Generation of Super-Resolution Video from Low Resolution Video Sequences: A Novel Approach

T.Madhusudhan[1] and Alwyn Roshan Pais[2]
[1] Student,Dept.of CSE, NITK Surathkal, INDIA. tmadhusudhan@gmail.com
[2] Faculty, Dept. of CSE, NITK Surathkal, INDIA. alwyn.pais@gmail.com

Abstract

The term "super-resolution" refers to the process of obtaining higher-resolution images from several lower-resolution ones, i.e. resolution enhancement. The quality improvement is caused by sub-pixel shifted from each other between images. Each low resolution image contains new information about the scene and super-resolution aims at combining these to give a higher resolution image. Super-resolution allows us in overcoming the limitations of the imaging system (resolving limit of the sensors) without the need for additional hardware.

This paper deals with high resolution video reconstruction from low resolution video. An algorithm for enhancing the resolution of video through histogram based segmentation and frequency domain registration is proposed. Segmentation of the video is done using Histogram method and frequency domain approach is used for registration of images. Bi-cubic interpolation is applied to this video to generate the super-resolution (SR) video. Based on the requirements additional brightness and temporal resolution can be added to the SR video. It is tested on indoor/outdoor real video data, demonstrating the feasibility of the approach, and then results are presented. Qualitative analysis is presented to analyze the quality of reconstructed Super resolution video comparing to input Low Resolution video. The results of this work can be used for object tracking and identification.

Keywords: SR(Super Resolution); HR (High Resolution); LR(Low Resolution);Temporal Video Segmentation.

1. Introduction

Over the past ten years, digital cameras have gone through a fast evolution towards extremely compact models, containing sensors with a steadily increasing number of pixels. From about 0.3 megapixels (million pixels) in 1993, the number of pixels on the CCD or CMOS sensor in a digital camera has increased to 39 megapixels in some of the latest professional models. This pixel count has become the major selling argument for the different camera manufacturers.

The number of pixels in a digital image is also often referred to as the resolution of an image. The ever-increasing demand for more pixels, or higher resolution, in combination with the availability of more and more computational power, has generated a large interest in super-resolution imaging. The goal in super-resolution imaging is to take multiple 'low' resolution images of the same scene, and combine them to generate a 'higher' resolution image. In this way, a photographer could for example take a series of four images using a four megapixel camera, and combine them to obtain an image as if it would be taken with a sixteen megapixel camera. And who would not be interested in such a feature?

In practice, such a combination of information from multiple images is not trivial. There are two main problems that need to be solved in a super-resolution algorithm. First, all the input images need to be correctly aligned with each other on a common grid. Next, an accurate, sharp image has to be reconstructed from the gathered information. If one of these two steps is not well done, the resulting image is not good, and no gain in resolution is obtained.

0-7695-3050-8/07 $25.00 © 2007 IEEE
DOI 10.1109/ICCIMA.2007.324

In this paper, we consider the problem of spatial resolution enhancement in video sequences. Since the typical resolution of display devices is often much higher than that of video, particularly in the case of video streamed over the internet, there is considerable interest in being able to exploit the high resolution of modern display devices when rendering video. The problem of spatial resolution enhancement of video sequences has been an area of active research. This considers the problem of HR video generation using a sequence of LR images of a scene. In the traditional single image restoration problem only a single input image is available for HR video generation. The task of obtaining a super resolved image from an under sampled and degraded image sequence can take advantage of the additional spatio-temporal data available in the image sequence [1]. In particular, camera and scene motion lead to frames in the video sequence containing similar, but not identical information [2]. This additional information content enables reconstruction of a super resolved image with a wider bandwidth than that of any of the individual LR frames [1]. The presented Super-resolution video generation method can be applied to different application domains, such as consumer digital video cameras, video surveillance and satellite imaging, etc.

2. Methodology

Algorithm 3.1: Super resolution video generation. A high resolution video sequence of frames $f_{HR,m}$ $m = (0,1,\ldots\ldots,M-1)$ is reconstructed from a sequence of low resolution video frames $f_{LR,m}$ $m = (0,1,\ldots\ldots,M-1)$ by using frequency domain registration and histogram based segmentation.

1. Histogram based segmentation was used to segment the video in temporal domain. This will give frames related to the same scene.
2. The individual video segments are converted to tiff images. Then these LR images are taken two (in sequence) at a time for motion estimation.
3. Rotation estimation: the rotation angle between next frame and reference frame in frame sequence is estimated using planar motion estimation.
4. Shift estimation: the horizontal and vertical shifts between next frame and reference frame in frame sequence are estimated using planar motion estimation.
5. Image reconstruction: a high resolution frame f_{HR} is reconstructed from the set of two registered frames.
    (a) For a frame in the frame sequence compute the coordinates of its pixels in the coordinates of its previous frame using estimated registered parameters.
    (b) From these known samples, interpolate the values on a uniform high resolution grid using bi-cubic interpolation to get super resolved frame.
6. All these super resolved images are combined to form super resolution video segments. In a similar manner the generated SR video segments are combined to form SR video corresponding to input LR video.
7. Brightness of the video can be improved by applying light improvement technique to each frame of total video sequence.
8. Temporal resolution of generated high resolution video is improved by using intermediate frame generation technique.

2.1 Temporal video segmentation

This is the first step towards Super resolution data generation of digital video sequences [5]. Its goal is to divide the video stream into a set of meaningful and manageable segments (shots) that are used as basic elements for SR algorithm. A shot is defined as an unbroken sequence of frames taken from one camera. There are two basic types of shot transitions: abrupt and gradual [5]. When two images are sufficiently dissimilar, there may be a cut. Gradual transitions are found by using cumulative difference measures and more sophisticated threshold schemes. Based on the metrics used to detect the difference between successive frames, the algorithms can be divided broadly into three categories: pixel based, block based and histogram comparisons [5]. So to divide the given input video to meaningful segments the histogram segmentation has been chosen, which is having the advantage of reducing sensitivity to camera and object movements by comparing the histograms of successive images [5].

2.1.1 Histogram based video segmentation

In the Global histogram comparison a gray level (color) histogram of a frame i is an n-dimensional vector $H_i(j)$, j=1,2,.. n, where n is the number of gray levels (colors) and $H_i(j)$ is the number of pixels from the frame i with gray level (color) j. A cut is declared if the absolute sum of histogram differences between two successive frames D(i, i+1) is greater than a threshold T as given in [5].

$$D(i,i+1) = \sum_{i=1}^{n} \left| H_i(j) - H_{i+1}(j) \right|$$

where $H_i(j)$ is the histogram value for the gray level j in the frame i, j is the gray value and n is the total number of gray levels.

2.2 Registration of images for super resolution

In this section, we present a planar motion estimation method for the registration images taken in sequence from a single obtained video segment from histogram based segmentation. As we explicitly use the Fourier transform, our method is limited to band-limited signals described in the Fourier basis. Although this limits the applicability of such a method, it is a very reasonable assumption in practice. The optical system of a digital camera typically acts as a low-pass filter, and attenuates or blocks all high frequencies. The captured image will therefore be an essentially band-limited signal. Our method computes the planar shift and rotation parameters between a pair of images. The difference between the shift and rotation estimation is shown in Section 2.2.1, and frequency domain rotation and shift estimation methods for images are described in Section 2.2.2 and Section 2.2.3, respectively.

2.2.1 Planar motion estimation

Fourier based image registration methods [6] only allow global motion in a plane parallel to the image plane. In such a case, the motion between two images can be described as a function of three parameters that are all continuous variables: horizontal and vertical shifts $x_{1,h}$ and $x_{1,v}$ and a planar rotation angle θ_1. A frequency domain approach allows us to estimate the horizontal and vertical shift and the (planar) rotation separately. Assume we have a continuous two-dimensional reference signal $f_0(x)$ and its shifted and rotated version $f1(x)$:

$$f_1(\boldsymbol{x}) = f_0(\mathbf{R}(\boldsymbol{x} + \boldsymbol{x}_1)),$$

$$\text{with } \boldsymbol{x} = \begin{pmatrix} x_h \\ x_v \end{pmatrix}, \boldsymbol{x}_1 = \begin{pmatrix} x_{1,h} \\ x_{1,v} \end{pmatrix}, \mathbf{R} = \begin{pmatrix} \cos\theta_1 & -\sin\theta_1 \\ \sin\theta_1 & \cos\theta_1 \end{pmatrix}$$

This can be expressed in Fourier domain as

$$\begin{aligned} F_1(\boldsymbol{u}) &= \iint_{\boldsymbol{x}} f_1(\boldsymbol{x}) e^{-j2\pi\mathbf{u}^T\boldsymbol{x}} d\boldsymbol{x} \\ &= \iint_{\boldsymbol{x}} f_0(\mathbf{R}(\boldsymbol{x}+\boldsymbol{x}_1)) e^{-j2\pi\mathbf{u}^T\boldsymbol{x}} d\boldsymbol{x} \\ &= e^{j2\pi\mathbf{u}^T\boldsymbol{x}_1} \iint_{\boldsymbol{x}'} f_0(\mathbf{R}\boldsymbol{x}') e^{-j2\pi\mathbf{u}^T\boldsymbol{x}'} d\boldsymbol{x}', \end{aligned}$$

with $F_1(u)$ the two-dimensional Fourier transform of $f_1(x)$ and the coordinate transformation $x' = x + x_1$. After another transformation $x'' = Rx'$, the relation between the amplitudes of the Fourier transforms can be computed as

$$\begin{aligned} |F_1(\boldsymbol{u})| &= \left| e^{j2\pi\mathbf{u}^T\boldsymbol{x}_1} \iint_{\boldsymbol{x}'} f_0(\mathbf{R}\boldsymbol{x}') e^{-j2\pi\mathbf{u}^T\boldsymbol{x}'} d\boldsymbol{x}' \right| \\ &= \left| \iint_{\boldsymbol{x}'} f_0(\mathbf{R}\boldsymbol{x}') e^{-j2\pi\mathbf{u}^T\boldsymbol{x}'} d\boldsymbol{x}' \right| \\ &= \left| \iint_{\boldsymbol{x}''} f_0(\boldsymbol{x}'') e^{-j2\pi\mathbf{u}^T(\mathbf{R}^T\boldsymbol{x}'')} d\boldsymbol{x}'' \right| \\ &= \left| \iint_{\boldsymbol{x}''} f_0(\boldsymbol{x}'') e^{-j2\pi(\mathbf{R}\boldsymbol{u})^T\boldsymbol{x}''} d\boldsymbol{x}'' \right| \\ &= |F_0(\mathbf{R}\boldsymbol{u})| . \end{aligned}$$

We can see that $|F_1(u)|$ is a rotated version of $|F_0(u)|$ over the same angle θ_1 as the spatial domain rotation (see Figure 3). $|F_1(u)|$ and $|F_0(u)|$ do not depend on the shift values x_1, because the spatial domain shifts only affect the phase values of the Fourier transforms. Therefore we can first estimate the rotation angle θ_1 from the amplitudes of the Fourier transforms $|F_0(u)|$ and $|F_1(u)|$. After compensation for the rotation, the shift x_1 can be computed from the phase difference between $|F_0(u)|$ and $|F_1(u)|$.

2.2.2. Rotation estimation

The rotation angle between $|F_0(u)|$ and $|F_1(u)|$ can be computed as the angle θ_1 for which the Fourier transform of the reference image $|F_0(u)|$ and the rotated Fourier transform of the image to be registered $|F_1(Ru)|$ have maximum correlation.

2.2.3 Shift estimation

A shift of the image parallel to the image plane can be expressed in Fourier domain as a linear phase shift:

$$F_1(u) = \iint_{x} f_1(x) e^{-j2\pi u^T x} dx = \iint_{x} f_0(x + x_1) e^{-j2\pi u^T x} dx$$
$$= e^{j2\pi u^T x_1} \iint_{x'} f_0(x') e^{-j2\pi u^T x'} dx' = e^{j2\pi u^T x_1} F_0(u).$$

It is well known that the shift parameters x_1 can thus be computed as the slope of the phase difference $\angle(F_1(u)/F_0(u))$. To make the solution less sensitive to noise, we fit a plane through the phase differences using a least squares method.

2.3. SR Image Reconstruction

The second main sub problem related to SR video reconstruction involves reconstructing HR images from under-sampled LR images in each video segment using an optimal interpolation Algorithm. This process includes image interpolation that has been used to increase the size of a single image. Although this field has been extensively studied [1,2], the quality of an image magnified from an aliased LR image is inherently limited. That is, single image interpolation cannot recover the high-frequency components lost or degraded during the LR sampling process. For this reason, image interpolation methods are not considered as SR techniques [2]. To achieve further improvements in this field, the next step requires the utilization of multiple data sets in which additional data constraints from several observations of the same scene can be used. The fusion of information from various observations of the same scene allows us SR reconstruction of the scene. Cubic interpolation has been used because of its lower computational complexity [7]. Based on the input image samples, the image values are interpolated on a regular high resolution grid.

The selection of number of input images depends on the parameter called registration accuracy. The optimal number of images used at a time to reconstruct the high resolution image is two. High resolution image is reconstructed from two consecutive Low resolution images in video sequence at a time. Coordinates of the pixels in the second frame are estimated using the registration parameters of the first frame. The high resolution image is obtained by interpolating these known samples to the high resolution grid.

2.4 Lightening the video

Lightening of the video is required for most of the surveillance applications [8]. In the proposed method the total video is first divided into RGB Images. Then every image is filled with light by multiplying each pixel RGB values with an integer lightening factor value which varies between 0 and 3. If any pixel RGB value exceeds more than max value 255 then the RGB value of that pixel is normalized by dividing with 255. Each frame in the video segment can be made still darker by using lightening factor value in between 0 and 1 and brighter by using lightening factor between 1 & 3.

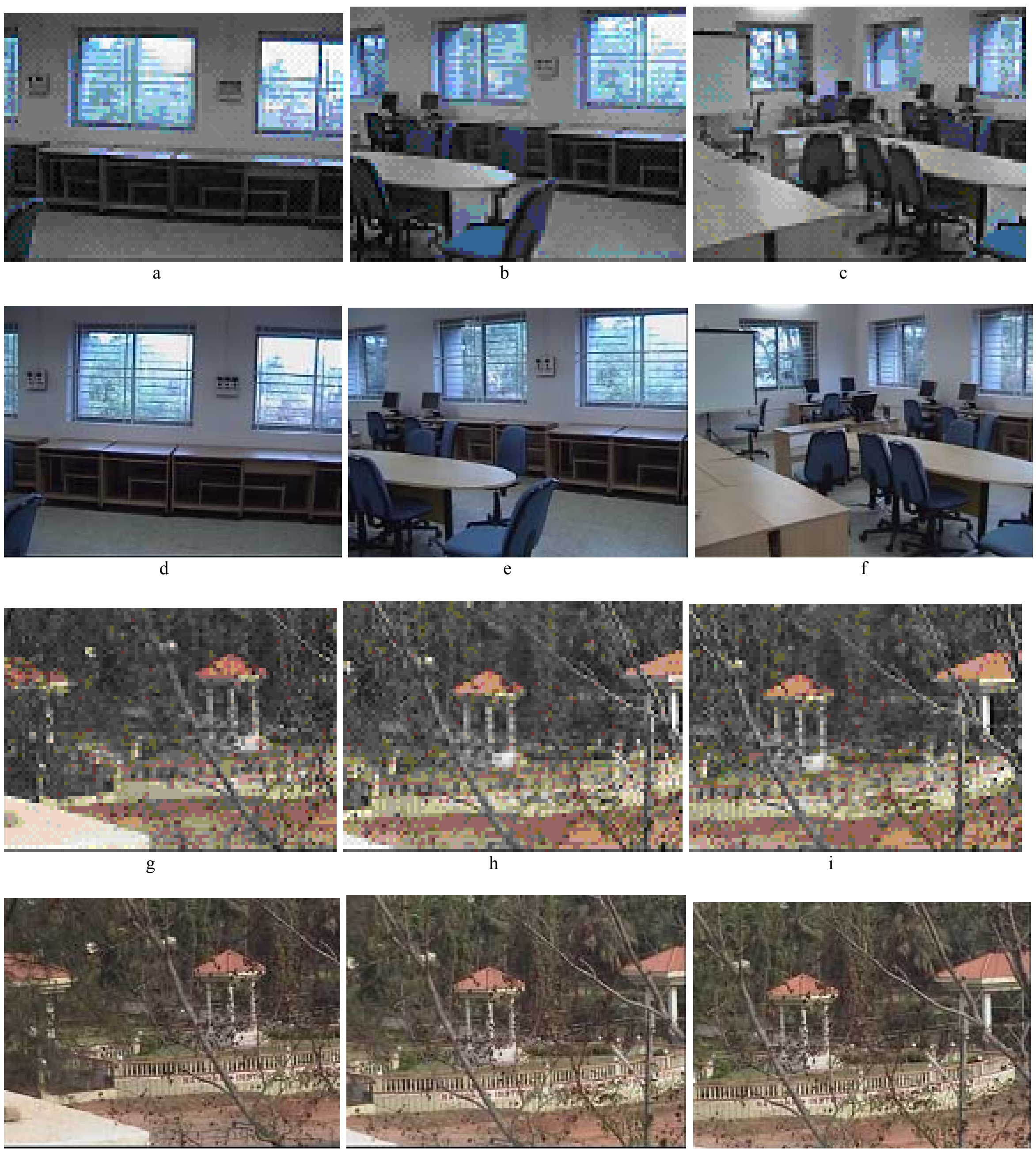
2.5 Generation of Intermediate Frame

Temporal Resolution needs to be improved for better clarity of the video. This can be achieved by generating intermediate frames. Global motion estimation between consecutive frames is calculated and an intermediate frame with half of the values of translation and rotation parameters is generated. This will give a video with improved temporal resolution.

3. Experimental Results & Qualitative Analysis

Two different kinds of video's i.e. indoor & outdoor videos were used to evaluate the performance of our algorithm. Histogram based segmentation was used to segment the video in temporal domain. This will give frames related to the same scene. The individual video segments are converted to tiff images. Then these LR images are taken two (in sequence) at a time for motion estimation. Then cubic interpolation is applied to get the super-resolved image. All these super resolved images are combined to form super resolution video segments. In a similar manner the generated SR video segments are combined to form SR video corresponding to input LR video. Brightness of the video can be modified by applying the light improvement algorithm.

The sample results are provided for two types of color videos: one for indoor environment and other one for the out door environment. It is observed that the proposed methodology provides good results for video frames wherein method [6] is studied only for images.



j k l

Fig. 1. Details of frames of video used for the experiments: a (1st frame), b (300th frame), and c (600th frame) are the original magnified low resolution color video frames in indoor environment and d, e, and f are their corresponding high resolution frames of the output SR video and g (1st frame), h (75th frame), and i (150th frame) are the original magnified low resolution color video frames in outdoor environment and j,k,l are their corresponding SR frames of the output video.

a b c

Fig. 2. Details of Generation of Intermediate Frames: a and b are consecutive frames in video sequence with motion parameters values i.e. translation x, translation y and rotation are 1,1, and 2 respectively and c is the intermediate frame with half of the calculated motion parameter values.

Qualitative Analysis of the obtained results:

We used MSU Video Quality Measurement Tool from Graphics & Media Lab, Moscow State University, Moscow, Russia for the qualitative analysis [9] of the generated SR videos compared to the input videos. We used two metrics for this comparison, they are Brightness Flicking Metric (BFM) and the Blurring measure. The obtained Qualitative Analysis is as follows:

Brightness Flicking Metric:

BFM metric is made to measure flicking quantity between neighboring frames of the sequence. Average brightness value is calculated for each frame. Metric's value is modulus of difference between average brightness values of previous and current frames.

Figure 3 and Figure 4 shows the graph drawn using MSU Video Quality Measurement Tool [9] for calculating Brightness Flicking Metric(BFM) of outdoor video and indoor video having the frames specified in Figure 1. It is observed that the BFM value of generated SR video is lesser compared to input LR video, which proves that Brightness Flickering of the videos can be reduced by applying our proposed SR video generation algorithm.

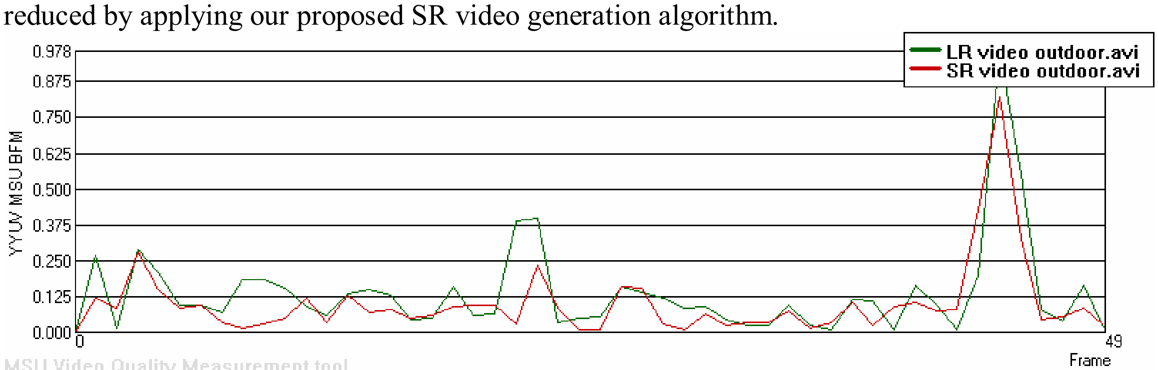


Figure 3: Comparison of Brightness Flicking Metric measure for the frames 0 to 49 of outdoor video sequences(LR Video and SR Video).

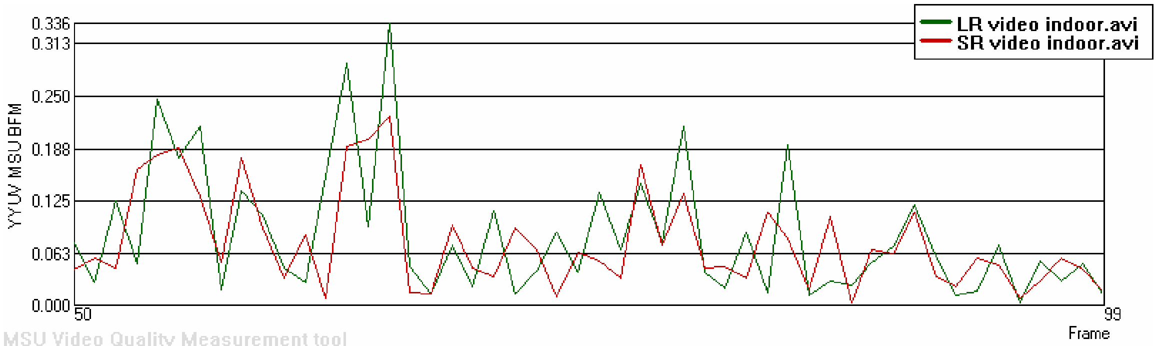


Figure 4: Comparison of Brightness Flicking Metric measure for the frames 50 to 99 of indoor video sequences(LR Video and SR Video).

4. Conclusion

A novel method for Super resolution video reconstruction using histogram based segmentation and frequency domain registration is presented in the paper. The proposed technique gives good results for global motion models and has robustness with respect to video changes.

5. References

[1] Sean Borman, Robert L.Stevenson, "Super-Resolution from image sequences - A Review," Dept of Electrical Engineering, University of Notre Darne,USA.

[2] S.C. Park, M. K. Park, and M. G. Kang, "Super-Resolution image reconstruction: A Technical Overview", IEEE Sig. Proc. Magazine No. 3, pp.21-36, May 2003.

[3] S. Chaudhuri, Ed., *Super- Resolution Imaging*. Norwell, MA: Kluwer, 2001.

[4] Marcelo Victor, Wiist Zibetti and Joceli Mayer, "Simultaneous Super-Resolution for Video Sequences," Digital Signal processing Research laboratory, Department of Electrical Engineering, Federal University of Santa Catarina, Brazil.

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

[6] 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.

[7]. Thomas M Lehmann, "Survey: Interpolation Methods in Medical Image Processing", *IEEE Transactions on MedicalImaging,* vol 18, no 11, November 1999.

[8] T. Madhu sudhan, Alwyn Roshan Pais, " Super – Resolution Video Generation Using Histogram Based Segmentation and Frequency Domain registration " , International Conference on Information Technology, Haldia, India, Mar19-21,2007.

[9]. http://www.compression.ru/video/quality_measure/info_en.html#start