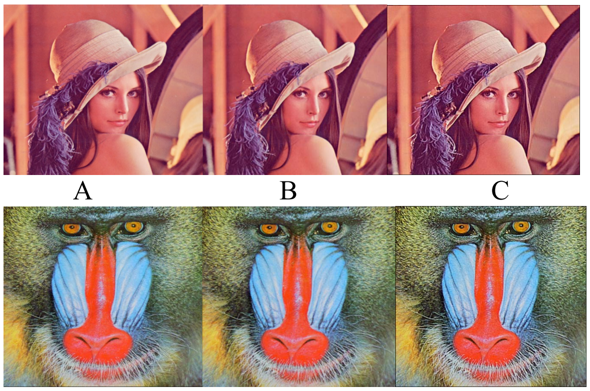

Figure 3. Shows the comparison of output of different methods with original image. A (original Lena image),B (Lena image by bicubic method), C (Lena imageby our method), D (Original Mandrill image) , E (Man-drill image by bicubic method), F (Mandrill image by our method).

4 Experimental Results.

Five Standard Images are taken for the experiment. These Standard Original images are down sampled. Then these images are subpixel shifted to create another set of images. So there will be 2 images for each original image. These down sampled images are used for super resolution.

Generated super-resolved images are compared with the original images (which are of same size as generated images) to check the error.

Also the output of our method is compared with images generated by bicubic interpolation method which is one of the most frequently used method for scaling images and video. Fig 3 shows the original image and image generated by bicubic interpolation and our method.

5. Quantitative Analysis.

5.1 Images

We used three metrics for checking the quality of generated images; RMSE (root mean squared error), PSNR (peak signal-to-noise ratio) and MSU Blurring Metric.

Different approaches exist for computing the RMSE and PSNR of a color image. The human eye is most sensitive to luma information so these metrics are computed for color images by converting the image to a color space that separates the intensity (luma) channel, such as YCbCr. The Y (luma), in YCbCr represents a weighted average of R, G, and B. G is given the most weight, again because the human eye perceives it most easily. With this consideration, above metrics are computed only on the luma channel. The results are tabulated in Table 1.

Table 1. Three metrics for checking the quality of super-resolved images.

Images	**Method**	**RMSE**	**PSNR**	**Blur index**
Lena	Bicubic	0.852	49.521	13.35
	Our Method	0.808	49.982	20.33
Man	Bicubic	0.8145	54.684	15.73
	Our Method	0.7903	54.946	22.96
Fruit	Bicubic	0.8346	49.701	12.42
	Our Method	0.7865	50.217	17.84
Plane	Bicubic	0.8683	49.357	15.2
	Our Method	0.8065	49.998	22.36
Monkey	Bicubic	0.7729	50.368	27.71
	Our Method	0.7535	50.588	46.52

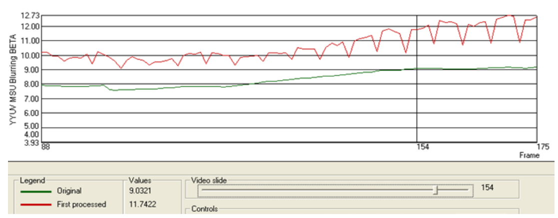

Figure 4. Shows the comparison of output of bicubic interpolation and our method for frames 88 to 175 of an indoor video. (Green line – Bicubic interpolation, Red line – Out Method)

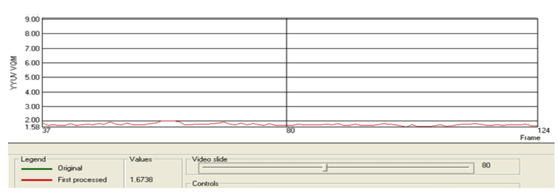

Figure 5. Visual Perception Quality Measurement for frames from 37 to 124. VQM value for frame no 80 is 1.67.

5.2 Video

We used MSU Video Quality Measurement Tool from Graphics & Media Lab, Moscow State University, Moscow, Russia for the quantitative analysis [9] of the generated SR videos compared to the input videos. We used two metrics for this comparison. They are the MSU Blurring Metric and DCT-based video quality metric (VQM).

5.2.1 MSU Blurring Metric - This metric allows you to compare sharpness of two images. If value of the metric for first picture is greater, than for second it means that first picture is sharper, than first.

Figure 4 shows that Blurring metric for frame no 64 of bicubic interpolation is around 6 and same for our method is around 8. i.e. video by our method is sharper than that of bicubic interpolation.

5.2.2 DCT-based video quality metric (VQM) -

Figure 5 shows the perception difference between the proposed method result and the original High Resolution Video which was taken as a reference video. Lesser values showed in results shows the less difference from the reference video in Visual Perception Quality.

6. Conclusion.

We have designed and implemented a system for super resolution of a color video. This system also tries to improve sharpness of super resolved video.

Experimental results show that output of our method is better than bicubic interpolation method. Some time output video is sharper than original image. Quantitative analysis shows that values for RMSE and PSNR for our method are better than that of bicubic interpolation. Results for blur index (which measures sharpness) for our methods are very good for some images like standard monkey image.

References

[1] Malay Kumar Nema, Subrata Rakshit, and Subhasis Chaudhuri, “Edge Model Based High Resolution Image Generation”,ICVGIP’06, Springer 2006.

[2] S. C. Park, M. K. Park, and M. G. Kang, “Super-resolution image reconstruction: A technical review”, IEEE Signal Processing Mag., vol. 20, pp. 21-36, May 2003.

[3] Subhasis Chaudhuri, “Super-Resolution Imaging”, Kluwer Academic publishers, 2001

[4] Subrata Rakshit and Malay k. Nema, “Symmetric residue pyramids: An extension to burt laplacian pyramids,” IEEE ICASSP, Hong Kong. 2003

[5] Nema, M.K., Rakshit, S., “ Edge-model based representation of laplacian subbands.”, Seventh Asian Conf. on Computer Vision (ACCV 7), Hyderabad, India,. Jan. 2006

[6] J.D. van Ouwerkerk, “Image super-resolution survey”, Image and Vision Computing 24, 2006

[7] C. Staelin, D. Greig, M. Fischer, R. Maurer, “Neural network image scaling using spatial errors,” HP Laboratories Israel, October 2003.

[8] Madhusudhan, T.; Pais, Alwyn Roshan, "Generation of Super-Resolution Video from Low Resolution Video Sequences: A Novel Approach," Conference on Computational Intelligence and Multimedia Applications, 2007. International Conference on , vol.3, no., pp.225-232, 13-15 Dec. 2007.

[9] Irena Koprinska, Sergio Carrato, “Temporal video segmentation: A survey,” Signal processing : Image Communication 16 (2001) 477-500.

[10] Patrick Vandewalle, Sabine Susstrunk, and Martin Vetterli, “A Frequency Domain Approach to Registration of Aliased Images with Application to Super-resolution,” EURASIP Journal on Applied Signal Processing, 2006.

[11] MSU Video Quality Measurement Tool, http : //compression.ru/video/quality measure/index en.html, Feb 2008