Wavelet Packet Based Video Super-Resolution Using Radial Basis Function Neural Network

PANN EI SAN[1], THEINT THEINT SOE[2], SAN SAN NAING[3]
[1] *Department of Electronic Engineering, Technological University (Thanlyin), Myanmar*
[2] *Faculty of Electronic Engineering, University of Technology (Yatanarpon Cyber City), Myanmar*
[3] *Department of Electronic Engineering, Technological University (Kyaukse), Myanmar*

Abstract- ***Video super-resolution uses a set of successive low-resolution (LR) images in a video sequence to reconstruct a high-quality and high-resolution (HR) image. In this paper, a video super-resolution image reconstruction method using local full search (LFS) motion estimation algorithm, wavelet packet transform (WPT) and radial basis function (RBF) neural network is proposed. LFS motion estimation algorithm is used to collect motion-trace volume which can eliminate deep object motion in the LR images. The HR input image is decomposed by using two-level two-dimension (2D) WPT. The resulted sub-images are employed to train the networks. RBF neural network is applied to each sub-band to learn the relationship between LR and HR images. The super-resolved (SR) image is finally produced by using the inverse WPT. Experiments are mainly focused on image sequences from real videos with different kind of motions. The objective and subjective quality assessments are carried out and compared with the conventional super-resolution methods. The experimental results give that the proposed method outperforms the conventional and recent super-resolution methods for real video sequences. It is well noticed that the proposed method has the best results in terms of both the objective quality and visual quality of the super-resolved image.***

Indexed Terms- ***Motion estimation, Radial basis function neural network, Super-resolution, Wavelet packet transform***

I. INTRODUCTION

Image/video super-resolution is the topic of great interest. Images with high-resolution are desired and more often required in most electronic imaging applications. Computers and mobile phones are essential devices for daily life. Because of multi-function and compact design, embedded cameras do not have enough quality to capture videos or images. Moreover, the current sensor technology has almost reached the limited level. One promising approach to cope up with strong demands of HR images is super-resolution (SR) image reconstruction method.

SR method is an image processing method that attempts to generate high quality and high-resolution (HR) images from one or more low-resolution images of a scene. There are two classes of super-resolution image reconstruction based on the number of observed images. Specifically, they are single-frame super-resolution (SFSR) and multi-frame super-resolution (MFSR). Video super-resolution is the same with MFSR which uses several low-resolution images to get a better quality HR image. . But, it requires a reliable, accurate and good-performed image registration or motion estimation method. There are two types of super-resolution: reconstruction-based and learning-based super-resolution.

Most of the super-resolution techniques are reconstruction-based [1]. These techniques can operate directly with the image pixel intensities and can super-resolve any image sequences provided between observation models. This reconstruction-based approach is generally a severely ill-posed problem because of the insufficient number of low-resolution images, ill-conditioned registration and unknown blurring operators. The performance of this reconstruction-based super-resolution algorithm degrades rapidly when the desired magnification factor is large or the available input

images are limited. In these cases, the results may be overly smooth; lacking important high-frequency details. However, the reconstruction-based SR techniques do not require training data set and therefore they do not depend on the observed images.

The iterative back projection (IBP) method is very easy to understand in applications. The IBP method provides a useful framework for solving the super-resolution problem by utilizing a mechanism for constraining the super-resolution restoration to conform to the observed data [2]. The iterative process comprises two steps: simulation of the observed images and back projection of the error to correct the estimate of the original scene. Irani and Peleg [3, 4] formulated the iterative back-projection SR reconstruction approach that is similar to the back-projection used in [5].

The concept of Projection onto convex sets (POCS) applied to the problem of super-resolution was introduced in [6]. Stark proposed the general technique for applying POCS in the field of image restoration. According to the method of POCS for super-resolution reconstruction, incorporating a priori information into the solution can be interpreted as restricting the solution to be a member of a closed convex set which characterize the desirable properties, such as fidelity to data, smoothness, sharpness etc., to be consistent in the final solution.

In recent years, a lot of interest has grown towards learning-based techniques which were initially introduced [7-9]. This technique generates a high-resolution image from one or more low-resolution images by learning from a collection of training images or strong image priors learned from data before. These training data sets may be from same or different scenes. This technique solves the problem differently by learning corresponding relationship between the LR images and the known HR images. Then the learned information is later utilized for the reconstruction of a HR image from LR images. This learning-based SR technique requires training data sets therefore it depends on the accuracy of matching between the input low-resolution frame and the training samples but have high performance when the magnification factor is large.

Learning-based super-resolution is employed in this work. In video super-resolution, a HR image can be generated from a set of successive LR images, instead of from just one LR frame. There are three main steps to achieve video super-resolution, namely, registration (motion estimation), interpolation, and restoration, which are implemented either sequentially or simultaneously [10]. Traditional sequential techniques largely depend on the availability of accurate motion estimation.

This paper is organized as follows. In Section II, motion estimation method is presented. Section III represents wavelet packet transform (WPT). Section IV is about radial basis function neural network. Section V describes the implementation procedure of the proposed method. Section VI gives the results and discussion. Section VII concludes this paper.

II. MOTION ESTIMATION METHOD

Most of the SR image reconstruction methods proposed in literatures consists of three stages: registration, interpolation and restoration. These steps can be implemented individually or simultaneously according to the reconstruction methods adopted. The estimation of motion information is referred to as registration; it is extensively studied in various fields of image processing.

Motion estimation plays a central role in the context of multi-frame super-resolution restoration. It is also an important tool in other imaging applications such as image stabilization, video compression using motion compensation, computer vision, restoration and motion detection. Motion estimation has a vast field and many approaches to the problem may be found in literatures. In [11], motion estimation is a very difficult problem due to its ill-posedness, the aperture problem and the presence of covered and uncovered regions. Frame-to-frame motions provide different views of the scenes or objects of interest. Frame motions may consist of a global motion from camera movement and local motion of object movements within the scenes.

Image registration algorithms can be categorized into two groups: spatial domain image registration and frequency domain image registration algorithms.

Keren algorithm is a very efficient and straightforward spatial domain image registration method [2]. This algorithm uses the Taylor series expansion of the spatial transformation. Vandewalle algorithm is a frequency domain registration technique [12], which precisely aligns a set of aliased images, based on their low-frequency components, to avoid aliasing effects and high-frequency noise.

III. WAVELET PACKET TRANSFORM

Wavelet is currently applied in many fields such as signal processing, image processing, computer vision, data compression and so on. The purpose of video super-resolution image reconstruction is to obtain a priori knowledge or lost information in generating HR image from multiple low-resolution images. In this case, wavelet is a key because of the characteristics which includes multi-resolution, directivity, anisotropy, relevance, low entropy and selection diversity of wavelet base [13]. The wavelet packet analysis [14] based on wavelet analysis can do multi-layers division in high frequency which the wavelet analysis cannot do.

The wavelet packet transform (WPT) is a generalization of the wavelet transform that provides good spectral and temporal resolutions in arbitrary regions of the time-frequency domain. In the case of 2D wavelet packet decomposition, the image can be divided into multi-layer classification to further decomposition which the wavelet analysis can't work in the high frequency.

In order to get decomposition at level *l*, decompose the approximations A_i^{l-1} and details $D_{i,h}^{l-1}$, $D_{i,v}^{l-1}$, $D_{i,d}^{l-1}$ as follows proposed by Perlibakas [15]:

$$A_i^{l-1} \rightarrow \left\{ A_{4i}^{l};\ D_{4i,h}^{l};\ D_{4i,v}^{l};\ D_{4i,\mathrm{d}}^{l} \right\},\ l > 0 \quad (1)$$

$$D_{i,h}^{l-1} \rightarrow \left\{ A_{4i+1}^{l};\ D_{4i+1,h}^{l};\ D_{4i+1,v}^{l};\ D_{4i+1,\mathrm{d}}^{l} \right\},\ l > 1 \quad (2)$$

$$D_{i,\mathrm{v}}^{l-1} \rightarrow \left\{ A_{4i+2}^{l};\ D_{4i+2,h}^{l};\ D_{4i+2,v}^{l};\ D_{4i+2,\mathrm{d}}^{l} \right\},\ l > 1 \quad (3)$$

$$D_{i,\mathrm{d}}^{l-1} \rightarrow \left\{ A_{4i+3}^{l};\ D_{4i+3,h}^{l};\ D_{4i+3,v}^{l};\ D_{4i+3,\mathrm{d}}^{l} \right\},\ l > 1 \quad (4)$$

where $i = 0, \ldots, (4^{(l-1)} - 1)$.

By using the above equations, we can decompose the two-dimensional image A_0^0 (level *l=0*) into approximation A_0^1, horizontal details D_{0h}^1, vertical details D_{0v}^1 and diagonal details D_{0d}^1 at level 1. Figure 1 is the wavelet packet tree for two level of decomposition.

The concept of wavelet packet is the same as wavelet. However, in wavelet packet analysis, the details as well as the approximation are split. Wavelet packet decomposition of an image is a useful technique for building compact and meaningful feature vectors. The high frequency components (Detailed information) are very useful in super-resolution image reconstruction. Therefore, in this research, wavelet packet transform is utilized to obtain some lost information in high frequency components. To sum up, this short section provides the concepts of wavelet packets.

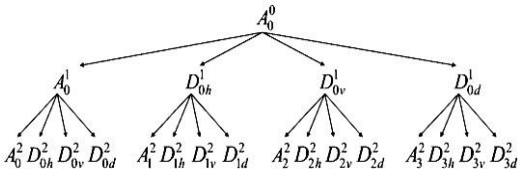

Figure 1. Two level wavelet packet decomposition tree

IV. RADIAL BASIS FUNCTION NEURAL NETWORK

Artificial Neural Network (ANN), usually called Neural Network (NN), is a mathematical model or computational model that is inspired by the structure or functional aspects of biological nervous systems such as human brain.

Radial basis function (RBF) neural network has become very popular, and is a strong rival to the multi-layer perceptron (MLP) neural network. It is a local approximation three-layered feed-forward neural network using a supervised training algorithm. It is typically configured with a single hidden layer of units whose activation function is selected from a class of functions called basis function. It is better than the traditional BP neural network in the approximation capability, classification ability and learning speed.

V. PROPOSED METHOD

In this section, the implementation procedure of the proposed video super-resolution method is represented. The required methodologies to implement the proposed SR framework such as motion estimation method, wavelet packet theory and RBF neural network have been explained in short. The whole process is implemented in MATLAB environment.

A. Local Full Search Block Matching Motion Estimation Method

In this section, Local Full Search Block Matching motion estimation method used in this paper is discussed in detail. This method has less computational complexity, fast processing time, high accuracy and compatibility with the proposed video super-resolution method.

Initially, motion-trace volumes are collected from the input low-resolution image sequence in order to get the input vector array for neural network. The purpose of motion-trace volumes is to track the unfathomable object motion in the LR frames to be removed. Choosing an optimal number of frames to be used for video super-resolution is also a very important part in motion estimation as it can lead to high computational cost because of unnecessary computations or low performance because of less information.

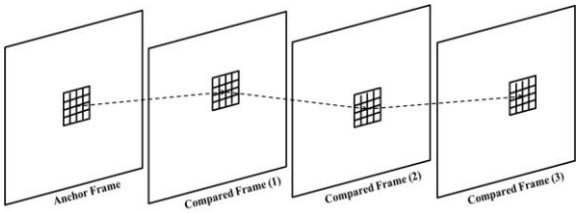

Figure 2. 4×4×4 motion-trace volume collection

In this work, four LR frames are chosen from a given video sequence. The first frame is set as Anchor Frame, and the rests are Compared Frames, as can be seen in figure 2. The Anchor Frame is divided into "n x n" blocks. For each block of Anchor Frame, a search window "SW" is set on every Compared Frame as shown in figure 3.

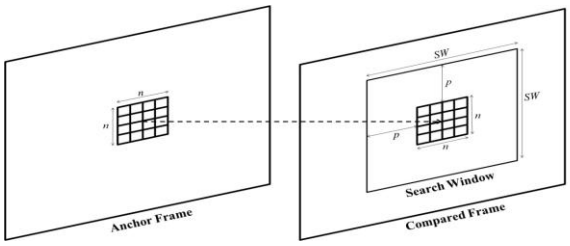

Figure 3. Illustration of setting a Search Range on a compared frame for each block of anchor frame

The value of "p" is the pixel space of the block, which is predetermined to search within the assumed best area for finding the best matching block. The search area is constrained up to "p" pixels on all four sides of the corresponding macro block in the frame to find the best match. The value of "p" is defined as the search parameter. The block size "n x n" is set to be "4×4" and the search parameter "p" is set to be "4" in order to get a rational block size and a fair search range. The searching starts from the left uppermost pixel of the search window, overlapping each block and moving pixel by pixel.

For each block of the Anchor Frame, the current block itself along with the best matched blocks from the Compared Frames is taken as a"4×4×4" motion-trace volume. The same procedure is applied for every block of the Anchor Frame and motion-trace volumes are collected. Finally, each motion-trace volume is converted into a column vector and put into an array which will be served as the input for training and interpolation.

B. Training Process

The block diagram of the training process is shown in figure 4.

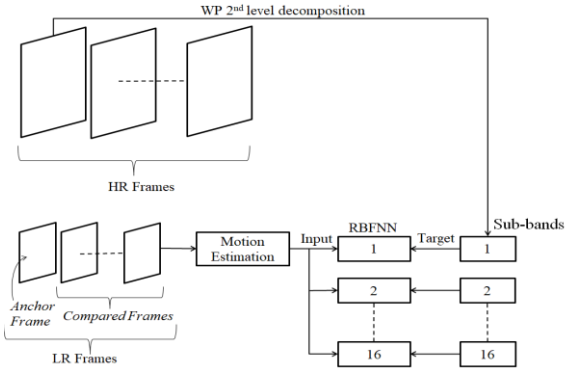

Figure 4: Training process of the proposed method

The steps of the training process are as follow:

1. HR frames are down-sampled to get LR Frames. The first LR frame is taken as the anchor frame, and the rests are used as compared frames.
2. Motion estimation method described above is employed. Let the image size be M x N and the block size, n x n. The number of blocks b can be calculated as follow:

$$b = \frac{M}{n} \times \frac{N}{n} \quad (5)$$

The size of search window is denoted as SW. Then, within a search widow, the minimum sum of absolute difference (SAD) value between anchor frame and compared frames is calculated by using SAD equation. The number of search blocks *s* within a search window can be calculated as follow:

$$s = (2p+1) \times (2p+1) \quad (6)$$

where p is the search parameter. The resultant motion vector array is dedicated as input for the training.

3. Let the size of the first HR frame be 2Mx2N and it is decomposed by using two-level two-dimensional wavelet packet transform to obtain 16 sub-bands. The resulted sub-bands are used as target of the network.
4. Finally, the input array is trained for each target and 16 networks are collected. These networks are used as the sub-band estimators to simulate in the interpolation process.

C. Interpolation Process

The block diagram of the interpolation process is shown in figure 5.

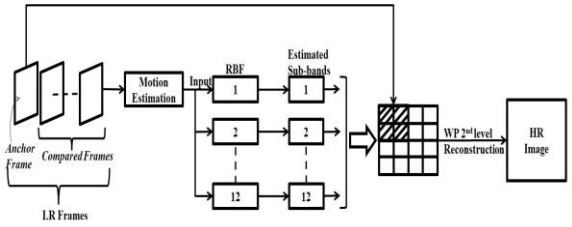

Figure 5. Interpolation process of the proposed method

It consists of the following steps:

1. Firstly, low-resolution frames to be processed are acquired.
2. The aforementioned motion estimation algorithm is employed as in training process to estimate motion between the input low-resolution frames and the vector array is collected as the input of neural networks.
3. The network is then simulated and the sub-band estimators estimate the 12sub-bands of the high frequency components of the wavelet packet decomposition. Instead of using four LL sub-bands, the anchor LR frame is employed because it has much more information than LL sub-bands to increase the quality of the super-resolved image.
4. Finally, the HR image is obtained by using inverse wavelet packet transform.

VI. RESULTS AND DISCUSSION

The test and analysis of the performance of the proposed video super-resolution is presented in real dataset. The real dataset are image sequences from the real video with different kind of motions. The proposed learning-based SR method is compared with the conventional super-resolution techniques using Keren image registration method with Iterated Back Projection (K_IBP) and Projection On Convex Sets (K_POCS) methods. Experiments have been carried out and tested in MATLAB R2018b.

In the training process, two real video sequences (Football and News) are employed. The number of frames in each of the video sequences is exactly 30 and the total number of frames is, therefore, 60. The video sequences are degraded and downscaled by a factor of 2 to produce LR version of the same sequences. The size of HR frames is 264×192 and the size of the LR frames is 132×96. Figure 6 shows the radial basis function neural network used in this work.

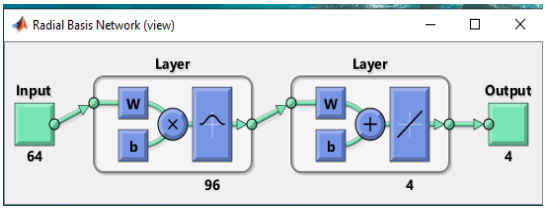

Figure 6. Radial basis function neural network used in the proposed method

For testing, five color image sequences of Akiyo, Flamingo, Ice, Car and Soda are used to evaluate the performance of the proposed method.

For performance evaluation, in addition to subjective visual quality judgments, the peak signal to noise ratio (PSNR) and mean structural similarity index (MSSIM) are used. These two image quality assessments are typical image quality measures. For an MxN image, PSNR can be calculated by using the following equations:

$$MSE=\frac{1}{MN}\sum_{i=1}^{M}\sum_{j=1}^{N}(x_{ij}-y_{ij})^2 \quad (7)$$

$$PSNR=10\log_{10}\frac{L^2}{MSE} \quad (8)$$

Where MxN is the size of the images and L is the maximum pixel value. PSNR is measured in decibels (dB).

Table 1. PSNR values for color real dataset

Methods / Images	K_IBP	K_POCS	WP_RBF
Akiyo	28.2467	25.4201	**29.6182**
Flamingo	26.2124	25.9722	**30.4577**
Ice	21.5829	24.2257	**28.4668**
Car	21.0037	22.7734	**24.4397**
Soda	21.2020	22.9380	**27.1815**

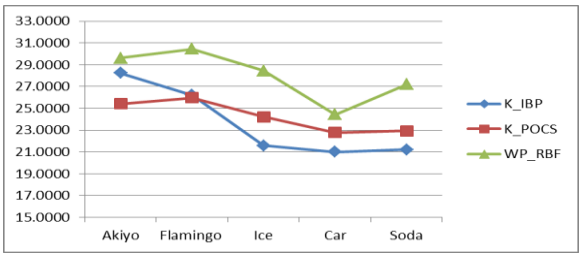

Figure 7. Comparison of PSNR values for color real dataset

Table 1 shows the PSNR values for color real dataset. From figure 7, the proposed method is better than the other two conventional methods in terms of PSNR. The highest PSNR value means that the quality of the image will be the best. Simultaneously, MSSIM index is calculated by using the following equations proposed in [16]:

$$SSIM(x,y)=\frac{(2\mu_x\mu_y+C_1)(2\sigma_{xy}+C_2)}{(\mu_x^2+\mu_y^2+C_1)(\sigma_x^2+\sigma_y^2+C_2)} \quad (9)$$

$$MSSIM(x,y)=\frac{1}{M}\sum_{j=1}^{M}SSIM(x_j,y_j) \quad (10)$$

Where x and y are the reference image and the resulted image, respectively. M is the total numbers of local window applied in the image.

Table 2. MSSIM values for color real dataset

Methods / Images	K_IBP	K_POCS	WP_RBF
Akiyo	0.8349	0.8395	**0.8670**
Flamingo	0.7665	0.7982	**0.8710**
Ice	0.6417	0.8184	**0.8615**
Car	0.5304	0.6702	**0.7048**
Soda	0.5147	0.6529	**0.7862**

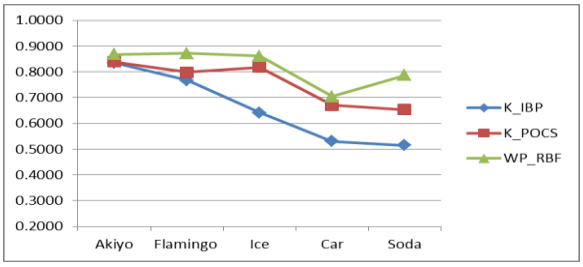

Figure 8. Comparison of MSSIM values for color real dataset

Table 2 describes the MSSIM values of color real dataset. As seen in figure 8, the proposed method has the highest values in terms of MSSIM. The higher the MSSIM value is, the better the quality of the image will be. The value of MSSIM ranges from 0 to 1.

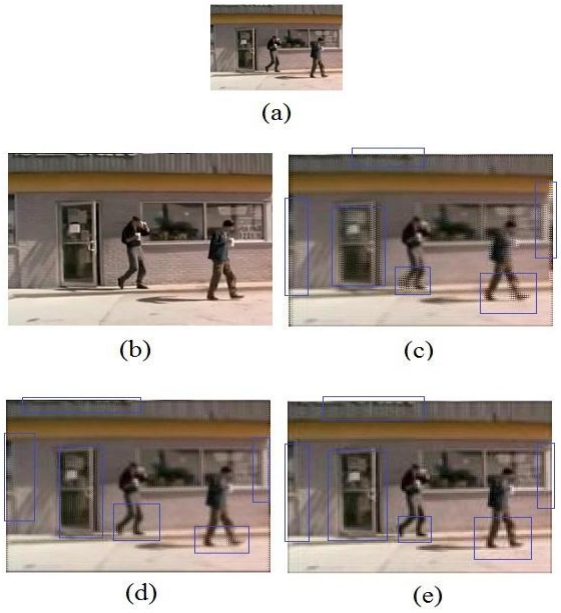

Figure 9. The 1st frame of Soda (a) LR image (b) Original image and super-resolved images using (c) K_IBP (d) K_POCS (e) WP_RBF

In order to assess visual quality, Soda is shown in figure 9. The regions of interest to be observed are marked with blue rectangles. It can be seen that the proposed method has better sharpness than the other methods.

VII. CONCLUSION

In this paper, video super-resolution image reconstruction methods have been examined. The goal is combining multiple low-resolution images from a set of video sequences to generate a single high-resolution image. It is observed that image registration or motion estimation is the most important step to determine the performance of super-resolution. Misaligned images may cause loss of detail or ghosts instead of improving the resolution. The purpose of this paper is to propose a novel video super-resolution method using local full search motion estimation algorithm, wavelet packet transform and RBF neural network. According to the experimental results, the proposed learning based SR method can improve PSNR and MSSIM values. Moreover, a better visual quality can be obtained in the super-resolved images.

REFERENCES

[1] R.Y. Tsai and T.S. Huang, "Multiple-frame image restoration and registration," Advances in Computer Vision and Image Processing. Greenwich, CT: JAI Press Inc., 1984, pp. 317-339.

[2] D. Keren, S. Peleg and R. Brada, "Image sequence enhancement using sub-pixel displacements," in Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pages 742–746 (June 1988).

[3] M. Irani and S. Peleg, "Improving resolution by image registration" CVGIP: Graphical Models and Image Processing, 53(3): 231–239 (May 1991).

[4] M. Irani and S. Peleg, Motion analysis for image enhancement: Resolution, occlusion and transparency. Journal of Visual Communications and Image Representation, 4: 324–335 (December 1993).

[5] D. Keren, S. Peleg and R. Brada, "Image sequence enhancement using sub-pixel displacements" in Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pages 742–746 (June 1988).

[6] H. Stark and P. Oskoui, "High-resolution image recovery from image-plane arrays, using convex projections," Jour. of Optical Society of America, vol. 6, no.11, pp. 1715-1726.

[7] Freeman, W.T., Pasztor, E.C., Carmichael, O.T.: "Learning Low-Level Vision," International Journal of Computer Vision 40(1), 25–47 (2000).

[8] Baker, S., Kanade, T.: "Limits on Super-Resolution and How to Break Them" IEEE Trans. Pattern Anal. and Machine Intell.24(9), 1167–1183 (2002).

[9] Freeman, W.T., Jones, T.R., Pasztor, E.C.: "Example-Based Super-Resolution" IEEE Computer Graphics and Applications 22(2), 56–65 (2002).

[10] J. C. Chan, J. Ma and F. Canters, "A comparison of super-resolution reconstruction methods for multi-angle CHRIS/Proba images," Proc. SPIE

Image and Signal Processing for Remote Sensing XIV, vol.7109, no. 1, Oct. 2008.

[11] Brian C. Tom, Aggelos T. Katsaggelos, "Resolution Enhancement of Video Sequences Using Motion Compensation".

[12] P. Vandewalle, S. S¨usstrunk, and M. Vetterli, "A frequency domain approach to registration of aliased images with application to super-resolution," accepted to EURASIP Journal on Applied Signal Processing, Special Issue on Super-Resolution Imaging, 2005.

[13] Wuqin Tong, Yongshun Ling, Chaochao Huang, Processing method of IR image based on mathematical morphology and wavelet transform[J], Opt. Precision Eng., 15(1), 2007, 138-144.

[14] M. Cocchi, R. Seeber, A. Ulrici et al., WPTER: Wavelet packet transform for efficient pattern recognition of signals [J], Chemo Metrics and Intelligent Laboratory Systems, 57(23), 2001, 97-119.

[15] Nyquist, H.: Certain topics in "Telegraph Transmission Theory" Trans. Amer. Inst. Elect. Eng. 47, 617–644 (1928).

[16] Z. Wang, A. C. Bovik, H. R. Sheikh and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," IEEE Trans. Image Processing, vol. 13, no. 4, pp. 600-612, 2004.